

Klasifikasi Gambar Asli dan Buatan AI Menggunakan SE-ResNet18 dan SE-ResNet50

Aurelia Catherine Lay^{*1}, Yuni Yamasari²

^{1,2}Program Studi Teknik Informatika, Universitas Negeri Surabaya, Surabaya
E-mail: ¹aurelia.22077@mhs.unesa.ac.id, ²yuniyamasari@unesa.ac.id

Abstrak

Perkembangan model generatif, seperti Generative Adversarial Networks (GAN) dan Diffusion Models, yang menghasilkan gambar berkualitas tinggi buatan AI yang sulit dibedakan dari gambar asli, menjadi pendorong utama penelitian ini. Tujuan penelitian ini adalah menggunakan arsitektur SE-ResNet50 dan SE-ResNet18 untuk membuat model klasifikasi yang mampu membedakan antara gambar asli dan gambar buatan AI. Teknik class weight dan optimasi hyperparameter yang memanfaatkan Optimasi Bayesian digunakan untuk mengatasi ketidakseimbangan data. Dataset Chameleon, yang berisi 26.033 gambar—11.170 gambar buatan AI dan 14.863 gambar asli digunakan. Berdasarkan hasilnya, SE-ResNet50 menunjukkan kemampuan generalisasi yang unggul dan mencapai akurasi 88,33%, presisi 91,60%, recall 87,58%, f1-score 0,88 ketika dipadukan dengan class weight dan Optimasi Bayesian. SE-ResNet18 mencapai akurasi 88,75%, presisi 0,9095%, recall 0,8917%, F1-score 0,88% pada skenario tanpa class weight dan Optimasi Bayesian. Penelitian ini menunjukkan bahwa arsitektur berbasis Squeeze-and-Excitation, yang dipadukan dengan teknik penyeimbangan data dan Optimasi Hyperparameter, terbukti menjadi pendekatan yang handal untuk klasifikasi gambar buatan AI ketika data tidak seimbang.

Kata Kunci: Deteksi Gambar, SE-ResNet, Class Weight, Optimasi Hyperparameter Bayesian,

1. PENDAHULUAN

Dalam beberapa tahun terakhir, kemajuan pesat telah terjadi dalam teknologi kecerdasan buatan generatif, mulai dari Generative Adversarial Networks (GANs) hingga Diffusion Models (DMs) yang menghasilkan gambar berkualitas tinggi menyerupai gambar asli [1]. Kemajuan teknologi ini, walaupun menawarkan peluang kreativitas yang besar, secara bersamaan menimbulkan kekhawatiran mendalam terkait keaslian visual dan penyebaran disinformasi. Ketika keaslian informasi visual tidak dapat terjamin lagi, kepercayaan publik terhadap media digital mulai menurun [2]. Upaya teknis dalam membedakan gambar asli dan gambar buatan AI menjadi mekanisme pertahanan yang sangat diperlukan dalam melindungi integritas informasi dan memerangi penyebaran deepfake.

Dampak nyata dari penyalahgunaan gambar buatan AI terlihat di berbagai bidang. Misalnya didalam bidang politik, gambar manipulatif yang memberikan disinformasi serta menurunkan kepercayaan publik terhadap informasi [3]. Di bidang hukum, pemalsuan bukti visual dapat mengancam integritas proses peradilan [4]. Selain itu, pada bidang digital, gambar AI dimanfaatkan untuk membuat profil media sosial palsu dalam berbagai kasus penipuan identitas [5].

Pendekatan berbasis Convolutional Neural Networks (CNN) dalam deteksi gambar AI telah mencapai hasil yang signifikan, akan tetapi menghadapi dua tantangan utama yaitu, menangani ketidakseimbangan kelas pada dataset dunia nyata, dan ketergantungan pada penyetelan hyperparameter secara manual yang tidak efisien [6], [7]. Keterbatasan ini mendorong perlunya kerangka kerja yang secara terpadu menangani bias distribusi data sekaligus mengotomatiskan proses optimasi arsitektur.

Penelitian sebelumnya telah mengembangkan berbagai pendekatan untuk deteksi gambar buatan AI. Mengusulkan model hybrid SE-ResNet50 yang memadukan ResNet50 dengan mekanisme Squeeze-and-Excitation (SE) serta membangun dataset CIFAKE yang terdiri dari 120,000 gambar dan menerapkan CNN [8], [9]. Selain itu, klasifikasi gambar wajah dengan fine-tuning pada ResNet-50 [10]. Meski demikian, penelitian tersebut umumnya dilakukan pada data yang seimbang dan belum secara khusus mengatasi ketidakseimbangan data maupun optimasi hyperparameter secara menyeluruh.

