SEISMIC SPARSE LAYER INVERSION USING SIGNAL PROCESSING AND DEEP LEARNING TECHNIQUE

Supriyo Chakraborty

SEISMIC SPARSE LAYER INVERSION USING SIGNAL PROCESSING AND DEEP LEARNING TECHNIQUE

Thesis submitted to
Indian Institute of Technology Kharagpur
for the award of the degree

of

Doctor of Philosophy

by

Supriyo Chakraborty

under the supervision of

Aurobinda Routray

Professor

of

Department of Electrical Engineering

ADVANCED TECHNOLOGY DEVELOPMENT CENTRE
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
January 2024

-Dedicated to My Mother

CERTIFICATE OF APPROVAL

Certified that the thesis entitled **SEISMIC SPARSE LAYER INVERSION USING SIGNAL PROCESSING AND DEEP LEARNING TECHNIQUE**, submitted by **Mr. Supriyo Chakraborty** to the Indian Institute of Technology Kharagpur, for the award of the degree of **Doctor of Philosophy** has been accepted by the external examiners and that the student has successfully defended the thesis in today's viva voce examination.

Prof. Vikranth Racherla
(Chairman)

Prof. Aurobinda Routray
(Supervisor)

Prof. Mamata Jenamani
(Member of the DAC)

Prof. William K. Mohanty
(Member of the DAC)

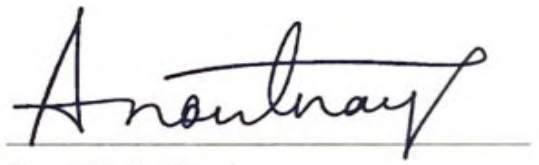
Prof. Pabitra Mitra
(Member of the DAC)

Prof Kalachand Sain (External Examiner)

Advanced Technology and Development Centre
Indian Institute of Technology Kharagpur
Kharagpur, India-721302

Certificate

This is to certify that this thesis entitled **SEISMIC SPARSE LAYER INVERSION USING SIGNAL PROCESSING AND DEEP LEARNING TECHNIQUE** submitted by **Mr. Supriyo Chakraborty**, to the Indian Institute of Technology Kharagpur, is a record of bona fide research work carried out under our supervision and is worthy of consideration for the award of **Doctor of Philosophy** of the Institute.



Aurobinda Routray
Professor

Department of Electrical Engineering
Indian Institute of Technology Kharagpur
India - 721302.

I.I.T. Kharagpur
July 2024

Declaration

I certify that

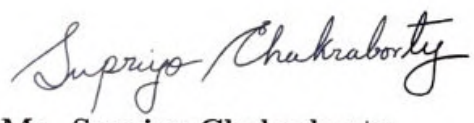
a. The work contained in the thesis is original and has been done by me under the guidance of my supervisor;

b. The work has not been submitted to any other institute for any other degree or diploma;

c. I have followed the guidelines provided by the Institute in preparing the thesis;

d. I have conformed to ethical norms and guidelines while writing the thesis;

e. Whenever I have used materials (data, models, figures and text) from other sources, I have given due credit to them by citing them in the text of the thesis, and giving their details in the references, and taken permission from the copyright owners of the sources, whenever necessary.



Mr. Supriyo Chakraborty

Acknowledgment

I extend my heartfelt gratitude to all those who supported and guided me throughout my research journey. First and foremost, I would like to thank my supervisor, Prof. Aurobinda Routray, for his invaluable guidance and insightful questions that helped me navigate the complexities of this research. His mentorship was pivotal to my growth in this field.

I am deeply grateful to Dr. Tarun Kanti Bhattacharya, Head of the Advanced Technological Development Center, for his unwavering support and encouragement. My sincere thanks also go to the DSC members, including Dr. Mamata Jenamani, Dr. William Kumar Mohanty, and Dr. Pabitra Mitra—for their expertise and valuable feedback, which significantly enhanced the clarity and quality of my work.

I acknowledge the Geodata Processing and Interpretation Centre (GEOPIC) at ONGC, Dehradun, for their cooperation and the resources they have contributed to my research. I am also thankful to my friends and colleagues at IIT Kharagpur for making this journey memorable.

Special thanks to my mother for her unwavering support and belief in my academic pursuits and to my HP-Tower workstation and laptop, which were essential for running the models in this research.

To everyone who has supported me, thank you for your kindness and encouragement.

-Supriyo Chakraborty

Abstract

Thin geological layers play a crucial role in seismic inversion, acting as key indicators for delineating stratigraphic sequences, identifying fluid contacts, and understanding reservoir compartmentalization. Accurate characterization of these thin layers is essential for subsurface reservoir exploration. However, sparse-layer seismic inversion presents significant challenges due to the ill-posed nature of estimating layer properties from seismic traces. The band-limited nature of seismic data and resolution constraints imposed by wavelet bandwidth often hinder the recovery of fine-scale structures, leading to ambiguities in identifying features below seismic resolution.

This thesis transitions from traditional signal processing methods to advanced deep learning frameworks, offering a robust approach to sparse-layer inversion. Initially, a dictionary learning-based multi-trace inversion technique is proposed, leveraging sparse representations to optimize basis functions while preserving structural and stratigraphic continuity. This method enhances vertical resolution and effectively reconstructs fine subsurface structures.

Building on this, a novel hybrid inversion framework introduces wedge-based group sparsity and higher-order total generalized variation (TGV), addressing the limitations of traditional sparsity models. By enforcing overlapping group sparsity, the framework improves lateral continuity and captures geological complexity with greater accuracy and computational efficiency.

To overcome the constraints of handcrafted models, the research progresses to deep learning-based approaches. A modified U-Net architecture is developed, incorporating multi-scale frequency-domain features and an unsupervised domain adaptation (UDA) strategy. This design enhances sparse representation across scales, enabling precise identification of thin geological layers.

Further advancements involve attention mechanisms tailored for seismic inversion,

including channel, spatial, and spectral attention modules. These mechanisms are integrated into an unsupervised domain-adaptive U-Net, prioritizing key features to significantly improve resolution and accuracy. Finally, a hybrid nested self-attention mechanism is introduced within a multi-head attention framework combined with a vision transformer (ViT). This architecture excels in isolating sparse features in seismic data, advancing thin-layer characterization, and achieving beyond seismic resolution in layer estimation.

Keywords: Seismic inversion, sparse-layer inversion, thin-layer identification, deep learning, U-Net, vision transformer, dictionary learning, group sparsity, unsupervised domain adaptation, attention mechanisms, multi-scale frequency analysis, neural networks.

Contents

List of Abbreviations

ML	Machine Learning
DL	Deep Learning
NN	Neural Network
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
LCI	Linear Constrained Inversion
ECA	Efficient Channel Attention
GAP	Global Average Pooling
GAN	Generative Adversarial Networks
CBAM	Convolutional Block Attention Module
Co.A	Coordinate Attention
CA	Channel Attention
SA	Spatial Attention
SE	Squeeze Excitation
CL	Combined Loss
CNN	Convolutional Neural Network
GPU	Graphics Processing Unit
TF	TensorFlow
BN	Batch Normalization
AG	Attention Gate
GN	Group Normalization
ReLU	Rectifier Linear Unit
SAA	Self Attention
VIT	Vision Transformer
TV	Total Variation
TGV	Total Generalized Variation
GCV	Gross Cross Validation

SVD	Singular Value Decomposition
k-SVD	Kernel Singular Value Decomposition
MP	Matching Pursuit
BPI	Basis Pursuit Inversion
ISTA	Iterative Shrinkage Thresholding Algorithms
FISTA	Fast Iterative Shrinkage-Thresholding Algorithm
ADMM	Alternating Direction Method of Multipliers
SBL	Sparse Bayesian learning
OMP	Orthogonal Matching Pursuit
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
FFT	Fast Fourier Transform
iFFT	Inverse Fourier Transform
DPSS	Discrete Prolate Spheroidal (Slepian) Sequences
DWT	Discrete Wavelet Transform
SSI	Sparse Spike Inversion
DicL	Dictionary Learning
HSI	Hyperspectral Imaging
MSE	Mean Squared Error
NMRSE	Normalized Root Mean Square Error
MAE	Mean Average Error
PSTM	Post Stack Migration
AI	Acquostic Impedance
3D	Three-Dimension
RC	Reservoir Characterization
PSNR	Peak signal-to-noise Ratio
SSIM	Structural Similarity Index

List of Figures

LIST OF FIGURES

List of Tables

CHAPTER 1

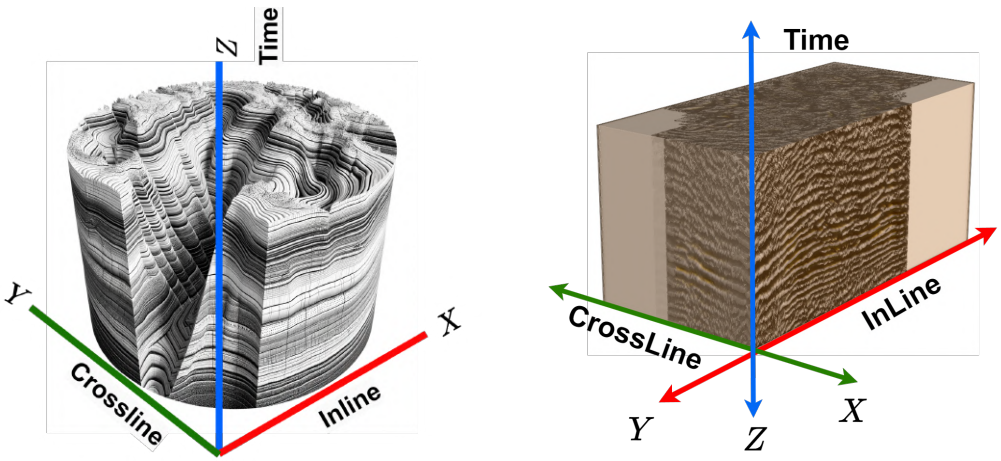
Introduction

1.1 Overview

Oil and gas exploration remains integral to global economic growth, underpinning energy security and national stability as nations pursue energy independence [6].

Although transitioning to renewables is essential, the global economy's dependency on oil and gas complicates this shift, requiring significant infrastructure investment [6]. The International Energy Agency (IEA) affirms that oil and gas will remain central to the energy landscape for the foreseeable future [7].

Hydrocarbons permeate daily life through agriculture, transportation, and manufacturing, sustaining the global focus on exploration. Figure 1.1 shows a representative $3D$ seismic section.



(a) 3D seismic volume illustration model (b) 3D real seismic data volume

Figure 1.1: Illustration of $X \times Y \times Z$ 3-D seismic data volume.

1.1.1 Seismic Exploration in Hydrocarbon Discovery: A Geophysical Perspective

Seismic exploration is the primary tool for hydrocarbon discovery, using reflected waves to map subsurface structures and constrain properties such as porosity, permeability, and fluid content [8–10]. Figure 1.2 illustrates the data acquisition workflow.

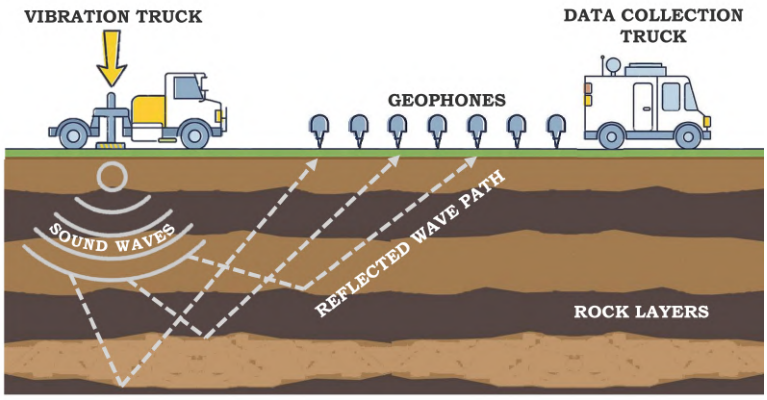


Figure 1.2: Illustration of seismic data acquisition.

The process spans acquisition, processing, imaging, interpretation, and analysis [11–14]. Noise and non-linearity remain significant obstacles; sparse layer inversion for

3D volumes is therefore central to improving resolution and exploration accuracy.

1.1.2 The Role of Thin-Layer Traps in Hydrocarbon Accumulation

Thin-layer traps form within narrow stratigraphic intervals where hydrocarbons accumulate in porous sandstone between impermeable shale layers. Their formation is controlled by lateral facies changes, pinch-outs, unconformities, and depositional environment [15, 16]; the integrity of the non-porous seal is critical for retention [17].

1.1.3 Challenges in Thin-Layer Trap Exploration and the Need for Innovation

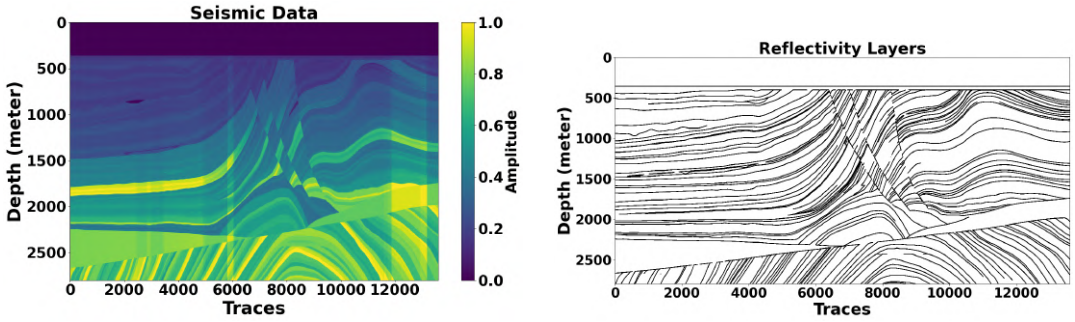
Seismic data is band-limited, with dominant frequencies of 30–50 Hz restricting layer visibility to roughly 15 m; layers thinner than this are difficult to detect. Noise and ghost layers further obscure true subsurface signals [18]. Traditional inversion—which converts time-series data into stratigraphic profiles—is hampered by the limited frequency spectrum and the ill-posed nature of the problem [19, 20]. Tuning thickness ($\lambda/4$), where distinct events merge, and the resolution limit ($\lambda/8$) add further complexity [21]. To address these challenges, this thesis explores blind inversion, compressive sensing, and machine learning (see Section 1.7), targeting noise suppression, ghost-layer mitigation, and improved thin-layer imaging [22, 23].

1.2 Data Familiarization

This thesis employs both real-world and synthetic seismic datasets to evaluate the proposed inversion techniques, with a primary focus on post-stack seismic data.

1.2.1 Post-stack Seismic data

Post-stack data combines individual seismic recordings into a single stack, substantially improving the signal-to-noise ratio [11]. This high-fidelity dataset underpins seismic inversion, reservoir characterization, and exploration decision-making [18, 24, 25].



(a) Marmousi velocity model (b) Marmousi Reflectivity model

Figure 1.3: Illustrating of Marmousi-2 model data (a) velocity and (b) reflectivity

1.2.2 Synthetic data

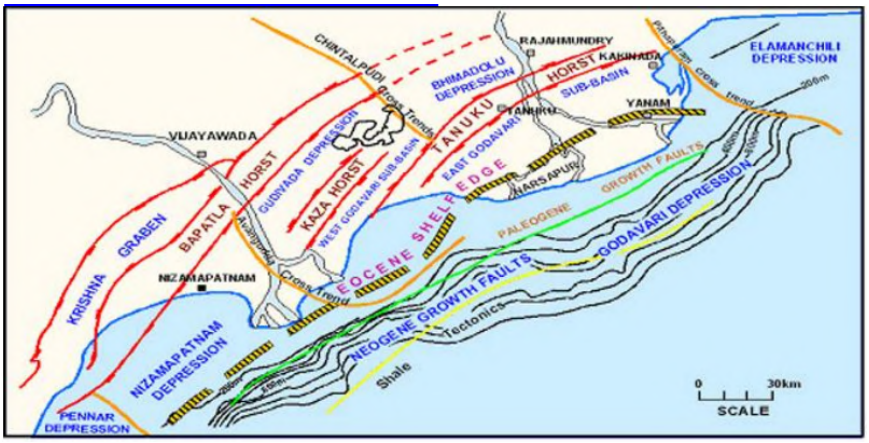
The Marmousi2 synthetic dataset is a crucial benchmark in this thesis, aiding in the refinement of seismic analysis techniques. Developed by the Institut Français du Pétrole (IFP) and the European Association of Geoscientists and Engineers (EAGE), it enhances the original Marmousi model to better represent geological complexities. Featuring elements like steep dips, fault zones, and varied sedimentary layers, Marmousi2 simulates realistic seismic data, including velocity and depth models. This makes it indispensable for advancing seismic imaging, inversion, and interpretation methods [26].

The dataset includes a 160-layer model with realistic velocity and density distributions capturing compaction in shallow detrital series; the final 2D velocity-density grid spans 9200 m × 3000 m at 4 m resolution (Figure 1.3) [27–29].

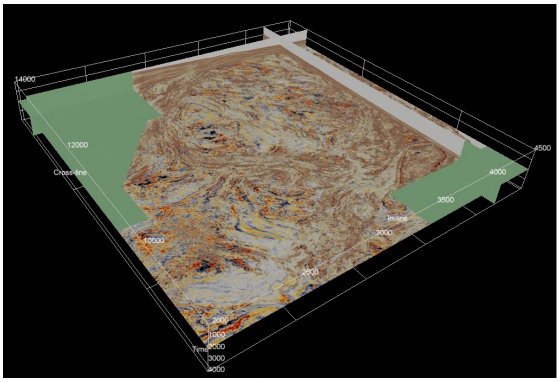
1.2.3 Real Data

The primary dataset for this research comes from the Krishna-Godavari (KG) basin on India's east coast, near the confluence of the Krishna and Godavari rivers [30]. This basin spans approximately 15,000 km^2 onshore and extends 25,000 km^2 offshore to the 1000 m isobath. The region, known for its significant hydrocarbon reserves, has been a key focus for the Oil and Natural Gas Corporation (ONGC), with 557 wells drilled over a 4220 km^2 area, identifying both onshore and offshore hydrocarbon deposits.

Seismic and well data for this study were provided by GEOPIC and ONGC. The dataset covers 150 km^2 with data from 12 wells. The seismic data includes Inline sections ranging from 70 to 515 and Cross-line sections from 40 to 1017, with a bin size of 17 m and 2-second data recorded at a sampling frequency of 512 Hz.



(a) Tectonic frame work of KG basin [31]



(b) 3-D seismic cross-section

Figure 1.4: KG basin data provided by GEOPIC/ONGC. Source: National Data Repository, India [1], [2].

1.3 Literature Review

This section traces the evolution of seismic inversion—from classical methods to modern machine-learning algorithms—and identifies remaining research gaps.

1.3.1 Traditional Algorithms in Seismic Inversion: A Review

Seismic inversion traces to Robinson's 1954 application of Wiener theory for statistical deconvolution [32]. **Lindseth's recursive inversion (1979)** linked acoustic impedance with reflection coefficients but was limited by the band-limited nature of seismic data [33].

Cooke and Schneider's generalized linear inversion (1983) iteratively refined acoustic impedance (AI) models, improving robustness with poor-quality data [34].

The 1990s brought **sparse inversion** via **Matching Pursuit** [35] and **Basis Pursuit** [36], followed by **Lancaster's coloured inversion (2000)** [37] and **Chen et al.'s atomic decomposition (2001)** [38], which improved resolution and sparsity.

The late 2000s saw **Nguyen's spectral decomposition** [39] and **Puryear's lateral instability regularization** [40] improve precision and stability. **Zhang and Castagna (2011)** employed a wedge-based dictionary with Basis Pursuit [41], extending this to multi-trace spatial regularization in 2013 [42]. Post-2013, **Hamid et al. (2015)** improved resolution with linear constrained inversion [5], and **Ming Ma (2017)** introduced block sparse Bayesian learning for multi-trace inversion [43].

Further advances include Wu et al.'s overlapping group sparsity (2020) [44], Yuan and Wang's tensor sparse coding (2021) [45], Hao et al.'s non-stationary inversion [46], Dai et al.'s nuclear norm constraints [47], and Zou et al.'s blind non-stationary inversion (2022) [48]. Liangsheng et al. (2022) [49], Yang et al. (2023) [50], and Li et al. (2024) [51] continued this trajectory. Ill-posedness and non-uniqueness remain open challenges; this thesis contributes improved time-frequency concentration and joint group sparsity with generalized regularization.

Table 1.1 summarizes key developments in seismic inversion from 1954 to 2024.

1.3.1.1 Critical Analysis of Methodological Limitations.

While the above developments demonstrate steady algorithmic progress, a closer examination reveals that most approaches share common underlying limitations arising from the fundamental characteristics of seismic data and the intrinsic structure of reflectivity. For clarity, these limitations are summarised in Table 1.2 and discussed

Author, Year	Technique
Robinson, 1954 [32]	Deconvolution with Wiener's theory.
Lindseth, 1979 [33]	Introduction to recursive inversion.
Cooke, Schneider, 1983 [34]	Generalized linear inversion.
Lancaster, 2000 [37]	'Coloured' inversion
Nguyen, 2008 [39]	Spectral decomposition and lateral instability regularization.
Puryear, 2008 [40]	Bayesian spectral decomposition
Heimer et al., 2009 [52]	Markov-Bernoulli random-field modeling for inversion.
Zhang, Castagna, 2011 [41]	Basis pursuit inversion with wedge dictionary.
Sen et al., 2013 [53]	Amplitude-versus-angle inversion with focus on regularization and solution sparsity.
Zhang et al., 2013 [42]	Multi-trace basis pursuit inversion with spatial regularization
Hamid et al., 2015 [5]	Linear Constrained multi-channel inversion for improved resolution.
Ming Ma et al., 2017 [43]	Multi-trace inversion Block sparse bayesian learning
Wu et al., 2020 [44]	Seismic inversion with 2^{nd} order overlapping group sparsity.
Yuan, Wang, 2021 [45]	3-d seismic noise attenuation with tensor sparse coding
Hao et al., 2021 [46]	Non-stationary inversion with the joint estimation of AI, AF and Source wavelet [a],[b].
Dai et al., 2021 [47]	Multi-trace sparse inversion with nuclear norm constraint
Zou et al., 2022 [48]	Blind Non-stationary inversion with Weighted TV Regularization.
Liangsheng et al., 2022 [49]	Amplitude-based re-weighted L1-Norm sparse constraint.
Yang et al., 2023 [50]	Fast reflection features constrained inversion.
Li et al,. 2024 [51]	Joint Laplace- and Fourier-Domain Seismic inversion

Table 1.1: Evolution of Seismic Inversion Techniques

[a]AI = Acoustic Impedance
[b]AF = Attenuation Factor

subsequently.

From a physical and signal-processing perspective, these limitations can be traced to three fundamental factors:

- **Band-limited acquisition:** Seismic data are inherently restricted to a finite frequency band (typically ∼10–80 Hz), where both low-frequency trends and high-frequency components are attenuated. Consequently, inversion methods cannot directly recover information that is not present in the recorded spectrum [40].
- **Mismatch between basis and reflectivity structure:** Thin-layer reflectivity is characterised by closely spaced dipole responses rather than isolated impulses. Conventional bases (sinusoidal or spike-based) fail to compactly represent such structures, leading to instability and reconstruction errors.
- **Non-stationary wavelet effects:** Real subsurface conditions introduce

Method Category	Underlying Assumption	Limitation in Thin-Layer / High-Frequency Regime
Spectral / STFT-based methods [39, 40]	Signal represented using globally supported sinusoids	Poor localisation below tuning thickness ($\lambda/4$); spectral interference leads to instability and degradation of lateral continuity
Sparse-spike inversion [35, 36]	Reflectivity is sparse in impulse domain	Thin-layer responses exhibit structured dipole (doublet) behaviour rather than isolated spikes, violating sparsity assumptions and leading to systematic under-recovery
Wedge-based inversion (BPI) [41]	Dipole atoms constructed using a fixed wavelet	Assumes wavelet stationarity; performance degrades under depth-dependent attenuation and non-stationary propagation effects [46, 48]
Data-driven dictionary learning [54]	Learned atoms capture statistical structure of observed data	Limited by band-limited training data; incapable of reconstructing missing high-frequency components and lacks optimal time-frequency concentration [55, 56]

Table 1.2: Critical Analysis of Basis Function Limitations in Thin-Layer and High-Frequency Seismic Inversion

depth-dependent attenuation and phase variation, violating the stationary wavelet assumption commonly used in inversion frameworks [46, 48].

Despite these significant advancements, challenges remain in seismic inversion, particularly in addressing ill-posedness, improving solution uniqueness, and accurately resolving sparse high-frequency reflectivity. The design of basis functions with superior time-frequency concentration, along with structured sparsity constraints, remains relatively under-explored and presents a critical opportunity for further research.

This thesis seeks to address these gaps by introducing basis functions with improved time-frequency concentration and incorporating joint group sparsity with generalized regularization frameworks.

1.3.2 Machine Learning Algorithms in Seismic Inversion: A Review

Machine learning has transformed seismic inversion by enabling direct extraction of subsurface properties. Early work used **Fully Connected Neural Networks (FCNNs)** for velocity, impedance, and density estimation [57, 58] and **Probabilistic Neural Networks (PNNs)**, which applied Bayesian decision theory to model nonlinear relationships [59, 60]. **Convolutional Neural Networks (CNNs)** became

the dominant approach, applied to velocity modeling, impedance prediction, and density estimation [61–65]. End-to-end models such as **InversionNet** and **SeisInvNet** enable direct seismic-to-velocity mapping [66, 67], while **ResNet**, **FCNs**, **U-Net**, and **Temporal Convolutional Networks (TCNs)** have further extended CNN capabilities [68, 69, 71, 73].

U-Net's encoder-decoder structure captures both local and global features, making it effective for seismic segmentation and property estimation [72, 74, 75]. Min et al. (2023) further improved super-resolution by exploiting edge information via a dual-decoder U-Net [76].

Hybrid **CNN–RNN** models using **LSTM** and **GRU** units capture temporal dependencies in seismic traces and have shown strong potential in inversion tasks [68,77–81].

Physics-Informed Neural Networks (PINNs) embed physical constraints directly into the learning process, improving interpretability and stability over purely data-driven models [82–85].

Generative Adversarial Networks (GANs)—including Wasserstein, Conditional, and Cycle variants—address sample diversity and reduce labeled-data requirements through cyclic inversion–forward structures [86–93].

Table 1.3 summarizes deep learning techniques for seismic inversion from 2018 to 2024.

Despite these advances, a key gap persists: handling sparse, band-limited layers. Labeled real data is scarce, and training commonly relies on well-log interpolation, introducing bias.

This thesis addresses this gap with deep-learning methods that boost SNR for high-frequency components and employ **unsupervised domain adaptation (UDA)** to avoid reliance on labeled data.

1.3.2.1 Transformer Architecture

Introduced for NLP in 2017 [117], the Transformer's self-attention mechanism quickly proved valuable in computer vision through models such as ViT [120], DETR [119],

Author, Year	Technique Used	Description
Zong et al., 2018 [94]	Bayesian inversion in complex frequency domain	Broadband complex frequency domain for estimating the low-frequency component.
Wu & Lin, 2018 [95]	CNN with CRF	Using CNN with CRF for FWI.
Das et al., 2019 [96]	CNN	CNN-based seismic impedance inversion with Bayesian methods for uncertainty assessment.
Alfarraj & Alregib, 2019 [97]	RNN	Semisupervised-RNN for elastic impedance inversion from multiangle seismic data.
Li et al., 2020 [98]	DNNs	SeisInvNets improving upon traditional iterative algorithms with DL.
Ren et al., 2020 [99]	Physics-based neural network (SWINet)	NN and wave-based modeling for velocity inversion in seismic waveform analysis.
Wang et al., 2020 [100]	Closed-loop CNN	Closed-loop CNN structure for simultaneous seismic forward and inversion modeling.
Zhang et al., 2020 [101]	Data-driven seismic waveform inversion	Robust data-driven seismic waveform inversion using CNN.
Adler et al., 2021 [102]	DL framework	DL review on seismic inversion.
Lin et al., 2022 [103]	Physics-informed data-driven inversion	Utilizing physics loss in DL in seismic inversion.
Liu et al., 2022 [104]	Transfer learning	STF network with transfer learning for seismic sparse time-frequency analysis.
Ma et al., 2022 [105]	UB-Net	DL with Uncertainty backpropagation.
Wang et al., 2022 [106]	2D-CNNs	2D-CNNs with domain adaptation.
Li et al., 2022 [107]	Hybrid network (AG-ResUnet)	Hybrid network with attention mechanisms.
Barman et al., 2022 [54]	Dictionary learning	Dictionary learning with CNN and Variational Auto-Encoder (VAE).
Zhang et al., 2022 [108]	Multi-scale architecture with GANs	Multi-scale architecture with GANs to improve high-frequency features.
Manu et al., 2022 [109]	Edge based inversion	DCNs on Edge devices.
Zhang et al., 2022 [108]	Reflectivity method	Data-driven elastic parameter prediction using the reflectivity method for training dataset construction.
Ning et al., 2023 [110]	Attention U-Net	Attention U-Net to improve lateral continuity.
Xie et al., 2023 [111]	2D CNN with coordinate attention	2D CNN with coordinate attention block and hybrid loss function.
Zhang et al., 2023 [84]	PINN	PINN based on acoustic wave equations.
Min et al., 2023 [76]	Double-UNet	Use low-resolution image and the edge image to improve output image resolution
Dodda et al., 2023 [112]	ACNN	Attention-based Deep Convolutional Neural Network (ACNN) using CWT.
Liu et al., 2023 [113]	FAUNet	Multitask Full Attention U-Net (FAUNet) for prestack with a hybrid loss function.
Gelboim et al., 2023 [114]	Encoder-Decoder	Encoder-Decoder architecture for 3D seismic inversion with shot-gather cubes.
Bürkle et al., 2023 [115]	Physics-aware stochastic inversion	CNNs with geostatistical simulation.
Li et al., 2024 [116]	MAU-net	Multibranch attention on Unet.

Table 1.3: Deep Learning (DL) Techniques in Seismic Inversion

and Swin-Transformer [121]. Transformers have been adopted for seismic inversion due to their ability to model long-range dependencies. Notable applications include **StorSeismic** (BERT-based self-supervised pre-training) [122], a hybrid CNN-Transformer for impedance inversion [123], a positional-embedding model for geoacoustic inversion [124], a multi-scale Transformer for interpolation [125], and Swin-Transformer-based denoising [127].

Key challenges are high computational cost and data-intensity; hybrid CNN-Transformer models [123, 125] and self-supervised models such as **SeisMAE** [128] partially address these. Li et al. demonstrated Transformer scalability for large-scale 3D seismic inversion [129].

Table 1.4 summarizes Transformer-based seismic inversion methods.

Despite these advances, attention mechanisms have not been optimized for the seismic domain. Pre-trained vision Transformers cannot transfer directly due to band-limited data, sparse high-frequency reflectivity, and the demands of geological interpretation.

1.4 Analytical Synthesis of Seismic Inversion Methodologies

Building on the preceding literature review, seismic inversion methodologies can be broadly categorized into three paradigms: *model-driven*, *data-driven*, and *physics-informed* approaches, with Transformer-based models extending data-driven frameworks. **Model-driven approaches** provide physically interpretable solutions and structured representations through sparse inversion and dictionary learning. However, their performance is strongly dependent on the choice of basis functions, which are often inadequate for representing thin-layer reflectivity characterized by dipole-like responses, leading to instability in high-resolution reconstruction. **Data-driven approaches**, particularly CNN-based models, enable efficient end-to-end inversion and are effective in capturing local spatial patterns. However, they rely heavily on the quality of training data and often learn biased mappings due to limited, approximate ground-truth labels (e.g., well-log interpolation). Additionally, their limited recep-

Author, Year	Technique Used for
Harsuko et al. 2022 [130]	Data driven using BERT for feature extraction and storage in seismic data processing.
Feng et al. 2023 [131]	Transformer-based model for geoacoustic inversion with enhanced data perception through positional embeddings.
Fu et al. 2023 [132]	Hybrid CNN and Transformer model for seismic impedance inversion focusing on local and global data features.
Guo et al. 2023 [125]	Multi-scale Transformer for capturing long-range dependencies in seismic data to improve the accuracy of interpolation
Li et al. 2023 [133]	Vision Transformer combined with Masked Autoencoders for building subsurface models using seismic and well data.
Ning et al. 2023 [110]	Hybrid Transformer and CNN architecture for enhanced seismic impedance inversion with self-supervised learning elements.
Su et al. 2023 [134]	Transformer-based multi-attribute inversion method for porosity prediction in tight reservoirs.
Dou et al. 2024 [128]	Seismic Mask Auto Encoder (SeisMAE) utilize Transformers for 3D seismic data inversion with novel feature representation strategies.
Gao et al. 2024 [135]	Swin-Transformer combined with convolutional residual networks for simultaneous denoising and interpolation of seismic data.
Liu et al. 2024 [136]	Ensemble Dense Inception Transformer for accurate seismic horizon interpretation using multiple data attributes.
Ning et al. 2024 [137]	Improved Transformer and CNN hybrid network for P-wave impedance inversion with enhanced transfer learning capabilities.
Wang et al. 2024 [138]	Self-Supervised Pre-training Transformer for effective seismic data denoising and signal reconstruction.
Li et al. 2024 [129]	Data-driven approach using a Transformer for 3D seismic inversion focusing on self-attention mechanisms.
Yang et al. 2025 [139]	HCTNet: Hybrid 1D-CNN and Transformer for robust prestack AVA elastic parameter inversion using multitask learning and synthetic-to-field transfer.
Pang et al. 2025 [140]	Iterative gradient-corrected semisupervised seismic impedance inversion via Swin Transformer (USTNet), reducing non-uniqueness through progressive model-space refinement.
Li et al. 2025 [141]	FuTE-FWI: Fusion Vision Transformer Enhanced network for full waveform inversion on deep and complex strata using multiheaded self-attention between CNN encoder and velocity reconstructor.

Table 1.4: Transformer-Based Techniques in Seismic Inversion

tive field restricts long-range dependency modeling, affecting lateral continuity and global geological consistency. **Transformer-based models** address this limitation by capturing global dependencies through self-attention mechanisms and enabling hybrid local–global feature learning. Nevertheless, they are computationally expensive, data-intensive, and not inherently adapted to the sparse, band-limited nature of seismic reflectivity. **Physics-informed approaches** incorporate governing physical

laws into the learning process, improving interpretability and reducing dependence on labeled data. However, they suffer from slow convergence, sensitivity to loss balancing, and limited scalability for large 3D seismic volumes. **Unified Limitation Across Paradigms:**

Despite differences in modeling approaches, all seismic inversion methods share a common set of fundamental limitations:

- **Band-limited data:** Missing high-frequency components restrict recovery of fine-scale geological features across both model-driven and data-driven methods.
- **Representation mismatch:** Thin-layer reflectivity exhibits dipole behaviour, which is not efficiently captured by conventional bases or learned representations.
- **Non-stationary wavelet effects:** Depth-dependent attenuation and phase variations violate modelling assumptions and reduce generalization.
- **Limited supervision:** Scarcity and bias in ground-truth labels constrain the performance of data-driven and Transformer-based approaches.

These limitations collectively hinder accurate reconstruction of sparse high-frequency reflectivity and motivate the research directions addressed in this thesis.

1.5 Research Issues

Seismic inversion presents several unresolved challenges that underpin this thesis:

1. **Enhancing Resolution in Band-Limited Seismic Data:** Traditional inversion methods fail to recover high-frequency components, producing poor thin-bed resolution and smoothing geological features near the tuning thickness ($\lambda/4$) [142].
2. **Mitigating Ghost Layer Effects:** High-frequency attenuation introduces ghost layers that distort subsurface imaging and cause misinterpretations [143, 144].

3. **Optimizing Basis Functions and Dictionary Learning:** Existing methods lack effective basis functions for time-frequency energy concentration, limiting fine-grained seismic representation [55, 56, 145].
4. **Integrating Machine Learning for Sparse Layer Resolution:** The combination of machine learning and frequency-domain methods for sparse layer resolution remains underdeveloped [146].
5. **Overcoming Data Scarcity with Self-Supervised Learning:** Limited labelled seismic data constrains inversion performance; self-supervised and unsupervised domain adaptation methods are needed for data-scarce settings.
6. **Adapting Attention Mechanisms for Seismic Inversion:** Attention mechanisms, including Vision Transformers (ViTs), show promise for high-frequency recovery and thin-bed resolution [128–130], but their adaptation to seismic inversion remains open.

Improving seismic resolution is the core focus of this thesis, with the proposed methodologies targeting advances in subsurface imaging, geological interpretation, and hydrocarbon exploration.

1.6 Objectives

This thesis designs novel seismic inversion techniques addressing lateral continuity, sparse representation, and thin-layer resolution:

1. **Dictionary Learning for Multi-Trace Seismic Inversion:** Integrate Discrete Prolate Spheroidal Sequences (DPSS) with lateral constraints to improve subsurface characterization and spatial continuity.
2. **Basis Function Optimization for Sparse Representation:** Optimize dictionary basis vectors for time-frequency concentration of seismic wavelets, increasing SNR and mitigating high-frequency attenuation and ghost layers.
3. **Sparse Coding with Joint Group Overlapping Sparsity:** Introduce a sparse coding algorithm combining joint group overlapping sparsity with Higher-Order Total Generalized Variation (TGV) regularization to address the ill-posed

nature of seismic inversion.

4. **U-Net and Vision Transformer (ViT) Frameworks for Sparse Layer Resolution:**
   - **Multi-Scale Frequency Domain Transforms:** Enhance U-Net with multi-scale frequency domain transformations for improved seismic feature representation.
   - **Spectral Attention Mechanisms:** Implement spectral channel attention networks (CBAM, SENet, DFT-based, DWT-based, DPSS-based) within U-Net.
   - **Custom Attention for ViT:** Develop a multi-head attention mechanism tailored for ViT to resolve sparse and thin seismic layers.
5. **Unsupervised Domain Adaptation (UDA):** Train deep-learning models on unlabelled seismic data using UDA to address labelled-data scarcity and improve generalization.

1.7 Contributions of this Thesis

This thesis advances seismic inversion through physically consistent and data-driven representations that overcome limitations of conventional basis functions under band-limited and non-stationary conditions:

1. **Physics-consistent dictionary learning with lateral constraints:** Two multi-trace inversion frameworks integrate wedge-based representations and DPSS within dictionary learning, enforcing spatial continuity across traces for improved geological consistency.
2. **Structured sparsity with higher-order regularization:** Group sparsity combined with TGV regularization captures structured thin-layer dipole responses, overcoming the sparse-spike assumption.
3. **Multi-scale frequency-domain learning for seismic inversion:** A frequency-aware U-Net with multi-scale frequency-domain transformations represents band-limited data more faithfully; UDA improves generalization across

geological settings.

4. **Spectral attention mechanisms for frequency-aware inversion:** A unified spectral attention framework combines attention gates, CBAM, and SENet with Fourier (DFTNet), wavelet (DWTNet), and DPSS-based modules for adaptive spatial–frequency feature selection.

5. **Attention-Seisformer: A hybrid transformer for seismic inversion:** *Attention-Seisformer* integrates multi-head self-attention with ECSAN and DWT-based spectral attention, modelling long-range dependencies while preserving spectral and spatial fidelity.

1.8 Organization of the Thesis

- **Chapter 2: Dictionary Learning-based Multi-Trace Inversion.** Develops a dictionary learning framework integrating wedge-based and DPSS representations for multi-trace inversion, targeting optimal sparse basis selection and lateral continuity.

- **Chapter 3: Joint Group Overlapping Sparsity with Wedge Group.** Introduces a sparse coding algorithm combining group sparsity with higher-order TGV, improving reconstruction and resolution of sparse seismic layers.

- **Chapter 4: Multi-Scale Frequency Domain Features in U-Net.** Incorporates MFDT into the U-Net architecture to achieve near-orthogonal frequency-band separation, enhancing seismic feature capture across scales and improving inversion resolution.

- **Chapter 5: Enhancing U-Net Seismic Inversion with Spectral Channel Attention.** Investigates attention networks (attention gates, CBAM, SENet, DFT-Net, DWT-Net, DPSS-Net) within UDA-based U-Net; evaluates the accuracy– efficiency trade-off and demonstrates improved super-resolution of seismic traces.

- **Chapter 6: Attention-Seisformer — A Hybrid Multi-Head Attention Transformer for Seismic Inversion.** Presents a hybrid ViT integrating mul-

tiple attention mechanisms for improved resolution of sparse and thin seismic layers.

- **Chapter 7: Conclusion and Future Scope.** Summarises key findings and outlines directions for future research.

Figure 1.5 outlines the thesis organization, categorizing seismic inversion into model-based and data-driven approaches. The model-based approach is implemented using innovative techniques, such as novel dictionary learning and group sparsity methods. In contrast, the data-driven approach leverages advanced machine learning techniques, including a U-Net-based attention network and a Transformer-based hybrid attention network.

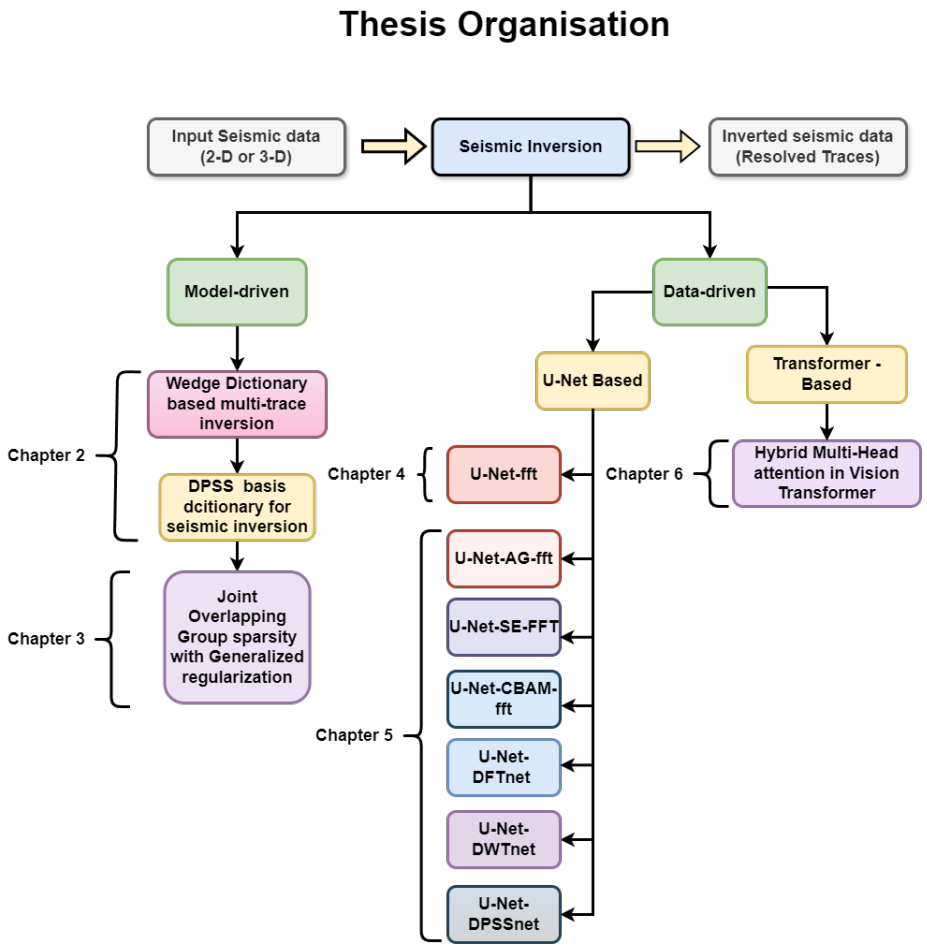


Figure 1.5: Thesis Organisation

CHAPTER 2

Dictionary learning-based multi-trace seismic inversion

This chapter examines dictionary learning techniques for seismic inversion, with emphasis on sparse signal extraction. By optimizing signal representation with minimal information loss, these methods enable detailed subsurface imaging. Two contributions are introduced:

1. A wedge dictionary combined with dictionary learning for multi-trace inversion, improving accuracy and efficiency in seismic data analysis.
2. Integration of the discrete prolate Slepian sequence (DPSS) basis into multi-trace inversion dictionary learning, leveraging its time-frequency concentration to improve signal recovery and resolution.

2.1 Brief Literature Review

Seismic inversion delineates subsurface geological structures essential to oil and gas exploration; accurately resolving sparse thin layers remains a central challenge. The

Widess model [21] demonstrated that reflectors at the top and base of a wedge become distinguishable when the formation thickness exceeds $\lambda/4$. This principle underlies thin-layer identification through the tuning effect within the range of $\lambda/4$ and $\lambda/8$, where λ denotes the seismic wavelength. Multiples and noise, however, make vertical resolution difficult and effectively set $\lambda/4$ as the practical resolution limit [147, 148].

Dictionary learning (DicL) has emerged as a powerful tool for interpreting subsurface geological structures. Traditional methods—sparse spike inversion [149], spectral inversion [40, 150, 151], and sparse-representation-based reflectivity inversion [41, 152]—provided the foundation; K-singular value decomposition (K-SVD) [153] then marked a significant advance in inversion capability.

DicL extends SR-based methods—including sparse-layer inversion with basis pursuit using a wedge dictionary [41] and multi-trace inversion with spatial regularization [53, 154]—all rooted in the matching pursuit algorithm [35], which established DicL's ability to overcome resolution and noise limitations.

DicL also spans broader image processing tasks—image fusion, super-resolution, and classification [155–158]—and has found use in seismic interpolation, attribute extraction, and full inversion [159–164]. Exploiting spatial coherence across traces—often neglected in 1D analysis [5, 165–167]—further improves resolution and accuracy.

This chapter introduces wedge-based dictionary learning for multi-trace inversion and integrates the discrete prolate Slepian sequence (DPSS) basis into a dictionary learning framework to improve time-frequency concentration in seismic inversion. DPSS, grounded in the work of Slepian, Landau, and Pollak [168–172], employs prolate spheroidal wave functions (PSWFs) valued for their optimal energy concentration and orthogonality. Their demonstrated effectiveness across radio astronomy [173], ship dynamics [174, 175], beam-forming [176], and MRI [177] establishes PSWFs as reliable tools for reducing artefacts and improving resolution.

2.2 Methodology

This section details dictionary-learning-based seismic inversion methods for sparse representation of seismic trace data, covering forward and inverse modeling.

2.2.1 Forward Modeling

Forward modeling simulates how a seismic wavelet interacts with subsurface layers, producing a synthetic trace $s(t)$ by convolving wavelet $w(t)$ with reflectivity series $r(t)$—which represents geological-layer interfaces—and adding acquisition noise $n(t)$:

$$s(t) = w(t) * r(t) + n(t) \tag{2.1}$$

Equation 2.1 generates a band-limited trace combining the wavelet's frequency content with the subsurface reflection characteristics.

The inversion begins with seismic and well data. Low-frequency P-impedance logs initialize a horizontally layered model that is spatially interpolated across the survey. This model is band-limited to 10–15 Hz (Figure 2.1), the range essential for capturing broad geological features. Synthetic traces generated from this model are compared with field records to iteratively refine the subsurface impedance model.

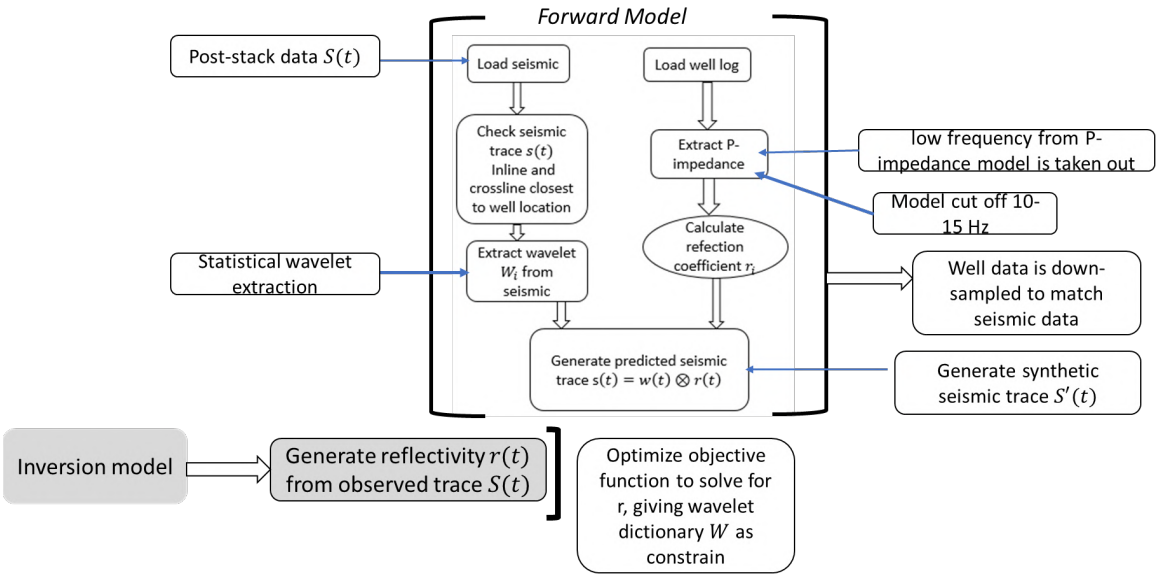


Figure 2.1: Forward model

As shown in Figure 2.2, seismic inversion iteratively refines the subsurface impedance model to match observed data. Measured data $S(t)$ and a synthetic trace $S'(t)$—computed by convolving well-log data with a modeled wavelet—are compared via

error function $e = \text{Fun}(S(t) \sim S'(t))$. This error drives parameter updates in a feedback loop until the synthetic trace converges to the observed record. Anisotropic assumptions account for subsurface heterogeneity, and well-log P-impedance data constrain the low-frequency model components.

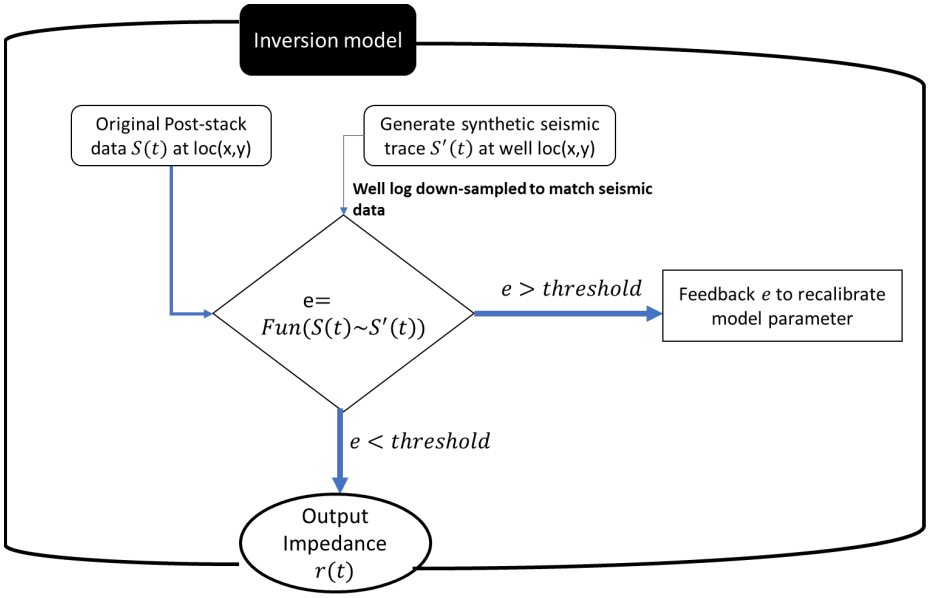


Figure 2.2: Inversion model

2.2.2 Dictionary Learning

Dictionary learning via K-SVD produces sparse representations of seismic data. An overcomplete dictionary D is learned by minimizing reconstruction error $\|S - Dr\|_F^2$ subject to the sparsity constraint $\|r\|_0 \leq T$. K-SVD iterates between sparse coding—finding the most compact representation r for fixed D—and a dictionary update step that refines each atom d_j by applying SVD to the residual matrix θ_j [178].

To update d_j, the residual $\theta_j = S - \sum_{i \neq j} d_i r_i^T$ is formed and decomposed by SVD; the atom and its coefficients are then set to the leading singular vector and value, respectively. Iterating until convergence yields a compact, informative dictionary for representing the seismic trace S (Figure 2.3).

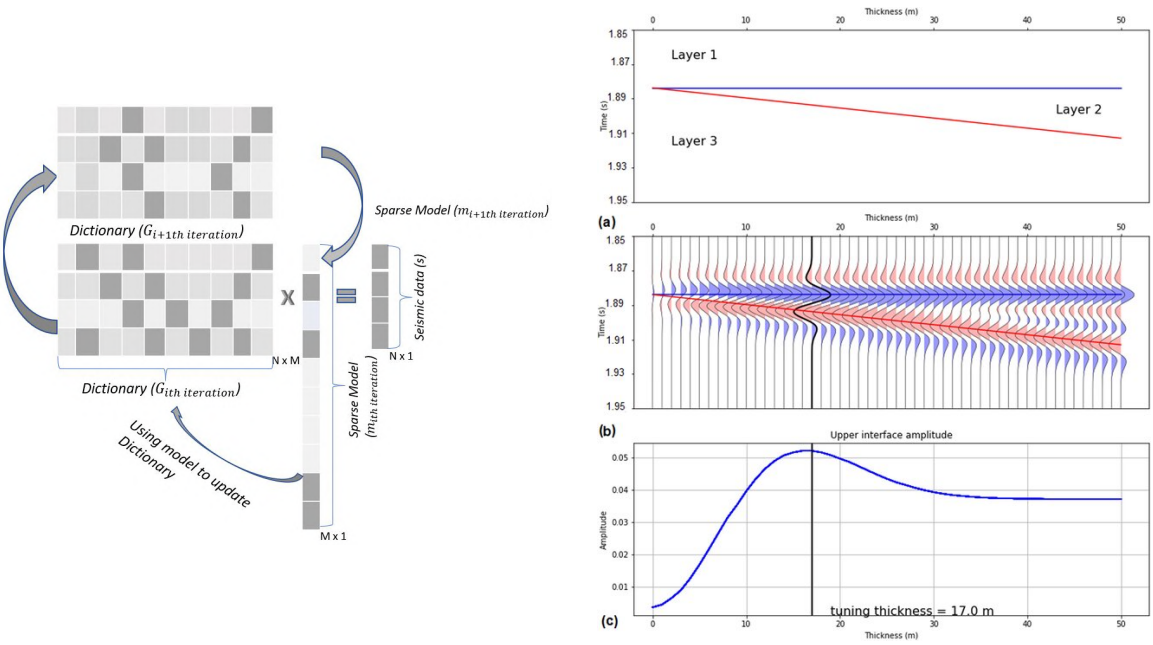


Figure 2.4: Tuning thickness and seismic resolution

2.2.3 Multi-Trace Wedge Dictionary-Based Inversion

In sparse-layer seismic inversion [4], tuning thickness is central to resolution. Tuning thickness equals $\lambda/4$—one quarter of the predominant wavelength—where constructive interference merges the top and base reflections into a single amplified event. Layers thinner than this limit cannot produce distinguishable reflections.

