

IDENTIFIKASI KONTEN VISUAL BUATAN AI DENGAN RESNET DAN FINE-GRAINED FEATURE EXTRACTION

Dzikri Maulana^{1*}, Nur Salamah Azzahrah², Reiza Alithian Syach³, Sopian Syaury⁴, Malik Syafi'i⁵, Chaerur Rozikin⁶

^{1,2,3,4,5,6} Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Indonesia; Telp. 081218669229

Keywords:

ResNet-50;
Fine-grained feature
extraction;
AI-generated image.

Correspondent Email:

dzikrimaulana1781945@gmail.com



Copyright © JITET (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstrak. Perkembangan pesat kecerdasan buatan (AI), khususnya dalam generasi citra digital, menghasilkan gambar yang semakin realistis dan sulit dibedakan dari gambar asli. Kondisi ini berisiko menimbulkan disinformasi, manipulasi opini publik, dan penurunan kepercayaan masyarakat. Penelitian ini mengusulkan sistem deteksi gambar palsu menggunakan arsitektur ResNet-50 yang diperkuat dengan teknik ekstraksi fitur granular melalui integrasi Squeeze-and-Excitation (SE) Block. Dataset yang digunakan berasal dari Kaggle, berisi citra asli dan buatan AI, dengan tahapan pra-proses berupa resize, normalisasi, dan augmentasi. Hasil eksperimen menunjukkan bahwa model ResNet-50 yang dimodifikasi mencapai akurasi validasi 96,7% dengan penurunan loss yang konsisten, menunjukkan proses optimisasi yang stabil dan kemampuan generalisasi yang baik tanpa indikasi overfitting. Model terbukti mampu membedakan citra asli dan buatan AI secara efektif, meskipun masih terdapat tantangan pada kasus borderline ketika gambar sintesis tampil sangat realistis. Penelitian ini menegaskan bahwa ekstraksi fitur granular mampu meningkatkan sensitivitas model terhadap detail tekstur halus, sehingga dapat menjadi pendekatan handal untuk forensik digital dan verifikasi konten visual.

Abstract. The rapid development of artificial intelligence (AI), particularly in image generation, has produced highly realistic visuals that are increasingly difficult to distinguish from authentic images. This poses risks of misinformation, public opinion manipulation, and erosion of trust. This study proposes a fake image detection system using the ResNet-50 architecture enhanced with fine-grained feature extraction through the integration of Squeeze-and-Excitation (SE) Blocks. The dataset consists of AI-generated and real images from Kaggle, preprocessed with resizing, normalization, and augmentation techniques. Experimental results demonstrate that the modified ResNet-50 achieved a validation accuracy of 96,7% and consistent loss reduction, indicating stable optimization and strong generalization without overfitting. The model successfully distinguished real and AI-generated images, although challenges remain in borderline cases where synthetic images are extremely realistic. These findings highlight the effectiveness of fine-grained feature extraction in improving sensitivity to subtle visual details, offering a reliable approach for digital forensics and content verification.

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence/AI) dalam bidang pengolahan citra telah memicu lahirnya fenomena baru yang dikenal sebagai *deepfake*. Teknologi ini memungkinkan manipulasi konten visual secara otomatis menggunakan metode pembelajaran mendalam untuk menghasilkan citra atau video sintesis yang sangat realistis [1]. Deepfake menggabungkan teknik *deep learning* dengan model generatif seperti *autoencoder* dan *Generative Adversarial Networks* (GAN) untuk melakukan *face swapping*, *lip-sync*, dan manipulasi ekspresi wajah (*puppet master*) dengan presisi tinggi. Seiring kemajuan perangkat keras dan ketersediaan dataset besar, proses pembuatan deepfake kini dapat dilakukan dengan mudah, bahkan oleh pengguna dengan kemampuan teknis terbatas. Kondisi ini memunculkan tantangan serius terhadap keamanan informasi, privasi individu, dan kepercayaan publik terhadap media digital [2].

