Karya Tulis Ilmiah Mahasiswa UMSIDA.docx

By User Upload

WORD COUNT	3275	TIME SUBMITTED	23-FEB-2026 07:12PM
		PAPER ID	120530094

Web-Based Application Design to Differentiate AI-Generated Videos from Original Videos Using Transformer Method
Perancangan Aplikasi Berbasis Web untuk Membedakan Video yang Dihasilkan oleh (AI) dengan Video Asli Menggunakan Metode Transformer

Eka Nugraha Saktifany Wicaksana[1)], Rohman Dijaya [*,2)] ,Irwan Alnarus Kautsar [3)], Nuril Lutvi Azizah [4)][1,2,3,4])Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

rohman.dijaya@umsida.ac.id

Abstract. *The advancement of deep learning and other artificial intelligence (AI) technologies has resulted in the emergence of synthetic videos, or deepfakes, that are increasingly realistic and difficult to distinguish from authentic videos. This situation poses a serious threat to the integrity of digital information due to their potential use to spread misinformation, fraud, and violate ethics, privacy, and security. Therefore, accurate, adaptive, and reliable deepfake detection methods are needed to address the ever-evolving visual manipulation techniques.*

This research proposes a deepfake detection model based on a hybrid architecture that combines ConvNeXt and Vision Transformer (ViT). ConvNeXt is used to extract local features of facial images, while Vision Transformer is used to capture global context more comprehensively. Furthermore, the model is equipped with two reconstruction paths, namely Autoencoder (AE) and Variational Autoencoder (VAE), to increase the model's sensitivity in detecting subtle visual manipulation artifacts.

The dataset used is derived from publicly available authentic and deepfake videos, with approximately one million extracted facial image frames. The data is divided into training, validation, and testing sets. The experimental results show that the model is able to achieve an accuracy level of up to 98% on the test data, as well as a high F1-Score value, for the dfdc (99.1), ff++ (95.5), Timit (98.3), Celeb DF v2 (91.6) datasets they averaged to 96.125..

Keywords - *Deepfake Detection, Deep Learning, ConvNeXt, Vision Transformer*

Abstrak. *Kemajuan metodologi kecerdasan buatan (AI), terutama dalam domain pembelajaran mendalam., telah mendorong munculnya video sintetis atau deepfake yang semakin realistis dan sulit dibedakan dari video asli. Kondisi ini menimbulkan ancaman serius terhadap integritas informasi digital karena berpotensi digunakan untuk menyebarkan misinformasi, penipuan, serta pelanggaran etika, privasi, dan keamanan. Oleh karena itu, diperlukan metode deteksi deepfake yang akurat, adaptif, dan andal dalam menghadapi teknik manipulasi visual yang terus berkembang.*

Penelitian ini mengusulkan model deteksi deepfake berbasis arsitektur hybrid yang mengombinasikan ConvNeXt dan Vision Transformer (ViT). ConvNeXt dimanfaatkan untuk mengekstraksi fitur lokal citra wajah, sementara Vision Transformer digunakan untuk menangkap konteks global secara lebih komprehensif. Selain itu, model dilengkapi dua jalur rekonstruksi, yaitu Autoencoder (AE) dan Variational Autoencoder (VAE), guna meningkatkan sensitivitas model dalam mendeteksi artefak manipulasi visual yang bersifat halus.

Dataset yang digunakan berasal dari video asli dan deepfake yang tersedia secara publik, dengan sekitar satu juta citra wajah hasil ekstraksi frame. Data dibagi menjadi set pelatihan, validasi, dan pengujian. Hasil eksperimen menunjukkan bahwa model mampu mencapai tingkat akurasi hingga 98% pada data pengujian, serta nilai F1-Score yang tinggi, untuk dataset dfdc (99,1),ff++(95,5),Timit(98,3),Celeb DF v2(91,6) mereka dirata-rata menjadi 96,125..

