

IMPLEMENTASI STABLE DIFFUSION DAN FINE-TUNING LOW RANK ADAPTATION UNTUK PEMBUATAN LOGO

Naela Mukaromah^{1*}, Sri Mulyono²

Universitas Islam Sultan Agung^{1,2}

naelamukaromah@std.unissula.ac.id^{1*} sri.m@unissula.ac.id²

Received: 15-01-2025

Revised: 30-01-2025

Approved: 11-02-2025

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem yang mampu menghasilkan logo berdasarkan deskripsi teks menggunakan model Stable Diffusion versi 1.5 dan teknik Fine-Tuning Low Rank Adaptation (LoRA). Studi literatur dilakukan untuk memahami pendekatan terbaik dalam implementasi text-to-image, model stable diffusion, dan metode evaluasi gambar. Dataset yang digunakan terdiri dari 50 gambar logo yang dikumpulkan dari berbagai sumber dan dilengkapi dengan image captioning. Dataset ini kemudian diproses melalui tahap preprocessing yang mencakup resize, standarisasi format, cropping, serta rotation dan flipping untuk meningkatkan variasi data. Pengembangan model melibatkan pelatihan Stable Diffusion 1.5 dengan fine-tuning LoRA, yang memungkinkan model beradaptasi dengan tugas spesifik tanpa memerlukan sumber daya komputasi yang besar. Proses pelatihan dilakukan dengan 400 iterasi untuk menghindari overfitting pada dataset yang terbatas. Pengujian sistem dilakukan dengan menghasilkan gambar berdasarkan teks deskripsi yang sama dengan caption data uji, kemudian dievaluasi menggunakan metrik CLIP-MMD dan Frechet Inception Distance (FID). Hasil penelitian menunjukkan bahwa model yang telah dilatih dapat menghasilkan gambar logo yang sesuai dengan deskripsi teks, dengan nilai CLIP-MMD dan FID yang menunjukkan kesesuaian gambar dengan dataset uji. Penelitian ini berkontribusi dalam pengembangan teknologi AI generatif untuk kebutuhan desain logo bagi pelaku UMKM, sehingga memungkinkan pembuatan logo secara efisien tanpa memerlukan keahlian desain grafis. Dengan penerapan teknik LoRA, penelitian ini juga membuktikan efektivitas fine-tuning dalam meningkatkan performa model generatif dengan sumber daya yang lebih rendah.

Kata Kunci: Text-to-Image, Stable Diffusion, LoRA Fine-Tuning, AI Generatif, Evaluasi Gambar

PENDAHULUAN

UMKM turut berperan dalam meratakan pendapatan dan meningkatkan kesejahteraan masyarakat di berbagai daerah, terutama di wilayah-wilayah yang masih memiliki tingkat kemiskinan yang tinggi(Halim, 2020). Tantangan yang sering dihadapi dibeberapa UMKM adalah keuangan terbatas dan kurangnya akses teknologi modern (Aftitah et al., 2025). Keterbatasan finansial seringkali menjadi penghalang bagi pelaku UMKM untuk mengembangkan dan memperkuat identitas merek mereka. Saat ini, kondisi produk UMKM di indonesia dalam memperkuat diri mereka atau branding masih beragam. Terdapat banyak UMKM yang berkualitas namun belum kuat dalam branding. Padahal potensi UMKM dalam mengembangkan usahanya melalui branding sangat memungkinkan, karena banyak bisnis kecil (UMKM) yang berubah menjadi brand besar karena bandingnya yang kuat. Meskipun terdapat beberapa UMKM yang berhasil dalam branding yang dilakukannya sendiri, ternyata masih banyak UMKM yang belum bisa melakuakan branding sendiri (Faiz Muntazori & Listya, 2021).

Teknologi artificial intellegence dapat dimanfaatkan untuk mengatasi permasalahan yang ada, teknologi ini diprediksi akan bertahan lama (Sony Maulana et al., 2023) dan memberikan dampak cukup besar (Arly et al., 2023) . Dengan AI dapat mengatasi kendala biaya dan sumber daya yang sering kali dihadapi dalam pembuatan identitas visual. AI memungkinkan pelaku UMKM untuk memproduksi identitas visual (logo) tanpa vendor. Sistem otomatis pembuat logo menggunakan AI dinilai efektif

(Cuevas et al., 2023). Logo merupakan salah satu bentuk identitas visual dari produk. Identitas visual produk bermanfaat untuk membedakan antara merk satu dengan lainnya. Logo digunakan untuk alat komunikasi yang membentuk identitas dari suatu perusahaan, lembaga, produk dan lain-lain (Aulia et al., 2021). Dengan adanya logo, orang akan mudah mengenali dan mengingat brand yang ditawarkan.

