



INTISARI

DETEKSI CITRA WAJAH GENERATIF MENGGUNAKAN FITUR *HYBRID* DOMAIN SPASIAL DAN FREKUENSI

Oleh

Zalfaa Aulia

22/497187/PA/21403

Perkembangan Generative Adversarial Networks membuat citra wajah sintesis semakin realistis sehingga sulit dibedakan dari citra nyata jika hanya mengandalkan petunjuk visual pada domain spasial. Penelitian ini mengusulkan metode deteksi citra wajah generatif berbasis arsitektur *multi-stream* yang menggabungkan fitur RGB (spasial), Fast Fourier Transform/FFT (frekuensi global), dan Discrete Wavelet Transform/DWT (frekuensi lokal). Untuk memperkuat informasi frekuensi, citra terlebih dahulu dikonversi ke ruang warna YCbCr sebelum ditransformasikan menggunakan FFT dan DWT, kemudian tiap stream diproses oleh backbone CNN (ResNet-18/34/50 dan MobileNetV2) dan digabungkan menggunakan *flatten-concatenation fusion* untuk klasifikasi biner *real* atau *fake*. Dataset yang digunakan merupakan subset dari Face Deepfake Detection Challenge berjumlah 10.000 citra (5.000 real dari CelebA dan FFHQ, serta 5.000 fake dari AttGAN, GDWCT, StarGAN, StyleGAN, dan StyleGAN2) dengan resolusi dinormalisasi menjadi 256×256 . Hasil pengujian pada skenario 5-GAN menunjukkan bahwa integrasi fitur frekuensi meningkatkan performa secara signifikan dibanding baseline spasial; konfigurasi ALL (RGB+FFT+DWT) mencapai akurasi 0,9813 (F1 0,9813) pada ResNet-18 dan 0,8993 (F1 0,8975) pada MobileNetV2, jauh di atas baseline spasial pada backbone yang sama. Pada skenario tambahan 1-GAN (StyleGAN2), performa model cenderung stabil dan tinggi, namun ALL tidak selalu menjadi yang terbaik pada semua backbone; pada beberapa kasus konfigurasi RGB+DWT memberikan hasil lebih optimal, khususnya pada backbone ringan.

Kata-kata kunci : *Fake image detection, Generative Adversarial Networks (GANs), Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT), Multi-stream architecture*



ABSTRACT

GENERATIVE FACE IMAGES DETECTION USING HYBRID SPATIAL AND FREQUENCY DOMAIN FEATURES

By

Zalfaa Aulia

22/497187/PA/21403

The rapid advancement of Generative Adversarial Networks has produced increasingly realistic synthetic face images, making them difficult to distinguish from real images when relying only on spatial-domain visual cues. This study proposes a multi-stream fake face image detection method that combines RGB (spatial) features with frequency-domain representations using the Fast Fourier Transform (FFT) for global spectral cues and the Discrete Wavelet Transform (DWT) for multi-resolution frequency information that preserves locality. To strengthen frequency-related characteristics, images are first converted into the YCbCr color space before applying FFT and DWT. Each stream is processed using CNN backbones (ResNet-18/34/50 and MobileNetV2), and the resulting features are fused using flatten–concatenation fusion for binary real/fake classification. The dataset is a subset of the Face Deepfake Detection Challenge consisting of 10,000 images (5,000 real images from CelebA and FFHQ, and 5,000 fake images generated by AttGAN, GDWCT, StarGAN, StyleGAN, and StyleGAN2), with all samples normalized to 256×256 resolution. Experimental results under the 5-GAN scenario show that incorporating frequency features consistently improves performance compared to the spatial-only baseline. The ALL (RGB+FFT+DWT) configuration achieves an accuracy of 0.9813 (F1 0.9813) with ResNet-18 and 0.8993 (F1 0.8975) with MobileNetV2, substantially outperforming the corresponding spatial baselines. In the additional 1-GAN (StyleGAN2) scenario, models generally yield stable and high performance; however, ALL is not always the best across all backbones, and in some cases RGB+DWT provides more optimal results, particularly for lightweight backbones.

Keywords : Fake image detection, Generative Adversarial Networks (GANs), Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT), Multi-stream architecture