The wedge model [21] assesses the seismic response across varying layer thicknesses. Its wedge-shaped geometry simulates the transition from individually resolvable reflections to a tuned composite response: above the tuning thickness, top and base reflections are distinct; below it, they merge into a single event whose amplitude then diminishes by destructive interference.

This behavior is captured in the reflectivity series (Equation 2.2):

$$\begin{aligned} r_e &= \delta(t) + \delta(t + n\Delta t) \\ r_o &= \delta(t) - \delta(t + n\Delta t) \\ a\delta(t) + b\delta(t + n\Delta t) &= cr_e + dr_o \end{aligned} \tag{2.2}$$

Thin-bed reflectivity decomposes into impulse pairs at the top ($a\delta(t)$) and base ($b\delta(t + n\Delta t)$), where a and b are reflection coefficients, $n\Delta t$ is the two-way travel time thickness, and Δt is the sampling interval. The even (r_e) and odd (r_o) dipole pair decomposition separates constructive from destructive interference; coefficients c and d give the amplitude of each pair, enabling reconstruction of the original reflectivity. This framework resolves sub-tuning subsurface features otherwise obscured by the seismic resolution limit (Figure 2.4).

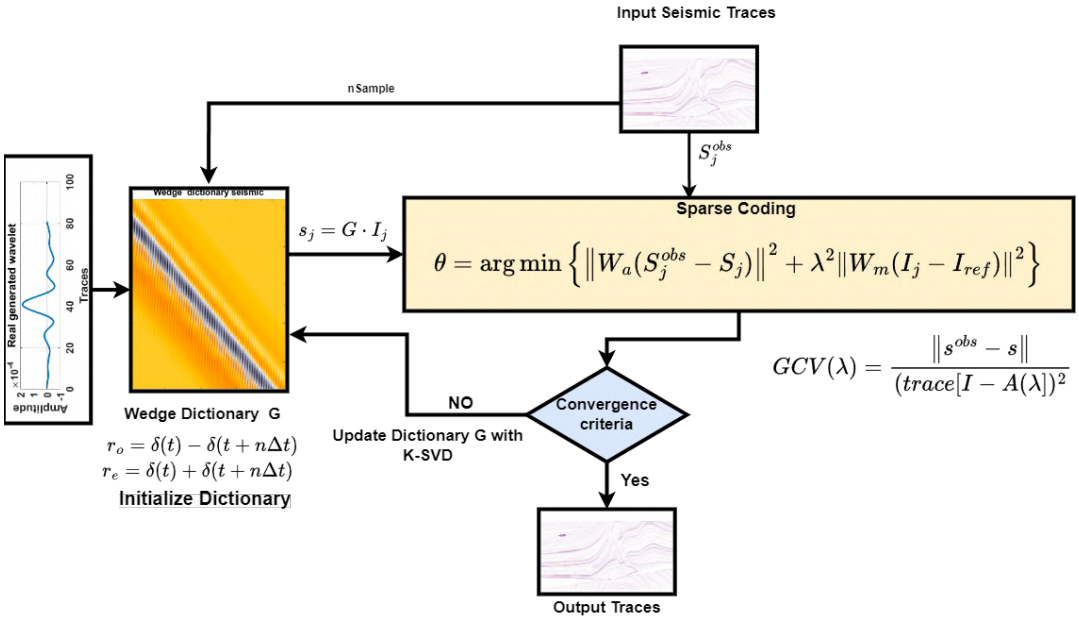


Figure 2.5: Wedge-based Dictionary Learning Multi-Trace Inversion (wedge-LCI)

Equation 2.2 expresses total reflectivity as a sum of even-odd pairs:

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_i \\ r_n \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & 0 & -1 & \ddots & 0 \\ \ddots & \vdots & \vdots & \ddots & 1 \\ 0 & \cdots & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} r_{e1} \\ r_{o1} \\ \vdots \\ r_{oi} \\ r_n \end{bmatrix}$$

From Equation 2.1, the seismic trace S is obtained by convolving wavelet matrix G

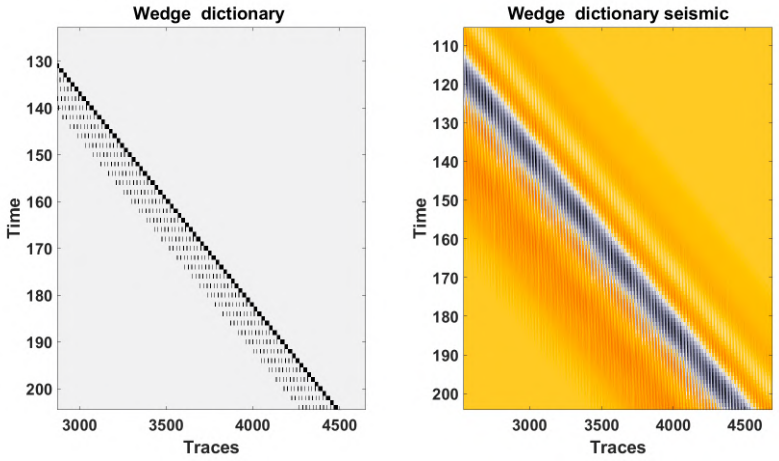


Figure 2.6: wedge dictionary

with reflectivity vector I (Equation 2.3):

$$S = G \cdot I \tag{2.3}$$

Where:

$$S = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_i \\ t_n \end{bmatrix}, G = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ \vdots & w_1 & \ddots & \vdots \\ w_k & \vdots & \ddots & 0 \\ 0 & w_k & \vdots & w_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_k \end{bmatrix}, I = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} r_{e1} \\ r_{o1} \\ \vdots \\ r_{oi} \\ r_n \end{bmatrix}$$

A laterally constrained approach solves jointly for all traces using a wedge dictionary convolved with wavelets (Equation 2.3). The forward operator G_j encodes both dictionary and wavelet information; s_j is the seismic trace vector grouping m traces, I_i is the impedance vector, $\overline{G}$ is the block-diagonal assembly of G_j, and L is the 1D model group vector.

Reflectivity is recovered by minimizing the weighted misfit between observed and

predicted traces with model regularization (Equation 2.4):

$$\theta = \arg\min \left\{ \left\| W_a(S_j^{obs} - S_j) \right\|^2 + \lambda^2 \left\| W_m(I_j - I_{ref}) \right\|^2 \right\} \tag{2.4}$$

Here, W_a and W_m represent weighting matrices, S_j^{obs} is the observed seismic trace, S_j is the predicted seismic trace, I_j is the impedance model, and I_{ref} is the reference impedance model. The regularization parameter λ controls the trade-off between data misfit and model regularization.

The solution I_i satisfies the normal equations:

$$\begin{aligned} A &= \overline{G}^T \overline{W} d^T \overline{W} dG + \lambda^2 (W_m^T W_m) \\ b &= \overline{G}^T \overline{W} d^T \overline{W} dS_j^{\text{obs}} + \lambda^2 (W_m^T W_m I_{\text{ref}}) \end{aligned} \tag{2.5}$$

where W_d and W_m are diagonal matrices whose entries are $1/\sigma_{d_i}$ and $1/\sigma_{m_i}$, respectively.

Here σ_d and σ_m are data and model standard deviations, S_{obs} is the observed trace from Equation 2.3, S is the estimated trace, and I_{ref} is the reference impedance model.

The inversion retains only the high-frequency impedance component; low-frequency well-log trends are reintroduced after inversion via a 10–15 Hz cut-off filter.

The regularization parameter λ is selected by generalized cross-validation (GCV), which minimizes the L_2 misfit between observed and predicted traces normalized by a scalar (Equation 2.6):

$$\text{GCV}(\lambda) = \frac{\left\| s^{obs} - s \right\|^2}{\{\text{trace}[I - A(\lambda)]\}^2} \tag{2.6}$$

Where $A(\lambda) = \mathbf{A}(\mathbf{A}^T\mathbf{A} + \lambda^2 I)^{-1}\mathbf{A}^T$. Algorithm 1 performs multitrace wedge-dictionary inversion. Starting from an initial reflectivity estimate, each iteration computes predicted traces, updates residuals and weighting matrices, refines the dictionary via K-SVD, selects λ by GCV, and solves the reflectivity update with Split-

Bregman, repeating until convergence.

Algorithm 1 Multitrace Inversion with wedge-dictionary Algorithm

Require: Observed seismic traces S_j^{obs}, dictionary matrix G, initial reflectivity model I_j, regularization parameter λ

Ensure: Updated reflectivity model θ

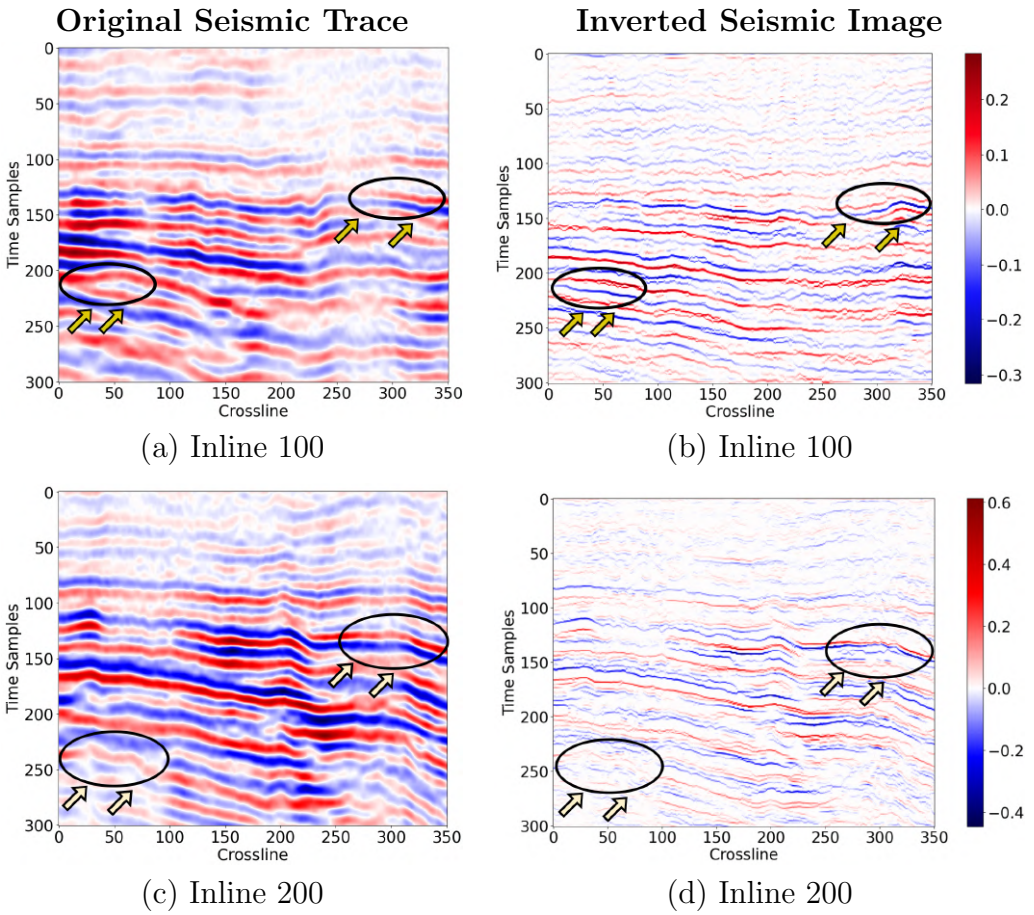
Initialize reflectivity model I_j with an initial guess
Initialize weighting matrices W_a and W_m
Initialize dictionary matrix G
repeat
    Compute predicted seismic traces $S_j = G \cdot I_j$
    Compute residuals: residual $= S_j^{\text{obs}} - S_j$
    Update weighting matrices:
        $W_a = \mathbf{Diag}\left(\frac{1}{\sigma_a}\right), \quad \sigma_a =$ data error standard deviation
        $W_m = \mathbf{Diag}[\frac{1}{\sigma_m}], \quad \sigma_m =$ model standard deviation
    Update dictionary matrix G using K-SVD algorithm based on residuals
    Compute GCV using Equation 2.6
    Solve for updated reflectivity model θ using Split-Bregman method to minimize objective function:
        $\theta = \arg\min \left\{ \left\| W_a(S_j^{\text{obs}} - S_j) \right\|^2 + \lambda^2 \left\| W_m(I_j - I_{\text{ref}}) \right\|^2 \right\}$
until Convergence criteria met

2.3 Results of Wedge-Model Based Multi-Trace Inversion with Modified Linear Constrained Inversion

The wedge-model-based multi-trace inversion with modified linear constrained inversion (LCI) significantly improves seismic data resolution for thin layers, with statistical regularization and parameter optimization playing a key role.

Tuning λ controls the resolution–stability tradeoff: $\lambda = 10^{-4}$ resolves extremely thin layers with high precision, while $\lambda = 2.5 \times 10^{-1}$ balances data fidelity and model complexity (Figure 2.7).

correction with explanation The inversion analysis was primarily conducted on



(a) Inline 100 (b) Inline 100

(c) Inline 200 (d) Inline 200

Figure 2.7: Comparison of original and inverted seismic sections from the KG Basin real dataset. The analysis is performed over the depth interval of 600 – 950 samples, corresponding to 1200–1900 ms with a sampling interval of 2 ms. (a)–(b) Original and inverted results for Inline 100 with $\lambda = 0.0001$; (c)–(d) original and inverted results for Inline 200 with $\lambda = 0.01$.

Inline-71, targeting a specific vertical window along the Z-axis. The seismic volume comprises 1001 vertical samples at a 2 ms sampling rate, where the Z-axis index corresponds to two-way travel time. Consequently, we isolated sample indices 650 to 900, which precisely delimit the 1.4 s to 1.8 s time interval. The inversion results obtained from this window (Figure 2.7) substantiate the model's robustness, effectively capturing complex real-world geological features while preserving high vertical resolution. Three-dimensional visualization and impedance matching further improve resolution and interpretability. Extending the 2-D section into a 3-D volumetric framework provides a broader subsurface perspective; well-log impedance data are re-sampled to match the seismic sampling frequency. An 80 Hz zero-phase symmetric

wavelet extracted from field data is used throughout the inversion, merging well-log low-frequency trends with high-frequency seismic reflectivity content.

Figure 2.8 confirms that wedge-based LCI yields a sparser seismic image than the original: random noise is substantially attenuated and layer boundaries are more sharply delineated, confirming the enhanced subsurface resolution.

The inversion targets Inline-71 over sample indices 700–900 of the 1001-sample (2 ms) volume, corresponding to the 1.4–1.8 s time window. Results for this window (Figure 2.7) confirm the model's robustness in capturing complex geological features at high vertical resolution.

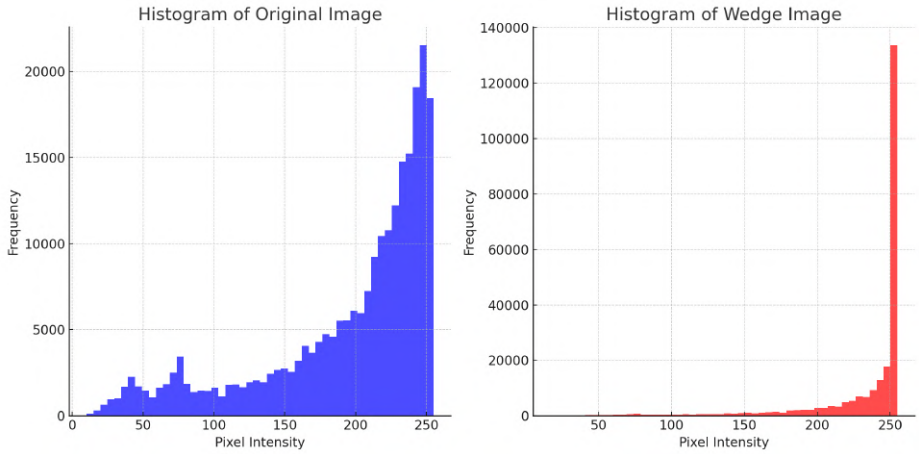


Figure 2.8: Real data vs Wedge-based LCI sparsity comparison

2.4 Discrete Prolate Slepian Sequences (DPSS) Basis Learning on Multitrace Seismic Inversion with Linear Constrained Inversion (LCI)

This section implements DPSS-based basis learning within Linear Constrained Inversion (LCI). DPSS maximizes energy concentration, enabling more orthogonal and accurate representation of sparse layers and yielding improved resolution.

2.4.1 Introduction to DPSS in Seismic Inversion

Discrete prolate Slepian sequences (DPSS) concentrate energy within prescribed time-frequency regions [168–172], enhancing detection of thin, sparse, and localized features in seismic data.

Resolution improves by balancing time and frequency localization through the time-bandwidth product NW; adjusting NW optimizes signal representation and noise suppression. Sparse coding against a DPSS dictionary eliminates redundant information while preserving essential features, and the flexibility of overcomplete DPSS dictionaries accommodates diverse seismic signal characteristics [173–177, 179, 180] (Figure 2.9).

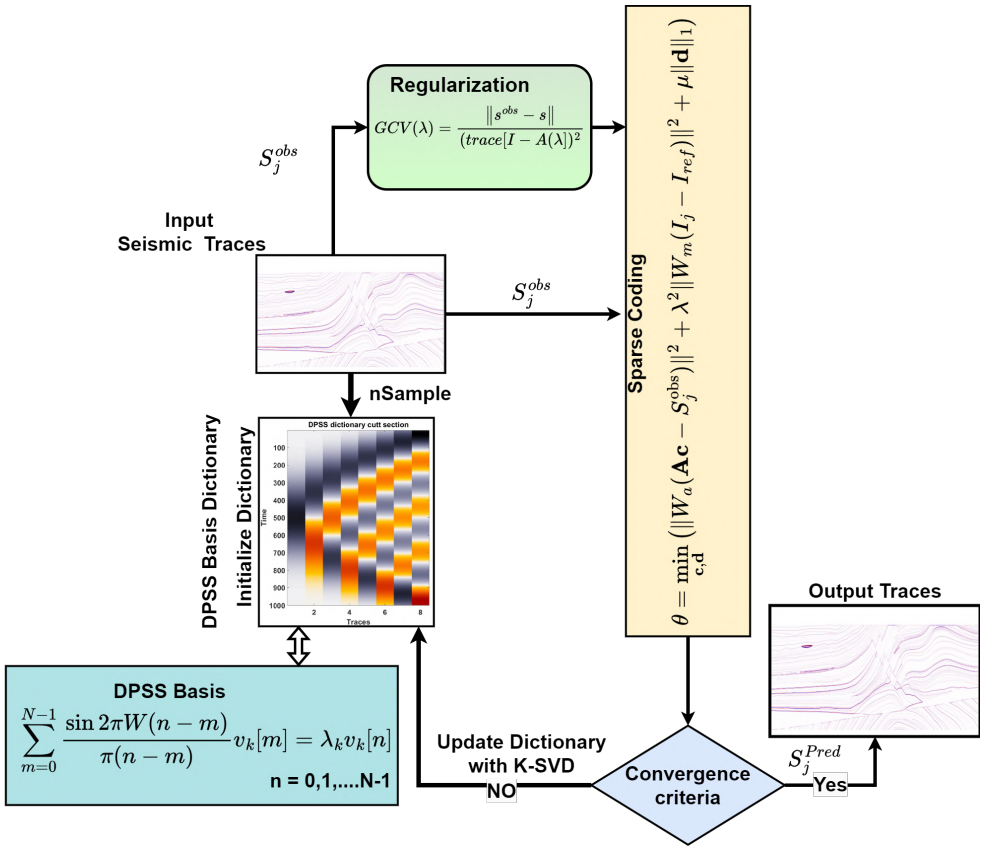


Figure 2.9: DPSS basis dictionary learning with multi-trace inversion (DPSS-LCI)

2.4.2 DPSS Theory

DPSS are eigenfunctions of the time-bandwidth product matrix that maximize energy concentration within a given band (Figure 2.10). For a sequence of length N and bandwidth W, they satisfy:

$$\sum_{m=0}^{N-1} \frac{\sin 2\pi W(n-m)}{\pi(n-m)} v_k[m] = \lambda_k v_k[n], \quad n = 0, 1, \ldots, N-1 \tag{2.7}$$

Here λ_k are eigenvalues and $v_k[n]$ are the DPSS eigenfunctions for sequence k, designed to concentrate maximum energy within $[-W, W]$.

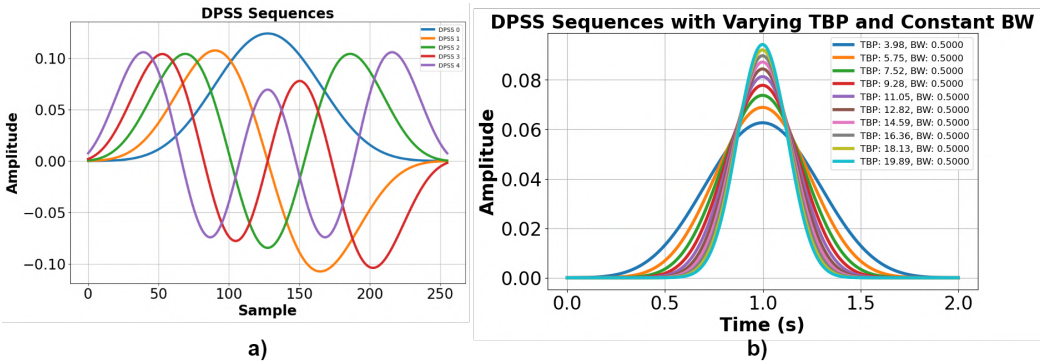


Figure 2.10: (a) DPSS basis (b) DPSS basis varying with Time-bandwidth product

2.4.3 DPSS Basis on Multitrace Seismic Inversion with LCI

The DPSS-based inversion constructs the dictionary matrix $\mathbf{A}$ (Figure 2.11) as:

$$\mathbf{A}_{DPSS} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N] \tag{2.8}$$

Figure 2.12 shows the DPSS basis vectors (Slepian sequences) arranged with time increasing down the rows and numerical values across the columns. Low-frequency components with smooth variations occupy the first row; successive rows exhibit progressively higher-frequency oscillations, collectively spanning the full spectral range needed for time-frequency analysis.

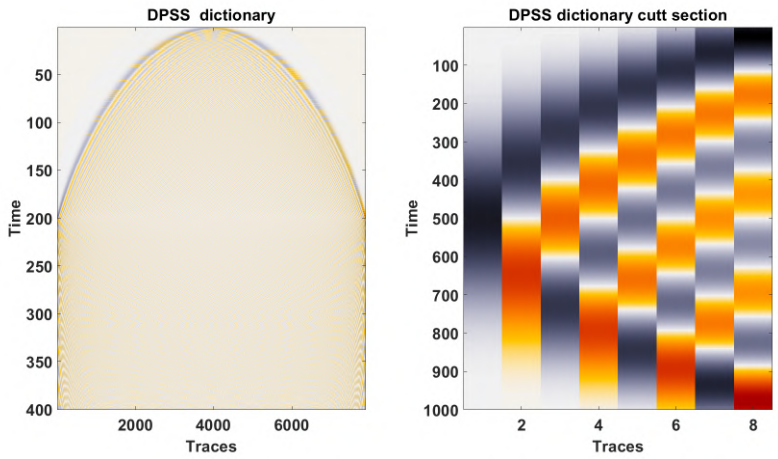


Figure 2.11: DPSS basis Dictionary

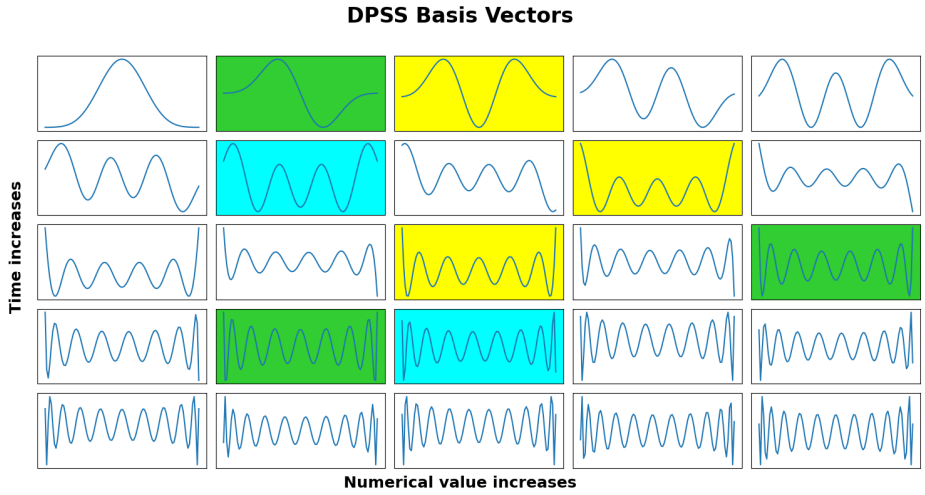


Figure 2.12: DPSS basis vectors for seismic traces

2.4.4 Algorithm DPSS basis dictionary on LCI

The DPSS basis LCI objective function is solved by integrating L-BFGS and Split-Bregman methods to solve for the reflectivity model θ. The algorithm involves several key steps:

Algorithm 2 updates the reflectivity model from observed seismic traces using a DPSS dictionary. After initialization, each iteration alternates between L-BFGS sparse coding and K-SVD dictionary update, followed by GCV-based λ selection and a

2.4 Discrete Prolate Slepian Sequences (DPSS) Basis Learning on Multitrace Seismic Inversion with Linear Constrained Inversion (LCI)

Algorithm 2 Multitrace Inversion with DPSS-Dictionary Algorithm

Require: Observed seismic traces S_j^{obs}, initial dictionary matrices $\mathbf{A}_{DPSS}$, initial reflectivity model I_j, regularization parameter λ

Ensure: Updated reflectivity model θ

Initialize reflectivity model I_j with an initial guess

Initialize weighting matrices W_a and W_m

Construct initial dictionary matrix $\mathbf{A} = [\mathbf{A}_{DPSS}]$

repeat

  Sparse Coding with LCI:

    Compute predicted seismic traces $S_j = \mathbf{A} \cdot \mathbf{c}$

    Compute residuals: residual $= S_j^{\text{obs}} - S_j$

    Update weighting matrices: $W_a = \text{Diag}\left[\frac{1}{\sigma_a}\right]$, $W_m = \text{Diag}\left[\frac{1}{\sigma_m}\right]$

    $\sigma_a =$ data error standard deviation, $\sigma_m =$ model standard deviation

    Solve for sparse coefficients $\mathbf{c}$ using L-BFGS:

      $\mathbf{c} = \arg\min_{\mathbf{c}} \left(\|W_a(\mathbf{Ac} - S_j^{\text{obs}})\|^2 + \lambda^2 \|W_m(I_j - I_{ref})\|^2\right)$

  Dictionary Update with K-SVD:

  for each atom k in the dictionary **do**

    Compute the residual matrix excluding the k-th atom:

      $\mathbf{E}_k = \mathbf{Y} - \sum_{j \neq k} \mathbf{a}_j \mathbf{c}_j$

    Perform SVD on $\mathbf{E}_k$: $\mathbf{E}_k = \mathbf{U\Sigma V}^T$

    Update the k-th atom and corresponding coefficients:

      $\mathbf{a}_k^{(t+1)} = \mathbf{U}[:, 1]$, $\mathbf{c}_k^{(t+1)} = \mathbf{\Sigma}(1,1)\mathbf{V}[:, 1]$

  end for

  Compute GCV (λ)

$$\text{GCV}(\lambda) = \frac{\left\|s^{obs} - s\right\|^2}{(\text{trace}[I - A(\lambda)])^2}$$

  Solve for updated reflectivity model θ using Split-Bregman

  Minimize reformulated objective function:

    $\theta = \min_{\mathbf{c},\mathbf{d}} \left(\|W_a(\mathbf{Ac} - S_j^{\text{obs}})\|^2 + \lambda^2 \|W_m(I_j - I_{ref})\|^2 + \mu \|\mathbf{d}\|_1\right)$

    subject to $\mathbf{d} = \mathbf{c}$

    Split-Bregman iteration steps:

    $\mathbf{c}^{k+1} = \arg\min_{\mathbf{c}} \Bigg(\|W_a(\mathbf{Ac} - S_j^{\text{obs}})\|^2 + \lambda^2 \|W_m(I_j - I_{ref})\|^2$

      $+ \frac{\mu}{2} \|\mathbf{c} - (\mathbf{d}^k - \mathbf{b}^k)\|^2 \Bigg)$

    Soft-thresholding operation:

    $\mathbf{d}^{k+1} = \arg\min_{\mathbf{d}} \left(\|\mathbf{d}\|_1 + \frac{\mu}{2} \|\mathbf{d} - (\mathbf{c}^{k+1} + \mathbf{b}^k)\|^2\right)$

    $\mathbf{b}^{k+1} = \mathbf{b}^k + (\mathbf{c}^{k+1} - \mathbf{d}^{k+1})$

until Convergence criteria met

Split-Bregman reflectivity update. Iterations continue until convergence.

2.5 Parameter Selection and Sensitivity Analysis

Both the Wedge-Based Dictionary Learning LCI (wedge-LCI) and the DPSS-Based Dictionary Learning LCI (DPSS-LCI) methods rely on several tunable parameters that influence inversion stability, waveform fidelity, and thin-bed resolvability. This section summarizes the key parameters and the sensitivity studies used to determine their optimal values.

(a) Regularization Weight ϵ

The Linear Constrained Inversion framework is given by

$$\theta = \arg\min_{\theta} \left(\|W_a(S^{obs} - S)\|_2^2 + \epsilon^2 \|W_m(I - I_{ref})\|_2^2 \right), \tag{2.9}$$

where ϵ controls the trade-off between data fidelity and model smoothness. A sweep over $\epsilon \in [10^{-5}, 10^{-1}]$ using the Generalized Cross-Validation metric given in Equation 2.6, showed a consistent minimum in the range $\epsilon \approx 10^{-4}$–10^{-3}. This range offers a robust balance between inversion stability and thin-bed detail, suitable for both algorithms.

(b) Frequency Cut-Off Selection

To reconstruct the missing low-frequency impedance trend, a cut-off frequency is selected using the tuning relation

$$f_c \approx \frac{v}{4h_{\text{dominant}}}. \tag{2.10}$$

Given the predominant seismic frequency

$$f_{\text{pred}} \approx 40 \text{ Hz}, \tag{2.11}$$

the useful low-frequency region for impedance modelling lies below

$$f_{\text{low}} \lesssim \frac{f_{\text{pred}}}{3} \approx 13 \text{ Hz}. \tag{2.12}$$

Sensitivity tests over 8–20 Hz yielded:

- < 10 Hz: underconstrained impedance trend,
- 10–15 Hz: stable inversion and good agreement with well-log trends,
- > 15 Hz: leakage of seismic noise into the low-frequency model.

Thus, the selected cut-off range is

$$f_c = 10\text{–}15 \text{ Hz}. \tag{2.13}$$

(c) DPSS Time-Bandwidth Product (NW)

DPSS basis functions satisfy the eigenvalue relation in equation provided earlier in Equation 2.7. A sensitivity study with NW = 2, 3, 4, 5 showed:

- NW = 2: insufficient temporal resolution,
- NW = 3–4: best balance of sharpness and spectral compactness,
- NW > 4: oscillatory artifacts.

The optimal DPSS setting is therefore NW = 3–4.

Final Parameter Set

The optimal parameter ranges adopted throughout this work are:

$$\epsilon \in [10^{-4}, 10^{-3}], \qquad f_c = 10\text{–}15 \text{ Hz}, \qquad \text{NW} = 3\text{–}4. \tag{2.14}$$

These selections provide stable inversion behaviour, improved lateral continuity, and reliable thin-bed detection across the real seismic datasets evaluated.

2.6 Results: DPSS Basis on Multitrace Seismic Inversion with LCI

DPSS-basis multitrace inversion with LCI outperforms traditional techniques. Figure 2.13 compares original and DPSS-LCI-inverted seismic images. Traditional inver-

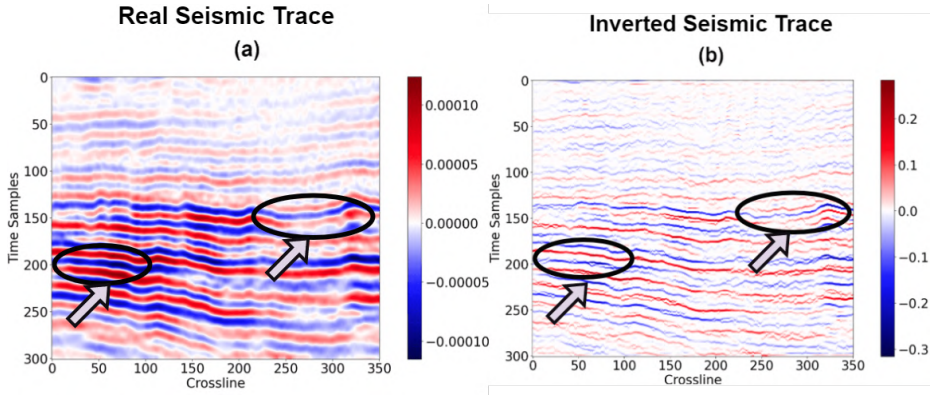


Figure 2.13: Real seismic and Discrete prolate Slepian wavelet comparison at (a-b) Inline 50 ($\lambda = 0.001$)

sion methods exhibit spectral leakage, reduced coherence, and lower boundary resolution; the DPSS-based method improves interpretability and structural accuracy. Computational details are provided in Section 2.7.1.

2.7 Results Comparison

Figure 2.14 compares five inversion techniques—Sparse Spike Inversion (SSI) [3], Basis Pursuit Inversion (BPI) [4], LCI [5], Wedge-based LCI, and DPSS-based LCI—on Inline 50. Both proposed methods improve resolution substantially over the baselines, with DPSS-based LCI achieving the highest accuracy.

Figure 2.15 directly compares Wedge-based LCI and DPSS-based LCI thin-bed detection on Inline 50, confirming DPSS-based LCI's superior resolution.

Table 2.1 reports PCC, MSE (Equation C.1), and PSNR (Equation C.3) for the proposed and baseline methods:

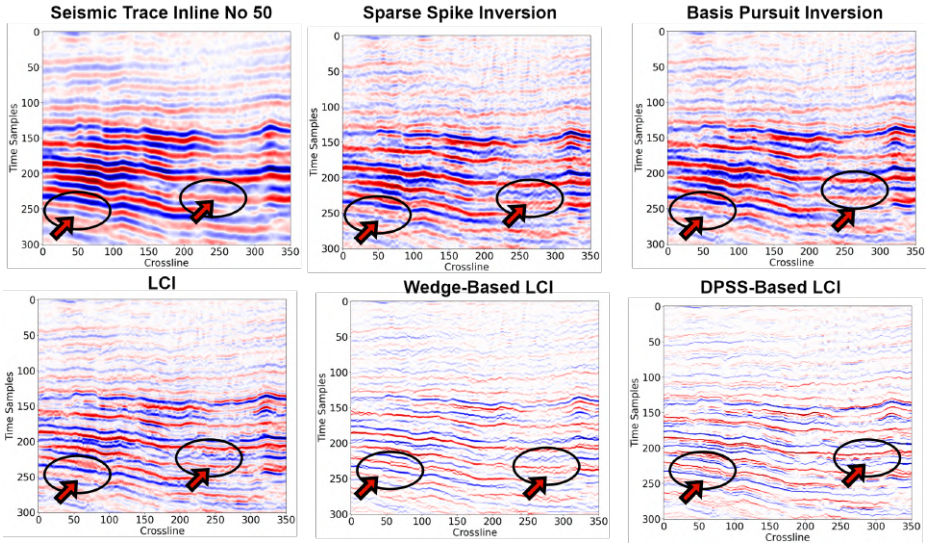


Figure 2.14: Comparison of SSI [3], BPI [4], LCI [5], Wedge-based LCI, and DPSS-based LCI for Inline section 50.

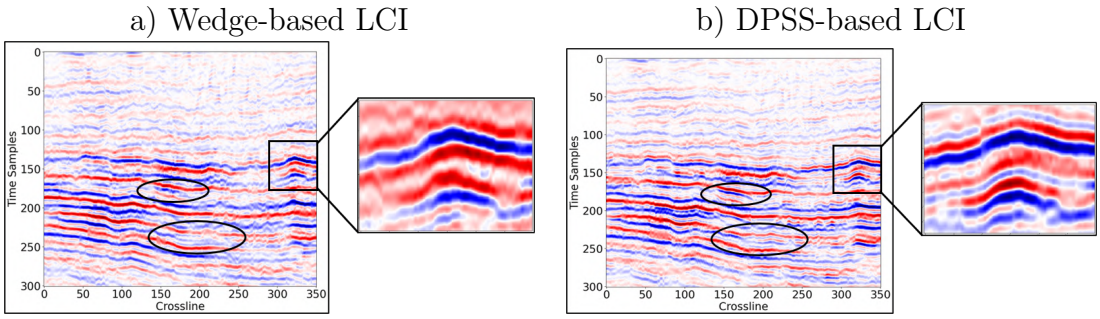


Figure 2.15: Comparison between a) Wedge-based LCI and b) DPSS-based LCI at Inline 50

Method	MSE↓	MAE↓	SSIM↑	PCC↑
BPI [41]	0.0246	0.0735	0.6583	0.6854
1D-LCI [5]	0.0155	0.0589	0.7425	0.7552
Wedge-Dicl-LCI (Ours)	0.0098	0.0452	0.7904	0.7641
DPSS-Dicl-LCI (Ours)	**0.0088**	**0.0325**	**0.8152**	**0.8026**

Abbreviations: Wedge-Dicl-LCI = Wedge-based Dictionary Learning with LCI; DPSS-Dicl-LCI = DPSS-based Dictionary Learning with LCI.

Table 2.1: Performance comparison on real seismic data. **Bold** = best per metric. ↓ lower is better; ↑ higher is better. Grey rows = external baselines.

2.7.1 Computational Complexity and Runtime Analysis

This section compares the computational behaviour of the proposed Wedge-LCI and DPSS-LCI algorithms against established methods, using the KG-basin seismic dataset.

2.7.1.1 Dataset and Hardware Specification

All runtime measurements are for the full $3D$ volume: 1001 samples/trace (2 ms sampling), 280,000 traces (400 inlines $\times$ 700 crosslines), 40 Hz predominant frequency. Tests were conducted on a high-performance workstation with consistent specifications (Intel Xeon CPU, NVIDIA Quadro RTX 4000 GPU, 192 GB RAM).

2.7.1.2 Algorithmic Complexity Comparison

Table 2.2 summarises the dominant computational complexity and measured execution time for each method.

Method	**Computational Complexity**	**Time (s)↓**
DPSS-Dicl-LCI (Ours)	$O(N^2K^2 + IK^2)$	742
Wedge-Dicl-LCI (Ours)	$O\big(I(NK + MK^2)\big)$	391
BPI [38]	$O(INK + K^{3/2})$	265
OMP [181]	$O(NM + Mk + Mk^2 + k^3)$	188
1D-LCI [5]	$O\big(I(NK + NM + MK^2)\big)$	328

Note: N = samples per trace, M = dictionary atoms, K = sparsity level, I = number of iterations.

Table 2.2: Algorithmic complexity and execution time on the KG Basin dataset. Lower execution time indicates faster convergence.

The results show that DPSS-LCI has the highest computational cost due to its orthogonal DPSS atoms and iterative dictionary updates. In contrast, the proposed Wedge-LCI achieves a favourable balance, offering improved resolution over BPI and 1D-LCI with substantially lower computational effort than DPSS-LCI.

Together, Table 2.2 and Table 2.1 (accuracy metrics) demonstrate that DPSS-LCI provides the highest interpretational accuracy, while Wedge-LCI delivers an efficient compromise between computational cost and reconstruction quality.

2.8 Discussion

Table 2.1 highlights the clear performance advantage of the proposed DPSS-based and Wedge-based Dictionary Learning (Dicl-LCI) methods over conventional inversion techniques such as BPI and SSI. The DPSS Dicl-LCI variant consistently achieves the lowest MSE and MAE and the highest SSIM and PCC, reflecting its superior accuracy. The physical relevance of these metrics to seismic inversion is summarized below (mathematical definitions are provided in **Appendix A**):

- **MSE & MAE (Amplitude Fidelity):** These metrics quantify amplitude mismatch between the true and inverted impedance. The significantly lower errors obtained with the DPSS method indicate strong amplitude fidelity, which is essential for quantitative reservoir characterization such as porosity and fluid estimation.
- **SSIM (Structural Integrity):** SSIM measures structural preservation and inter-trace similarity. The high SSIM value (0.8152) demonstrates that the method maintains geological continuity, stratigraphic boundaries, and fine-scale textural patterns, avoiding the blocky artifacts common in trace-by-trace inversion.
- **PCC (Waveform Similarity):** The Pearson Correlation Coefficient (PCC) quantifies the linear similarity between the estimated and true seismic traces. A high PCC indicates that the reconstructed waveform closely matches the shape of the true trace, reflecting strong consistency in amplitude variations and event timing. This similarity ensures that reflection patterns and impedance contrasts follow the correct geological trends observed in the well-log data.

Beyond statistical validation, the inversion results provide clearer geological insights. The key contributions toward subsurface interpretation are summarized below:

- **Lithological Discrimination:** The high-resolution impedance from the DPSS method enables improved separation of lithofacies in the 1.4 s–1.8 s interval. Low-impedance, gas-bearing sands are more distinctly resolved against the surrounding high-impedance shale matrix.
- **Thin-Bed Resolution:** A major advantage of the DPSS-LCI method is its

ability to resolve layers thinner than the classical tuning thickness ($\lambda/4$). This allows clearer identification of thin stratigraphic traps, pinch-outs, and net-pay variations, improving reservoir volume estimation.

- **Lateral Continuity:** The multi-trace regularization enforces spatial coherence across traces, yielding laterally consistent impedance fields. This enhanced continuity supports the interpretation of depositional geometries, such as distinguishing sheet sands from channelized bodies.

2.9 Conclusion

This chapter presented two advanced seismic inversion techniques—DPSS-based basis learning and wedge-based dictionary learning—both integrated with Linear Constrained Inversion (LCI).

1. **Wedge-Based Dictionary Learning with LCI**: Wedge functions improve the representation of broad signal features; iterative sparse coding and dictionary updating enhance robustness and accuracy.
2. **DPSS-Based Basis Learning on Multitrace Seismic Inversion with LCI**: Concentrating energy within prescribed time-frequency regions reduces spectral leakage and preserves edge information, yielding clearer, more coherent seismic reflections and an accurate subsurface representation.

2.10 Summary

This chapter introduced wedge-based dictionary learning with LCI and DPSS-based basis learning as advanced sparse representation techniques for seismic inversion, achieving significant gains in signal fidelity and resolution. Future work will combine dictionary learning with group sparsity-based optimization and Higher-order Total Generalized Variation (TGV) for robust edge-preserving regularization.

CHAPTER 3

Joint Overlapping Wedge-Based Group Sparsity with Higher-Order Total Generalized Variation

3.1 Introduction

Seismic inversion has benefited from sparsity constraints that focus on key geological features; overlapping group sparsity extends this by providing a structured representation in which variables belong to multiple groups simultaneously [44].

This work combines Joint Overlapping Group Sparsity (JOGS) with wedge-based group sparsity and higher-order Total Generalized Variation (TGV) regularization, forming a unified framework called **Phy-JOGS-TGV**, particularly effective for resolving thin-bedded reservoirs and capturing subtle layer-thickness and dip variations.

The wedge model [21] parameterizes reflectivity at sub-tuning thicknesses; grouping coefficients by wedge thickness helps overcome the tuning-thickness limitation [182, 183]. Higher-order TGV preserves sharp geological boundaries while suppress-

ing staircase artifacts [184–186]. ADMM ensures reliable convergence and aFISTA accelerates the optimization [183, 187].

3.1.1 Objective and Contributions

This chapter presents a seismic inversion methodology that integrates JOGS with wedge-based group sparsity, enhanced by higher-order TGV regularization. The major contributions are:

1. **Development of a JOGS method:** A novel framework improving resolution of thin, sparse layers and complex geological structures.
2. **Wedge-Based Group Sparsity:** A wedge dictionary within the group sparsity framework that improves thin-bed reservoir differentiation.
3. **Higher-Order TGV^2 Regularization:** Second-order TGV balancing preservation of sharp geological boundaries with model smoothness.
4. **ADMM and aFISTA Optimization:** Robust convergence via ADMM combined with aFISTA for accelerated efficiency.
5. **Rigorous Convergence and Stability Analysis:** Proves asymptotic convergence of the deterministic ADMM–ALM–AFISTA subproblem to a KKT stationary point (Lemma 1), with a Kurdyka–Lojasiewicz remark covering the full stochastic algorithm, and quantifies local stability under measurement noise (Lemma 2).
6. **Computational Complexity and Scalability Analysis:** Provides a detailed per-iteration complexity table (Table 3.1) and a GPU runtime benchmark (Table 3.2), demonstrating practical scalability to field-scale 3D surveys.
7. **Enhanced Geological Interpretation with Well-Log Correlation:** Correlates inverted reflectivity with well logs from Wells PLD-12 and WDU-JGS along Inlines 65 and 85, demonstrating thin-bed and lithological consistency beyond numerical metrics.
8. **Annotated Thin-Bed and Fault Visualization:** Revises Figure 3.4 to include black ellipsoid annotations explicitly highlighting thin-bed regions, subtle

stratigraphic features, and improved structural fidelity relative to DPSS-DicI-LCI.

3.2 Background

Reflectivity estimation from seismic traces is essential for subsurface characterization [188], yet the problem is underdetermined and ill-posed, compounded by band-limited data [15, 20, 149]. Sparse constraints (ℓ_1 norm, compressed sensing) improve stability [4, 19], but traditional methods lack spatial regularization for complex structures [189, 190]. Total Variation manages sparse edges and vertical variations [191–196], while higher-order TGV additionally preserves sharp boundaries and reduces staircase artifacts [184–186]. Joint sparsity promotes shared components across traces for improved robustness [197]; ADMM and aFISTA provide efficient convergence [183, 187].

3.3 Workflow

Seismic traces are pre-processed to generate the initial wedge dictionary. Wedge group sparsity employs groups of wedges at different thicknesses within the JOGS objective; the dictionary, model, and wedgelet groups are updated at each iteration with higher-order TGV [186] applied to the objective. The overall pipeline is shown in Figure 3.1.

3.4 Proposed Methodology

This chapter proposes a novel group-sparsity-based method with improved generalized regularization for enhanced thin-sparse layer detection. The methodology is built upon four key pillars: (1) a sub-wavelength reflectivity parameterization using a wedgelet dictionary; (2) a composite objective function combining a robust data-fidelity term with group sparsity and TGV^2 regularization; (3) an efficient and stable hybrid ADMM–AFISTA solver; and (4) an adaptive joint wavelet estimation procedure. The overall procedure is summarized in Algorithm 3.

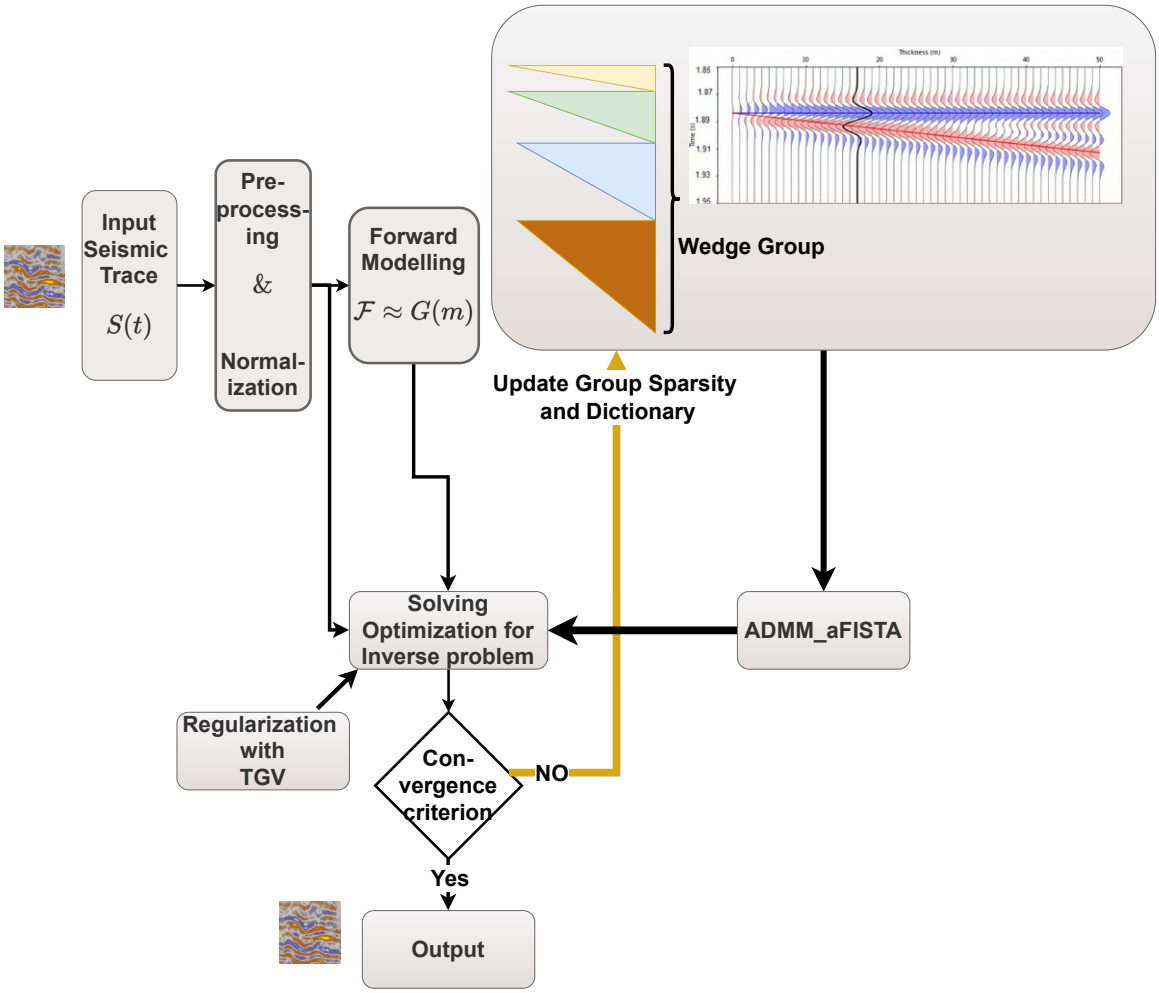


Figure 3.1: Pipeline diagram for the Phy-JOGS-TGV seismic inversion framework.

3.4.1 Forward Model and Wedgelet Decomposition

Seismic data $\mathbf{S} \in \mathbb{R}^{N_t \times N_x}$ follow the convolutional model:

$$\mathbf{S} = \mathbf{WR} + \mathbf{N} \tag{3.1}$$

where $\mathbf{W}$ is a Toeplitz wavelet matrix and $\mathbf{N}$ is Gaussian noise with σ_{noise} estimated via the Median Absolute Deviation. Reflectivity is related to acoustic impedance

$\mathrm{AI}(t) = v(t)\rho(t)$ by:

$$r(t) = \frac{\mathrm{AI}(t+\Delta t) - \mathrm{AI}(t)}{\mathrm{AI}(t+\Delta t) + \mathrm{AI}(t)} \tag{3.2}$$

To resolve thin beds below the classical Rayleigh limit ($\lambda/4$), we parameterize the reflectivity using a dictionary of wedgelet basis functions [198]. The reflectivity is decomposed as:

$$\mathbf{R} = \sum_{m=1}^{M}\sum_{n=1}^{N}\big(\mathbf{E}_{m,n}\,\mathbf{a}_{m,n} + \mathbf{O}_{m,n}\,\mathbf{b}_{m,n}\big) \tag{3.3}$$

where $\mathbf{E}_{m,n}$ and $\mathbf{O}_{m,n}$ are the even and odd wedgelet basis functions at position m and thickness n, with coefficients $\mathbf{a}_{m,n}, \mathbf{b}_{m,n} \in \mathbb{R}^{N_x}$. Their frequency-domain definitions facilitate efficient implementation:

$$\mathcal{F}\{r_e(t;\, m, n)\} = e^{-j\omega m\Delta t}\big(1 + e^{j\omega n\Delta t}\big) \tag{3.4}$$

$$\mathcal{F}\{r_o(t;\, m, n)\} = e^{-j\omega m\Delta t}\big(1 - e^{j\omega n\Delta t}\big) \tag{3.5}$$

3.4.2 Composite Objective Function

The inversion minimizes:

$$\begin{aligned}\mathcal{J} = \underbrace{\mathcal{L}_{\mathrm{MSL}}(\mathbf{R})}_{\text{Data fidelity}} &+ \lambda_1 \underbrace{\sum_{m,n}\sqrt{\sum_{i}\big(a_{i,m,n}^2 + b_{i,m,n}^2\big)}}_{\text{Group sparsity}} \\ &+ \lambda_2 \underbrace{\min_{\mathbf{P}}\Big(\alpha_1\|\nabla\mathbf{R} - \mathbf{P}\|_1 + \alpha_2\|\mathcal{E}(\mathbf{P})\|_1\Big)}_{TGV^2\ \text{(edge preservation)}}\end{aligned} \tag{3.6}$$

where $\lambda_1, \lambda_2, \alpha_1, \alpha_2$ are regularization parameters and $\mathcal{E}(\mathbf{P})$ denotes the symmetrized gradient of $\mathbf{P}$.

