

PEMBUATAN LANSKAP FANTASI BERBASIS *DIFFUSION MODELS* DENGAN *FINE-TUNING LOW-RANK ADAPTATION (LORA)*

Tsabit Gholib^{1*}, Sri Mulyono²
Universitas Islam Sultan Agung^{1,2}
tsabitg@std.unissula.ac.id sri.m@unissula.ac.id²

Received: 18-01-2025

Revised: 06-02-2025

Approved: 11-02-2025

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model generatif berbasis *diffusion models* dalam menghasilkan gambar lanskap fantasi dari deskripsi teks. Metode penelitian yang digunakan meliputi studi literatur, pengumpulan dataset, preprocessing dataset, pelatihan model, serta pengujian dan evaluasi model. Studi literatur dilakukan untuk memahami konsep *diffusion models*, pemetaan teks ke gambar, serta teknik *fine-tuning* yang relevan. Pengumpulan dataset mencakup akuisisi gambar lanskap fantasi dari berbagai sumber daring, serta proses captioning untuk menghubungkan gambar dengan deskripsi teks. Dataset kemudian dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Tahap preprocessing dataset mencakup berbagai teknik augmentasi seperti *resize*, *random resized crop*, *horizontal flip*, *color jitter*, dan *normalization* untuk meningkatkan variasi serta kualitas data pelatihan. Model *Stable Diffusion 2.1* digunakan sebagai dasar pengembangan, dengan *fine-tuning* menggunakan *Low-Rank Adaptation (LoRA)* untuk meningkatkan efisiensi komputasi dan adaptasi terhadap dataset spesifik. Evaluasi model dilakukan menggunakan metrik kuantitatif seperti *CLIP-MMD* dan *FID*, serta penilaian kualitatif berbasis analisis visual. Hasil penelitian menunjukkan bahwa model yang telah dilatih mampu menghasilkan gambar lanskap fantasi yang sesuai dengan deskripsi teks, dengan peningkatan akurasi dan detail visual dibandingkan model dasar. Teknik *LoRA* terbukti efektif dalam meningkatkan spesifikasi visual dengan efisiensi komputasi yang lebih baik. Simpulan, penelitian ini memberikan kontribusi dalam pengembangan model generatif yang lebih adaptif untuk aplikasi *text-to-image*, khususnya dalam pembuatan lanskap fantasi.

Kata Kunci: *Diffusion Models, Stable Diffusion, Fine-Tuning, Lora, Text-To-Image, Lanskap Fantasi*

PENDAHULUAN

Dalam pembuatan konten kreatif, khususnya video *game-*, desain lingkungan dunia imajiner memainkan peran yang sangat krusial dalam membangun suasana dan memperkuat narasi. Seiring dengan perkembangan zaman, Jumlah pemain video game dari tahun ke tahun terus meningkat, mencakup berbagai kalangan mulai dari anak-anak hingga orang dewasa. Pada tahun 2024, jumlah pemain game diseluruh dunia mencapai 3.32 miliar (Howarth, 2024), yang mencerminkan pesatnya perkembangan industri video game. Tidak mengherankan pendapatan dari sektor video game sangat tinggi. Pendapatan dari video game di Indonesia sendiri diperkirakan mencapai 3,667 Juta USD di tahun 2024. Tentu ini menggambarkan betapa tingginya potensi pasar industri game di Indonesia (Purwanto et al., 2024). Hal ini ini membuat industri video game berlomba-lomba menciptakan pengalaman bermain yang tidak hanya menyenangkan, tetapi juga menarik dan imersif. Hal ini sangat terlihat pada genre fantasi, yang memberikan ruang bagi pengembang untuk mengekspresikan imajinasi dan kreativitas dalam menciptakan dunia baru yang memukau melalui *concept art* yang diciptakan.

Salah satu elemen penting dalam *concept art* video game adalah lanskap, yang tidak hanya berfungsi sebagai latar belakang, tetapi juga sebagai elemen yang memperkuat cerita dan juga *gameplay*. Dalam game bergenre fantasi, lanskap berperan penting dalam membangun atmosfer dan memberikan pemain rasa keajaiban serta

keterhubungan dengan dunia yang mereka jelajahi. Hal ini terlihat jelas pada beberapa nominasi *Game of the Year 2024*, seperti "*Astro Bot*", "*Black Myth: Wukong*", dan "*Final Fantasy VII Rebirth*", di mana lanskap fantasi tidak hanya menjadi elemen estetis, tetapi juga menghidupkan cerita dan pengalaman bermain (Awards, 2024).

