

KLASIFIKASI CITRA GENERASI ARTIFICIAL INTELLIGENCE MENGGUNAKAN METODE FINE TUNING PADA RESIDUAL NETWORK

Sulthan Abiyyu Hakim^{*1}, Muhammad Ubaidillah², Aditya Rizky Ramadhan³, Rahmada Zulvia Azzahra Hawari⁴, Audhinata Bebytama Rizky⁵, Raniyah Lutfi⁶, Putri Tsania Maulidia Hermanto⁷, Novanto Yudistira⁸

^{1,2,3,4,5,6,7,8}Universitas Brawijaya, Malang

Email: ¹sabiyyuhakim@student.ub.ac.id, ²devvevan@student.ub.ac.id, ³adityarizky1020@student.ub.ac.id, ⁴rahmadazulvia@student.ub.ac.id, ⁵audhinata@student.ub.ac.id, ⁶raniyahlutfi3@student.ub.ac.id, ⁷putritsania999@student.ub.ac.id, ⁸yudistira@ub.ac.id

*Penulis Korespondensi

(Naskah masuk: 04 Juni 2024, diterima untuk diterbitkan: 14 Juni 2024)

Abstrak

Citra generasi AI memiliki beragam manfaat yang signifikan, baik dalam bidang penelitian maupun industri. Namun, penggunaan citra generasi AI juga memiliki dampak negatif dalam konteks hukum, politik dan berbagai aspek lain dalam kehidupan. Penelitian ini menitik beratkan klasifikasi citra generasi AI yang dapat mendeteksi keaslian dari suatu citra. Metode yang diusulkan adalah menggunakan model *residual network* yang telah dilakukan *fine tuning*. Teknik *fine tuning* yang dilakukan meliputi penggunaan *learning rate scheduler* berbasis *warm up* yang diikuti dengan *linear scheduler*, akumulasi gradien, dan augmentasi citra. Penelitian menunjukkan bahwa model *residual network* 152 menghasilkan performa terbaik dengan *f1 score* 0.963 dan *loss* 0.08.

Kata kunci: klasifikasi citra, generasi AI, fine tuning, residual network

AI GENERATED IMAGE CLASSIFICATION USING FINE TUNING ON RESIDUAL NETWORK

Abstract

AI-generated images have various significant benefits, both in the realm of research and industry. However, the use of AI-generated images also has negative impacts in legal, political, and various other aspects of life. This research focuses on the classification of AI-generated images that can detect the authenticity of an image. The proposed method involves using a fine-tuned residual network model. The fine-tuning techniques applied include the use of a warm-up-based learning rate scheduler that followed by linear scheduler, gradient accumulation, and image augmentation. The research demonstrates that the Residual Network 152 model achieves the best performance with a *f1 score* of 0.963 and a *loss* of 0.08.

Keywords: image classification, AI generated, fine tuning, residual network

1. PENDAHULUAN

Artificial Intelligence (AI) atau Kecerdasan Buatan dapat diartikan sebagai kemampuan suatu sistem untuk mereplika kerangka berpikir, menyelesaikan masalah, memahami konsep, dan belajar dengan efektif (Morandín-Ahuerma, 2022). Kemampuan AI sangat luar biasa, karena mampu memproses dan menganalisis data dalam jumlah besar dengan cepat serta belajar dari pengalaman-pengalaman sebelumnya. Hal ini memungkinkan AI untuk melakukan tugas-tugas seperti pengenalan wajah, penerjemahan bahasa, pengambilan keputusan, dan bahkan menciptakan karya seni atau musik secara otomatis. Kemajuan dalam AI telah membawa dampak signifikan di berbagai industri,

termasuk kesehatan, otomotif, dan finansial, dan terus menjadi salah satu bidang riset yang paling menarik dan berpotensi mengubah paradigma dalam teknologi komputer.

Model *Generative Adversarial Network* (GAN), telah memberikan kontribusi yang luar biasa dalam dunia komputasi cerdas. GAN memiliki kemampuan untuk menghasilkan gambar-gambar yang belum pernah ada sebelumnya, termasuk wajah-wajah yang realistis (Goodfellow et al., 2014). Selain itu, GAN juga digunakan dalam mentransfer gaya gambar (Chen et al., 2019). Tak hanya itu, GAN juga terbukti berguna dalam meningkatkan resolusi gambar melalui teknologi *Super Resolution* (Ledig et al., 2017).

Namun, perkembangan dalam bidang generasi gambar tidak berhenti di sini. Model *diffusion* (Ho et al., 2020) juga muncul dan menunjukkan kemampuan generasi gambarnya yang begitu luar biasa. Penelitian ini menunjukkan bahwa model *diffusion* mampu melakukan generasi gambar dengan hasil yang lebih superior daripada GAN (Dhariwal & Nichol, 2021).

Hasil citra generasi AI memiliki manfaat yang signifikan dalam berbagai bidang penelitian dan aplikasi. Salah satu manfaatnya adalah augmentasi data citra (Perez & Wang, 2017). Dengan menggunakan citra yang dibuat oleh AI, peneliti dapat memperluas kumpulan data mereka secara substansial, meningkatkan keragaman data, dan memperbaiki kualitas data pelatihan untuk model prediksi lainnya. Selain itu, citra generasi AI juga digunakan sebagai alat penyuntingan (*editing tool*) yang efektif (Wu et al., 2017) yang memungkinkan pengguna untuk melakukan penyuntingan gambar dengan lebih mudah dan efisien. Selain itu, citra generasi AI juga diterapkan dalam dunia seni, di mana seniman menggunakan teknologi ini untuk menciptakan karya seni yang inovatif dan unik (Chen et al., 2020). Dengan demikian, citra yang dihasilkan oleh AI memiliki beragam potensi manfaat yang signifikan di berbagai bidang.

Dalam era perkembangan teknologi informasi dan komunikasi yang pesat, citra generasi AI telah menjadi perangkat yang sangat berpengaruh dalam berbagai aspek kehidupan manusia. Sebagaimana yang dicatat oleh (Westerlund, 2019), AI telah menjadi alat yang digunakan oleh berbagai pelaku, termasuk aktivis politik, kriminal, bahkan individu biasa. Namun, kehadiran AI tidak selalu memberikan manfaat yang positif. Penggunaan citra generasi AI dalam konteks hukum dapat merusak kepercayaan pada bukti dan merugikan nilai probatifnya (Maras & Alexandrou, 2019). Lebih jauh lagi, penelitian oleh (Zannettou et al., 2019) menunjukkan bahwa citra generasi AI dapat digunakan untuk menyebarkan informasi palsu, terutama dalam konteks politik. Hal ini dapat memiliki dampak serius, terutama dalam peristiwa penting seperti pemilihan umum, di mana pandangan terhadap aktivis politik dapat berubah secara signifikan akibat penyebaran informasi yang tidak benar. Oleh karena itu, penting bagi kita untuk memahami implikasi dari penggunaan AI dalam berbagai bidang dan mengembangkan strategi yang tepat untuk mengatasi tantangan yang timbul dari perkembangan teknologi ini.

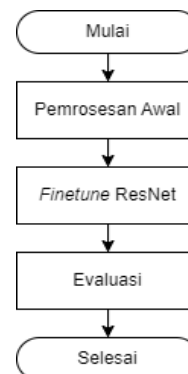
Fokus utama penelitian ini adalah masalah performa *fine tuning Residual Network* dalam mengatasi permasalahan klasifikasi citra generasi AI. Dari permasalahan tersebut, penelitian ini memiliki tujuan untuk menganalisis dan mengevaluasi performa metode *fine tuning Residual Network* dalam mengatasi tantangan klasifikasi citra generasi AI. Metode usulan untuk menghadapi permasalahan yang telah diidentifikasi, peneliti menawarkan pengembangan sebuah metode klasifikasi citra

generasi AI yang memanfaatkan arsitektur *Convolutional Neural Network* (CNN) yang telah dilakukan *pretrained* sebelumnya, yaitu model *Residual Network* atau ResNet (He dkk., 2015). ResNet merupakan salah satu jenis CNN yang sangat dalam dengan banyak lapisan atau blok residual. Pemilihan ResNet didasarkan pada penelitian sebelumnya yang telah mengungkapkan keunggulan arsitektur ini dalam mengatasi masalah klasifikasi citra kompleks.