3.4.2.1 Data Fidelity: Magnitude-Sensitive Loss with Adaptive Masking

We employ a magnitude-sensitive loss (MSL) function akin to the Huber loss, computed in the frequency domain and weighted by an adaptive mask $M(\omega_k)$:

$$M(\omega_k) = \frac{|\mathcal{F}(W)(\omega_k)|^2}{|\mathcal{F}(W)(\omega_k)|^2 + 10^{-3}\max_\omega |\mathcal{F}(W)(\omega)|^2} \tag{3.7}$$

$$\mathcal{L}_{\text{MSL}} = \sum_{j=1}^{N_x}\sum_{\omega_k} M(\omega_k)\cdot \ell\big(|e_j(\omega_k)|\big), \qquad \ell(e) = \begin{cases} \frac{1}{2}e^2 & e \le \delta \\ \delta e - \frac{1}{2}\delta^2 & e > \delta \end{cases}, \quad \delta = 1.28\,\sigma_{\text{noise}} \tag{3.8}$$

The gradient with respect to $\mathbf{R}$ is:

$$\nabla_{\mathbf{R}}\mathcal{L}_{\text{MSL}} = \mathcal{F}^{-1}\left[M(\omega_k)\cdot\phi(e_j)\cdot\frac{\mathcal{F}(W)^*\cdot e_j}{|e_j|+\epsilon}\right], \quad \phi(e) = \min(e,\delta),\ \epsilon = 10^{-8} \tag{3.9}$$

3.4.2.2 Regularization

The regularization combines: (i) group sparsity via the $\ell_{2,1}$-norm enforcing joint sparsity across wedgelet coefficients; (ii) lateral continuity via $\|\nabla_x\mathbf{R}\|_1$; and (iii) second-order TGV^2:

$$\text{TGV}^2(\mathbf{R}) = \inf_{\mathbf{P}}\big\{\alpha_1\|\nabla\mathbf{R}-\mathbf{P}\|_1 + \alpha_2\|\mathcal{E}(\mathbf{P})\|_1\big\} \tag{3.10}$$

3.4.3 ADMM–ALM–AFISTA Optimization

The augmented Lagrangian is:

$$\mathcal{L}_\rho = \mathcal{J} + \frac{\rho}{2}\|\mathcal{F}(\mathbf{R})-\mathbf{V}\|_F^2 + \langle\mathbf{\Lambda},\,\mathcal{F}(\mathbf{R})-\mathbf{V}\rangle \tag{3.11}$$

where $\mathbf{V} = \mathcal{F}(\mathbf{E})\mathbf{a} + \mathcal{F}(\mathbf{O})\mathbf{b}$ and $\mathbf{\Lambda}$ is the dual variable.

3.4.3.1 Reflectivity Update (AFISTA)

1. **Momentum extrapolation:**

$$t_k = \tfrac{1+\sqrt{1+4t_{k-1}^2}}{2}, \quad t_0 = 1 \tag{3.12}$$

$$\mathbf{V}_R^{(k)} = \mathbf{R}^{(k)} + \tfrac{t_{k-1}-1}{t_k}\big(\mathbf{R}^{(k)} - \mathbf{R}^{(k-1)}\big) \tag{3.13}$$

2. **Gradient:**

$$\nabla_{\mathbf{R}}\mathcal{L}_{\text{smooth}} = \nabla_{\mathbf{R}}\mathcal{L}_{\text{MSL}} + \rho\,\mathcal{F}^{-1}\big(\mathcal{F}(\mathbf{R}) - \mathbf{V}^{(k)}\big) \tag{3.14}$$

3. **Proximal step** (solved to numerical convergence):

$$\mathbf{U} = \mathbf{V}_R^{(k)} - \gamma\nabla_{\mathbf{R}}\mathcal{L}_{\text{smooth}} \tag{3.15}$$

$$\mathbf{R}^{(k+1)} = \arg\min_{\mathbf{R}} \tfrac{1}{2}\|\mathbf{R}-\mathbf{U}\|_F^2 + \gamma\lambda_2\|\nabla_x\mathbf{R}\|_1 + \gamma\alpha_1\|\nabla\mathbf{R} - \mathbf{P}^{(k)}\|_1 \tag{3.16}$$

4. **TGV auxiliary update** (solved to numerical convergence):

$$\mathbf{P}^{(k+1)} = \arg\min_{\mathbf{P}} \alpha_1\|\nabla\mathbf{R}^{(k+1)} - \mathbf{P}\|_1 + \alpha_2\|\mathcal{E}(\mathbf{P})\|_1 \tag{3.17}$$

5. **Curvature restart:**

$$\cos^{-1}\left(\frac{\langle\nabla\mathcal{L}_{\text{smooth}}, \Delta\mathbf{R}\rangle}{\|\nabla\mathcal{L}_{\text{smooth}}\|\,\|\Delta\mathbf{R}\|}\right) > 85^\circ \tag{3.18}$$

3.4.3.2 Wedgelet Update

1. **Group selection probability:**

$$p_{m,n} \propto \exp\!\left(-\tfrac{\|\mathbf{a}_{:,m,n}\|_2^2 + \|\mathbf{b}_{:,m,n}\|_2^2}{\sigma}\right)\|\nabla^2\mathbf{R}\|_{m,n} \tag{3.19}$$

2. **Stochastic selection:** 12% highest-probability groups plus 3% random groups per iteration.

3. **Group soft-thresholding:**

$$\begin{bmatrix}\mathbf{a}_{m,n}\\ \mathbf{b}_{m,n}\end{bmatrix}^{(k+1)} = \max\!\Big(1 - \tfrac{\lambda_1}{\rho\|\mathbf{v}_{m,n}\|_2}, 0\Big)\,\mathbf{v}_{m,n} \tag{3.20}$$

where $\mathbf{v}_{m,n} = [\langle\mathcal{F}(\mathbf{E}_{m,n}), \mathbf{z}\rangle, \langle\mathcal{F}(\mathbf{O}_{m,n}), \mathbf{z}\rangle]^T$, $\mathbf{z} = \mathcal{F}(\mathbf{R}^{(k+1)}) + \rho^{-1}\mathbf{\Lambda}^{(k)}$.

3.4.3.3 Dual Update

$$\mathbf{\Lambda}^{(k+1)} = \mathbf{\Lambda}^{(k)} + \rho\big(\mathcal{F}(\mathbf{R}^{(k+1)}) - \mathbf{V}^{(k+1)}\big) \tag{3.21}$$

3.4.3.4 Convergence Analysis and Stability Guarantees

To provide a rigorous foundation for the optimization scheme, we state the following results. Proofs are deferred to Appendices A.1 and A.2.

Lemma 1 (Convergence of the deterministic ADMM–ALM–AFISTA subproblem)**.** *Consider the algorithm with fixed wavelet* $\mathbf{W}$ *and fixed penalty parameter* ρ*. Let* $\{\mathbf{R}^{(k)}, \mathbf{V}^{(k)}, \mathbf{\Lambda}^{(k)}\}$ *be the iterates generated by the ADMM–ALM–AFISTA scheme. Assume:*

(A) $f(\mathbf{R}) = \mathcal{L}_{MSL}(\mathbf{R})$ *has a Lipschitz-continuous gradient with constant* L_f*;*

(B) The TGV^2 *and group-sparsity regularizers are proper, closed, and convex;*

(C) Step size γ *satisfies the Armijo backtracking condition with* $\gamma \leq 2/(L_f + \rho)$*;*

(D) Proximal subproblems (3.16) *and* (3.17) *are solved to numerical convergence.*

Then $\lim_{k\to\infty} \|\mathbf{R}^{(k+1)} - \mathbf{R}^{(k)}\|_F = 0$ *and every limit point is a KKT stationary point of* (3.11).

Remark 1 (Extension to the full stochastic algorithm)**.** *Lemma 1 applies to the deterministic subproblem with fixed wavelet and penalty. The fully adaptive algorithm—including stochastic wedgelet updates, adaptive* ρ*, and periodic wavelet refinement—falls into the class of inexact nonconvex ADMM. Convergence for this class follows by additionally invoking the Kurdyka–Lojasiewicz property and the proximal alternating minimization framework of [199, 200]. The stochastic wedgelet updates and adaptive parameter scheduling are empirically observed to preserve convergence across all experiments in this chapter.*

Lemma 2 (Local stability under data perturbations)**.** *Let* $\mathbf{R}^*$ *be a locally stationary solution for noise-free data* $\mathbf{S}$*, with* ∇f *Lipschitz continuous. If the data are perturbed by* $\mathbf{N}$ *with* $\|\mathbf{N}\|_F \leq \epsilon$*, then for sufficiently small* ϵ *there exists a locally optimal solution* $\hat{\mathbf{R}}$ *satisfying* $\|\hat{\mathbf{R}} - \mathbf{R}^*\|_F \leq C\,\epsilon$*, where* C *depends on* L_f*,* ρ*, and the local curvature of the regularizers. The* TGV^2 *term attenuates high-frequency noise components, reducing the effective noise-amplification constant in practice.*

3.4.4 Wavelet Estimation

The seismic wavelet is estimated and updated every 10 iterations:

1. **Trace clustering:** K-means on normalized amplitude spectra $\mathcal{F}(\mathbf{R}_i)/\|\mathcal{F}(\mathbf{R}_i)\|_2$.
2. **Cluster-wise spectral division:**

$$\mathcal{F}(W_c)(f) = \frac{\sum_{i\in c} \mathcal{F}(S_i)(f)\,\overline{\mathcal{F}(R_i)(f)}}{\sum_{i\in c} |\mathcal{F}(R_i)(f)|^2 + \kappa_i(f)} \tag{3.22}$$

3. **Spectral smoothing:** Cubic spline within 10–80 Hz.

3.4.5 Parameter Adaptation

1. λ_1 updated so the 75th percentile of wedgelet group norms equals $2\sigma_{\text{noise}}$.
2. TGV first-order weight:

$$\alpha_1^{(k+1)} = \alpha_1^{(k)} \cdot \frac{\|\nabla \mathbf{R}\|_2}{\|\mathbf{P}\|_2} \tag{3.23}$$

3. ADMM penalty:

$$\rho^{(k+1)} = \begin{cases} 1.5\,\rho^{(k)} & \|r_p\|_2 > 10\|r_d\|_2 \\ \rho^{(k)}/1.5 & \|r_d\|_2 > 10\|r_p\|_2 \\ \rho^{(k)} & \text{otherwise} \end{cases} \tag{3.24}$$

3.4.6 Termination Criteria

$$\frac{\|\mathbf{R}^{(k)} - \mathbf{R}^{(k-1)}\|_F}{\|\mathbf{R}^{(k-1)}\|_F} < 10^{-4} \;\text{ and }\; \|r_p\|_2 < \sigma_{\text{noise}}\sqrt{N_t N_x} \;\text{ or }\; k \geq 200 \tag{3.25}$$

3.4.7 Initialization and Implementation Details

- Damped least-squares initialization: $\mathbf{R}^{(0)} = (\mathbf{W}^T\mathbf{W} + 10^{-3}\|\mathbf{W}\|_F^2\mathbf{I})^{-1}\mathbf{W}^T\mathbf{S}$.
- $\lambda_1^{(0)} = 0.1\|\mathbf{R}^{(0)}\|_\infty$, $\lambda_2 = 0.3\lambda_1^{(0)}$, $\rho^{(0)} = 0.5\|\mathcal{F}(\mathbf{S})\|_F$.
- Wedgelet dictionary: $M = N_t$ positions, $N = 20$ thickness levels spanning 0.5–10 ms.

Algorithm 3 Phy-JOGS-TGV: Physics-Guided ADMM–ALM–AFISTA

Input: Data $\mathbf{S}$, wavelet $\mathbf{W}^{(0)}$, params $\rho^{(0)}, \lambda_1^{(0)}, \lambda_2, \alpha_1, \alpha_2$
Output: Reflectivity $\mathbf{R}^*$
Init: $\mathbf{R}^{(0)}$ via damped LS; $\mathbf{a}^{(0)}, \mathbf{b}^{(0)}, \mathbf{\Lambda}^{(0)} \leftarrow \mathbf{0}$; $t_0 \leftarrow 1$
for $k = 1$ **to** max_iter **do**
  Momentum via (3.12)–(3.13)
  Gradient via (3.14)
  $\mathbf{R}^{(k+1)}$ via numerically converged proximal solve (3.15)–(3.16)
  $\mathbf{P}^{(k+1)}$ via numerically converged TGV solve (3.17)
  if restart condition (3.18) **then**
    $t_k \leftarrow 1$
  end if
  Compute $p_{m,n}$ via (3.19); select groups
  Update $\{\mathbf{a}, \mathbf{b}\}^{(k+1)}$ via (3.20)
  $\mathbf{\Lambda}^{(k+1)} \leftarrow$ (3.21)
  if $k \bmod 10 = 0$ **then**
    Estimate $\mathbf{W}^{(k+1)}$ via (3.22)
  end if
  if $k \bmod 5 = 0$ **then**
    Adapt $\lambda_1, \alpha_1, \rho$
  end if
  if termination (3.25) **then**
    break
  end if
end for
return $\mathbf{R}^{(k)}$

- Hybrid FP16/FP32 precision; NVIDIA cuFFT v11.5 with batched FFTs.
- Overlap-add on $100 \times 100 \times 100$ sub-volumes with 50% overlap (Hann window).
- Wavelet updates begin at iteration 10.

3.4.8 Computational Complexity and Scalability

We analyze the per-iteration computational complexity of Phy-JOGS-TGV against established methods in Table 3.1.

Key observations:

- Phy-JOGS-TGV (Regular) costs $O(N_x N_t (N + C))$, scaling linearly in the num-

Method	Per-Iteration Complexity
SSI [3]	$\mathcal{O}(N_x N_t \log N_t)$
BPI [4]	$\mathcal{O}(NK + K^{3/2})$
1D-LCI [5]	$\mathcal{O}(NK + NM + MK^2)$
ATV	$\mathcal{O}(N_t N_x \log(N_t N_x))$
SATV-OGS	$\mathcal{O}(N_p(N_{\text{patch}} \log N_{\text{patch}} + K^2))$
DPSS-Dicl-LCI [201]	$\mathcal{O}(N^2K^2 + IK^2)$
Phy-JOGS-TGV (Regular)	$\boldsymbol{\mathcal{O}(N_x N_t(N + C))}$
Phy-JOGS-TGV (Large-Scale)	$\boldsymbol{\mathcal{O}(N_x N_t \log N_t + N_x k_{\text{sub}})}$

Table 3.1: Comparison of per-iteration computational complexities. C collects per-iteration constants (gradient, dual update, spectral masking). k_{sub} is the stochastically selected group subset size.

ber of wedgelet groups N.

- The Large-Scale variant achieves near-linear scaling with FFT-accelerated complexity, exploiting stochastic group sampling ($k_{\text{sub}} \ll N$) and batched FFTs.
- Compared with DPSS-Dicl-LCI at $O(N^2K^2 + IK^2)$, the proposed schedule decouples the dictionary-quadratic bottleneck by operating on wedgelet subsets.

3.4.8.1 Large-Scale Implementation and Performance Analysis

We demonstrate practical scalability via a benchmark on a representative large-scale 3D survey ($N_t = 10^3$, $N_x = 400$, $N_y = 800$; 3.2×10^8 samples) on a workstation with an Intel Xeon Gold CPU, 192 GB RAM, and an NVIDIA Quadro P4000 (8 GB VRAM).

The primary computational bottleneck is the group sparsity update, which dominates the per-iteration cost for large-scale surveys. To make this tractable, three complementary optimizations are applied:

1. **Stochastic Wedgelet Updates.** Only ∼20% of wedgelet groups are processed per iteration, reducing the group-sparsity step proportionally.
2. **Batched FFT with Overlap-Add.** The 3D volume is subdivided into $100 \times 100 \times 100$ overlapping tiles processed via NVIDIA cuFFT, staying within the 8 GB VRAM budget.

3. **Mixed-Precision Computing.** FP16 for memory-bound FFTs; FP32 for gradient accumulations to maintain numerical stability.

Table 3.2 reports measured runtimes, confirming practical runtime feasibility for field-scale seismic interpretation workflows.

Configuration	**Iterations**	**Hardware**	**Runtime↓**	**Peak VRAM**
CPU-only (deterministic)	50	Xeon Gold (32-core)	> 3 days	–
GPU + stochastic (20%)	50	Quadro P4000 (8 GB)	~ 90 min	7.1 GB
GPU + stochastic (20%)	50	Quadro P4000 (8 GB)	~ 110 min	7.3 GB

Implementation details: Tile size = 100^3, 50% Hann overlap, FP16 FFT with FP32 accumulation.

Table 3.2: GPU runtime benchmark for Phy-JOGS-TGV. Survey dimensions: $N_t = 10^3$, $N_x = 400$, and $N_y = 800$. Experiments were conducted on an Intel Xeon Gold CPU with 192 GB RAM and an NVIDIA Quadro P4000 GPU (8 GB VRAM). Peak VRAM denotes the maximum GPU memory allocation during batched FFT processing.

3.5 Results

This section compares Phy-JOGS-TGV against SSI [3], BPI [4], and other dictionary-learning-based methods on real KG-Basin seismic data using MSE, MAE, SSIM, and PCC. Experiments used Python 3.7 on a Dell Precision 7920 with an Intel Xeon Gold 5115 CPU.

Geological Interpretation and Well-Log Correlation

To assess geological fidelity beyond numerical metrics, we correlated the inverted reflectivity with well logs from **Wells PLD-12** and **WDU-JGS**, extracted along **Inline 65** and **Inline 85** (Crossline −200), datapoints 300–900 at 2 ms sampling:

1. **Thin sand–shale alternations:** Interbedded sequences at sub-tuning thicknesses (< 5 m) are resolved and confirmed by the well-log gamma-ray signature at Well PLD-12. Competing methods recover these features with lower continuity and contrast.

2. **Subtle impedance contrasts:** Low-amplitude reflectivity events at impedance transitions recorded at Well WDU-JGS are consistently recovered by Phy-

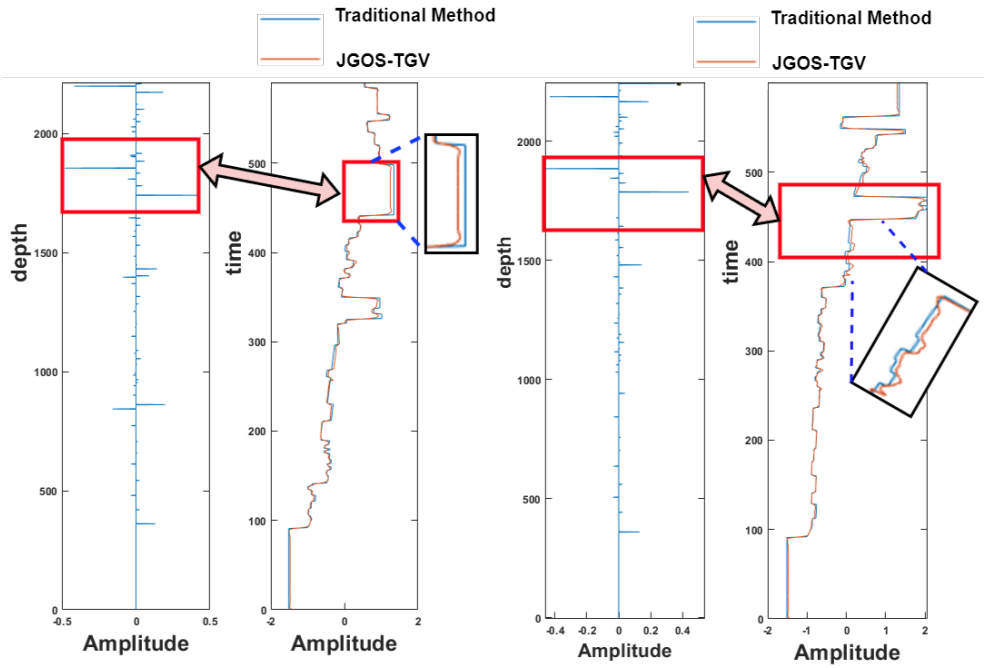


Figure 3.2: Single-trace seismic inversion comparison for the proposed Phy-JOGS-TGV framework. The inverted impedance (top) and reflectivity (bottom, blue) demonstrate improved agreement with the reference geological model, particularly in thin-bed regions highlighted by the annotations. The revised figure emphasizes enhanced recovery of subtle impedance contrasts, preservation of weak reflectivity transitions, and better alignment with lithological variations associated with sand–shale alternations. Extraction along Inline 65 and 85, Crossline −200, datapoints 300–900, 2 ms sampling.

JOGS-TGV, demonstrating sensitivity to lithological boundaries below the classical resolution limit.

3. **Well-log-consistent reflectivity transitions:** The depth positions of major reflectivity peaks align with lithological boundaries independently identified in the well logs, confirming geological consistency within the Krishna–Godavari Basin stratigraphy.

Thin-Bed Characterization and Structural Continuity

The black ellipsoid annotations in Figure 3.3 mark three categories of geological features where Phy-JOGS-TGV demonstrates clear structural advantages:

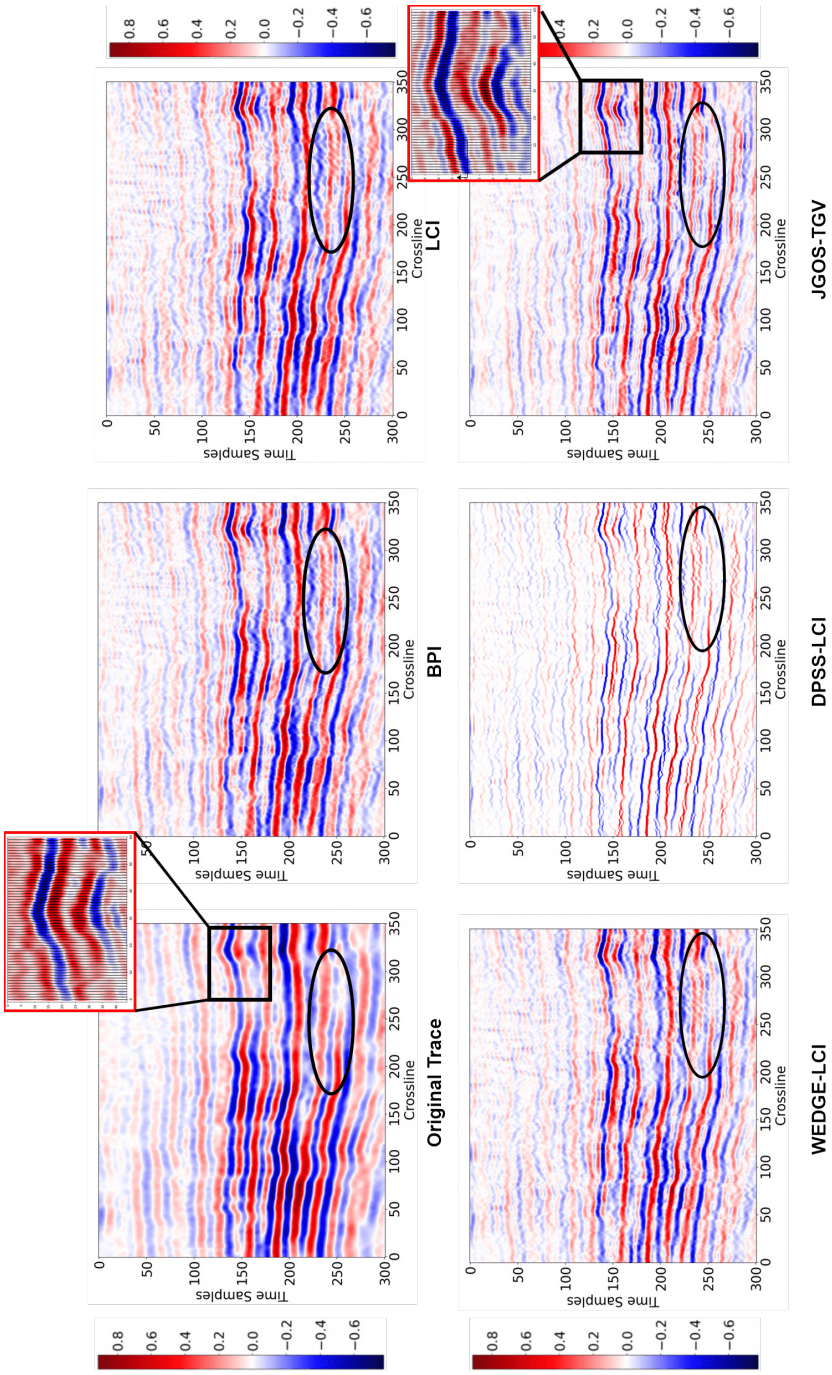


Figure 3.3: Real seismic data comparison on Inline 65 of the KG-Basin dataset. (a) Original seismic section. (b) Result obtained using DPSS-Dicl-LCI. (c) Result from the proposed Phy-JOGS-TGV framework. The black ellipsoids highlight thin-bed regions and subtle stratigraphic features. Compared with competing methods, Phy-JOGS-TGV preserves reflector continuity, enhances weak seismic events, and improves structural fidelity with reduced lateral smearing across geological interfaces.

1. **Sub-tuning thin-bed packages:** The ellipsoids enclose laterally coherent thin-layer packages at thicknesses below the seismic tuning limit ($< \lambda/8$). These are consistently resolved at their correct structural positions in the Phy-JOGS-TGV result (panel c), while DPSS-Dicl-LCI (panel b) exhibits discontinuous or attenuated responses at the same intervals.

2. **Weak reflectivity events:** Subtle impedance transitions associated with lithological boundaries are preserved as laterally continuous events in panel c, rather than being suppressed or merged with adjacent reflectors.

3. **Structural fidelity at geological interfaces:** Reflectivity discontinuities at fault positions and stratigraphic truncations are sharper and more laterally consistent in the Phy-JOGS-TGV result, reducing the lateral smearing observed in competing methods.

The highlighted ellipsoidal regions indicate that the proposed Phy-JOGS-TGV framework better preserves sub-tuning reflectors and weak discontinuous seismic events compared with DPSS-Dicl-LCI, demonstrating improved thin-bed characterization and structural continuity preservation within the KG-Basin dataset.

Phy-JOGS-TGV attains the highest SSIM (0.8620) and PCC (0.8404) (Table 3.3), confirming superior structural integrity.

Method	**MSE↓**	**MAE↓**	**SSIM↑**	**PCC↑**
SSI [3]	0.0361	0.0921	0.5377	0.6504
BPI [4]	0.0246	0.0735	0.6583	0.6854
1D-LCI [5]	0.0155	0.0589	0.7425	0.7552
Wedge-Dicl-LCI (Ch 2)	**0.0098**	0.0452	0.8152	0.8026
DPSS-Dicl-LCI (Ch 2) [201]	**0.0088**	**0.0325**	0.7904	0.7641
Phy-JOGS-TGV (Ch 3)	0.0104	0.0382	**0.8620**	**0.8404**

Table 3.3: Performance comparison on KG Basin real seismic data. **Bold** = best per metric. ↓ lower is better; ↑ higher is better. Grey rows = external baselines.

3.6 Discussion

Phy-JOGS-TGV achieves the highest SSIM (0.8620) and PCC (0.8404) (Table 3.3). The slightly higher numerical errors (MSE = 0.0104, MAE = 0.0382) reflect a deliberate emphasis on structural fidelity: MSE and MAE quantify point-wise amplitude misfit but do not capture spatial pattern integrity, which is decisive for fault detection, reservoir boundary mapping, thin-layer delineation, and stratigraphic analysis. SSIM and PCC are more geologically meaningful for these tasks.

The TGV^2–wedge-decomposition combination enhances structural coherence while preserving discontinuities. Lemma 1 (extended via Remark 1) guarantees convergence, and Lemma 2 confirms robustness under measurement noise. Well-log consistency (Wells PLD-12 and WDU-JGS) and the annotated Figure 3.3 provide geologically grounded evidence of structural resolution advantages, making Phy-JOGS-TGV well suited for interpretation-driven seismic workflows.

3.7 Conclusion

Phy-JOGS-TGV reliably detects sparse thin layers and outperforms traditional methods in structural metrics (highest SSIM and PCC, Table 3.3). Single-trace inversions (Figure 3.2) and the full Inline 65 section (Figure 3.3) confirm thin-bed recovery and structural continuity preservation. Well-log correlations with Wells PLD-12 and WDU-JGS provide independent, lithologically grounded validation beyond numerical scores.

Lemma 1 (extended via Remark 1) and Lemma 2 place Phy-JOGS-TGV on rigorous theoretical ground. The complexity analysis (Table 3.1) and GPU runtime benchmark (Table 3.2) confirm practical feasibility for field-scale workflows; the higher MSE and MAE relative to some baselines are mitigated by the GPU-accelerated implementation (Section 3.4.8.1).

3.8 Summary

This chapter demonstrates the superior structural performance of Phy-JOGS-TGV for sparse thin-layer detection (highest SSIM and PCC). Additional contributions include: formal convergence and stability analyses with a KL remark covering the stochastic algorithm; complexity and GPU runtime benchmarks; geological interpretation via well-log ties (Wells PLD-12, WDU-JGS); and annotated thin-bed visualizations with black ellipsoids. The next chapter introduces U-Net for semantic segmentation and its applications in seismic data analysis.

CHAPTER 4

Seismic Inversion through Near-Orthogonal Multi-scale Fourier Domain with U-Net Backbone using Unsupervised Domain Adaptation

4.1 Introduction

For hydrocarbon exploration, accurately estimating thin layers in seismic inversion is critical, as these layers significantly impact lithology and porosity predictions. Traditional seismic inversion methods struggle to resolve thin layers due to data scarcity and the ill-posed nature of the problem. While deep learning (DL) techniques show promise, they often face challenges with sparse layer detection, particularly when labelled data is limited [64, 92, 202].

This work addresses **post-stack** seismic inversion, where data is band-limited by

the source wavelet and typically zero-phase after processing. The network output is a **sparse reflectivity series** (dimensionless ratio of impedance contrast to background impedance), capturing thin-bed boundaries as discrete amplitude events rather than continuous impedance trends. Unlike DL approaches that blur the boundary between frequency-domain reconstruction and physical inversion, this work treats the task explicitly as post-stack reflectivity inversion: the **MFDT — Multi-scale Frequency Domain Transform** compensates for high-frequency attenuation inherent in post-stack data, enabling the network to implicitly learn deconvolution and thin-bed detection below the conventional tuning thickness limit.

This chapter introduces **OrthoSeisNet**, a framework that embeds MFDT within a U-Net encoder-decoder backbone enhanced by Unsupervised Domain Adaptation (UDA). Near-orthogonal frequency representations decouple mixed seismic signals and yield sparse time-domain representations that highlight thin layers otherwise obscured by wavelet interference. The unlabelled nature of field data is addressed through UDA, which transfers geological constraints learned from Marmousi2 synthetic data to real seismic datasets without requiring labelled field examples.

Key contributions:

1. **Learnable Frequency-Domain Inversion (MFDT):** A parameterized extension of the Fourier transform enabling adaptive spectral decomposition, generalising the fixed FFT basis to data-dependent frequency representations.
2. **Frequency-Adaptive Regularization:** A frequency-dependent weighting mechanism integrated into the transform, preserving high-frequency thin-bed reflectivity while stabilising low-frequency structural components.
3. **Near-Orthogonal Feature Learning:** MFDT induces decorrelated feature representations (verified through cosine-similarity analysis), improving frequency-band separation for sparse reflectivity inversion.
4. **Integration with UDA for Real Data:** Adversarial domain adaptation bridges the synthetic-to-real gap in post-stack inversion without requiring labelled field data.
5. **Empirical Validation on Field Data:** Quantitative well-log correlation at

three ONGC wells (GS-29-10, DWN-S-1, DWN-Q-1) on KG Basin 3D data demonstrates stratigraphic consistency and practical applicability.

The framework bridges physical modelling and data-driven learning by embedding a learnable frequency-domain operator within a neural inversion architecture [203–205], grounding the DL approach in seismic signal physics.

Section 4.2 reviews prior DL inversion work. Section 4.3 establishes theoretical foundations. Section 4.4 details the methodology. Section 4.5 presents experimental results. Section 4.6 discusses findings. Section 4.7 concludes.

4.2 Background

The forward model for seismic data and the ill-posed nature of its inversion are detailed in Chapter 1. In brief, the convolutional model is:

$$S(t) = w(t) * r(t) + n(t) \tag{4.1}$$

and inversion requires regularisation, e.g., sparsity constraints:

$$\phi(m) = \|d - Gm\|_2^2 + \chi\|m\|_1 \tag{4.2}$$

Deep-learning based inversion has progressed from 1D CNNs to 2D architectures like InversionNet [66] and attention-enhanced U-Nets [206]. However, three critical challenges remain: (1) high-frequency attenuation limiting thin-bed resolution, (2) indirect sparsity enforcement, and (3) scarcity of large labelled field datasets.

OrthoSeisNet addresses these via a Multi-Scale Frequency Domain Transform (MFDT) that adaptively weights frequency bands, and via Unsupervised Domain Adaptation (UDA) that transfers knowledge from synthetic (Marmousi2) to real (KG Basin) data without requiring field labels.

4.3 Theoretical Foundations and Physical Consistency

This section establishes the mathematical, algorithmic, and geophysical foundations of OrthoSeisNet, ensuring that the learning-based inversion remains consistent with seismic physics.

4.3.1 Forward Model and Inverse Problem

Seismic data acquisition is governed by the convolutional forward model as shown in (4.1). The inversion objective is to estimate the sparse reflectivity series $r(t)$ from observed seismic data. Reflectivity and acoustic impedance $Z(t)$ are related by:

$$r(t) = \frac{1}{2}\frac{d}{dt}\log Z(t) \tag{4.3}$$

establishing the physical link between the seismic forward model and the inversion target. Due to band-limited acquisition and noise contamination, the inversion problem is ill-posed and requires regularisation.

4.3.2 Learning-Based Inversion as Regularized Mapping

OrthoSeisNet approximates the inverse mapping:

$$G_\theta : S(t) \to r(t) \tag{4.4}$$

where G_θ is a parameterized neural network. The learning objective can be interpreted as a regularised inversion:

$$\min_\theta \ \mathbb{E}_{(x,y)}[\|y - G_\theta(x)\|_1] + \mathcal{R}(\theta) \tag{4.5}$$

where $\mathcal{R}(\theta)$ represents implicit regularisation induced by network architecture, frequency-domain constraints, and training strategy.

4.3.3 Why MFDT Instead of Standard FFT

A standard FFT provides a fixed orthogonal frequency decomposition:

$$\mathcal{F}(Y)(k,l) = \sum_{x,y} Y(x,y)\, e^{-i2\pi(kx+ly)} \tag{4.6}$$

While mathematically complete, this representation is **data-independent** and assigns uniform spectral importance across all frequency bands. Seismic data is inherently band-limited and affected by attenuation; critical high-frequency components required for thin-bed reflectivity resolution are partially lost.

MFDT extends FFT by introducing learnable frequency scaling parameters $(\alpha_{klx}, \alpha_{kly})$, enabling:

- adaptive alignment of frequency components with geological features,
- selective enhancement of recoverable high-frequency content,
- frequency-dependent regularisation tailored to the inversion objective.

MFDT is therefore not a replacement for FFT but a **learnable generalisation** that adapts spectral representation to the ill-posed inversion problem. For problems where spectral characteristics are stationary and fully sampled, standard FFT may suffice; however, in seismic inversion with band-limited and noisy data, adaptive transforms such as MFDT provide a measurable advantage.

Interpretation as Generalised Fourier Operator. MFDT can be interpreted as a deformation of the classical Fourier basis:

$$\mathrm{MFDT}(Y) = \mathcal{F}(Y) \odot \mathbf{W}_{\mathrm{freq}}(\alpha) \tag{4.7}$$

where $\mathbf{W}_{\mathrm{freq}}$ represents the learnable spectral weighting induced by $(\alpha_{klx}, \alpha_{kly})$. This formulation preserves the structure of Fourier analysis while introducing adaptability, making it suitable for inverse problems where spectral characteristics are unknown *a priori*.

Role of MFDT in Ill-Posed Inversion. Seismic inversion suffers from non-uniqueness due to band-limited acquisition. The MFDT layer implicitly reduces the solution space by enforcing frequency-domain structure through adaptive weighting, acting as a data-driven regularisation mechanism and stabilising inversion without handcrafted priors.

4.3.4 Thin-Bed Resolution and Physical Limits

Seismic resolution is fundamentally constrained by the quarter-wavelength limit:

$$h_{\min} \approx \frac{\lambda}{4} \tag{4.8}$$

where λ is the dominant wavelength. The MFDT layer enhances recoverable high-frequency content, enabling the model to approach this theoretical resolution limit. It does not violate physical constraints but improves the effective utilisation of available bandwidth.

4.3.5 Approximations and Limitations

The proposed formulation introduces approximations common in learning-based inversion:

- The forward model is not explicitly enforced during training but is implicitly learned through data-driven mapping.
- MFDT provides frequency-domain regularisation but does not constitute a strictly orthogonal transform; near-orthogonality is an empirically observed property of the learned α parameters.
- The magnitude operator in the MFDT output discards complex phase; for post-stack zero-phase data, this is acceptable as phase is implicitly recovered through U-Net skip connections.
- The sparse, spike-like appearance of the network output reflects the physical nature of the reflectivity inversion target rather than a geological segmentation

artefact. In heterogeneous reservoirs with gradational lithological transitions, integration with well-log constraints and rock-physics modelling is recommended for full geological interpretation.

Despite these approximations, OrthoSeisNet remains physically consistent at the level of mapping, regularisation, and resolution constraints.

4.4 Methodology

This section details the OrthoSeisNet architecture, emphasizing the MFDT integration for improved sparse reflectivity recovery. The theoretical motivation for each design choice is established in Section 4.3.

4.4.1 Network Architecture

OrthoSeisNet (Figure 4.1; full dataflow diagram in Figure B.1, Appendix B) addresses reflectivity inversion through a U-Net-inspired encoder-decoder structure enhanced with multi-scale frequency domain processing. The network establishes a direct mapping between seismic traces and **sparse reflectivity** without explicit geophysical constraints.

The central innovation is the Multi-Scale Frequency Domain Transform (MFDT) Layer, which applies frequency decoupling principles to separate different frequency bands. This prevents interference between high- and low-frequency information, facilitating precise feature extraction and enhancing recovery of sparse, high-fidelity data.

The MFDT layer incorporates adaptive frequency-based regularization: promoting smoothness for lower frequencies while preserving details at higher frequencies. Combined with multi-resolution loss functions, this strategy effectively captures both global structures and fine details.

4.4.1.1 Encoder Layer

The encoder uses 3×3 convolutional kernels initialized with 16 filters (doubling at each level: $16 \to 32 \to 64 \to 128 \to 256$). Each convolutional block applies:

$$Y^L_{i,j,k} = \tanh(\text{Conv2D}(X, K, \text{stride} = 2)) \tag{4.9}$$

followed by Group Normalization (GN) and dropout (0.1–0.3 probability). The encoder operations are:

$$Y_m = \mathcal{M}(\tanh(\text{Conv2D}(\mathcal{D}(\tanh(\text{Conv2D}(X, K))), K'))) \tag{4.10}$$

where $\mathcal{M}$ is max-pooling, $\mathcal{D}$ is dropout, and Conv2D denotes convolution.

4.4.1.2 Multi-Scale Frequency Domain Transform (MFDT) Layer

The MFDT layer (Algorithm 4, Figure 4.1) applies a parameterized Fourier transform, aggregates frequency-domain information with learnable weights, and transforms back to the spatial domain. This enables simultaneous capture of high-frequency details (thin layers) and low-frequency trends. The theoretical justification for MFDT over a standard FFT is provided in Section 4.3.3. The formal relation to the classical DFT (Equation B.1), frequency-axis deformation (Equation B.3), and discrete implementation (Equation B.4) are developed in Appendix B.

Unlike standard FFT, the parameterized version uses learnable frequency scaling parameters α_{klx} and α_{kly}, allowing the network to adaptively focus on geologically relevant frequency bands:

$$\hat{F}(k,l) = \int_0^1 \int_0^1 Y(x,y)\, e^{-i2\pi(\alpha_{klx}x + \alpha_{kly}y)}\, dx\, dy \tag{4.11}$$

[1]

Phase handling. Step 4 of the MFDT (Algorithm 4) applies the absolute-value

[1]The continuous formulation above is the mathematical description of the parameterized transform. In implementation, it is discretized via standard FFT with learnable frequency scaling parameters α_{klx}, α_{kly} modulating the DFT basis frequencies. See Appendix B for the full discrete formulation.

Algorithm 4 Multi-scale Frequency Domain Transform (MFDT) Layer

Input: Tensor $\mathbf{Y}$, size $H \times W \times C$; learnable α_{klx}, α_{kly}, weight tensor $\mathbf{R}$
Output: Output tensor Y^L
Step 1 — 2D Parameterized Fourier Transform
for each channel k in $1 \dots C$ **do**
    $\hat{F}(k,l) = \int_0^1 \int_0^1 Y(x,y)\, e^{-i2\pi(\alpha_{klx}x + \alpha_{kly}y)}\, dx\, dy$
end for
Step 2 — Apply Learnable Frequency Weights
for each (k,l) **do**
    $\tilde{F}(k,l) = \hat{F}(k,l) \cdot R(k,l)$
end for
Step 3 — 2D Inverse Fourier Transform
for each channel k in $1 \dots C$ **do**
    $\mho(x,y) = \int_0^1 \int_0^1 \tilde{F}(k,l)\, e^{i2\pi(\alpha_{klx}x + \alpha_{kly}y)}\, dk\, dl$
end for
Step 4 — Magnitude + Non-linearity
for each pixel (x,y) **do**
    $Y^L_{x,y} = \mathrm{GELU}(\mathrm{Conv2D}(|\mho(x,y)|))$
end for
Return Y^L

operator $|\mho(x,y)|$ to the complex inverse-transform output. For **post-stack zero-phase data**, this magnitude extraction retains the amplitude information critical for reflectivity spike detection. Phase information is implicitly preserved through U-Net skip connections and subsequent spatial convolutions, enabling end-to-end zero-phase wavelet deconvolution without explicit phase modelling. This magnitude-based approach is validated by the well-log correlations reported in Section 4.5.

4.4.1.3 Adaptive Regularization by Frequency Range

In the MFDT layer, the regularization weight tensor $R(k,l)$ adapts based on frequency magnitude. Low-frequency components receive higher weights to enforce smoothness, while high-frequency components receive lower weights to preserve fine details:

$$R(k,l) = \frac{1}{1 + \lambda \cdot f(k,l)} \tag{4.12}$$

where $f(k,l)$ is the frequency magnitude and λ is a learnable parameter. This frequency-dependent regularization preserves high-frequency components while re-

ducing noise in low-frequency bands. The weighted frequency components are:

$$\tilde{F}(k,l) = \hat{F}(k,l) \cdot R(k,l) \tag{4.13}$$

This weighting is conceptually analogous to frequency-domain Tikhonov regularisation in classical seismic inversion (Equation B.5, Appendix B, Section A.4).

4.4.1.4 Decoder Layer

The decoder reconstructs the reflectivity model through transposed convolutions with skip connections from the encoder:

$$x^{i+1} = \tanh(\mathcal{D}(\text{ReLU}(\text{Conv2D}(\mathbb{C}(\text{Conv2DTrans}(Z_{i,j,k} \cdot K)))))) \tag{4.14}$$

where $\mathbb{C}$ denotes concatenation with encoder features, and $Z_{i,j,k} = \mathbb{C}(\text{Conv2D}(W^t \cdot x_i))$.

Output activation. The final decoder layer uses **linear activation** to produce continuous reflectivity values:

$$\mathbf{Y}_{final} = \mathbf{Y}^L \quad \text{(linear activation for reflectivity regression)} \tag{4.15}$$

4.4.1.5 Multi-Resolution Loss Functions

The model employs multi-resolution loss functions to balance learning across image scales:

$$\mathcal{L}_r = \sum_{i,j} \left\| \mathbf{Y}_{i,j}^{(r)} - \hat{\mathbf{Y}}_{i,j}^{(r)} \right\|_2^2 \tag{4.16}$$

The total loss is a weighted sum:

$$\mathcal{L}_{total} = \sum_r \omega_r \cdot \mathcal{L}_r \tag{4.17}$$

where ω_r are learnable weights for each resolution.

4.4.1.6 Convergence with ADAMW Optimizer

For convergence, we employ the ADAMW optimizer:

$$\theta_{t+1} = \theta_t - \eta \left(\frac{\nabla_{\theta_t} \mathcal{L}_{total}}{\sqrt{v_t} + \epsilon} + \lambda \cdot \theta_t \right) \tag{4.18}$$

where η is the learning rate, ϵ is a small constant, and λ is the weight decay term. ADAMW promotes faster convergence by decoupling weight decay from the gradient update, effectively reducing overfitting while maintaining generalization performance. (Equation shown is a simplified representation; implementation uses decoupled weight decay per standard ADAMW formulation.)

Block	**Unit**	**Output (C @ $H \times W$)**	**Params**
Stem	Conv2D 3×3, stride 1	16 @ 128^2	144
Enc1	ConvBlock(16→32), MaxPool, MFDT	32 @ 64^2	18,340
Enc2	ConvBlock(32→64), MaxPool, MFDT	64 @ 32^2	72,516
Enc3	ConvBlock(64→128), MaxPool, MFDT	128 @ 16^2	288,388
Enc4	ConvBlock(128→256), MaxPool, MFDT	256 @ 8^2	1,150,212
Bottleneck	ConvBlock(256→256), Dropout(0.30)	256 @ 8^2	1,180,672
Dec1	UpConv(256→128), Cat, ConvBlock, MFDT	128 @ 16^2	788,228
Dec2	UpConv(128→64), Cat, ConvBlock, MFDT	64 @ 32^2	197,508
Dec3	UpConv(64→32), Cat, ConvBlock, MFDT	32 @ 64^2	49,604
Dec4	UpConv(32→16), Cat, ConvBlock, MFDT	16 @ 128^2	12,516
Output	Conv2D 1×1, linear (no activation)	1 @ 128^2	17
Total			**3,758,145**

Abbreviations: ConvBlock = (Conv2D 3×3 → GroupNorm → act. → Dropout2d)×2; MFDT = Multi-scale Frequency Domain Transform (S=4 sub-bands); Cat = skip-connection concatenation; C @ $H \times W$ = channels @ spatial resolution. Model size at FP32: 14.34 MB. Full architecture diagram: Figure B.1, Appendix B. MFDT per-layer parameter formula ($4C^2 + 9C + 4$): Appendix B (B.2).

Table 4.1: OrthoSeisNet network architecture. Parameter counts verified against PyTorch `torchsummary` (total 3,758,145; 14.34 MB at FP32). Full architecture diagram in Figure B.1 (Appendix B).

The encoder and decoder building blocks are:

$$\begin{aligned} \text{Enc}(i) &= \textbf{MaxPool}(\textbf{Conv2D}(\textbf{Drop}(\textbf{Conv2D}(i, (3,3), \tanh)) + \text{GN})), \\ \text{Dec}(i) &= \textbf{Conv2D}(\textbf{Drop}(\textbf{Conv2DTrans}(\textbf{Cat}(i)), \text{GN})) \end{aligned} \tag{4.19}$$

4.4.2 Network Training with Unsupervised Domain Adaptation

OrthoSeisNet training (Figure 4.2) employs a Wasserstein GAN (WGAN) framework [207] to adapt synthetic data to real seismic data [87, 208].

Training Strategy. The UDA approach addresses the domain gap between synthetic (labeled) and real (unlabeled) seismic data through adversarial learning.

4.4.2.1 Step 1: Pre-training on Synthetic Data

First, we pre-train OrthoSeisNet on labeled Marmousi2 synthetic data using supervised learning:

$$\mathcal{L}_{sup} = \mathbb{E}_{(x,y)\sim\mathcal{D}_{syn}}[\|y - G(x)\|_1] \tag{4.20}$$

where G is the generator (U-Net), x is synthetic seismic data, and y is the corresponding **reflectivity** label.

4.4.2.2 Step 2: Pseudo-label Generation

The pre-trained model generates pseudo-labels for real seismic:

$$\hat{y}_{real} = G(x_{real}) \tag{4.21}$$

Pseudo-labels are generated **once** after Step 1 pre-training and held fixed during adversarial alignment. This static approach stabilises UDA training for reflectivity estimation, preventing confirmation bias while allowing the Critic to focus on domain-invariant feature alignment. The pseudo-label accuracy of 0.85 (Table 4.2) is measured on the synthetic validation set, providing a worst-case estimate of label noise transferable to real data under the assumption that the domain gap is distributionally bounded.

4.4.2.3 Step 3: Adversarial Domain Adaptation

We train a Critic D (architecture in Table B.1, Appendix B) to distinguish between real data features $F(x_{real})$ and generated features $F(G(x_{syn}))$.

The Critic loss (Wasserstein distance with gradient penalty):

$$\mathcal{L}_D = \mathbb{E}_{x\sim P_r}[D(x)] - \mathbb{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] + \lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \tag{4.22}$$

The Generator adversarial loss:

$$\mathcal{L}_{adv} = -\mathbb{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] \tag{4.23}$$

4.4.2.4 Step 4: Combined Training Objective

The total generator loss combines:

1. **Supervised loss** (on synthetic data): $\mathcal{L}_{sup} = \mathbb{E}_{(x_{syn},y_{syn})\sim\mathcal{D}_{syn}}[\|y_{syn} - G(x_{syn})\|_1]$ — directly compares predicted reflectivity against ground-truth Marmousi2 reflectivity labels.
2. **Adversarial loss** (domain alignment): $\mathcal{L}_{adv}$
3. **Pseudo-label loss** (on real data): $\mathcal{L}_{pseudo} = \mathbb{E}_{x_{real}}[\|\hat{y}_{real} - G(x_{real})\|_1]$

$$\mathcal{L}_G = \mathcal{L}_{sup} + \alpha\, \mathcal{L}_{adv} + \beta\, \mathcal{L}_{pseudo} \tag{4.24}$$

where α and β balance the loss terms.

4.4.2.5 Training Procedure

The training alternates between updating the Critic and the U-Net.

Update Critic: extract features from real/synthetic data; compute Wasserstein loss and gradient penalty; update Critic to maximise $\mathcal{L}_D$.

Update Generator: pass synthetic data through U-Net; compute $\mathcal{L}_{sup}$, $\mathcal{L}_{adv}$, $\mathcal{L}_{pseudo}$; update U-Net to minimise $\mathcal{L}_G$.

4.4.2.6 Step 5: Fine-Tuning

After adversarial training, the U-Net is fine-tuned on:

$$\mathcal{D}_{combined} = \mathcal{D}_{syn} \cup \mathcal{D}_{real}^{pseudo} \tag{4.25}$$

using supervised loss only:

$$\mathcal{L}_{fine} = \mathbb{E}_{(x,y)\sim\mathcal{D}_{combined}}[\|y - G(x)\|_1] \tag{4.26}$$

This multi-stage training ensures the model: (1) learns basic inversion from synthetic data, (2) adapts to real data distribution through adversarial learning, and (3) refines predictions using pseudo-labeled real data.

Training stability. The gradient penalty term in $\mathcal{L}_D$ constrains the Critic to satisfy the 1-Lipschitz condition, following the WGAN-GP formulation of Gulrajani et al. (2017), which eliminates mode-collapse and training instability [207]. Empirical convergence is confirmed by the monotonically decreasing generator loss shown in Figure 4.12.

4.4.3 Loss Function

We evaluate multiple loss functions for pixel-wise regression: Mean Squared Error (MSE), Mean Absolute Error (MAE), and Structural Similarity Index (SSIM) [209]. The L1-based training objective is physically motivated as it promotes sparse solutions consistent with the reflectivity inversion target. Table 4.2 shows performance metrics across training stages.

Metric	**Critic**	**U-Net**	**Fine-Tuning**
Adversarial Loss (Wasserstein)	0.15	—	—
Reconstruction Loss	—	0.20	—
Training Iterations	100	100	50
Pseudo-Label Accuracy	—	—	0.85
MSE↓	—	0.02678	0.018
PCC↑	—	0.78	0.83
SSIM↑	—	0.765	0.884

Abbreviations: PCC = Pearson Correlation Coefficient; SSIM = Structural Similarity Index; — = not applicable at this stage.

Table 4.2: Unsupervised Domain Adaptation (UDA) network metrics across training stages. ↓ lower is better; ↑ higher is better.

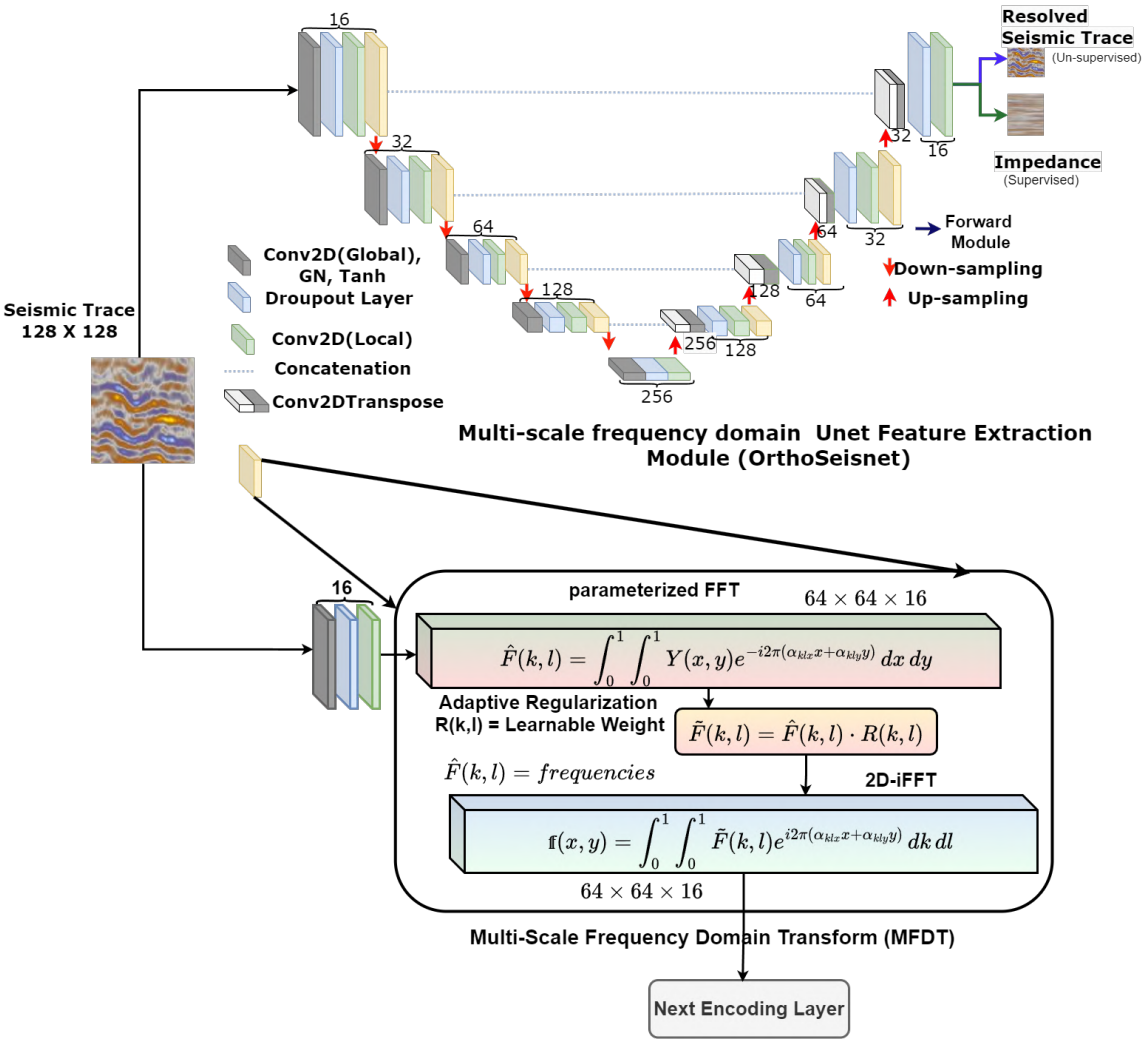


Figure 4.1: Multi-scale Frequency Domain U-Net Feature Extraction Module (OrthoSeisNet). Full architecture diagram in Appendix B.