Text to Image merupakan metode yang mengubah deskripsi teks menjadi representasi visual, yang digunakan untuk menghasilkan gambar berdasarkan deskripsi teks yang diberikan. Model AI yang sering digunakan dalam proses generalisasi adalah GANs, Variational Autoencoders (VAEs), dan Diffusion Models. Masing-masing model tersebut memiliki peran penting dalam menghasilkan gambar yang realistik dan akurat dan mampu menggeneralisasi teks ke dalam gambar lebih mendetail. Hal ini menjadi salah satu alasan penggunaan dari model AI tersebut. Alasan penggunaan AI ini adalah karena menghasilkan gambar yang lebih realistik dan akurat (Ramzan et al., 2022). Diffusion Models adalah model AI yang melakukan proses generalisasi untuk menghasilkan gambar dari data yang bersifat acak menjadi gambar yang lebih jelas. Cara kerja Diffusion Models melalui proses pembalikan noise dari distribusi Gaussian yang terdiri dari dua tahap, *forward process* (menambahkan noise) dan *reverse process* (mengurangi noise). Metode ini dapat menghasilkan citra yang lebih realistik dengan kompleksitas komputasi lebih rendah, namun diperlukan penyetelan lebih lanjut untuk dataset agar distribusi dapat lebih representatif (Huang et al., 2024). Model ini terkenal karena dapat menghasilkan sampel berkualitas tinggi dan melakukan tugas-tugas seperti sintesis gambar dan penghilangan noise (Naseem, 2024).

Stable Diffusion 1.5, yang merupakan bagian dari diffusion models, adalah model kecerdasan buatan yang dikembangkan oleh Stability AI menggunakan teknik Latent Diffusion Model (LDM), dan memiliki kemampuan untuk mengkonversi teks menjadi gambar berkualitas tinggi. Struktur stable diffusion diantaranya terdiri dari U-Net, CLIP Text Encoder, dan Variational Autoencoder (VAE) (B. Liu et al., 2024). Dalam machine learning, fine tuning adalah teknik yang digunakan untuk memperbaiki kinerja model dengan cara bobot dan bias disesuaikan kembali pada model yang telah dilatih sebelumnya pada tugas baru. Pada teknik ini biasanya digunakan pada model yang dilatih pada dataset besar kemudian digunakan pada tugas pengenalan citra atau teks yang lebih kecil. Tahapan fine tuning dimulai dari pret-trained model atau mengambil model yang sudah dilatih pada dataset besar sebagai model dasar. Selanjutnya, lapisan terakhir model diganti untuk menyesuaikannya dengan tugas baru. Lapisan terakhir adalah lapisan keluaran atau kelas. Melatih ulang lapisan terakhir selama proses *fine-tuning* digunakan untuk memastikan bahwa lapisan lain tetap menggunakan bobot dan bias yang sudah diatur sebelumnya. (Pradana et al., 2024).

Penelitian yang dilakukan oleh Xiao pada tahun 2024, melakukan pembuatan sistem berbasis AI untuk membuat logo tipografi semantik. Sistem ini menggunakan generatif AI diffusion model dan *Contrastive Language Image Pre-training* CLIP untuk menggabungkan elemen gambar dan menghasilkan logo. Diffusion model digunakan untuk menghasilkan gambar penggabungan antara elemen huruf dan gambar, serta mendukung penggabungan pada berbagai tingkat kedekatan huruf. Sedangkan CLIP digunakan untuk menyelaraskan teks, gambar, dan hasil desain, serta mengevaluasi seberapa sesuai *output* dan *input* teks. Sistem yang dibuat disebut Typedance, dimana elemen huruf dan gambar digabungkan menjadikannya desain yang bermakna dan estetis. Hasil penelitian ini menunjukkan sistem yang dibuat dapat membantu desainer untuk menciptakan logo yang menggabungkan elemen tipografi dengan elemen

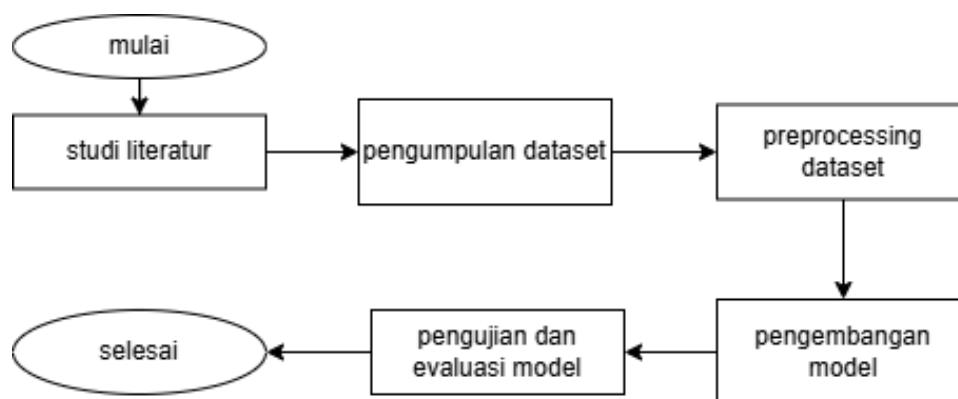
gambar(Xiao et al., 2024). Penelitian yang berjudul "*Text-to-Image Generation Using Deep Learning*" membahas teknologi untuk menghasilkan gambar dari deskripsi teks menggunakan model deep learning. Salah satu pendekatan yang digunakan adalah *Recurrent Convolutional Generative Adversarial Network* (RC-GAN). Teknologi ini memungkinkan komputer mengubah deskripsi teks menjadi gambar visual yang sesuai, seperti menciptakan foto bunga dari deskripsi teks. Model ini dilatih dengan dataset gambar bunga dan deskripsi teks yang relevan. RC-GAN terdiri dari dua bagian utama, yaitu generator yang membuat gambar dari teks dan discriminator yang memeriksa apakah gambar yang dihasilkan realistik. Eksperimen menunjukkan bahwa model ini dapat menghasilkan gambar dengan kualitas yang baik, mengungguli metode sebelumnya dalam beberapa metrik seperti Inception Score (IS) dan PSNR (*Peak Signal-to-Noise Ratio*) (Ramzan et al., 2022). Berdasarkan penelitian tersebut memungkinkan untuk pembuatan logo dari deskripsi teks, yang termasuk dari text to image