Kata Kunci - *Deepfake Detection, Deep Learning, ConvNeXt, Vision Transformer*

I. PENDAHULUAN

Fenomena deepfake atau video sintetis berbasis kecerdasan buatan (AI) menjadi isu serius dalam perkembangan teknologi informasi. Dengan kemajuan deep learning, AI kini mampu menghasilkan video yang sangat menyerupai kondisi nyata, sehingga meningkatkan risiko penyebaran misinformasi di ruang digital. Teknologi deepfake tidak hanya dimanfaatkan untuk hiburan, tetapi juga menimbulkan persoalan etika, privasi, dan keamanan informasi. Manipulasi visual semacam ini berpotensi merusak integritas komunikasi, menyulitkan proses verifikasi berita oleh jurnalis, serta menurunkan kepercayaan publik terhadap sumber informasi resmi.Menanggapi permasalahan tersebut, diperlukan pendekatan sistematis untuk mendeteksi keaslian video secara efektif. Penelitian ini mengusulkan model

deteksi deepfake berbasis arsitektur Transformer yang memiliki kemampuan kuat dalam memahami hubungan spasial melalui mekanisme self-attention. Untuk memperkuat ekstraksi fitur lokal, digunakan ConvNeXt sebagai pengembangan jaringan konvolusional yang efisien dan akurat. Sementara itu, Vision Transformer (ViT) berperan dalam menangkap konteks global, sehingga mampu mendeteksi artefak manipulasi yang bersifat halus dan tersebar. Kombinasi ConvNeXt dan ViT diharapkan dapat menghasilkan sistem deteksi deepfake yang lebih akurat, adaptif, dan andal dalam menghadapi kompleksitas manipulasi visual yang semakin meningkat.

II. METODE

Penelitian ini menggunakan data sekunder berupa video asli dan video deepfake yang tersedia secara publik dengan variasi resolusi, sudut pengambilan gambar, dan kondisi pencahayaan. Video-video tersebut diekstraksi menjadi frame gambar dengan fokus pada area wajah menggunakan algoritma deteksi wajah berbasis Python, kemudian dilakukan proses cropping dan verifikasi kualitas. Dari proses ini dihasilkan sekitar 1.000.000 citra wajah yang terbagi seimbang antara kelas real dan fake. Selanjutnya, dataset Dataset dipartisi menjadi subset pelatihan (80%), validasi (15%), dan pengujian (5%).

2.1. Pengumpulan Dataset

Dataset DFDC terdiri dari lebih dari 100.000 video beresolusi tinggi yang mencakup video asli (real) dan video manipulasi (deepfake). Dataset ini melibatkan 3.426 partisipan dan direkam dalam beragam kondisi pencahayaan, latar belakang, serta sudut pengambilan gambar. Proses manipulasi dilakukan menggunakan delapan teknik deepfake berbeda, sehingga menghasilkan variasi data yang tinggi dan bersifat representatif. Komposisi data menunjukkan rasio sekitar 6:1 antara video deepfake dan video asli, dengan dominasi jumlah video manipulasi..

Dataset FaceForensics++ terdiri dari 1.000 video asli yang dikumpulkan dari YouTube dan dimanipulasi menggunakan empat metode manipulasi wajah otomatis, yaitu DeepFakes, Face2Face, FaceSwap, dan NeuralTextures. Dataset ini tersedia dalam berbagai tingkat kompresi, seperti c23 (kompresi sedang) dan c40 (kompresi tinggi), serta mendukung beragam resolusi video untuk keperluan evaluasi deteksi deepfake.

Dataset TIMIT (TM) mencakup 4.380 video deepfake dan 2.563 video asli yang dirancang untuk menguji sistem deteksi manipulasi audio–visual. Meskipun ukurannya lebih kecil dibandingkan DFDC, dataset ini digunakan secara eksklusif pada tahap pelatihan untuk memperluas variasi teknik manipulasi yang dipelajari oleh mode

Dataset Celeb-DF (v2) terdiri dari 890 video asli dan 5.639 video deepfake dengan kualitas manipulasi yang tinggi. Dataset ini menampilkan wajah figur publik, sehingga sangat sesuai untuk pengujian performa model pada skenario dunia nyata

