INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student : Sadu Chiranjeevi

Roll Number : 146101026

Programme of Study : Ph.D.

Thesis Title: Detection Methods against Digital Image Attacks for Secure Computer Vision

Name of Thesis Supervisor(s) : Prof. Pradip K. Das

Thesis Submitted to the Department/ Center : CSE

Date of completion of Thesis Viva-Voce Exam : 14/06/2023

Key words for description of Thesis Work : Digital Images, Digital Image Attacks, Face Swap Attacks, Copy-move Forgery Attacks, Adversarial Attacks, Detection Methods, Computer Vision.

SHORT ABSTRACT

In today's digital age, our everyday life is filled with digital multimedia data as one of the primary forms for communication. As a result, Computer Vision (CV) systems supported by Machine Learning (ML) and Deep Learning (DL) techniques are now pervasive to process such multimedia. However, with modern technologies in sophisticated editing tools and DL models, it becomes a critical task to protect CV systems from digital image attacks. This thesis focuses on detecting a spectrum of digital attacks at the image level.

Digital images play a pivotal role for carrying important information in many real-world fields. With developments in modern technologies, attacking digital images becomes an easy task. Therefore, authentication of digital images is necessary. The thesis mainly focuses on (i) detection of face swap attacks (ii) detection of Copy-Move Forgery (CMF) attacks and (iii) detection of facial adversarial attacks.

Face swapping transfers the face of a source image to the face of a destination image or vice-versa while preserving photo-realism. We propose a method to create face swap attacks on original images and a technique to defend against them. Augmented 81-facial landmark points are extracted for creating the face swap attacks. The features are provided to Support Vector Machines (SVMs). The proposed detection method detects face swap attacks with 95% accuracy on a real-world dataset.

In CMF attacks, the attacker copies some regions of the image and pastes them into one or more regions of the same image. We propose a detection method for such forgery regions based on Binary Robust Invariant Scalable Keypoints (BRISK) and Speeded Up Robust Features (SURF) descriptors. Both fused features are matched and clustering is performed to reduce false positives. The proposed method is tested on real-world copy-move datasets. Experimental results show that our method is robust against various geometric transformations and precisely determines the forged regions.

ML models and especially DL models have impressively performed on perceptual tasks over the past few years. However, these models remain vulnerable to carefully crafted adversarial attacks. Therefore, the detection of adversarial attacks is essential for the rightful and confident usage of DL-based solutions in the real world. As face provides a rich source of information, we propose novel defense methods to detect different types of adversarial facial attacks. The proposed defense methods are evaluated on real-world datasets and experimental results show that they are robust against a wide range of adversarial face attacks.