Adapun penelitian yang membahas penyesuaian (*fine-tuning*) model *Stable Diffusion* guna menghasilkan ikon bergaya tertentu untuk kebutuhan komersial, seperti ikon sekrup dan kabinet dapur. Penulis bereksperimen dengan berbagai cara pelatihan dan ukuran teks deskripsi, seperti teks pendek, teks panjang, dan penggunaan gambar referensi. Hasil yang didapatkan dalam penelitian ini model *stable diffusion* dapat disesuaikan untuk menghasilkan ikon dengan gaya tertentu dan berkualitas baik. Model yang dilatih menggunakan deskripsi teks pendek dan gambar referensi menghasilkan gambar yang lebih akurat secara teknis, namun hasilnya tidak selalu sesuai dengan ekspektasi manusia dalam hal detail dan konsistensi gaya. Sebaliknya, model dengan deskripsi teks panjang lebih mampu menghasilkan ikon yang sesuai dengan deskripsi dan lebih mendekati gaya yang diinginkan, meskipun skornya sedikit lebih rendah (Sultan et al., 2024). Dalam penerapannya, *Low Rank Adaption* (LoRA) muncul sebagai salah satu teknik *fine-tuning* yang dapat digunakan untuk menghasilkan logo dari teks. Salah satu studi yang menunjukkan potensi ini dilakukan oleh Mahendra pada tahun 2022, yang berhasil meningkatkan efisiensi model bahasa besar LLaMA 2-7B dengan teknik LoRA. Penulis memulai dengan mengumpulkan data berupa 60.500 artikel tentang kesehatan, pendidikan, sejarah, dan teknologi melalui teknik web scraping.

Setelah itu, data tersebut diproses agar bersih dari elemen yang tidak penting, memiliki panjang teks yang seragam, dan diubah menjadi bentuk numerik supaya bisa dipahami oleh komputer. Kemudian penulis mengaplikasikan LoRA untuk mengurangi ukuran parameter model sebesar empat kali lipat, sehingga model menjadi lebih ringan dan lebih cepat digunakan tanpa mengurangi kualitasnya. Hasilnya, model yang telah dioptimalkan dengan LoRA hanya membutuhkan memori GPU sebesar 4,74 GB, jauh lebih kecil dibandingkan model aslinya yang memerlukan 23,1 GB. Selain itu, model ini juga mampu memproses data hingga 8 kali lebih cepat, sehingga sangat hemat waktu dan daya komputasi (Mahendra, 2022). Kemudian pada penelitian yang berjudul "A LoRA is Worth a Thousand Pictures", metode yang digunakan adalah LoRA fine-tuning dan stable diffusion 1.5 untuk membahas hubungan antara bobot LoRA dan representasi gaya seni. Dataset yang digunakan berupa seni kecil, seperti karya seni dari genre seperti Art Nouveau, Baroque, Post-Impressionism, Renaissance, dan Ukiyo-e. Fine-tuning dilakukan pada u-net bagian self-attention, cross-attention, dan feed-forward neural network. Hasil penelitian menunjukkan bahwa LoRA dinilai merepresentasikan gaya seni lebih baik dibanding metode CLIP atau DINO. Selain itu, hasil clustering dan retrieval lebih presisi untuk pengelompokan seni(C. Liu et al., 2024).

Berdasarkan beberapa penelitian di atas, penulis dapat menyimpulkan bahwa *Stable diffusion* dan teknik *fine-tuning* Low-Rank Adaptation (LoRA) dapat dikombinasikan untuk membantu pembuatan logo dari teks. *Stable diffusion* berperan sebagai model generatif yang mampu mengubah deskripsi teks menjadi gambar dengan detail tinggi dan kualitas visual yang baik, sehingga memungkinkan pembuatan logo sesuai dengan deskripsi teks. LoRA berperan dalam proses pelatihan model. Dengan LoRA, model dapat dengan mudah disesuaikan untuk menambahkan gaya visual tertentu atau memenuhi kebutuhan spesifik logo tanpa harus melatih ulang seluruh model.

METODE PENELITIAN



Gambar 1. Alur metode penelitian

1. Studi Literatur

Melakukan kajian literatur terhadap beberapa penelitian yang relevan serta memahami pendekatan yang baik dalam proses pembuatan sistem yang akan dilakukan. Kajian literatur yang dilakukan terhadap penelitian-penelitian yang membahas mengenai text-to-image, model stable diffusion, LoRA fine-tuning, dan matriks evaluasi untuk gambar.

2. Pengumpulan dataset

Mengumpulkan dataset berupa gambar logo sebanyak 50 gambar. Dataset diperoleh dari berbagai platform, kemudian disimpan pada google drive. Pada pengumpulan dataset, dilakukan image captioning yang menghasilkan file txt. Image captioning adalah tahapan pemberian deskripsi pada gambar berdasarkan elemen yang terdapat didalamnya. Kemudian dataset tersebut dibagi menjadi data pelatihan dan data pengujian 40 data pelatihan 10 data pengujian. Dimana data pelatihan digunakan sistem untuk belajar sehingga dapat menghasilkan gaya yang sesuai dengan dataset, dan data pengujian digunakan sebagai pembanding.