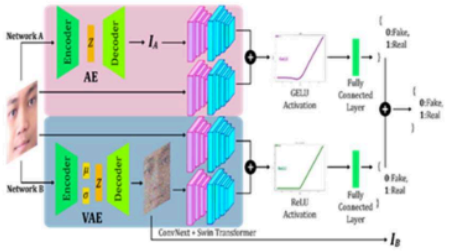
2.2. Pre-Processing Data Video

Proses pre-processing diimplementasikan secara metodis guna menjamin kualitas serta relevansi data yang ada dalam penelitian ini studi dan evaluasi Pendekatan. deteksi deepfake. Tahapan dimulai dengan ekstraksi region wajah dari setiap video menggunakan library computer vision, sehingga model difokuskan pada area visual yang paling informatif. Selanjutnya, citra wajah disesuaikan ke ukuran standar 224 × 224 piksel dengan format RGB guna menjaga keseimbangan antara detail visual dan efisiensi komputasi.

Setelah itu, dilakukan verifikasi kualitas gambar untuk memastikan wajah terdeteksi dengan jelas dan utuh, serta menghindari keberadaan data bermasalah yang dapat menimbulkan noise label. Frame yang tidak relevan atau gagal terdeteksi wajahnya dihapus untuk menjaga konsistensi dan integritas dataset. Pada tahap akhir, dataset hasil pre-processing yang berjumlah sekitar 1.000.000 citra wajah dipisahkan menjadi data studi, validasi, dan pengujian menggunakan distribusi label yang seimbang antara kelas real dan fake. Proses ini bertujuan untuk mendukung pembelajaran model yang adil, stabil, dan memiliki kemampuan generalisasi yang baik.

2.3. Perancangan Arsitektur

Dalam penelitian ini menerapkan pendekatan arsitektur hybrid dengan mengombinasikan kemampuan ekstraksi fitur lokal dari ConvNeXt dan pemahaman konteks global dari Vision Transformer untuk deteksi video deepfake. Model dirancang dengan dua alur pemrosesan utama, yaitu Network A yang memanfaatkan Autoencoder (AE) untuk rekonstruksi citra dan Network B yang menggunakan Variational Autoencoder (VAE) berbasis distribusi probabilistik. Kedua alur tersebut dikombinasikan dengan ConvNeXt–Swin Transformer sebagai modul ekstraksi fitur dan klasifikasi. Hasil prediksi dari kedua jaringan kemudian digabungkan untuk menghasilkan keputusan klasifikasi akhir, sehingga meningkatkan akurasi dan ketahanan model terhadap variasi manipulasi visual.



Gambar 1. perancangan model

Gambar menunjukkan arsitektur model hybrid untuk deteksi video deepfake yang mengombinasikan Autoencoder (AE) dan Variational Autoencoder (VAE) dengan Transformer. Network A dan Network B memproses citra melalui rekonstruksi dan representasi laten, kemudian fitur diekstraksi dan diklasifikasikan, sebelum hasil prediksi digabungkan untuk menghasilkan keputusan akhirGambar 3.4 Diagram arsitektur model [16]

2.4. Autoencoder (AE) dan Variational Autoencoder (VAE)

Autoencoder (AE) dan Variational Autoencoder (VAE) berfungsi untuk merekonstruksi citra dan mempelajari representasi laten, sehingga membantu model menangkap karakteristik spasial dan distribusi data yang membedakan citra asli dan deepfake

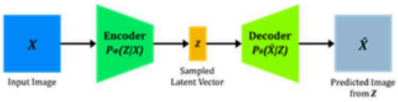
Tabel 2.4 Konfigurasi encoder AE dan VAE

Layer	Kernel	Stride	Channel AE (Encoder)	Channel VAE (Decoder)
1	3×3	1	3 → 16	3 → 16
2	3×3	1	16 → 32	16 → 32
3	3×3	1	32 → 64	32 → 64
4	3×3	1	64 → 128	64 → 128
5	3×3	1	128 → 256	128 → 256

Tabel 2.4 menunjukkan konfigurasi encoder pada model Autoencoder (AE) dan Variational Autoencoder (VAE). Setiap layer menggunakan kernel 3×3 dengan stride 1, sementara jumlah channel meningkat secara bertahap dari 3 hingga 256 untuk mengekstraksi fitur citra yang semakin kompleks sebagai representasi laten

2.5. Autoencoder (AE)

Autoencoder terdiri dari Encoder dan Decoder yang bekerja secara berurutan untuk mengompresi citra input ke dalam representasi laten dan merekonstruksinya kembali, sebagaimana diilustrasikan pada Gambar



