Abstract

The democratization of easy–to–use, sophisticated image editing software through rapid advancement in digital technology has led to the circulation of enormous number of fake and manipulated images through the Internet and social media. Fake imagery misrepresents the truth and manipulates viewers, thus calling for analysis methods to detect and uncover the history of questionable multimedia contents. Over the past decade, *Multimedia Forensics* has grown dynamically to build up techniques that provide ways to test the authenticity and integrity of an image to identify forgeries. Multimedia forensics analyses an image using image processing methods and domain expertise to reconstruct the processing history of the image since its acquisition, and detect manipulations (if any). In the first part of this thesis, we investigate finding traces of manipulation and using the benefits of decision–making performance of Machine Learning and Deep Learning algorithms to develop multimedia forensics tools. An adversary may perform copy–paste forgery where portions from JPEG image files are copy–pasted into another image file (JPEG, TIFF, etc.) and then save it in uncompressed format (TIFF) or compressed format (JPEG). Multiple compression traces in an image would suggest the possibility of image manipulations. We first develop two blind JPEG–based forgery detection tools by identifying traces of double JPEG compression; first, by computing an optimal error–image matrix that depicts the existence of forgeries and secondly, by extracting Discrete Cosine Transform Residual (DCTR) features and using a Support Vector Machine (SVM) model for classification. However, a manipulated image is often made to undergo post–manipulation processing techniques to conceal detectable traces of forgeries. These efforts of removing detectable traces of forgeries are termed "anti–forensics". Next, we resort to devising a counter anti–forensic method that exploits the traces left by the anti–forensics methods using a Deep Learning (DL) algorithm that locally and globally allows the extraction of residual features for classification. We specifically target to develop a counter anti–forensic method on median filtered JPEG images. In another category of multimedia forensics, i.e., the identification of 3D rendered Computer Generated (CG) images, most state–of–the–art Convolutional Neural Network (CNN) based techniques come at the cost of high computational overhead, both during training and testing. In this thesis, we develop an efficient DL technique that, instead of processing an entire image, is trained to adaptively integrate features extracted from the sequence of selected patches of the input image. Thus, resulting in fewer computational parameters and lesser computation time for performing the classification.

Malware Classification is currently an active field of research related to computer security. Classification of malware is helpful for the analyst to get a better insight into the functioning of the malware. In the next part of this thesis, we develop a malware classification method based on malware visualization using DL technique. The main insight behind the use of malware visualization is to use image processing methods to identify visual patterns in the layout and texture of a binary malware program so as to infer its family. We investigate a CNN architecture that can effectively train and classify malware program files represented in RGB and grayscale image forms. Using a multi–scale attention based technique, the network is guided to focus its learning on those spatial parts that contain information relevant to the input.

Keywords: Multimedia Forensics, JPEG–based Forgery, Median Filtering Attacks, Computer Generated Images, Malware, Deep Learning