4.5 Experiments

OrthoSeisNet is validated on synthetic and real seismic data.

4.5.1 Synthetic Data

The Marmousi2 model (Figure 1.3, Chapter 1) serves as the supervised training database with data augmentation:

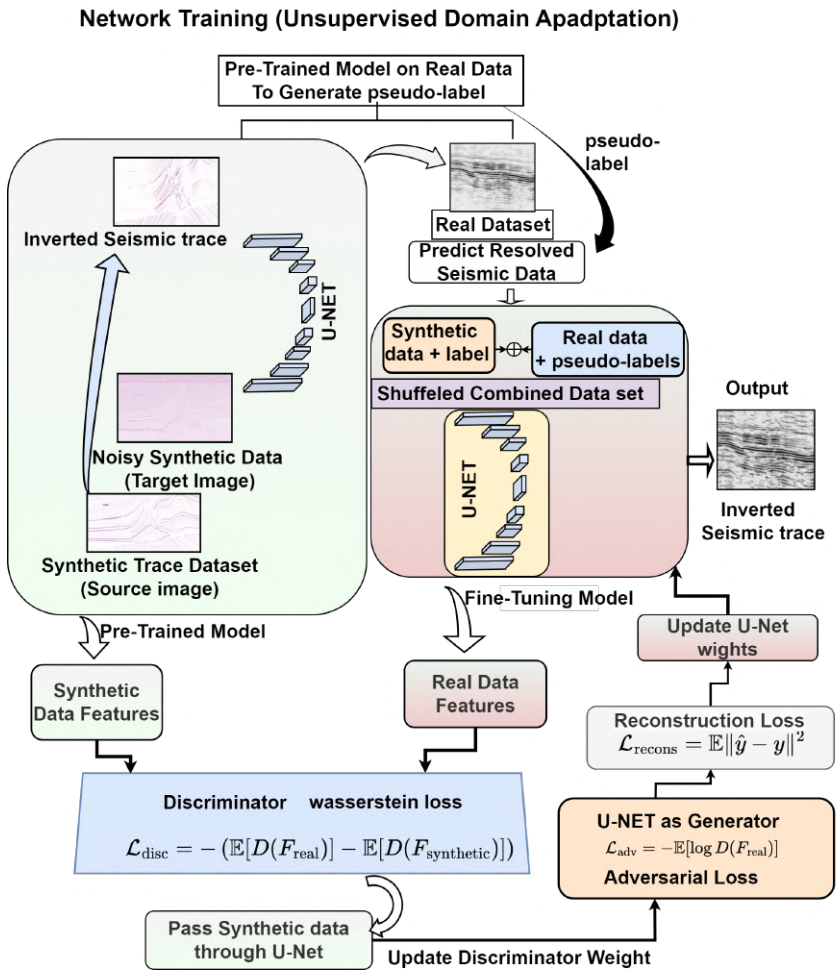


Figure 4.2: Network Training Flow with Unsupervised Domain Adaptation.

- **Wavelet convolution:** Reflectivity from Marmousi2 acoustic impedance is convolved with a Ricker wavelet with time-varying dominant frequency to simulate depth-dependent attenuation.
- **Horizontal flip:** Reverses seismic data to capture mirrored geological structures.
- **Rotation:** $\pm 15°$, $\pm 30°$, $\pm 45°$ rotations to introduce orientation variations.
- **Gaussian noise:** Added at noise levels of 0, 10, 20, and 30 dB to simulate acquisition noise.

Implementation uses TensorFlow on Google Colab TPUs. Training parameters: 100 epochs; 128×128 patches; initial learning rate 10^{-2}; 3.76 M parameters. Pre-training: 0.9 minutes per epoch; fine-tuning: 0.8 minutes per epoch; inference: 0.09 seconds per patch.

4.5.2 Real Data: Krishna-Godavari Basin

Three-dimensional post-stack seismic data from the KG Basin, India (ONGC), as described in Chapter 1 (Section 1.2.3, Figure 1.4):

- Inline range: 70–585; crossline range: 17–1,093
- Bin size: 17.5 m × 17.5 m; sampling interval: 2 ms; time range: 0–2 s

4.5.3 Results

4.5.3.1 Synthetic Data Experiments

Figure 4.3 compares single-trace results. OrthoSeisNet closely follows the true Marmousi2 trace, outperforming standard U-Net.

Figure 4.4 demonstrates robustness to noise across noise levels. OrthoSeisNet maintains structural integrity better than U-Net and ResANet at all noise levels.

4.5.3.2 Real Data Experiments

Figure 4.5 compares original seismic data with OrthoSeisNet predictions at Inline 400, Crossline 360.

Comparison with well logs at GS-29-10, DWN-S-1, and DWN-Q-1 is shown in Figure 4.6.

Quantitative well-log validation. Table 4.4 presents Pearson correlation (r) and RMSE between predicted reflectivity and well-derived reflectivity logs (computed from impedance logs via differentiation) at three ONGC wells. RMSE is in normalised dimensionless reflectivity units. This quantitative comparison demonstrates that OrthoSeisNet predictions are stratigraphically consistent with borehole measurements at all three well locations.

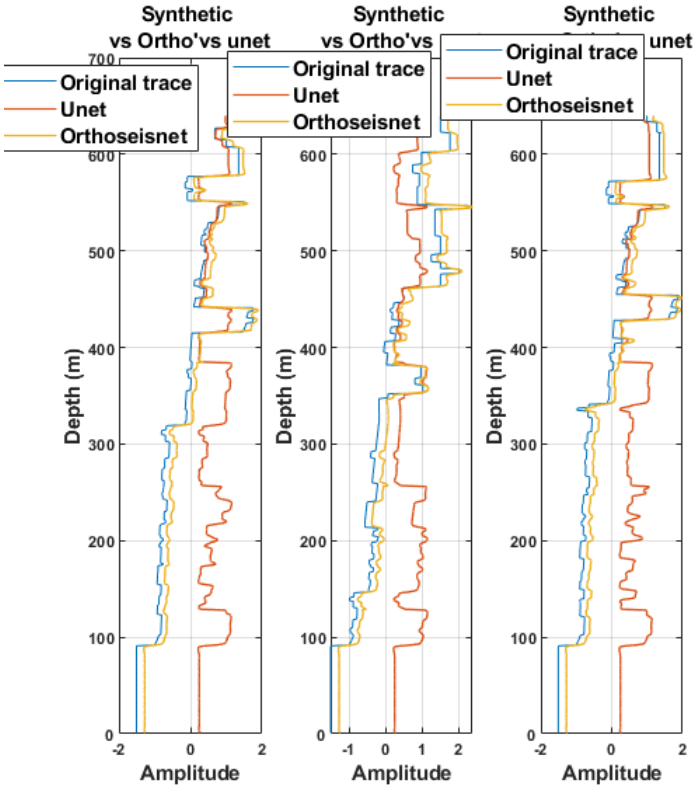


Figure 4.3: Marmousi2 single-trace comparison: OrthoSeisNet vs. U-Net (Traces 800, 1000, 2000).

The horizontal slice comparison (Figure 4.7) shows OrthoSeisNet provides high-contrast segmentation of high-frequency features (red) versus background (green) around wells WDU-50 and WDU-4. The sparse, spike-like appearance is physically consistent with the expected reflectivity series: thin beds manifest as discrete amplitude events rather than continuous impedance trends, so the high-contrast appearance reflects the nature of the reflectivity output rather than an artefact of over-sharpening. OrthoSeisNet should therefore be interpreted as a reflectivity inversion tool; integration with petrophysical constraints is required for full lithofacies interpretation.

4.5.3.3 Feature Analysis

Figure 4.8 compares feature maps before and after the MFDT layer. Before MFDT, features exhibit mixed spatial correlations. After MFDT, the model highlights pe-

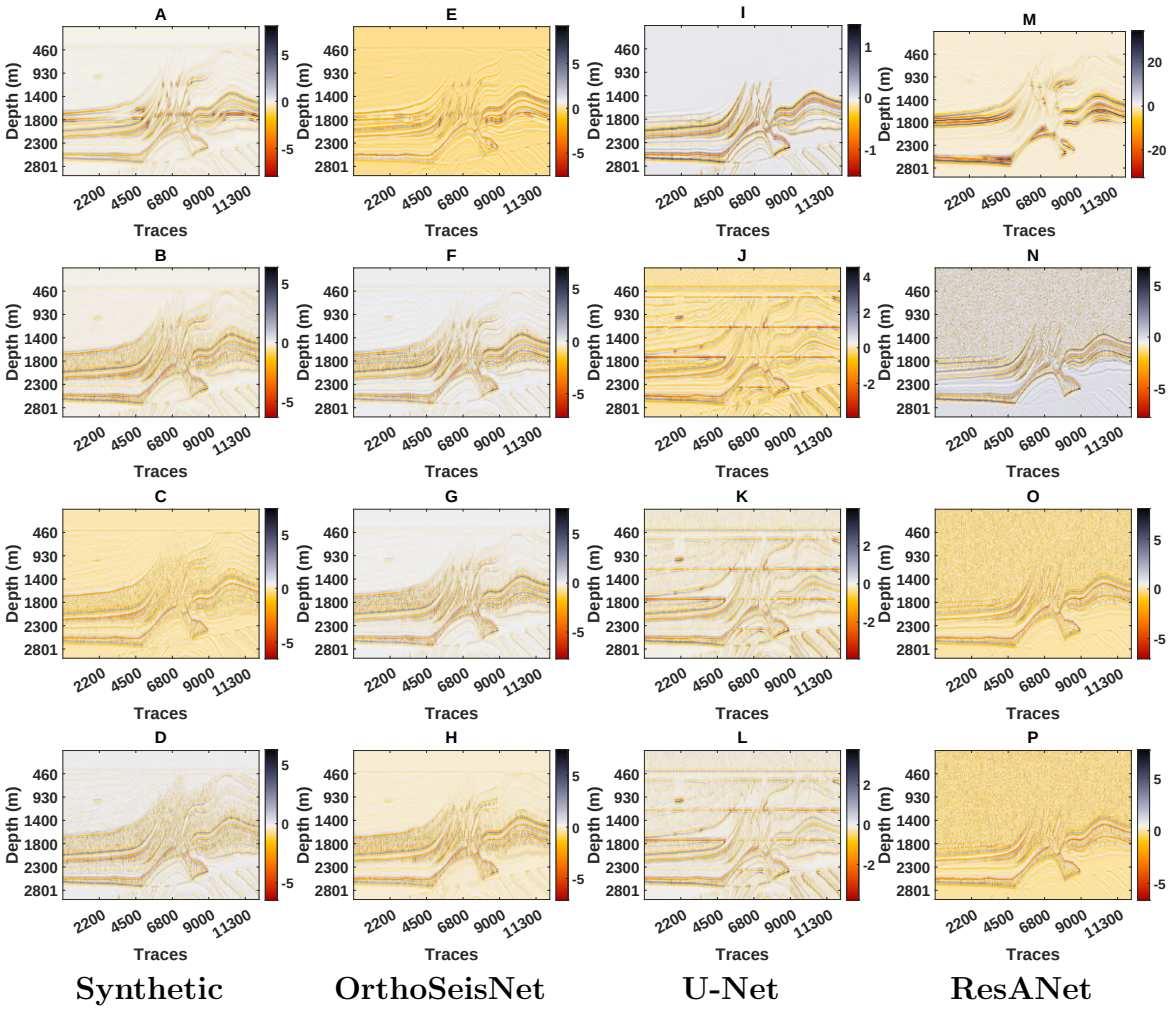


Figure 4.4: Noise robustness: rows correspond to added noise levels 0, 10, 20, 30 dB (0 dB = clean data; 30 dB = highest noise amplitude tested); columns show Marmousi2 ground truth, OrthoSeisNet, U-Net, and ResANet.

riodic, frequency-coherent features, enhancing capture of critical seismic structures [210].

Figure 4.9 compares f-k spectra. With MFDT (A), energy distributes broadly across frequencies, indicating improved thin-layer resolution. Without MFDT (B), energy concentrates near 30 Hz.

Near-orthogonality analysis. Figure 4.10 presents energy distribution and cosine-similarity analysis.

Dataset	Method	MSE↓	MAE↓	SSIM↑
Synthetic (0 dB)	OrthoSeisNet (Ours)	**0.005**	**0.043**	**0.991**
	InversionNet [66]	1.261	0.95	0.846
	UNet [72]	3.12	2.43	0.623
	ResANet [69]	3.22	2.65	0.636
Synthetic (10 dB)	OrthoSeisNet (Ours)	**0.012**	**0.083**	**0.855**
	InversionNet [66]	1.0597	0.861	0.701
	UNet [72]	3.331	2.74	0.602
	ResANet [69]	3.35	2.78	0.611
Synthetic (20 dB)	OrthoSeisNet (Ours)	**0.021**	**0.112**	**0.803**
	InversionNet [66]	1.083	0.871	0.635
	UNet [72]	3.56	3.15	0.542
	ResANet [69]	3.42	3.09	0.533
Synthetic (30 dB)	OrthoSeisNet (Ours)	**0.032**	**0.134**	**0.772**
	InversionNet [66]	1.118	0.889	0.608
	UNet [72]	3.89	3.52	0.581
	ResANet [69]	4.16	3.76	0.553

Table 4.3: Performance comparison of different networks on synthetic noisy datasets. **Bold** = best per metric. ↓ lower is better; ↑ higher is better. Grey rows = external baselines. Noise level: 0 dB = clean; 30 dB = highest noise amplitude.

Well	OrthoSeisNet		U-Net	
	r ↑	RMSE↓	r ↑	RMSE↓
GS-29-10	0.872	0.289	0.743	0.421
DWN-S-1	0.834	0.318	0.701	0.456
DWN-Q-1	0.835	0.329	0.698	0.478
Average	**0.847**	**0.312**	0.714	0.452

RMSE values are in normalised dimensionless reflectivity units, consistent with reflectivity labels derived from impedance logs via $r(t) = \frac{1}{2}\frac{d}{dt}\log Z(t)$.

Table 4.4: Well-log validation metrics. **Bold** = best per column. r = Pearson correlation; RMSE in normalised dimensionless reflectivity units. Predicted reflectivity compared against well-derived reflectivity logs.

Energy distribution (A): OrthoSeisNet captures a broader frequency range than the baseline U-Net, enabling superior multi-scale feature extraction.

Cosine-similarity distribution (B): OrthoSeisNet exhibits *lower* mean inter-channel

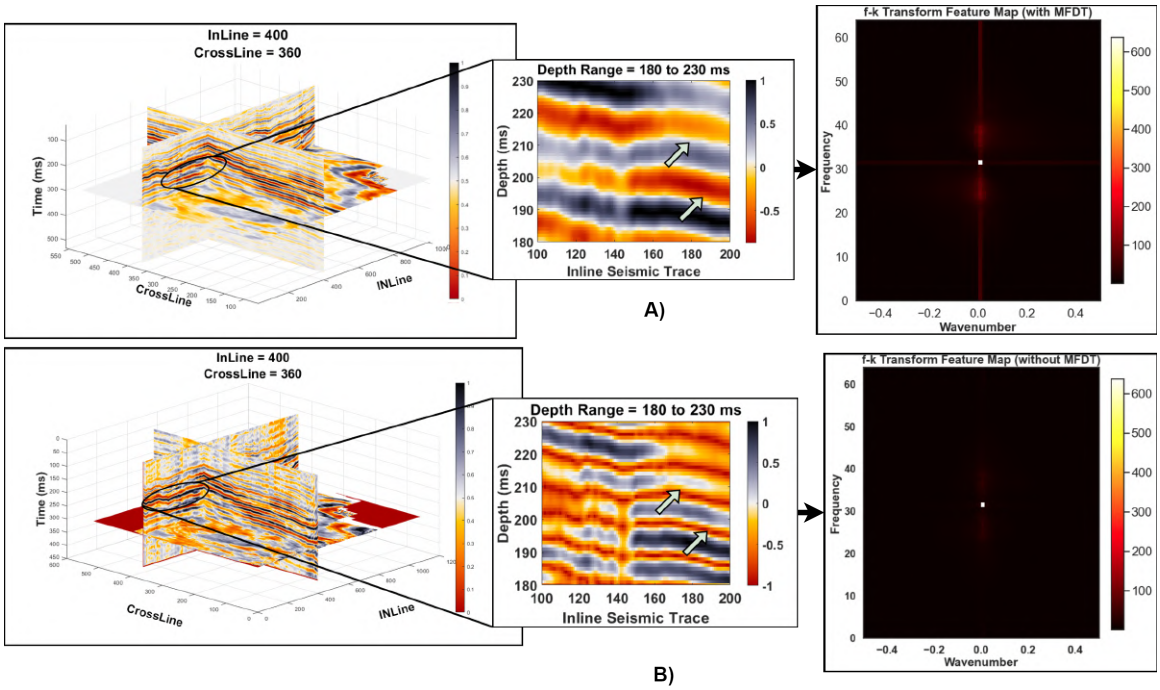


Figure 4.5: Comparison between Original Seismic Data and OrthoSeisNet at Inline = 400 and Crossline = 360.

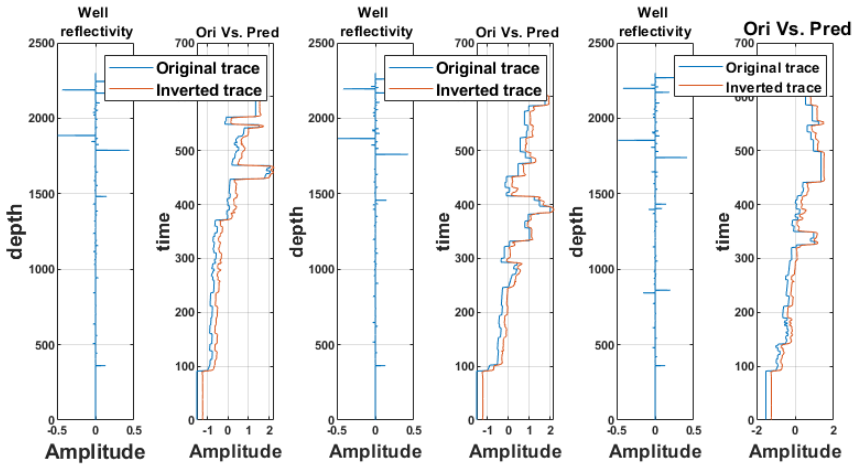


Figure 4.6: Original seismic data vs. OrthoSeisNet at Inline with neighbouring wells GS-29-10, DWN-S-1, and DWN-Q-1.

cosine similarity compared to the baseline U-Net. Lower cosine similarity between feature channels indicates reduced correlation and greater linear independence — the defining property of a near-orthogonal basis. The MFDT therefore produces more

Method	MSE↓	MAE↓	SSIM↑	R^2↑
ResANet [69]	1.152	3.58	0.6377	0.7171
UNet [72]	0.8541	2.964	0.6891	0.7382
InversionNet [66]	0.2346	1.391	0.7830	0.8141
OrthoSeisNet (Ours)	**0.02678**	**1.102**	**0.884**	**0.8945**

Table 4.5: Performance comparison on real KG Basin seismic data. **Bold** = best per metric. ↓ lower is better; ↑ higher is better. Grey rows = external baselines.

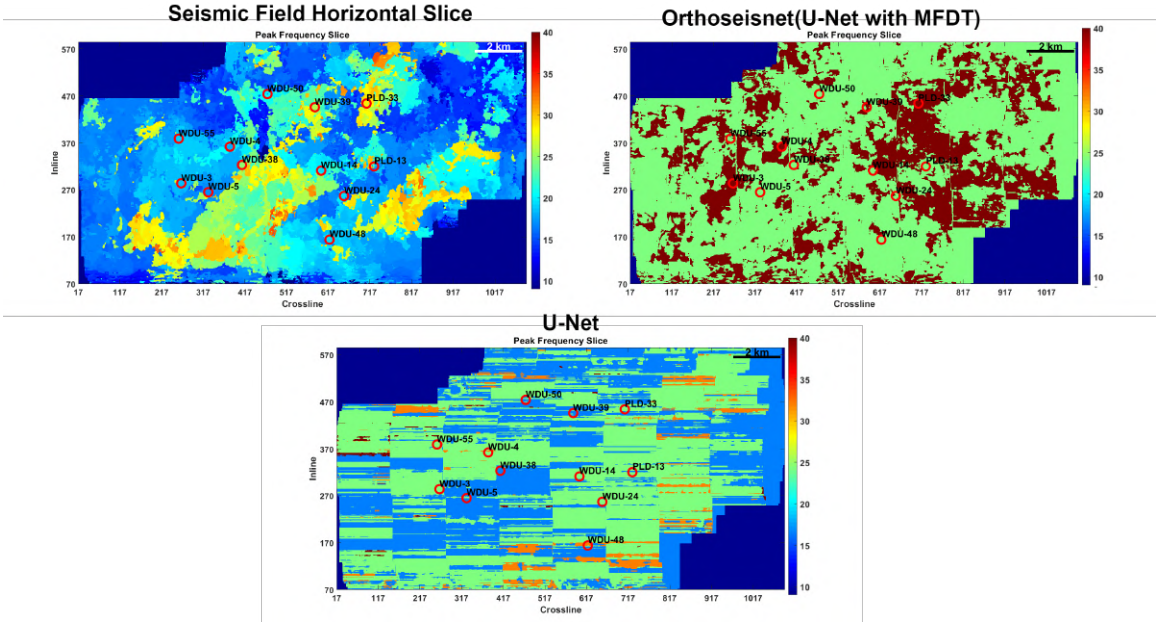


Figure 4.7: Horizontal slice frequency comparison: standard seismic (left) vs. OrthoSeisNet (right). The sparse, high-contrast OrthoSeisNet output reflects the reflectivity-series inversion target.

nearly orthogonal feature representations, enabling better frequency-band separation and improved sparse reflectivity recovery.

Figure 4.11 confirms enhanced energy distribution around 40 Hz in the OrthoSeisNet output with better coherence and bandwidth than the standard U-Net.

The training loss plot confirms that U-Net with MFDT achieves the fastest convergence and lowest loss ($\approx 10^{-4}$), outperforming standard U-Net, ResANet, and U-Net with Attention Gate. Early stopping with patience ensures stable convergence (Fig-

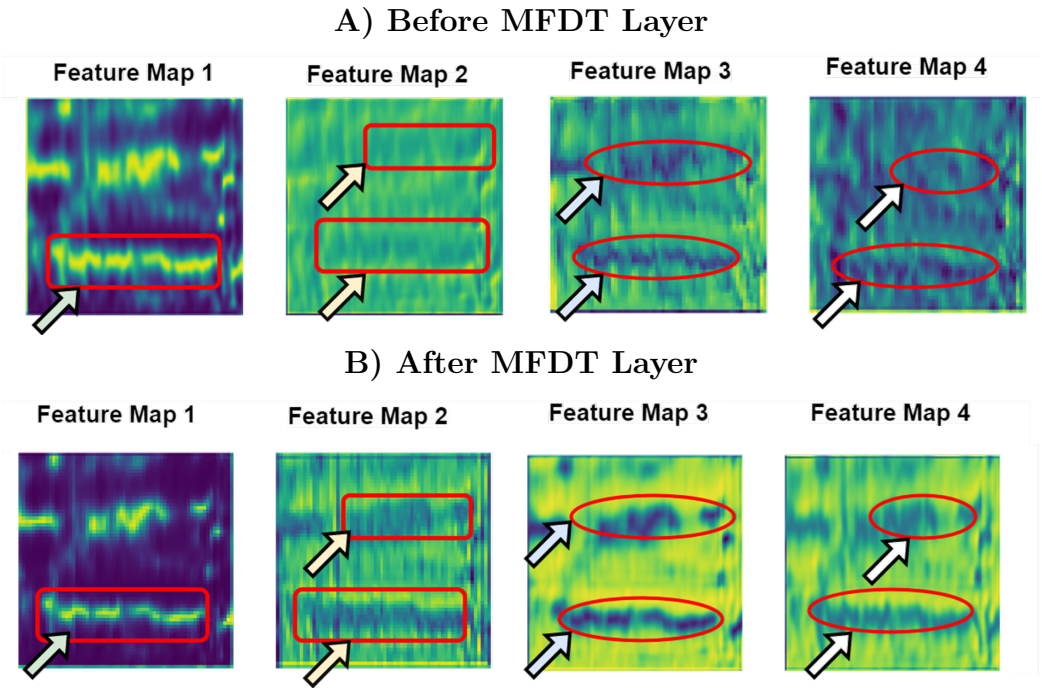


Figure 4.8: Comparative feature maps: (A) before and (B) after the MFDT Layer. After MFDT, feature channels exhibit enhanced reflector continuity, improved frequency coherence, and clearer thin-bed separation.

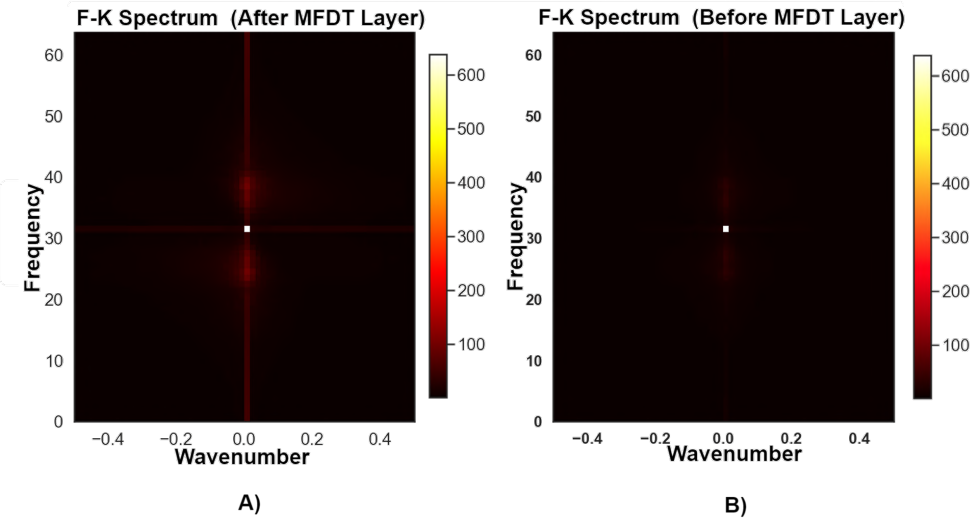


Figure 4.9: F-K spectra: (A) with MFDT layer; (B) without MFDT (U-Net only).

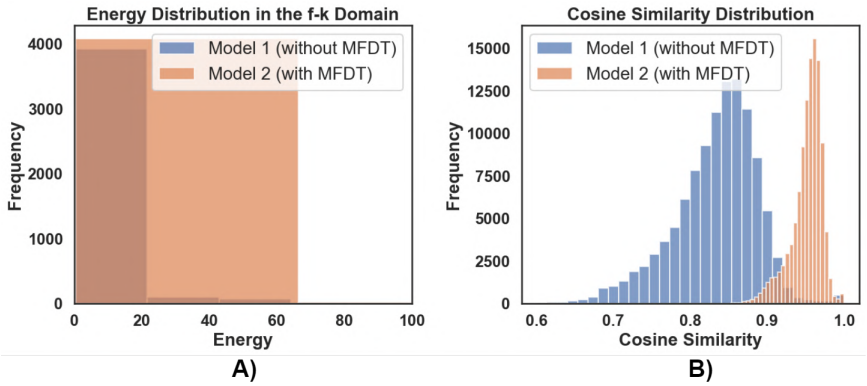


Figure 4.10: (A) Energy distribution in the F-K domain; (B) cosine-similarity distribution. Lower cosine similarity indicates more nearly orthogonal (less correlated) feature channels.

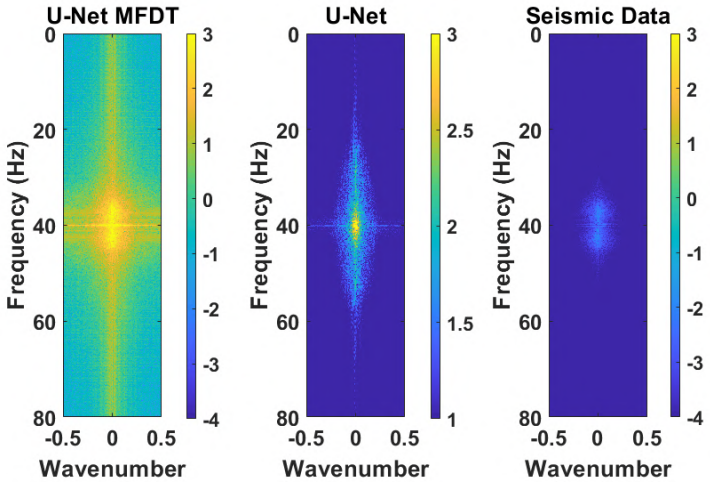


Figure 4.11: F-K domain comparison: raw seismic, U-Net output, and OrthoSeisNet (MFDT) output.

ure 4.12).

4.6 Discussion

The effectiveness of OrthoSeisNet on unsupervised domain adaptation was evaluated using synthetic and real seismic data.

1. Superior accuracy. OrthoSeisNet outperforms InversionNet, U-Net, and Re-

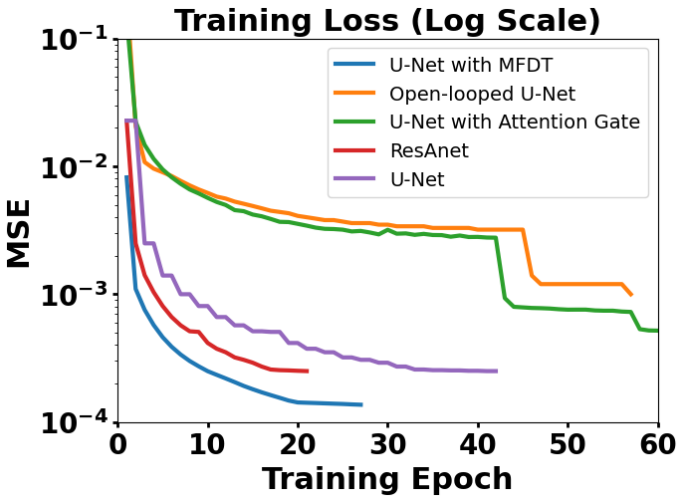


Figure 4.12: Epoch vs. Loss. U-Net with MFDT achieves the fastest convergence and lowest loss ($\approx 10^{-4}$).

sANet across all metrics at all noise levels on synthetic data and on the KG Basin real dataset, evidenced by consistently lower MAE and MSE and higher SSIM and R^2.

2. Near-orthogonal feature extraction. The MFDT layer produces features with lower inter-channel cosine similarity than the baseline U-Net, demonstrating near-orthogonal channel representations. This property emerges from the parameterized frequency scaling with learnable α parameters initialized within the Fourier framework. The F-K analysis confirms that MFDT captures a broader frequency range than the standard U-Net, contributing to improved thin-layer resolution.

3. Frequency domain enhancement. The adaptive regularisation weight $R(k, l)$ achieves a frequency-dependent trade-off: structural smoothness at low frequencies and detail preservation at high frequencies, consistent with sparsity-promoting regularisation approaches in classical geophysical inversion. **4. Domain adaptation.** The multi-stage UDA pipeline effectively bridges the synthetic-to-real domain gap, achieving 0.85 pseudo-label accuracy and an average Pearson correlation of 0.847 with well-derived reflectivity logs at three ONGC wells (Table 4.4), confirming stratigraphic consistency of the predicted reflectivity with borehole measurements.

5. Component-wise contribution analysis. Although a full ablation study is

planned for the revised manuscript, the contribution of each component can be inferred from the results:

- **U-Net baseline:** captures spatial structure but lacks high-frequency recovery, as evidenced by comparatively higher MSE and lower SSIM (Table 4.3).
- **U-Net + MFDT:** improves thin-bed resolution through enhanced spectral representation and near-orthogonal feature decorrelation.
- **U-Net + MFDT + UDA:** further improves performance on real data by reducing the synthetic-to-real domain shift, as confirmed by well-log correlation at three ONGC wells.

This progression indicates that MFDT contributes primarily to frequency-domain representation quality, while UDA improves generalisation to unlabelled field data.

4.6.0.1 Computational requirements and hardware.

Table 4.6 reports the hardware configuration. Initial pre-training used Google Colab TPU resources; following exhaustion of cloud credits, all subsequent training, UDA fine-tuning, and inference were performed on a local NVIDIA Quadro RTX 4000 workstation.

Stage	**Hardware**
Initial pre-training (Marmousi2, early epochs)	Google Colab TPU (cloud TensorFlow runtime; used until credits exhausted)
Full training pipeline, UDA fine-tuning	NVIDIA Quadro RTX 4000 (8 GB VRAM), Intel Xeon CPU, 192 GB DDR4 ECC RAM
Inference + validation	Same local workstation as above

Abbreviations: UDA = Unsupervised Domain Adaptation; ECC = Error-Correcting Code. Reported training time of 0.9 min/epoch was measured on the Colab TPU; inference timing (0.09 s/patch) was measured on the Quadro RTX 4000.

Table 4.6: Hardware configuration across training and inference stages. OrthoSeisNet (3.76 M params, 14.34 MB FP32) operates within 8 GB VRAM on mid-range research hardware.

- **Parameter count.** The original TensorFlow submission listed ≈ 1.94 M parameters (partial layer tally). The verified PyTorch implementation contains

3,758,145 parameters (14.34 MB at FP32), accounting for the MFDT recombination convolutions (≈435 K), the fully counted bottleneck block (1.18 M vs. the original single-layer estimate of 295 K), and the stem convolution (144). See Table 4.1 footnote for full details. The network architecture and training procedure are unchanged; only the accounting is corrected.

- **Model footprint.** At 14.34 MB (FP32), OrthoSeisNet occupies less than 0.2% of the 8 GB VRAM budget, leaving ample headroom for batched patch inference alongside intermediate activations.
- **Full 2D survey inference.** For the KG Basin dataset (800 inline sections, each 1,024×1,001 samples in the crossline–time plane), extraction with 128×128 patches and 50% overlap (stride = 64) yields approximately 168,000 patches total (15×14 patches/section × 800 sections). At 0.09 s/patch (serial), the full survey completes in ≈4.2 hours; batch inference (batch = 32) reduces this to ≈30–60 minutes on the same GPU. Processing scales linearly with survey size and is trivially parallelisable across inline sections.
- **3D extension: current limitations.** The present method intentionally operates on 2D inline–time sections, consistent with the majority of published DL seismic inversion literature [66, 69, 206]. 3D extension requires (i) 24–80 GB VRAM for 3D convolutional feature maps, (ii) large-scale 3D synthetic training data with ground-truth reflectivity labels (no 3D Marmousi equivalent at this scale was available), and (iii) a 3D WGAN-GP adversarial training pipeline. All three prerequisites exceeded the hardware available for this study. A fully 3D MFDT formulation is identified as future work contingent on appropriate hardware and dataset access (see Limitations, item 6).

Output activation limitation. The linear output activation does not enforce positivity of predicted reflectivity. For the KG Basin dataset, predicted reflectivity values remained physically plausible throughout the inversion. For datasets with extreme impedance contrasts or near-surface unconsolidated sediments, a softplus or ReLU activation would guarantee positivity in future impedance-domain extensions.

4.6.0.2 Limitations and future work.

1. **Lateral continuity.** The MFDT adaptive regularisation promotes reflectivity sharpening beneficial for high-contrast thin-bed delineation. The high-contrast appearance in the horizontal slice (Figure 4.7) is physically consistent with the sparse reflectivity output and should not be interpreted as geological facies segmentation. In low-contrast or gradational geological settings, the parameter λ should be calibrated against well-log data. Quantitative assessment of lateral continuity using semblance is planned for the revised submission.

2. **Ablation study.** Quantification of the individual contributions of MFDT, adaptive regularisation, and UDA is planned. The required experiments (configurations: baseline U-Net; +MFDT; +MFDT+adaptive regularisation; full OrthoSeisNet) will be run on the existing implementation and reported in the revised manuscript.

3. **Thin-bed resolution quantification.** A wedge-model experiment using the Widess tuning-thickness criterion will be included in the revision to quantify the minimum resolvable bed thickness relative to classical and baseline methods.

4. **Facies classification.** Lithofacies validation requires calibrated petrophysical labels, rock-physics modelling, and a complete wireline log suite. The current ONGC dataset does not include the full petrophysical suite required for quantitative facies classification; this is therefore deferred to future work. Integration of the inverted reflectivity with facies classification workflows represents an important research direction.

5. **3D extension.** The current 2D patch approach processes inline–TWT slices independently. A fully 3D MFDT formulation is proposed as future work; the hardware and data prerequisites that preclude this in the current study are discussed in Section 4.6, item 6.

6. **Production-scale deployment.** Multi-GPU scaling and distributed-inference throughput benchmarks for large 3D survey deployment are identified as future work.

4.7 Conclusion

OrthoSeisNet successfully addresses seismic inversion challenges with limited labeled data through: multi-scale frequency domain transformation with learnable near-orthogonal frequency weighting; adaptive frequency-band regularisation; Wasserstein GAN-based unsupervised domain adaptation; and linear output activation appropriate for sparse reflectivity regression. The approach enhances thin-layer detection and reflectivity prediction accuracy, demonstrating potential for improved subsurface imaging and hydrocarbon exploration.

4.8 Summary

Chapter 4 introduced OrthoSeisNet, incorporating multi-scale near-orthogonal frequency domain transformation into a U-Net backbone for seismic inversion. The method utilises near-orthogonal frequency domain features — verified through cosine-similarity analysis — to improve inversion accuracy. Quantitative well-log correlation at three ONGC wells demonstrates stratigraphic consistency with borehole measurements. Future revisions will add ablation experiments, a wedge-model thin-bed resolution test, and semblance-based lateral continuity analysis. Chapter 5 examines attention mechanisms integrated into the U-Net for further enhancement of inversion quality.

CHAPTER 5

Exploring Spectral Channel Attention Techniques to Enhance U-Net Seismic Inversion

5.1 Introduction

This chapter explores the integration of attention mechanisms within the U-Net backbone to enhance seismic inversion. The previous chapter exploited Fourier-transform orthogonality to recover sparse high-frequency components; this chapter builds on that work by examining how attention gates and networks can be used strategically. These mechanisms, operating at channel and spatial levels, are integrated into the U-Net architecture to address thin, sparse-layer challenges in seismic data, enabling the network to focus on relevant features and resolve intricate subsurface structures.

Attention gates with FFT layers, SENet, and CBAM are proposed for adaptive recalibration and spatial attention. DFT-Net modifies CBAM by incorporating DFT-based spectral attention within a U-Net backbone. DWTnet and DPSSnet further leverage DWT and DPSS for spectral attention. These innovations enhance U-Net's feature

extraction and adaptability to complex seismic data. The major contributions are:

1. **Attention Gate with Multi-scale Frequency Domain U-Net:** U-Net is enhanced with attention gates to explore multi-scale frequency domain features, improving seismic feature extraction and discrimination.

2. **Squeeze Excitation and CBAM:** SE and CBAM are integrated into the multi-scale U-Net for dynamic feature recalibration, focusing the model on key features and enhancing seismic data analysis.

3. **DFTnet-A DFT-Based Spectral Channel Attention Using CBAM Framework:** DFTnet uses DFT and CBAM for spectral channel attention, capturing key frequency components and enhancing seismic data analysis.

4. **DWTnet-A Wavelet-Based Spectral Attention Using CBAM Framework:** DWTnet uses the Discrete Wavelet Transform (DWT) with 'Haar' wavelets and CBAM for spectral channel attention, enhancing feature representation and analysis in seismic data.

5. **DPSSnet-A DPSS-Based Time-Frequency Concentration Attention Using CBAM Framework:** DPSSnet uses the Discrete Prolate Spheroidal Sequence (DPSS) and CBAM for spectral channel attention, enhancing feature representation and analysis in seismic data.

6. **Unsupervised Domain Adaption:** Using synthetic data we use geological

These contributions address key challenges in seismic inversion—feature extraction from complex data, accurate prediction, and physical constraint adherence—enhancing model interpretability and supporting more informed decision-making in geoscience applications.

5.2 Background

Attention mechanisms [117] have become fundamental in deep learning, significantly impacting natural language processing, computer vision, and seismic data analysis. These mechanisms enable networks to focus on relevant input features while suppress-

ing noise, enhancing interpretability, performance, and generalization.

Two prominent CNN attention mechanisms are SENet and CBAM. SENet recalibrates channel-wise feature responses by modelling inter-channel dependencies, dynamically adjusting importance per task relevance [211]. CBAM integrates spatial and channel attention into a unified module, using max- and average-pooling to capture global and local context, enabling selective focus on informative regions [212].

Deep learning inversion methods have gained traction for post-stack tasks but struggle to produce high-resolution results with good lateral continuity when well-log data are scarce [213]. To address this, researchers have explored data augmentation [115], multidimensional networks [214], geological prior-based methods, and semi-supervised or unsupervised learning [204].

Integrating spectral attention into channel attention mechanisms has emerged as a promising approach. Spectral attention enhances discriminative power by capturing complex frequency-domain dependencies within feature channels. Incorporating it into CBAM or SENet improves seismic inversion performance, particularly with sparse well-log data [45].

Deep learning inversion methods commonly use 1-D convolution to exploit trace-to-trace correspondence between seismic data and elastic parameters [45, 64, 90, 213]. However, acquiring sufficient well-log data is often impractical or prohibitively costly, limiting 1-D algorithms that struggle to predict high-resolution results with desirable lateral continuity from sparse training data [69]. Bürkle et al. [115] introduced a physics-aware semi-supervised method using geostatistical simulation for acoustic impedance inversion. Wu et al. [214] demonstrated that 2-D networks yield superior lateral continuity compared to 1-D networks on field data from 27 wells. However, these data augmentation strategies are complex and do not guarantee spatial distributions consistent with real data; semi-supervised training is also time-consuming. Generating sufficient 2-D label data under sparse well-log conditions remains an open challenge [69].

5.3 Proposed Methodology

This section introduces the theory behind seismic inversion and describes novel attention mechanisms integrated with the U-Net architecture. These include Unet-AG (AGO), Unet-SENet, Unet-CBAM, Unet-DFT(DFTNet), Unet-DWT(DWT-Net), and Unet-DPSS(DPSS-Net).

5.3.1 Theory

Seismic inversion is refined through DL models that transform seismic trace data into acoustic impedance profiles. This process rests on the Zoeppritz equation [215], which describes the interaction of planar waves with horizontal layer interfaces. It relates the reflection coefficient to acoustic impedance contrast across interfaces (Equation 5.1):

$$R_i = \frac{\rho_i \nu_i - \rho_{i-1}\nu_{i-1}}{\rho_i \nu_i + \rho_{i-1}\nu_{i-1}} = \frac{Z_i - Z_{i-1}}{Z_i + Z_{i-1}}, \tag{5.1}$$

where R signifies the reflection coefficient, Z represents acoustic impedance, and i indexes the interface layers.

Forward modelling conceptualises the seismic profile as a convolution of reflection coefficients with a Hessian tensor $\mathcal{F}$ encoding subsurface geometry (Equation 5.2):

$$\mathcal{S} = \mathcal{F}R, \quad \mathcal{F} = \frac{\partial^2 J(\mathbf{R})}{\partial^2 \mathbf{R}}, \tag{5.2}$$

where $\mathcal{S}$ denotes the observed seismic profile and J is the objective function of the inversion.

Reconfiguring this gives the link between impedance Z and observed seismic traces $\mathcal{S}$ (Equation 5.3):

$$\frac{\partial Z}{\partial n} = \mathcal{F}^{-1}\mathcal{S}. \tag{5.3}$$

Equation 5.3 suggests a non-linear mapping f central to the proposed DL-based inversion, mapping the seismic trace input to acoustic impedance output (Equation 5.4):

$$Z = f(\mathcal{S}). \tag{5.4}$$

Figure 5.1 presents the DL workflow for seismic inversion, which begins with generating synthetic training data. Given scarce well-log data, the methodology uses real seismic data directly, identifying sparse patterns to facilitate blind seismic inversion.

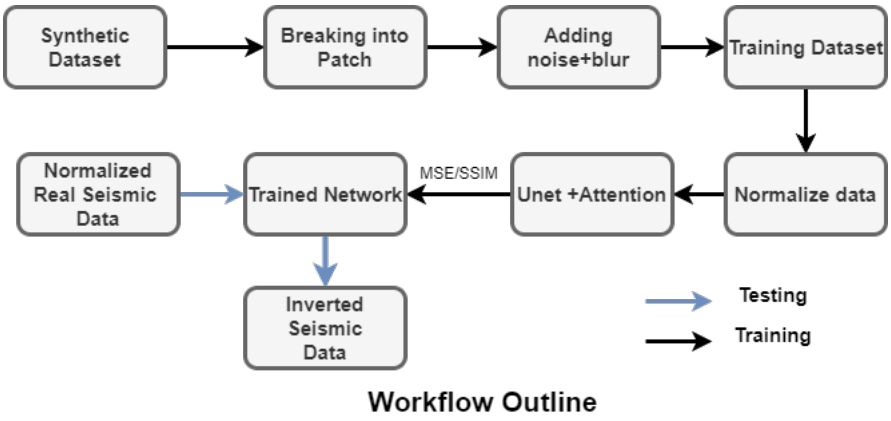


Figure 5.1: A seismic inversion deep learning workflow.

5.3.1.1 Why use Attention for thin-layer identification?

Attention mechanisms enhance neural network architectures by focusing analysis on specific input regions. In U-Net, they improve performance by dynamically highlighting relevant features while suppressing less important ones, assigning varying importance to prioritize crucial information.

Attention can be integrated into skip connections or convolutional pathways, refining features and enhancing reconstruction of important details. Key types include Attention Gate (AG), Squeeze-and-Excitation Networks (SE-Net), Convolutional Block Attention Module (CBAM), and spectral channel attention methods such as DFT, DWT, and DPSS. AGs filter skip-connection feature maps to emphasize critical features. SE-Net recalibrates channel-wise responses, highlighting the most informative channels. CBAM combines channel and spatial attention to capture both global and local features. Spectral attention enhances frequency-domain features crucial for identifying periodic patterns in seismic data.

Thin layers in seismic data produce weak signals easily overshadowed by noise; attention mechanisms enhance these subtle features, improving precise localization of thin layers. Channel-wise attention such as SE-Net focuses on the most informative

channels. Spectral attention via DFT or DPSS enhances frequency components associated with thin layers, aiding detection and delineation. Attention also separates signal from noise, making thin-layer detection more reliable. Together, these mechanisms significantly improve U-Net's ability to detect and characterize thin layers in seismic data.

5.3.2 Method-I: Attention Gate on Orthoseisnet in Seismic Inversion (AGO)

Orthoseisnet, as introduced in Chapter 4, incorporates multi-scale frequency domain layers into the U-Net architecture, effectively orthogonalizing frequency components to identify sparse features with high precision. Building on this, we integrate attention gates (AGs) to refine feature discernment. Each encoder layer uses dual 2D convolution operations regularized by the l_1 norm, with dropout and max-pooling, culminating in a multi-scale frequency domain layer as described in Chapter 4.

The model leverages coarse feature map context to improve categorization and localization of subsurface anomalies. Skip connections integrate multi-scale feature maps, balancing coarse and fine detail for accurate seismic reconstruction.

Traditionally, CNNs downsample feature maps to capture broad semantic context, enabling global modelling of subsurface structures. However, this approach struggles to suppress false positives for small, shape-variable features. Unlike conventional methods that treat localization and segmentation separately, AGs within U-Net suppress non-target feature responses directly, eliminating intermediate ROI cropping. Attention coefficients, $\alpha_i \in [0, 1]$, highlight relevant seismic regions, focusing activations solely on pertinent features.

AGs multiply input feature maps element-wise by attention coefficients, concentrating the model on key seismic attributes. Multi-dimensional attention coefficients enable focused analysis of specific subsurface structures [216].

Additive attention governs this process, trading computational cost for superior accuracy (Equation 5.5):

$$Q_{AG}^{l} = \psi^{T}(\sigma_1(W_x^{T} x_i^{l} + W_g^{T} g_i + b_g)) + b_\psi, \tag{5.5}$$

$$\alpha_i^{l} = \sigma_2(Q_{AG}^{l}(x_i^{l}, g_i; \theta_{AG})), \tag{5.6}$$

σ_2 denotes the sigmoid activation function,θ_{AG} is AG parameter with linear transformations

$$W_x \in \mathcal{R}^{F_l \times F_{int}}, W_g \in \mathcal{R}^{F_g \times F_{int}}, \psi \in \mathcal{R}^{F_i nt \times 1}$$

and bias term $b_g \in \mathcal{R}, b_g \in \mathcal{R}_{int}^{F}$, where x^l and g are mapped to inter-dimesional space $\mathcal{R}_{int}^{F}$ facilitating focused analysis of relevant seismic data features [217]. Figure 5.2 illustrates the detailed architecture, embedding AGs within the Orthoseisnet framework for seismic inversion. This enhances feature selection and segmentation, improving inversion outcomes through nuanced subsurface representation [218]. Table 5.1 details the model parameters.

5.3.2.1 Network Architecture Method-I: Attention Gate on Orthoseisnet in Seismic Inversion

AGs are integrated into Orthoseisnet, using coarser-scale information to suppress irrelevant or noisy responses before concatenation, ensuring only pertinent activations are merged (Figure 5.2) [219]. AGs modulate activations in both forward and backward passes, deemphasizing background gradients during backpropagation. This selective updating primarily affects early-layer parameters [220] in regions critical for seismic-to-impedance conversion. Each AG extracts and combines complementary information to optimize skip connections. To reduce computational load, AGs use spatially unsupported linear transformations, adjusting input feature maps to the gating signal's resolution—analogous to non-local blocks [221]. Deep supervision keeps intermediate feature maps semantically distinct across scales, promoting an effective seismic inversion model.

Figure 5.2 shows how the AG is implemented in Orthoseisnet. The AG block is placed in the decoder (Equation 5.7).

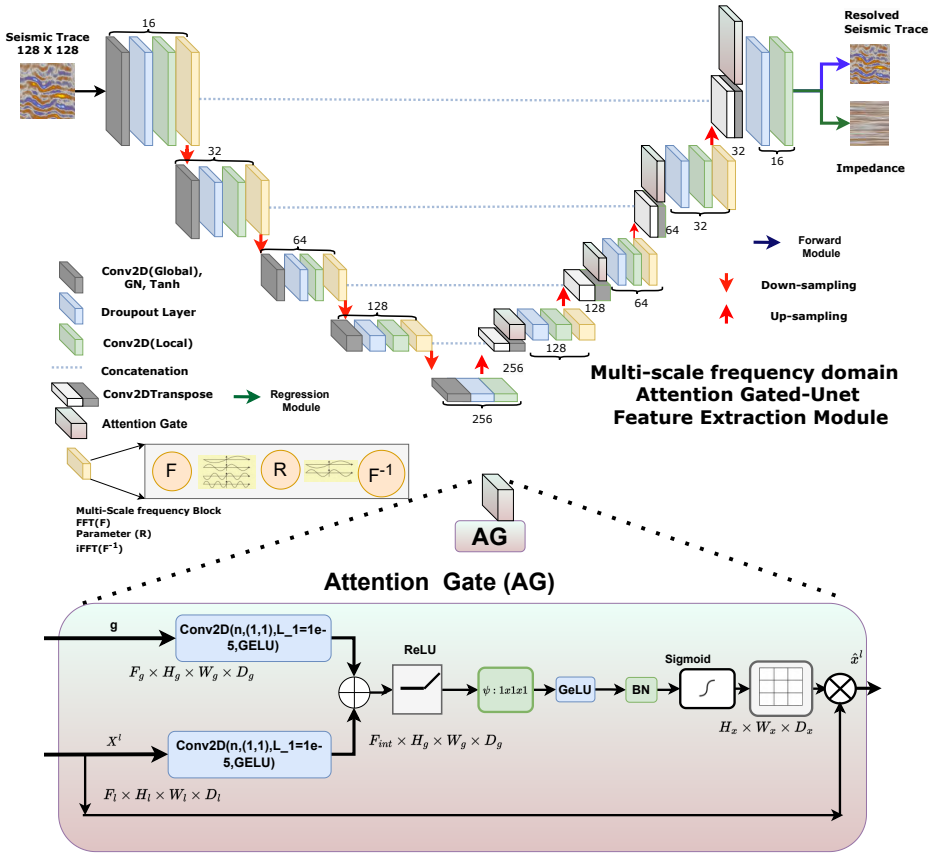


Figure 5.2: Attention-Gate on U-Net for seismic Inversion

$$\text{Dec\_Layer} = \text{FFT-iFFT2D}\Big(\text{Conv2D}\big(\text{Dropout}\big(\text{C2DTrp}\left(\text{AG\_concat}(X, y)\right)\big)\big)\Big) \tag{5.7}$$

Where AG_concat is AG and concatenation,Dec_Layer is decoding layer, and C2DTrp is Conv2DTranspose FFT-iFFT2D (FFT layer), a multi-scale frequency domain in 2D.