Gambar 2 Autoenconder

Autoencoder merupakan arsitektur yang tersusun atas dua komponen utama, yaitu encoder dan decoder.yang bekerja secara berurutan untuk mengompresi citra input ke dalam representasi laten dan merekonstruksinya kembali, sebagaimana diilustrasikan pada Gambar X

2.6. Evaluasi dan Ablasi Studi

Tahap pelatihan dilakukan dengan mengintegrasikan seluruh komponen arsitektur yang telah dirancang guna mencapai performa optimal dalam deteksi video deepfake. Proses ini mencakup pemilihan backbone, pengaturan hyperparameter, serta strategi optimasi. Model menggunakan ConvNeXt Tiny sebagai backbone berbasis CNN untuk mengekstraksi fitur spasial lokal secara efisien, dan Swin Transformer Tiny sebagai modul Vision Transformer untuk menangkap konteks global melalui mekanisme window-based self-attention. Untuk optimasi, digunakan AdamW optimizer dengan learning rate 0,0001 dan weight decay 0,0001 guna memastikan konvergensi yang stabil sekaligus mengurangi risiko overfitting

2.7. Pengujian Validasi

Pengujian dan validasi dilakukan setelah proses pelatihan untuk mengevaluasi kemampuan model untuk menggeneralisasi terhadap data yang belum familiar. Penilaian dilakukan dengan menggunakan test set yang terdiri dari 50.000 citra wajah dengan distribusi seimbang antara kelas real dan fake. Kinerja model diukur menggunakan metrik evaluasi standar Ini biasanya digunakan dalam penelitian deep learning untuk menentukan skor klasifikasi secara komprehensif

.

III. HASIL DAN PEMBAHASAN

3.1. Pengaturan Parameter Model

Penelitian ini menetapkan Variabel pelatihan standar yang diimplementasikan secara konsisten pada seluruh eksperimen untuk memastikan proses pelatihan berjalan stabil, terukur, dan dapat direplikasi, serta menghasilkan model dengan kemampuan generalisasi yang baik..

3.2. Hasil Pelatihan Model

Proses pelatihan dilakukan menggunakan dua jaringan, yaitu Network A berbasis Autoencoder (AE) dan Network B berbasis Variational Autoencoder (VAE), yang dilatih secara terpisah selama 30 epoch dengan bantuan GPU NVIDIA Tesla V100. Waktu pelatihan Network A sekitar 60 menit dan Network B sekitar 93 menit. Evaluasi model dilakukan pada empat dataset, yaitu DFDC, FF++, DeepfakeTIMIT, dan Celeb-DF v2, dengan membandingkan kinerja model ensemble terhadap masing-masing jaringan untuk mengukur ketahanannya terhadap berbagai teknik manipulasi deepfake

Tabel 3.2 Hasil pelatihan model

Dataset	Model (AE + VAE)			AE			VAE		
	A	R	F	A	R	F	A	R	F
1.DFDC	98.5	98.7	98.4	97.5	98.7	97.2	98.4	98.7	95.5
2.FF++	97.0	95.5	98.5	95.5	94.1	95.5	96.8	95.5	98.0
3.TIMIT	98.2	-	98.2	97.6	-	97.5	97.8	-	97.8
4.Celeb-DF (v2)	90.9	92.4	98.5	94.0	87.6	95.9	94.2	83.0	97.6

Pada tabel diatas dijelaskan setiap dataset dimasukan dan uji coba dengan metode ae vae dan keduanya menghasilkan niali sebagai berikut

1.Dataset DFDC pada pengujian metode keduanya model mendeteksei 98, Real nilai 98 dan fake dideteksi 98

2.Dataset FF++ pada pengujian metode keduanya model mendeteksi 97, realnya rata-rata 95 dan Fake dideteksi rata-rata98

3.Dataset Timit pada pengujian metode keduanya model mendeteksi 98 sementara realnya rata-rata 98dan Fake rata-rata dideteksi 97

4.Dataset Celeb b-DF (v2) pada pengujian metode keduanya model mendeteksi 90 sementara realnya 92 dan fakenya rata-rata 98