Teknologi deteksi deepfake berkembang seiring peningkatan kualitas teknik generatif. Pendekatan awal mengandalkan metode forensik multimedia seperti analisis *Photo Response Non-Uniformity* (PRNU) dan artefak kompresi JPEG, namun metode ini terbukti kurang tangguh terhadap teknik manipulasi modern [2]. Dalam konteks Indonesia, penelitian awal berfokus pada pemanfaatan arsitektur *Convolutional Neural Network* (CNN) standar untuk mendeteksi manipulasi pada citra statis. Misalnya, CNN diterapkan untuk mengidentifikasi ciri visual deepfake melalui tahapan pelatihan konvolusional dan klasifikasi biner [3]. Selanjutnya, kombinasi CNN dengan Capsule Network dikaji untuk meningkatkan kemampuan representasi fitur spasial dan struktur lokal citra sintesis. Pendekatan ini menunjukkan potensi dalam meningkatkan akurasi deteksi terhadap variasi teknik deepfake [4].

Berbagai penelitian internasional telah mengembangkan arsitektur CNN yang lebih kompleks untuk mendeteksi deepfake secara lebih andal. Studi komparatif terhadap arsitektur VGG19, ResNet50, DenseNet121, dan MobileNetV2 menunjukkan bahwa ResNet50 memiliki performa kompetitif dalam mendeteksi gambar deepfake, berkat kemampuannya dalam mengekstraksi fitur

visual pada berbagai tingkat kedalaman [5]. Penelitian lainnya mengevaluasi efektivitas pendekatan pembelajaran mendalam dengan memanfaatkan dataset besar seperti CelebA-HQ dan citra hasil model GAN, yang menunjukkan bahwa ResNet50 mencapai akurasi tertinggi (99,43%) dibandingkan arsitektur lainnya dalam klasifikasi wajah palsu dan asli [6]. Selain itu, studi terkait klasifikasi citra medis menunjukkan efektivitas arsitektur ResNet50 yang dioptimasi menggunakan algoritma Adam dan Stochastic Gradient Descent (SGD), yang mampu mencapai akurasi hingga 95% dalam klasifikasi kanker mulut [7]. Temuan ini menunjukkan bahwa ResNet tidak hanya unggul dalam mendeteksi citra sintesis, tetapi juga sangat efektif dalam mengidentifikasi fitur lokal yang kompleks.

Meskipun demikian, terdapat kesenjangan penelitian yang cukup jelas. Sebagian besar pendekatan deteksi deepfake masih berfokus pada ekstraksi fitur global, sementara fitur halus (*fine-grained features*) yang sering kali menjadi pembeda penting antara citra asli dan buatan belum dimanfaatkan secara optimal. Selain itu, banyak penelitian yang menggunakan arsitektur CNN tanpa proses *fine-tuning* mendalam, sehingga performa model menjadi kurang adaptif terhadap variasi teknik generatif yang terus berkembang [6]. Dalam konteks ini, pendekatan *fine-tuning* terhadap model residual network seperti ResNet152 menjadi sangat penting. Penelitian terbaru menunjukkan bahwa penerapan strategi *warm-up learning rate scheduler*, augmentasi citra, dan akumulasi gradien pada proses *fine-tuning* dapat menghasilkan performa klasifikasi citra buatan AI dengan skor F1 mencapai 0,963 [8]. Hal ini membuktikan bahwa penggabungan arsitektur ResNet dengan ekstraksi fitur tingkat lanjut dapat meningkatkan ketangguhan deteksi terhadap konten sintesis.

Berdasarkan uraian tersebut, penelitian ini berfokus pada **identifikasi konten visual buatan AI menggunakan arsitektur ResNet yang dikombinasikan dengan teknik fine-grained feature extraction**. Penelitian ini bertujuan untuk menjawab keterbatasan pendekatan sebelumnya yang masih menekankan fitur global dan pelatihan standar, dengan menghadirkan metode yang lebih adaptif dan akurat dalam menghadapi beragam teknik generatif modern. Selain memberikan

kontribusi teoretis terhadap pengembangan teknik deteksi deepfake, penelitian ini juga berpotensi mendukung keamanan informasi dan validasi konten digital di era media berbasis AI.

2. TINJAUAN PUSTAKA

Tinjauan pustaka ini menyajikan teori dan konsep dasar yang digunakan sebagai landasan dalam penelitian, mencakup Arsitektur Residual Network (ResNet), *fine-grained feature extraction* dan metodologi proyek CRISP-DM.

2.1 Arsitektur ResNet

Residual Network (ResNet) adalah arsitektur *deep learning* yang digunakan sebagai model inti dalam penelitian ini. ResNet dirancang untuk mengatasi masalah degradasi akurasi dan vanishing gradient pada jaringan saraf yang sangat dalam [9]. Solusi kuncinya adalah penggunaan residual learning melalui penambahan shortcut connections yang memungkinkan informasi melewati beberapa lapisan tanpa perubahan. Pendekatan ini memfasilitasi pelatihan jaringan yang sangat dalam secara efektif [10]. Studi telah menunjukkan bahwa ResNet unggul dalam ekstraksi fitur dan klasifikasi gambar yang presisi, menjadikannya pilihan yang kuat untuk tugas computer vision [9].