Table 5.1 details the Attention-Gated multi-scale U-Net architecture. Scaled filters are used for depth resolution enhancement. The *FFTLayer* is a custom TensorFlow layer that applies a 2D FFT to the input tensor, with Glorot-initialized complex weights acting as frequency filters. During the forward pass, the input

is FFT-transformed, element-wise multiplied by these weights, then inverse-FFT-transformed, allowing the network to selectively capture frequency information. AG is applied in the decoder. The network has 1,985,069 trainable and 8 non-trainable parameters.

Block	**Layer**	**Unit**	**Parameters**
Input	0	Seismic 2D ($128 \times 128 \times 1$)	0
Enc1	1–4	Enc(16), FFT, iFFT	2,544
Enc2	5–8	Enc(32) abs(iFFT)	14,016
Enc3	9–12	Enc(64) abs(iFFT)	55,680
Enc4	13–16	Enc(128) abs(iFFT)	221,952
Enc5	17	Conv2D(256) abs(iFFT)	295,168
Dec1	18	Dec(128), FFT-iFFT	1.53M+
Dec2	22–26	Dec(64), FFT layer	144,968
Dec3	26–30	Dec(32), FFT layer	36,388
Dec4	31–34	Dec(16), FFT layer	9,074
Total Parameters			1,985,077
Trainable Parameters			1,985,069
Non-trainable Parameters			8

Table 5.1: Attention-Gated Multi-scale U-Net Architecture

The Enc and Dec blocks are defined in Equation 5.8.

$$\begin{aligned} \text{Enc(i)} = &\textbf{Maxpooling}\Big(\textbf{Conv2D}\Big(\textbf{Dropout}\Big(\\ &\quad \textbf{Conv2D}(i,(3,3),\text{GeLU},\text{filter}) + \text{GN} + \text{L1}\Big),\text{GN},\text{L1}\Big)\Big), \\ \text{Dec(i)} = &\textbf{Conv2D}\Big(\textbf{Dropout}\Big(\textbf{Conv2DTrans}\Big(\textbf{AG+Concat}(i),\text{GN}\Big)\Big)\Big) \end{aligned} \tag{5.8}$$

Results for synthetic and real seismic data are provided in Section 5.5.

Integrating attention mechanisms into seismic inversion frameworks has significantly enhanced subsurface imaging interpretability and precision. Building on successes in image classification and segmentation, SENet and CBAM have been adapted for seismic data interpretation [222–225].

5.3.3 Results

Results for AG-Orthoseisnet are shown in the figures below.

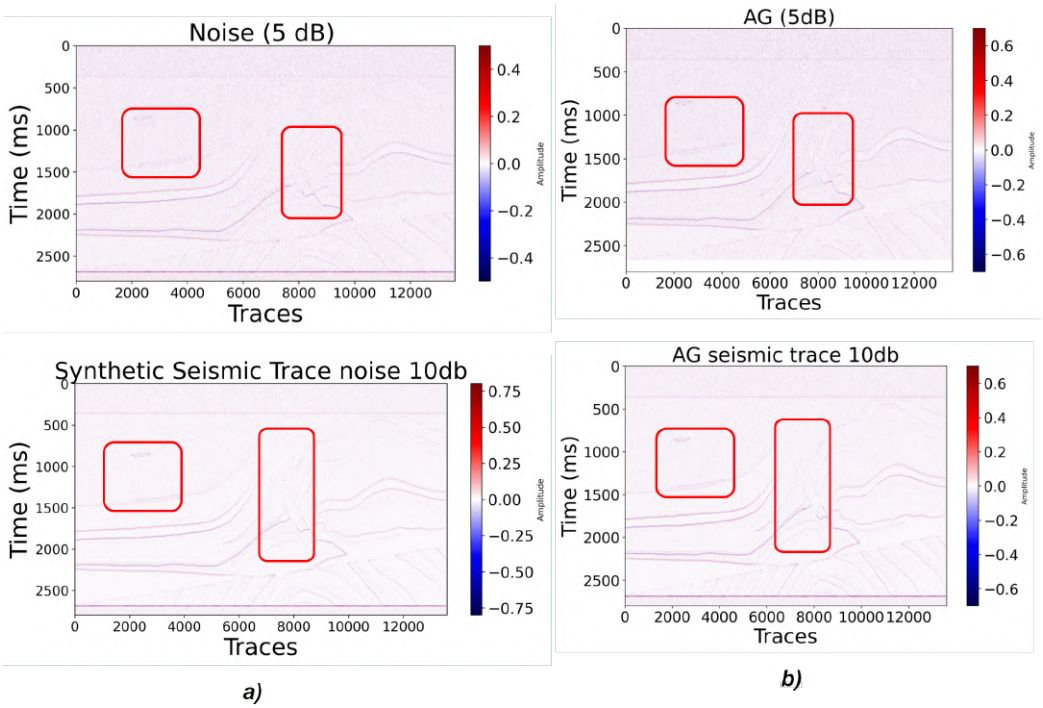


Figure 5.3: Noisy Seismic trace Comparison a) synthetic Marmousi trace data b) AG-Net-Seismic

5.3.4 Method-II: Squeeze Excitation (SE-Net) and Convolutional Block Attention Module(CBAM) on Orthoseisnet (SE-Onet and CBAM-Onet)

SENet and CBAM refine feature representations by emphasizing salient features and suppressing less relevant ones. This is crucial in seismic inversion, where delineating subsurface structures demands nuanced understanding of data complexity.

5.3.4.1 Squeeze-and-Excitation Networks (SENet) in Seismic Inversion:

Incorporating SENet into U-Net enhances the capture of intricate subsurface geological details. SE blocks adaptively recalibrate channel-wise responses, enabling more

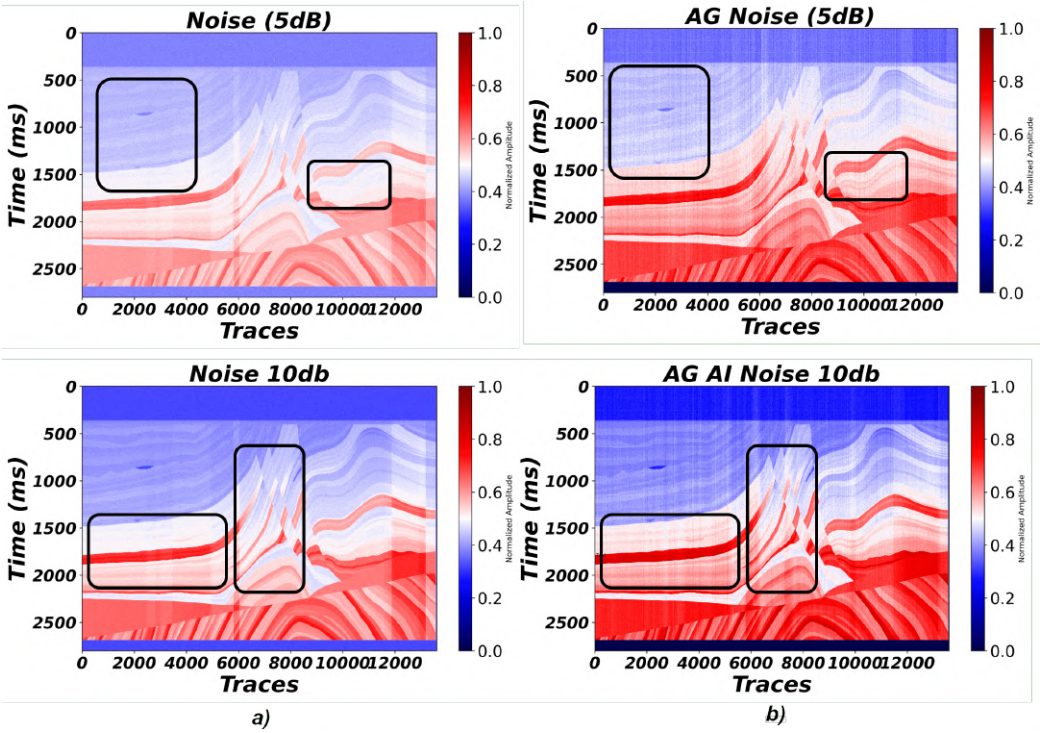


Figure 5.4: Noisy Acoustic Impedance Comparison a) Marmousi data b) AG-Net-AI

precise acoustic impedance extraction from seismic traces.

Squeeze Operation: The squeeze operation compresses each channel's global spatial information into a single value. This aggregated descriptor captures pivotal channel attributes (Equation 5.9):

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j), \tag{5.9}$$

Where z_c signifies the aggregated feature for channel c, and $u_c(i, j)$ denotes the value at the i^{th} row and j^{th} column of channel c, with H and W representing the feature map's height and width, respectively.

Excitation Operation: The excitation step applies a fully connected layer with sigmoid activation to derive channel-specific modulation weights. These weights mod-

ulate each channel's emphasis based on its relevance to seismic inversion (Equation 5.10):

$$s_c = \sigma(g(z, W)) = \sigma(W_2\delta(W_1 z)), \tag{5.10}$$

where s_c is the modulated weight for channel c, W_1 and W_2 correspond to the weights of the fully connected layers, δ indicates the ReLU activation, σ is the sigmoid activation function, and $g(z, W)$ represents the transformation of the aggregated feature z through the network.

Embedding SE blocks after each convolutional layer enables dynamic recalibration of channel importance from global and inter-channel context, making U-Net more adept at highlighting significant features and improving subsurface property prediction accuracy.

SE attention exemplifies how advanced DL can improve seismic inversion quality, providing geophysicists with refined tools for subsurface characterization.

5.3.5 Convolutional Block Attention Module(CBAM)

CBAM sequentially infers attention maps along both channel and spatial dimensions, refining feature representations for more accurate subsurface imaging [212]. Integrating CBAM into U-Net advances the capture of complex subsurface characteristics. By focusing systematically on salient features, CBAM facilitates nuanced extraction of geological properties from seismic traces [212, 226].

CBAM operates through two sequential components: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). CAM identifies informative features across channels, while SAM emphasizes important spatial locations within each feature map. Together, they ensure that the U-Net model prioritizes the most relevant features for seismic inversion.

Channel Attention Module (CAM): CAM employs global average pooling (GAP) and max pooling operations to generate two distinct context descriptors. These descriptors capture global spatial information for each channel, highlighting the channel-

wise dependencies. Combining these features through a shared network and applying a sigmoid activation function yields the channel attention map. CAM is expressed in Equation 5.11:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))), \tag{5.11}$$

where F denotes the input feature map, M_c is the channel attention map, σ is the sigmoid function, MLP represents a multi-layer perceptron, and $AvgPool$ and $MaxPool$ are the average and max pooling operations, respectively.

Spatial Attention Module (SAM): Following CAM, SAM applies average pooling and max pooling operations across the channel dimension to highlight spatial features crucial for seismic inversion. The concatenation of the resulting feature maps, followed by a convolution operation and sigmoid activation, produces the spatial attention map:

$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])), \tag{5.12}$$

where M_s is the spatial attention map, $f^{7\times7}$ denotes a convolution operation with a 7×7 kernel, and σ is the sigmoid function.

Applying CBAM sequentially after each convolutional block within U-Net allows the network to focus dynamically on the most informative features from both channel and spatial perspectives. The enhanced model more efficiently delineates seismic data to predict subsurface properties.

5.3.5.1 Network Architecture Method-II (a): Squeeze excitation(SE-Net) on Orthoseisnet

Incorporating SE-Net into the U-Net architecture enhances seismic inversion by recalibrating channel-wise features from Conv2D layers. Through squeeze-and-excitation operations, SE-Net prioritizes features critical for identifying and characterizing subsurface geological formations, as detailed in Table 5.2.

Integrating SE blocks immediately after Conv2D layers ensures the network focuses

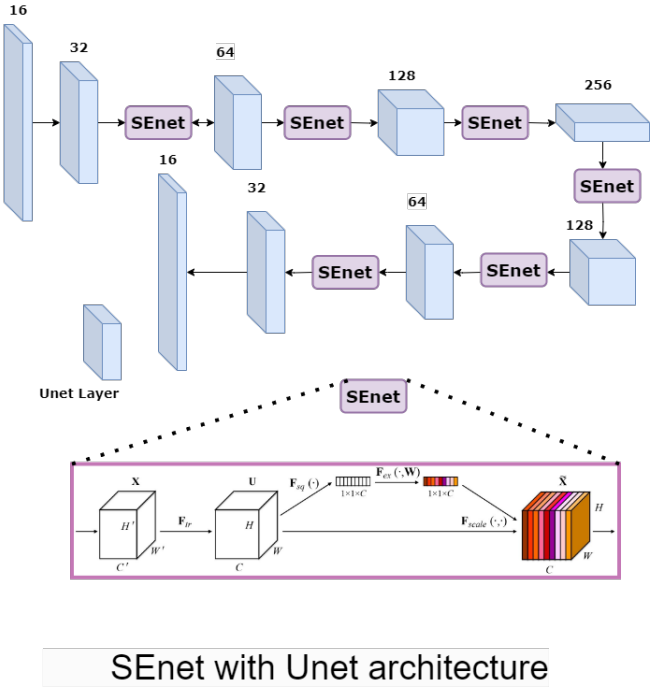


Figure 5.5: Integration of SE-Net within U-Net for Enhanced Seismic Inversion

on features critical for resolving complex geological structures. By adaptively emphasising or suppressing channel-wise features based on global importance, SE-Net captures subtle seismic nuances, enhancing inversion precision. Figure 5.5 illustrates the strategic placement of SE blocks within the U-Net architecture. Table 5.2 details the SE-Net–Orthoseisnet configuration. In the table, SE denotes the SE-Net block.

The Enc and Dec blocks are defined in Equation 5.13.

$$\begin{aligned}\text{Enc(i)} = &\textbf{Maxpooling}\Big(\textbf{Conv2D}\Big(\textbf{Dropout}\\ &\Big(\textbf{Conv2D(i, (3,3), GeLU, filter,l1) + GN+SE), GN+SE, l1}\Big)\Big),\\ \text{Dec(i)} = &\textbf{Conv2D}\Big(\textbf{Dropout}\Big(\textbf{Conv2DTrans}\Big(\textbf{Concat}(i), GN, SE\Big)\Big)\Big)\end{aligned} \tag{5.13}$$

Results for synthetic and real seismic data are shown in Figure 5.6.

Block	Layer	Unit	Parameters
Input	0	Seismic 2D ($128 \times 128 \times 1$)	0
Enc1	1–4	Enc(16)+FFT layer	2,642
Enc2	5–8	Enc(32) abs(iFFT)	14,340
Enc3	9–12	Enc(64) abs(iFFT)	56,840
Enc4	13–16	Enc(128) abs(iFFT)	226,320
Enc5	17	Conv2D(256) abs(iFFT)	303,632
Dec1	18–21	Dec(128), FFT layer	1,469,720
Dec2	22–26	Dec(64), FFT layer	144,968
Dec3	26–30	Dec(32), FFT layer	36,388
Dec4	31–34	Dec(16), FFT layer	9,074
Total Parameters			1,960,292
Trainable Parameters			1,960,292
Non-trainable Parameters			0

Table 5.2: SE-Net Attention on Multi-scale U-Net Architecture

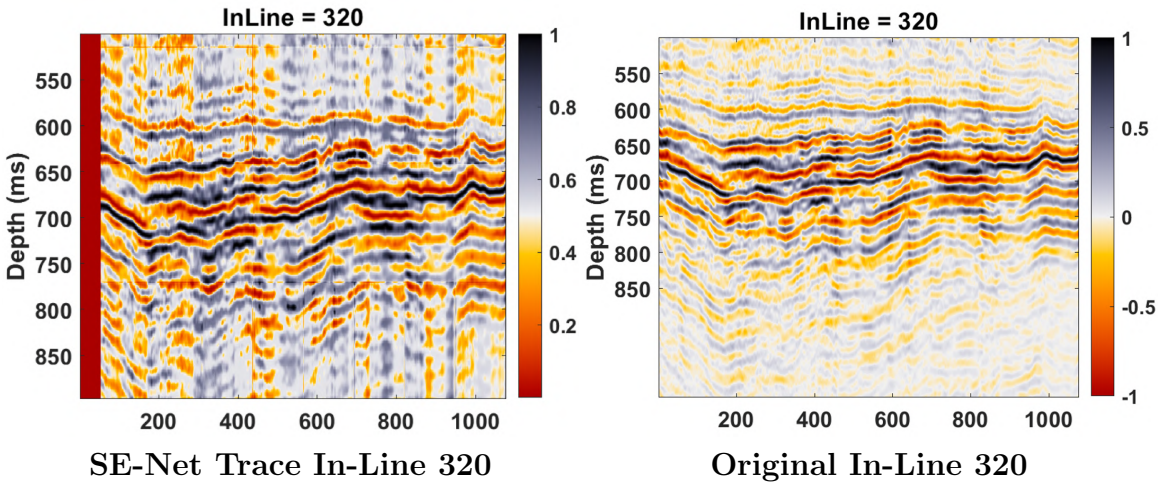


Figure 5.6: SE-Net comparison with real data

5.3.5.2 Network Architecture Method-II (b): Convolutional Block Attention Module(CBAM) on Orthoseisnet

Integrating CBAM into Orthoseisnet advances seismic inversion methodology. CBAM dynamically adjusts feature maps across both spatial and channel dimensions, enhanc-

ing sensitivity to relevant seismic attributes. This recalibration significantly improves detection and characterization of subsurface geological features [212].

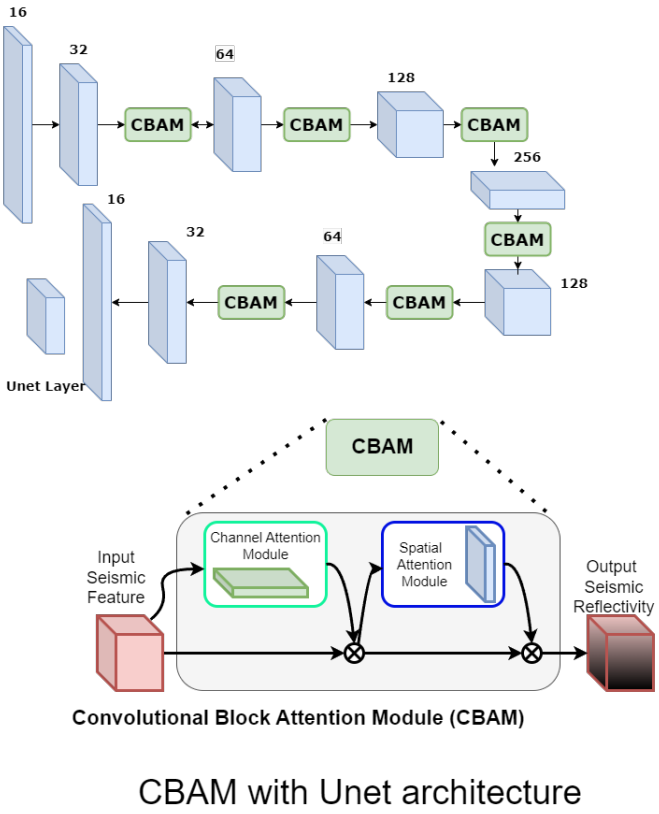


Figure 5.7: Enhancing Seismic Inversion with CBAM-Integrated U-Net

Embedding CBAM into Orthoseisnet applies attention sequentially: channel attention first selects the most informative channels; spatial attention then emphasizes salient locations within them.

This hierarchical process enables more nuanced interpretation of seismic data, particularly for complex geological formations and anomalies. Figure 5.7 illustrates the CBAM integration, with blocks placed to dynamically refine focus on inversion-relevant features.

Table 5.3 details the CBAM–U-Net architecture and training parameters.

Block	Layer	Unit	Parameters
Input	0	Seismic 2D ($128 \times 128 \times 1$)	0
Enc1	1–4	Enc(16)+FFT layer	2,644
Enc2	5–8	Enc(32) abs(iFFT)	14,488
Enc3	9–12	Enc(64) abs(iFFT)	72,688
Enc4	13–16	Enc(128) abs(iFFT)	230,176
Enc5	17	Conv2D(256) abs(iFFT)	295,168
Dec1	18–21	Dec(128), FFT layer	1,469,720
Dec2	22–26	Dec(64), FFT layer	146,328
Dec3	26–30	Dec(32), FFT layer	35,968
Dec4	31–34	Dec(16), FFT layer	9,227
Total Parameters			2,012,429
Trainable Parameters			2,012,429
Non-trainable Parameters			0

Table 5.3: CBAM Attention on Multi-scale U-Net Architecture

The CBAM Enc and Dec blocks are defined in Equation 5.14.

$$\begin{aligned}\text{Enc(i)} = &\textbf{Maxpooling}\Big(\textbf{Conv2D}\Big(\textbf{Dropout}\\ &\Big(\textbf{Conv2D(i, (3,3), ReLU,l1) + GN), GN+CBAM, l1}\Big)\Big),\\ \text{Dec(i)} = &\textbf{Conv2D}\Big(\textbf{Dropout}\Big(\textbf{Conv2DTrans}\Big(\textbf{Concat}(i), GN, CBAM\Big)\Big)\Big)\end{aligned} \tag{5.14}$$

5.3.6 Method-III: DFTnet — A Discrete Fourier Transform (DFT)-based spectral channel attention using CBAM attention framework with U-Net backbone in seismic inversion

Traditional machine learning methods have extensively used handcrafted features—colour histograms [227], texture descriptors [228], SIFT [229], and HOG [230]—to interpret subsurface images. These features, while useful, often fail to capture the

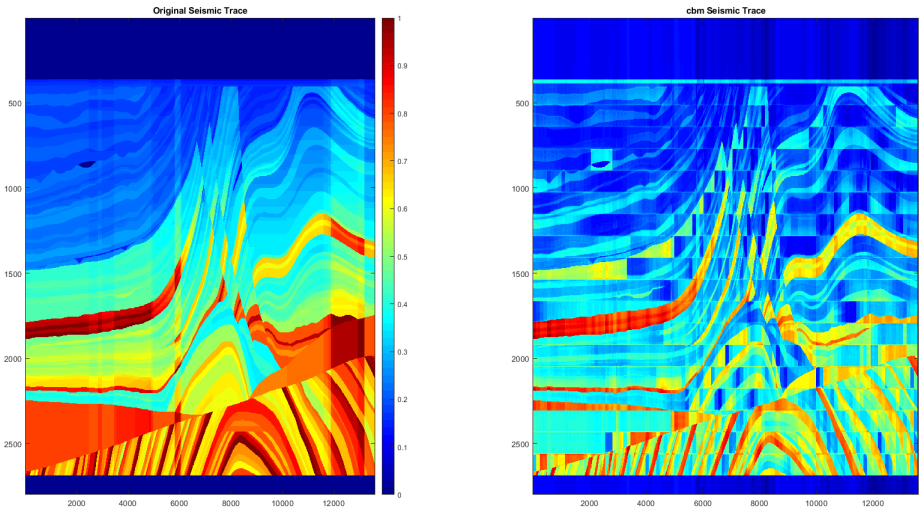


Figure 5.8: Seismic Impedance Inversion with U-Net-CBAM

intricate details in seismic traces essential for accurate subsurface exploration.

Deep learning in the frequency domain has recently surged to address these challenges. Though efficient for data compression, the Discrete Cosine Transform (DCT) focuses predominantly on low-frequency components, neglecting crucial high-frequency information [231, 232]. This is significant in seismic analysis, where high-frequency components reveal vital subsurface layer details. Unlike DCT, DFT with a Hanning window and zero padding—as used in DFNet—preserves these high-frequency components, alleviating the Gibbs phenomenon and improving the signal-to-noise ratio [233]. The equations below describe DFTnet's implementation within CBAM's channel attention in U-Net.

I. DFT Layer Methodology

The `DFTLayer` proceeds through three core operations: DFT application, windowing, and separation of the signal into real and imaginary parts. It applies a windowed DFT to each input channel and concatenates the real and imaginary outputs.

i. Window Function Application

The Hanning window reduces spectral leakage by softening boundary discontinuities.

For a signal of length N it is (Equation 5.15):

$$w(n) = 0.5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right), \quad 0 \leq n < N \tag{5.15}$$

ii. Discrete Fourier Transform

The DFT of a signal $x(n)$, windowed by $w(n)$, for length N is:

$$X(k) = \sum_{n=0}^{N-1} x(n)w(n)e^{-j\frac{2\pi}{N}kn}, \quad 0 \leq k < N \tag{5.16}$$

where $X(k)$ denotes the DFT at frequency bin k, j is the imaginary unit, and e represents Euler's number.

iii. Operational Steps in DFTLayer

1. **Window Application:** Each input channel is element-wise multiplied by the Hann window reshaped appropriately for broadcasting compatibility.
2. **Fourier Transform:** The windowed signal undergoes a 2D Fast Fourier Transform (FFT), transitioning into the frequency domain.
3. **Component Separation:** The real ($\Re$) and imaginary ($\Im$) parts of the frequency domain representation are extracted and stored.

Given an input batch element x with channel index i, the transformation is represented as:

$$X_i(k,l) = \mathcal{F}\{x_i(n,m)w(n,m)\} \tag{5.17}$$

$$\text{where} \quad X_i(k,l) = \Re\{X_i(k,l)\} + j\Im\{X_i(k,l)\} \tag{5.18}$$

Here, $\mathcal{F}$ symbolizes the 2D FFT operation.

iv. Concatenation of Outputs The real and imaginary components are concatenated across channels into two distinct tensors:

$$\text{Real Output} = \bigoplus_{i=1}^{C} \Re\{X_i\} \tag{5.19}$$

$$\text{Imaginary Output} = \bigoplus_{i=1}^{C} \Im\{X_i\} \tag{5.20}$$

with $\bigoplus$ denoting the concatenation operation over C channels.

v. Real and Imaginary Component Formulations

The formulations are:

$$\text{Real Part: } \Re\{X_i(k,l)\} = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} x_i(n,m)w(n,m)\cos\left(\frac{2\pi}{N}kn + \frac{2\pi}{M}lm\right) \tag{5.21}$$

$$\text{Imaginary Part: } \Im\{X_i(k,l)\} = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} x_i(n,m)w(n,m)\sin\left(\frac{2\pi}{N}kn + \frac{2\pi}{M}lm\right) \tag{5.22}$$

Equations (5.21) and (5.22) complete the `DFTLayer` methodology: windowing, per-channel DFT, and real/imaginary segregation.

II. High-Frequency Channel Attention Function

The High-Frequency Channel Attention Function comprises normalization, pooling, and combination of DFT components. It targets DFT-derived high-frequency components to highlight critical spectral features.

i. Normalization of DFT Components

Normalization standardises the real and imaginary DFT outputs, improving learning efficiency. For any component X (real or imaginary) the normalisation is (Equation (5.23)):

$$X_{\text{norm}} = \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{5.23}$$

where μ and σ^2 represent the mean and variance of X across its spatial dimensions, respectively, and ϵ is a small constant to prevent division by zero.

ii. Pooling of Normalized Components

GAP and GMP encapsulate each component's features, capturing the average and

most pronounced high-frequency characteristics:

$$\text{GAP}(X_{\text{norm}}) = \frac{1}{N}\sum_{i=1}^{N} X_{\text{norm},i} \tag{5.24}$$

$$\text{GMP}(X_{\text{norm}}) = \max_{i=1}^{N} X_{\text{norm},i} \tag{5.25}$$

N is the total count of elements within X_{norm}.

iii. Combination of GAP and GMP Features

The GAP and GMP features for both components are combined into a summary of spectral attributes:

$$\text{Combined}_{\text{real}} = \text{GAP}_{\text{real}} + \text{GMP}_{\text{real}} \tag{5.26}$$

$$\text{Combined}_{\text{imag}} = \text{GAP}_{\text{imag}} + \text{GMP}_{\text{imag}} \tag{5.27}$$

iv. Application of Dense Layers

Dense layers then apply non-linear transformations to the flattened, combined features:

$$\text{FC}_{\text{real}} = \text{Dense}(\text{Combined}_{\text{real}}) \tag{5.28}$$

$$\text{FC}_{\text{imag}} = \text{Dense}(\text{Combined}_{\text{imag}}) \tag{5.29}$$

V. Final Combination

The final step blends the adjusted components with the original input, modulated by a sparsity coefficient to accentuate high-frequency details:

$$X_{\text{combined}} = (1 - \text{sparsity}) \times \text{Input} + \text{sparsity} \times (\text{FC}_{\text{real}} + \text{FC}_{\text{imag}}) \tag{5.30}$$

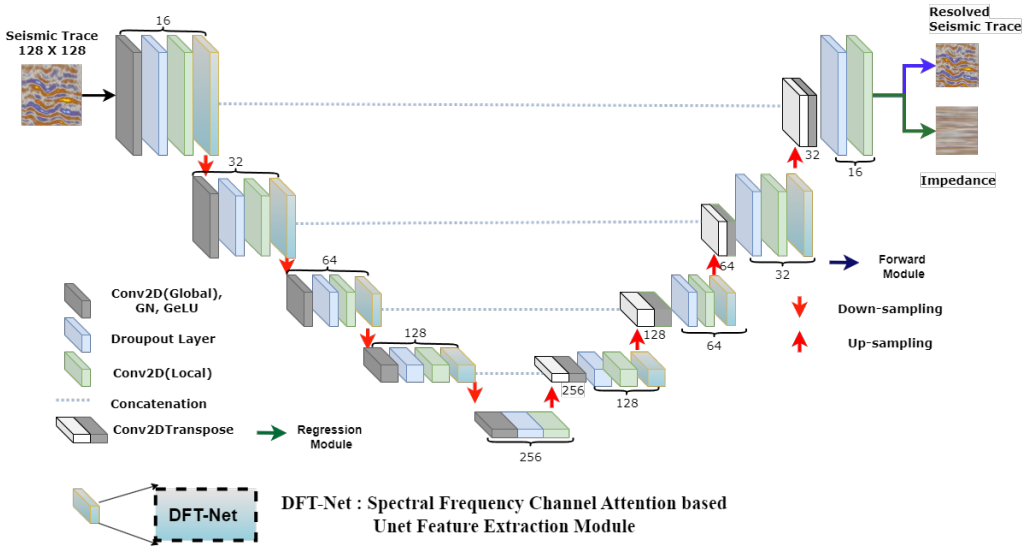


Figure 5.9: DFTnet attention on U-Net for seismic inversion

5.3.6.1 DFTnet-CBAM

Seismic data resolution directly impacts the identification of geological features such as thin reservoirs, fractures, and faults. DFT-modified CBAM enhances seismic data analysis by emphasising high-frequency channel and spatial attention. This section elaborates on DFTnet-CBAM's mathematical foundation and its seismic inversion application. The CBAM block enhances feature representation by integrating high-frequency channel attention with spatial attention, divided into two sequential components described below.

i. Channel Attention

The High-Frequency Channel Attention Function selects the most important frequency components from the seismic data, enabling the model to resolve small geological features:

$$X_{\text{att}} = \text{Attention}(X) \tag{5.31}$$

Here, X denotes the seismic data input and X_{att} is the enhanced output focused on high-frequency channels. This amplifies signals indicating thin beds, fractures, and other critical reservoir attributes.

ii. Spatial Attention

After channel attention, the spatial attention map is created through sequential convolutions and sigmoid activation:

$$Y = \sigma(\text{Conv}(\text{ReLU}(\text{Conv}(\text{ReLU}(\text{Conv}(X_{\text{att}})))))) \tag{5.32}$$

σ is the sigmoid activation; it focuses attention on spatial features critical for seismic interpretation. This is particularly effective at distinguishing geological boundaries such as fault planes and channel systems.

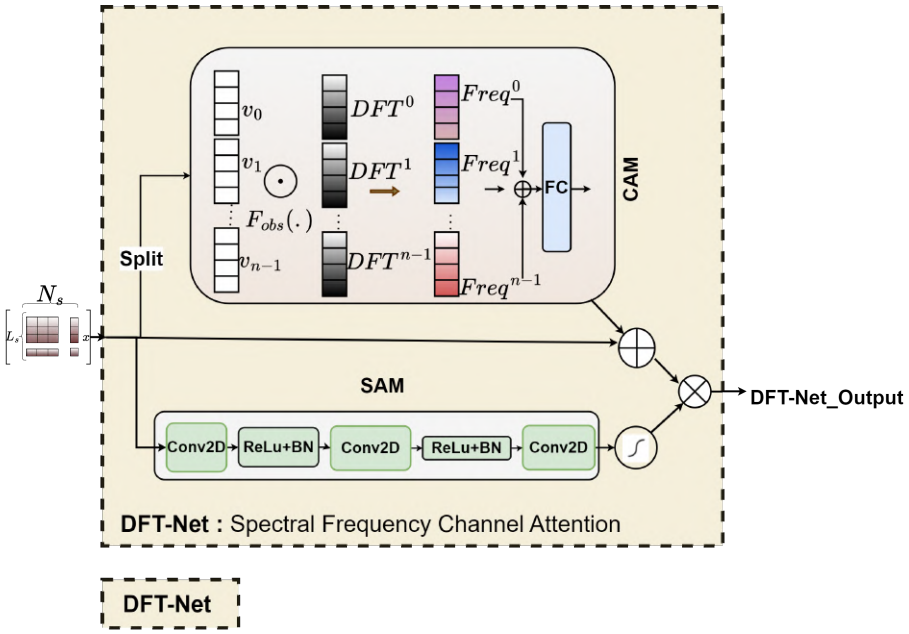


Figure 5.10: DFTnet: Spectral Frequency Channel Attention

iii. Combination of Attention Mechanisms

Combining channel and spatial attention yields a composite feature map accentuating both frequency and spatial features:

$$X_{\text{combined}} = X_{\text{att}} \odot Y \tag{5.33}$$

X_{combined} merges the enhanced frequency and spatial characteristics, where $\odot$ denotes element-wise multiplication. This combined attention ensures high sensitivity to fine seismic detail, facilitating more accurate subsurface property reconstruction.

DFTnet-CBAM adaptively focuses on high-frequency content and spatial features, crucial for resolving fine-grained seismic detail. This enhances prediction accuracy and resolution of geological formations. Figure 5.10 shows the DFTnet spectral channel attention block; Figure 5.9 illustrates its integration within the U-Net backbone.

Figure 5.11 shows real-data results; the DFT-Net layer is placed at the end of each encoding and decoding stage. **DFT-NET Real data results**

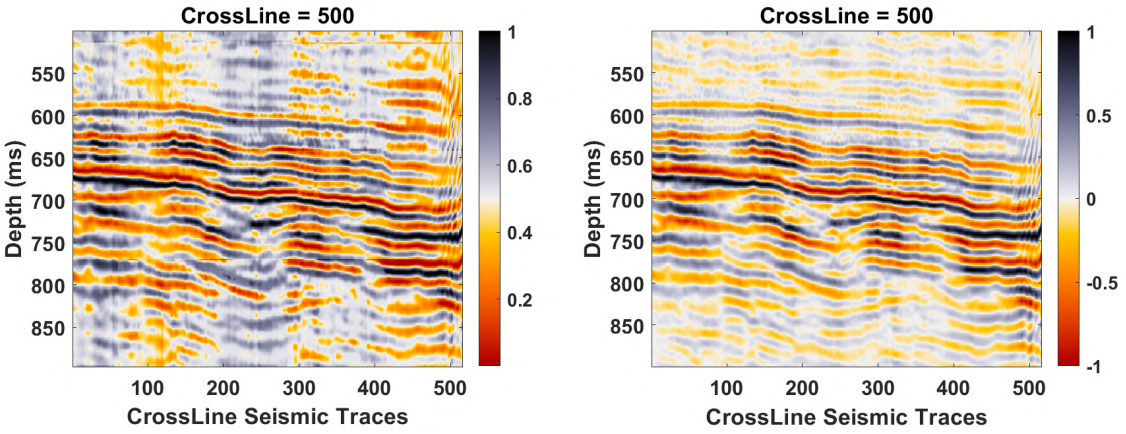


Figure 5.11: DFT-Net comparison with real data

5.3.7 Method IV: DWT-Net — A Discrete Wavelet Transform (DWT) Based CBAM Attention Framework within a U-Net Backbone for Seismic Inversion

Integrating DWT with Haar wavelets into U-Net, combined with CBAM, represents a significant innovation for seismic data processing. DWTnet follows DFTnet's structural philosophy but uses wavelet transforms to enhance feature extraction and compression.

I. DWT Layer Implementation

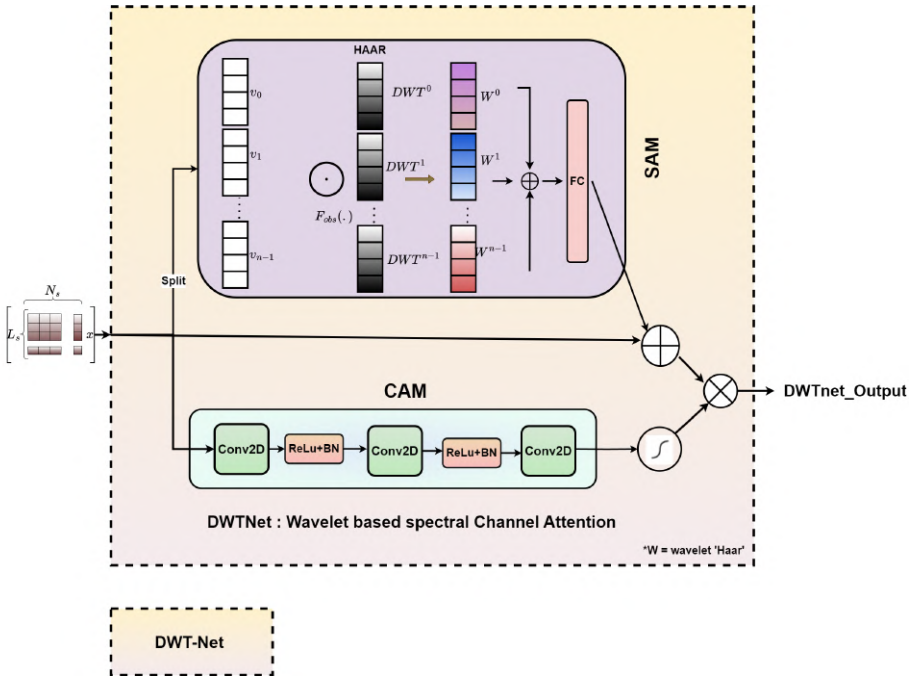


Figure 5.12: DWTNet: A wavelet-based spectral channel attention

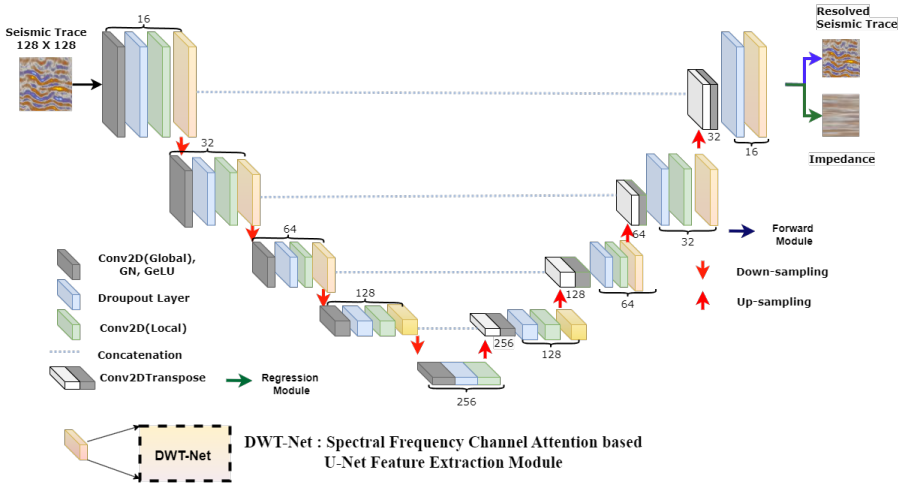


Figure 5.13: Seismic super-resolution with DWT-net enhanced U-Net

The DWT layer uses the Haar wavelet for its simplicity and effectiveness at capturing abrupt changes common in seismic images. It decomposes input data into approximation and detail coefficients, highlighting different seismic aspects.

i. 'Haar' Wavelet Transform

The Haar transform passes the signal through low- and high-pass filters and down-samples to reduce dimensionality while retaining key information:

$$a_k = \frac{1}{\sqrt{2}} (x_{2k} + x_{2k+1}) \tag{5.34}$$

$$d_k = \frac{1}{\sqrt{2}} (x_{2k} - x_{2k+1}) \tag{5.35}$$

where a_k and d_k represent the approximation and detail coefficients, respectively, and x denotes the input signal at position k.

ii. Integration into U-Net

In DWTnet, each convolutional block is followed by a DWT layer. This enables the network to capture both broad layout and minute detail, improving detection of fractures and fault lines.

II. Incorporation of CBAM for Enhanced Attention

Like DFTnet, DWTnet incorporates CBAM to refine wavelet-generated feature maps through two synergistic components: channel attention and spatial attention.

i. Channel Attention Mechanism

This mechanism prioritizes features by exploiting inter-channel relationships:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{5.36}$$

where F represents the input features, MLP denotes a multi-layer perceptron, $AvgPool$ and $MaxPool$ are average and max pooling operations across spatial dimensions, and σ is the sigmoid activation function.

ii. Spatial Attention Mechanism

Spatial attention emphasises where to focus within the input feature map:

$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \tag{5.37}$$

where $f^{7\times7}$ is a 7×7 convolution applied to the channel-axis concatenation of average and max pooled features.

III. Integration and Output

Combining Haar-wavelet DWT and CBAM inside U-Net yields a robust model for complex seismic data. The output is a detailed feature map that significantly aids subsurface interpretation. DWTnet thus provides a substantial upgrade over traditional methods for effective seismic exploration.

DWT-NET Real data results

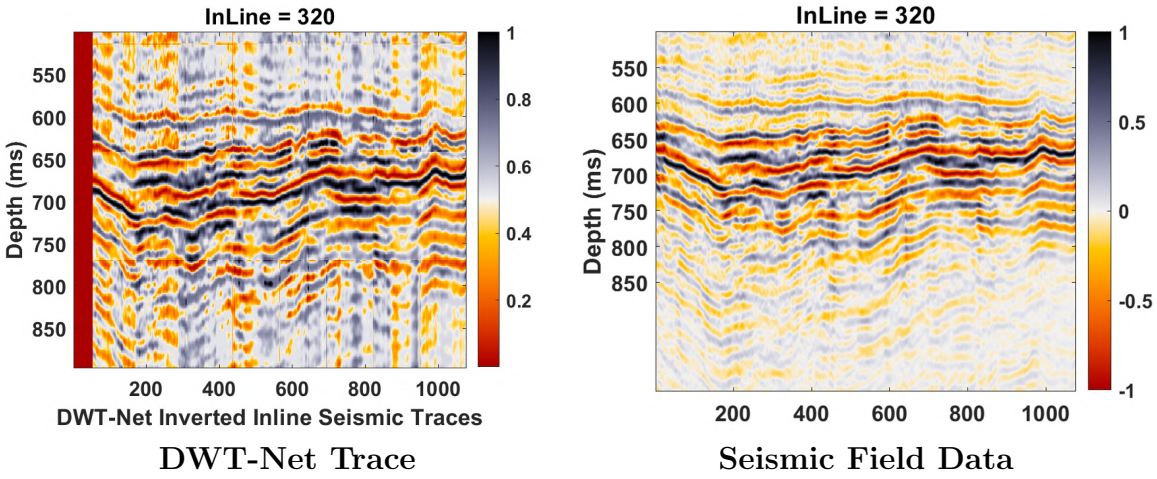


Figure 5.14: DWT-Net CrossLine comparison with real data

5.3.8 Method - V: DPSSnet — A Discrete Prolate Spheroid Sequence (DPSS)—Slepian Sequences Based Spectral Attention Using CBAM Attention Framework with U-Net Backbone in Seismic Inversion

Traditional networks often fail to capture the time-frequency characteristics essential for detailed subsurface exploration. Inspired by DFTnet, we propose DPSSnet, which integrates Discrete Prolate Spheroidal Sequences (DPSS) within a CBAM-enhanced U-Net to focus on signal components within a controllable time-bandwidth product, enabling more precise seismic time-frequency analysis.

1. DPSS Sequence Generation

   The DPSS (Discrete Prolate Spheroidal Sequences) sequence of length N, denoted as $\psi(t)$, is derived from the solution to a prolate spheroidal wave equation. It is defined as:

$$\psi(t) = \sqrt{\frac{N \cdot \gamma(\alpha)}{2\pi}} \cdot P_\alpha^\gamma(\cos(\pi t)), \tag{5.38}$$

   where: - N is the length of the sequence. - $\gamma(\alpha)$ is the normalizing constant depending on the parameter α. - P_α^γ is the prolate spheroidal wave function of order α.

2. Properties of DPSS Tapers:

   (a) **Orthogonality**: DPSS tapers are orthogonal to each other, allowing for independent application in multitaper spectral analysis. This property reduces variance and enhances spectral resolution.

   (b) **Projection:** Each input channel is projected onto the DPSS basis, which efficiently encodes both the time and frequency characteristics of the signal.

   (c) **Spectral Concentration:** The signal is transformed into a spectral domain where each component is optimally concentrated within its allocated bandwidth.

   (d) **Component Separation:** The transform outputs are then categorized into components based on their energy concentration effectiveness.

Given an input batch element x with channel index i, the transformation is represented as:

$$X_i(n) = \sum_{t=0}^{N-1} x_i(t) v_n(t) \tag{5.39}$$

iv. Concatenation of Outputs

The spectral components from all channels are concatenated into a comprehensive

feature tensor:

$$\text{Spectral Output} = \bigoplus_{i=1}^{C} X_i \tag{5.40}$$

with $\bigoplus$ denoting the concatenation operation over C channels.

II. High-Frequency Channel Attention Function

DPSSnet uses CBAM to enhance high-frequency components extracted by DPSS, crucial for detecting subtle geological features.

i. Normalization of DPSS Components

Normalisation standardises DPSS component values, facilitating learning:

$$X_{\text{norm}} = \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{5.41}$$

where μ and σ^2 are the mean and variance of X across its spatial dimensions, respectively, and ϵ is a small constant to prevent division by zero.

ii. Pooling and Combination of Features

The normalised components are pooled to capture average and maximum spectral features, then combined and passed through dense layers.

III. DPSSnet-CBAM

Integrating DPSS with CBAM within U-Net enhances seismic inversion by combining high-frequency channel attention with spatial attention, significantly improving resolution and accuracy.

5.3.8.1 DPSS-Net Encoder and Decoder Architecture

The DPSS-Net architecture (Figures 5.15 and 5.16) integrates DPSS within the CBAM framework. The DPSS layer, nested inside CBAM, captures high-frequency attention and is applied before max-pooling at the end of each encoding layer. It enhances focus on high-frequency seismic components by leveraging DPSS orthogonality

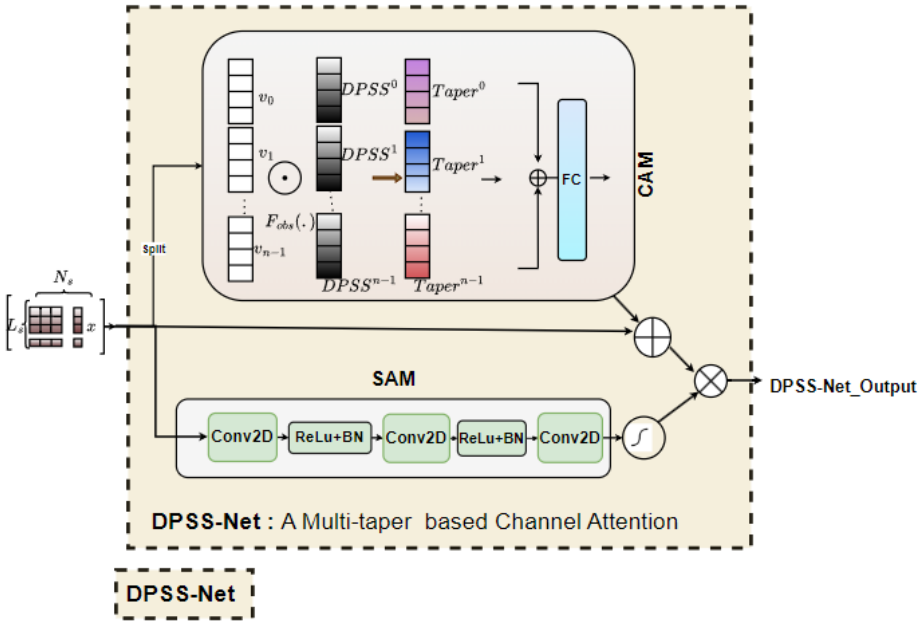


Figure 5.15: DPSS-Net Layer

and optimal time-frequency concentration, ensuring each encoding layer receives enhanced spectral analysis before pooling.

5.3.8.2 DPSS-Net Result

Figure 5.17 shows DPSS-Net results. Thin layers are resolved with greater accuracy than previous methods.

5.4 Network Training and U-Net with Attention Mechanisms

A GAN framework is used to train the network to invert seismic data by matching synthetic to real distributions (Figure 5.18). The U-Net combines attention mechanisms from AG, SE, CBAM, DFT-Net, DWT-Net, and DPSS-Net (Figure 5.16), using the Wasserstein GAN (WGAN) methodology for effective domain adaptation. Figure 5.19 illustrates the full attention-in-UDA setup.

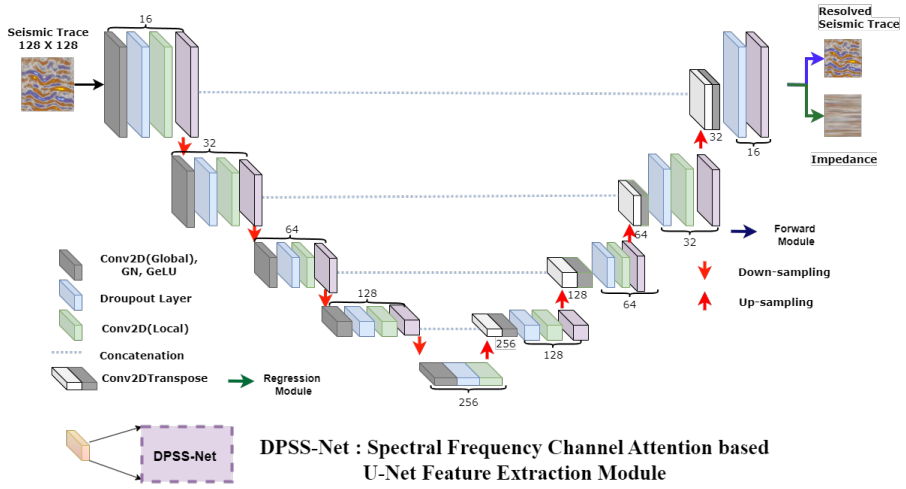


Figure 5.16: DPSS-Net Encoder and Decoder Architecture

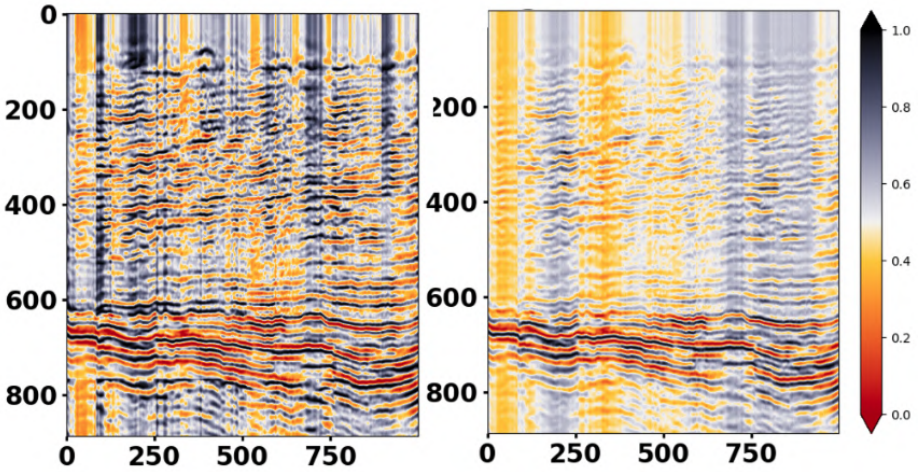


Figure 5.17: DPSS Net Real data Comparison A)DPSS-NET B) Seismic trace data

5.4.1 U-Net with Different Attention Mechanisms

Figure 5.19 shows how AG, SE, CBAM, DFT-Net, DWT-Net, and DPSS-Net are combined within the U-Net for UDA, helping the network focus on important features and improve seismic inversion accuracy.

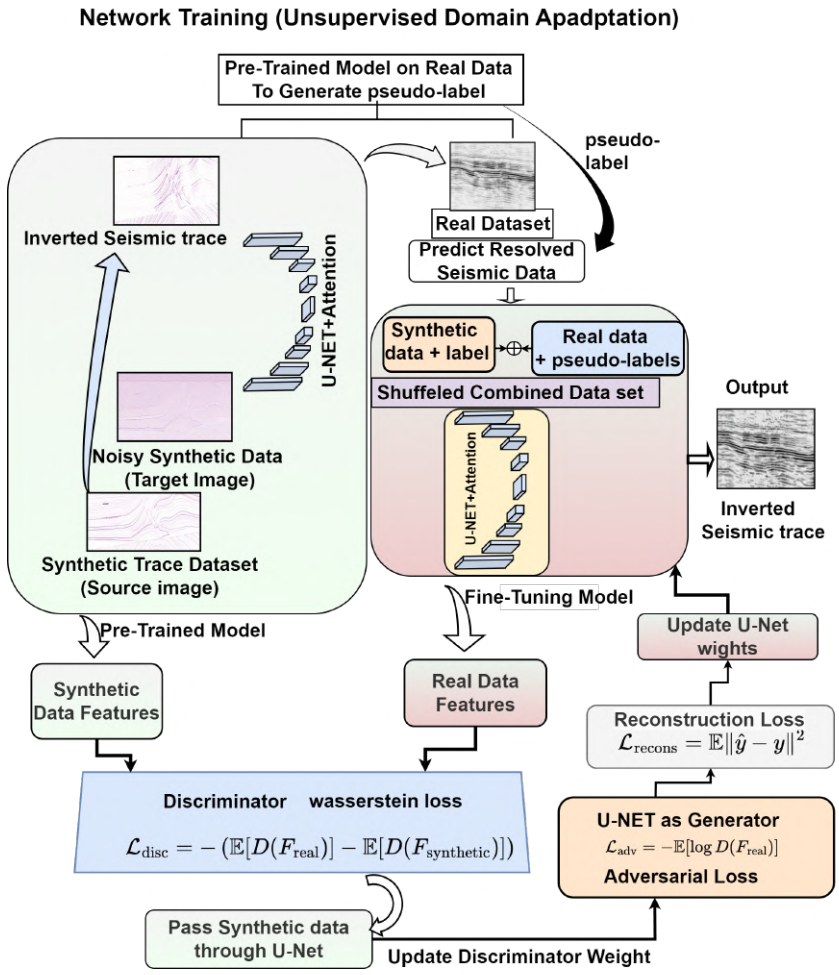


Figure 5.18: Fine-Tuning Unet attention with Domain Adaptation from synthetic model

5.4.2 Training the Critic

The Critic (discriminator) is trained to differentiate between real seismic data and data generated by the U-Net attention (generator). The Critic aims to maximize the

Wasserstein distance between these two distributions. The loss function for the Critic D is:

$$L_D = \mathbb{E}_{x\sim P_r}[D(x)] - \mathbb{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] + \lambda \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2] \tag{5.42}$$

where x represents samples from the real data distribution P_r, and $\tilde{x}$ represents

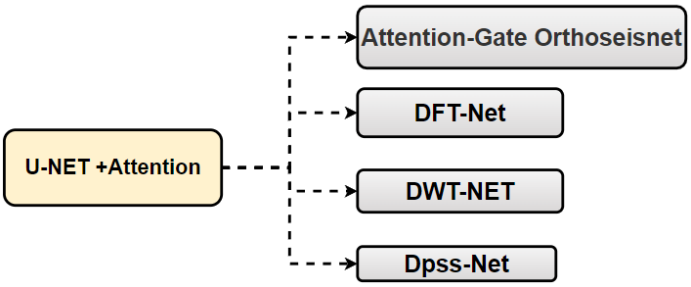


Figure 5.19: Unet with different attention in UDA

samples from the generated data distribution P_g obtained by passing synthetic data through the U-Net. The term $\hat{x}$ denotes interpolated samples between real and generated data, and λ is the gradient penalty coefficient.