Perhitungan ini dihasilkan otputnya diteksi deepleraning sendiri perkalian bobot weigth penambahan bias kemudian dilakuakan fungsi aktifasi relu kemudian masuk bagian hidden layer setelah masuk hiden layer masuk ke output

3.3. Pengujian Model

Tabel 3.3 pengujian model

Parameter	Nilai / Konfigurasi
Optimizer	Adam
Learning rate	0.0001
Weight decay	0.0001
Epoch	30
Batch size (AE)	32
Batch size (VAE)	16
Augmentation	Albumentations
Input image	224 × 224 × 3

PARAMETER PELATIHAN DISEBUT OPTIMAL KARENA DIPILIH PADA TITIK KESEIMBANGAN ANTARA NILAI YANG TERLALU BESAR DAN TERLALU KECIL, SEHINGGA MENGHASILKAN PROSES PELATIHAN YANG STABIL DAN EFEKTIF. SECARA EMPIRIS, KONFIGURASI INI MENUNJUKKAN PENURUNAN LOSS TRAINING DAN VALIDATION YANG KONSISTEN, TANPA DIVERGENSI MAUPUN OVERFITTING, SERTA MENGHASILKAN AKURASI DAN F1-SCORE TERTINGGI DIBANDINGKAN KONFIGURASI PARAMETER LAINNYA.

SECARA TEORETIS, PEMILIHAN PARAMETER TELAH DISESUAIKAN DENGAN KARAKTERISTIK ARSITEKTUR CNN–TRANSFORMER SERTA KOMPLEKSITAS AUTOENCODER (AE) DAN VARIATIONAL AUTOENCODER (VAE), DI MANA

OPTIMIZER ADAM, LEARNING RATE KECIL, DAN WEIGHT DECAY MEMBANTU MENJAGA STABILITAS PEMBARUAN BOBOT SERTA MENCEGAH MEMORISASI DATA. SELAIN ITU, PENGGUNAAN PARAMETER YANG SAMA PADA BERBAGAI DATASET (DFDC, FF++, TIMIT, DAN CELEB-DF) TETAP MEMBERIKAN PERFORMA YANG STABIL DAN KONSISTEN, SEHINGGA MENUNJUKKAN BAHWA KONFIGURASI INI BERSIFAT ROBUST DAN BUKAN HASIL KEBETULAN.

HTTPS://WWW.KAGGLE.COM/CODE/CLONESANG/DEEPFAKE-VIDEO-DETECTION-MAJOR

DATASET YANG DIAMBIL PADA SITUS INI TERDIRI DARI 1000 VIDEO ASLI YANG DIAMBIL DARI YOUTUBE DAN KEMUDIAN DIMANIPULASI MENGGUNAKAN EMPAT METODE MANIPULASI WAJAH OTOMATIS SEPERTI DEEPFAKES, FACE2FACE, FACESWAP, DAN NEURALTEXTURES. DATASET INI DISEDIAKAN DALAM BEBERAPA TINGKAT KOMPRESI DI ANTARANYA C23 (COMPRESSION MEDIUM) DAN C40 (COMPRESSION HIGH), SERTA BERBAGAI RESOLUSI VIDEO

HTTPS://WWW.KAGGLE.COM/CODE/KERNELER/STARTER-DFDC-FRAME-150-6C547762-E/OUTPUT

DATASET DFDC BERISI LEBIH DARI 100.000 VIDEO RESOLUSI TINGGI YANG MENCAKUP VIDEO ASLI DAN DEEPFAKE, MELIBATKAN 3.426 PARTISIPAN. VIDEO DIREKAM DALAM BERAGAM KONDISI PENGAMBILAN DAN DIMANIPULASI MENGGUNAKAN DELAPAN METODE DEEPFAKE, SEHINGGA MEMILIKI TINGKAT VARIASI YANG TINGGI. KOMPOSISI DATA DIDOMINASI OLEH VIDEO MANIPULASI DENGAN RASIO SEKITAR 6:1 DIBANDINGKAN VIDEO ASLI

3.4.Use casepenggunaan system

Aplikasi deteksi deepfake dirancang untuk pengguna umum dan diuji berdasarkan alur interaksi yang realistis. Hubungan antara pengguna dan sistem digambarkan melalui Diagram Kasus Penggunaan sebagaimana terlihat pada Gambar 3.4