2.2 Fine-Grained Feature Extraction

Fine-grained feature extraction merupakan pendekatan pengolahan citra yang menekankan penangkapan detail visual halus untuk membedakan objek dengan perbedaan antar kelas yang sangat kecil. Berbeda dari ekstraksi fitur global yang hanya menyoroti pola bentuk atau warna umum, teknik ini berfokus pada tekstur mikro, distribusi noise, dan pola lokal yang sering tak terlihat secara kasat mata [11].

Dalam klasifikasi citra, metode ini umum digunakan pada tugas dengan tingkat kemiripan tinggi seperti pengenalan spesies burung atau jenis kendaraan serupa [12]. Strategi yang sering diterapkan meliputi *part localization*, *multi-scale feature fusion*, dan *attention mechanism* untuk mengarahkan fokus model pada area paling diskriminatif [13]. Salah satu bentuk efektifnya ialah *Squeeze-and-Excitation (SE) Block*, yang menyesuaikan bobot saluran fitur agar jaringan lebih responsif terhadap

detail penting, sehingga meningkatkan efektivitas ekstraksi fitur halus pada arsitektur seperti ResNet [14].

2.3 Metodologi CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah model proses standar dan terbuka yang digunakan untuk memandu proyek data. Model ini dipilih sebagai kerangka kerja penelitian karena relevansinya yang berkelanjutan dalam praktik industri. CRISP-DM terdiri dari enam fase utama yang bersifat iteratif:

1. *Business Understanding* (Memahami tujuan dan kebutuhan bisnis).
2. *Data Understanding* (Mengumpulkan dan memahami karakteristik data).
3. *Data Preparation* (Mempersiapkan data untuk pemodelan, termasuk pembersihan dan transformasi).
4. *Modeling* (Menerapkan teknik dan membangun model).
5. *Evaluation* (Menguji kinerja model dan validasi).
6. *Deployment* (Mengimplementasikan hasil model).

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan rancangan eksperimen, serta mengadopsi kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai panduan dalam pengembangan. Fokus penelitian diarahkan pada pengembangan dan evaluasi model *Deep Learning* untuk klasifikasi citra digital, dengan tujuan utama membedakan antara citra asli dan citra hasil generasi kecerdasan buatan.

3.1 Business Understanding

Tujuan penelitian ini adalah mengembangkan sistem deteksi citra buatan AI dengan memanfaatkan pendekatan kuantitatif dan rancangan eksperimen. Sistem yang dibangun berfokus pada perbedaan citra asli dan citra hasil generasi AI secara otomatis dengan akurasi tinggi, menggunakan model *Deep Learning* ResNet yang diperkuat dengan teknik *fine-grained feature extraction*.

3.2 Data Understanding

Dataset yang digunakan berasal dari AI-Generated Images vs Real Images [15] yang

tersedia secara publik melalui Kaggle. Dataset terdiri dari 79.950 citra digital yang terbagi secara seimbang ke dalam dua kelas: real (0): 39.975 citra (50%) dan fake/AI-Generated (1): 39.975 citra (50%). Format citra bervariasi, meliputi JPEG, PNG, dan WEBP, dengan resolusi serta mode warna yang beragam. Untuk menjaga reproduktibilitas, dataset dibagi secara acak menggunakan seed tetap (seed=42) menjadi dua subset, yaitu data latih sebesar 80% atau 63.960 citra dan data uji sebesar 20% atau 15.990 citra.

3.3 Data Preparation

Pra-pemrosesan dilakukan secara berbeda untuk data latih dan data uji. Pada data uji, citra diubah ukurannya menjadi 224×224 piksel dan dilakukan normalisasi menggunakan nilai rata-rata serta deviasi standar dari ImageNet, yaitu mean sebesar (0.485, 0.456, 0.406) dan standar deviasi sebesar (0.229, 0.224, 0.225). Sementara itu, pada data latih dilakukan langkah yang sama ditambah dengan teknik augmentasi untuk meningkatkan kemampuan generalisasi model. Augmentasi yang diterapkan meliputi pemotongan acak pada ukuran 224×224 piksel, pembalikan horizontal dengan probabilitas 0,5, penyesuaian kecerahan, kontras, saturasi, serta hue dengan probabilitas 0,7, dan penyesuaian acak terhadap kecerahan serta kontras dengan probabilitas 0,5.