3. Preprocessing dataset

Preprocessing dataset adalah mengolah data mentah menjadi data siap digunakan untuk training. Tujuan dilakukan preprocessing ini untuk meningkatkan variasi dataset dan mempermudah sistem untuk mengenali dataset karena data bersih dan konsisten. Dari data pelatihan yang sudah ada berupa gambar, selanjutnya di proses sebagai berikut :

- a. Resize adalah mengubah ukuran gambar yang ada secara keseluruhan, baik itu memperbesar atau memperkecil tanpa memotong. Tujuan dari adanya resize adalah untuk menyamakan resolusi dataset.
- b. Standarisasi format adalah menyamakan jenis ekstensi file untuk jenis ekstensi

- file untuk memastikan bahwa semua data dimasukkan dalam format yang sama. Tujuannya adalah untuk menghindari error pada saat data diproses.
- c. Cropping adalah proses memotong bagian tertentu pada gambar. Tujuan dari proses ini adalah untuk mengkonsentrasi perhatian pada bagian tertentu dari gambar dan menghilangkan bagian yang tidak penting.
 - d. Rotation dan flipping gambar, melakukan pemutaran dan pembalikkan pada gambar dataset yang ada. Jenis pembalikkan yang terjadi pada proses ini yaitu horizontal flip, vertical flip , dan diagonal flip. Bertujuan agar dataset lebih bervariasi.
4. Pengembangan model
- a. Stable diffusion 1.5
- Stable Diffusion 1.5, yang merupakan bagian dari diffusion models, yaitu model generasi untuk menghasilkan gambar berdasarkan teks. Cara kerja stable diffusion 1.5 yaitu dimulai dengan random noise, kemudian melewati U-Net untuk dihilangkan noise nya secara bertahap dan mulai berbentuk gambar berdasarkan teks yang diberikan. Pada tahap ini model membutuhkan beberapa langkah untuk menghilangkan dan memperbaiki noise, sehingga mencapai hasil akhir yang diinginkan.
- b. Fine tuning Low Rank Adaptation
- Pada tahapan ini dilakukan proses fine-tuning atau training terhadap model stable diffusion versi 1.5. Fine-tuning adalah teknik yang digunakan untuk memperbaiki kinerja model dengan cara bobot dan bias disesuaikan kembali pada model yang telah dilatih sebelumnya pada tugas baru. *Fine-tuning* bertujuan untuk melatih pre-trained model yang berada pada *top layer* dan berfungsi untuk mengenali feature dari dataset yang ada (Sasongko et al., 2023). Fine tuning memiliki kemampuan untuk mengurangi jumlah daya komputasi yang mahal dan data berlabel yang diperlukan untuk menghasilkan model besar yang dapat disesuaikan dengan berbagai kasus penggunaan.
- Pada penelitian ini, penulis menggunakan Low Rank Adaptation fine-tuning. Low Rank Adaptation fine-tuning adalah teknik untuk menyesuaikan model kecerdasan buatan (AI) yang sudah dilatih sebelumnya agar bisa digunakan untuk tugas atau bidang tertentu. Cara kerja LoRA dengan menambahkan menambahkan komponen sederhana (matriks kecil) ke dalam model, sehingga model bisa belajar hal-hal baru tanpa perlu mengubah strukturnya secara besar-besaran. Teknik ini memungkinkan pelatihan model menjadi lebih cepat dan hemat, karena hanya bagian kecil dari model yang dilatih ulang. Sementara itu, bagian utama model tetap tidak berubah, sehingga hasil sebelumnya tidak hilang. Dengan cara ini, LoRA dapat mengurangi kebutuhan perangkat keras (seperti memori atau GPU) hingga tiga kali lipat dibandingkan metode pelatihan biasa. Selain itu, LoRA menghasilkan kualitas yang sebanding dengan pelatihan penuh, tetapi dengan biaya dan waktu yang lebih efisien. Dengan menerapkan struktur low-rank pada pembaruan matriks, LoRA juga mampu menghasilkan hasil yang lebih baik, sekaligus mempertahankan atau bahkan menghasilkan kualitas yang setara dengan *fine-tuning* penuh (Hu et al., 2022).
5. Pengujian dan Evaluasi
- a. Pengujian
- Setelah melalui tahap pengembangan, selanjutnya adalah tahapan pengujian.

Dimana pada tahap ini sistem akan menghasilkan gambar dari model yang sudah dilatih. Untuk menghasilkan gambar, prompt yang digunakan berupa teks yang sama dengan caption dari data pengujian. Alasan penggunaan teks yang sama adalah untuk memudahkan proses perbandingan subjektif antara gambar hasil model baru dan gambar data pengujian.

b. Evaluasi

Pada evaluasi model, menggunakan matriks CLIP score dan FID (Frechet Inception Distance) dengan penjelasan sebagai berikut :

1) CLIP-MMD (CLIP Maximum Mean Discrepancy)

CLIP-MMD digunakan untuk menghitung kesesuaian gambar yanghasilkan terhadap deskripsi teks. Metriks evaluasi ini memanfaatkan embedding dari CLIP, model yang sudah dilatih untuk memetakan teks dan gambar sehingga dapat memahami hubungan kesesuaian antara teks dan gambar. Penggabungan embedding dari CLIP menggunakan konsep Maximum Mean Discrepancy. Cara kerja pada sistem yaitu, folder hasil generate model baru dihitung CLIP-MMD menggunakan deskripsi teks atau prompt yang digunakan saat menghasilkan gambar. Kemudian dibandingkan dengan gambar yang ada pada dataset pengujian. CLIP-MMD dengan nilai yang rendah menunjukkan bahwa gambar yang dihasilkan mendekati distribusi gambar referensi (dataset pengujian) yang menunjukkan bahwa kualitas gambar baik, teks deskripsi sesuai (Jayasumana et al., 2023).