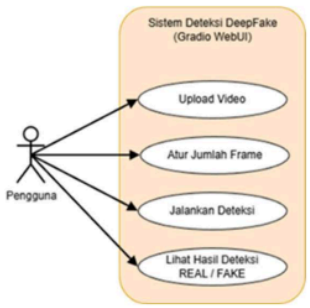


DIAGRAM 3.USECASE PENGGUNAAN SYSTEM

Gambar 3 mengilustrasikan Use Case Diagram untuk aplikasi deteksi deepfake yang menunjukkan hubungan antara pengguna dan sistem dalam keseluruhan proses penggunaan aplikasi.

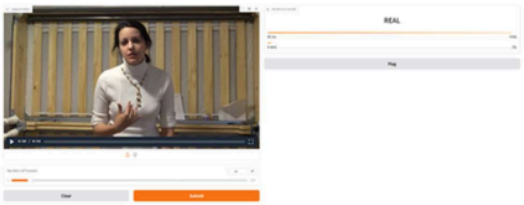
3.5.Senario Pengujian Melalui Antarmuka Gradio

Pengujian sistem dilakukan melalui aplikasi web berbasis Gradio yang memungkinkan pengguna mengunggah video, menjalankan proses deteksi, dan melihat hasil prediksi secara langsung. Gradio dipilih karena kemampuannya dalam menyajikan pipeline deep learning secara intuitif serta dapat dijalankan baik secara lokal maupun melalui layanan hosting. Proses pengujian meliputi inisialisasi model pre-trained, pengunggahan video oleh pengguna, pengaturan jumlah frame yang dianalisis, serta eksekusi deteksi yang mencakup ekstraksi frame, deteksi wajah, praproses data, inferensi model berbasis AE dan VAE, hingga perhitungan skor probabilitas akhir.

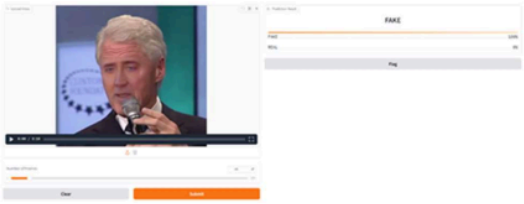
Hasil deteksi ditampilkan dalam bentuk label real atau fake beserta nilai confidence score, sehingga pengguna dapat memperoleh hasil klasifikasi secara jelas dan informatif.

3.6.Hasil Pengujian pada antarmuka gradio

Pengujian dilakukan menggunakan sejumlah video real dan deepfake untuk memvalidasi performa model dalam lingkungan aplikasi. Contoh tampilan antarmuka pengujian ditampilkan pada Gambar 4.5 dan Gambar 3.6



Gambar 4. Hasil Proses Deteksi Video Asli (REAL)



Gambar 5. Tampilan Hasil Deteksi Video Manipulasi (FAKE)

3.7.Evaluasi Metrik F1-Score

Selain pengujian melalui antarmuka aplikasi, performa model juga dievaluasi menggunakan metrik F1-Score yang mengombinasikan precision dan recall. Evaluasi ini dilakukan pada model ensemble (AE + VAE) serta masing-masing jaringan internal untuk mengukur kemampuan klasifikasi pada kedua kelas

Tabel 3.7 Nilai F1-Score Berdasarkan Dataset

Dataset	Model (AE + VAE)	AE	VAE
DFDC	99.1	98.4	98.4
FF++	95.5	94.9	96.8
TIMIT	98.3	97.5	97.8
Celeb-DF (v2)	91.6	95.2	89.0

Tabel 4.7 menyajikan nilai F1-Score pada beberapa dataset untuk model ensemble (AE + VAE) serta masing-masing model AE dan VAE. Hasil ini menunjukkan bahwa pendekatan ensemble secara umum memberikan performa yang lebih stabil dan unggul dalam mendeteksi deepfake pada berbagai dataset. untuk dataset dfdc (99,1),ff++(95,5),Timit(98,3),Celeb DF v2(91,6) mereka dirata-rata menjadi 96,125

V. PENUTUP DAN REKOMENDASI

Penelitian ini mampu merancang dan mengimplementasikan deteksi deepfake web menggunakan pendekatan arsitektur hybrid yang mengombinasikan ConvNeXt dan Vision Transformer (ViT), serta diperkuat dengan dua jalur rekonstruksi menggunakan Autoencoder (AE) dan Variational Autoencoder (VAE). Pendekatan ini memungkinkan model mengekstraksi fitur lokal secara efektif sekaligus memahami konteks global melalui mekanisme self-attention, sehingga mampu mendeteksi artefak manipulasi visual yang bersifat halus.