3.4 Modeling

Model yang digunakan adalah ResNet-50 tanpa bobot *pretrained* yang dimodifikasi menjadi Fine-Grained ResNet melalui penambahan Squeeze-and-Excitation (SE) Block setelah lapisan keempat (layer4). ResNet-50 berfungsi sebagai backbone dalam mengekstraksi fitur spasial, sedangkan SE Block berperan sebagai mekanisme *channel attention* yang memperkuat fitur penting dalam membedakan citra asli dengan citra buatan AI. Bagian akhir model berupa classifier head yang terdiri atas lapisan linear berukuran 2048 menuju 512, fungsi aktivasi ReLU, lapisan dropout dengan probabilitas 0,3, dan lapisan linear terakhir berukuran 512 menuju 2 kelas. Model dilatih menggunakan framework PyTorch dengan dukungan Multi-GPU melalui DataParallel dan Mixed Precision Training untuk meningkatkan efisiensi komputasi.

Proses pelatihan menggunakan fungsi kerugian CrossEntropyLoss yang dipadukan dengan label smoothing sebesar 0,05. Optimizer yang digunakan adalah AdamW dengan nilai learning rate sebesar 1×10^{-4} dan weight decay sebesar 1×10^{-5} , sementara pengaturan laju pembelajaran mengikuti CosineAnnealingLR dengan nilai Tmax sebesar 15. Ukuran batch yang digunakan adalah 128 dengan jumlah maksimum epoch sebanyak 15. Regulasi dilakukan melalui penerapan dropout, label smoothing, dan weight decay. Selain itu, teknik early stopping juga digunakan untuk mencegah overfitting, dengan menghentikan pelatihan jika akurasi validasi tidak meningkat selama lima epoch berturut-turut, serta menyimpan bobot terbaik berdasarkan akurasi validasi tertinggi.

3.5 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan pada data uji yang tidak digunakan dalam pelatihan. Beberapa metrik digunakan dalam tahap ini, yaitu akurasi sebagai ukuran utama keberhasilan prediksi, matriks kebingungan untuk memvisualisasikan distribusi hasil klasifikasi, serta laporan klasifikasi yang mencakup presisi, recall, dan F1-score untuk masing-masing kelas.

3.6 Deployment

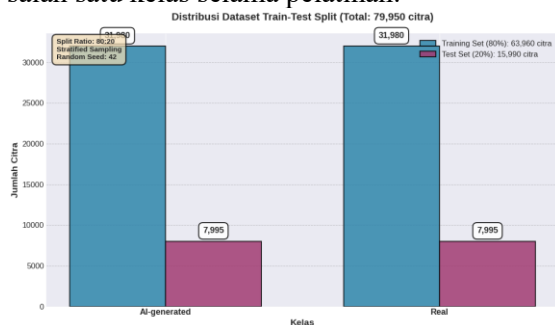
Model yang telah dilatih dan dievaluasi kemudian disiapkan untuk digunakan melalui antarmuka sederhana berbasis Jupyter Notebook. Tahap deployment dilakukan dengan menyimpan bobot model dalam format .pth menggunakan PyTorch, kemudian memanggil kembali bobot tersebut pada tahap *inference* dengan *pipeline* yang sama persis dengan pelatihan model.

Pendekatan ini dipilih karena lebih sesuai untuk kebutuhan penelitian dan eksplorasi awal, sehingga memudahkan proses debugging, analisis visual, dan dokumentasi hasil.

5. HASIL DAN PEMBAHASAN

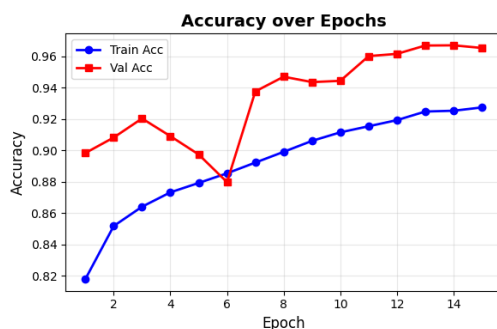
Penelitian ini menghasilkan sistem klasifikasi citra berbasis *Fine-Grained ResNet-50* yang dilengkapi dengan *Squeeze-and-Excitation (SE) Block* untuk meningkatkan kemampuan model dalam menangkap perbedaan fitur halus antara citra asli dan citra buatan AI. Dataset yang digunakan adalah *AI vs. Human-Generated*