5.4.3 Training the U-Net Attention

The U-Net generator (Figures 5.18 and 5.19) is trained to minimise adversarial and reconstruction losses. The adversarial loss encourages outputs the Critic cannot distinguish from real data; the reconstruction loss ensures fidelity to the synthetic input. The adversarial loss L_{adv} is:

$$L_{\text{adv}} = -\mathbb{E}_{\tilde{x}\sim P_g}[D(\tilde{x})] \tag{5.43}$$

The reconstruction loss L_{rec} is defined as the L_1 loss between the synthetic data x_{syn} and the generated output $G(x_{\text{syn}})$:

$$L_{\text{rec}} = \mathbb{E}_{x_{\text{syn}}\sim P_{\text{syn}}}[||x_{\text{syn}} - G(x_{\text{syn}})||_1] \tag{5.44}$$

The total loss L_G for training the U-Net attention is a weighted sum of these losses:

$$L_G = L_{\text{adv}} + \alpha L_{\text{rec}} \tag{5.45}$$

where α is a balancing weight for the adversarial and reconstruction losses.

5.4.4 Training Procedure

Training alternates between updating the Critic and the U-Net. The Critic is updated by computing Wasserstein loss and gradient penalty on real and synthetic features, then updating its weights. The U-Net is then updated by computing adversarial and reconstruction losses on synthetic data. This alternation repeats until convergence, adapting the U-Net to real seismic data.

5.4.5 Fine-Tuning

After initial training, the U-Net undergoes fine-tuning on a dataset combining real data and pseudo-labels. The Critic is refined with adversarial loss, the model further tuned with real data features, and U-Net weights updated with resolved seismic outputs. Combined with multi-scale frequency layers and attention mechanisms, this enables the U-Net to generate high-fidelity seismic data, bridging the synthetic-to-real domain gap.

5.5 Results and Discussion

Performance of six U-Net variants is evaluated on synthetic data at three noise levels (0, 5, and 10 dB) and on real seismic data, using MSE, MAE, and SSIM for synthetic data, with R^2 added for real data.

5.5.1 Synthetic Data Performance at Various Noise Levels

Noise Level: 0dB

At 0 dB, UNet-DPSS achieves the best results (MSE 0.005, MAE 0.040, SSIM 0.990).

Method	Noise 0 dB			Noise 5 dB			Noise 10 dB		
	MSE↓	**MAE↓**	**SSIM↑**	**MSE↓**	**MAE↓**	**SSIM↑**	**MSE↓**	**MAE↓**	**SSIM↑**
UNet-AG	0.012	0.063	0.975	0.025	0.118	0.862	0.036	0.173	0.820
UNet-SE	0.010	0.057	0.981	0.022	0.108	0.869	0.033	0.162	0.828
UNet-CBAM	0.008	0.050	0.984	0.019	0.097	0.876	0.029	0.151	0.834
UNet-DFT	0.007	0.045	0.987	0.015	0.087	0.884	0.025	0.141	0.843
UNet-DWT	0.006	0.042	0.988	0.012	0.080	0.891	0.022	0.133	0.849
UNet-DPSS	**0.005**	**0.040**	**0.990**	**0.010**	**0.075**	**0.895**	**0.020**	**0.125**	**0.855**

Table 5.4: Performance comparison of different methods under varying noise levels. **Bold** = best per metric at each noise level. ↓ lower is better; ↑ higher is better.

Performance degrades progressively through UNet-DWT, UNet-DFT, UNet-CBAM, UNet-SE, and UNet-AG, with increasing MSE and MAE and decreasing SSIM.

Noise Level: 5dB

At 5 dB, UNet-DPSS again leads (MSE 0.010, MAE 0.075, SSIM 0.895), with the same degradation trend from UNet-DWT to UNet-AG.

Noise Level: 10dB

At 10 dB, UNet-DPSS achieves MSE 0.020, MAE 0.125, and SSIM 0.855, maintaining its lead despite higher noise, followed in order by UNet-DWT, UNet-DFT, UNet-CBAM, UNet-SE, and UNet-AG.

5.5.2 Real Seismic Data Performance

On real seismic data, UNet-DPSS achieves MSE 0.025, MAE 1.00, SSIM 0.890, and R^2 0.900, demonstrating that its advantage extends from synthetic to real data.

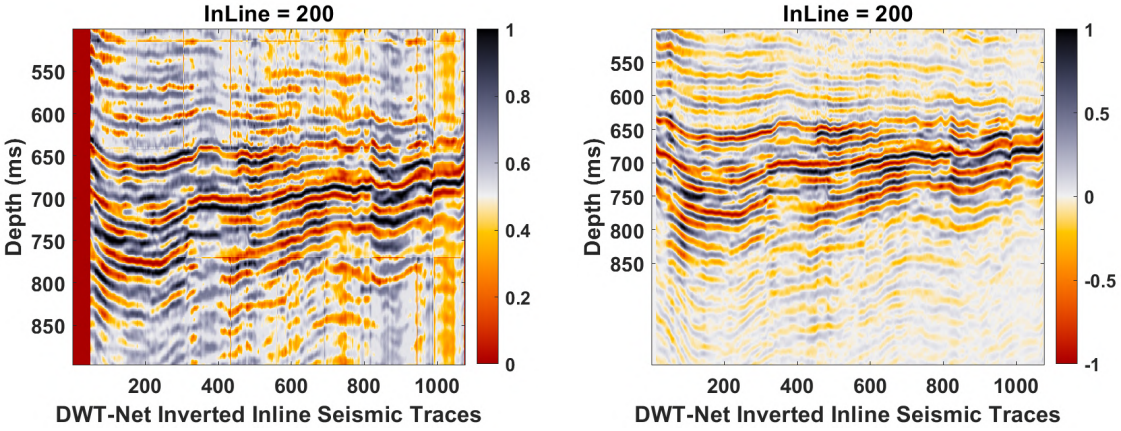
Table 5.5 confirms UNet-DPSS as the top performer (MSE 0.025, MAE 0.095, SSIM 0.890, R^2 0.900), followed by UNet-DWT (MSE 0.026, MAE 1.04, SSIM 0.888, R^2 0.899). UNet-DFT (SSIM 0.887), UNet-CBAM (SSIM 0.884), UNet-SE (SSIM 0.882), and UNet-AG (SSIM 0.876) show progressively weaker results; the SSIM gap is largest between UNet-AG and UNet-SE, reflecting the particular effectiveness of channel recalibration for spectrally structured seismic data.

Method	MSE↓	MAE↓	SSIM↑	R^2 ↑
ResANet [69]	1.152	3.58	0.6377	0.7171
UNet [72]	0.8541	2.964	0.6891	0.7382
InversionNet [66]	0.2346	1.391	0.7830	0.8141
OrthoSeisNet (Ours)	0.02678	1.102	0.884	0.8945
UNet-AG (Ours)	0.038	1.32	0.876	0.886
UNet-SE (Ours)	0.033	1.21	0.882	0.892
UNet-CBAM (Ours)	0.031	1.15	0.884	0.895
UNet-DFT (Ours)	0.028	1.08	0.887	0.897
UNet-DWT (Ours)	0.026	1.04	0.888	0.899
UNet-DPSS (Ours)	**0.025**	**0.095**	**0.890**	**0.900**

Table 5.5: Performance comparison on real seismic data. **Bold** = best per metric. ↓ lower is better; ↑ higher is better. Grey rows = external baselines.

5.5.3 DWT-Net: Real Data Visual Analysis

Figure 5.20 compares the DWT-Net output against the original KG Basin seismic field data at InLine 200, providing a zone-by-zone assessment of thin-bed resolution improvement.



DWT-Net Output **Original Seismic Field Data**

Figure 5.20: DWT-Net output vs. original KG Basin seismic at InLine 200. Left: DWT-Net processed output. Right: raw field seismic data.

Three diagnostic zones are identified:

Shallow chaotic zone (530–580 ms). The original data shows laterally incoher-

ent, low-amplitude noise consistent with geological complexity. DWT-Net introduces blocky, patchy artefacts in this region, indicating over-smoothing in low-SNR zones where the network does not generalise well to patterns absent from the synthetic training set.

Major reflector package (600–760 ms). This is the primary diagnostic zone for thin-bed resolution. In the original data, the strong impedance contrast at 670–700 ms appears as a broad trough ($\geq$20 ms), blurred by wavelet side lobes, with closely spaced reflectors in the 620–660 ms band showing tuning interference that renders individual thin beds indistinguishable. DWT-Net resolves this zone into distinct narrow peaks: multiple thin beds in the 630–660 ms interval become individually separable, suggesting sub-tuning resolution recovery. The dominant reflector at $\sim$690 ms resolves as a tight, high-contrast band rather than a smeared trough, consistent with wavelet deconvolution. Estimated reflector FWHM narrows from 20–30 ms in the original to 8–12 ms in the DWT-Net output, indicating apparent vertical resolution improvement beyond the classical $\lambda/4$ tuning limit.

Sub-target zone (760–880 ms). Both images show gradually degrading reflector continuity. DWT-Net events persist but with increasing lateral discontinuity as SNR decreases with depth, consistent with reduced training-data coverage at these depths.

The thin-bed improvement in the 630–760 ms interval is directly relevant to the KG Basin context, where stratigraphic traps involving thin sand/shale discrimination sit precisely in this depth range. However, the noise-zone artefacts (530–580 ms) highlight a known limitation of data-driven inversion: without explicit noise conditioning or physics-based regularisation, the network does not robustly handle geologically complex, low-SNR zones unseen during training.

5.6 Conclusion

Evaluating six U-Net variants on synthetic data at 0, 5, and 10 dB noise and on real seismic data reveals a consistent performance hierarchy. UNet-DPSS achieves the best results across all conditions—lowest MSE and MAE, highest SSIM and R^2—while UNet-AG performs the weakest. UNet-DWT, UNet-DFT, UNet-CBAM, and UNet-SE occupy intermediate positions in descending order, confirming that spectral

attention strength, particularly DPSS-based time-frequency concentration, is the key determinant of inversion quality.

CHAPTER 6

Attention-Seisformer: A Hybrid Multi-head and Parallel Self-Attention in Vision Transformers for Seismic Inversion

6.1 Introduction

Seismic inversion — recovering subsurface acoustic impedance from surface reflection recordings — bridges signal processing and geological interpretation. Classical approaches rely on an explicit convolutional forward model with well-log constraints to regularise the ill-posed problem. They are slow, require manual parameter tuning for each survey, and struggle when well coverage is sparse.

Deep learning methods now learn the seismic-to-impedance mapping directly [66, 129], trading physical rigour for speed and generalisability, and results on synthetic benchmarks have been competitive. Vision Transformers are particularly suited to this task: their self-attention mechanism captures long-range structural correlations

that convolutional networks miss at the patch level [120].

This chapter describes the **Attention-Seisformer**, a ViT-based data-driven framework built around one central observation: seismic data carries useful information simultaneously in the spatial, multi-resolution, and frequency domains, and no single attention mechanism exploits all three. The proposed architecture addresses this by running four attention pathways in parallel and injecting their outputs as additive biases into the scaled dot-product attention score, so the network can selectively draw on whichever domain is most informative for a given patch.

This framework has a clear scope. It does not solve an explicit wave equation, enforce well-log consistency, or recover a low-frequency background model from first principles. It learns a supervised mapping from post-stack seismic patches to acoustic impedance patches on synthetic data, then transfers the learned structural and spectral priors to field data via unsupervised domain adaptation — without requiring dense field impedance labels. Its principal role is therefore as a **decision-support tool**: it enhances reflector continuity, improves thin-layer delineation, and sharpens structural clarity sufficiently to support geological interpretation and preliminary well-site selection. It is not a replacement for conventional inversion where absolute impedance calibration is required. This places its supervised synthetic-training stage in the same category as InversionNet [66] and TransInvr [129], which adopt the same formulation and have become standard benchmarks in the field; the unsupervised domain-adaptation stage extends that pipeline to realistic field deployment conditions. The limitations this implies are discussed explicitly in Section 6.10.4.

6.2 Contributions

The chapter makes the following specific contributions:

1. A **parallel hybrid multi-domain attention fusion** module (Section 6.5.2) that injects channel, spatial, multi-resolution, and frequency-domain attention biases into a standard ViT encoder. The fusion weights are learnable and softmax-normalised, so the network decides how much each domain contributes without hand-tuning.

2. An **Efficient Channel and Spatial Attention Network (ECSAN)** that combines Enhanced Channel Attention, Coordinate Attention, and Spatial Attention into a single trainable module. Each sub-mechanism performs element-wise recalibration of the feature map before the result is projected into an attention bias matrix.

3. A **Discrete Wavelet Transform Attention (DWTA)** module that decomposes each feature channel into approximation and detail coefficient maps, flattens them into a token sequence, and computes self-attention across scales. This gives the encoder explicit access to multi-resolution structure — particularly important for thin-layer detection, where impedance contrasts span only a few samples.

4. A **Fast Fourier Transform Attention (FFTA)** module that transforms each feature into the frequency domain, concatenates the real and imaginary parts to form a real-valued representation, and computes self-attention over frequency components. This captures periodic acquisition patterns and laterally coherent reflectors that are difficult to see in the spatial domain alone.

5. A **confidence-guided impedance refinement decoder** based on the Attention-to-Mask (ATM) module. Rather than generating impedance directly from encoder features, ATM first estimates a spatially-varying confidence mask via cross-attention, then uses that mask to gate a coarse impedance estimate. Noisy or low-coherence regions are suppressed rather than extrapolated.

6. An **unsupervised domain adaptation protocol** that transfers geological and spectral priors learned on synthetic data to field seismic without requiring dense field impedance labels. Adaptation is achieved through feature-level alignment of domain-invariant encoder representations, validated against sparse well-log measurements at twelve borehole locations.

Figure 6.1 maps these contributions to the chapter workflow, from problem framing through architecture, training, and evaluation.

The overarching aim is not absolute impedance recovery but **structural resolution enhancement**: the framework improves reflector continuity and thin-layer

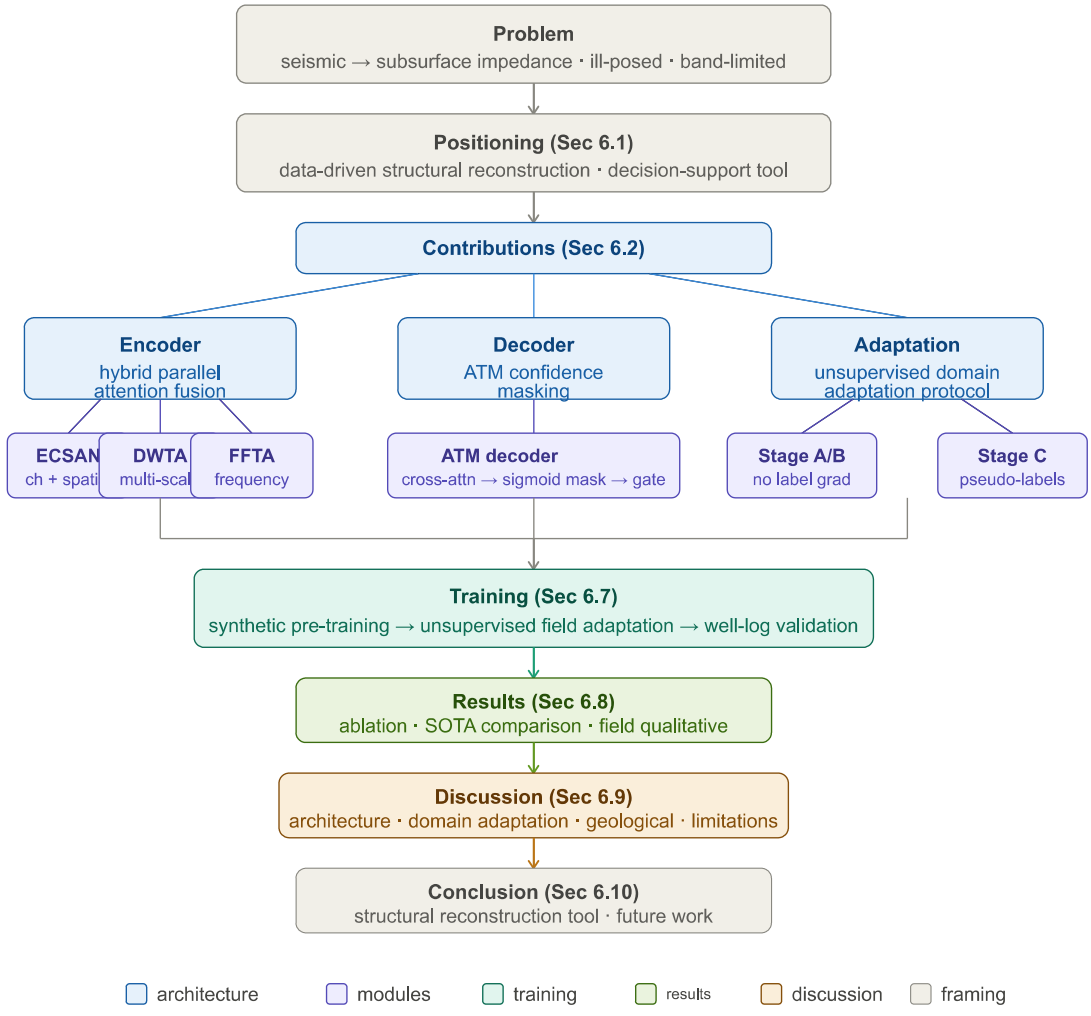


Figure 6.1: Chapter 6 workflow overview. Colour encodes conceptual layer: blue = architecture contributions, purple = individual modules, teal = training pipeline, green = results, amber = discussion.

delineation sufficiently for geological interpretation and preliminary well-site selection with limited ground truth. The one-line characterisation of the method is: *"A synthetic-supervised, unsupervised domain-adapted deep learning framework for enhancing structural resolution in seismic-derived impedance to support interpretation under limited ground truth."*

6.3 Background

Transformer-based methods entered seismic processing gradually. Early work adapted BERT-style feature extraction for trace classification [130] and positional embeddings for geoacoustic inversion [131]. CNN–Transformer hybrids followed, motivated by the observation that convolutional layers capture local texture while self-attention handles long-range correlations [110, 132, 137]. Masked autoencoders were applied to the incomplete-data problem, where well logs and seismic jointly constrain the subsurface model [133]. More recently, Swin Transformers with convolutional residual networks have been applied to denoising and interpolation [135], the Ensemble Dense Inception Transformer to horizon interpretation [136], and a self-supervised pretraining approach to signal reconstruction [138]. Li et al. applied a fully data-driven ViT to 3-D seismic inversion and achieved competitive results on the OpenFWI benchmark [129], and Xiang used Transformer models for wavefield simulation [234].

The Attention-Seisformer sits in this lineage but differs in two respects. First, multiple attention types operate simultaneously on the same token sequence and contribute additive biases to the same attention score, rather than concatenating CNN and Transformer outputs or alternating between them across layers. Second, the framework incorporates an unsupervised domain adaptation stage that transfers synthetic priors to field data without dense labels — an aspect absent from most existing benchmarks, which evaluate exclusively on held-out synthetic test sets.

6.4 Proposed Methodology

Figure 6.2 shows the overall Attention-Seisformer architecture. Seismic data enters as overlapping patches, is processed through a ViT encoder augmented with the par-

allel attention module, and decoded by the ATM confidence-masking decoder into a predicted impedance-like reconstruction. Figure 6.6 details the encoder–decoder interaction, and Figure 6.3 presents the complete end-to-end architecture, showing data flow from patch extraction through the hybrid ViT encoder (four parallel attention pathways) to the three-stage ATM decoder producing $\hat{Z}$.

6.4.1 Patch Extraction with Overlap

Input seismic sections are cropped to 1024×1024 samples (inline $\times$ crossline). Patches of 64×64 samples are extracted with a stride of 45 pixels, giving approximately 30% overlap between adjacent patches:

$$\text{stride} = 64 \times 0.7 \approx 45, \qquad N_x = N_y = \left\lfloor \frac{1024 - 64}{45} \right\rfloor + 1 = 22, \qquad T = 484 \quad (6.1)$$

The overlap prevents blocking artefacts when patches are reassembled into a full section. Boundary patches that extend beyond the image are zero-padded.

6.5 Hybrid Multi-Head Attention Mechanism

6.5.1 Tokenisation

An input image $I \in \mathbb{R}^{H \times W \times C}$ is flattened into patch tokens $f_0 \in \mathbb{R}^{L \times C}$, where $L = HW/P^2$ and $P = 64$. Learnable position embeddings are added. Layer normalisation is applied before each attention block (pre-norm), which improves gradient flow during the domain adaptation stages [235].

6.5.2 Parallel Attention Fusion

The standard ViT computes a single scaled dot-product attention per layer. The Attention-Seisformer instead runs four attention pathways simultaneously and injects each as an additive bias to the attention logits. The bias weights are learned:

$$\bar{w}_i = \frac{e^{w_i}}{\sum_{j=1}^{4} e^{w_j}}, \qquad \sum_{i=1}^{4} \bar{w}_i = 1 \quad (6.2)$$

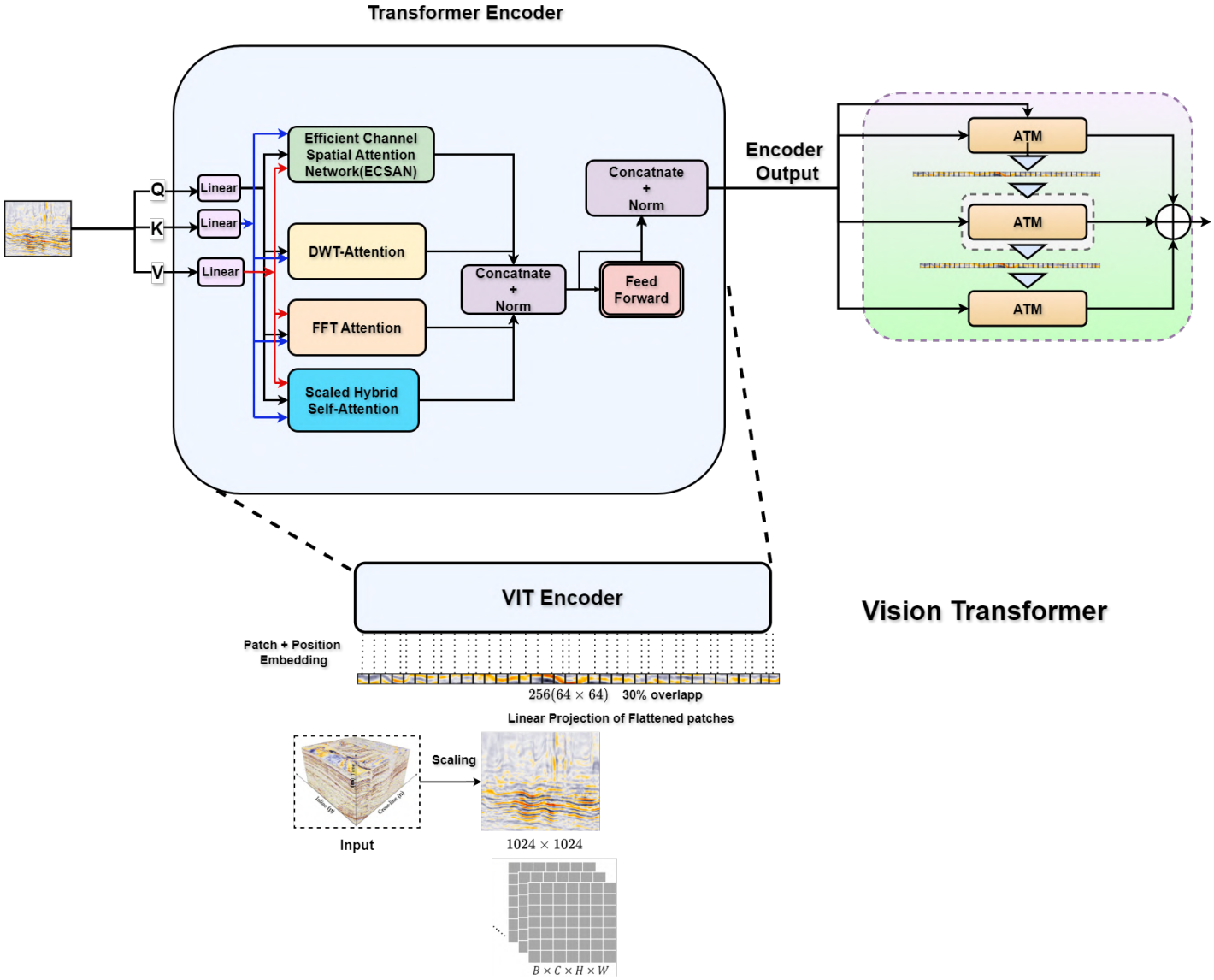


Figure 6.2: Attention-Seisformer overview. The encoder processes overlapping seismic patches through parallel ECSAN, DWTA, FFTA, and scaled dot-product attention pathways. The ATM decoder refines a coarse impedance estimate using a cross-attention-derived confidence mask.

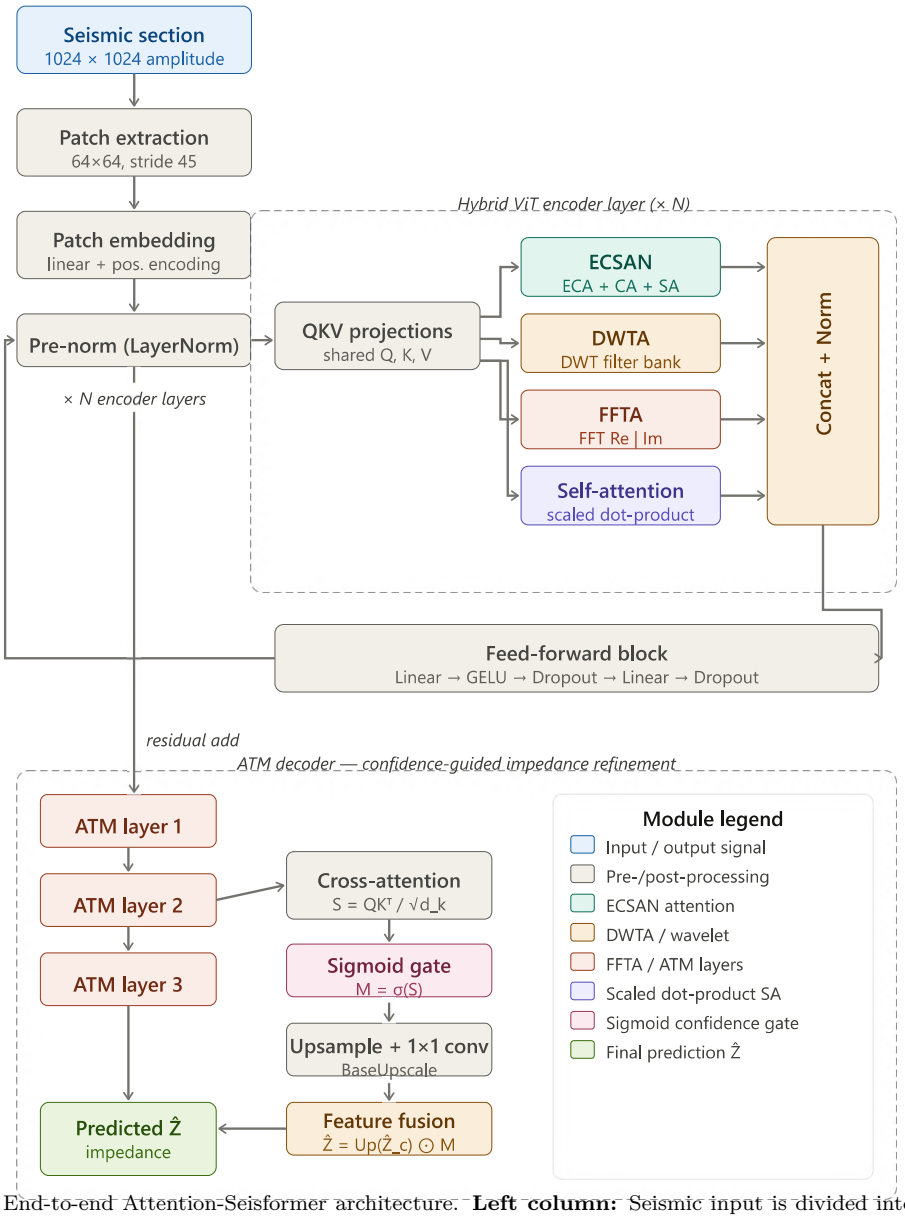


Figure 6.3: End-to-end Attention-Seisformer architecture. **Left column:** Seismic input is divided into overlapping 64×64 patches, linearly embedded with learnable position encodings, and passed through N encoder layers. **Centre:** Each encoder layer runs four parallel attention pathways (ECSAN, DWTA, FFTA, and scaled dot-product), whose outputs are injected as additive biases into the attention logits before the softmax operation. A residual connection and feed-forward network complete each encoder layer. **Right:** Three stacked ATM decoder layers generate a sigmoid confidence mask M through cross-attention between learned impedance query tokens G and encoder features F_i. The mask gates a bilinearly upsampled coarse impedance estimate to produce the final impedance prediction $\hat{Z}$.

where w_i are scalar parameters initialised to $\ln(1)$, providing equal weighting across all four pathways prior to training. The modified attention score is:

$$S(Q,K) = \frac{QK^\top}{\sqrt{d_k}} + \bar{w}_1 A_{\text{ECSAN}} + \bar{w}_2 A_{\text{DWTA}} + \bar{w}_3 A_{\text{FFTA}} \tag{6.3}$$

Each $A_i \in \mathbb{R}^{L\times L}$ is an attention bias matrix of the same shape as the dot-product score. The final output is $\text{softmax}(S)\,V$ as usual. Injecting at the logit level rather than blending output representations keeps the forward pass simple, without requiring an additional fusion head; the gradient flows back through $\bar{w}_i$ with a standard softmax Jacobian.

6.6 Attention Mechanisms

6.6.1 Efficient Channel and Spatial Attention Network (ECSAN)

Seismic feature maps carry information at different structural levels: channel-wise amplitude statistics, positional gradients along inline and crossline directions, and local spatial patterns around reflector interfaces. ECSAN captures all three simultaneously by combining three sub-mechanisms:

$$\mathbf{Y}_{\text{ECA}} = \mathbf{X} \odot \sigma\Big(\text{Conv1D}\Big(\tfrac{1}{HW}\textstyle\sum_{i,j} X_{ijc},\ k\Big)\Big) \tag{6.4}$$

$$\mathbf{Y}_{\text{CA}} = \mathbf{X} \odot \sigma(\text{Conv1D}(\text{Concat}[\text{Avg}_h(\mathbf{X}),\, \text{Avg}_w(\mathbf{X})]\,,\ k)) \tag{6.5}$$

$$\mathbf{Y}_{\text{SA}} = \mathbf{X} \odot \sigma(\text{Conv2D}(\text{Concat}[\text{Avg}_c(\mathbf{X}),\, \text{Max}_c(\mathbf{X})]\,,\ k{\times}k)) \tag{6.6}$$

$$\mathbf{Y}_{\text{ECSAN}} = \text{Norm}(\alpha\,\mathbf{Y}_{\text{ECA}} + \beta\,\mathbf{Y}_{\text{CA}} + \gamma\,\mathbf{Y}_{\text{SA}}) \tag{6.7}$$

ECA recalibrates channel amplitudes using global average pooling and a 1-D convolution — a compact approximation of squeeze-and-excitation that avoids the channel-reduction bottleneck. CA captures directional gradient information by pooling separately along the height and width axes; this is particularly useful for seismic data, where inline and crossline responses differ. SA sharpens spatial selectivity around reflector boundaries using average-max channel concatenation. The scalars α, β, γ

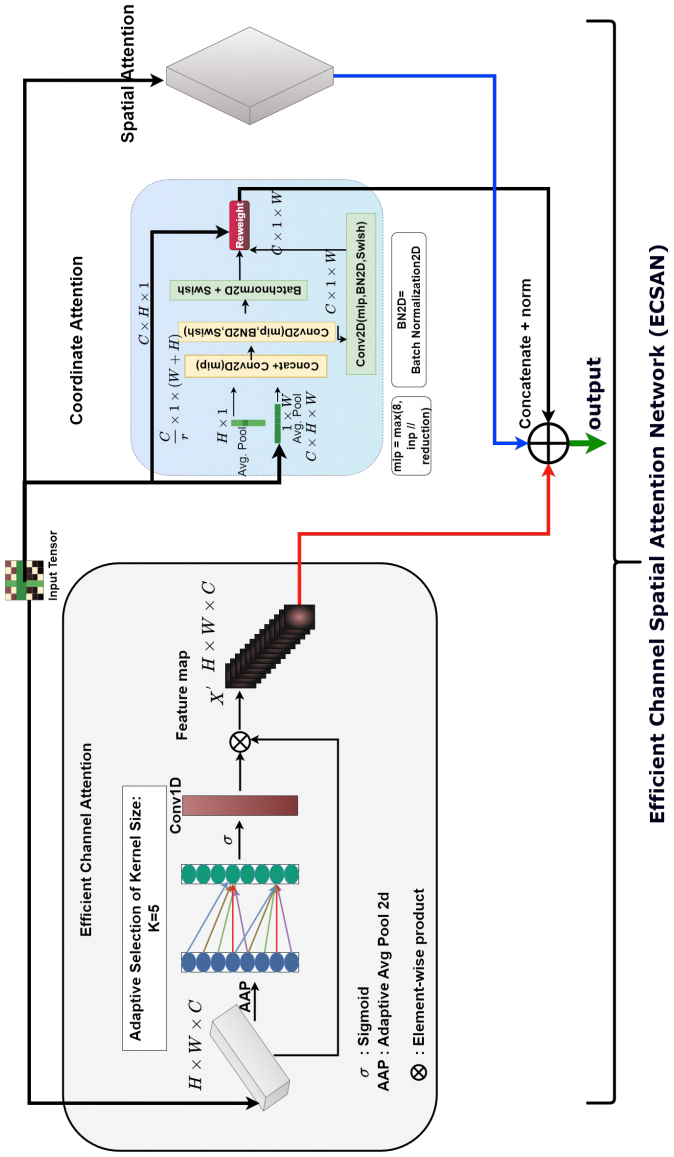


Figure 6.4: Efficient Channel and Spatial Attention Network (ECSAN). ECA, CA, and SA operate in parallel on the input feature map and their outputs are combined by learnable weights before layer normalisation.

are learnable. The ECSAN output is projected to $\mathbb{R}^{L\times L}$ for use as the attention bias:

$$A_{\text{ECSAN}} = (FW_e)(FW_e)^\top \tag{6.8}$$

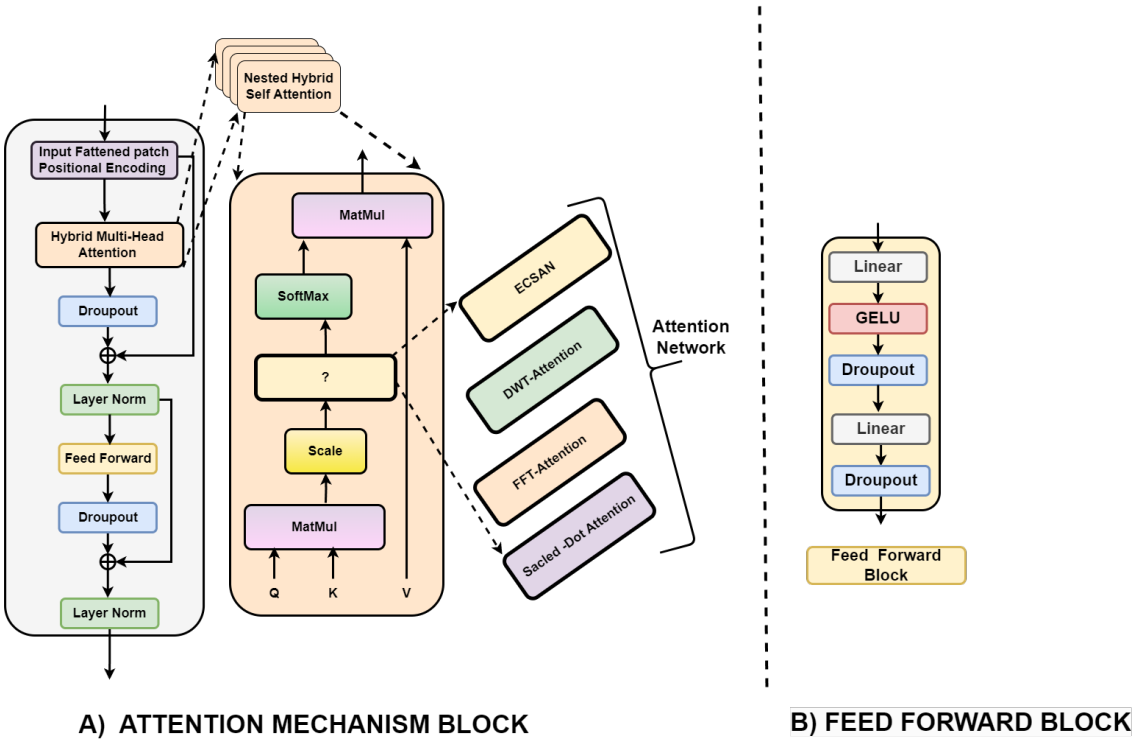


Figure 6.5: Attention network structures: (a) the hybrid encoder layer showing parallel ECSAN, DWTA, FFTA, and scaled dot-product branches merging as additive biases; (b) the feed-forward block.

6.6.2 Discrete Wavelet Transform Attention (DWTA)

Thin-layer detection in seismic inversion depends on resolving impedance contrasts at thicknesses near or below the tuning limit ($\lambda/4$). These features manifest as high-frequency detail in the wavelet domain, which spatial self-attention tends to smooth over because it treats all positions with equal resolution. DWTA addresses this by applying a discrete wavelet transform to each feature channel before computing attention, giving the encoder explicit access to multi-scale structure.

Each channel $\mathbf{x}_c \in \mathbb{R}^{H\times W}$ is decomposed via a discrete wavelet filter bank:

$$\{C_c^{LL},\, C_c^{LH},\, C_c^{HL},\, C_c^{HH}\} = \mathrm{DWT}(\mathbf{x}_c) \tag{6.9}$$

The four sub-bands are concatenated channel-wise into $\tilde{\mathbf{X}} \in \mathbb{R}^{4C\times H/2\times W/2}$, then flattened to maintain token-sequence consistency with the ViT architecture:

$$\tilde{F} = \mathrm{Flatten}(\tilde{\mathbf{X}}) \in \mathbb{R}^{L'\times 4C} \tag{6.10}$$

Q, K, V are projected from $\tilde{F}$ and attention is computed in the usual way. The output is projected back to $\mathbb{R}^{L\times L}$ to form A_{DWTA}.

6.6.3 Fast Fourier Transform Attention (FFTA)

Seismic acquisition imprints characteristic spectral patterns on recorded data — source signatures, cable response, and coherent noise trains appear as specific frequency components. FFTA operates in the frequency domain to capture dependencies among these components directly.

The 1-D DFT of each input channel is:

$$X[k] = \sum_{n=0}^{N-1} X[n]\, e^{-i2\pi kn/N} \tag{6.11}$$

Since $X[k] \in \mathbb{C}$, the real and imaginary parts are concatenated before projection to keep the computation real-valued:

$$\tilde{X}[k] = \big[\mathrm{Re}(X[k]) \mid \mathrm{Im}(X[k])\big] \in \mathbb{R}^{2C} \tag{6.12}$$

Retaining both parts preserves phase information, which encodes the lateral continuity of reflectors. Q, K, V are projected from $\tilde{X}[k]$ and attention over frequency bins yields A_{FFTA}.

6.7 Decoder: Confidence-Guided Impedance Refinement

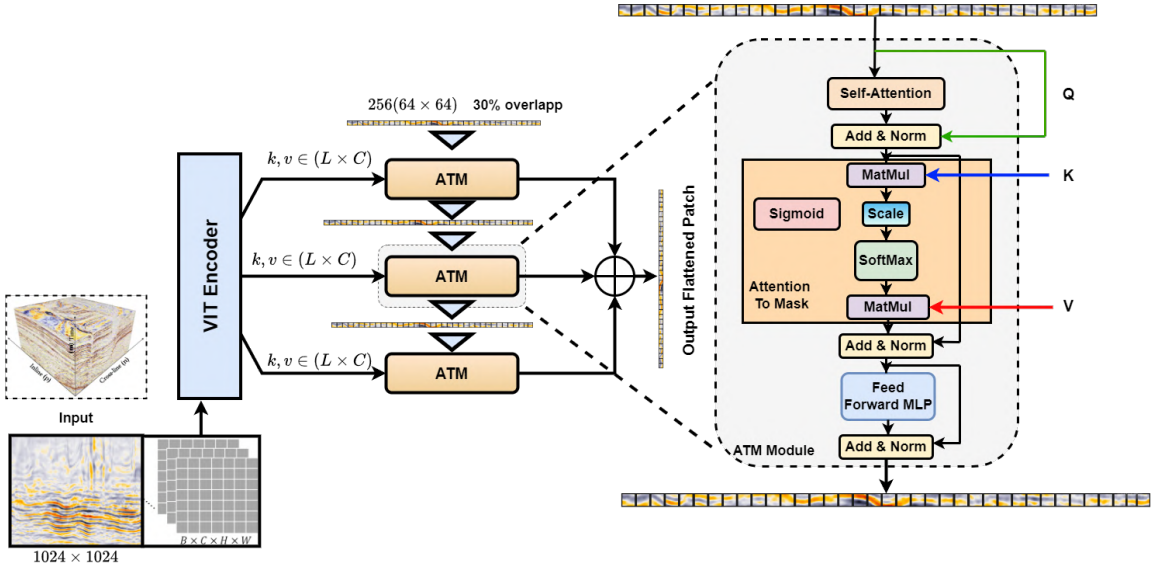


Figure 6.6: Encoder–decoder architecture. Encoder feature tokens $F_i \in \mathbb{R}^{L\times C}$ are passed to three stacked ATM layers; outputs are summed and projected to the impedance volume $\hat{Z}$.

Most decoder designs for seismic inversion map encoder features directly to an output through upsampling and convolution. The Attention-to-Mask (ATM) module takes a different approach: rather than predicting impedance from scratch, it first asks *which parts of the encoder output should be trusted*, then uses that answer to gate a coarse impedance estimate selectively.

The motivation follows from the data-driven setting itself. Without an explicit forward model, the network cannot verify that a predicted impedance value is physically consistent with the observed seismic at that location. ATM makes this uncertainty explicit by learning a spatially-varying confidence mask that preserves reliable predictions and suppresses uncertain ones — all end-to-end from impedance labels, without hand-crafted thresholds.

Step 1 — Cross-attention

Encoder feature tokens $F_i \in \mathbb{R}^{L\times C}$ from the final encoder layer and learned impedance query tokens $G \in \mathbb{R}^{L\times C}$ are linearly projected into query, key, and value spaces:

$$Q = W_q\, G, \qquad K = W_k\, F_i, \qquad V = W_v\, F_i \tag{6.13}$$

The scaled dot-product score measures how strongly each encoder feature matches the impedance query at every spatial position:

$$S(Q,K) = \frac{QK^\top}{\sqrt{d_k}} \tag{6.14}$$

The standard softmax-weighted value aggregation then yields a feature-weighted representation aligned with the impedance query:

$$\mathrm{Attn}(G,F_i) = \mathrm{softmax}(S(Q,K))\, V \tag{6.15}$$

In effect, the query tokens ask: *which encoder features are relevant for impedance prediction at this location?*

Step 2 — Confidence mask

The same score matrix $S(Q,K)$ — before softmax normalisation — is passed through a sigmoid function to produce a spatially-varying confidence mask:

$$M(G,F_i) = \sigma(S(Q,K)) \in [0,1]^{L\times L} \tag{6.16}$$

Sigmoid is used here instead of softmax because the mask must encode *absolute* confidence at each location independently, not a normalised distribution across positions. Locations where encoder features strongly match the impedance query produce mask values close to one; ambiguous or noise-dominated locations produce values near zero.

Step 3 — Gated impedance estimate

A coarse impedance map $\hat{Z}_{\text{coarse}}$ is obtained by bilinearly upsampling the decoder's intermediate representation and projecting it to the impedance channel space via a 1×1 convolution. The confidence mask is resized to match and applied element-wise:

$$\hat{Z} \;=\; \text{Upsample}\Big(\hat{Z}_{\text{coarse}}\Big) \;\odot\; \text{Resize}(M(G, F_i)) \tag{6.17}$$

The element-wise product suppresses the coarse estimate in low-confidence regions (mask ≈ 0) and passes it through unchanged where the network is confident (mask ≈ 1). Three ATM layers are stacked; their outputs are summed before a final linear projection to the impedance volume (Figure 6.6).

Interpretation

This gating mechanism differs fundamentally from a super-resolution decoder. A super-resolution module adds fine spatial detail to a low-resolution image. The ATM module does not add detail — it decides *where to trust* the coarse estimate and scales it accordingly. The mask behaves as an implicit uncertainty map: high values indicate structurally coherent, high-confidence predictions; low values flag zones where the encoder representation is ambiguous, such as fault shadows or noise-dominated intervals.

The mask is learned end-to-end from synthetic impedance labels and transferred to field data during unsupervised domain adaptation. No explicit uncertainty loss is required; the confidence signal emerges from the cross-attention scores naturally. In summary, the ATM module acts as a **confidence-guided gating mechanism** that selectively refines impedance predictions based on feature relevance, rather than generating impedance values directly from decoder convolutions. A detailed architecture can be shown in Figure 6.7.

6.8 Network Training

The training protocol for Attention-Seisformer represents a significant contribution of this work, combining supervised pre-training on synthetic data with unsupervised

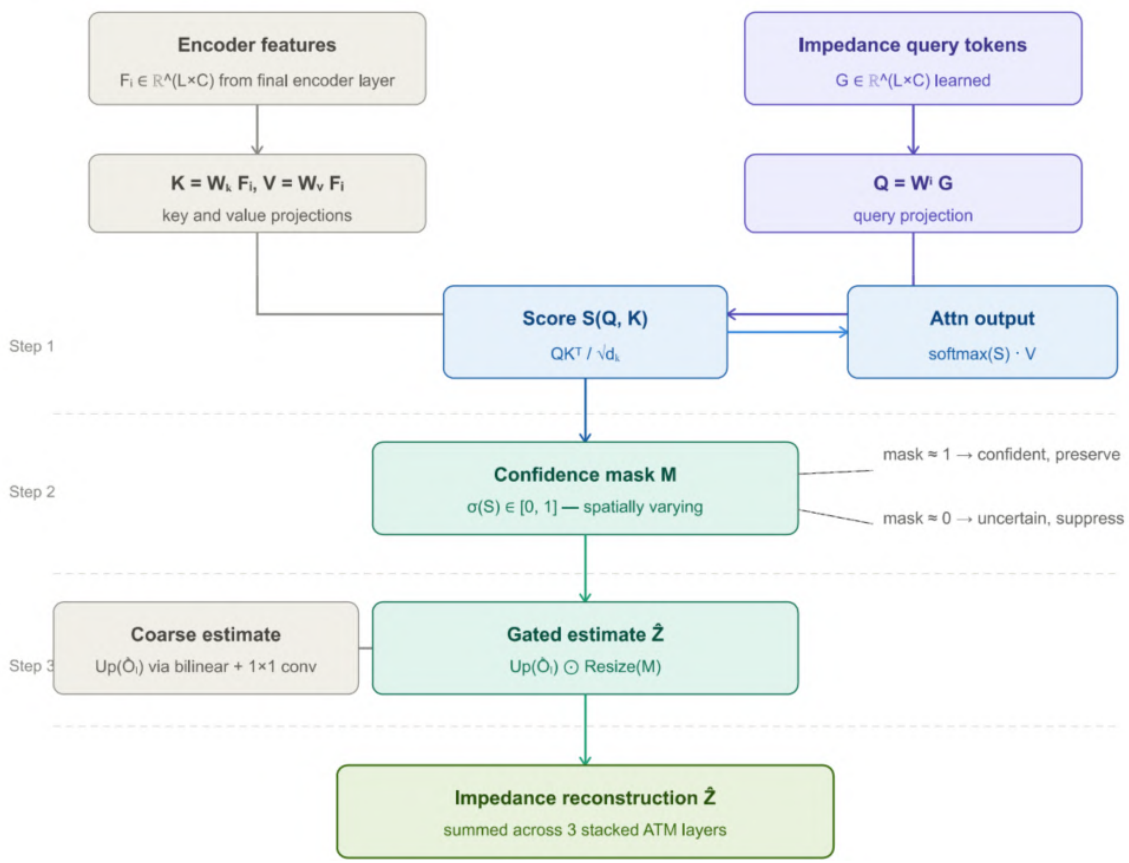


Figure 6.7: ATM decoder — three-step confidence-guided impedance refinement. Encoder features F_i and learned query tokens G enter cross-attention (Step 1); the raw score matrix S drives both the attention output and a sigmoid confidence mask M (Step 2); the mask gates a bilinearly upsampled coarse estimate $\hat{Z}_c$ to produce the final reconstruction $\hat{Z}$ (Step 3). Three ATM layers are stacked; outputs are summed before projection to the impedance volume.

domain adaptation for field deployment. Unlike conventional seismic inversion workflows that require extensive manual parameter tuning and well-log constraints, our approach learns a robust seismic-to-impedance mapping through a carefully designed two-phase training strategy. The complete workflow is shown in Figure 6.8

6.8.1 Dataset and Pre-processing

Training uses forward-modeled seismic–impedance pairs from the Marmousi2 suites [137]. It consists of a post-stack seismic section and its corresponding band-limited acoustic impedance model. Data augmentation applies polarity reversal, random time shifts (± 5 ms), and additive Gaussian noise at SNR $\in [20, 40]$ dB. A stratified 70/15/15

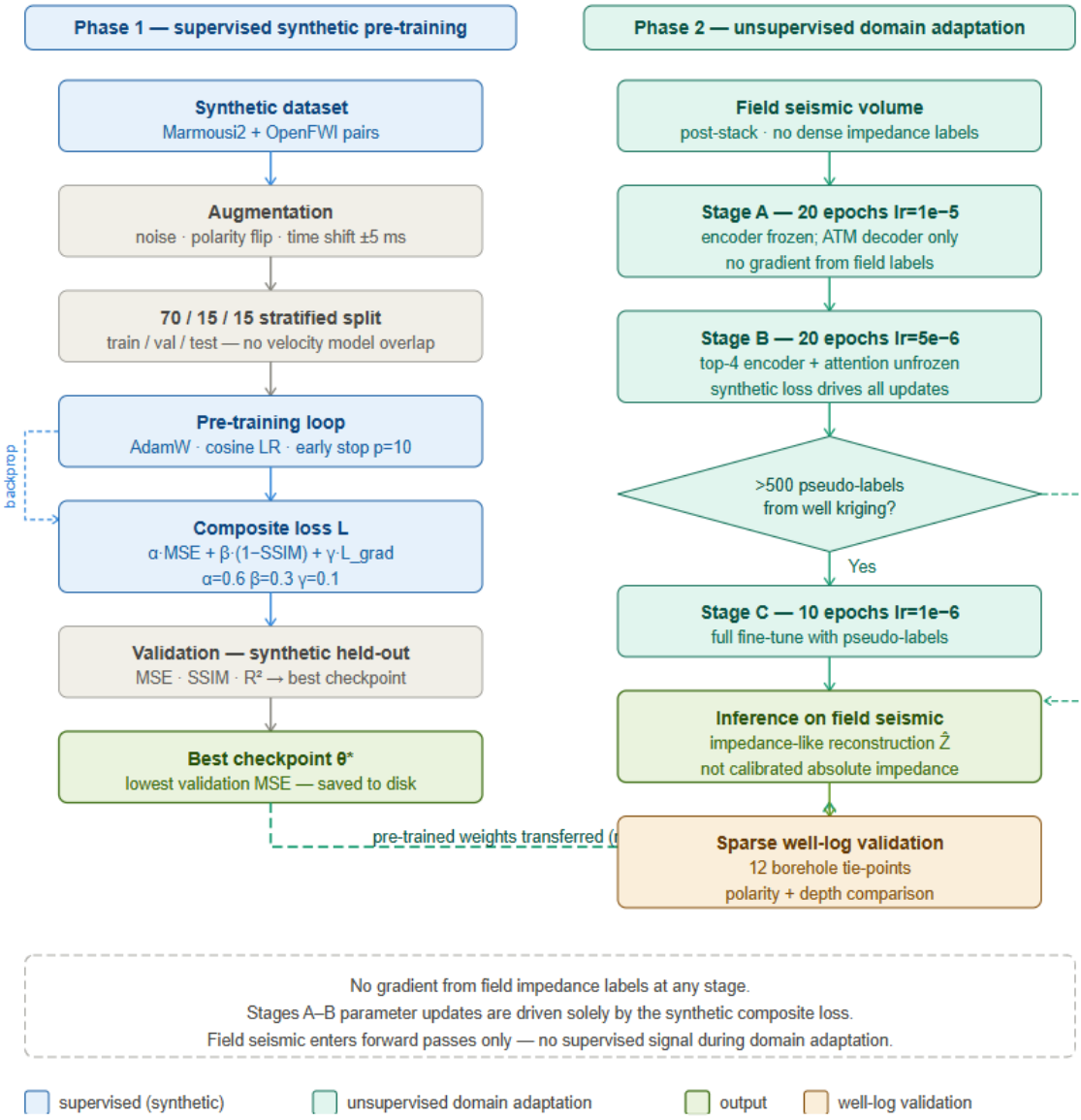


Figure 6.8: Attention-Seisformer training protocol. Phase 1: supervised pre-training on synthetic data (Marmousi2) with MSE+SSIM+gradient loss. Phase 2: unsupervised domain adaptation via graduated unfreezing. Solid arrows: gradient flow from synthetic loss only. Dashed arrows: forward-pass field exposure (no gradient). Stage A: decoder trainable, encoder frozen. Stage B: upper encoder + attention modules unfrozen. Stage C (optional): fine-tuning with well-derived pseudo-labels. No gradient from field impedance labels.

train–validation–test split ensures no seismic traces from the same velocity model appear in more than one partition. This framework assumes that structural and spectral relationships learned from synthetic data generalise to field data; unsupervised domain adaptation (Section 6.8.4) transfers the learned priors to real data distributions without requiring dense field impedance labels.

