Image Artifact Analysis for Deepfake Detection Using GAN with EfficientNet-B4 and Self-Attention

Mrs. S.Sarawathi
Assistant Professor, JSS Institute of Speech and Hearing, Mysore, India

Lakshmi S
Department of Computer Science, MES College of Arts, Commerce and Science Bangaluru, India

Abstract:

The rapid advancement of generative models has led to the creation of highly convincing deepfake images, raising significant concerns in security, media integrity, and digital forensics. This paper presents a robust deepfake detection framework leveraging a hybrid architecture combining EfficientNet-B4 and Self-Attention mechanisms within a Generative Adversarial Network (GAN) framework. The proposed system introduces an enhanced pre-processing pipeline including Error Level Analysis (ELA) and edge detection to expose visual inconsistencies commonly present in manipulated images. The discriminator, enriched with EfficientNet-B4 and Self-Attention, effectively distinguishes between real and fake images, achieving high classification accuracy. Experiments were conducted on a balanced dataset comprising over 13,000 real and fake images. The results demonstrate significant improvements in detection accuracy and artifact localization. This study contributes a novel integration of CNN-transformer hybrid models with artifact-focused pre-processing, offering a promising direction for future deepfake detection systems.

Keywords:

Deepfake Detection, EfficientNet-B4, Self-Attention, GAN, Image Artifacts, Error Level Analysis, Edge Detection, CNN, Transformer.