Berdasarkan hasil pengujian pada empat dataset publik, yaitu DFDC, FF++, DeepfakeTIMIT, dan Celeb-DF v2, model ensemble (AE + VAE) menunjukkan performa yang stabil dan unggul dibandingkan masing-masing jaringan secara terpisah. Model mencapai nilai F1-Score sebesar 99,1% pada DFDC, 95,5% pada FF++, 98,3% pada TIMIT, dan 91,6% pada Celeb-DF v2, dengan rata-rata keseluruhan sebesar 96,125%. Hasil yang diperoleh menunjukkan bahwa pendekatan hybrid yang diusulkan menunjukkan kemampuan generalisasi yang baik terhadap berbagai variasi teknik manipulasi deepfake

PERNYATAAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih kepada pihak Universitas Muhammadiyah Sidoarjo karena yang telah menyediakan sarana dan prasarana yang diperlukan, khususnya laboratoriumkomputer, yang memfasilitasi pelaksanaan upaya penelitian ini. Apresiasi juga diberikan kepada administrasi kampus atas dukungan administrasi dan bantuan mereka yang tak ternilai selama proses penelitian.

REFERENSI

[1] R. A. Prawiratama, 'Design of a Generative AI Image Similarity Test Application and Handmade Images Using Deep Learning Methods', Telematika, vol. 20, no. 3, p. 326, Nov. 2023, doi: 10.31315/telematika.v20i3.10096.
[2] M. A. I. H. Khusna and S. Pangestuti, 'DEEPFAKE, TANTANGAN BARU UNTUK NETIZEN (DEEPFAKE, A NEW CHALLENGE FOR NETIZEN)', PROMEDIA (PUBLIC RELATION DAN MEDIA KOMUNIKASI), vol. 5, Jun. 2019, doi: 10.52447/promedia.v5i2.2300.
[3] Y. Arif Fernandes and Y. Fatma, 'METODE DEEP LEARNING DALAM TEKNOLOGI DEEPFAKE : SYSTEMATIC LITERATURE REVIEW', JATI (Jurnal Mahasiswa Teknik Informatika), vol. 9, no. 2, pp. 3403–3410, Apr. 2025, doi: 10.36040/jati.v9i2.12987.
[4] I. Leliana, G. Irhamdhika, A. Haikal, R. Septian, and E. Kusnadi, 'ETIKA DALAM ERA DEEPFAKE: BAGAIMANA MENJAGA INTEGRITAS KOMUNIKASI', Jurnal Visi Komunikasi, vol. 22, no. 02, p. 234, Jan. 2024, doi: 10.22441/visikom.v22i02.24229.
[5] D. Putra, S. Sania, and A. Mitrin, 'Pengaruh Deepfake terhadap Kredibilitas Media Tradisional: Tantangan dan Implikasi di Era Digital', Sagara Komunika, vol. 1, pp. 13–18, 2024, doi: 10.25311/sagara/Vol1.Iss1.2022.
[6] Y. Hao, L. Dong, F. Wei, and K. Xu, 'Self-Attention Attribution: Interpreting Information Interactions Inside Transformer', Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 14, pp. 12963–12971, May 2021, doi: 10.1609/aaai.v35i14.17533.
[7] J. Feng, H. Tan, W. Li, and M. Xie, 'Conv2NeXt: Reconsidering Conv NeXt Network Design for Image Recognition', in 2022 International Conference on Computers and Artificial Intelligence Technologies (CAIT), IEEE, Nov. 2022, pp. 53–60. doi: 10.1109/CAIT56099.2022.10072172.
[8] K. Han et al., 'A Survey on Vision Transformer', IEEE Trans Pattern Anal Mach Intell, vol. 45, no. 1, pp. 87–110, Jan. 2023, doi: 10.1109/TPAMI.2022.3152247.
[9] T. Raharjo et al., 'ANALISIS FORENSIK DEEPFAKE BERBASIS CONVOLUTIONAL NEURAL NETWORK (CNN) UNTUK DETEKSI INKONSISTENSI TEKSTUR DAN POLA PADA CITRA WAJAH', JATI (Jurnal Mahasiswa Teknik Informatika), vol. 9, no. 2, pp. 2731–2738, Mar. 2025, doi: 10.36040/jati.v9i2.13058.
[10] J. Mu, M. Adrezo, and A. N. Haikal, 'Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode Convolutional Neural Network', Teknika, vol. 13, no. 1, pp. 45–50, Jan. 2024, doi: 10.34148/teknika.v13i1.705.
[11] Wawan Kurniawan, A. Kurniasih, and Muhamad Abdul Ghani, 'Real or Deepfake Face Detection in Images and Video Data using YOLO11 Algorithm', Journal of Artificial Intelligence and Engineering Applications (JAIEA), vol. 4, no. 2, pp. 1514–1521, Feb. 2025, doi: 10.59934/jaiea.v4i2.939.
[12] M. I. Abidin, I. Nurtanio, and A. Achmad, 'Deepfake Detection in Videos Using Long Short-Term Memory and CNN ResNext', ILKOM Jurnal Ilmiah, vol. 14, no. 3, pp. 178–185, Dec. 2022, doi: 10.33096/ilkom.v14i3.1254.178-185.