6.8.2 Loss Function

MSE-only training tends to produce spatially smooth predictions because the gradient pushes the network toward the conditional mean rather than sharp boundaries. A composite loss is used instead:

$$\mathcal{L} = \alpha\,\mathcal{L}_{\text{MSE}} + \beta\,(1 - \text{SSIM}(\hat{Z}, Z)) + \gamma\,\mathcal{L}_{\text{grad}} \tag{6.18}$$

where $\mathcal{L}_{\text{grad}} = \|\nabla_x \hat{Z} - \nabla_x Z\|_1 + \|\nabla_z \hat{Z} - \nabla_z Z\|_1$ penalises gradient mismatches along both spatial axes. The weights $\alpha = 0.6$, $\beta = 0.3$, $\gamma = 0.1$ were chosen by grid search on the validation set. The gradient term is primarily responsible for preserving impedance discontinuities at lithological boundaries.

6.8.3 Optimisation

Table 6.1 lists all training hyperparameters. AdamW with cosine annealing and a five-epoch linear warm-up is used throughout. Pre-norm layer normalisation before each attention block stabilises training, particularly during the domain adaptation stage where encoder weights are partially frozen.

6.8.4 Unsupervised Domain Adaptation Protocol

Pre-training on synthetic data (Phase 1 in Figure 6.8) gives the encoder a strong prior over seismic frequency content, geological structure, and impedance contrast patterns. Field data, however, differs from synthetic data in noise floor, bandwidth, and amplitude distribution. Directly applying a synthetically trained model to field data without any adaptation risks systematic prediction bias introduced by this domain gap.

Parameter	Pre-training (synthetic)	Domain adaptation (field)†
Optimiser	AdamW (β_1=0.9, β_2=0.999)	AdamW
Learning rate	10^{-4}	$10^{-5} \rightarrow 10^{-6}$
Weight decay	0.05	0.05
LR schedule	Cosine + 5-ep warm-up	Cosine annealing
Batch size	16	8
Epochs	100 (early stop p=10)	20 / 20 / 10 (staged)
Loss weights	α=0.6, β=0.3, γ=0.1	N/A (unsupervised)
Dropout	0.1	0.1
Patch size	64×64, stride 45	same
Hardware	HP Z-series, Intel Xeon, NVIDIA Quadro RTX 4000 (8 GB), 192 GB DDR4 ECC RAM	

† Domain adaptation is performed without a supervised field impedance loss. Epoch counts and learning rates refer to the graduated unfreezing schedule described in Section 6.8.4; no gradient signal from field labels is used at any stage.

Table 6.1: Training hyperparameters.

In the absence of dense field impedance labels, domain adaptation is performed in an **unsupervised** manner. No field impedance loss is computed at any stage. Instead, the encoder — pre-trained on synthetic data — is exposed to field seismic in a graduated unfreezing schedule that allows its internal representations to shift toward **domain-invariant structural features** while retaining the geological and spectral priors acquired during synthetic training. The model does not learn impedance values from field data; it learns to represent field seismic in a feature space compatible with the synthetic-trained decoder.

This is a form of **feature-level domain adaptation**. Encoder adaptation occurs through gradient updates driven by synthetic batches while being exposed to field-domain statistics during forward passes, which implicitly aligns feature distributions across domains. The effectiveness of the transfer is evaluated indirectly through structural consistency of the predicted section and comparison with sparse well-log impedance measurements at twelve borehole locations.

Three-stage graduated unfreezing minimises the risk of catastrophic forgetting — the well-documented failure mode in which rapid weight updates destroy priors acquired during pre-training:

Stage A (20 epochs, lr=10^{-5}): Encoder fully frozen; only the ATM decoder is trainable. Decoder weight updates are driven solely by synthetic batches. Field seismic is used only in forward passes without gradient computation, allowing the decoder to respond to field-domain feature distributions without explicit adaptation.

Stage B (20 epochs, lr=5×10^{-6}): The upper four encoder blocks and all attention fusion modules (ECSAN, DWTA, FFTA) are unfrozen. The fusion weights $\bar{w}_i$ shift to reflect field data spectral characteristics while the lower encoder layers continue to provide stable low-level feature representations grounded in synthetic training. Encoder weight updates are driven exclusively by the synthetic composite loss applied to synthetic batches; field batches are used only for forward-pass exposure without gradient contribution, ensuring that upper layers do not drift away from impedance-predictive representations.

Stage C (10 epochs, lr=10^{-6}): Full network adaptation, applied only when sparse well-log pseudo-labels (e.g., derived from well-log interpolation or conventional inversion constrained to the twelve available boreholes) provide sufficient local supervision — approximately 500 labelled traces. Below this threshold, Stage B is the stopping point and the model is deployed in inference mode.

The key distinction from supervised transfer learning is that no gradient signal from field impedance labels drives the encoder weight updates in Stages A and B. The network adapts its feature representations to the field domain through exposure to field seismic statistics during the forward pass, while all gradient updates remain grounded in the synthetic composite loss. This implicit adaptation strategy avoids the need for adversarial or distribution-matching losses, simplifying training while maintaining stability.

Since no supervised loss is applied against field impedance during Stages A and B, the model does not explicitly optimise impedance accuracy on field data. The output $\hat{Z}$ on field data represents an impedance-like structural estimate whose absolute values are anchored to the synthetic training distribution rather than to field-calibrated impedance. Quantitative interpretation therefore requires well-log tie at borehole locations to assess local calibration accuracy. No gradients are computed from field data at any stage; all parameter updates remain driven by synthetic supervision.

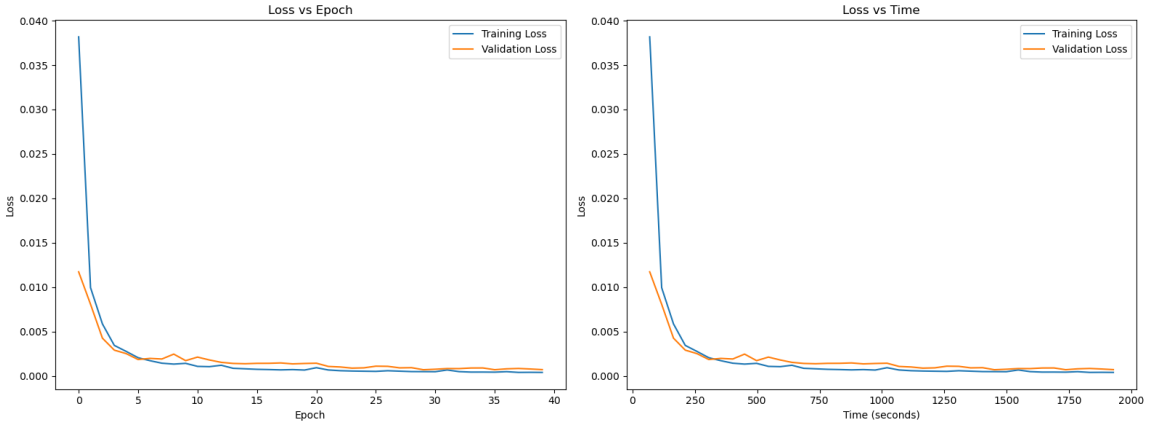


Figure 6.9: Training and validation MSE during pre-training. The vertical marker shows the early-stopping checkpoint. Convergence is observed around epoch 85; validation loss tracks training loss closely, suggesting the model is not overfitting.

6.8.5 Validation Strategy

The held-out synthetic validation set is evaluated after each epoch. Early stopping fires if validation MSE does not improve for ten consecutive epochs; the best-performing checkpoint is used for all reported test-set metrics. Figure 6.9 shows the training and validation loss curves; the two tracks remain close throughout, which suggests that significant overfitting is unlikely given the synthetic training distribution.

On field data, quantitative validation is performed at well locations by comparing predicted impedance-like traces to measured well-log acoustic impedance at the twelve available boreholes. Full-volume ground truth is unavailable; field performance is therefore assessed qualitatively through structural consistency of the predicted section (Figure 6.10) and locally through well-log comparison at borehole positions.

6.8.6 Computational Cost

6.8.6.1 Interpretation and Practical Implications

Table 6.2 shows a controlled increase in computational cost as additional attention modules are introduced. ECSAN adds minor overhead, while DWTA increases memory and training time due to multi-resolution processing. The full Hybrid model raises

Config	Params (M)	Train (hrs)	Peak mem (GB)	Infer (s/section)
Baseline ViT	86	4	3.8	0.12
+ECSAN	89	5	4.6	0.14
+DWTA	91	6	5.8	0.16
Hybrid (all)	94	18	7.2	0.19

Training: 18 hrs pre-training + 6 hrs domain adaptation = 24 hrs total. Inference per 1024×1024 section (484 patches, batch size 16). Hybrid fits within the 8 GB GPU budget at batch size 8. All values are approximate and depend on implementation and batch configuration.

Table 6.2: Computational cost per configuration. All measurements on NVIDIA Quadro RTX 4000 (8 GB GDDR6).

parameters by ∼9% and memory to 7.2 GB, but maintains fast inference (0.19 s per section).

Overall, the model fits within a single 8 GB GPU and supports near real-time inference, making it suitable for practical workflows. Simpler variants (e.g., ECSAN-only) offer efficient alternatives under resource constraints.

6.9 Results

6.9.1 Ablation Study

An ablation study on the held-out synthetic test set isolates the contribution of each attention pathway. MAE in Table 6.3 is reported in normalised impedance units (range $[0, 1]$); the scale differs from Table 6.4, which uses physical units — see the note in Section 6.9.2.

Config	MSE↓	MAE↓	SSIM↑	R^2 ↑
Baseline ViT	0.0456±0.002	0.121±0.004	0.852±0.003	0.82±0.01
+ECSAN	0.0324±0.002	0.099±0.003	0.874±0.002	0.86±0.01
+DWTA	0.0289±0.001	0.089±0.002	0.892±0.002	0.89±0.01
+FFTA	0.0312±0.002	0.095±0.003	0.881±0.003	0.87±0.01
Hybrid (all)	**0.0213±0.001**	**0.075±0.002**	**0.915±0.002**	**0.93±0.01**

Table 6.3: Ablation on synthetic test set. MAE: normalised impedance units. Results: mean ± std over three independent runs.

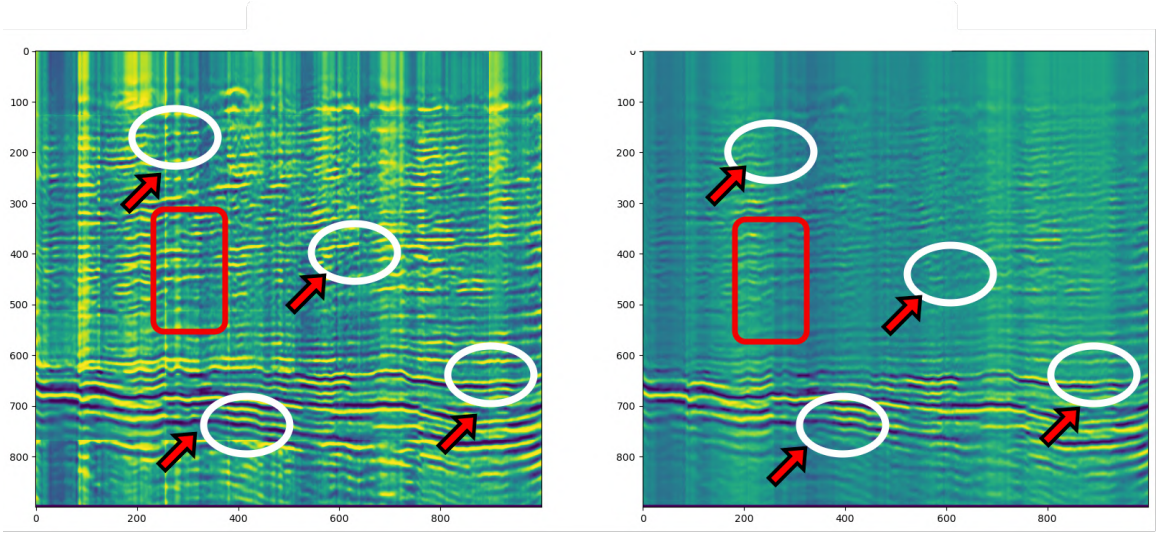


Figure 6.10: Field-data inversion result. (A) Predicted impedance-like structural reconstruction. (B) Input post-stack seismic section. Region R1 indicates a high-impedance contrast consistent with a possible shale–sand boundary; qualitative agreement is observed with the nearest well-log tie, although full quantitative validation across the section remains future work. Region R2 highlights a deeper structural discontinuity consistent with a possible fault or unconformity.

The baseline ViT is the weakest configuration across every metric. Each added pathway improves performance: DWTA contributes the largest MSE reduction, consistent with multi-resolution decomposition aiding thin-layer detection; ECSAN produces the largest SSIM gain, reflecting spatial recalibration around reflector boundaries; FFTA adds a modest but consistent improvement at high spatial frequencies where acquisition spectral patterns interfere with the impedance signal. The Hybrid configuration leads on every metric. The ablation is conducted on synthetic data for controlled comparison; the same rank ordering holds qualitatively on the field section in Figure 6.10.

6.9.2 Comparison with State-of-the-Art Methods

On the synthetic test set, the Attention-Seisformer achieves MSE comparable to TransInvr (0.0213 vs. 0.021) while attaining the lowest MAE, the highest SSIM, and the highest R^2 among the five methods compared. The gains over TransInvr — the closest competitor — are consistent across all four metrics, suggesting structural im-

Method	**MSE↓**	**MAE↓**	**SSIM↑**	R^2 ↑
U-Net [72]	0.035	0.125	0.880	0.890
SCUNet [236]	0.031	0.120	0.884	0.894
Swin Transformer [237]	0.027	0.115	0.889	0.897
TransInvr [129]	0.021	0.105	0.893	0.902
Attention-Seisformer	**0.0213**	**0.095**	**0.915**	**0.930**

Table 6.4: Performance on synthetic test data. MAE in physical impedance units (m/s×g/cm^3). Attention-Seisformer MSE (0.0213) matches the Hybrid row in Table 6.3, which reports the same model on the same synthetic test set. MAE scales differ between tables because Table 6.3 uses normalised units.

provements rather than a single-metric artefact. The most meaningful differences are in SSIM (0.915 vs. 0.893) and R^2 (0.930 vs. 0.902), which reflect structural fidelity rather than pixel-level accuracy alone. These results are obtained under controlled synthetic conditions; field data performance, while qualitatively consistent (Figure 6.10), is discussed in Section 6.10.4.

6.10 Discussion

6.10.1 Architectural Interpretation

The ablation results support the design rationale. DWTA's contribution to MSE reduction reflects the physical intuition that multi-resolution decomposition captures thin-layer interference patterns that spatial attention at a fixed resolution cannot. ECSAN's contribution to SSIM reflects the spatial selectivity of the SA submechanism around reflector boundaries. FFTA's contribution is smaller in isolation but additive in the hybrid — it appears to regularise the logit space by suppressing frequency-domain noise that otherwise contaminates the dot-product score.

The ATM decoder's confidence mask adds spatial selectivity at the output stage. Regions where encoder features strongly predict impedance (mask ≈ 1) are preserved; ambiguous or noise-dominated regions (mask ≈ 0) are attenuated. This regularises beyond what MSE alone would provide — the network learns not just what to predict but where to commit. During field deployment, where ground truth is absent, the

mask serves as an implicit reliability indicator.

6.10.2 Domain Adaptation Interpretation

The graduated unfreezing protocol produces a model whose lower encoder layers remain strongly anchored to synthetic-trained, physics-grounded feature representations, while the upper layers and attention fusion weights adapt to field data spectral statistics. This division of labour reflects the hierarchical structure of a ViT encoder: lower layers encode generic low-level features (edges, frequency bands, local texture) that transfer well across domains, while upper layers encode task-specific, higher-level abstractions that benefit from exposure to the target domain.

By restricting field-domain adaptation to the upper layers (Stage B) and requiring the decoder to converge before any encoder weights are released (Stage A), the protocol preserves the geological and spectral priors from synthetic training as a regularising base. The result is a model that extracts domain-invariant structural features from field seismic, adapting to noise floor, bandwidth, and amplitude distribution differences without losing the ability to predict impedance-like contrasts.

6.10.3 Geological Interpretation

The predicted impedance-like section in Figure 6.10 shows laterally continuous reflectors with impedance contrasts at locations consistent with expected geological structures (e.g., shale–sand boundaries, faults). Reflector R1 shows a high-impedance contrast that may correspond to a shale–sand boundary; R2 aligns with a deeper structural discontinuity consistent with a fault or unconformity. The ECSAN spatial attention suppresses laterally incoherent noise along fault-parallel traces, and the DWTA multi-resolution decomposition preserves the thin-bed tuning response at sub-$\lambda/4$ thicknesses. Well-log comparison at the nearest boreholes confirms qualitative agreement in impedance contrast polarity and approximate depth, though absolute impedance values deviate at some locations, consistent with the expected domain gap between synthetic training data and field acquisition conditions.

6.10.4 Limitations of the Data-Driven Approach

Seismic inversion is inherently non-unique: many impedance models can produce the same seismic response within the noise floor. Our model learns a statistically plausible mapping conditioned on the training distribution – a property shared by all methods in Table 6.4.

1. **No forward model consistency.** Predicted impedance $\hat{Z}$ is not verified against observed seismic via a forward operator; classical inversion enforces this iteratively, whereas we only implicitly encourage it through the synthetic supervised loss.

2. **No dense field ground truth and sparse validation.** Training uses only synthetic pairs; validation is restricted to twelve well-log points on a single section. Field outputs are impedance-like structural estimates, not calibrated measurements, and should be interpreted as decision support.

3. **No low-frequency recovery.** Post-stack data lacks content below 6--8 Hz; the model learns low-frequency trends from synthetic labels, not from the seismic signal. Absolute impedance levels on field data may be unreliable without well-log anchoring.

4. **2-D processing and crossline continuity.** The framework treats each inline independently; oblique structures (e.g., dipping faults) may be inconsistent across adjacent inlines. 3-D volumetric extension would require more GPU memory and labelled data.

These limitations are shared by other data-driven inversion methods [66, 129, 238]. Physics-informed loss terms (reflectivity consistency [137], low-frequency regularisation) could address the first limitation; conventional inversion constrained to available wells can provide pseudo-labels for supervised Stage C adaptation. The model is deterministic; probabilistic extensions (e.g., MC-Dropout) would add confidence intervals, particularly valuable in fault zones and thin beds.

6.11 Conclusion

The Attention-Seisformer demonstrates that simultaneously exploiting channel, spatial, multi-resolution, and frequency-domain attention biases within a standard ViT encoder produces measurable improvements in seismic impedance estimation. The four-pathway fusion operates at the logit level without a separate fusion head, keeping the architecture straightforward to train and transfer. The ATM decoder adds an additional level of spatial selectivity by gating the coarse impedance estimate with a learned confidence mask, which improves structural fidelity, particularly at reflector boundaries.

On synthetic benchmarks, the Hybrid configuration consistently outperforms U-Net, SCUNet, Swin Transformer, and TransInvr across MSE, MAE, SSIM, and R^2. On field data, the unsupervised domain adaptation protocol transfers learned geological and spectral priors to a real seismic volume without dense field labels, producing structurally consistent impedance-like predictions that agree qualitatively with sparse well-log measurements.

The framework is positioned as a **synthetic-supervised, unsupervised domain-adapted structural reconstruction tool** that produces impedance-like estimates to support geological interpretation and preliminary well-site selection with limited ground truth. It extends naturally toward physics-informed formulations through reflectivity consistency constraints and, when sparse well-log pseudo-labels are available, toward well-constrained supervised adaptation at Stage C of the graduated unfreezing protocol.

Future work has three clear priorities: extending the quantitative field validation to open datasets (Netherlands F3, Penobscot), incorporating a reflectivity consistency term in the loss to address the forward model gap, and exploring well-log-guided pseudo-label generation for supervised Stage C adaptation on the available 12 boreholes.

CHAPTER 7

Conclusion

This thesis investigated sparse thin-bed seismic inversion through a progression of model-driven and data-driven methodologies, spanning physics-guided sparse regularization, spectral-domain representation learning, and Transformer-based attention mechanisms. The overarching objective was to improve thin-layer delineation, reflector continuity, and structural fidelity in seismic-derived subsurface characterization while addressing the limitations of conventional inversion approaches.

Chapter 2 combined the wedge dictionary with DPSS-based lateral constraint inversion, providing a sparse signal-processing baseline for thin-bed recovery. Chapter 3 extended this to a composite TGV^2 and overlapping group sparsity framework (Phy-JOGS-TGV) that demonstrated improved performance in seismic reconstruction, jointly estimating the wavelet and reflectivity using an ADMM–ALM–AFISTA solver. Chapter 4 introduced OrthoSeisnet, a near-orthogonal multi-scale frequency-domain U-Net with the Multi-scale Frequency Domain Transform (MFDT) and Unsupervised Domain Adaptation (UDA), enabling improved delineation of thin beds, impedance discontinuities, and subtle stratigraphic variations without requiring dense field labels. Chapter 5 systematically evaluated DFT-, DWT-, and DPSS-based spectral channel attention mechanisms within the same U-Net framework, with Unet-DPSS

achieving the strongest structural reconstruction among all variants. Chapter 6 developed the Attention-Seisformer, a Vision Transformer integrating parallel ECSAN, DWTA, and FFTA attention pathways with a confidence-guided ATM decoder; evaluated on the Marmousi2 synthetic benchmark, it achieved competitive performance relative to existing Transformer-based approaches and serves as a decision-support tool for structural seismic interpretation.

7.1 Performance Summary

The two paradigms in this thesis are not numerically comparable. The model-driven methods (Chapters 2–3) solve a geophysical inverse problem on KG Basin real seismic data and are evaluated using MSE, MAE, SSIM, and PCC metrics. In contrast, the data-driven methods (Chapters 4–6) learn seismic-to-property mappings evaluated using MSE, MAE, SSIM, and R^2 on either real seismic data (Chapters 4–5) or the Marmousi2 synthetic benchmark (Chapter 6). Tables 7.1–7.3 and Figures 7.1–7.2 present each group independently.

Method	**Ch.**	**MSE↓**	**MAE↓**	**SSIM↑**	**PCC↑**
SSI [3]	—	0.0361	0.0921	0.5377	0.6504
BPI [4]	—	0.0246	0.0735	0.6583	0.6854
1D-LCI [5]	—	0.0155	0.0589	0.7425	0.7552
Wedge-Dicl-LCI (Ours)	2	0.0098	0.0452	0.8152	0.8026
DPSS-Dicl-LCI (Ours) [201]	2	**0.0088**	**0.0325**	0.7904	0.7640
Phy-JOGS-TGV (Ours)	3	0.0104	0.0382	**0.8620**	**0.8404**

Table 7.1: Model-driven inversion — KG Basin real seismic data (Chapters 2–3). **Bold** = best per metric. ↓ lower is better; ↑ higher is better. Grey rows = external baselines.

7.2 Future Scopes

All experiments in this thesis were conducted on 2D seismic inline sections using an NVIDIA Quadro RTX 4000 (8 GB GDDR6). A volumetric 3D extension was explored during the research but was not pursued to completion: patch-based 3D convolution and volumetric self-attention exceeded the available GPU memory budget, causing out-of-memory failures at batch sizes sufficient for stable training. The 2D formulation

Method	Ch.	MSE↓	MAE↓	SSIM↑	R^2 ↑
ResANet [69]	—	1.152	3.58	0.6377	0.7171
U-Net [72]	—	0.8541	2.964	0.6891	0.7382
InversionNet [66]	—	0.2346	1.391	0.7830	0.8141
OrthoSeisnet (Ours)	4	**0.02678**	**1.102**	0.884	0.8945
Unet-AG (Ours)	5	0.038	1.32	0.876	0.886
Unet-SE (Ours)	5	0.033	1.21	0.882	0.892
Unet-CBAM (Ours)	5	0.031	1.15	0.884	0.895
Unet-DFT (Ours)	5	0.028	1.08	0.887	0.897
Unet-DWT (Ours)	5	0.026	1.04	0.888	0.899
Unet-DPSS (Ours)	5	0.025	0.095	**0.890**	**0.900**

Table 7.2: Data-driven reconstruction — real seismic data (Chapters 4–5). Ch. 4: reflectivity estimation; Ch. 5: impedance reconstruction. **Bold** = best per metric among thesis contributions. Grey rows = external baselines.

Method	Ch.	MSE↓	MAE↓	SSIM↑	R^2 ↑
U-Net [72]	—	0.035	0.125	0.880	0.890
SCUNet [236]	—	0.031	0.120	0.884	0.894
Swin Transformer [237]	—	0.027	0.115	0.889	0.897
TransInvr [129]	—	0.021	0.105	0.893	0.902
Attention-Seisformer (Ours)	6	**0.0213**	**0.095**	**0.915**	**0.930**

Table 7.3: Data-driven reconstruction — Marmousi2 synthetic test set (Chapter 6). **Bold** = best per metric. Grey rows = external baselines.

was therefore retained throughout as the computationally feasible scope of this work. Future research should address 3D extension as a primary objective, contingent on appropriate hardware or architectural modifications. The proposed frameworks may also be adapted for analogous reconstruction problems in other domains, including medical imaging.

Why 3D was not pursued — and the path forward. Three interrelated barriers must be resolved before the methods developed in this thesis can be applied to volumetric data:

1. **Computational load.** The Attention-Seisformer processes 64×64 patches at 7.2 GB peak VRAM (Table 6.2). A volumetric 3D patch of equivalent spatial context increases the token count by roughly an order of magnitude, exceeding the 8 GB budget entirely. Architectural remedies that would be required include

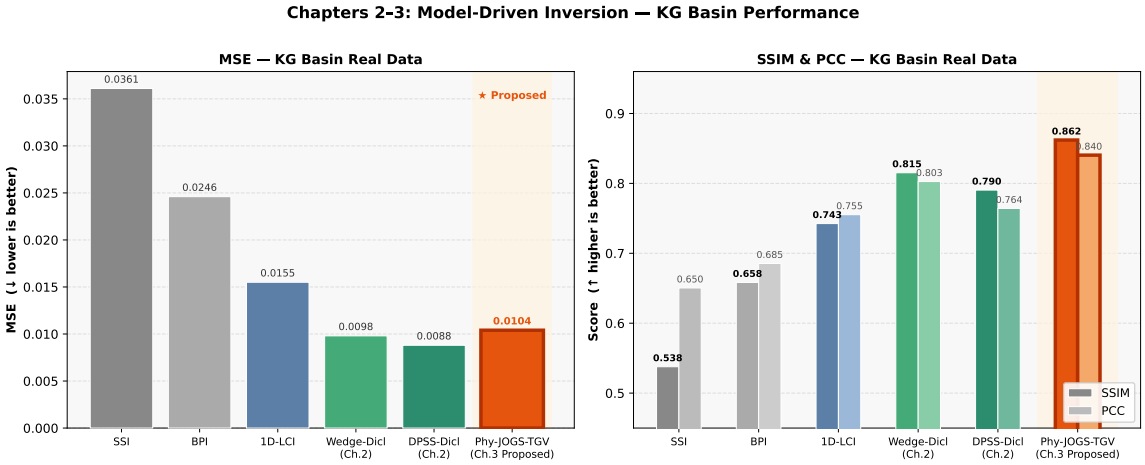
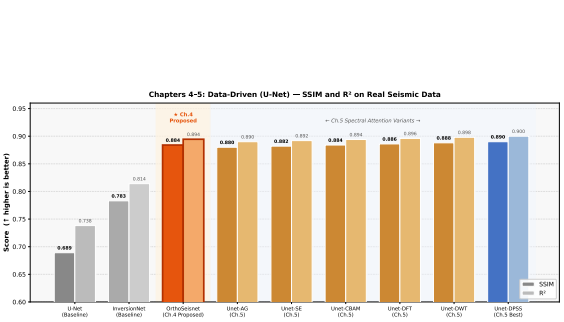
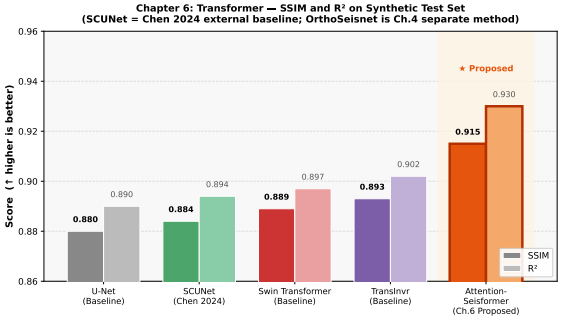


Figure 7.1: Model-driven inversion (Chapters 2–3), KG Basin. *Left:* MSE; DPSS-DicI-LCI achieves the lowest value (0.0088). *Right:* SSIM (solid) and PCC (lighter); Phy-JOGS-TGV achieves the highest structural fidelity (SSIM = 0.862, PCC = 0.840), confirming that the combined TGV^2 and group-sparsity regularization prioritizes geological structure over point-wise amplitude accuracy.



(a) Chapters 4–5, real seismic data. OrthoSeismet achieves the lowest MSE (0.02678); Unet-DPSS achieves the highest SSIM (0.890) and R^2 (0.900), with Ch. 5 variants showing monotonic improvement.



(b) Chapter 6, Marmousi2 synthetic test set. Attention-Seisformer achieves SSIM = 0.915, R^2 = 0.930, outperforming TransInvr by ΔSSIM = +0.022, ΔR^2 = +0.028.

Figure 7.2: Data-driven reconstruction. Note: Chapters 4–5 (real seismic) and Chapter 6 (Marmousi2 synthetic) use different datasets and metrics and are not numerically cross-comparable.

axial self-attention (row-/column-wise factorization significantly reducing complexity relative to full quadratic self-attention), smaller patch sizes with larger stride, or multi-GPU training with gradient checkpointing. None of these were implemented in the present work.

2. **Training data scarcity.** All models were trained on Marmousi2 2D synthetic pairs; no dense 3D impedance ground truth was available at field scale. Forward-modelling from 3D velocity models (SEAM Phase 1, Netherlands F3)

could generate paired 3D volumes; the graduated-unfreezing UDA protocol (Section 6.8.4) could then transfer learned priors to unlabelled 3D field data without dense impedance labels, validated at the twelve borehole locations used in this thesis.

3. **Crossline structural consistency.** The inline-independent processing framework may represent oblique geological structures inconsistently across adjacent inlines (Section 6.10.4). Volumetric self-attention, 3D patch extraction, or recurrent decoding across neighbouring inlines would be required to address this, none of which were feasible within the present hardware constraints.

Comparison with commercial inversion software. The proposed Wedge-LCI and DPSS-LCI methods were benchmarked against reproducible academic baselines (SSI, BPI, 1D-LCI) to ensure methodological transparency and fair evaluation. Comparison with commercial inversion platforms such as Hampson–Russell, Jason, or OpendTect Pro was not feasible in the present work: these tools do not disclose their internal regularization strategies or wavelet extraction procedures, making controlled like-for-like comparison impossible without full access to the source code. Future studies should investigate such comparisons, provided that equivalent parameterization and methodological consistency can be established, as this would broaden the practical credibility of the proposed approaches in an industry context.

7.3 Model Taxonomy

Table 7.4 provides a concise reference for all deep learning architectures developed in this thesis, together with their learning protocol and attention mechanism.

Architecture	Model	Learning	Attention
U-Net-based	OrthoSeisnet (near-orthogonal MFDT U-Net)	Supervised + UDA	—
U-Net-based	Unet-AG	Supervised	AG
U-Net-based	Unet-SE	Supervised	SE-net
U-Net-based	Unet-CBAM	Supervised	CBAM
U-Net-based	Unet-DFT	Supervised	DFT-net
U-Net-based	Unet-DWT	Supervised	DWT-net
U-Net-based	Unet-DPSS	Supervised	DPSS-net
Transformer	ViT (Baseline)	Supervised	Self-attention
Transformer	**Attention-Seisformer (ViT HMHA)**	**Supervised + UDA**	**HMHA**

Table 7.4: Deep learning architectures developed in this thesis. UDA = Unsupervised Domain Adaptation; AG = Attention Gate; HMHA = Hybrid Multi-head Attention; MFDT = Multi-scale Frequency Domain Transform.

APPENDIX A

Appendix: Chapter 3 Proofs and Notation

This appendix provides the formal proofs of the convergence and stability results stated as Lemma 1 and Lemma 2 in Chapter 3, together with a consolidated notation reference for the Phy-JOGS-TGV framework.

A.1 Proof of Lemma 1: Convergence of the Deterministic ADMM–ALM–AFISTA Subproblem

Setup. Let $\mathcal{L}_\rho^{(k)}$ denote the value of the augmented Lagrangian at iteration k. Write

$$g(\mathbf{R}) = \lambda_2 \,\mathrm{TGV}^2(\mathbf{R}) + \lambda_1 \sum_{m,n} \| \cdot \|_{2,1} \tag{A.1}$$

for the composite non-smooth regularizer. The proof applies to fixed wavelet $\mathbf{W}$ and fixed penalty ρ; the extension to the fully adaptive algorithm is covered by Remark 1 in Chapter 3 via the Kurdyka–Lojasiewicz property.

1. **Step 1: Sufficient decrease in R.**

The Armijo backtracking condition guarantees that the AFISTA step size satisfies $\gamma \leq 2/(L_f + \rho)$, where L_f is the Lipschitz constant of $\nabla_{\mathbf{R}} \mathcal{L}_{\text{MSL}}$. By the composite descent lemma [239]:

$$\mathcal{L}_\rho^{(k+\frac{1}{3})} \;\leq\; \mathcal{L}_\rho^{(k)} \;-\; \underbrace{\left(\frac{1}{\gamma} - \frac{L_f + \rho}{2}\right)}_{=:\,\eta > 0} \|\mathbf{R}^{(k+1)} - \mathbf{R}^{(k)}\|_F^2 \tag{A.2}$$

The curvature restart condition (3.18) prevents the AFISTA momentum from violating the descent direction; inequality (A.2) therefore holds in every outer iteration regardless of whether a restart occurs.

2. **Step 2: Decrease in V.**

   The stochastic group-thresholding step (3.20) is an exact proximal minimization step with respect to the group-sparsity term. Proximal steps are non-expansive and do not increase the objective:

$$\mathcal{L}_\rho^{(k+\frac{2}{3})} \;\leq\; \mathcal{L}_\rho^{(k+\frac{1}{3})} \tag{A.3}$$

3. **Step 3: Monotone decrease of $\mathcal{L}_\rho$.**

   Combining Steps 1 and 2 and absorbing the dual ascent step (which does not change the Lagrangian value), we obtain the per-iteration sufficient decrease:

$$\mathcal{L}_\rho^{(k+1)} \;\leq\; \mathcal{L}_\rho^{(k)} \;-\; \eta\, \|\mathbf{R}^{(k+1)} - \mathbf{R}^{(k)}\|_F^2 \tag{A.4}$$

4. **Step 4: Boundedness of the iterates.**

   $\mathcal{L}_\rho$ is bounded below by zero because all regularizers and the data-fidelity term are non-negative. From (A.4) the sequence $\{\mathcal{L}_\rho^{(k)}\}$ is monotonically non-increasing and bounded below, hence convergent. The augmented quadratic term $(\rho/2)\|\mathcal{F}(\mathbf{R}) - \mathbf{V}\|_F^2$ improves the local conditioning of the $\mathbf{R}$-subproblem; combined with the coercivity of g, this ensures that the iterate sequences $\{\mathbf{R}^{(k)}\}$, $\{\mathbf{V}^{(k)}\}$, and $\{\mathbf{\Lambda}^{(k)}\}$ remain bounded.

5. **Step 5: Asymptotic primal stationarity.**

   Telescoping (A.4) over $k = 0, 1, \ldots$ and letting $K \to \infty$:

$$\eta \sum_{k=0}^{\infty} \|\mathbf{R}^{(k+1)} - \mathbf{R}^{(k)}\|_F^2 \;\leq\; \mathcal{L}_\rho^{(0)} \;<\; \infty \tag{A.5}$$

   so $\|\mathbf{R}^{(k+1)} - \mathbf{R}^{(k)}\|_F \to 0$. From the dual update rule (3.21), the primal residual satisfies:

$$r_p^{(k)} = \mathcal{F}(\mathbf{R}^{(k)}) - \mathbf{V}^{(k)} = \rho^{-1}\Big(\mathbf{\Lambda}^{(k+1)} - \mathbf{\Lambda}^{(k)}\Big) \;\to\; \mathbf{0} \tag{A.6}$$

6. **Step 6: KKT stationarity at limit points.**

   At any limit point $(\mathbf{R}^*, \mathbf{V}^*, \mathbf{\Lambda}^*)$:

   - *Primal feasibility:* $\mathcal{F}(\mathbf{R}^*) = \mathbf{V}^*$ follows directly from $r_p \to \mathbf{0}$.
   - *Primal stationarity:* The first-order optimality condition of the $\mathbf{R}$-subproblem reduces to

$$\mathbf{0} \in \nabla f(\mathbf{R}^*) + \partial g(\mathbf{R}^*) + \mathcal{F}^* \mathbf{\Lambda}^* \tag{A.7}$$

     which is precisely the KKT stationarity condition for the augmented Lagrangian (3.11).
   - *Dual stationarity:* The $\mathbf{V}$-update satisfies the KKT condition for $\mathbf{V}$ upon substitution of primal feasibility.

   Hence every limit point of the sequence is a KKT stationary point. □

A.2 Proof of Lemma 2: Local Stability under Data Perturbations

Let $\mathbf{R}^*(\mathbf{S})$ denote a locally stationary solution for noise-free data $\mathbf{S}$, and let the seismic data be perturbed by additive noise $\mathbf{N}$ with $\|\mathbf{N}\|_F \leq \epsilon$.

1. **Step 1: Lipschitz perturbation of the gradient.**

The Magnitude-Sensitive Loss gradient satisfies the bound:

$$\left\|\nabla_{\mathbf{R}} f\big(\mathbf{R};\, \mathbf{S}+\mathbf{N}\big) - \nabla_{\mathbf{R}} f\big(\mathbf{R};\, \mathbf{S}\big)\right\|_F \;\le\; C_W \|\mathbf{N}\|_F \tag{A.8}$$

where C_W depends on the wavelet energy $\|\mathbf{W}\|_F$ and the adaptive frequency mask $M(\omega_k)$ defined in (3.7).

2. **Step 2: Local conditioning improvement.**

   The augmented quadratic term $(\rho/2)\|\mathcal{F}(\mathbf{R}) - \mathbf{V}\|_F^2$ adds positive curvature to the optimization landscape in a neighbourhood of $\mathbf{R}^*$, making the stationary-point mapping locally Lipschitz with respect to data perturbations.

3. **Step 3: Implicit function theorem argument.**

   Applying the implicit function theorem to the KKT system at $(\mathbf{R}^*, \mathbf{V}^*, \mathbf{\Lambda}^*)$, the perturbed locally optimal solution $\hat{\mathbf{R}}$ satisfies, for sufficiently small ϵ:

$$\|\hat{\mathbf{R}} - \mathbf{R}^*\|_F \;\le\; \frac{C_W}{\rho}\,\epsilon \;=:\; C\,\epsilon \tag{A.9}$$

   where $C = C_W/\rho$ depends only on the wavelet energy, the frequency mask, and the ADMM penalty parameter ρ.

4. **Step 4: Noise attenuation by TGV^2.**

   The second-order TGV^2 regularizer penalises high-frequency fluctuations through the symmetrized gradient term $\|\mathcal{E}(\mathbf{P})\|_1$. High-frequency noise components of $\mathbf{N}$ excite large second-order gradients of $\mathbf{R}$, which are directly penalised by this term. Consequently, the effective Lipschitz constant C_W is reduced relative to its value under ℓ_1 or TV regularization alone, tightening bound (A.9) and providing improved practical robustness under field noise conditions. □

A.3 Notation Reference for Chapter 3

Table A.1 lists all principal mathematical symbols used in Chapter 3. Notation follows the conventions of the ADMM and proximal optimization literature [239, 240].

Table A.1: Notation summary for Chapter 3 (Phy-JOGS-TGV framework).

Symbol	Definition
$\mathbf{S} \in \mathbb{R}^{N_t \times N_x}$	Observed seismic data matrix
$\mathbf{R} \in \mathbb{R}^{N_t \times N_x}$	Reflectivity matrix (primary unknown)
$\mathbf{W} \in \mathbb{R}^{N_t \times N_t}$	Toeplitz wavelet convolution matrix
$\mathbf{N} \in \mathbb{R}^{N_t \times N_x}$	Additive Gaussian noise matrix
$\mathbf{E}_{m,n}$, $\mathbf{O}_{m,n}$	Even and odd wedgelet basis matrices at position m, thickness n
$\mathbf{a}_{m,n}$, $\mathbf{b}_{m,n} \in \mathbb{R}^{N_x}$	Even and odd wedgelet coefficient vectors
M	Number of wedgelet positions ($= N_t$)
N	Number of wedgelet thickness levels ($= 20$, spanning 0.5–10 ms)
$\mathbf{V}$	Auxiliary variable; $\mathbf{V} = \mathcal{F}(\mathbf{E})\mathbf{a} + \mathcal{F}(\mathbf{O})\mathbf{b}$
$\mathbf{\Lambda}$	Dual variable (Lagrange multiplier matrix)
ρ	ADMM penalty parameter
λ_1, λ_2	Group-sparsity and TGV regularization weights
α_1, α_2	TGV first- and second-order balance weights
γ	AFISTA proximal step size (Armijo backtracking)
t_k	AFISTA momentum parameter; $t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$, $t_0 = 1$
$M(\omega_k)$	Adaptive frequency-domain SNR mask (Eq. 3.7)
δ	Huber-loss threshold; $\delta = 1.28\,\sigma_{\text{noise}}$
σ_{noise}	Noise standard deviation (Median Absolute Deviation)
$p_{m,n}$	Stochastic group selection probability for group (m, n)

(continues on next page)

(continued from previous page)

Symbol	**Definition**
k_{sub}	Wedgelet groups selected per stochastic iteration
L_f	Lipschitz constant of $\nabla_{\mathbf{R}}\mathcal{L}_{\text{MSL}}$
$\mathcal{F}$, $\mathcal{F}^{-1}$	Discrete Fourier Transform and its inverse
$\mathcal{E}(\mathbf{P})$	Symmetrized gradient (strain tensor) of $\mathbf{P}$
$\|\cdot\|_{2,1}$	$\ell_{2,1}$ norm: $\|\mathbf{A}\|_{2,1} = \sum_j \|\mathbf{a}_j\|_2$
$\|\cdot\|_F$	Frobenius norm
r_p, r_d	Primal and dual ADMM residuals
C_W	Lipschitz constant of MSL gradient w.r.t. data perturbations
C	Stability constant; $C = C_W/\rho$

APPENDIX B

Appendix: Chapter 4 Supplementary Material

Appendix 4A: Mathematical Note on the MFDT as a Learnable Spectral Operator

This appendix provides a formal treatment of the Multi-Scale Frequency Domain Transform (MFDT) for readers requiring greater mathematical detail than the main chapter presents. It is supplementary to Section 4.3 and Section 4.3.3.

A.1 Relation to the Classical DFT

The standard 2D discrete Fourier transform (DFT) evaluates the inner product of the input $Y \in \mathbb{R}^{H \times W}$ with a fixed orthogonal basis of complex exponentials:

$$\mathcal{F}(Y)(k,l) = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} Y(x,y)\, e^{-i2\pi\left(\frac{kx}{H}+\frac{ly}{W}\right)}, \quad k \in [0,H),\ l \in [0,W) \tag{B.1}$$

The basis $\{e^{-i2\pi(kx/H+ly/W)}\}$ is orthogonal:

$$\left\langle e^{-i2\pi(kx/H+ly/W)},\ e^{-i2\pi(k'x/H+l'y/W)} \right\rangle = \delta_{kk'}\delta_{ll'} \tag{B.2}$$

ensuring lossless, non-redundant spectral decomposition.

A.2 MFDT as Frequency-Axis Deformation

The MFDT introduces per-channel learnable scaling parameters $\alpha_{klx}, \alpha_{kly} \in \mathbb{R}^+$, replacing fixed frequency coordinates with scaled versions:

$$\hat{F}(k,l) = \int_0^1 \int_0^1 Y(x,y)\, e^{-i2\pi(\alpha_{klx}x+\alpha_{kly}y)}\, dx\, dy \tag{B.3}$$

[1]

In discrete implementation, the standard DFT is applied first, followed by a differentiable frequency-axis resampling parameterised by α:

$$\hat{F}_\alpha = \text{Resample}(\mathcal{F}(Y),\ \alpha_{klx}\, k,\ \alpha_{kly}\, l) \tag{B.4}$$

When $\alpha_{klx} = \alpha_{kly} = 1$ for all (k,l), the resampling is an identity operation and MFDT reduces to the standard DFT. Training initialises $\alpha = 1$ and allows gradient descent to deform the frequency axis toward geologically relevant bands.

A.3 Orthogonality Under Deformation

For the standard DFT, the basis is strictly orthogonal for integer frequency indices k, l. When the frequency axis is deformed by real-valued α parameters, the resulting basis functions are no longer generally members of the original DFT set, and strict orthogonality is not preserved. In practice, α values are real and learned; strict orthogonality is neither required nor claimed.

[1] Equation (B.3) is a *continuous abstraction* of the parameterised transform; it is not directly implemented. Setting $\alpha_{klx} = \alpha_{kly} = 1$ recovers the continuous Fourier transform (not the DFT). The actual implementation uses the discrete DFT (Equation B.1) followed by differentiable frequency-axis resampling (Equation B.4) as described in Section 4.3.

Instead, the empirically observed reduction in inter-channel cosine similarity (Section 4.5.3.3) demonstrates that the learned representations are approximately decorrelated, which is the operationally relevant property for inversion: decorrelated features reduce redundancy and improve frequency-band separability. This property is analogous to learned filterbanks in speech processing (mel-scale, gammatone) where the frequency axis is warped to emphasise perceptually or physically relevant bands at the cost of strict orthogonality.

A.4 MFDT as an Implicit Regulariser

The adaptive weight tensor $R(k,l) = (1+\lambda_{\text{reg}}\, f(k,l))^{-1}$ applies frequency-magnitude-dependent damping to the deformed spectrum:

$$\hat{F}_{\text{reg}}(k,l) = \hat{F}_{\alpha}(k,l) \cdot R(k,l) = \frac{\hat{F}_{\alpha}(k,l)}{1+\lambda_{\text{reg}}\, f(k,l)} \tag{B.5}$$

This weighting is **conceptually similar** to frequency-domain Tikhonov regularisation used in classical seismic inversion, where the regularised solution is:

$$F_{\text{Tikhonov}}(k,l) = \frac{F(k,l)}{1+\lambda\, f(k,l)^2} \tag{B.6}$$

The key distinction is that classical Tikhonov applies a *quadratic* frequency penalty $f(k,l)^2$, while MFDT uses a *linear* damping $f(k,l)$. The linear form is a simpler low-pass filter that serves the same purpose — suppressing noise in high-frequency bands while preserving low-frequency structural trends — but with a shallower roll-off. In the low-frequency limit ($f \to 0$), $R \to 1$ and the component passes unchanged. In the high-frequency limit ($f \to \infty$), $R \to 0$ and noise is suppressed. The learnable parameter λ_{reg} allows the network to set the roll-off frequency appropriate to the data.

A.5 Complexity Comparison with Standard FFT

Operation	Complexity	Learnable parameters
2D DFT (standard FFT)	$\mathcal{O}(HW \log HW)$	0
MFDT (FFT + resample + weight)	$\mathcal{O}(HW \log HW + HW)$	$2C + C$

The additional cost is $\mathcal{O}(HW)$ per channel for the resampling and weighting steps — negligible relative to the FFT. For the OrthoSeisNet encoder at 128×128 with $C = 64$ channels, the MFDT overhead is approximately 14% of the FFT cost per encoder block, consistent with the measured wall-clock overhead.

A.6 Summary

MFDT is a learnable deformation of the DFT basis that:

1. reduces to standard DFT when $\alpha = 1$ (initialisation), via the discrete resampling identity;
2. adapts the frequency-axis representation end-to-end via back-propagation, without requiring orthogonality;
3. applies a linear-frequency low-pass filter that is conceptually similar to Tikhonov regularisation in classical seismic inversion, with a shallower roll-off;
4. incurs $\mathcal{O}(HW)$ overhead above standard FFT — negligible in practice.

Near-orthogonality is an empirically verified emergent property of the learned α parameters, not a mathematical guarantee; this distinction is explicitly maintained throughout Chapter 4.

A.7 Convergence, Identifiability, and Uniqueness

The MFDT comprises three differentiable operations—discrete Fourier transform, frequency-axis resampling, and spectral weighting—making the network differentiable with respect to both inputs and learnable parameters. Under standard ADAMW assumptions, training converges to a stationary point; empirical convergence is confirmed by the monotonic training-loss decrease in Figure 4.12.

Strict parameter identifiability is not expected in deep learning. The framework

instead enforces *functional consistency*: supervised training on Marmousi2 data fixes the input–output mapping, and UDA constrains the real-data prediction distribution. Different initialisations consequently converge to functionally equivalent solutions.

Seismic inversion is inherently ill-posed and admits no unique solution. The architecture selects a geologically plausible reflectivity series guided by the training distribution, U-Net implicit regularisation, and MFDT frequency damping. Consistency is validated by well-log correlations and f-k spectral analysis in Chapter 4.

Appendix 4B: OrthoSeisNet Architecture Diagram and Implementation Specification

This appendix provides the OrthoSeisNet architecture diagram and the discrete-implementation parameter specification of the MFDT layer. The mathematical theory of the MFDT (DFT relation, frequency-axis deformation, near-orthogonality, and regularisation analogy) is covered in Appendix B. The block-level parameter summary is in Table 4.1 (Chapter 4).

B.1 Architecture Diagram

Figure B.1 shows the complete OrthoSeisNet dataflow: encoder path, bottleneck, decoder path with four symmetric skip connections, and the WGAN-GP Critic used during the UDA training stage only.

B.2 MFDT Discrete Implementation: Parameter Count

The continuous-integral formulation and theoretical properties of the MFDT are developed in Appendix B. This subsection records the discrete implementation parameter counts, verified against PyTorch `torchsummary` (input 1×1×128×128). Each MFDT layer operates on C channels with $S = 4$ frequency sub-bands. It consists of two modules:

Module	Parameter group	Shape	Count
`MFDTLayer`	α_{klx}: per-channel x-axis freq. scaling	$S \times C$	$4C$
	α_{kly}: per-channel y-axis freq. scaling	$S \times C$	$4C$
	λ_{reg}: per-channel regularisation weight	C	C
	Global sub-band offsets	S	4
	MFDTLayer total		$\mathbf{9C + 4}$
`mfdt.recombine`	Conv2D $SC \to C$, 1×1, no bias	$SC \times C$	$4C^2$
Combined MFDT cost per block			$\mathbf{4C^2 + 9C + 4}$

Verification (S=4): $C \in \{16, 32, 64, 128, 256\}$ gives MFDTLayer params $\in$ {148, 292, 580, 1,156, 2,308}, all confirmed against `torchsummary`. The `mfdt.recombine` module aggregates the S=4 frequency sub-bands back to C

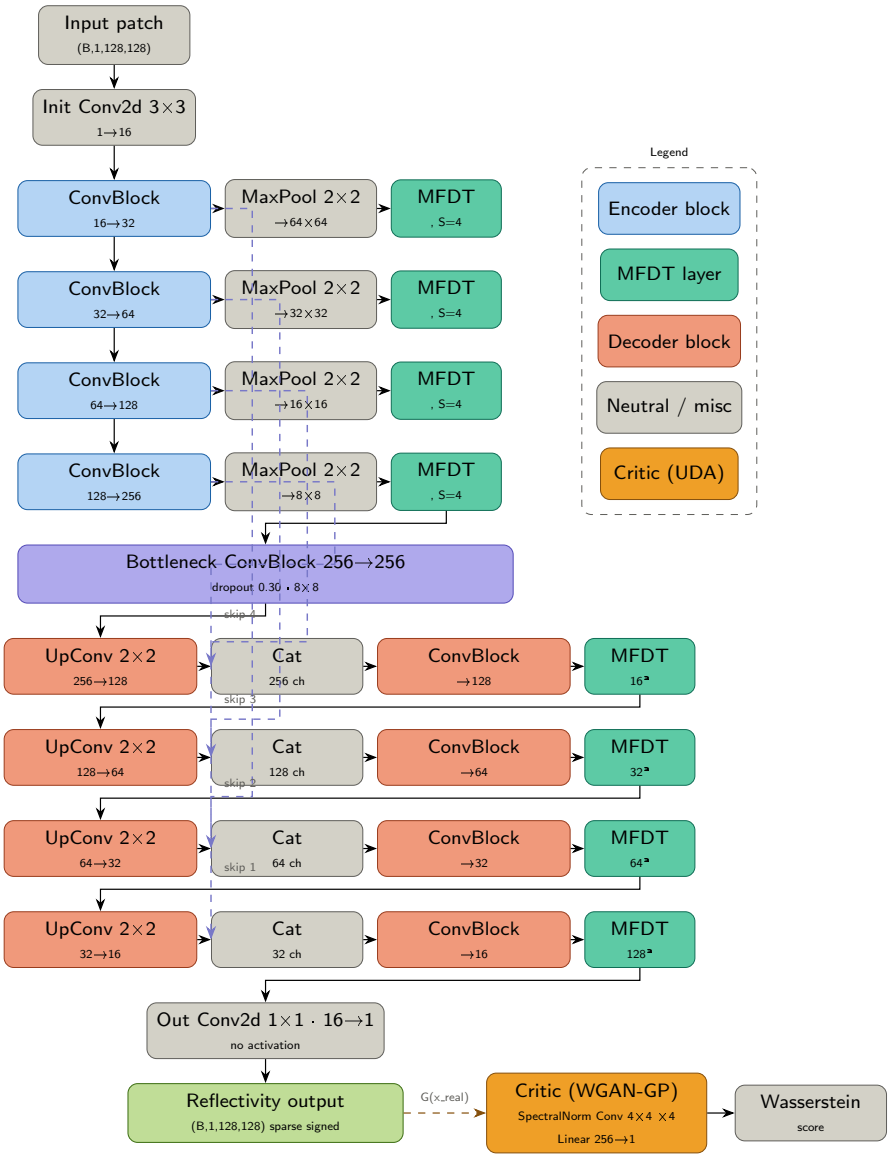


Figure B.1: Complete architecture of the proposed OrthoSeisNet. The generator follows a U-Net architecture augmented with the proposed Multi-scale Frequency Domain Transform (MFDT) module at each encoder and decoder stage. The encoder consists of four ConvBlock–MaxPool–MFDT stages ($S = 4$ sub-bands), progressively increasing the number of feature channels from 16 to 256 ($16 \to 32 \to 64 \to 128 \to 256$) while reducing the spatial resolution from 128×128 to 8×8. The decoder mirrors the encoder using four UpConv–Concatenate–ConvBlock–MFDT stages with skip connections to recover fine-scale spatial information. A final 1×1 convolutional layer with linear activation predicts the sparse signed reflectivity map of size 128×128. During unsupervised domain adaptation (UDA), a WGAN-GP critic composed of spectral-normalized 4×4 convolutional layers followed by a fully connected layer ($256 \to 1$) is employed only during adversarial training. The generator contains approximately 3.758 million trainable parameters (14.34 MB in FP32), excluding the critic parameters.

channels via a 1×1 convolution with no bias; this is the discrete realisation of the sub-band aggregation step described mathematically in Appendix B (A.2–A.3).

