

ABSTRAK

GRACHIELLA DIAN MALEMUKUR. Deteksi *Deepfake* pada Citra Wajah Menggunakan Arsitektur *Hybrid CNN-Transformer*. Dibimbing oleh MARTALELI BETTIZA dan MARISHA PERTIWI

Penyebaran konten *deepfake* yang semakin masif menimbulkan ancaman serius terhadap integritas informasi visual, sehingga dibutuhkan sistem deteksi yang akurat dan andal. Penelitian ini bertujuan membangun sistem deteksi *deepfake* pada citra wajah menggunakan arsitektur *hybrid* yang dibangun dengan menggabungkan ResNet-34 sebagai *backbone* CNN untuk mengekstraksi fitur lokal dan *Data-efficient Image Transformer (DeiT)-Small* sebagai *backbone Transformer* untuk mengekstraksi fitur global, yang kemudian digabungkan menggunakan metode *feature fusion concatenation* sebelum diklasifikasikan. Dataset pelatihan yang digunakan merupakan dataset multi-source yang terdiri dari 20.776 citra gabungan dari GRAVEX-200K, FaceForensics++, dan Deepfake and Real Images dengan komposisi kelas yang seimbang. Hasil eksperimen menunjukkan bahwa model *hybrid* yang dilatih mencapai 96,88%, *precision* 96,89%, *recall* 96,88%, dan *F1-score* 96,88%, pada data uji internal. Pengujian lintas dataset menggunakan StyleGAN–StyleGAN2 Deepfake Face Images menunjukkan penurunan *accuracy* menjadi 74,75% akibat *domain shift*.

Kata Kunci: CNN-Transformer, *deepfake*, deteksi citra wajah, DeiT-Small, ResNet-34

ABSTRACT

GRACHIELLA DIAN MALEMUKUR. Deepfake Detection on Facial Images Using a Hybrid CNN–Transformer Architecture. Supervised by MARTALELI BETTIZA and MARISHA PERTIWI.

The rapid proliferation of deepfake content poses a serious threat to the integrity of visual information, creating an urgent need for accurate and reliable detection systems. This study aims to develop a deepfake detection system for facial images using a hybrid architecture that combines ResNet-34 as a CNN backbone for local feature extraction and Data-efficient Image Transformer (DeiT)-Small as a Transformer backbone for global feature extraction. The features extracted by both backbones are fused using a feature fusion concatenation method before being classified. The training dataset consists of a balanced multi-source dataset containing 20,776 images collected from GRAVEX-200K, FaceForensics++, and Deepfake and Real Images. Experimental results show that the proposed hybrid model achieves an accuracy of 96,88%, precision of 96,89%, recall of 96,88%, and F1-score of 96,88% on the internal test set. Cross-dataset evaluation using the StyleGAN–StyleGAN2 Deepfake Face Images dataset reveals a decrease in accuracy to 74.75%, indicating the impact of domain shift between the training and testing data distributions.

Keywords: *CNN–Transformer, deepfake, facial image detection, DeiT-Small, ResNet-34.*