2) FID (Frechet Inception Distance)

FID digunakan untuk mengukur kualitas visual gambar dari model generatif. FID menghitung jarak distribusi statistik antara fitur gambar yang dihasilkan model generatif dan gambar asli (Yu et al., 2021). Pada penerapan di sistem, gambar logo yang dihasilkan oleh model dan dataset uji (10 gambar logo) dihitung menggunakan FID. Apabila nilai yang dihasilkan FID rendah menunjukkan bahwa gambar yang dihasilkan lebih mirip dengan dataset uji, dengan demikian kualitas gambar lebih tinggi atau lebih bagus.

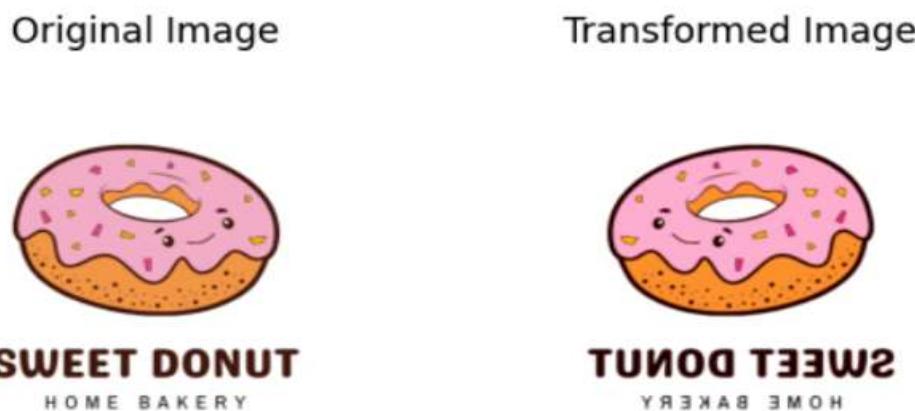
HASIL KEGIATAN DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan sistem yang mampu menghasilkan logo berdasarkan deskripsi teks menggunakan Stable Diffusion versi 1.5 dan Fine-Tuning Low Rank Adaption (LoRA). Dengan menggunakan model ini, pelaku UMKM diharapkan dapat membuat desain logo untuk produknya dengan lebih mudah tanpa memerlukan vendor. Kajian literatur yang dilakukan pada penelitian ini mencakup penelitian yang membahas mengenai teknologi AI generative models untuk menghasilkan gambar berdasarkan teks deskripsi. Beberapa penelitian terkait dijadikan rujukan oleh penulis, penelitian tersebut membahas tentang text to image, stable diffusion, LoRA fine-tuning. Dataset yang berhasil dikumpulkan oleh penulis berupa gambar logo yang diperoleh diberbagai platform. Dataset telah dibagi menjadi 2, data pelatihan dan data pengujian. Selain itu sudah dilakukan captioning dan dipasangkan dengan file .txt yang didalamnya berisi deskripsi teks berdasarkan gambar dengan maksimal teks 77 token, dalam bahasa inggris. Tabel 1 menunjukkan total dataset beserta dengan caption.

Tabel 1.
Jumlah Dataset

No	Dataset	Jumlah	Jumlah caption
1.	Data Pelatihan	40	40
2.	Data pengujian	10	10
	Total	50	50

Dataset yang berhasil diberikan caption selanjutnya dilakukan preprocessing untuk meningkatkan variasi dataset dan mempermudah sistem untuk mengenali dataset karena data bersih dan konsisten.



Gambar 2. Hasil preprocessing

Gambar 2 menunjukkan hasil preprocessing dataset yang sudah dilakukan preprocessing. Gambar sudah melalui tahapan seperti distandarisasi format, cropping, resize, Rotation dan flipping.

Pada tahap pengembangan model penulis telah melakukan fine-tuning pada model stable diffusion. Tujuan dilakukan proses ini adalah untuk melatih agar model dapat menghasilkan gambar yang sesuai dengan dataset. Adapun penulis melakukan uji coba terhadap 400 iterasi training pada model stable diffusion. Alasan penggunaan 400 iterasi agar tidak mengalami overfitting karena jumlah dataset tidak terlalu banyak.

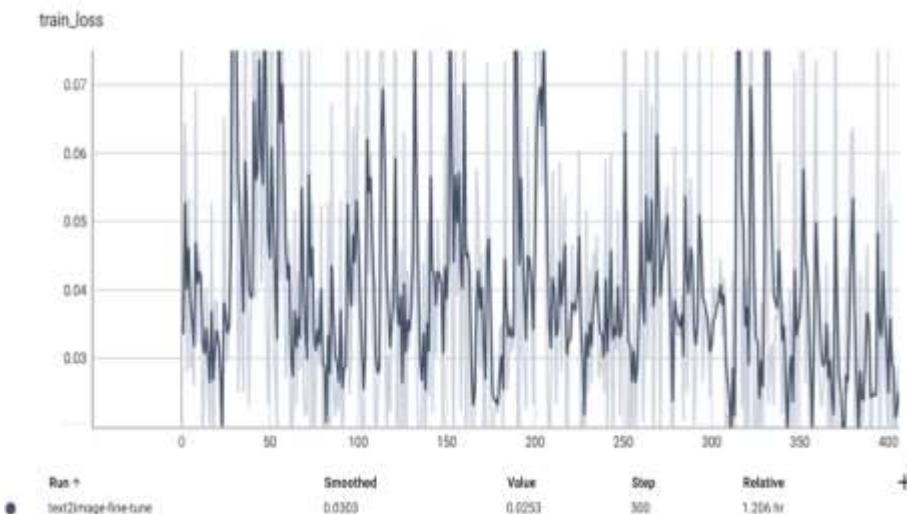
Tabel 2.

Parameter pelatihan model

Parameter pelatihan model

dataloader_num_workers	8
resolution	512
train_batch_size	1
gradient_accumulation_steps	4
max_train_steps	400
learning_rate	5e-04
max_grad_norm	1
lr_scheduler	cosine
lr_warmup_steps	0
checkpointing_steps	100
seed	1337

Tabel 2 menunjukkan parameter yang digunakan saat pelatihan model. Parameter tersebut berfungsi untuk mengontrol proses pelatihan model dan berpengaruh pada kesesuaian bobot terhadap data.



Gambar 3. Grarik Monitoring Loss

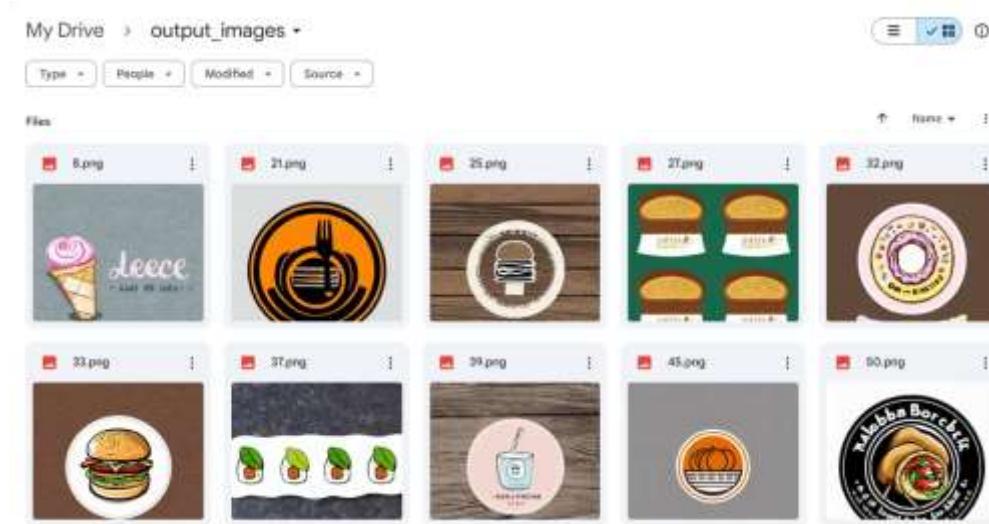
Gambar 3 menunjukkan grafik loss pada saat proses training berlangsung. Berdasarkan grafik tersebut, terlihat bahwa nilai loss mengalami penurunan secara bertahap beriringan dengan bertambahnya langkah iterasi. Ini menunjukkan bahwa proses training model berhasil mempelajari data yang ada. Namun juga terdapat kondisi dimana loss mengalami fluktuasi, dimana pada proses training nilai loss naik dan turun secara drastis. Ini menunjukkan adanya ketidakstabilan dalam proses model mempelajari data. Secara keseluruhan, grafik tersebut menunjukkan bahwa model belajar dengan baik, tetapi masih mengalami fluktuasi loss yang dapat mengakibatkan terhambatnya konvergensi model.



Gambar 4. Hasil Generate

Gambar 4 merupakan hasil generate dari model yang telah dilatih menggunakan teknik LoRA fine-tuning. Gambar tersebut menggunakan prompt "cartoon-style burger logo featuring a juicy cheeseburger with lettuce, tomato, and a sesame seed bun. Encircled text reads 'Delicious Burgers' in bold retro font". Secara umum, model berhasil menghasilkan suatu logo dengan bentuk dan warna yang cukup konsisten. Namun teks yang ditampilkan pada logo tidak terbaca dengan jelas.

Pada tahap pengujian, model akan melakukan generate gambar berdasarkan dataset pengujian. Model akan menggunakan teks deskripsi yang sama dengan dataset pengujian. Model akan menghasilkan 10 gambar, sesuai dengan dataset pengujian yang ada.

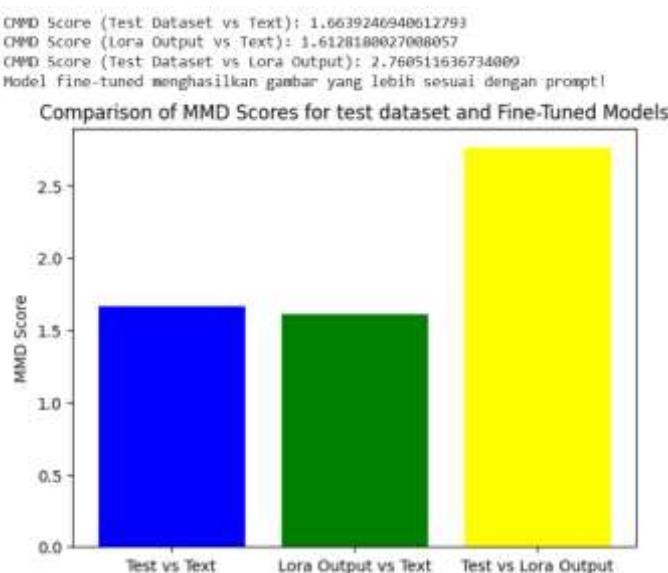


Gambar 5. Hasil pengujian

Gambar 5 menunjukkan hasil pengujian menggunakan model baru. Secara visual, gambar yang dihasilkan sudah menyerupai logo. Ini menunjukkan model mampu menghasilkan desain yang berasiasi. Namun, pada bagian teks model masih susah untuk mengenaliinya sehingga membuat teks yang dimunculkan tidak terbaca dengan baik. Pada bagian Evaluasi, gambar yang dihasilkan oleh model pada tahap pengujian dilakukan evaluasi. Evaluasi berfungsi untuk menilai seberapa sesuai dan bagus kualitas dari gambar yang dihasilkan model.

1) CLIP-MMD (CLIP Maximum Mean Discrepancy)

Score yang diperoleh ketika dilakukan evaluasi menggunakan CLIP-MMD direpresentasikan pada diagram dibawah.

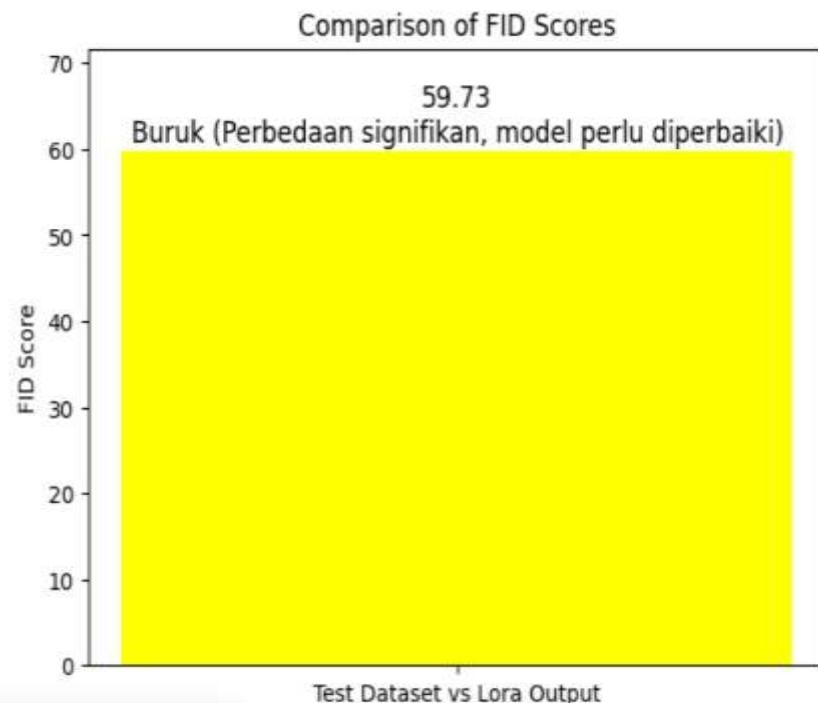


Gambar 6. Diagram hasil CLIP-MMD

Gambar 6 menunjukkan score yang diperoleh CLIP-MMD dalam menghitung kualitas gambar yang dihasilkan oleh model baru. Dataset pengujian mendapatkan score 1.663 sedangkan model baru menghasilkan score 1.612 . Hal ini menunjukkan bahwa gambar yang dihasilkan oleh model baru menghasilkan gambar yang lebih sesuai dengan teks deskripsi, meskipun hanya memiliki perbedaan yang tipis.

2) FID Score

FID Score (Test Dataset vs Lora Output): 59.73
Interpretasi: Buruk (Perbedaan signifikan, model perlu diperbaiki)



Gambar 7. Hasil evaluasi FID

Gambar 7 menunjukkan hasil matriks evaluasi menggunakan FID. Score yang diperoleh 59,73 dimana nilai tersebut menunjukkan bahwa model belum mampu menghasilkan gambar dengan kualitas yang baik dan model perlu diperbaiki dalam segi kualitas visual. Score yang masih kurang baik dapat disebabkan oleh beberapa faktor seperti kurangnya iterasi, jumlah dataset, dan parameter training kurang sesuai.

KESIMPULAN

Bawa model stable diffusion dapat dilakukan fine-tuning dengan teknik LoRA. Model yang di fine-tuning dapat menghasilkan gambar berupa logo berdasarkan deskripsi teks. Serta model dinilai dapat mempelajari pola dari dataset pelatihan yang ada dan menghasilkan gambar logo yang cukup baik. Hal ini ditunjukkan dengan hasil evaluasi menggunakan metriks CLIP-MMD yang memperoleh score 1,612. Dimana score tersebut menggambarkan bahwa hasil gambar oleh model baru sesuai dengan teks deskripsi yang ada. Namun, pada matriks evaluasi menggunakan FID, gambar yang dihasilkan memperoleh score 59,73 yang artinya kualitas gambar belum cukup baik. Selain itu model mengalami kesulitan untuk memunculkan teks pada gambar.

Kekurangan dalam penelitian ini seperti kualitas visual gambar yang dihasilkan dan akurasi teks, diharapkan dapat diperbaiki pada penelitian selanjutnya.

DAFTAR PUSTAKA

- Aftitah, F. N., K. J. L., Hasanah, K., Lailatul, N., Bina, U., & Informatika, S. (2025). *Pengaruh UMKM Terhadap Pertumbuhan Ekonomi Di Indonesia Pada Tahun 2023 Pemerintah mendukung UMKM melalui program seperti Kredit Usaha Rakyat (KUR), meskipun penyalurannya tahun 2023 belum memenuhi target. UMKM kini terus*. 3, 32–43.
- Arly, A., Dwi, N., & Andini, R. (2023). Implementasi Penggunaan Artificial Intelligence Dalam Proses Pembelajaran Mahasiswa Ilmu Komunikasi di Kelas A. *Prosiding Seminar Nasional*, 362–374.
- Aulia, F., Afriwan, H., & Faisal, D. (2021). Konsistensi Logo Dalam Membangun Sistem Identitas. *Gorga : Jurnal Seni Rupa*, 10(2), 439. <https://doi.org/10.24114/gr.v10i2.28131>
- Cuevas, E., Avalos, O., & Gálvez, J. (2023). *Analysis and Comparison of Metaheuristics*. 1063(November), 21–24.
- Faiz Muntazori, A., & Listya, A. (2021). Branding UMKM Produk Kopi Bang Sahal melalui Desain Logo. *SENADA : Semangat Nasional Dalam Mengabdi*, 1(3), 342–351.
- Halim, A. (2020). Pengaruh Pertumbuhan Usaha Mikro, Kecil dan Menengah Terhadap Pertumbuhan. *Ekonomi*, 2(September), 158.
- Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). Lora: Low-Rank Adaptation of Large Language Models. *ICLR 2022 - 10th International Conference on Learning Representations*, 1–26.
- Huang, P., Gao, X., Huang, L., Jiao, J., Li, X., Wang, Y., & Guo, Y. (2024). Chest-Diffusion: A Light-Weight Text-to-Image Model for Report-to-CXR Generation. *Proceedings - International Symposium on Biomedical Imaging*, 1–5. <https://doi.org/10.1109/ISBI56570.2024.10635417>
- Jayasumana, S., Ramalingam, S., Veit, A., Glasner, D., Chakrabarti, A., & Kumar, S. (2023). *Rethinking FID: Towards a Better Evaluation Metric for Image Generation*. <https://doi.org/10.1109/CVPR52733.2024.00889>
- Liu, B., Wang, C., Cao, T., Jia, K., & Huang, J. (2024). *Towards Understanding Cross and Self-Attention in Stable Diffusion for Text-Guided Image Editing*. <https://doi.org/10.1109/CVPR52733.2024.00747>
- Liu, C., Takikawa, T., & Jacobson, A. (2024). *A LoRA is Worth a Thousand Pictures*.
- Mahendra, A. (2022). *Implementasi low-rank adaptation of large language model (lora) untuk effisiensi large language model*. 8(3), 110–118.
- Naseem, B. (2024). *Temporal Scene Generation with Stable Diffusion*. Huggingface.Co.
- Pradana, A. G., Setiadi, D. R. I. M., & Muslikh, A. R. (2024). Fine tuning model Convolutional Neural Network EfficientNet-B4 dengan augmentasi data untuk klasifikasi penyakit kakao. *Journal of Information System and Application Development*, 2(1), 01–11. <https://doi.org/10.26905/jisad.v2i1.11899>
- Ramzan, S., Iqbal, M. M., & Kalsum, T. (2022). Text-to-Image Generation Using Deep Learning †. *Engineering Proceedings*, 20(1), 1–6. <https://doi.org/10.3390/engproc2022020016>
- Sasongko, T. B., Haryoko, H., & Amrullah, A. (2023). Analisis Efek Augmentasi Dataset dan Fine Tune pada Algoritma Pre-Trained Convolutional Neural Network (CNN). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(4), 763–768. <https://doi.org/10.25126/jtiik.20241046583>

- Sony Maulana, M., NurmalaSari, Rheno Widianto, S., Dewi Ayu Safitri, S., & Maulana, R. (2023). Pelatihan Chat Gpt Sebagai Alat Pembelajaran Berbasis Artificial Intelligence Di Kelas. *Jurnal Penelitian Dan Pengabdian Masyarakat Jotika*, 3(1), 16–19. <https://doi.org/10.56445/jppmj.v3i1.103>
- Sultan, Y., Ma, J., & Liao, Y.-Y. (2024). *Fine-Tuning Stable Diffusion XL for Stylistic Icon Generation: A Comparison of Caption Size*.
- Xiao, S., Wang, L., Ma, X., & Zeng, W. (2024). TypeDance: Creating Semantic Typographic Logos from Image through Personalized Generation. *Conference on Human Factors in Computing Systems - Proceedings*, 1–24. <https://doi.org/10.1145/3613904.3642185>
- Yu, Y., Zhang, W., & Deng, Y. (2021). Frechet inception distance (fid) for evaluating gans. *Researchgate.Net, September*, 1–7.