Images dari Kaggle oleh Alessandra Sala dkk., yang terdiri atas 79.950 citra data gambar. Dataset dibagi secara acak menjadi data latih sebesar 80% dan data uji sebesar 20% dengan seed tetap agar hasil dapat direproduksi. Gambar 1 memperlihatkan distribusi kelas pada kedua subset tersebut, yang menunjukkan keseimbangan jumlah data antara kelas *Real* dan *AI-generated*, masing-masing sekitar 50%. Keseimbangan ini memastikan bahwa model tidak memiliki kecenderungan bias terhadap salah satu kelas selama pelatihan.



Gambar 1. Distribusi Kelas pada Set Pelatihan & Pengujian

Proses pelatihan model dilakukan selama 15 epoch menggunakan konfigurasi optimizer AdamW, fungsi kerugian CrossEntropyLoss dengan label smoothing 0,05, dan learning rate scheduler CosineAnnealingLR. Hasil pelatihan menunjukkan peningkatan performa yang konsisten seiring bertambahnya epoch. Akurasi data latih meningkat dari 81,8% pada epoch pertama menjadi 92,7% pada epoch ke-15, sedangkan nilai loss menurun stabil dari 0,44 menjadi 0,26. Akurasi validasi juga mengalami peningkatan signifikan, dari 89,8% menjadi 96,7% pada akhir pelatihan. Gambar 2 menunjukkan tren akurasi dan loss selama proses pelatihan dan validasi yang menandakan model tidak mengalami overfitting, karena kedua kurva bergerak searah dengan selisih yang kecil.



Gambar 2. Metrik Akurasi dan Loss pada Training & Validation

Evaluasi dilakukan pada data uji yang terdiri atas 15.990 citra. Model mencapai **akurasi keseluruhan sebesar 96,7%**, menandakan kemampuan generalisasi yang sangat baik. *Confusion matrix* pada Gambar 3 memperlihatkan bahwa model mampu mengklasifikasikan kedua kelas dengan tingkat kesalahan yang rendah. Dari 8.124 citra asli, sebanyak 7.779 citra diklasifikasikan dengan benar, sedangkan 345 citra salah dikenali sebagai citra AI. Sebaliknya, dari 7.866 citra AI, sebanyak 7.684 citra terklasifikasi dengan benar, sedangkan 182 citra salah dikenali sebagai citra asli.

Confusion Matrix:

```
[[7779  345]
 [ 182 7684]]
```

Gambar 3. *Confusion Matrix* pada Data Uji

Hasil *classification report* pada Gambar 4 memperkuat temuan ini. Model memperoleh precision sebesar 0,98 untuk kelas real dan 0,96 untuk kelas AI-generated, dengan recall masing-masing 0,96 dan 0,98. Nilai F1-score untuk kedua kelas berada pada 0,97, menandakan keseimbangan yang baik antara kemampuan deteksi dan ketepatan prediksi. Nilai-nilai ini menggambarkan bahwa model tidak hanya mampu mengenali citra dengan benar, tetapi juga jarang melakukan kesalahan prediksi yang fatal.

Classification Report:				
	precision	recall	f1-score	support
Real	0.98	0.96	0.97	8124
Fake	0.96	0.98	0.97	7866
accuracy			0.97	15990
macro avg	0.97	0.97	0.97	15990
weighted avg	0.97	0.97	0.97	15990

Gambar 4. *Classification Report*

Secara visual, performa model ini juga menunjukkan konsistensi selama proses pelatihan, di mana penurunan loss yang stabil mengindikasikan konvergensi parameter yang efektif. Hasil ini membuktikan bahwa penerapan *fine-grained feature extraction* dan *Squeeze-and-Excitation Block* pada ResNet-50 memberikan dampak positif terhadap peningkatan kemampuan diskriminatif model dalam mendeteksi ciri-ciri halus pada citra buatan AI.