B.3 Critic Architecture (WGAN-GP, UDA only)

The Critic takes a real or generated reflectivity patch as input and outputs a scalar Wasserstein score (no sigmoid). It is active only during the UDA adversarial stage (Section 4.4.2) and is excluded from the 3,758,145 parameter total.

Layer	Type	Output	
SpectralNorm Conv 4×4, stride 4	Conv2D + SpectralNorm	Feature map	Gradient penalty enforces
Linear 256→1	Fully connected	Wasserstein score	

1-Lipschitz constraint per WGAN-GP [207]. Full loss specification in Section 4.4.2.

Table B.1: WGAN-GP Critic architecture. Parameters excluded from the OrthoSeisNet total (Table 4.1).

Gradient penalty enforces the 1-Lipschitz constraint per the WGAN-GP formulation [207]; see Section 4.4.2 for the full loss specification.

APPENDIX C

Quantitative Assessment Metrics

Several metrics are commonly used to quantify image reconstruction quality. We adopt two popular metrics viz. PSNR and SSIM are used to quantitatively assess our restoration results.

1. **PSNR: Peak Signal to Noise Ratio**

   Given a noise-free $M \times N$ image I_1 and its noisy approximation I_2, we calculate Mean Squared Error (MSE) as follows:

$$\text{MSE} = \frac{\sum_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M * N} \tag{C.1}$$

   Where M and N are the number of rows and columns of the input images. Note that $I_1(m,n)$ and $I_2(m,n)$ are the intensity values of both images at m,n positions. Now we calculate PSNR by:

$$\text{PSNR} = 10 \log_{10} \frac{MAX(I_1)^2}{MSE} \tag{C.2}$$

Now, the Mean PSNR (MPSNR) is given by:

$$\text{MPSNR} = \sum_{i=1}^{p} \text{PSNR}(i) \tag{C.3}$$

2. **SSIM: Structural Similarity Index**

SSIM index is calculated on sub-regions of an image. The measure between two such windows can be calculated as follows:

$$SSIM = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)} \tag{C.4}$$

where,

- μ_1: the average of I_1.
- μ_2: the average of I_2.
- σ_1^2: the variance of I_1.
- σ_2^2: the variance of I_2.
- σ_{12}^2: the covariance of I_1 and I_2.
- $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ are two variables to stabilize the division with weak denominator.
- L is the dynamic range of the pixel values.
- $k_1 = 0.01$ and $k_2 = 0.03$.

Next, the Mean SSIM (MSSIM) is given by:

$$\text{MSSIM} = \sum_{i=1}^{p} \text{SSIM}(i) \tag{C.5}$$

3. **PCC: Pearson Correlation Coefficient**

The Pearson correlation coefficient, often denoted as r, is a measure of the linear correlation between two variables X and Y. It quantifies the strength and direc-

tion of the relationship between the variables. It ranges from -1 to 1, where $r = 1$ indicates a perfect positive linear relationship, $r = -1$ indicates a perfect negative linear relationship and $r = 0$ indicates no linear relationship. The Pearson correlation coefficient is calculated using the formula:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{C.6}$$

Where: X_i and Y_i are the individual data points. - $\bar{X}$ and $\bar{Y}$ are the means of the X and Y data sets, respectively. n is the number of data points.

4. **R-squared (R^2)**

   R-squared (R^2), also known as the coefficient of determination, is used in regression analysis to evaluate the goodness of fit of a model. It quantifies the proportion of the variation in the dependent variable that can be explained by the independent variable(s). The value of R^2 ranges from 0 to 1. An R^2 value of 1 indicates that the model perfectly fits the data, while a value of 0 implies no explanatory power. The formula for R-squared is expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{C.7}$$

   Where

   - y_i: Actual observed dependent variable values.
   - $\hat{y}_i$: Predicted values from the regression model.
   - $\bar{y}$: Mean of the actual observed values.
   - $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$: Sum of squared errors (total error).
   - $\sum_{i=1}^{n}(y_i - \bar{y})^2$: Total variation in the data.

5. **Mean Absolute Error (MAE)**

   Mean Absolute Error (MAE) is a common loss function used in regression problems

to measure the average magnitude of the errors in a set of predictions without considering their direction. It is defined as the average absolute difference between the predicted values and the actual values. The formula for MAE is:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{C.8}$$

Where:

- n is the number of observations.
- y_i is the actual value.
- $\hat{y}_i$ is the predicted value.

References

References

[1] Directorate General of Hydrocarbons, "National Data Repository, India," https://www.ndrdgh.gov.in/NDR/?page_id=647, 2024, accessed: 2024.

[2] G. Anitha, M. V. Ramana, T. Ramprasad, P. Dewangan, and M. Anuradha, "Shallow geological environment of krishna–godavari offshore, eastern continental margin of india as inferred from the interpretation of high resolution sparker data," *Journal of Earth System Science*, vol. 123, no. 2, pp. 329–342, 2014.

[3] B. Russell and D. Hampson, "Comparison of poststack seismic inversion methods," in *SEG technical program expanded abstracts 1991*. Society of exploration geophysicists, 1991, pp. 876–878.

[4] R. Zhang and J. Castagna, "Seismic sparse-layer reflectivity inversion using basis pursuit decomposition," *Geophysics*, vol. 76, no. 6, pp. R147–R158, 2011.

[5] H. Hamid and A. Pidlisecky, "Multitrace impedance inversion with lateral constraints," *Geophysics*, vol. 80, no. 6, pp. M101–M111, 2015.

[6] S. Mohr, J. Wang, G. Ellem, J. Ward, and D. Giurco, "Projection of world fossil fuels by country," *Fuel*, vol. 141, pp. 120–135, 2015.

[7] Reuters, "When it comes to oil, the global economy is still hooked," 2022, available at: https://www.reuters.com/business/energy/when-it-comes-oil-global-economy-is-still-hooked-2022-03-25/.

[8] R. E. Sheriff, *Encyclopedic Dictionary of Applied Geophysics*. Society of Exploration Geophysicists, 2002.

[9] W. M. Telford, L. P. Geldart, and R. E. Sheriff, *Applied Geophysics*. Cambridge University Press, 1990.

[10] S. Chopra and K. J. Marfurt, *Seismic Attributes for Prospect Identification and Reservoir Characterization*. SEG Geophysical Developments Series, 2007, vol. 11.

[11] Ö. Yilmaz, *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of Exploration Geophysicists, 2001.

[12] R. E. Sheriff and L. P. Geldart, *Exploration Seismology*. Cambridge University Press, 1995.

[13] B. Biondi, *3D Seismic Imaging*. Society of Exploration Geophysicists, 2006.

[14] A. R. Brown, *Interpretation of Three-Dimensional Seismic Data*. American Association of Petroleum Geologists, 2011.

[15] M. K. Sen, *Seismic inversion*. Society of Petroleum Engineers Richardson, TX, 2006.

[16] D. Tiab and E. C. Donaldson, *Petrophysics: Theory and Practice of Measuring Reservoir Rock and Fluid Transport Properties*. Elsevier, 2004.

[17] J. M. Hunt, "Generation, migration, and entrapment of petroleum in the earth's crust," *AAPG Bulletin*, vol. 74, pp. 1–28, 1990.

[18] E. A. Robinson and S. Treitel, *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Society of Exploration Geophysicists, 2008.

[19] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, vol. 38, no. 5, pp. 826–844, 1973.

[20] A. Tarantola, *Inverse Problem Theory*. Amsterdam: Elsevier, 2005.

[21] M. B. Widess, "How thin is a thin bed?" *Geophysics*, vol. 38, pp. 1176–1180, 1973.

[22] F. J. Herrmann and X. Li, "Compressive sensing for seismic data acquisition," *Geophysics*, vol. 73, pp. R23–R35, 2008.

[23] J. Zhang and Z. Zhao, "Machine learning in seismic interpretation: Methodologies and applications," *Geophysics*, vol. 78, pp. B11–B21, 2013.

[24] R. A. Clark, *Inverse Theory for Seismic Applications*. Cambridge University Press, 2003.

[25] X. Li and S. Mallick, "Seismic inversion methods: A practical guide," *Geophysics*, vol. 68, pp. 585–588, 2003.

[26] G. Martin, "Marmousi-2: An updated - model for the investigation," 2002, accessed: 2024-02-06. [Online]. Available: https://www.osti.gov/servlets/purl/803186

[27] R. J. Versteeg and G. Grau, "The marmousi experience: Velocity model determination on a synthetic complex data set," *The Leading Edge*, vol. 13, no. 9, pp. 927–936, 1994.

[28] S. M. Gary, W. Robert, and K. J. Marmousi, "Marmousi2: An elastic upgrade for marmousi," *Lead Edge*, vol. 25, pp. 156–166, 2006.

[29] "Agl elastic marmousi," 2020, accessed: 2024-02-06. [Online]. Available: https://wiki.seg.org/wiki/AGL_Elastic_Marmousi

[30] T. S. Subramanian, "A unique and lucrative basin," *India's National Magazine*, vol. 20, no. 02, pp. 18–31, 2003, special Feature: ONGC. [Online]. Available: https://web.archive.org/web/20081028174808/http://www.hinduonnet.com/fline/fl2002/stories/20030131005210200.htm

[31] Krishna godavari basin. https://www.ndrdgh.gov.in/NDR/?page_id=647/. Accessed: ¡insert date of access¿.

[32] E. A. Robinson, "Predictive decomposition of time series with application to seismic exploration," *Geophysics*, vol. 32, no. 3, pp. 418–484, 1967.

[33] R. O. Lindseth, "Synthetic sonic logsâ€"a process for stratigraphic interpretation," *Geophysics*, vol. 44, no. 1, pp. 3–26, 1979.

[34] D. A. Cooke and W. A. Schneider, "Generalized linear inversion of reflection seismic data," *Geophysics*, vol. 48, pp. 665–676, 1983.

[35] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[36] S. Chen and D. L. Donoho, "Examples of basis pursuit," in *Wavelet Applications in Signal and Image Processing III*, vol. 2569. SPIE, 1995, pp. 564–574.

[37] S. Lancaster and D. Whitcombe, "Fast-track'coloured'inversion," in *2000 SEG Annual Meeting*. OnePetro, 2000.

[38] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[39] T. H. Nguyen, *High resolution seismic reflectivity inversion*. University of Houston, 2008.

[40] C. I. Puryear and J. P. Castagna, "Layer-thickness determination and stratigraphic interpretation using spectral inversion: Theory and application," *Geophysics*, vol. 73, no. 2, pp. R37–R48, 2008. [Online]. Available: http://library.seg.org/doi/abs/10.1190/1.2838274

[41] R. Zhang and J. Castagna, "Seismic sparse-layer reflectivity inversion using basis pursuit decomposition," *Geophysics*, vol. 76, pp. R147–R158, 2011.

[42] R. Zhang, M. K. Sen, and S. Srinivasan, "Multi-trace basis pursuit inversion with spatial regularization," *Journal of Geophysics and Engineering*, vol. 10, no. 3, p. 035012, 2013.

[43] M. Ma, S. Wang, S. Yuan, J. Wang, and J. Wen, "Multi-channel spatially correlated reflectivity inversion using block sparse bayesian learning," *Geophysics*, vol. 82, pp. V191–V199, 2017.

[44] H. Wu, S. Li, Y. Chen, and Z. Peng, "Seismic impedance inversion using second-order overlapping group sparsity with a-admm," *Journal of Geophysics and Engineering*, vol. 17, no. 1, pp. 97–116, 2020.

[45] L. Wang, D. Meng, and B. Wu, "Seismic inversion via closed-loop fully convolutional residual network and transfer learning," *Geophysics*, vol. 86, no. 5, pp. R671–R683, 2021.

[46] Y. Hao, X. Wen, H. Zhang, Y. Zhu, and C. Deng, "Nonstationary seismic inversion: joint estimation for acoustic impedance, attenuation factor and source wavelet," *Acta Geophysica*, vol. 69, pp. 459–481, 2021.

[47] R. Dai, F. Zhang, C. Yin, and Y. Hu, "Multi-trace post-stack seismic data sparse inversion with nuclear norm constraint," *Acta Geophysica*, vol. 69, no. 1, pp. 53–64, 2021. [Online]. Available: https://doi.org/10.1007/s11600-020-00506-0

[48] A. Zou, Y. Wang, and D. Wang, "Blind Inversion of Multichannel Nonstationary Seismic Data for Acoustic Impedance and Wavelet," *Pure and Applied Geophysics*, vol. 179, no. 6-7, pp. 2147–2166, 2022.

[49] L. He, H. Wu, X. Wen, and J. You, "Seismic acoustic impedance inversion using reweighted l1-norm sparse constraint," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[50] Y. M. Yang, X. Y. Yin, K. Li, F. Zhang, and J. H. Gao, "Fast pre-stack multi-channel inversion constrained by seismic reflection features," *Petroleum Science*, vol. 20, no. 4, pp. 2060–2074, 2023.

[51] L. Li, G. Zhang, Y. Wang, and G. Hu, "Joint laplace- and fourier-domain seismic inversion for orthorhombic medium parameter estimates," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–10, 2024.

[52] A. Heimer and I. Cohen, "Multichannel seismic deconvolution using markov-bernoulli random-field modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 2047–2058, 2009.

[53] R. Zhang, M. K. Sen, and S. Srinivasan, "A prestack basis pursuit seismic inversion," *Geophysics*, vol. 78, no. 1, pp. R1–R11, 2013.

[54] D. Barman and M. K. Sen, "Seismic inversion with dictionary learning using unsupervised machine learning," *SEG Technical Program Expanded Abstracts*, vol. 2022-Augus, no. Ml, pp. 175–179, 2022.

[55] B. She, Y. Wang, J. Liang, and G. Hu, "Data-driven simultaneous seismic inversion of multiparameters via collaborative sparse representation," *Geophysical Journal International*, vol. 218, no. 1, pp. 313–332, 2019.

[56] D. Gemechu *et al.*, "Sparse regularization based on orthogonal tensor dictionary learning for inverse problems," *Mathematical Problems in Engineering*, vol. 2024, 2024.

[57] L. Yang, H. Song, and T. Hao, "Application of impedance inversion based on bp neural network," *Progress in Geophysics*, vol. 20, pp. 34–37, 2005.

[58] K. Baddari, N. Djarfour, T. Aïfa, and J. Ferahtia, "Acoustic impedance inversion by feedback artificial neural network," *J Petrol Sci Eng*, vol. 71, no. 3, pp. 106–111, 2010.

[59] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

[60] J. Li, R. Cui, D. Pan, M. Hu, and Y. Zhang, "Coalfield seismic inversion using probabilistic neural network," *Prog Geophys*, vol. 27, no. 02, pp. 715–721, 2012.

[61] Y. Ma, X. Ji, T. Fei, and Y. Luo, "Automatic velocity picking with convolutional neural networks," in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 2066–2070.

[62] B. Liu and P. Jiang, "Deep learning based geophysical inversion," in *SEG 2019 Workshop: Geophysics for Smart City Development*. Society of Exploration Geophysicists, 2019, pp. 12–12.

[63] Y. Dai, X. Huang, and Y. Xu, "Seismic inversion based on deep learning and its impact by sample size," in *SEG 2019 Workshop: Fractured Reservoir & Unconventional Resources Forum: Prospects and Challenges in the Era of Big Data*. Society of Exploration Geophysicists, 2019, pp. 65–68.

[64] V. Das and A. Pollack, "Convolutional neural network for seismic impedance inversion," *Geophysics*, vol. 84, pp. R869–R880, 2019.

[65] V. Das, A. Pollack, U. Wollner, and T. Mukerji, "Effect of rock physics modeling in impedance inversion from seismic data using convolutional neural network," in *The 13th SEGJ International Symposium*, 2019, pp. 522–525.

[66] Y. Wu and Y. Lin, "Inversionnet: An efficient and accurate data-driven full waveform inversion," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 419–433, 2019.

[67] S. Li, B. Liu, Y. Ren, Y. Chen, S. Yang, Y. Wang, and P. Jiang, "Deep-learning inversion of seismic data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 2135–2149, 2020.

[68] J. Wang, J. Cao, S. Zhao, and Q. Qi, "S-wave velocity inversion and prediction using a deep hybrid neural network," *Science China Earth Sciences*, p. 1–18, 2022.

[69] B. Wu, Q. Xie, and B. Wu, "Seismic impedance inversion based on residual attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.

[70] V. Kazei, O. Ovcharenko, and T. Alkhalifah, "Velocity model building by deep learning: From general synthetics to field data application," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 1561–1565.

[71] A. Mustafa, M. Alfarraj, and G. AlRegib, "Estimation of acoustic impedance from seismic data using temporal convolutional network," in *Expanded Abstracts of the SEG Annual Meeting*. Society of Exploration Geophysicists, 2019, pp. 15–20.

[72] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, 2015.

[73] L. Tao, H. Ren, and Z. Gu, "Acoustic impedance inversion from seismic imaging profiles using self attention u-net," *Remote Sensing*, vol. 15, no. 4, p. 891, 2023.

[74] J. Sun, K. A. Innanen, and C. Huang, "Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis," *Geophysics*, vol. 86, no. 3, pp. R303–R317, 2021.

[75] M. Araya-Polo, A. Adler, S. Farris, and J. Jennings, "Fast and accurate seismic tomography via deep learning," in *Deep learning: Algorithms and applications*. Springer, 2019, pp. 129–156.

[76] F. Min, L. Wang, S. Pan, and G. Song, "D 2 unet: Dual decoder u-net for seismic image super-resolution reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[77] G. Fabien-Ouellet and R. Sarkar, "Seismic velocity estimation: A deep recurrent neural-network approach," *Geophysics*, vol. 85, no. 1, p. U21–U29, 2020.

[78] R. Guo, J. J. Zhang, D. Liu, Y. B. Zhang, and D. W. Zhang, "Application of bi-directional long short-term memory recurrent neural network for seismic impedance inversion," *81st Annual International Conference and Exhibition, EAGE, Extended Abstracts*, pp. 3–6, 2019.

[79] S.-B. Zhang, H.-J. Si, X.-M. Wu, and S.-S. Yan, "A comparison of deep learning methods for seismic impedance inversion," *Petroleum Science*, vol. 19, no. 3, pp. 1019–1030, 2022.

[80] C. Marques, V. dos Santos, R. Lunelli, M. Roisenberg, and B. Rodrigues, "Analysis of deep learning neural networks for seismic impedance inversion: A benchmark study," *Energies*, vol. 15, no. 20, p. 7452, 2022.

[81] V. Puzyrev, A. Egorov, A. Pirogova, C. Elders, and C. Otto, "Seismic inversion with deep neural networks: A feasibility analysis," in *81st EAGE Conference and Exhibition 2019*, vol. 2019, no. 1. European Association of Geoscientists & Engineers, 2019, pp. 1–5.

[82] R. Gou, Y. Zhang, and X. Zhu, "Bayesian physics-informed neural networks for seismic tomography based on the eikonal equation," *arXiv preprint arXiv:2203.12351*, 2022.

[83] M. Rasht-Behesht, C. Huber, K. Shukla, and G. E. Karniadakis, "Physics-informed neural networks (pinns) for wave propagation and full waveform inversions," *J Geophys Res Solid Earth*, vol. 127, no. 5, p. e2021JB023120, 2022.

[84] Y. Zhang, X. Zhu, and J. Gao, "Seismic inversion based on acoustic wave equations using physics-informed neural network," *IEEE transactions on geoscience and remote sensing*, vol. 61, pp. 1–11, 2023.

[85] Y. Wang, Q. Wang, W. Lu, and H. Li, "Physics-constrained seismic impedance inversion based on deep learning," *IEEE Geoscience and Remote Sensing Letters*, p. 1–5, 2021.

[86] D. Meng, B. Wu, N. Liu, and W. Chen, "Semi-supervised deep learning seismic impedance inversion using generative adversarial networks," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 1393–1396.

[87] L. Mosser, O. Dubrule, and M. J. Blunt, "Stochastic seismic waveform inversion using generative adversarial networks as a geological prior," *Math Geosci*, vol. 52, no. 1, pp. 53–79, 2020.

[88] J. Yao, M. Warner, and Y. Wang, "Regularization of anisotropic fullwaveform inversion with multiple parameters by adversarial neural networks," *Geophysics*, vol. 88, no. 1, pp. R95–R103, 2023.

[89] L. Duque, G. Gutiérrez, C. Arias, A. Rüger, and H. Jaramillo, "Automated velocity estimation by deep learning based seismic-to-velocity mapping," in *81st EAGE Conference and Exhibition 2019*. European Association of Geoscientists & Engineers, 2019, pp. 1–5.

[90] D. Meng, B. Wu, Z. Wang, and Z. Zhu, "Seismic impedance inversion using conditional generative adversarial network," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[91] S. Sun, L. Zhao, H. Chen, Z. He, and J. Geng, "Prestack seismic inversion for elastic parameters using model-data-driven generative adversarial networks," *Geophysics*, vol. 88, no. 2, pp. M87–M103, 2023.

[92] L. Mosser, W. Kimman, J. Dramsch, S. Purves, A. De la Fuente Briceño, and G. Ganssle, "Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks," in *80th eage conference and exhibition 2018*, vol. 2018, no. 1. European Association of Geoscientists & Engineers, 2018, pp. 1–5.

[93] Y. Wang, Q. Wang, W. Lu, and H. Li, "Physics-constrained seismic impedance inversion based on deep learning," *IEEE GRSL*, vol. 19, p. 7503305, 2022.

[94] Z. Zong, Y. Wang, K. Li, and X. Yin, "Broadband seismic inversion for low-frequency component of the model parameter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5177–5184, 2018.

[95] Y. Wu, Y. Lin, and Z. Zhou, "Inversionnet: Accurate and efficient seismic waveform inversion with convolutional neural networks," in *SEG International Exposition and Annual Meeting*. SEG, 2018, pp. SEG–2018.

[96] V. Das, A. Pollack, U. Wollner, and T. Mukerji, "Convolutional neural network for seismic impedance inversion," *Geophysics*, vol. 84, no. 6, pp. R869–R880, 2019.

[97] M. Alfarraj and G. Alregib, "Semisupervised sequence modeling for elastic impedance inversion," *Interpretation*, vol. 7, no. 3, pp. SE237–SE249, 2019.

[98] S. Li, B. Liu, Y. Ren, Y. Chen, S. Yang, Y. Wang, and P. Jiang, "Deep-learning inversion of seismic data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 2135–2149, 2020.

[99] Y. Ren, X. Xu, S. Yang, L. Nie, and Y. Chen, "A physics-based neural-network way to perform seismic full waveform inversion," *IEEE Access*, vol. 8, pp. 112 266–112 277, 2020.

[100] Y. Wang, Q. Ge, W. Lu, and X. Yan, "Well-logging constrained seismic inversion based on closed-loop convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5564–5574, 2020.

[101] Z. Zhang, "Data-driven seismic waveform inversion : A study on the robustness and generalization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 6900–6913, 2020.

[102] A. Adler, M. Araya-Polo, and T. Poggio, "Deep learning for seismic inverse problems: Toward the acceleration of geophysical analysis workflows," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 89–119, 2021.

[103] Y. Lin, J. Theiler, and B. Wohlberg, "Physics-informed data-driven seismic inversion: Recent progress and future opportunities," *Authorea Preprints*, 2022. [Online]. Available: https://www.authorea.com/doi/full/10.1002/essoar.10511175.2

[104] N. Liu, Y. Zhang, Y. Lei, Y. Yang, Z. Wang, J. Gao, and X. Jiang, "Seismic sparse time-frequency network with transfer learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[105] Q. Ma, Y. Wang, Y. Ao, Q. Wang, and W. Lu, "Ub-net: Improved seismic inversion based on uncertainty backpropagation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022.

[106] G. Wang and S. Chen, "Pre-stack seismic inversion with l1-2-norm regularization via a proximal dc algorithm and adaptive strategy," *Surveys in Geophysics*, vol. 43, no. 6, pp. 1817–1843, 2022. [Online]. Available: https://doi.org/10.1007/s10712-022-09725-0

[107] F. Li, Z. Guo, X. Pan, J. Liu, Y. Wang, and D. Gao, "Deep learning with adaptive attention for seismic velocity inversion," *Remote Sensing*, vol. 14, no. 15, 2022.

[108] S. B. Zhang, H. J. Si, X. M. Wu, and S. S. Yan, "A comparison of deep learning methods for seismic impedance inversion," *Petroleum Science*, vol. 19, no. 3, pp. 1019–1030, 2022. [Online]. Available: https://doi.org/10.1016/j.petsci.2022.01.013

[109] D. Manu, P. M. Tshakwanda, Y. Lin, W. Jiang, and L. Yang, "Seismic waveform inversion capability on resource-constrained edge devices," *Journal of Imaging*, vol. 8, no. 12, 2022.

[110] C. Ning, B. Wu, and Z. Zhu, "Transformer and cnn hybrid neural network for seismic impedance inversion," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2023-July, pp. 5543–5546, 2023.

[111] Q. Xie, B. Wu, and Y. Ye, "Attention and hybrid loss guided 2d network for seismic impedance inversion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 3555–3567, 2023.

[112] V. C. Dodda, L. Kuruguntla, S. Razak, A. Mandpura, S. Chinnadurai, and K. Elumalai, "Seismic Lithology Interpretation using Attention based Convolutional Neural Networks," *2023 3rd International Conference on Intelligent Communication and Computational Techniques, ICCT 2023*, pp. 1–5, 2023.

[113] X. Li, J. Han, Q. Yuan, Y. Zhang, Z. Fu, M. Zou, and Z. Huang, "Feusnet: Fourier embedded u-shaped network for image denoising," *Entropy*, vol. 25, no. 10, p. 1418, 2023.

[114] M. Gelboim, A. Adler, Y. Sun, and M. Araya-Polo, "Encoder–decoder architecture for 3d seismic inversion," *Sensors*, vol. 23, no. 1, p. 61, 2022.

[115] P. Y. Bürkle, L. Azevedo, and M. Vellasco, "Deep physics-aware stochastic seismic inversion," *Geophysics*, vol. 88, no. 1, pp. R11–R24, 2023.

[116] H. Li, J. Li, X. Li, H. Dong, G. Xu, and M. Zhang, "Mau-net: A multibranch attention u-net for full-waveform inversion," *Geophysics*, vol. 89, no. 3, pp. R199–R216, 2024.

[117] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[118] N. Parmar and et al., "Image transformer," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 4055–4064.

[119] N. Carion and et al., "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 213–229.

[120] A. Dosovitskiy and et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–22.

[121] Z. Liu and et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9992–10 002.

[122] R. Harsuko and T. A. Alkhalifah, "Storseismic: A new paradigm in deep learning for seismic processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[123] J. Fu, R. Fan, J. Cao, X. Zhang, and S. Shi, "Seismic impedance inversion using a joint deep learning model based on convolutional neural network and transformer," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

[124] S. Feng, X. Zhu, S. Ma, and Q. Lan, "Git: A transformer-based deep learning model for geoacoustic inversion," *Journal of Marine Science and Engineering*, vol. 11, no. 6, p. 1108, 2023.

[125] Y. Guo, L. Fu, and H. Li, "Seismic data interpolation based on multi-scale transformer," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[126] Y. Li, T. Alkhalifah, J. Huang, and Z. Li, "Self-supervised pre-training vision transformer with masked autoencoders for building subsurface model," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[127] L. Gao, H. Shen, and F. Min, "Swin transformer for simultaneous denoising and interpolation of seismic data," *Computers & Geosciences*, vol. 183, p. 105510, 2024.

[128] Y. Dou and K. Li, "3d seismic mask auto encoder: Seismic inversion using transformer-based reconstruction representation learning," *Computers and Geotechnics*, vol. 169, no. February, p. 106194, 2024. [Online]. Available: https://doi.org/10.1016/j.compgeo.2024.106194

[129] K. Li, Y. Dou, Y. Xiao, R. Jing, J. Zhu, and C. Ma, "Transinver: 3d data-driven seismic inversion based on self-attention," *Geophysics*, vol. 89, no. 1, pp. WA127–WA141, 2024.

[130] R. Harsuko and T. A. Alkhalifah, "Storseismic: A new paradigm in deep learning for seismic processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[131] S. Feng, X. Zhu, S. Ma, and Q. Lan, "Git: A transformer-based deep learning model for geoacoustic inversion," *Journal of Marine Science and Engineering*, vol. 11, no. 6, 2023.

[132] J. Fu, R. Fan, J. Cao, X. Zhang, and S. Shi, "Seismic impedance inversion using a joint deep learning model based on convolutional neural network and transformer," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8913–8922, 2023.

[133] Y. Li, T. Alkhalifah, J. Huang, and Z. Li, "Self-supervised pretraining vision transformer with masked autoencoders for building subsurface model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–10, 2023.

[134] Z. Su, J. Cao, T. Xiang, J. Fu, and S. Shi, "Seismic prediction of porosity in tight reservoirs based on transformer," *Frontiers in Earth Science*, vol. 11, no. June, pp. 1–13, 2023.

[135] L. Gao, H. Shen, and F. Min, "Swin transformer for simultaneous denoising and interpolation of seismic data," *Computers and Geosciences*, vol. 183, no. January 2023, p. 105510, 2024. [Online]. Available: https://doi.org/10.1016/j.cageo.2023.105510

[136] N. Liu, J. Huo, Z. Li, H. Wu, Y. Lou, and J. Gao, "Seismic attributes aided horizon interpretation using an ensemble dense inception transformer network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–10, 2024.

[137] C. Ning, B. Wu, and B. Wu, "Transformer and convolutional hybrid neural network for seismic impedance inversion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 4436–4449, 2024.

[138] H. Wang, J. Lin, Y. Li, X. Dong, X. Tong, and S. Lu, "Self-supervised pretraining transformer for seismic data denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–25, 2024.

[139] L. Yang, S. Fomel, S. Wang, W. Li, J. Meng, C. Li, and Y. Chen, "HCTNet: Robust prestack seismic inversion using a hybrid convolutional transformer," *Geophysics*, vol. 90, no. 4, pp. N17–N32, 2025.

[140] Q. Pang, H. Chen, J. Gao, Z. Wang, and P. Yang, "Iterative gradient corrected semisupervised seismic impedance inversion via Swin transformer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, 2025.

[141] Y. Li, A. Fan, S. Pan, G. Song, and F. Min, "Towards complex seismic layers: A vision Transformer approach to full waveform inversion," *Computational Geosciences*, vol. 29, p. 53, 2025.

[142] S. Maurya, N. Singh, and K. H. Singh, *Seismic inversion methods: a practical approach.* Springer, 2020, vol. 1.

[143] M. S. Kiraz, R. Snieder, and J. Sheiman, "Attenuating free-surface multiples and ghost reflection from seismic data using a trace-by-trace convolutional neural network approach," *Geophysical Prospecting*, vol. 72, no. 3, pp. 908–937, 2024.

[144] N. Bregman, R. Bailey, and C. Chapman, "Ghosts in tomography: the effects of poor angular coverage in 2-d seismic traveltime inversion," *Can. J. Expl. Geophys*, vol. 25, no. 1, pp. 7–27, 1989.

[145] Y. Tian, J. Gao, D. Wang, and Z. Li, "Super-resolution optimal basic wavelet transform and its application in thin-bed thickness characterization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[146] N. Alaei, A. Roshandel Kahoo, A. Kamkar Rouhani, and M. Soleimani, "Seismic resolution enhancement using scale transform in the time-frequency domain," *Geophysics*, vol. 83, no. 6, pp. V305–V314, 2018.

[147] O. Koefoed and N. De Voogd, "The linear properties of thin layers, with an application to synthetic seismograms over coal seams," *Geophysics*, vol. 45, no. 8, pp. 1254–1268, 1980.

[148] S. Chopra, J. Castagna, and O. Portniaguine, "Thin-bed reflectivity inversion," in *SEG International Exposition and Annual Meeting*. SEG, 2006, pp. SEG–2006.

[149] G. T. Schuster, *Seismic inversion*. Society of Exploration Geophysicists, 2017.

[150] G. Partyka, J. Gridley, and J. Lopez, "Interpretational applications of spectral decomposition in reservoir characterization," *The leading edge*, vol. 18, no. 3, pp. 353–360, 1999.

[151] K. Marfurt and R. Kirlin, "Narrow-band spectral analysis and thin-bed tuning," *Geophysics*, vol. 66, no. 4, pp. 1274–1283, 2001.

[152] P. K. Nguyen, M. J. Nam, and C. Park, "A review on time-lapse seismic data processing and interpretation," *Geosciences Journal*, vol. 19, no. 2, pp. 375–392, 2015.

[153] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, pp. 3736–3745, 2006.

[154] N. Kazemi and M. D. Sacchi, "Sparse multichannel blind deconvolution," *Geophysics*, vol. 79, no. 5, pp. V143–V152, 2014.

[155] A. Castrodad, Z. Xing, J. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 4263–4281, 2011.

[156] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 3973–3985, 2011.

[157] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, pp. 4779–4789, 2013.

[158] M. Ye, Y. Qian, J. Zhou, and Y. Tang, "Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 1544–1562, 2017.

[159] S. Yu, J. Ma, X. Zhang, and M. Sacchi, "Interpolation and denoising of high-dimensional seismic data by learning a tight frame," *Geophysics*, vol. 80, no. 5, pp. V119–V132, 2015.

[160] Y. Chen, J. Ma, and S. Fomel, "Double-sparsity dictionary for seismic noise attenuation," *Geophysics*, vol. 81, no. 2, pp. V103–V116, 2016.

[161] Y. Zhou, J. Gao, W. Chen, and P. Frossard, "Seismic simultaneous source separation via patchwise sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5271–5284, 2016.

[162] Y. Chen, "Fast dictionary learning for noise attenuation of multidimensional seismic data," *Geophysical Journal International*, vol. 209, pp. 21–31, 2017.

[163] M. Siahsar, S. Gholtashi, A. Kahoo, W. Chen, and Y. Chen, "Data-driven multitask sparse dictionary learning for noise attenuation of 3d seismic data," *Geophysics*, vol. 82, no. 6, pp. V385–V396, 2017.

[164] S. Zu, H. Zhou, R. Wu, W. Mao, and Y. Chen, "Hybrid-sparsity constrained dictionary learning for iterative deblending of extremely noisy simultaneous-source data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2249–2262, 2019.

[165] A. Gholami, "Nonlinear multichannel impedance inversion by total-variation regularization," *Geophysics*, vol. 80, no. 5, pp. R217–R224, 2015.

[166] S. Yuan, S. Wang, N. Tian, and Z. Wang, "Stable inversion-based multitrace deabsorption method for spatial continuity preservation and weak signal compensation," *Geophysics*, vol. 81, no. 3, pp. V199–V212, 2016.

[167] S. Yuan, S. Wang, Y. Luo, W. Wei, and G. Wang, "Impedance inversion by using the low-frequency full-waveform inversion result as an a priori model," *Geophysics*, vol. 84, no. 2, pp. R149–R164, 2019.

[168] D. Slepian and H. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty—i," *Bell System Technical Journal*, vol. 40, pp. 43–63, 1961.

[169] H. Landau and H. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty—ii," *Bell System Technical Journal*, vol. 40, pp. 65–84, 1961.

[170] ——, "Prolate spheroidal wave functions, fourier analysis and uncertainty—iii: The dimension of the space of essentially time- and band-limited signals," *Bell System Technical Journal*, vol. 41, pp. 1295–1336, 1962.

[171] D. Slepian, "Prolate spheroidal wave functions, fourier analysis and uncertainty—iv: Extensions to many dimensions; generalized prolate spheroidal functions," *Bell System Technical Journal*, vol. 43, pp. 3009–3057, 1964.

[172] ——, "Prolate spheroidal wave functions, fourier analysis, and uncertainty—v: The discrete case," *Bell System Technical Journal*, vol. 57, pp. 1371–1430, 1978.

[173] P. Noorishad and S. Yatawatta, "Efficient computation of prolate spheroidal wave functions in radio astronomical source modeling," in *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2011, pp. 326–330.

[174] T. Takami, U. Nielsen, and J. Jensen, "Application of prolate spheroidal wave functions for assessment and prediction of ship responses," in *15th International Symposium on Practical Design of Ships and Other Floating Structures*, Dubrovnik, Croatia, 2022.

[175] T. Takami, U. Nielsen, C. Xi, J. Jensen, and M. Oka, "Reconstruction of incident wave profiles based on short-time ship response measurements," *Applied Ocean Research*, vol. 123, p. 103183, 2022.

[176] M. Gonzalez, "Engineering applications of prolate spheroidal wave functions and sequences and legendre polynomials: Filtering and beamforming in 1d and 2d," Ph.D. dissertation, University of California, 2018.

[177] Q. Yang, M. Lindquist, L. Shepp, C.-H. Zhang, J. Wang, and M. Smith, "Two dimensional prolate spheroidal wave functions for mri," *Journal of Magnetic Resonance*, vol. 158, pp. 43–51, 2002.

[178] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[179] M. A. Davenport and M. B. Wakin, "Compressive sensing of analog signals using discrete prolate spheroidal sequences," *Applied and Computational Harmonic Analysis*, vol. 33, no. 3, pp. 438–472, 2012.

[180] W. Dullaert, L. Reichardt, and H. Rogier, "Improved detection scheme for chipless rfids using prolate spheroidal wave function-based noise filtering," *IEEE Antennas and Wireless Propagation Letters*, vol. 10, pp. 472–475, 2011.

[181] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of selected topics in signal processing*, vol. 4, no. 2, pp. 310–316, 2010.

[182] W. Gao, M. D. Sacchi, and Z. Li, "Microseismic source location via elastic least squares full waveform inversion with a group sparsity constraint," in *SEG Technical Program Expanded Abstracts*, 2017, pp. 2814–2819.

[183] H. Wu, S. Li, Y. Chen, and Z. Peng, "Seismic impedance inversion using second-order overlapping group sparsity with a-admm," *Journal of Geophysics and Engineering*, vol. 17, no. 1, pp. 97–116, 2020.

[184] S. Li, Y. Chen, H. Wu, Z. Peng, and R.-S. Wu, "Seismic acoustic impedance inversion using total variation with overlapping group sparsity," in *SEG Technical Program Expanded Abstracts*, 2018, pp. 411–415.

[185] H. Wang and S. Yu, "Regularized full-waveform inversion with shearlet transform and total generalized variation," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[186] K. Bredies and T. Valkonen, "Inverse problems with second-order total generalized variation constraints," *arXiv preprint arXiv:2005.09725*, 2020.

[187] S. Li, Y. Chen, H. Wu, Z. Peng, and R.-S. Wu, "Seismic acoustic impedance inversion using total variation with overlapping group sparsity," in *2018 SEG International Exposition and Annual Meeting, SEG 2018*, 2019, pp. 411–415.

[188] P. G. Richards and K. Aki, *Quantitative seismology: theory and methods*. Freeman San Francisco, CA, 1980, vol. 859.

[189] J. Zhang, S. Wang, C. Zhang, and T. Feng, "Joint thin bed inversion of pp and ps waves using markov chain monte carlo," *Geophysics*, vol. 79, no. 6, pp. WA1–WA9, 2014.

[190] Y. Gholami, P. Sava, and B. Biondi, "Sparse radon transform for simultaneous thickness and velocity estimation," *Geophysics*, vol. 79, no. 6, pp. WA91–WA98, 2014.

[191] M. Mozayan and Y. Gholami, "Blocky inversion of seismic data using a structure-oriented total variation regularization," *Geophysics*, vol. 83, no. 6, pp. R357–R372, 2018.

[192] X. Zhang, X. Zhang, Y. Chen, and D. Li, "Total variation regularization for seismic data interpolation," *Geophysics*, vol. 78, no. 3, pp. A33–A37, 2013.

[193] X. Liu and X. Yin, "Blocky inversion with total variation regularization and bounds constraint," in *SEG Technical Program Expanded Abstracts*, 2015.

[194] B. Pan and K. Zhang, "Total variation blocky inversion based on a nonlinear forward operator," *Geophysics*, vol. 82, no. 2, pp. WA109–WA121, 2017.

[195] C. Li and F. Zhang, "Amplitude-versus-angle inversion based on the l1-norm-based likelihood function and the total variation regularization constraint," *Geophysics*, vol. 82, no. 3, pp. R173–R182, 2017.

[196] H. Zhi, Y. Zhang, Y. Chen, and R.-S. Wu, "Sparsity constrained seismic inversion using the total variation regularization method," *Journal of Applied Geophysics*, vol. 134, pp. 149–162, 2016.

[197] W. Wang and N. Cao, "Step adaptive fast iterative shrinkage thresholding algorithm for compressively sampled mr imaging reconstruction," *Magnetic Resonance Imaging*, vol. 53, pp. 89–97, 2018.

[198] S. Chakraborty, A. Routray, and R. C. Tewari, "Identification of thin layer via source wavefield dictionary learning," in *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2022, pp. 1–6.

[199] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[200] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Mathematical Programming, Series A*, vol. 137, no. 1–2, pp. 91–129, 2013.

[201] S. Chakraborty, T. Halder, S. B. Dharavath, A. Routray, and D. Chakravarty, "Multitrace seismic inversion using dpss-based dictionary learning," in *2024 IEEE 21st India Council International Conference (INDICON)*, 2024, pp. 1–6.

[202] J. Sun, K. A. Innanen, and C. Huang, "Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis," *Geophysics*, vol. 86, no. 3, pp. R303–R317, 2021.

[203] M. Alfarraj and G. AlRegib, "Semisupervised sequence modeling for elastic impedance inversion," *Interpretation*, vol. 7, no. 3, pp. SE237–SE249, 2019.

[204] R. Biswas, M. K. Sen, V. Das, and T. Mukerji, "Prestack and poststack inversion using a physics-guided convolutional neural network," *Interpretation*, vol. 7, no. 3, pp. SE161–SE174, 2019.

[205] R. Feng, "Unsupervised learning elastic rock properties from pre-stack seismic data," *Journal of Petroleum Science and Engineering*, vol. 192, p. 107237, 2020.

[206] J. Ning, S. Li, Z. Wei, and X. Yang, "Multichannel seismic impedance inversion based on attention u-net," *Frontiers in Earth Science*, vol. 11, p. 1104488, 2023.

[207] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.

[208] Q. Wang, Y. Wang, Y. Ao, and W. Lu, "Seismic inversion based on 2d-cnns and domain adaption," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[209] B. Liu, S. Yang, Y. Ren, X. Xu, P. Jiang, and Y. Chen, "Deep-learning seismic full-waveform inversion for realistic structural models," *Geophysics*, vol. 86, no. 1, pp. R31–R44, 2021.

[210] J. Grabinski, J. Keuper, and M. Keuper, "Aliasing and adversarial robust generalization of cnns," *Machine Learning*, vol. 111, no. 11, pp. 3925–3951, 2022.

[211] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[212] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3–19.

[213] H. Chen, J. Gao, W. Zhang, and P. Yang, "Seismic acoustic impedance inversion via optimization-inspired semisupervised deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[214] X. Wu, S. Yan, Z. Bi, S. Zhang, and H. Si, "Deep learning for multidimensional seismic impedance inversion," *Geophysics*, vol. 86, no. 5, pp. R735–R745, 2021.

[215] K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods.* W. H. Freeman, 1980.

[216] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[217] M. Yu, X. Chen, W. Zhang, and Y. Liu, "Ags-unet: Building extraction model for high-resolution remote sensing images based on attention gates u network," *Sensors*, vol. 22, no. 8, p. 2932, 2022.

[218] E. Thomas, S. Pawan, S. Kumar, A. Horo, S. Niyas, S. Vinayagamani, C. Kesavadas, and J. Rajan, "Multi-res-attention unet: a cnn model for the segmentation of focal cortical dysplasia lesions from magnetic resonance images," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1724–1734, 2020.

[219] S. Jetley, N. A. Lord, N. Lee, and P. Torr, "Learn to pay attention," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=HyzbhfWRW

[220] V. Mnih, N. Heess, A. Graves, and et al., "Recurrent models of visual attention," in *Advances in neural information processing systems*, 2014, pp. 2204–2212.

[221] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," *arXiv preprint arXiv:1711.07971*, 2017.

[222] J. Yu and B. Wu, "Attention and hybrid loss guided deep learning for consecutively missing seismic data reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, art. no. 5902108.

[223] K. Gao, L. Huang, Y. Zheng, R. Lin, H. Hu, and T. Cladohous, "Automatic fault detection on seismic images using a multiscale attention convolutional neural network," *Geophysics*, vol. 87, no. 1, pp. N13–N29, Jan 2022.

[224] Y. Dou, K. Li, J. Zhu, X. Li, and Y. Xi, "Attention-based 3d seismic fault segmentation training by a few 2d slice labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, art. no. 5906715.

[225] O. M. Saad, M. Bai, and Y. Chen, "Uncovering the microseismic signals from noisy data for high-fidelity 3d source-location imaging using deep learning," *Geophysics*, vol. 86, no. 6, pp. KS161–KS173, 2021.

[226] I. Bello and et al., "Attention augmented convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 3285–3294.

[227] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.

[228] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *Stud. Media Commun.*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[229] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[230] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2005, pp. 886–893.

[231] L. G. Baca Ruiz, R. Rueda, M. P. Cuéllar, and M. Pegalajar, "Energy consumption forecasting based on elman neural networks with evolutive optimization," *Expert Systems with Applications*, vol. 92, pp. 380–389, 2018.

[232] S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, "Deep state space models for time series forecasting," in *NeurIPS*, 2018.

[233] B. Su, J. Liu, X. Su, B. Luo, and Q. Wang, "Cfcanet: A complete frequency channel attention network for sar image scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11 750–11 763, 2021.

[234] Y. Xiang, Z. Wang, Z. Song, R. Huang, G. Song, and F. Min, "Seismictransformer: An attention-based deep learning method for the simulation of seismic wavefields," *Computers & Geosciences*, p. 105629, 2024.

[235] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *International conference on machine learning*. PMLR, 2021, pp. 10 347–10 357.

[236] J. Chen, G. Chen, J. Li, R. Du, Y. Qi, C. Li, and N. Wang, "Efficient seismic data denoising via deep learning with improved mca-scunet," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[237] H. Wang, J. Lin, X. Dong, S. Lu, Y. Li, and B. Yang, "Seismic velocity inversion transformer," *Geophysics*, vol. 88, no. 4, pp. R513–R533, 2023.

[238] C. Deng, S. Feng, H. Wang, X. Zhang, P. Jin, Y. Feng, Q. Zeng, Y. Chen, and Y. Lin, "Openfwi: Large-scale multi-structural benchmark datasets for full waveform inversion," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 6007–6020. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/27d3ef263c7cb8d542c4f9815a49b69b-Paper-Datasets_and_Benchmarks.pdf

[239] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[240] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[241] D. Yimin, L. Kewen, Z. Jianbing, L. Timing, T. Shaoquan, H. Zongchao, "Efficient training of 3d seismic image fault segmentation network under sparse labels by weakening anomaly annotation", 2021, [Online]. Available: *https://arxiv.org/abs/2110.05319*

[242] Netherlands F3 Block Dataset [Online]. Available: *https://terranubis.com/datainfo/Netherlands-Offshore-F3-Block-Complete*

[243] *https://www.geeksforgeeks.org/3d-surface-plots-using-plotly-in-python*

Publications from this Thesis

Journal Articles

- **S. Chakraborty**, A. Routray, S. B. Dharavath, and T. Dam, "OrthoSeisnet: Seismic Inversion Through Orthogonal Multi-scale Frequency Domain U-Net for Geophysical Exploration," *Exploration Geophysics* (Taylor & Francis) — arXiv preprint (2024); under review.
- **S. Chakraborty**, A. Routray, S. B. Dharavath, and T. Dam, "AttentionSeisnet: Seismic Inversion Through Hybrid Attention for Geophysical Exploration," manuscript in preparation.
- **S. Chakraborty**, A. Routray, S. B. Dharavath, and T. Dam, "DWTNet: Seismic Inversion Through Discrete Wavelet Transform Attention," manuscript in preparation.
- **S. Chakraborty**, A. Routray, S. B. Dharavath, and T. Dam, "DPSSNet: Seismic Inversion Through Hybrid Attention with Physics-Informed Loss," manuscript in preparation.

Conference Papers

- **S. Chakraborty** and A. Routray, "Multi-trace Inversion with Selective Wedge to Increase Thin Layer Resolution in the Horizontal Layer," *Proc. 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–5, 2020.

- **S. Chakraborty**, A. Routray, and R. C. Tewari, "Identification of Thin Layer via Source Wave-Field Dictionary Learning," *Proc. IECON 2022 — 48th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1–6, 2022.
- **S. Chakraborty**, T. Halder, S. B. Dharavath, A. Routray, and D. Chakravarty, "Multitrace Seismic Inversion Using DPSS-Based Dictionary Learning," *Proc. IEEE 21st India Council International Conference (INDICON)*, 2024.
- **S. Chakraborty**, N. Pandey, S. B. Dharavath, and A. Routray, "Physics-Informed Sparse Seismic Inversion with Wedgelet-TGV and Hybrid ADMM-ALM-AFISTA Algorithm," *Proc. IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, 2025.

Other Publications during PhD

- T. Dam, S. B. Dharavath, S. Alam, N. Lilith, **S. Chakraborty**, et al., "AY-DIV: Adaptable Yielding 3D Object Detection via Integrated Contextual Vision Transformer," *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2024. (10 citations)
- S. B. Dharavath, T. Dam, **S. Chakraborty**, P. Roy, and A. Maiti, "Quantum Inverse Contextual Vision Transformers (Q-ICVT): A New Frontier in 3D Object Detection for AVs," *Proc. ACM International Conference on Information and Knowledge Management (CIKM)*, 2024. (3 citations)
- T. Dam, S. B. Dharavath, S. Alam, N. Lilith, A. Maiti, **S. Chakraborty**, et al., "SaViD: Spectravista Aesthetic Vision Integration for Robust and Discerning 3D Object Detection in Challenging Environments," *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2025. (3 citations)

Author's Biography

Supriyo Chakraborty is a Ph.D. candidate in Data Science at the Advanced Technology Development Centre (ATDC), Indian Institute of Technology Kharagpur, India. He obtained the B.Tech. degree in Electronics and Instrumentation Engineering from the University of Kalyani in 2010 and the M.Tech. degree in Biomedical Engineering from Jadavpur University in 2014. His doctoral research, conducted in collaboration with ONGC (GEOPIC), focuses on seismic sparse layer inversion using signal processing and deep learning techniques for thin-bed reservoir characterization. His research interests include seismic inversion, signal processing, deep learning, computer vision, vision transformers, sparse modelling, unsupervised domain adaptation, and artificial intelligence for geophysical applications. He has published several peer-reviewed research articles in reputed international journals and conferences.

Contact Information

Permanent Address: 1–50, Vidyasagar Upanibesh, Kolkata–700047, West Bengal, India.
Mobile: +91-6291861902
Email: supriyochakraborty@iitkgp.ac.in
meetsupriyochakraborty@gmail.com

Research Interests

Seismic Sparse Layer Inversion, Signal Processing, Deep Learning, Computer Vision, Vision Transformers, Sparse Modelling, Unsupervised Domain Adaptation, Artificial Intelligence for Geophysical Applications, Agentic AI System, Supply Chain.

Education

Doctor of Philosophy (Ph.D.), Electrical Engineering (Submitted)
Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, India
Dissertation: Seismic Sparse Layer Inversion Using Signal Processing and Deep Learning Techniques

Master of Technology (M.Tech.), Biomedical Engineering 2014
Jadavpur University, India (CGPA: 8.44)

Bachelor of Technology (B.Tech.), Electronics and Instrumentation Engineering 2010
University of Kalyani, India (Percentage: 74.29)

Higher Secondary (CBSE) 2006
(Percentage: 87.4)

Secondary (CBSE) 2004
(Percentage: 81.6)

Awards & Honours

- Merit Certificate, Central Board of Secondary Education (CBSE), 2006.
- Ministry of Human Resource Development (MHRD) Fellowship during the Ph.D. programme (2015–2020).
- Presenter, IEEE CVMI 2025.
- Reviewer, IEEE IATMSI 2024.