Penelitian ini bertujuan untuk mengevaluasi efektivitas penerapan class weight dalam mengatasi bias model pada dataset yang tidak seimbang. Mengukur efisiensi dan signifikansi Optimasi Hyperparameter Bayesian dalam menentukan konfigurasi yang optimal untuk menstabilkan konvergensi model. Menganalisis perbandingan kinerja secara komprehensif antara arsitektur SE-ResNet18, dan SE-ResNet50 pada berbagai skenario pengujian pada dataset chameleon

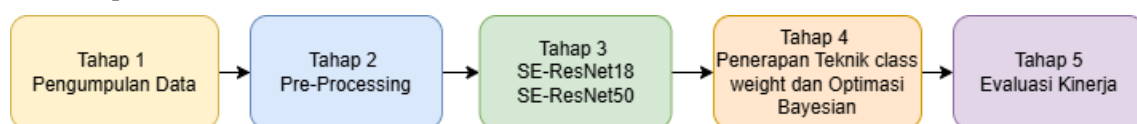
Penelitian ini menggunakan metode pada dua arsitektur deep learning yaitu SE-ResNet18, dan SE-ResNet50 [8], [11] pada dataset Chameleon yang terdiri dari 26.033 gambar asli dan gambar buatan AI [1]. Pengujian dilakukan pada tiga skenario yaitu tanpa class weight dan Optimasi Bayesian, dengan class weight, serta class weight dengan Optimasi Bayesian dan lima variasi pembagian dataset. Evaluasi menggunakan akurasi, presisi, recall, dan AUC.

Penelitian ini secara khusus mengevaluasi arsitektur berbasis Squeeze-and-Excitation, yaitu SE-ResNet18 dan SE-ResNet50, sebagai pendekatan CNN yang termasuk efisien secara komputasi untuk mendeteksi gambar buatan AI, sementara perbandingan dengan arsitektur berskala besar seperti EfficientNet ataupun arsitektur berbasis transformer seperti Vision Transformer (ViT) tidak termasuk dalam ruang lingkup penelitian ini dikarenakan perbedaan kebutuhan sumber daya dan kompleksitas pelatihan yang cukup jelas. Selain itu, dataset yang digunakan dirancang sehingga seluruh gambar buatan AI yang berasal dari Midjourney dan DALL-E 3 dapat lolos dari penilaian persepsi manusia, sehingga penelitian ini lebih difokuskan pada evaluasi mendalam terhadap performa model pada rentang generator tersebut, sedangkan potensi generalisasi terhadap model generatif AI yang lebih baru akan dibahas sebagai bagian dari keterbatasan penelitian.

Kebaruan penelitian ini terletak pada perbandingan menyeluruh dua arsitektur deep learning yaitu SE-ResNet18, dan SE-ResNet50 dengan menggunakan dataset Chameleon [1] yang dirancang sehingga seluruh gambar buatan AI dapat lolos dari penilaian persepsi manusia. Pendekatan yang diusulkan mengintegrasikan class weight untuk mengurangi bias akibat distribusi data yang tidak seimbang, serta Optimasi Hyperparameter Bayesian [12] untuk penyetelan model secara otomatis dan efisien.

2. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Alur Penelitian

Penelitian ini mencakup lima tahapan sebagaimana pada Gambar 1. Tahap pertama adalah pengumpulan data, yaitu pengunduhan dataset Chameleon yang terdiri dari 26.033 gambar, baik asli maupun hasil AI. Tahap kedua adalah pre-processing meliputi resize gambar, augmentasi data, dan normalisasi. Tahap ketiga adalah penerapan arsitektur SE-ResNet18 dan SE-ResNet50. Tahap keempat adalah penerapan teknik class weight dan Optimasi Bayesian (Baseline, Class

weight, dan Class weight + Bayesian). Tahap kelima adalah evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, F1-score, AUC, dan confusion matrix.

2.2 Sumber Data

Penelitian ini menggunakan Dataset Chameleon [1] yang terdiri dari 26.033 gambar dan memiliki dua kelas yaitu gambar buatan AI dan gambar asli. 14.863 gambar asli dari Unplash (57,1%) dan 11.170 gambar buatan AI dari ArtStation, Civitai, dan Liblib (42,9%), yang dibuat oleh model generatif seperti Midjourney dan DALL-E 3. Dataset ini dirancang sebagai benchmark yang menantang karena seluruh gambar AI telah melewati proses kurasi ketat, termasuk filter resolusi, filter kualitas (hash comparison), filter semantik (CLIP), dan anotasi oleh 20 anotator independen, sehingga semua gambar buatan AI dapat menipu persepsi manusia. Tabel 1 menyajikan distribusi data.

Tabel 1. Distribusi Dataset

Label Kelas	Jumlah Gambar	Proporsi
Gambar Asli	14.863	57,1%
Gambar buatan AI	11.170	42,9%
Total	26.033	100%

2.3 Pre-Processing

Gambar diresize ke 224×224 piksel sesuai standar input ResNet berbasis ImageNet [8], [10]. Augmentasi data meliputi rotasi acak $\pm 20^\circ$, pergeseran horizontal dan vertikal 20%, shear 20%, zoom 20%, serta horizontal flip hanya diterapkan pada data latih untuk meningkatkan variasi dan mencegah overfitting. Normalisasi piksel menggunakan mean dan standar deviasi RGB dari ImageNet[13]. Pembagian dataset menjadi lima varian yaitu 70:15:15, 80:10:10, 60:20:20, 75:15:10, 70:20:10, dilakukan sebelum pre-processing guna mencegah kebocoran data [14]. Data validasi dan uji hanya dinormalisasi, tanpa augmentasi.

2.4 Arsitektur Model

Penelitian ini mengimplementasikan dua arsitektur yaitu SE-ResNet18 merupakan *backbone* ResNet-18 dengan kedalaman lapisan 18 *layer* [11], [15] dan penambahan blok Squeeze-and-Excitation pada setiap residual block. Adapun SE-ResNet50 adalah *backbone* ResNet-50 dengan kedalaman lapisan 50 *layer* [10] dan integrasi blok Squeeze-and-Excitation yang sama.

Mekanisme *Squeeze-and-Excitation* terdiri dari dua proses berurutan. Pada tahap *squeeze*, tahap ini bertujuan memadatkan informasi spasial dari setiap saluran fitur (dengan tinggi H dan lebar W) menjadi sebuah deskriptor skalar tunggal. Kompresi ini dicapai melalui operasi *Global Average Pooling* (GAP), yang secara matematis ditunjukkan pada persamaan (1):

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

Dimana Z_c adalah luaran skalar untuk saluran ke- c , dan u_c merepresentasikan peta fitur pada saluran tersebut.

Tahap *excitation* menggunakan dua lapisan *fully connected* (FC) yang dipisahkan oleh fungsi aktivasi *ReLU*, diikuti oleh fungsi sigmoid, untuk menyesuaikan tingkat signifikansi setiap saluran. Vektor ini kemudian digunakan untuk melakukan penskalaan ulang terhadap peta fitur secara adaptif [8], [16]. Fungsi ini merumuskan vektor pembobotan adaptif pada persamaan (2):

$$s = F_{ex}(z, W) = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \quad (2)$$

2.5 Class Weight

Class weight adalah nilai yang diberikan untuk setiap kelas selama proses pelatihan model, dirancang untuk meningkatkan kinerja model pada kelas dengan jumlah data lebih sedikit. Untuk mengurangi ketidakseimbangan kelas [17]. Diterapkan menggunakan rumus pada persamaan (3) [18]:

$$W_c = \frac{N}{C \times n_c} \quad (3)$$

Class weight dipilih sebagai teknik penyeimbangan data dikarenakan penerapannya tidak menambah jumlah data maupun mengubah distribusi asli data, sehingga sesuai dengan tujuan penelitian yang lebih memfokuskan pada evaluasi arsitektur model daripada rekayasa data. Adapun perbandingan dengan teknik penyeimbangan lain tidak dilakukan dalam penelitian ini dan bisa dijadikan arah untuk penelitian lanjutan.

2.6 Optimasi Hyperparameter Bayesian

Optimasi Hyperparameter Bayesian [12], [19] diimplementasikan guna menentukan konfigurasi model yang optimal secara efisien. Berbeda dengan *grid search* atau *random search*, pendekatan ini secara strategis memanfaatkan hasil iterasi sebelumnya untuk mengarahkan pencarian menuju parameter optimal. Mekanisme tersebut menjadikan optimasi jauh lebih terarah dan hemat komputasi. Penelitian ini hanya menerapkan Bayesian Optimization sebagai metode optimasi hyperparameter, tanpa membandingkan langsung secara empiris dengan *grid search* maupun *random search* pada dataset yang digunakan, perbandingan tersebut bisa dijadikan arah penelitian lanjutan.

2.7 Sumber Daya Komputasi

Pelatihan model dilakukan pada platform Google Colaboratory dengan runtime Python 3 dan akselerasi hardware GPU NVIDIA Tesla T4. SE-ResNet18 memiliki total 11.543.161 parameter sekitar 44,03 MB, sedangkan SE-ResNet50 memiliki total 26.381.105 parameter sekitar 100,64 MB yang dimana lebih dua kali lipat dibandingkan SE-ResNet18. Pada konfigurasi terbaik masing-masing arsitektur, proses pelatihan berlangsung selama 20 *epoch* yang memiliki rata-rata waktu per *epoch* sekitar 5,9 menit untuk SE-ResNet18 dengan total kurang lebih 63 menit dan 6,7 menit untuk SE-ResNet50 dengan total kurang lebih 71 menit, yang menunjukkan bahwa arsitektur yang lebih dalam membutuhkan beban komputasi lebih besar. Untuk SE-ResNet50, proses tuning hyperparameter dengan Optimasi Bayesian dilakukan sebanyak 6 trial, dengan total waktu 21 menit 57 detik sebelum pelatihan akhir dilakukan menggunakan konfigurasi optimal yang diperoleh. Waktu pelatihan pada skenario dan variasi split data lainnya dapat berbeda-beda, tergantung pada proporsi data pelatihan yang digunakan.

2.8 Evaluasi Kinerja Model

Kinerja model diukur berdasarkan pada analisis *confusion matrix* yang terdiri atas empat komponen dasar: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dari keempat komponen tersebut, diturunkan lima metrik evaluasi sebagai berikut pada persamaan (4), (5), (6), (7), (8) [20]:

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

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \quad (7)$$

$$AUC = TPR = \frac{TP}{TP + FN} \quad (8)$$

$$FPR = \frac{FP}{FP + TN}$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pre-Processing



Gambar 2. Hasil Pre-processing dan Augmentasi

Berdasarkan pada Gambar 2 menampilkan contoh *pre-processing* dan augmentasi data menggunakan ImageDataGenerator diterapkan pada data pelatihan. Augmentasi digunakan secara acak pada setiap gambar selama proses pelatihan. Tampak bahwa gambar-gambar tersebut mengalami sejumlah perubahan, termasuk rotasi, pergeseran posisi, pembesaran, dan pembalikan horizontal. Untuk meningkatkan keragaman data pelatihan secara artifisial guna mencegah overfitting dan meningkatkan kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya [13], [21].

3.2 Hasil Evaluasi Kinerja SE-ResNet18

Model SE-ResNet-18 menunjukkan performa yang bervariasi dengan konfigurasi terbaik pada split data 70:15:10 dengan skenario *baseline* (tanpa class weight + bayesian) memiliki akurasi 0,8875, presisi 0,9095, *recall* 0,8917, F1-score 0,88. Konfigurasi terendah pada split data 70:15:15 dengan skenario *baseline* memiliki akurasi 0,8098, presisi 0,9498, *recall* 0,7040, F1-score 0,81. Tabel 2 menyajikan konfigurasi terbaik dan terendah SE-ResNet18.

Tabel 2. Konfigurasi Terbaik dan Terendah SE-ResNet18

Konfigurasi	Split / Skenario	Akurasi	Presisi	Recall	F1-score
Terbaik	75:15:10 / Baseline	0.8875	0.9095	0.8917	0.88
Recall Tertinggi	70:20:10 / Balancing	0.8829	0.8685	0.88	0.9368
Terendah	70:15:15 / Baseline	0.8098	0.9498	0.7040	0.81

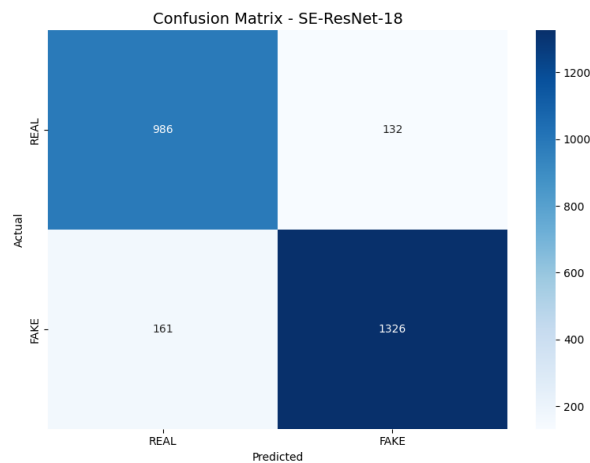
3.3 Hasil Evaluasi Kinerja SE-ResNet50

Berdasarkan Tabel 3 menunjukkan performa yang bervariasi dengan konfigurasi terbaik pada split data 70:15:15 dengan skenario Class weight + Bayesian memiliki akurasi 0,8833, presisi 0,9160, *recall* 0,8758, F1-score 0,88. Konfigurasi terendah pada split data 60:20:20 dengan skenario *baseline* (tanpa class weight + bayesian) memiliki akurasi 0,8018, presisi 0,9281, *recall* 0,7077, F1-score 0,80. Tabel 3 menyajikan konfigurasi terbaik dan terendah SE-ResNet50 beserta grafik pelatihan masing-masing konfigurasi.

Tabel 3. Konfigurasi Terbaik dan Terendah SE-ResNet50

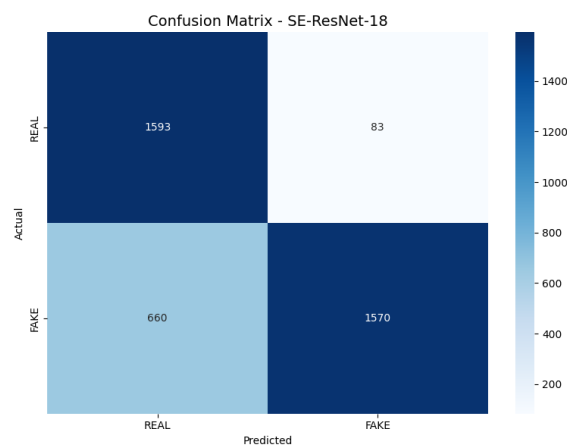
Konfigurasi	Split / Skenario	Akurasi	Presisi	Recall	F1-score
Terbaik	70:15:15 / Class weight+Bayesian	0.8833	0.9160	0.8758	0.88
Terendah	60:20:20 / Baseline	0.8018	0.9281	0.7077	0.80

3.4 Hasil Evaluasi Confusion Matrix



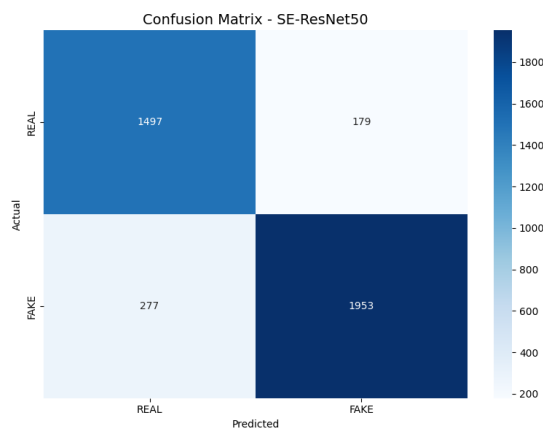
Gambar 3. Confusion Matrix model SE-ResNet18 performa terbaik

Berdasarkan confusion matrix yang ditunjukkan pada Gambar 3, model mengidentifikasi 986 gambar asli (True Negative), 1.326 gambar buatan AI (True Positive). Sementara itu, terdapat 161 gambar buatan AI salah diklasifikasi sebagai gambar asli (False Negatives), dan 132 gambar asli salah diklasifikasi sebagai gambar buatan AI (False Positive).



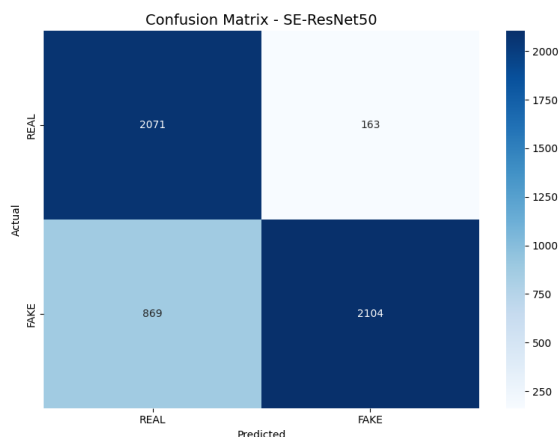
Gambar 4. Confusion Matrix model SE-ResNet18 performa terendah

Berdasarkan confusion matrix yang ditunjukkan pada Gambar 4, model mengidentifikasi 1593 gambar asli (True Negatives), 1570 gambar buatan AI (True Positive). Sementara itu, terdapat 660 gambar buatan AI salah diidentifikasi sebagai gambar asli (False Negatives), dan 83 gambar asli salah diidentifikasi sebagai gambar buatan AI (False Positive).



Gambar 5. Confusion Matrix model SE-ResNet50 performa terbaik

Berdasarkan Gambar 5, model SE-ResNet50 mengidentifikasi 1.497 gambar asli (True Negatives) dan 1.953 gambar buatan AI (True Positive). Di sisi lain, 277 gambar buatan AI salah diidentifikasikan sebagai asli (False Negatives), sementara 179 gambar asli salah diidentifikasi sebagai palsu (False Positives).



Gambar 6. Confusion Matrix model SE-ResNet50 performa terendah

Berdasarkan confusion matrix yang ditunjukkan pada Gambar 6, model SE-ResNet50 mengidentifikasi 2071 gambar asli (True Negatives) dan 2104 gambar buatan AI (Positive). Namun demikian, 869 gambar buatan AI (Negatives) secara keliru diidentifikasi sebagai asli, dan 163 gambar asli (False Positives) secara keliru diidentifikasi sebagai palsu.

3.5 Interpretasi Kinerja SE-ResNet18

Hasil penelitian SE-ResNet18 menunjukkan akurasi optimal 88,75% dengan *recall* 89,17% pada konfigurasi terbaik pada split data 75:15:10 dengan skenario *baseline* tanpa class weight maupun Optimasi Bayesian. Hal ini menunjukkan bahwa mekanisme *Squeeze-and-Excitation* pada ResNet18 yang sederhana mampu menyesuaikan bobot antarkarnal secara adaptif [11], [15]. Sehingga model dapat membedakan fitur penting antara gambar asli dan buatan AI meskipun data tidak seimbang. Mekanisme berbasis kanal memungkinkan jaringan menekan fitur yang kurang

relevan dan memperkuat fitur yang lebih informatif [22]. Karakteristik yang bermanfaat pada dataset dengan distribusi tidak seimbang.

3.6 Interpretasi Kinerja SE-ResNet50

Hasil penelitian SE-ResNet50 menunjukkan konfigurasi terbaiknya pada split data 70:15:15 dengan skenario class weight + bayesian, akurasi optimal 88,33% dengan presisi 91,60%. Hal ini menandakan bahwa ketika diatur secara optimal, SE-ResNet50 dapat menghasilkan prediksi positif dengan tingkat keakuratan yang sangat baik. Kedalaman arsitektur dengan 50 lapisan residual memang menyediakan ruang representasi fitur yang lebih luas, akan tetapi juga meningkatkan jumlah parameter yang harus disesuaikan. Kondisi ini membuat model lebih rentan terhadap perubahan distribusi data saat pelatihan.

Meskipun SE-ResNet18 dan SE-ResNet50 menunjukkan performa yang cukup baik dengan akurasi sekitar 88%, masih ada sejumlah kesalahan klasifikasi yang perlu diperhatikan. Pada konfigurasi terbaik SE-ResNet18, masih terdapat 132 gambar asli yang salah diklasifikasikan sebagai gambar buatan AI dan 161 gambar buatan AI yang diklasifikasikan sebagai gambar asli. Kesalahan klasifikasi lebih banyak ditemukan pada konfigurasi dengan performa yang lebih rendah. Hal ini terlihat pada Gambar 4 dan Gambar 6, yang menunjukkan bahwa ratusan gambar buatan AI masih belum berhasil dideteksi. Hasil tersebut menunjukkan bahwa meskipun model memiliki akurasi yang baik secara keseluruhan, proses deteksi gambar buatan AI pada dataset Chameleon masih menghadapi kendala. Kondisi ini berkaitan dengan karakteristik dataset Chameleon yang secara khusus menggunakan gambar buatan AI yang dirancang agar sulit dibedakan dari gambar asli oleh manusia.

3.7 Dampak Class Weight dan Optimasi Hyperparameter Bayesian

Pengaruh class weight cukup signifikan dan berbeda-beda antar arsitektur. Pada SE-ResNet18, pemberian class weight mampu menaikkan *recall* dari 70,40% pada konfigurasi terendah tanpa penyeimbangan menjadi 93,68% pada konfigurasi terbaik dengan penyeimbangan, peningkatan sebesar 23,28 poin persentase yang merupakan dampak terbesar di antara semua intervensi yang diuji. Class weight terbukti mampu meningkatkan kerentanan model terhadap kelas minoritas secara konsisten [17]. Memberikan penalti berdasarkan proporsi setiap kelas dalam data, sehingga mekanisme *Squeeze-and-Excitation* pada SE-ResNet18 dapat menyesuaikan bobot antarkanal secara lebih proporsional selama proses pelatihan berlangsung.

Pada SE-ResNet50, penggunaan class weight saja belum mampu untuk menghasilkan kinerja optimal, jumlah hyperparameter yang jauh lebih besar pada arsitektur ini dapat mengurangi manfaat penyeimbangan jika tidak diatur dengan tepat. Optimasi Bayesian mengatasi keterbatasan ini dengan menelusuri ruang hyperparameter secara lebih terarah, sehingga menghasilkan konfigurasi yang optimal [12]. Pendekatan ini terbukti efisien dibandingkan metode pencarian biasa seperti grid search maupun random search [19]. Selain itu, penggunaan teknik class weight yang dipadukan dengan optimasi hyperparameter secara umum diketahui mampu meningkatkan akurasi CNN hingga 3-5 poin persentase pada data dengan distribusi kelas yang tidak merata [7].

Kombinasi antara class weight dan optimasi bayesian tidak selalu memberikan peningkatan yang sebanding. Pada SE-ResNet18, beberapa konfigurasi yang menggabungkan kedua teknik menunjukkan kinerja sedikit lebih rendah dibanding dengan class weight saja. Hal ini menandakan adanya interaksi kompleks antara class weight dan hyperparameter tertentu seperti dropout. Kondisi ini juga pernah terjadi, dimana penerapan teknik secara berlebihan dapat mengganggu keseimbangan yang telah terbentuk melalui pendekatan yang lebih sederhana [23]. Peningkatan kinerja yang tidak konsisten pada beberapa konfigurasi gabungan menunjukkan bahwa efektivitas Bayesian Optimization dipengaruhi oleh ruang pencarian hyperparameter dan

jumlah percobaan yang digunakan. Jumlah percobaan yang terbatas, seperti pada penelitian ini, dapat membuat proses optimasi berhenti pada optimum lokal dan belum mencapai optimum global. Kondisi tersebut dapat menyebabkan penerapan Bayesian Optimization bersama class weight memberikan hasil yang berbeda pada setiap konfigurasi. Oleh sebab itu, penggunaan kedua teknik tersebut perlu disertai dengan penentuan ruang pencarian hyperparameter yang tepat agar peningkatan kinerja dapat diperoleh secara lebih konsisten.

3.8 Analisis Karakteristik Visual Gambar Kesalahan Klasifikasi



Gambar 7. Karakteristik Visual Gambar Kesalahan Klasifikasi

Pengamatan awal terhadap beberapa sampel prediksi SE-ResNet18 yang salah menunjukkan pola yang serupa pada dua kasus, seperti terlihat pada Gambar 7. Dimana pada gambar kanan berupa jalanan kota dengan anjing dibagian depan, sedangkan gambar kiri berupa menampilkan vas keramik dengan motif dekoratif. Keduanya merupakan gambar asli (REAL), tetapi diprediksi sebagai gambar buatan AI dengan tingkat keyakinan masing-masing sebesar 0,89 dan 0,67. Kedua gambar tersebut memiliki beberapa karakteristik yang serupa, seperti warna yang cukup kuat, komposisi yang terlihat artistik, serta pencahayaan yang tampak profesional. Karakteristik tersebut dapat menyerupai tampilan gambar yang dihasilkan oleh model generatif. Kondisi ini diduga membuat model lebih sulit membedakan gambar asli dengan gaya editorial dari gambar buatan AI. Namun, pengamatan ini masih bersifat awal dan hanya didasarkan pada beberapa sampel kesalahan klasifikasi SE-ResNet. Analisis visual terhadap kesalahan klasifikasi pada SE-ResNet50 belum dilakukan, sehingga pola kesalahan pada kedua arsitektur belum dapat disimpulkan secara menyeluruh.

3.9 Perbandingan dengan Penelitian Terdahulu

Hasil penelitian ini menunjukkan mencapai akurasi 88,33% terlihat lebih rendah secara angka, jika dibandingkan dengan hasil penelitian sebelumnya yang mencapai akurasi 96,12 menggunakan SE-ResNet50 pada data seimbang [8]. Namun, perbedaan ini dikarenakan perbedaan kondisi percobaan. Penelitian tersebut menggunakan dataset yang diseimbangkan, sedangkan penelitian ini menggunakan dataset Chameleon yang tidak seimbang dan dirancang khusus agar semua gambar AI sulit dibedakan secara kasar mata oleh manusia, sebuah tantangan yang berat dan mencerminkan kondisi nyata. Oleh karena itu, pentingnya evaluasi model deep learning pada kondisi nyata yang bervariasi dan tidak ideal untuk mengukur ketahanan model secara mendalam [24].

3.10 Keterbatasan Penelitian

Penelitian ini memiliki keterbatasan yang mencakup tiga aspek. Pertama, dataset Chameleon yang digunakan mencakup gambar buatan AI dari Midjourney dan DALL-E 3 saja, sehingga kemampuan generalisasi model terhadap gambar dari model generatif lain belum dapat dijamin. Kedua, teknik penyeimbangan data yang digunakan class weight tanpa perbandingan dengan metode alternatif lain. Ketiga, Optimasi *hyperparameter* menggunakan *Bayesian Optimization*,

tanpa perbandingan menyeluruh dengan pendekatan mendalam dengan hyperparameter lainnya.

4. KESIMPULAN

Penelitian ini membandingkan dua arsitektur yaitu SE-ResNet18 dan SE-ResNet50, untuk mendeteksi gambar buatan AI pada dataset Chameleon yang tidak seimbang. SE-ResNet18 mencapai akurasi terbaik 88,75% dengan recall 89,17% dan kesalahan prediksi yang relatif seimbang. Hal ini menjadikannya arsitektur yang unggul dalam menjaga keseimbangan antara akurasi, kemampuan generalisasi, dan keyakinan prediksi. Mekanisme *Squeeze-and-Excitation* pada model yang sederhana terbukti efektif meskipun tanpa class weight, akan tetapi tetap menunjukkan respons besar terhadap class weight, terbukti dengan kenaikan *recall* hingga 93,68%. SE-ResNet50 menunjukkan kemampuan generalisasi yang stabil tanpa tanda overfitting pada berbagai konfigurasi. Model ini mencapai presisi tertinggi 91,60% ketika dikombinasikan dengan class weight dan Optimasi Hyperparameter Bayesian. Hal ini menegaskan bahwa arsitektur yang lebih mendalam membutuhkan strategi pelatihan yang lebih terencana agar dapat bekerja secara optimal. Kedua arsitektur ini menunjukkan keunggulan yang saling melengkapi. SE-ResNet18 baik dalam menjaga keseimbangan antara akurasi dan *recall* serta stabilitas prediksi, sementara SE-ResNet50 unggul dalam hasil presisi dan kestabilan konvergensi dalam jangka panjang.

5. SARAN

Penelitian lanjutan disarankan untuk (1) evaluasi kedua arsitektur dengan dataset yang diperbarui secara berkala dengan gambar-gambar dari model generatif terbaru agar sistem pendeteksian tetap relevan di tengah pesatnya perkembangan teknologi AI; (2) Mengkaji beberapa metode penyeimbangan data, seperti *SMOTE* dan *Focal Loss*; (3) Membandingkan kinerja dengan model-model terkini seperti Vision Transformer (ViT) dan EfficientNet; (4) Mengembangkan sistem deteksi secara *real-time* agar dapat diterapkan dalam skenario deteksi secara langsung; (5) Mengimplementasikan model ke dalam aplikasi praktis seperti verifikasi konten digital atau platform media sosial.