Selain itu, manfaat yang dapat diperoleh dari penelitian ini sangat beragam. Bagi industri teknologi dan AI, implementasi *fine tuning* pada model ResNet memiliki potensi untuk meningkatkan akurasi sistem klasifikasi citra, memberikan kesempatan bagi industri teknologi untuk menghasilkan produk dan layanan yang lebih canggih dan efisien. Selanjutnya, bagi akademisi dan peneliti, hasil dari penelitian ini akan membuka peluang untuk penelitian lanjutan di bidang pengenalan citra generasi AI, di mana pendekatan *fine tuning* yang digunakan dapat dijadikan sebagai dasar untuk eksperimen atau pengembangan model yang lebih kompleks. Sementara itu, bagi masyarakat luas, kemajuan teknologi klasifikasi citra generasi AI memiliki dampak positif pada berbagai sektor, seperti kesehatan, keamanan, dan industri. Dengan adanya inovasi dalam pengenalan citra, berbagai solusi lebih aman, efisien, dan inovatif dapat dihadirkan. Contohnya dalam konteks politik, teknologi ini memiliki potensi besar untuk meningkatkan kemampuan mendeteksi citra generasi AI yang digunakan untuk menyebarkan kebencian.

2. METODE PENELITIAN

Penelitian ini dilaksanakan dengan pendekatan metodologi yang sistematis untuk mengembangkan dan mengevaluasi model arsitektur *Residual Network*. Diagram alir penelitian diperlihatkan pada Gambar 1.



Gambar 1: Diagram Alir Penelitian

2.1 Dataset

Penelitian dimulai dari mengunduh *dataset* citra. *Dataset* klasifikasi citra generasi AI terdiri dari dua kelas yang berbeda, yaitu gambar asli/*real* yang berarti gambar merupakan hasil tangkapan gambar dari kamera dan gambar palsu/*fake* yang berarti

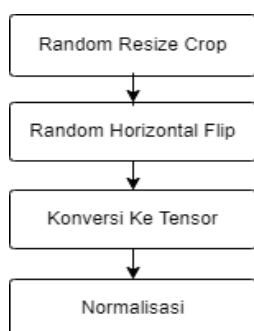
gambar tersebut merupakan citra generasi AI. *Dataset* yang digunakan dalam penelitian ini bernama CIFAKE (Bird & Lotfi, 2023) dapat diakses melalui tautan [berikut](#).



Gambar 2: Sampel Dataset

Dataset latih dan uji berjumlah 100000 dan 20000 secara berurutan. Pada data latih, terdapat 50000 gambar untuk kelas *real* dan 50000 gambar untuk kelas *fake*. Kemudian pada data uji, terdapat 10000 gambar untuk kelas *real* dan 10000 gambar untuk kelas *fake*. Gambar 2, menunjukkan sampel gambar untuk setiap kelasnya. Pada kedua kelas, kategori gambar sama persis yaitu makhluk hidup dan kendaraan. Gambar dalam *dataset* memiliki ukuran yang kecil yaitu 32 x 32 *pixels* dengan resolusi vertikal dan horizontal sebesar 96 dpi.

2.2 Pemrosesan Awal



Gambar 3: Diagram Alir Pemrosesan Awal

Dalam tahap awal pengolahan, gambar yang diberikan dimasukkan ke dalam prosedur *Random Resize Crop*, yang berfungsi untuk mengubah ukuran dan memotong gambar secara acak. Setelah itu, gambar tersebut dapat mengalami proses *Random Horizontal Flip*, dimana gambar dapat dibalik secara horizontal. Proses ini bertujuan untuk meningkatkan variasi dalam *dataset* dengan harapan model akan mampu mengidentifikasi fitur penting dari gambar dalam berbagai skala dan orientasi.

Setelah gambar telah mengalami augmentasi, langkah selanjutnya adalah melakukan konversi gambar ke dalam bentuk *tensor*. Dalam konteks ini, *tensor* dapat diartikan sebagai *array* multi-dimensi yang digunakan oleh perangkat keras komputasi untuk memproses data dengan efisiensi. Tahap terakhir dalam proses pengolahan ini adalah melakukan normalisasi. Dalam proses ini, nilai *pixel* gambar disesuaikan sehingga memiliki rata-rata dan

simpangan baku yang sesuai dengan spesifikasi model *residual network*.

2.3 Finetune ResNet

Bobot *pretrained* dari model *residual network* didapati dari platform *Huggingface*. Jenis model *residual network* yang digunakan adalah ResNet50, ResNet101, dan ResNet152. Sedangkan untuk pembandingan, digunakan model EfficientNet, VGG16, dan CNN (Bird & Lotfi, 2023). *Hyperparameter* yang digunakan untuk melakukan *fine tuning* didefinisikan pada tabel 1.

Tabel 1. Konfigurasi Hyperparameter

Hyperparameter	nilai
Batch Size	32
Learning Rate	5×10^{-5}
Epoch	5
Gradient Accumulation	4
warm up step	10%

2.4 Evaluasi

Metrik evaluasi yang digunakan dalam penelitian ini adalah metrik klasifikasi seperti *F1 Score*, *Precision*, *Recall*, dan Akurasi. Meskipun terdapat empat metrik, tetapi metrik yang digunakan adalah *F1 Score* agar dapat dibandingkan dengan hasil dari penelitian sebelumnya (Bird & Lotfi, 2023). Evaluasi dilakukan setiap 600 *steps* dan pada setiap *step* kelipatan 1200 akan dilakukan penyimpanan model sebagai *checkpoint*. Evaluasi tidak setiap *step* latih selesai dikarenakan dapat memperlambat proses *fine tuning*.

Precision adalah rasio antara jumlah benar positif dengan jumlah keseluruhan yang diperkirakan sebagai positif. Dapat dihitung menggunakan persamaan (1).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall adalah rasio antara jumlah benar positif dengan jumlah keseluruhan yang sebenarnya adalah positif. Dapat dihitung menggunakan persamaan (2).

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

F1 Score adalah rata-rata harmonik dari *Precision* dan *Recall*. Ini adalah cara untuk mendapatkan satu ukuran yang menggabungkan kedua metrik tersebut. Dapat dihitung menggunakan persamaan (3).

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Akurasi adalah rasio dari jumlah prediksi yang benar dengan jumlah keseluruhan prediksi. Dapat dihitung menggunakan persamaan (4).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Dimana TP (*True Positives*), TN (*True Negatives*), FP (*False Positives*), dan FN (*False*

Negatives). TP mengacu pada jumlah kasus positif yang benar-benar terdeteksi dengan benar sebagai positif, yang mencerminkan ketepatan dalam mengidentifikasi kasus positif. Di sisi lain, TN mengukur jumlah kasus negatif yang juga benar-benar terdeteksi dengan benar sebagai negatif, mengindikasikan kemampuan sistem dalam mengenali situasi yang sebenarnya negatif. Namun, FP merupakan jumlah kasus negatif yang keliru diperkirakan sebagai positif, menggambarkan seberapa sering sistem memberikan peringatan palsu. Sedangkan FN adalah jumlah kasus positif yang keliru diperkirakan sebagai negatif, menunjukkan ketidakmampuan sistem dalam mendeteksi kasus positif yang sebenarnya.

2.5 Tools dan Perangkat Keras

Dalam penelitian ini, bahasa pemrograman Python dipilih sebagai instrumen utama mengingat kelengkapan pustakanya yang mendukung implementasi *deep learning*. Pustaka PyTorch, khususnya sub pustaka torchvision, diadopsi untuk mendukung aspek-aspek tertentu dari analisis dan pemrosesan data citra. Selain itu, untuk mendokumentasikan serta menganalisis metrik pelatihan dan evaluasi, pustaka *Weight and Bias* diintegrasikan ke dalam kerangka kerja penelitian.

Penelitian dijalankan menggunakan layanan komputasi awan dari Kaggle. Layanan ini menyediakan sebuah GPU NVIDIA P100 dengan VRAM sebesar 16GB. CPU yang digunakan adalah Xeon Processors @2.2Ghz dengan 1 core dan 2 threads. Terakhir, RAM yang digunakan berukuran sebesar 13GB.

3. KAJIAN PUSTAKA

Artificial Intelligence (AI) atau Kecerdasan Buatan adalah suatu sistem di mana kemampuan komputer mampu melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia (Morandín-Ahuerma, 2022). Hal ini mencakup kemampuan untuk memahami bahasa manusia, mengenali pola dalam data, membuat keputusan, dan bahkan belajar dari pengalaman. Dapat dilihat dari peran-peran tersebut bahwa AI memiliki potensi besar untuk mengatasi masalah yang semakin kompleks dan memungkinkan untuk menghasilkan dampak positif bagi masyarakat.

Salah satu arsitektur *deep neural network* buatan yang sangat sukses adalah *Residual Network* atau ResNet (He dkk., 2015) yang memiliki banyak varian berdasarkan kedalaman lapisannya, di antara nya adalah ResNet-18, ResNet-50, ResNet-101, dan ResNet-152. ResNet adalah arsitektur jaringan saraf yang mampu melatih jaringan yang sangat dalam dan mengatasi masalah pelatihan jaringan yang dangkal. Salah satu ciri utama dari arsitektur ResNet adalah upaya untuk mengatasi kendala yang muncul dalam masalah hilangnya gradien. Kendala ini muncul ketika kita bekerja dengan jaringan yang sangat

dalam karena penurunan nilai fungsi kerugian (*loss*) yang diperlukan untuk menyesuaikan bobot-bobot yang sesuai.

EfficientNet (Tan & Le, 2019) merupakan keluarga arsitektur jaringan saraf yang dioptimalkan untuk keseimbangan antara efisiensi komputasi dan akurasi. Pendekatan yang digunakan adalah dengan menyesuaikan skala jaringan saraf secara bersamaan dalam tiga dimensi: kedalaman (*depth*), lebar (*width*), dan resolusi gambar (*resolution*). Konsep scaling ini bertujuan untuk mengatasi masalah yang sering terjadi pada model-model yang hanya mengoptimalkan satu atau dua dimensi saja, yang bisa menyebabkan performa tidak optimal atau memerlukan komputasi yang sangat tinggi. EfficientNet telah menunjukkan performa yang unggul dalam berbagai tugas klasifikasi gambar dengan memanfaatkan pendekatan yang efisien dalam penggunaan sumber daya komputasi.

VGGNet (Simonyan & Zisserman, 2014) adalah arsitektur jaringan saraf dalam yang sangat terkenal dan digunakan secara luas. Model ini terkenal karena desainnya yang sederhana namun sangat efektif, dengan menggunakan lapisan-lapisan konvolusi yang lebih kecil (3x3) tetapi dalam jumlah yang lebih banyak. Meskipun VGGNet memiliki performa yang baik dalam klasifikasi gambar, model ini memerlukan sumber daya komputasi yang sangat besar, baik dalam hal memori maupun daya komputasi, karena jumlah parameternya yang sangat besar.

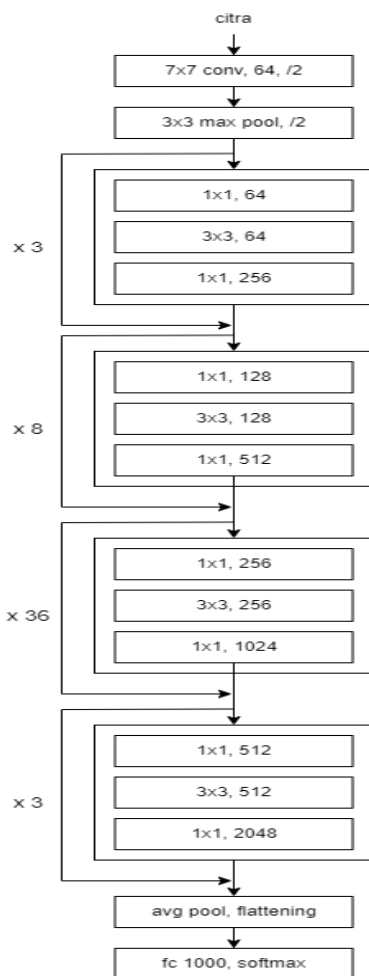
Inception Network, atau yang dikenal juga sebagai GoogLeNet (Szegedy dkk., 2015), adalah arsitektur jaringan saraf dalam yang memperkenalkan konsep "Inception Module". Modul ini memungkinkan penggunaan beberapa filter dengan ukuran yang berbeda dalam satu lapisan, yang kemudian digabungkan (*concatenate*) untuk menghasilkan output yang lebih kaya. Pendekatan ini memungkinkan jaringan untuk menangkap berbagai fitur dengan skala yang berbeda dalam satu tahap komputasi, sehingga meningkatkan efisiensi dan akurasi tanpa harus menambah kedalaman jaringan secara signifikan.

DenseNet (Huang dkk., 2017) atau Densely Connected Convolutional Networks adalah arsitektur jaringan saraf yang memperkenalkan konsep koneksi yang sangat padat antara lapisan-lapisan jaringan. Dalam DenseNet, setiap lapisan mendapatkan input dari semua lapisan sebelumnya dan memberikan output ke semua lapisan berikutnya. Pendekatan ini memungkinkan gradien mengalir dengan lebih baik dan mengurangi masalah hilangnya gradien yang sering terjadi pada jaringan yang sangat dalam. Selain itu, DenseNet dapat menggunakan jumlah parameter yang lebih sedikit dibandingkan dengan model-model lain yang memiliki kedalaman yang sama.

EfficientNet, seperti namanya, dirancang untuk efisiensi komputasi dengan memanfaatkan scaling yang optimal. Dalam konteks ini, model usulan harus dibandingkan dengan EfficientNet dalam hal

penggunaan sumber daya komputasi dan kecepatan inferensi. VGGNet dan DenseNet dikenal dengan kemampuan mereka dalam mencapai akurasi yang tinggi dalam klasifikasi gambar. Model usulan harus dievaluasi dalam hal akurasi dan kemampuan generalisasi dibandingkan dengan model-model ini. Inception Network (GoogLeNet) menawarkan kompleksitas yang lebih tinggi dalam menangkap fitur melalui Inception Module. Perbandingan dengan model ini akan menunjukkan seberapa baik model usulan dapat menangkap fitur yang kompleks dan beragam dalam data.

Dalam konteks pembelajaran mendalam (*deep learning*), perkalian berulang-ulang dari bobot-bobot ini dapat mengakibatkan gradien semakin kecil seiring dengan peningkatan jumlah lapisan, hingga akhirnya gradien ini menjadi sangat kecil dan hampir tidak berkontribusi. Selain itu, pelatihan jaringan saraf yang sangat dalam dengan banyak parameter membutuhkan penyesuaian *learning rate* yang signifikan. Oleh karena itu, ResNet mengusulkan konsep yang mengizinkan aliran langsung dari beberapa lapisan ke lapisan berikutnya, serta memanfaatkan hasil aktivasi dari lapisan sebelumnya sebagai salah satu caranya untuk mengatasi masalah ini.



Gambar 4: Arsitektur ResNet-152

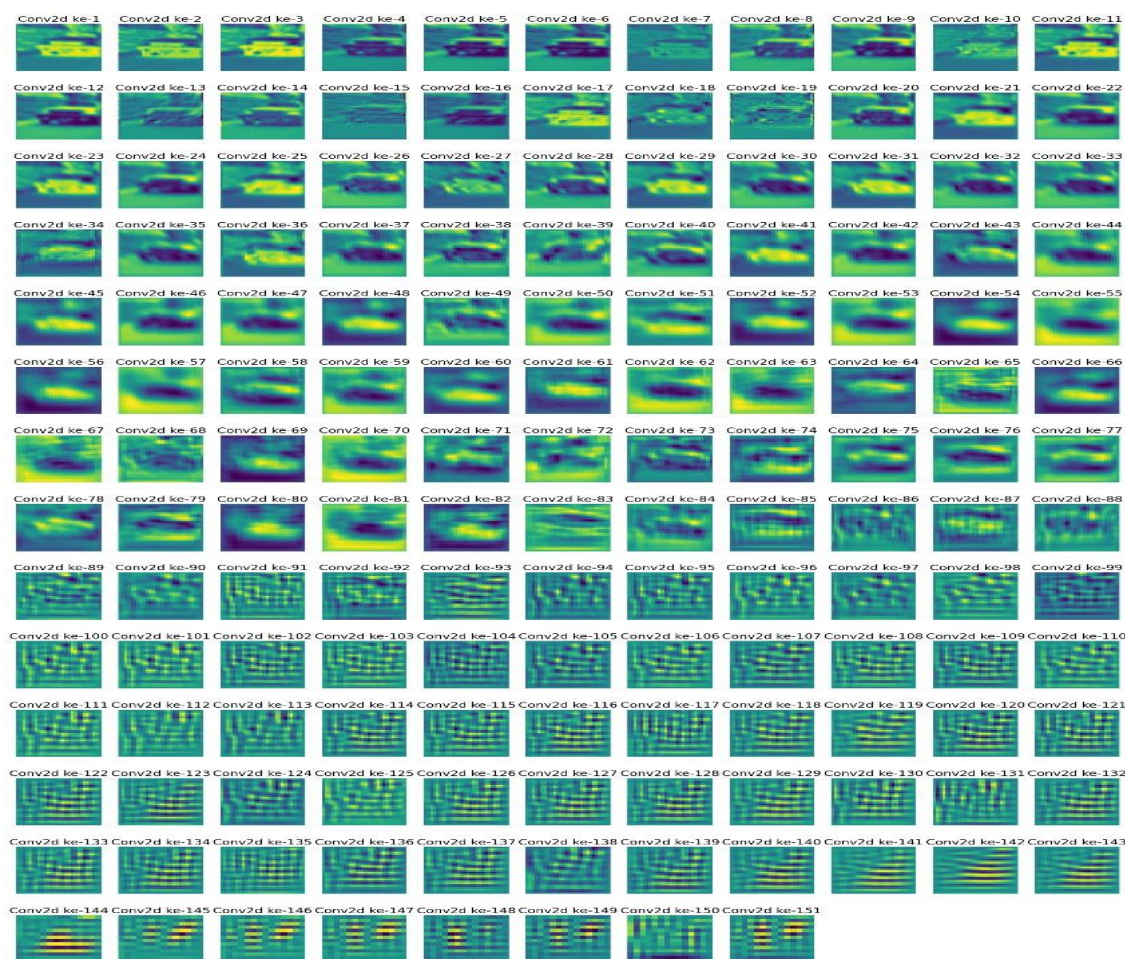
Gambar 4 adalah arsitektur dari ResNet-152 yang terdiri dari 152 lapisan yang kompleks, yang menjadikannya salah satu jaringan saraf konvolusional (CNN) terdalam yang umum digunakan. Dalam ResNet-152, ada dua jenis blok utama yang digunakan: blok konvolusi dan blok identitas. Blok konvolusi terdiri dari beberapa *layer* konvolusi berturut-turut yang digunakan untuk mengekstraksi fitur dari gambar. Di antara blok konvolusi, terdapat lapisan *pooling* yang digunakan untuk mengurangi dimensi gambar.

Pada awalnya, gambar input dengan ukuran 224x224x3 (panjang x lebar x kanal warna/RGB) diproses melalui lapisan konvolusi awal dengan filter ukuran 7x7, 64 filter, dan stride 2, menghasilkan feature map berukuran 112x112x64. Selanjutnya, lapisan pooling maksimal 3x3 dengan stride 2 mengurangi dimensi feature map menjadi 56x56x64. Kemudian, data ini melalui serangkaian blok residual yang terdiri dari blok konvolusi dan blok identitas.

Blok konvolusi residual pertama (Conv2_x) menggunakan kombinasi lapisan konvolusi 1x1, 3x3, dan 1x1. Proses ini menghasilkan feature map berukuran 56x56x256 setelah melewati 3 blok residual (1 blok konvolusi dan 2 blok identitas). Conv3_x block menurunkan ukuran feature map menjadi 28x28x512 dengan total 8 blok residual (1 blok konvolusi dan 7 blok identitas). Conv4_x block lebih lanjut mengurangi ukuran feature map menjadi 14x14x1024 melalui 36 blok residual (1 blok konvolusi dan 35 blok identitas). Terakhir, Conv5_x block mengurangi ukuran feature map menjadi 7x7x2048 dengan 3 blok residual (1 blok konvolusi dan 2 blok identitas).

Setiap konvolusi dalam ResNet-152 diikuti oleh batch normalization dan fungsi aktivasi ReLU. Batch normalization menormalkan input dari setiap lapisan, membantu menjaga distribusi nilai yang stabil selama pelatihan, sementara ReLU memperkenalkan non-linearitas dengan mengatur nilai negatif menjadi nol, yang membantu jaringan belajar representasi yang lebih kompleks. Setelah melalui semua blok residual, feature map 7x7x2048 diproses melalui lapisan global average pooling yang mengambil rata-rata dari setiap channel fitur, menghasilkan vektor fitur akhir berukuran 1x1x2048.

Vektor fitur ini kemudian masuk ke fully connected (FC) layer yang mengubahnya menjadi probabilitas kelas, menghasilkan prediksi akhir. Jumlah keluaran dari FC layer ini tergantung pada jumlah kelas yang dilatih, misalnya 1000 kelas untuk dataset ImageNet. Dengan kombinasi desain yang dalam dan shortcut connections, ResNet-152 mampu mengatasi masalah vanishing gradient, memungkinkan pelatihan jaringan yang sangat dalam dengan performa yang unggul dalam tugas-tugas klasifikasi gambar.



Gambar 5 Keluaran Feature Map masing-masing layer pada ResNe

Visualisasi feature map (Gambar 5) yang dihasilkan dalam ResNet dengan 5 lapisan memberikan wawasan tentang bagaimana jaringan mengekstraksi dan memproses informasi dari gambar input secara bertahap. Pada lapisan pertama, feature map menunjukkan deteksi tepi dan tekstur dasar dengan pola-pola sederhana seperti garis dan sudut. Setelah melalui lapisan pooling, dimensi feature map berkurang tetapi fitur yang paling signifikan tetap dipertahankan, menunjukkan bahwa jaringan mulai fokus pada area-area penting dari gambar. Saat gambar diproses melalui blok residual, visualisasi feature map menampilkan fitur yang lebih kompleks, dengan shortcut connections memungkinkan kombinasi fitur dari lapisan sebelumnya tanpa kehilangan informasi kritis. Pada lapisan konvolusi terakhir dari lima lapisan ini, feature map mengungkapkan representasi yang lebih abstrak dari gambar asli, di mana detail-detail halus telah digantikan oleh pola-pola yang lebih tinggi yang relevan untuk tugas klasifikasi. Warna-warna dan intensitas dalam visualisasi feature map ini menunjukkan aktivasi neuron yang berbeda, membantu mengidentifikasi bagian mana dari gambar yang memberikan kontribusi signifikan terhadap output akhir jaringan.

Kemudian, *residual connection* adalah blok identitas. Blok ini memungkinkan informasi untuk mengalir melalui jaringan tanpa gangguan yang berlebihan. Dalam blok identitas, *shortcut* (jalan pintas) diperkenalkan, yang memungkinkan aliran informasi melalui blok tanpa modifikasi berlebihan. Ini membantu mencegah masalah pelatihan yang dikenal sebagai *vanishing gradient*. Selain itu, ResNet-152 juga menggunakan *batch normalization* untuk mempercepat pelatihan dan mengurangi *overfitting*. Selain itu, aktivasi ReLU (*Rectified Linear Unit*) digunakan sebagai fungsi aktivasi untuk memperkenalkan non-linearitas ke dalam jaringan.

Fine Tuning merupakan teknik yang digunakan untuk melatih kembali model yang digunakan pada beberapa proses sebelumnya (Tang & de Sa, 2020). *Fine Tuning* digunakan untuk meningkatkan ragam kualitas model yang dihasilkan dengan melihat nilai akurasi model menggunakan semua *layer* ekstraksi fitur. Proses *fine tuning*, secara umum hanya pada lapisan *fully connected* saja, yaitu dengan menggantikan *output neuron* sesuai dengan jumlah kelas pada *dataset* yang digunakan. Apabila *dataset* yang digunakan berbeda dengan *dataset ImageNet* (*pre-trained*), maka dapat dilakukan proses *fine tuning* pada lapisan-lapisan yang berada diatas. Hal

ini bertujuan untuk menyesuaikan model *dataset* yang digunakan, agar fitur-fitur yang didapatkan bersifat spesifik terhadap *dataset*.

AdamW adalah sebuah algoritma optimasi yang memperoleh popularitas tinggi untuk pelatihan *deep learning* karena kemampuannya mencapai kinerja yang bagus dengan *hyperparameter* yang lebih sedikit (Zhuang et al., 2022). Ini improvisasi dari algoritma optimasi Adam, yang secara luas digunakan dalam pelatihan CNN, transformer, dan MLP (Wightman & Touvron, 2021). AdamW memisahkan gradien dari regularisasi dari aturan pembaruan Adam-2, menghasilkan konvergensi yang lebih baik dibandingkan dengan Adam (Zhuang et al., 2022). Salah satu keuntungan menggunakan AdamW adalah konvergensi yang lebih cepat dan kinerja superior dibandingkan dengan algoritma optimasi lain seperti Adam (Kim et al., 2023).

Learning rate scheduler dengan *warm up* adalah teknik yang secara bertahap meningkatkan *learning rate* dari nilai yang kecil ke *learning rate* awal yang diinginkan selama *epoch* atau step tertentu. Strategi ini sangat berguna ketika berhadapan dengan pelatihan *batch* besar atau saat menggunakan *learning rate* awal yang tinggi, karena membantu mengurangi ketidakstabilan numerik dan meningkatkan stabilitas proses pelatihan (Li et al., 2021). Dengan meningkatkan *learning rate* secara bertahap, strategi *warm up* memungkinkan model untuk menjelajahi ruang parameter lebih efektif dan menemukan solusi yang lebih baik. Telah terbukti efektif dalam berbagai domain, termasuk klasifikasi gambar, pemodelan bahasa, dan pelatihan jaringan saraf mendalam (He et al., 2018; Xiong, 2020; Li et al., 2021).

Penggunaan *linear learning rate scheduler* memiliki beberapa manfaat yang signifikan dalam pelatihan model *deep learning*. Ditemukan bahwa penggunaan *linear learning rate scheduler* dapat meningkatkan laju konvergensi dan performa akhir model (Gotmare et al., 2018). Dengan demikian, penggunaan *linear learning rate scheduler* dapat membantu meningkatkan efisiensi dan performa pelatihan model *deep learning*.

Gradient accumulation adalah teknik mengumpulkan gradien dari beberapa *batch* sebelum melakukan pembaruan bobot. Teknik ini memungkinkan penggunaan ukuran *batch* yang lebih besar tanpa mempengaruhi penggunaan memori yang berlebihan, sehingga dapat meningkatkan efisiensi pelatihan (Yan et al., 2021). Dengan demikian, penggunaan *gradient accumulation* dapat membantu meningkatkan stabilitas, efisiensi, dan performa pelatihan *deep neural network*.

4. HASIL DAN DISKUSI

Hasil yang didapatkan dari hasil eksperimentasi dengan alur yang sudah dijelaskan sebelumnya adalah sebagai berikut.

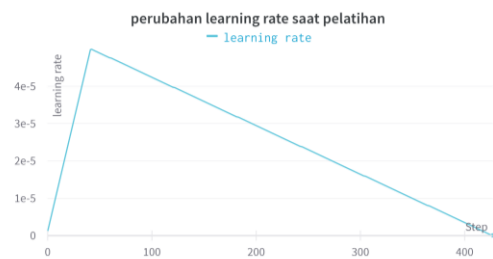
Tabel 2. Perbandingan Model

model	loss	F1 Score	Parameter (juta)
ResNet18	0.19	0.928	11.7
ResNet50	0.16	0.931	25.6
ResNet101	0.16	0.933	44.5
ResNet152	0.08	0.963	60.2
CNN (Bird, 2023)	0.17	0.936	-
VGG16	0.22	0.833	138.4
EfficientNetB4	0.17	0.941	19.3
EfficientNetB3	0.20	0.880	12.2

Tabel 3. Evaluasi Metrik ResNet152

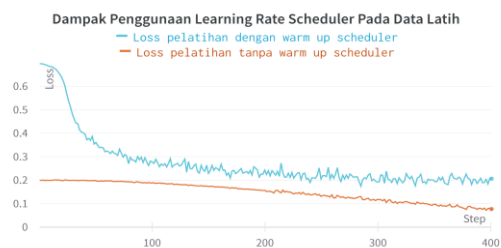
metrik	skor
F1 Score	0.9634
Recall	0.9522
Precision	0.9749
Akurasi	0.9632

Dari Tabel 2 menunjukkan perbandingan antara berbagai model dalam hal nilai *loss* dan *F1 Score*. Model-model yang dibandingkan antara lain ResNet50, ResNet101, ResNet152, CNN (Bird, 2023), VGG16, dan EfficientNet. Dari semua model yang dibandingkan, ResNet152 menunjukkan nilai *loss* terendah, yaitu sebesar 0.08 dan *F1 Score* tertinggi sebesar 0.963. Hal ini mengindikasikan ResNet152 memiliki kinerja yang paling baik di antara semua model yang dibandingkan. ResNet152, menunjukkan kinerja yang unggul karena arsitektur ResNet dirancang dengan *residual connections* yang memungkinkan untuk membuat jaringan yang sangat dalam tanpa permasalahan gradien yang hilang (He et al., 2018). Tentunya, semakin dalam jaringan maka semakin tinggi kemampuan dalam dapat menangkap pola-pola *non linear* pada suatu data.

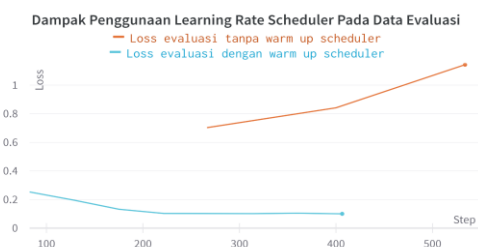


Gambar 6: Perubahan Nilai Learning Rate

Pada gambar 6, learning mengalami perubahan sepanjang waktu pelatihan. Dalam fase *warm up*, khususnya selama 10% data dari *total step*, *learning rate* bernilai 0 kemudian mencapai *learning rate* yang ditargetkan, yaitu $5e-5$. Setelah mencapai titik ini, *learning rate* mulai menurun secara linear hingga akhirnya mendekati nol. Ini menunjukkan suatu pendekatan adaptif dalam menyesuaikan *learning rate*, memastikan bahwa model tidak hanya cepat konvergen di awal, namun juga mampu mempertahankan stabilitas seiring dengan berjalannya waktu pelatihan.



Gambar 7: Dampak Learning Rate Scheduler Pada Loss Pelatihan



Gambar 8: Dampak Learning Rate Scheduler

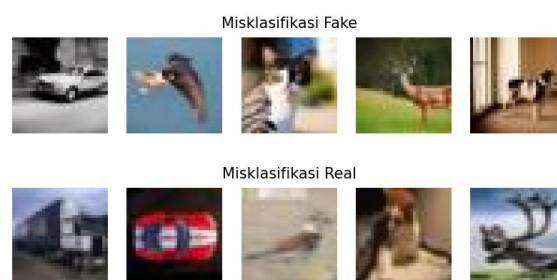
Gambar 7 dan 8 menunjukkan perbandingan performa dari ResNet152 yang diberi *learning rate scheduler* dan tidak. Jika dari awal *learning rate* diatur sebesar $5e-5$ maka pada *step* 0 *loss* pelatihan akan sangat rendah bahkan perlahan mendekati 0. Hasil *loss* pelatihan yang terlihat sangat bagus ini tidak berbanding lurus dengan *loss* pada data evaluasi yang ternyata semakin naik. Hal ini menunjukkan jika model ResNet152 mengalami *overfitting*. Sementara jika menggunakan *learning rate scheduler*, *loss* pada data pelatihan dan evaluasi berbanding lurus yang menunjukkan model tidak *overfitting* maupun *underfitting*. Perlu dicatat pada tahap evaluasi, terdapat perbedaan poin *step* evaluasi disebabkan karena perubahan kelipatan evaluasi pada proses pelatihan.

Sebenarnya	Prediksi	
	Positif	Negatif
Positif	48.44%	1.24%
Negatif	2.43%	47.89%

Gambar 9: Confusion Matrix

Gambar 9 menunjukkan bahwa *True Positive* (TP) lebih banyak daripada *True Negative* (TN). Artinya, model lebih sering benar mengidentifikasi kelas positif (*Fake*) dibandingkan dengan kelas negatif (*Real*). Model ResNet152 lebih cenderung atau lebih sensitif terhadap prediksi kelas positif. Ini juga bisa menunjukkan bahwa kelas positif mungkin lebih mudah diprediksi atau lebih dominan dalam *dataset*. Kemudian, *False Negative* (FN) lebih

banyak daripada *False Positive* (FP). Hal ini menunjukkan bahwa model ResNet152 lebih sering salah dalam mengidentifikasi kelas positif (menganggapnya sebagai negatif) daripada salah dalam mengidentifikasi kelas negatif (menganggapnya sebagai positif). Dalam kata lain, model cenderung lebih konservatif dalam memprediksi kelas positif dan mungkin kehilangan beberapa contoh positif yang sebenarnya. Hal ini bisa berdampak pada *recall* kelas positif, yang mungkin lebih rendah daripada presisi kelas positif. TP dan FN nya sangat tinggi menunjukkan bahwa model ResNet memiliki kecenderungan tinggi untuk memprediksi gambar sebagai kelas positif.



Gambar 10: Sampel Misklasifikasi

Gambar 10 menampilkan beberapa contoh sampel yang mengalami misklasifikasi oleh model yang mengalami misklasifikasi oleh model ResNet152. Bagian yang bertuliskan "Misklasifikasi *Fake*" menunjukkan gambar-gambar yang seharusnya diklasifikasikan sebagai *real* tetapi diprediksi sebagai *fake* oleh model. Sebaliknya, bagian "Misklasifikasi *Real*" menampilkan gambar-gambar yang seharusnya diklasifikasikan sebagai *fake* tetapi diprediksi sebagai *real*.

Salah satu tantangan yang dihadapi oleh model adalah adanya gambar dengan warna yang mencolok. Model ResNet152 juga cenderung terkecoh dengan gambar *real* yang memiliki tekstur tidak detail. Selain itu, terdapat kecenderungan model untuk memprediksi gambar sebagai *real* ketika gambar tersebut memiliki latar belakang yang polos. Dengan demikian, meskipun ResNet152 telah menghasilkan skor evaluasi yang cukup baik, namun masih ada ruang untuk peningkatan, terutama dalam mengatasi kasus-kasus misklasifikasi seperti yang ditunjukkan.

ResNet18, dengan 11.7 juta parameter, menunjukkan *loss* 0.19 dan F1 Score 0.928. Model ini, meskipun efisien dalam jumlah parameter, tidak memiliki kedalaman yang cukup untuk menangkap fitur yang sangat kompleks, yang menjelaskan performanya yang lebih rendah dibandingkan dengan varian ResNet yang lebih dalam. ResNet50, dengan 25.6 juta parameter, memperbaiki performa dengan *loss* 0.16 dan F1 Score 0.931. Ini menunjukkan bahwa menambah kedalaman jaringan dengan lebih banyak lapisan konvolusi meningkatkan kemampuan model untuk menangkap fitur yang lebih kompleks, meskipun tidak signifikan jika dibandingkan dengan ResNet18.

ResNet101, dengan 44.5 juta parameter, memperlihatkan hasil yang sedikit lebih baik dengan loss 0.16 dan F1 Score 0.933. Penambahan parameter dan kedalaman lebih lanjut meningkatkan performa, tetapi margin peningkatannya mulai menurun. Ini menunjukkan bahwa ada titik di mana menambah kedalaman jaringan tidak lagi memberikan peningkatan performa yang signifikan tanpa strategi pelatihan yang tepat dan arsitektur yang lebih efisien.

ResNet152, dengan 60.2 juta parameter, menunjukkan performa terbaik dengan loss 0.08 dan F1 Score 0.963. Kedalaman tambahan memungkinkan model untuk menangkap fitur yang lebih kompleks dan abstrak dari data. Penggunaan shortcut connections atau identitas blok dalam ResNet152 memungkinkan gradien mengalir lebih bebas melalui jaringan, mengurangi masalah vanishing gradient. Selain itu, batch normalization dan fungsi aktivasi ReLU setelah setiap konvolusi menstabilkan distribusi nilai input dan memperkenalkan non-linearitas, yang membantu jaringan belajar representasi yang lebih mendalam dan kompleks.

VGG16, dengan 138.4 juta parameter, menunjukkan performa yang jauh lebih rendah dengan loss 0.22 dan F1 Score 0.833. Meskipun memiliki jumlah parameter yang sangat besar, efisiensi model dalam memanfaatkan parameter ini jauh lebih rendah dibandingkan ResNet152. Hal ini karena VGG16 tidak menggunakan shortcut connections, yang membuatnya lebih rentan terhadap masalah vanishing gradient dalam jaringan yang sangat dalam.

EfficientNetB3, dengan 12.2 juta parameter, memiliki loss 0.20 dan F1 Score 0.880. EfficientNetB4, dengan 19.3 juta parameter, menunjukkan performa yang lebih baik dengan loss 0.17 dan F1 Score 0.941. Meskipun EfficientNet dikenal karena efisiensi komputasinya dan penggunaan parameter yang lebih sedikit untuk mencapai performa yang baik, dalam hal akurasi, ResNet152 masih unggul. Hal ini menunjukkan bahwa meskipun EfficientNet sangat efisien, ResNet152 memiliki keunggulan dalam menangkap dan memproses fitur yang lebih kompleks berkat kedalamannya yang lebih besar dan arsitektur yang lebih optimal.

Dari hasil perbandingan ini, dapat disimpulkan bahwa ResNet152 outperform model-model lainnya karena kombinasi kedalaman jaringan, penggunaan shortcut connections, batch normalization, dan ReLU. Meskipun memiliki jumlah parameter yang besar, efisiensi dalam memanfaatkan parameter tersebut dan kemampuan menangkap fitur yang sangat kompleks menjadikan ResNet152 sebagai model yang superior dalam tugas-tugas klasifikasi gambar. Performanya yang unggul dalam hal loss dan F1 Score menunjukkan bahwa ResNet152 dapat memberikan hasil yang lebih akurat dan dapat

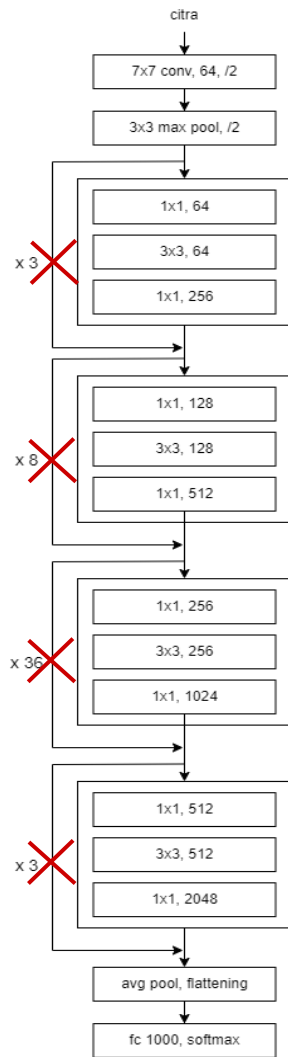
diandalkan dibandingkan model-model lain yang dianalisis.

4.1 Studi Ablasi

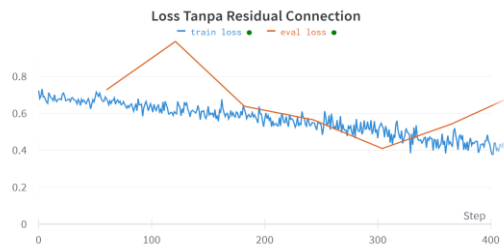
ResNet152 jauh lebih unggul jika dibandingkan dengan model lainnya karena memiliki *residual connection* yang memungkinkan untuk membangun arsitektur jaringan yang sangat dalam. Untuk lebih memahami bagaimana *residual connection* dan arsitektur jaringan yang sangat dalam ini dapat mempengaruhi kinerja dari model ResNet152, dilakukan studi ablasi dengan tiga kondisi. Pertama, model ResNet152 dijalankan tanpa adanya *residual connection*. Pada kondisi ini, diselidiki lebih jauh tentang pengaruh dari *residual connection* dengan cara membandingkan antara ResNet152 dengan *residual connection* tanpa bobot *pretrained* dan ResNet152 dengan *residual connection* tanpa bobot *pretrained*. Penghilangan bobot *pretrained* ini bertujuan untuk mengurangi pengaruh bobot *pretrained* ke performa model. Sehingga dapat diketahui pengaruh *residual connection* pada kinerja ResNet152 dalam menyelesaikan permasalahan klasifikasi citra generasi AI. Kedua, model ResNet152 dibandingkan dengan model ResNet lainnya yang memiliki kedalaman jaringan yang jauh lebih dangkal. Ketiga, model-model *residual network* dengan kedalaman yang beragam dibandingkan dengan tiga kondisi, yaitu model *residual network* tanpa *residual connection* dan tanpa bobot *pretrained*; model *residual network* tanpa *residual connection* dengan bobot *pretrained*; dan model *residual network* dengan *residual connection* dan dengan bobot *pretrained*. Tujuan dari studi ini adalah melihat pengaruh kombinasi kedalaman jaringan, bobot *pretrained*, dan *residual connection* terhadap kinerja.

Untuk menghilangkan *residual connection*, perlu dilakukan perubahan pada tiap lapisan model *pretrained*. Untuk setiap blok besar konvolusi atau yang biasa disebut dengan *stage*, memiliki satu *residual connection* seperti yang ditunjukkan pada gambar 11. Pada lapisan *residual connection*, dihilangkan untuk setiap blok besar atau *stage*-nya. Kemudian, untuk perhitungan *forward pass* juga perlu dimodifikasi agar tidak melakukan perhitungan untuk bagian *residual connection*. Dengan begini, akan menghasilkan model ResNet152 tanpa *residual connection*.

Loss pelatihan yang didapatkan untuk model ResNet152 tanpa *residual connection* sangatlah tidak stabil yang ditunjukkan dengan *loss* yang fluktuatif seperti yang ditunjukkan pada Gambar 12. Model ini juga kesulitan untuk mencapai konvergensi yang tercermin melalui nilai *loss* yang tidak menunjukkan penurunan yang signifikan, serta evaluasi *loss* yang tidak mengindikasikan tren menurun. Fenomena ketidakmampuan untuk mencapai konvergensi ini dapat mengimplikasikan adanya bobot yang tidak mendapat nilai pembaruan secara efektif.



Gambar 11: Arsitektur ResNet152 Tanpa Residual Connection



Gambar 12: Loss Tanpa Residual Connection

Salah satu potensi penyebab dari fenomena ini adalah masalah *vanishing gradient* di mana gradien yang bertanggung jawab untuk memperbarui bobot mendekati nilai nol, sehingga pembaruan yang signifikan pada bobot menjadi minim. Hal ini dapat terjadi dikarenakan arsitektur model yang sangat mendalam tetapi tidak terdapat *residual connection*. Arsitektur yang sangat dalam memungkinkan terjadi penumpukan gradien, yang mana suatu fungsi aktivasi tersaturasi dihitung berulang kali sehingga mendapatkan hasil yang sangat kecil. *Residual connection* memungkinkan aliran gradien melompati suatu jaringan yang menumpuk sehingga tidak menghilangkan gradien tetapi tetap dapat menangkap

pola-pola kompleks karena kedalaman dari arsitekturnya. Lebih jauh, tabel 4 adalah metrik evaluasi yang dihasilkan oleh model ResNet152 tanpa *residual connection*. Hasilnya menunjukkan kinerja yang buruk.

Tabel 4 Evaluasi Metrik Tanpa Residual Connection

metrik	skor
F1 Score	0.6287
Recall	0.4660
Precision	0.9657
Akurasi	0.7226

Tabel 5 Evaluasi Metrik Dalam Berbagai Skenario

Skenario	F1 Score	Recall	Precision	Akurasi
1	0.5743	0.4178	0.9137	0.6857
2	0.5927	0.4230	0.9321	0.7013

Tabel 5 merupakan hasil evaluasi dari dua skenario pelatihan. Skenario satu adalah pelatihan model ResNet152 tanpa *residual connection* dan tanpa bobot *pretrained*. Sementara skenario dua adalah pelatihan model ResNet152 dengan *residual connection* dengan bobot *pretrained*. Hasil evaluasi untuk setiap pengukuran metrik, menunjukkan bahwa skenario kedua lebih unggul daripada skenario pertama. Meskipun selisih skornya hanya sedikit, namun hal ini membuktikan bahwa *residual connection* dapat membantu meningkatkan kinerja model ResNet152. Skor yang rendah ini menunjukkan bahwa kemungkinan lapisan konvolusi masih belum bisa melakukan ekstraksi fitur dengan baik karena belum konvergen, mengingat jumlah data dan jumlah *epoch* yang cenderung terbatas. Akan tetapi, bobot *pretrained* saja tanpa adanya *residual connection* juga menunjukkan hasil yang kurang memuaskan seperti pada Tabel 4. Lapisan *residual connection* memungkinkan model ResNet152 memiliki ruang pembelajaran untuk pola kompleks lebih jauh daripada tanpa *residual connection*, tetapi bukan berarti hanya menambahkan *residual connection* tanpa pelatihan yang cukup sudah bisa menaikkan kinerja model ResNet152. Sehingga dapat dipelajari bahwa bobot *pretrained* dan lapisan *residual connection* bekerja dengan sinergis dan memiliki andilnya masing-masing dalam meningkatkan kinerja dari model ResNet152.

Tabel 6 Perbandingan Model

model	loss	F1 Score	Parameter (juta)
ResNet18	0.19	0.928	11.7
ResNet50	0.16	0.931	25.6
ResNet101	0.16	0.933	44.5
ResNet152	0.08	0.963	60.2

Studi berikutnya adalah pada kasus jaringan yang sama-sama memiliki *residual connection* namun berbeda kedalamannya. Tujuan dari studi ablasi ini adalah untuk mengetahui pengaruh dari kedalaman model dengan *residual connection* dalam memahami pola-pola citra yang kompleks. Tabel 6 menunjukkan perbandingan berbagai model,

termasuk ResNet dengan variasi kedalaman yang berbeda, yaitu versi ResNet18, ResNet32, ResNet50, ResNet101, dan ResNet152. Dari *F1 score* didapati bahwa ResNet18 sebagai jaringan yang paling dangkal memiliki skor terburuk. Dari pola *F1 score* yang didapatkan, jumlah kedalaman lapisan cenderung linear dengan skornya, yang mana model terkecil memperoleh skor terburuk dan model terbesar memperoleh skor terbaik. Hasil ini menunjukkan model yang lebih dalam lebih memungkinkan untuk mempelajari pola-pola kompleks dari citra. Meskipun begitu, perbedaan skor yang didapatkan tidak terlalu signifikan, terutama jika dibandingkan dengan studi ablasi sebelumnya yang mana lapisan *residual connection* dieliminasi.

Tabel 7 merupakan hasil evaluasi metrik untuk berbagai skenario pada beragam variasi kedalaman model *residual network*. Skenario satu adalah ketika model *residual network* dilatih tanpa *residual connection* dan tanpa bobot *pretrained*. Selanjutnya skenario dua adalah ketika model *residual network* dilatih tanpa *residual connection* namun dengan bobot *pretrained*

Tabel 7 Evaluasi Metrik Dalam Berbagai Skenario dan Kedalaman Model *Residual Network*

Model	Skenario	Loss	F1 Score
ResNet18	1	0.798	0.479
	2	0.532	0.498
	3	0.19	0.928
ResNet50	1	0.671	0.504
	2	0.445	0.529
	3	0.16	0.931
ResNet101	1	0.572	0.521
	2	0.378	0.542
	3	0.16	0.933
ResNet152	1	0.558	0.574
	2	0.281	0.592
	3	0.08	0.963

Sementara skenario ketiga adalah untuk setiap model *residual network* dengan *residual connection* dan dengan bobot *pretrained*. Untuk setiap percobaan, *F1 score* dan *loss* sinergis yang ditandai dengan ketika *loss*-nya semakin rendah maka *F1 score* nya semakin tinggi. Selain itu, skenario ketiga tentu menjadi yang paling unggul untuk setiap kedalaman model. Kemudian disusul dengan skenario kedua yang jauh lebih unggul daripada skenario pertama untuk setiap variasi kedalaman model *residual network*. Hal ini menunjukkan bahwa kombinasi model *residual network* yang dalam ditambah dengan *residual connection* ditambah dengan bobot *pretrained* dapat memaksimalkan kinerja dalam studi kasus klasifikasi citra generasi AI. Hal ini sinergis dengan temuan-temuan sebelumnya pada studi ablasi yang menunjukkan jika *residual connection*, bobot *pretrained*, dan kedalaman model memungkinkan model *residual network* untuk lebih memahami dan mengekstraksi fitur-fitur penting pada citra, yang mempermudah kepala klasifikasi untuk melakukan pemisahan antara kelas *real* dan *fake*.

5. KESIMPULAN

Penelitian ini membahas tentang klasifikasi citra generasi AI menggunakan metode *fine tuning* pada *Residual Network* (ResNet). *Dataset* yang digunakan adalah CIFAKE yang terdiri dari dua kelas, yaitu gambar asli dan gambar generasi AI. Proses *fine tuning* dilakukan pada tiga jenis ResNet, yaitu ResNet50, ResNet101, dan ResNet152, dengan menggunakan AdamW sebagai algoritma optimasi, *linear learning rate scheduler* dengan *warm up*, dan *gradient accumulation*. Hasilnya menunjukkan bahwa ResNet152 memiliki performa terbaik di antara semua model yang dibandingkan, dengan *F1 Score* sebesar 0.963. Beberapa faktor yang mempengaruhi kinerja model, seperti warna, tekstur, dan latar belakang gambar. Studi ablasi menunjukkan bahwa kinerja model yang sangat baik ini dipengaruhi adanya *residual connection* yang memungkinkan membuat suatu arsitektur yang sangat dalam sehingga dapat memberi informasi lebih terhadap pola-pola citra yang kompleks. Sehingga, jaringan dapat memprediksi jauh lebih baik. Ditunjukkan juga bahwa bobot *pretrained* dan *residual connection* bekerja dengan sinergis dalam memperbaiki kinerja dari model ResNet152 dalam menyelesaikan permasalahan klasifikasi citra generasi AI.

DAFTAR PUSTAKA

- BIRD, J. J., & LOTFI, A. 2023. CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images (arXiv:2303.14126). arXiv. <https://doi.org/10.48550/arXiv.2303.14126>
- CHEN, X., XU, C., YANG, X., SONG, L., & TAO, D. 2019. Gated-GAN: Adversarial Gated Networks for Multi-Collection Style Transfer. *IEEE Transactions on Image Processing*, 28(2), 546–560. <https://doi.org/10.1109/TIP.2018.2869695>
- CHEN, Z., CHEN, L., ZHAO, Z., & WANG, Y. 2020. AI Illustrator: Art Illustration Generation Based on Generative Adversarial Network. 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC), 155–159. <https://doi.org/10.1109/ICIVC50857.2020.9177494>
- DHARIWAL, P., & NICHOL, A. 2021. Diffusion Models Beat GANs on Image Synthesis (arXiv:2105.05233). arXiv. <https://doi.org/10.48550/arXiv.2105.05233>
- GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A., & BENGIO, Y. 2014. Generative Adversarial Networks (arXiv:1406.2661). arXiv. <https://doi.org/10.48550/arXiv.1406.2661>
- GOTMARE, A., KESKAR, N. S., XIONG, C., & SOCHER, R. 2018. A Closer Look at Deep Learning Heuristics: Learning rate restarts,

- Warmup and Distillation.
<https://doi.org/10.48550/ARXIV.1810.13243>
- HE, K., ZHANG, X., REN, S., & SUN, J. 2015. Deep Residual Learning for Image Recognition (arXiv:1512.03385). arXiv.
<https://doi.org/10.48550/arXiv.1512.03385>
- HO, J., JAIN, A., & ABBEEL, P. 2020. Denoising Diffusion Probabilistic Models (arXiv:2006.11239). arXiv.
<https://doi.org/10.48550/arXiv.2006.11239>
- HUANG, G., LIU, Z., VAN DER MAATEN, L., & WEINBERGER, K. Q. 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
<https://doi.org/10.1109/CVPR.2017.243>
- KIM, J., KANG, H., YANG, J., JUNG, H., LEE, S., & LEE, J. 2023. Multitask Deep Learning for Human Activity, Speed, and Body Weight Estimation Using Commercial Smart Insoles. *IEEE Internet of Things Journal*, 10(18), 16121–16133.
<https://doi.org/10.1109/JIOT.2023.3267335>
- LEDIG, C., THEIS, L., HUSZAR, F., CABALLERO, J., CUNNINGHAM, A., ACOSTA, A., AITKEN, A., TEJANI, A., TOTZ, J., WANG, Z., & SHI, W. 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network (arXiv:1609.04802; Versi 5). arXiv.
<https://doi.org/10.48550/arXiv.1609.04802>
- LI, X., XIA, Y., LONG, X., LI, Z., & LI, S. 2021. Exploring Text-Transformers in AAAI 2021 Shared Task: COVID-19 Fake News Detection in English. Dalam T. Chakraborty, K. Shu, H. R. Bernard, H. Liu, & M. S. Akhtar (Ed.), *Combating Online Hostile Posts in Regional Languages during Emergency Situation* (Vol. 1402, hlm. 106–115). Springer International Publishing. https://doi.org/10.1007/978-3-030-73696-5_11.
- MARAS, M.-H., & ALEXANDROU, A. 2019. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23(3), 255–262.
<https://doi.org/10.1177/1365712718807226>
- MORANDÍN-AHUERMA, F. 2022. What is Artificial Intelligence. 3(12). *International Journal of Research Publication and Reviews*. DOI: 10.55248/gengpi.2022.31261
- SIMONYAN, K., & ZISSERMAN, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint.
<https://doi.org/10.48550/arXiv.1409.1556>
- SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., & RABINOVICH, A. 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9). doi: 10.1109/CVPR.2015.7298594
- PEREZ, L., & WANG, J. 2017. The Effectiveness of Data Augmentation in Image Classification using Deep Learning (arXiv:1712.04621). arXiv.
<https://doi.org/10.48550/arXiv.1712.04621>
- TANG, S., & DE SA, V. R. 2020. Deep Transfer Learning with Ridge Regression (arXiv:2006.06791). arXiv.
<https://doi.org/10.48550/arXiv.2006.06791>
- WESTERLUND, M. 2019. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 39–52.
<https://doi.org/10.22215/timreview/1282>
- WIGHTMAN, R., TOUVRON, H., & JÉGOU, H. 2021. ResNet strikes back: An improved training procedure in timm.
<https://doi.org/10.48550/ARXIV.2110.00476>
- WU, X., XU, K., & HALL, P. 2017. A survey of image synthesis and editing with generative adversarial networks. *Tsinghua Science and Technology*, 22(6), 660–674.
<https://doi.org/10.23919/TST.2017.8195348>
- XIONG, R., YANG, Y., HE, D., ZHENG, K., ZHENG, S., XING, C., ZHANG, H., LAN, Y., WANG, L., & LIU, T.-Y. 2020. On Layer Normalization in the Transformer Architecture.
<https://doi.org/10.48550/ARXIV.2002.04745>
- YAN, M., YANG, J., CHEN, C., ZHOU, J. T., PAN, Y., & ZENG, Z. 2022. Enhanced gradient learning for deep neural networks. *IET Image Processing*, 16(2), 365–377.
<https://doi.org/10.1049/ipr2.12353>
- ZANNETTOU, S., SIRIVIANOS, M., BLACKBURN, J., & KOURTELLIS, N. 2019. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. *Journal of Data and Information Quality*, 11(3), 1–37.
<https://doi.org/10.1145/3309699>
- ZHUANG, Z., LIU, M., CUTKOSKY, A., & ORABONA, F. 2022. Understanding AdamW through Proximal Methods and Scale-Freeness.
<https://doi.org/10.48550/ARXIV.2202.00089>