Deep Learning-Based Hybrid Model for Detecting AI-Generated Fake Images

Zulfiye Beyza Metin, Fatih Ozkaynak

Firat University, Department of Software Engineering, Elazig, Turkey.

Abstract

Fake images are digitally generated or altered using artificial intelligence, often through advanced deep learning techniques like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Models. These images pose significant risks, particularly in credibility-sensitive areas such as fake news and biometric security. The failure of security systems to detect fake images can lead to identity fraud, data manipulation, and misinformation, ultimately compromising information security and public trust. In this study, we proposed a deep learning-based hybrid method for fake image detection. Our approach leverages ResNet-50 and Vision Transformer (ViT) models for feature extraction. The extracted features are then combined into a single feature vector, which is processed by the XGBoost algorithm for classification. The model was trained and tested on the CIFAKE dataset, consisting of 60,000 real and 60,000 fake images. Experimental results demonstrate that our method achieves a high accuracy of 98.77% in detecting fake images. These findings highlight the effectiveness and reliability of our approach, surpassing existing methods in performance.

Keywords: Artificial Image Forensics, Feature Extraction, Generative Adversarial Networks, Synthetic Image, Vision Transformer