[13] C. P. Prasetia, 'JITE (Journal of Informatics and Telecommunication Engineering) Efficient Real and Fake Face detection Using ResNet18', JITE, vol. 4, no. 2, 2025, doi: 10.31289/jite.v9i1.15128.
[14] M. Patrick, C. Lubis,) Agus, and B. Dharmawan, 'Jurnal Ilmu Komputer dan Sistem Informasi PENDETEKSIAN CITRA DEEPFAKE WAJAH DI SMARTPHONE MENGGUNAKAN MOBILENETV3-SMALL DAN LBP'.
[15] C. Qin, L. Chen, Z. Cai, M. Liu, and L. Jin, 'Long short-term memory with activation on gradient', Neural Networks, vol. 164, pp. 135–145, Jul. 2023, doi: 10.1016/j.neunet.2023.04.026.
[16] X. Liu, L. Hu, L. Tie, L. Jun, X. Wang, and X. Liu, 'Integration of Convolutional Neural Network and Vision Transformer for gesture recognition using sEMG', Biomed Signal Process Control, vol. 98, p. 106686, Dec. 2024, doi: 10.1016/j.bspc.2024.106686.
[17] Y. Liu et al., 'Generative artificial intelligence and its applications in materials science: Current situation and future perspectives', Journal of Materiomics, vol. 9, no. 4, pp. 798–816, Jul. 2023, doi: 10.1016/j.jmat.2023.05.001.
[18] Y. Li, N. Miao, L. Ma, F. Shuang, and X. Huang, 'Transformer for object detection: Review and benchmark', Eng Appl Artif Intell, vol. 126, p. 107021, Nov. 2023, doi: 10.1016/j.engappai.2023.107021.
[19] L. Gao, J. Zhang, C. Yang, and Y. Zhou, 'Cas-VSwin transformer: A variant swin transformer for surface-defect detection', Comput Ind, vol. 140, p. 103689, Sep. 2022, doi: 10.1016/j.compind.2022.103689.

Conflict of Interest Statement:
The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Karya Tulis Ilmiah Mahasiswa UMSIDA.docx

ORIGINALITY REPORT

20%
SIMILARITY INDEX

PRIMARY SOURCES

1	archive.umsida.ac.id Internet	602 words — 17%
2	www.mdpi.com Internet	39 words — 1%
3	pels.umsida.ac.id Internet	19 words — 1%
4	cmsdata.iucn.org Internet	14 words — < 1%
5	dcase.community Internet	12 words — < 1%
6	berbura.bangka.go.id Internet	10 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF