

Anime AI-Generated Artwork Detector Using MobileNetV3Large

Seagata Ade Pratama Barus (1301210371)
seagata@student.telkomuniversity.ac.id

Abstract — Melakukan klasifikasi pada Anime AI-Generated Artwork dengan menggunakan feature extraction berupa Gabor Filter, Local Binary Pattern, dan Feature Maps. Model CNN yang digunakan ialah MobileNetV3Large dengan input layer yang dinaikkan menjadi 560×560 dan 2 layer atas yang di modif sedikit, dataset yang digunakan dikumpulkan dari berbagai gambar populer dengan sumber, rentang waktu, dan artstyle yang beragam menghasilkan akurasi 81%.

I. Introduction

Topik tugas besar saya dalam mata kuliah Pengolahan Citra Digital adalah fenomena terkini mengenai ***AI-Generated Artwork***, khususnya sejak Jason M. Allen memenangi kompetisi seni digital di Colorado State Fair Fine Arts 2022 dengan karya berbasis AI, *MidJourney*. Peristiwa ini memicu perdebatan luas di media dan masyarakat daring.

Pada 2024, karya AI, terutama dengan gaya anime, semakin umum ditemukan, meski menuai pro dan kontra. Di luar isu etika dan hak cipta, muncul masalah berupa penipuan dan *deepfake*, yang menimbulkan kebutuhan mendesak akan alat deteksi karya berbasis AI. Saya mengusulkan pendekatan menggunakan CNN ***MobileNetV3Large***, yang ringan dan cepat untuk perangkat seluler, dengan metode *feature*

extraction Gabor Filter, dibandingkan dengan Gabor Filter, LBP, dan *Feature Maps*, guna menentukan metode yang paling efektif.

II. Related Work

[1] Melakukan penelitian mengenai AI-generated image, namun penelitian tersebut lebih general dalam hal citra yang akan dideteksi, dan lebih terfokus pada image yang telah di modifikasi oleh AI atau yang biasa disebut dengan in-paint, dimana image di modifikasi dengan sebuah prompt tertentu, yang berikutnya akan di generate oleh AI. [2] melakukan penelitian yang mirip namun lebih terfokus kepada aspek *art* daripada citra secara umum, dengan hasil akurasi yang tinggi dengan pendekatan yang cukup rumit. Untuk anime art sendiri [3] melakukan penelitian yang mirip dengan tugas besar ini, dengan menggunakan *MobileNetV2* & *MobileNetV3* menghasilkan akurasi 96,8% dan 97,2%, yang membedakan tugas akhir ini dengan penelitian tersebut adalah dataset yang digunakan, dimana dataset yang saya gunakan adalah dataset yang lebih modern dan lebih sulit dikenali/rumit dan dengan berbagai ukuran citra dan dari berbagai sumber serta model generasi, dimana [3] menggunakan dataset ai artwork yang cukup lawas dan mudah dikenali serta hanya di generate dengan NovelAI belaka.

III. Data

Untuk mendeteksi *AI-Generated Anime Artwork*, diperlukan dataset yang relevan. Dataset akan diambil dari Danbooru, Civitai, Zerochan, dan Pixiv, mencakup 500 citra AI dan 500 citra buatan manusia. Citra dipilih berdasarkan popularitas (like, rating, bookmark, dan waktu) guna meningkatkan akurasi dan memastikan model dapat mendeteksi AI dari yang paling modern hingga yang cukup lawas..

Preprocessing berupa image enhancement dan augmentation data tidak dilakukan, dikarenakan betapa sensitifnya citra yang akan digunakan dengan manipulasi citra, oleh sebab itu hanya akan dilakukan *feature extraction* pada citra. Terdapat 3 metode feature extraction yang digunakan pada tugas ini yaitu Gabor Filter, LBP, dan feature extraction default dari CNN yaitu *Feature Maps*. Preprocessing masih tetap akan dilakukan tapi hanya pada model Feature Maps karena dibutuhkannya normalisasi, dengan menggunakan library preprocess_input dari tensorflow MobileNetV3.



Proses Pembagian dataset akan dilakukan dengan rasio standar 80:20, 80% train dan 20% validation.

IV. Methods

Seperti penjelasan pada introduction implementasi model ini menggunakan pre-trained *MobileNetV3Large* pada dataset imagenet karena arsitektur CNN tersebut dianggap ringan dan cukup mendekati *state-of-the-art* (SOTA) saat ini. Optimizer yang digunakan adalah SGD pada Gabor Filter dan Feature Maps dan Adam pada Local Binary Pattern, yang dalam beberapa jurnal terbukti menjadi salah satu optimizer terbaik selain Adagrad. Sementara itu, *feature extraction* yang akan digunakan adalah *Gabor Filters*, *Local Binary Pattern* (LBP), dan *Feature Maps*. Proses evaluasi model ini akan mencakup perbandingan antara model yang menggunakan teknik *image processing* dan model tanpa teknik tersebut. Evaluasi ini akan dilakukan melalui metrik *confusion matrix*, *accuracy*, *precision*, *recall*, dan F1-Score.

Model akan menggunakan Arsitektur MobileNetV3Large yang akan di modif layer

atasnya dan layer bawahnya, dimana layer bawah akan diubah input size nya dari yang awalnya 224×224 diubah menjadi 560×560 , dikarenakan 224×224 dinilai terlalu kurang untuk menjadi input size untuk citra pada dataset ini. Layer atas pada dataset ini ditambahkan 2 layer lagi yaitu GlobalAveragePooling2D dan Dense layer dengan aktivasi softmax.

V. Experiments

a. Gabor Filter

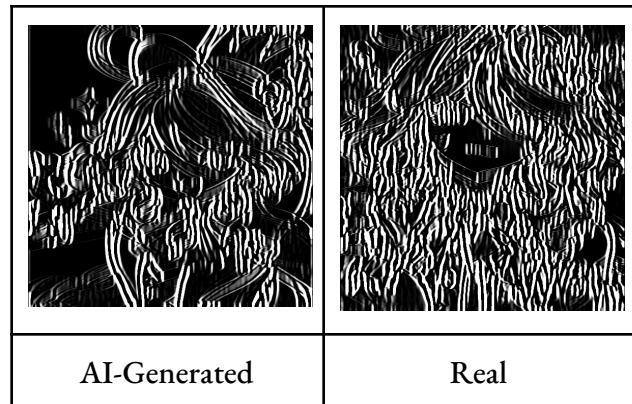
Hyperparameter Model:

- Epochs = 10
- Batch size = 64
- Loss = CategoricalCrossentropy
- Optimizer = SGD
- Learning rate = 0.001
- Momentum = 0.9

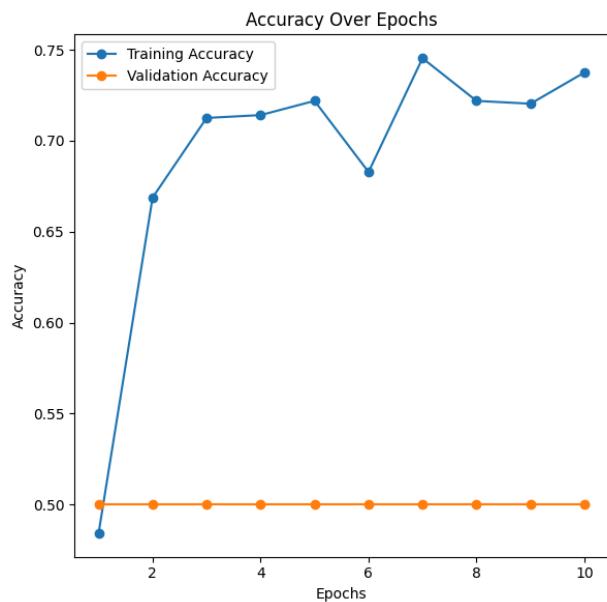
Eksperimentasi pada gabor filter dilakukan dengan mengaplikasikan gabor filter pada setiap citra dengan menggunakan library cv2, Parameter yang digunakan untuk mengaplikasikan Gabor Filter adalah sebagai berikut:

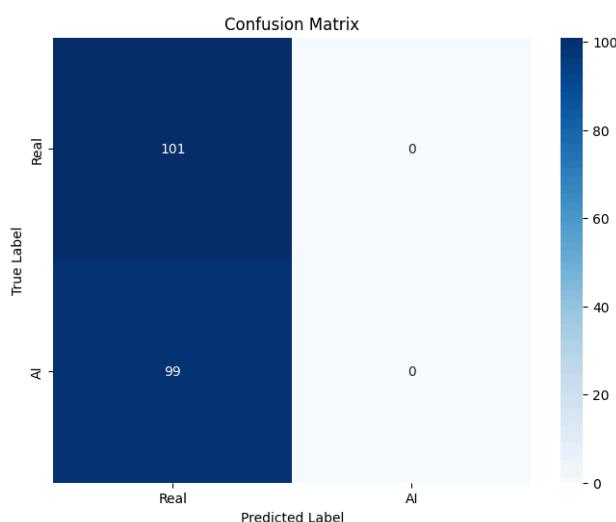
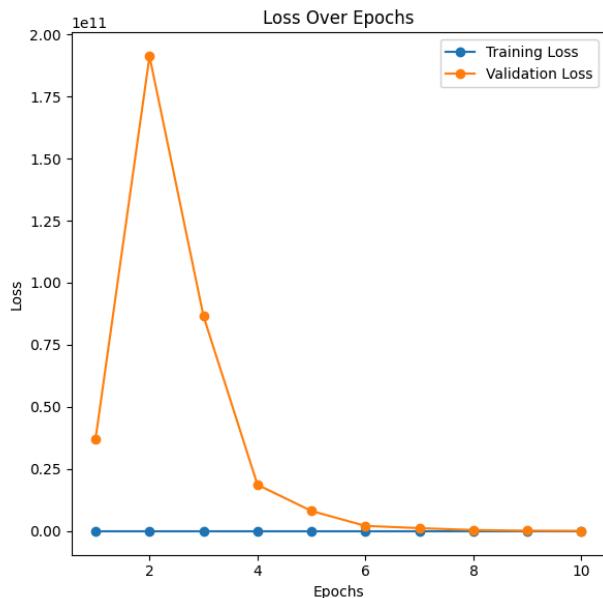
- Kernel size = 21
- Sigma = 5.0
- Theta = 0
- Lambda = 10.0
- Gamma = 0.5

Dengan hasil citra sebagai berikut:



Setelah melakukan pelatihan model, ditemukan hasil seperti dibawah ini:





	precision	recall	f1-score	support
Real	0.51	1.00	0.67	101
AI	0.00	0.00	0.00	99
accuracy			0.51	200
macro avg	0.25	0.50	0.34	200
weighted avg	0.26	0.51	0.34	200

Hasilnya cukup aneh, yaitu 51% akurasi dimana sepertinya model tidak bekerja dengan benar, entah masalah pada coding atau preprocessing atau memang MobileNetV3 tidak cocok menggunakan array untuk melatih

modelnya, hal itu juga yang menjadi alasan epochs nya hanya 10, karena dinilai tidak layak untuk dilanjutkan. Loss pada model juga menandakan model tidak cocok untuk tipe data ini karena keduanya mendekati 0.

b. Local Binary Pattern (LBP)

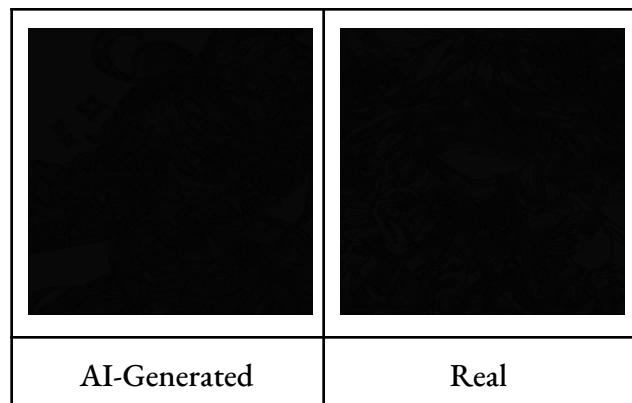
Hyperparameter Model:

- Epochs = 10
- Batch size = 64
- Loss = CategoricalCrossentropy
- Optimizer = Adam
- Learning rate = 0.01

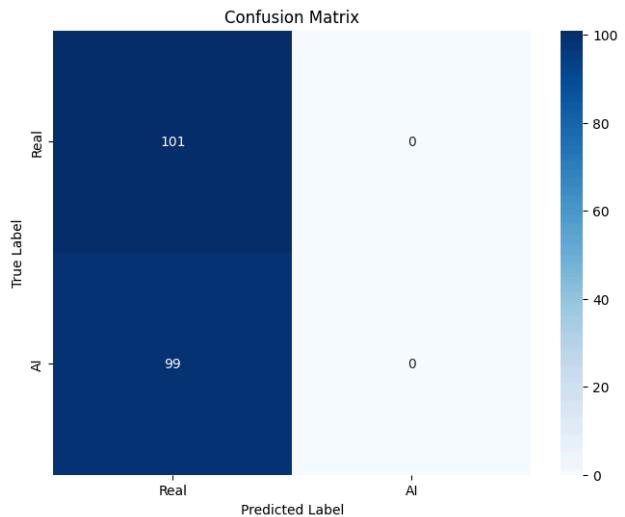
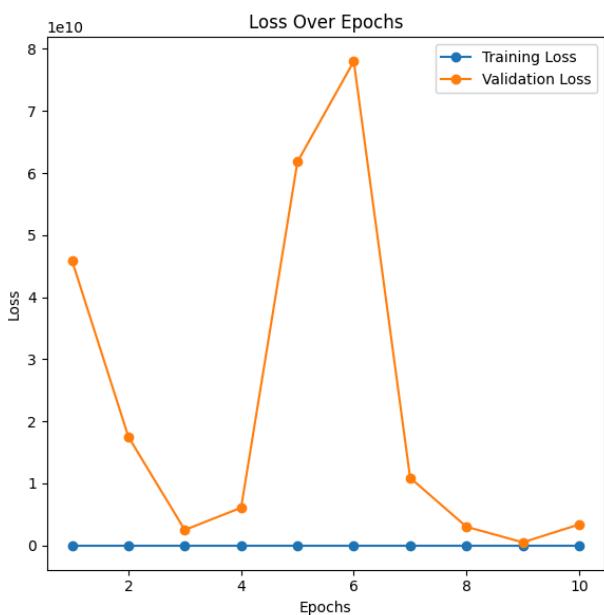
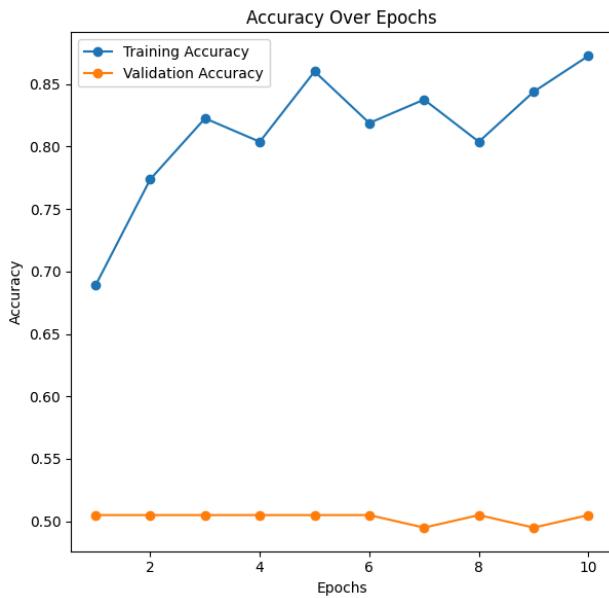
Eksperimentasi pada LBP dilakukan dengan mengaplikasikan LBP pada setiap citra dengan menggunakan library skimage.feature. Parameter yang digunakan untuk mengaplikasikan LBP adalah sebagai berikut:

- Color = gray
- P = 8
- R = 1
- Method = Uniform

Dengan hasil citra sebagai berikut:



Setelah melakukan pelatihan model, ditemukan hasil seperti dibawah ini:



	precision	recall	f1-score	support
Real	0.51	1.00	0.67	101
AI	0.00	0.00	0.00	99
accuracy			0.51	200
macro avg	0.25	0.50	0.34	200
weighted avg	0.26	0.51	0.34	200

Hasilnya sama dengan sebelumnya, yaitu akurasi 51% dimana sepertinya model tidak bekerja dengan benar, entah masalah pada coding atau preprocessing ataukah MobileNetV3 tidak cocok dalam menggunakan array untuk melatih modelnya, hal itu juga yang menjadi alasan epochs nya hanya 10, karena dinilai tidak layak untuk dilanjutkan. Hal yang sama juga terjadi pada loss di LBP seperti sebelumnya, diperlukan analisis lebih lanjut untuk mengatasinya.

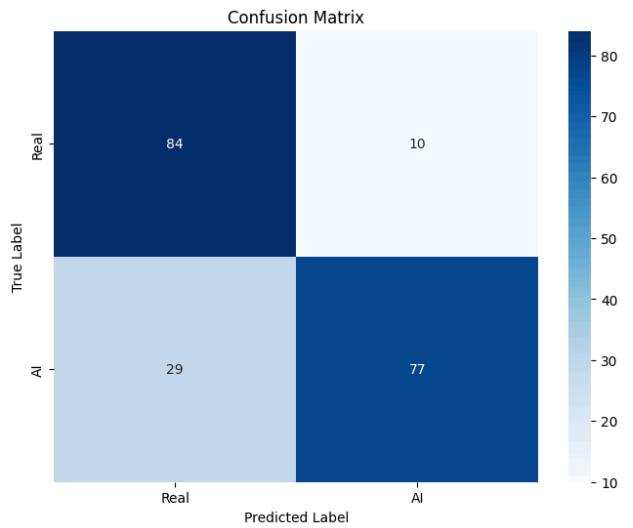
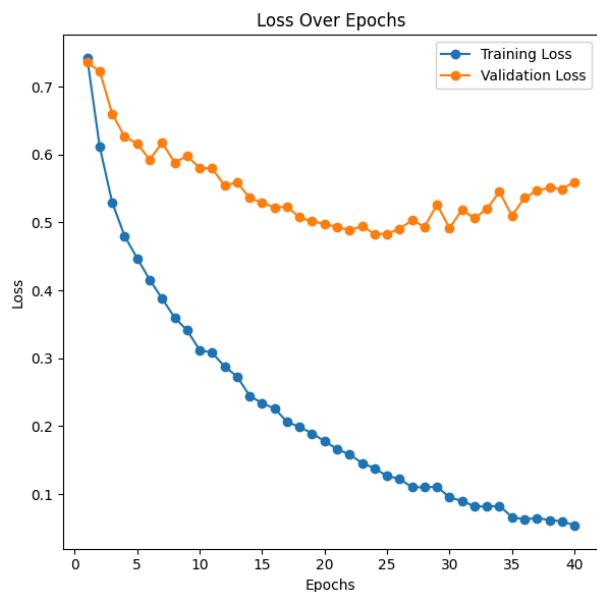
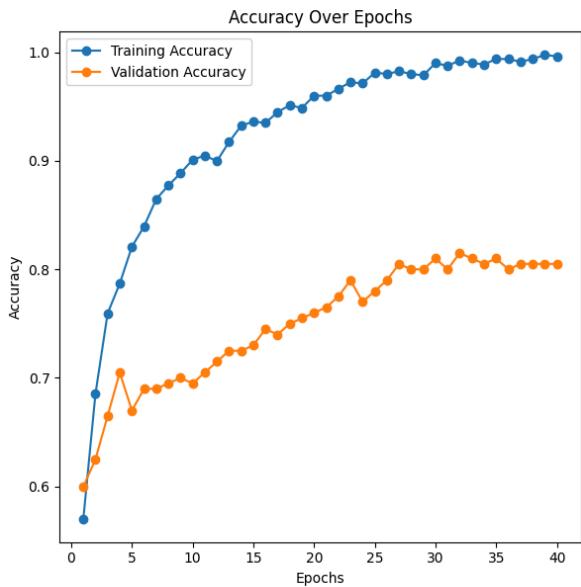
c. Feature Maps

Hyperparameter Model:

- Epochs = 40
- Batch size = 64
- Loss = CategoricalCrossentropy
- Optimizer = SGD
- Learning rate = 0.001
- Momentum = 0.9

Eksperimen pada feature maps atau default feature extraction dari CNN dilakukan dengan langsung melakukan pelatihan pada model setelah menggunakan library preprocess_input untuk menyiapkan citra dengan cara menormalisasikan dan melakukan hal lainnya.

Setelah melakukan pelatihan model, ditemukan hasil seperti dibawah ini:



	precision	recall	f1-score	support
Real	0.74	0.89	0.81	94
AI	0.89	0.73	0.80	106
accuracy			0.81	200
macro avg	0.81	0.81	0.80	200
weighted avg	0.82	0.81	0.80	200

Hasil dari model ini adalah hasil terbaik dari berbagai experiment yang dilakukan dengan akurasi 81%, dimana setelah dianalisa issue dari modelnya adalah dataset memprediksi real image dibanding AI, mungkin hal ini dapat diatasi dengan menambahkan dataset yang lebih banyak atau menambahkan weight pada prediksi AI, dengan memberatkan keputusan pada prediksi AI dibanding real. Jika diberikan 10 epochs lagi model mungkin akan mengalami stagnasi di kisaran 82%, dikarenakan loss dari training sudah mendekati 0 namun validation loss masih tinggi, hal ini kemungkinan besar adalah overfitting karena datasetnya yang masih kurang.

VI. Conclusion

Kesimpulan yang dapat ditarik dari tugas besar ini adalah tidak selalu feature extraction dapat membantu dalam klasifikasi object, dalam kasus ini hal itu terlihat jelas dengan betapa

kacaunya model yang dihasilkan Gabor Filter dan LBP, yang bisa saja diakibatkan oleh penggunaan CNN yang sebenarnya kurang tepat untuk input array, ataupun code yang salah, akan tetapi bukan berarti Gabor Filter dan LBP sudah lawas, namun bisa saja untuk gambar yang digunakan pada dataset ini penggunaan feature maps unggul dikarenakan pattern yang dapat dipelajari dari tiap citra tidaklah sama akibat banyaknya artstyle dan model generasi. Dibalik itu model MobileNetV3Large yang sedikit dimodif ini dapat mencapai hasil akurasi 81% walaupun mendekati overfitting, untuk mengatasinya dataset yang lebih banyak dapat meningkatkan akurasinya menjadi lebih akurat, dan penggunaan learning rate yang lebih rendah mungkin akan membantu pelatihan model.

References

- [1] D. C. Epstein, I. Jain, O. Wang, and R. Zhang, “Online Detection of AI-Generated Images,” -, pp. 382–392, Oct. 2023, doi: 10.1109/iccvw60793.2023.00045.
- [2] Bianco, Tommaso & Castellano, Giovanna & Scaringi, Raffaele & Vessio, Gennaro. (2023). Identifying AI-Generated Art with Deep Learning.
- [3] S. W. Kusuma, F. Natalia, C. S. Ko, and S. Sudirman, “Detection of AI-Generated Anime Images Using Deep Learning,” 革新的コンピューティング・情報・制御に関する速報 — B:応用, vol. 15. ICIC International学会, 2024. doi: 10.24507/icicelb.15.03.295.