ABSTRACT

The advancement of artificial intelligence (AI) technology based on visual content manipulation, particularly in videos, has become increasingly sophisticated. One prominent example is deepfake, a method that uses neural networks to generate or alter videos realistically. While this technology is beneficial in fields like entertainment and education, its potential misuse—such as creating deceptive or reputation-damaging fake videos—poses a serious threat. To address this issue, an approach based on Generative Adversarial Networks (GAN) offers a promising solution. GAN consists of a generator that produces fake data and a discriminator that differentiates between real and fake data. In the context of deepfake classification, GAN is competitively trained to adaptively recognize video manipulations by leveraging unique patterns in deepfakes that are challenging for humans to classify. This approach not only enhances classification accuracy as data complexity increases but also enables a faster process compared to conventional methods. The implementation of GAN in AI video manipulation classification has significant implications for digital and social security, protecting against the spread of false information and increasing public trust in digital media. This research represents a crucial step in safeguarding information integrity in the digital era.

Keywords : AI, GAN, Deepfake