Analisis kesalahan menunjukkan bahwa sebagian besar kesalahan klasifikasi terjadi pada citra buatan AI dengan tingkat realisme tinggi. Model masih menghadapi tantangan dalam membedakan citra buatan AI yang memiliki pencahayaan alami, tekstur halus, dan komposisi realistis menyerupai hasil kamera profesional. Hal ini menandakan bahwa tantangan utama terletak pada kemampuan model untuk mengenali pola mikroskopik dan anomali visual yang halus, terutama pada hasil generatif dengan kualitas tinggi.

Untuk menguji kemampuan generalisasi di luar domain data pelatihan, model diuji menggunakan 16 sampel citra eksternal yang tidak termasuk dalam dataset Kaggle, sebagaimana ditampilkan pada Tabel 1. Sampel terdiri atas delapan citra asli (real images) dan delapan citra buatan AI (AI-generated images) yang diambil dari berbagai sumber publik daring.

Tabel 1. Tabel hasil Pengujian dengan gambar dari sumber eksternal

Gambar	Label	Prediksi	Probabilitas
	Asli	Asli	0.91

	Asli	Asli	0.92
	Asli	Asli	0.98
	Asli	Asli	0.91
	Asli	Asli	0.72
	Asli	Asli	0.84
	Asli	Asli	0.95
	Asli	Asli	0.94
	AI	AI	0.62
	AI	AI	0.97
	AI	AI	0.99
	AI	AI	0.67
	AI	AI	0.98
	AI	Asli	0.95
	AI	Asli	0.93
	AI	Asli	0.97

Dari total 16 sampel eksternal, sebanyak 13 citra diklasifikasikan dengan benar dan 3 citra mengalami kesalahan klasifikasi, seluruhnya merupakan false negative, yakni citra buatan AI yang diprediksi sebagai citra asli.

Menariknya, meskipun terjadi kesalahan klasifikasi, **tingkat probabilitas prediksi model tetap tinggi** (berkisar antara 0,90–0,95) pada kasus *false negative*. Hal ini menunjukkan bahwa model memiliki keyakinan kuat terhadap keputusannya, tetapi masih kesulitan mengenali pola visual yang sangat realistis pada citra hasil generasi AI. Citra-citra yang salah diklasifikasikan umumnya memiliki pencahayaan lembut, detail tekstur halus, dan komposisi alami.

Secara keseluruhan, hasil pengujian eksternal ini memperlihatkan bahwa model telah memiliki kemampuan deteksi yang kuat dan stabil, namun tetap menghadapi tantangan pada *borderline cases*, yakni ketika citra buatan AI memiliki tingkat realisme yang hampir sempurna. Menariknya, perkembangan teknologi generatif saat ini membuat AI mampu menghasilkan citra dengan pencahayaan alami, tekstur kulit yang realistis, dan detail visual yang menyerupai hasil tangkapan kamera profesional.

Berbeda dengan model generatif generasi awal yang sering menampilkan artefak visual seperti tepi buram, tekstur tidak konsisten, atau pola simetris berulang, model modern seperti Midjourney, DALL·E, dan Stable Diffusion kini telah mampu meniru distribusi pencahayaan dan kedalaman spasial yang sangat alami. Dengan demikian, indikator konvensional seperti ketidakwajaran tekstur atau pencahayaan buatan tidak lagi menjadi ciri khas yang dapat diandalkan untuk mendeteksi konten buatan AI.

Fenomena ini menunjukkan bahwa pendekatan deteksi berbasis fitur visual murni mulai menghadapi keterbatasan ketika berhadapan dengan citra hasil model generatif canggih yang menggunakan diffusion-based learning atau attention-driven rendering. Oleh karena itu, strategi deteksi di masa depan perlu diarahkan pada kombinasi analisis fitur halus dan analisis distribusi statistik tingkat tinggi seperti frekuensi spektral atau artefak domain latar belakang, yang tidak dapat sepenuhnya ditiru oleh generator AI.

Hasil ini sekaligus menegaskan bahwa model *Fine-Grained ResNet-50* dengan SE Block masih memiliki potensi besar, namun perlu diperluas dengan mekanisme adaptif yang mampu menangkap anomali generatif pada level spasial dan kontekstual yang lebih dalam. Dengan pengujian lanjutan menggunakan hasil generatif dari Midjourney, DALL·E, dan Stable Diffusion, diharapkan model dapat dikembangkan menjadi sistem deteksi lintas domain yang tetap tangguh terhadap kemajuan teknologi AI generatif yang semakin realistis.