DAFTAR PUSTAKA

- [1] S. Yan *et al.*, "A Sanity Check for AI-generated Image Detection," Feb. 2025, [Online]. Available: <http://arxiv.org/abs/2406.19435>
- [2] Y. Wang, Y. Hao, and A. X. Cong, "Harnessing Machine Learning for Discerning AI-Generated Synthetic Images," May 2024, [Online]. Available: <http://arxiv.org/abs/2401.07358>
- [3] V. Dan, "Deepfakes as a Democratic Threat: Experimental Evidence Shows Noxious Effects That Are Reducible Through Journalistic Fact Checks," *International Journal of Press/Politics*, vol. 31, no. 3, pp. 525–550, Sep. 2019, doi: 10.1177/19401612251317766.
- [4] J. Tan, A. E. Cinar, and K. Benyekhle, "Can You Trust What You See? Human and AI Detection of Synthetic Legal Evidence," 2026.
- [5] "Internet Crime Complaint Center (IC3) | Criminals Use Generative Artificial Intelligence to Facilitate Financial Fraud." Accessed: Aug. 17, 2026. [Online]. Available: <https://www.ic3.gov/PSA/2024/PSA241203>
- [6] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, "A survey on imbalanced learning: latest research, applications and future directions," *Artif. Intell. Rev.*, vol. 57, no. 6, Jun. 2024, doi: 10.1007/s10462-024-10759-6.
- [7] W. Hussain *et al.*, "Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-024-76178-3.
- [8] A. R. Gunukula, H. Das Gupta, and V. S. Sheng, "Detecting AI-Generated Images Using a Hybrid ResNet-SE Attention Model," *Applied Sciences (Switzerland)*, vol. 15, no. 13, Jul. 2025, doi: 10.3390/app15137421.
- [9] J. J. Bird and A. Lotfi, "CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images," Mar. 2023, [Online]. Available: <http://arxiv.org/abs/2303.14126>

-
- [10] K. A. Abdal Mashkoor, "Classifying Real and AI-Generated Images Using Fine-Tuned ResNet50," *Journal of Advanced Artificial Intelligence, Engineering and Technology*, vol. 01, no. 02, 2025, doi: 10.56147/aaiet.1.2.7.
 - [11] A. Asare, A. Adjei Broni, A. K. Asare Dickson, M. Sagoe, and J. M. Cudjoe, "Performance of ResNet-18 and InceptionResNetV2 in Automated Detection of Diabetic Retinopathy," *Medicine Advances*, vol. 3, no. 3, pp. 231–241, Sep. 2025, doi: 10.1002/med4.70023.
 - [12] M. Subramanian *et al.*, "Diagnosis of Retinal Diseases Based on Bayesian Optimization Deep Learning Network Using Optical Coherence Tomography Images," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/8014979.
 - [13] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
 - [14] N. D. Pham and N. Chilamkurti, "Data Leakage Threats and Protection in Split Learning: A Survey," Association for Computing Machinery (ACM), Dec. 2023, pp. 141–147. doi: 10.1145/3659154.3659189.
 - [15] H. Mo, Z. Kuang, H. Wang, X. Cai, and K. Cheng, "Enhancing Burn Diagnosis through SE-ResNet18 and Confidence Filtering," *Journal of Imaging Informatics in Medicine*, Feb. 2025, doi: 10.1007/s10278-025-01495-6.
 - [16] M. Khan and Y. Hossni, "A comparative analysis of LSTM models aided with attention and squeeze and excitation blocks for activity recognition," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-88378-6.
 - [17] W. Setiawan, M. Koeshardianto, E. M. S. Rochman, A. Rachmad, and T. Herawan, "Class Weighting in Imbalanced Data for Leukocyte Classification using Fine-Tuning Inception-v3," *Iraqi Journal for Computer Science and Mathematics*, vol. 6, no. 3, pp. 380–395, 2025, doi: 10.52866/2788-7421.1283.
 - [18] M. Abdelhamid and A. Desai, "Balancing the Scales: A Comprehensive Study on Tackling Class Imbalance in Binary Classification," Sep. 2024, [Online]. Available: <http://arxiv.org/abs/2409.19751>
 - [19] M. Masum *et al.*, "Bayesian Hyperparameter Optimization for Deep Neural Network-Based Network Intrusion Detection."
 - [20] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min.*, vol. 16, no. 1, Dec. 2023, doi: 10.1186/s13040-023-00322-4.
 - [21] D. Stojnev and A. Stojnev Ilić, "Preprocessing Image Data for Deep Learning," Singidunum University, Oct. 2020, pp. 312–317. doi: 10.15308/sinteza-2020-312-317.
 - [22] M. H. Guo *et al.*, "Attention mechanisms in computer vision: A survey," Sep. 01, 2022, *Tsinghua University*. doi: 10.1007/s41095-022-0271-y.
 - [23] I. Araf, A. Idri, and I. Chair, "Cost-sensitive learning for imbalanced medical data: a review," *Artif. Intell. Rev.*, vol. 57, no. 4, Apr. 2024, doi: 10.1007/s10462-023-10652-8.
 - [24] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Machine Learning with Applications*, vol. 6, p. 100134, 2021, doi: 10.24433/CO.0411648.v1.
-