5. KESIMPULAN

Berdasarkan hasil penelitian, pengujian, dan analisis yang telah dilakukan terhadap model *Fine-Grained ResNet-50* dengan integrasi *Squeeze-and-Excitation (SE) Block*, dapat disimpulkan beberapa hal penting yang mencakup hasil utama, keunggulan pendekatan yang diusulkan, keterbatasan yang dihadapi, serta arah pengembangan penelitian di masa mendatang sebagai berikut:

- Penelitian ini berhasil mengembangkan sistem deteksi citra buatan AI menggunakan arsitektur *Fine-Grained ResNet-50* yang diperkaya dengan integrasi *Squeeze-and-Excitation (SE) Block*. Model mencapai akurasi validasi sebesar **96,7%** dan loss akhir sebesar **0,26**, menandakan proses optimisasi yang stabil dan kemampuan generalisasi yang baik tanpa indikasi *overfitting*.
- Pendekatan *fine-grained feature extraction* terbukti meningkatkan sensitivitas model terhadap tekstur mikro dan detail halus yang sering menjadi pembeda antara citra asli dan citra buatan AI. Integrasi *SE Block* membantu jaringan beradaptasi dengan fitur paling informatif di tiap saluran, sehingga memperkuat ketepatan klasifikasi. Selain itu, hasil pengujian eksternal menunjukkan bahwa model tetap mampu mempertahankan akurasi yang cukup baik terhadap data dari luar domain pelatihan.
- Meskipun performa tinggi, model masih menghadapi kesulitan dalam mengenali *borderline cases*, yaitu citra buatan AI dengan tingkat realisme sangat tinggi. Hal ini terjadi karena perkembangan teknologi generatif seperti Midjourney, DALL·E, dan Stable Diffusion telah memungkinkan AI

menghasilkan citra dengan pencahayaan alami, tekstur kulit realistis, dan kedalaman spasial menyerupai hasil kamera profesional. Kondisi tersebut menyebabkan fitur visual konvensional tidak lagi cukup untuk membedakan citra buatan dan asli.

- d. Kemungkinan pengembangan penelitian selanjutnya difokuskan pada peningkatan kemampuan model dalam menghadapi variasi data yang lebih luas dan kompleks. Langkah pertama adalah melakukan pengujian lintas domain (cross-domain testing) menggunakan citra dan video yang dihasilkan oleh berbagai model generatif modern seperti Midjourney, DALL·E, dan Stable Diffusion. Pengujian ini bertujuan untuk menilai ketahanan (robustness) model terhadap perbedaan distribusi data visual yang dihasilkan oleh generator dengan karakteristik estetika dan struktur tekstur yang beragam.

Selain itu, penelitian ke depan dapat diarahkan pada eksplorasi arsitektur jaringan lain, seperti DenseNet, EfficientNet, maupun pendekatan ensemble hybrid yang mengombinasikan keunggulan beberapa model untuk memperoleh kinerja yang lebih optimal dalam hal efisiensi komputasi dan akurasi deteksi. Pengembangan berikutnya juga dapat difokuskan pada peningkatan kemampuan real-time melalui penerapan teknik model compression, quantization, dan optimasi GPU, sehingga sistem dapat diimplementasikan secara langsung pada perangkat forensik digital atau sistem pemantauan konten daring dengan latensi yang minimal.

Upaya pengembangan juga mencakup perluasan dataset pelatihan dengan menambahkan video sintesis dan data multimodal, seperti kombinasi citra, teks, dan audio, agar model memiliki kemampuan generalisasi yang lebih baik terhadap berbagai format konten visual buatan AI. Selain itu, penting untuk melakukan analisis terhadap serangan adversarial dan manipulasi digital guna menguji ketahanan model terhadap upaya pengelabuan sistem deteksi. Dengan pengembangan-pengembangan tersebut, diharapkan sistem deteksi yang dihasilkan

tidak hanya unggul dalam akurasi, tetapi juga tangguh, adaptif, dan siap diterapkan dalam konteks dunia nyata yang dinamis.

- e. Penelitian ini memberikan kontribusi penting bagi bidang *digital forensics* dan *content authenticity verification*. Namun, penerapan teknologi ini harus tetap memperhatikan aspek etika, privasi, dan transparansi agar hasil deteksi tidak disalahgunakan dan tetap mendukung keamanan informasi secara bertanggung jawab.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis menyampaikan terima kasih kepada dosen pembimbing dan seluruh pengajar di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang, yang telah memberikan arahan, bimbingan, dan dukungan selama proses perkuliahan dan penyusunan penelitian ini.

Ucapan terima kasih juga penulis sampaikan kepada rekan-rekan mahasiswa, tim riset, serta semua pihak yang telah memberikan bantuan, masukan, dan semangat dalam setiap tahap penelitian, mulai dari pengumpulan data, pelatihan model, hingga penyusunan laporan. Penulis juga berterima kasih kepada penyedia dataset serta komunitas open-source yang menyediakan sumber daya penting dalam mendukung pelaksanaan penelitian ini. Semoga karya ini dapat memberikan kontribusi yang bermanfaat bagi pengembangan ilmu pengetahuan dan penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] A. A. Abu-Ein, O. M. Al-Hazaimah, A. M. Dawood, and A. I. Swidan, "Analysis of the current state of deepfake techniques-creation and detection methods," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 3, pp. 1659–1667, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1659-1667.
- [2] W. J. Hadi, S. M. Kadhém, and A. R. Abbas, "A survey of deepfakes in terms of deep learning and multimedia forensics," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 4, pp. 4408–4414, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4408-4414.

- [3] S. A. Gulo, A. A. Pertiwi, S. Putri, S. Nasution, and H. Syahputra, "DETEKSI DEEPFAKE DALAM CITRA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN)," 2025.
- [4] D. Suharyanto, C. Lubis,) Agus, and B. Dharmawan, "Jurnal Ilmu Komputer dan Sistem Informasi PENERAPAN CONVOLUTIONAL NEURAL NETWORK DAN CAPSULE NETWORKS DALAM MENDETEKSI DEEPFAKE."
- [5] H. Kusniyati, "Komparasi Performa VGG19, ResNet50, DenseNet121 dan MobileNetV2 Dalam Mendeteksi Gambar Deepfake Performance Comparison of VGG19, ResNet50, DenseNet121 and MobileNetV2 in Detecting Deepfake Image," 2024. [Online]. Available: www.jurnal.unimed.ac.id
- [6] M. Berrahal, M. Boukabous, M. Yandouzi, M. Grari, and I. Idrissi, "Investigating the effectiveness of deep learning approaches for deep fake detection," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3853–3860, Dec. 2023, doi: 10.11591/eei.v12i6.6221.
- [7] M. Fauzan Novriandy, B. Rahmat, and A. Junaidi, "KLASIFIKASI CITRA PENYAKIT KANKER MULUT MENGGUNAKAN ARSITEKTUR RESNET50 OPTIMASI ADAM DAN SGD," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4732.
- [8] S. A. Hakim *et al.*, "KLASIFIKASI CITRA GENERASI ARTIFICIAL INTELLIGENCE MENGGUNAKAN METODE FINE TUNING PADA RESIDUAL NETWORK AI GENERATED IMAGE CLASSIFICATION USING FINE TUNING ON RESIDUAL NETWORK," vol. 11, no. 3, pp. 655–666, 2024, doi: 10.25126/jtiik.2024118118.
- [9] X. Jin, "Analysis of Residual Block in the ResNet for Image Classification," *INSTICC*, Jun. 2024, pp. 253–257. doi: 10.5220/0012800400003885.
- [10] Y. Wang, "Research on Image Classification Based on ResNet," 2023, pp. 971–979. doi: 10.2991/978-94-6463-300-9_98.
- [11] T. Zhou *et al.*, "Fine-grained image recognition method for digital media based on feature enhancement strategy," *Neural Comput Appl*, vol. 36, no. 5, pp. 2323–2335, Feb. 2024, doi: 10.1007/s00521-023-08968-1.
- [12] J. Fu, H. Zheng, and T. Mei, "Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 4476–4484. doi: 10.1109/CVPR.2017.476.
- [13] J. Ren, C. Li, Y. An, W. Zhang, and C. Sun, "Few-Shot Fine-Grained Image Classification: A Comprehensive Review," *AI*, vol. 5, no. 1, pp. 405–425, Mar. 2024, doi: 10.3390/ai5010020.
- [14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745.
- [15] Alessandra Sala, Manuela Jeyaraj, Toma Ijatom, and Margarita Pitsiani, "AI vs. Human-Generated Images," <https://www.kaggle.com/datasets/alessandrasala79/ai-vs-human-generated-dataset>. Accessed: Oct. 07, 2025. [Online]. Available: <https://www.kaggle.com/datasets/alessandrasala79/ai-vs-human-generated-dataset>