Pembuatan lanskap fantasi dituntut dalam penggabungan elemen visual fantastis, seperti elemen mitologi atau pemandangan luar biasa yang tidak ada di dunia nyata, dengan kebutuhan fungsional yang mendalam (Afshar et al., 2022). Tantangan besar yang dihadapi industri video game adalah menciptakan *concept art* yang tidak hanya estetis tetapi juga fungsional (Bahia, 2023). Tantangan tersebut meliputi memastikan kejelasan visual, menyeimbangkan detail artistik dengan performa teknis, serta mengatasi keterbatasan waktu dan anggaran produksi. Desain dunia yang buruk dapat mengganggu keterlibatan pemain, sementara desain yang terlalu rumit dapat menurunkan performa teknis. Selain itu, salah satu tantangan utama dalam desain lanskap fantasi adalah memastikan konsistensi artistik dalam setiap elemen lingkungan, karena perbedaan gaya atau pencahayaan yang tidak sesuai dapat merusak imersif pemain. Oleh karena itu, seniman dan desainer harus menciptakan keseimbangan antara imajinasi kreatif dan kebutuhan teknis agar dunia fantasi terasa hidup dan imersif (Mehrafrooz, 2024).

Teknologi *text-to-image* telah menjadi topik utama dalam perkembangan kecerdasan buatan, terutama dalam upaya menghasilkan gambar dari deskripsi teks dengan presisi dan estetika yang tinggi (Zhou et al., n.d.). Kemampuan ini memungkinkan model kecerdasan buatan menerjemahkan ide-ide konseptual menjadi visualisasi yang nyata, sebuah lompatan besar dalam aplikasi *Computer Vision* dan NLP (Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, 2024). Pada awalnya, teknik seperti *Generative Adversarial Networks* (GANs) digunakan untuk tugas ini, namun GANs memiliki kelemahan, seperti ketidakstabilan selama pelatihan dan akumulasi kesalahan pada gambar yang kompleks. Dalam beberapa tahun terakhir, *Diffusion models* muncul sebagai alternatif yang lebih stabil dan efektif dibandingkan GANs untuk generasi gambar berbasis teks. Penelitian sebelumnya menunjukkan bahwa *Diffusion models* lebih unggul dibandingkan model generatif lainnya, seperti *Generative Adversarial Networks* (GANs), dalam hal stabilitas dan kualitas gambar. GANs sering menghadapi masalah ketidakstabilan pelatihan dan *mode collapse*, sedangkan *Diffusion models* mampu menghasilkan gambar yang lebih realistis dan kaya akan detail, terutama ketika menangani latar belakang dan tekstur yang kompleks (Peng, 2024).

Meskipun *diffusion model* kuat dalam menghasilkan gambar, diperlukan teknik tambahan untuk mencapai personalisasi yang lebih dalam yaitu berupa teknik *fine-tuning*. Salah satu pendekatan yang menonjol adalah *Low-Rank Adaptation* (LoRA). LoRA memungkinkan penambahan fitur atau gaya baru dengan cara yang hemat sumber daya, menjadikannya ideal untuk kebutuhan spesifik seperti pembuatan komik dengan gaya artistik tertentu (Jin & Song, 2023). Sebagai contoh, penerapan LoRA pada *stable diffusion* memungkinkan model menghasilkan gambar dengan detail gaya tertentu yang lebih tajam dan koheren setelah proses *fine-tuning*. Hasil pengujian menunjukkan bahwa DataInf dapat meningkatkan performa model dibandingkan model awal, terutama dalam metrik visualisasi FID yang lebih rendah, menunjukkan peningkatan kualitas gambar (Kwon et al., 2024). Selain itu, integrasi *Diffusion models* dengan teknik *fine-tuning* seperti LoRA mendukung kebutuhan personalisasi dalam industri kreatif.

Dalam konteks pembuatan lanskap fantasi, personalisasi ini memungkinkan model menyesuaikan *output* berdasarkan gaya seniman atau estetika tertentu yang diinginkan

pengguna, seperti dunia magis ala cerita rakyat atau latar futuristik. Pengguna dapat menggabungkan elemen-elemen unik seperti pola warna tertentu, tata cahaya spesifik, hingga bentuk geometri yang tidak biasa, sehingga menghasilkan visualisasi yang lebih kaya dan imersif (Wang et al., 2024). Penggunaan *Diffusion models* dengan *fine-tuning* melalui LoRA menawarkan pendekatan yang menjanjikan untuk menghasilkan lanskap fantasi secara otomatis dari deskripsi teks. Teknik ini tidak hanya memungkinkan kontrol yang lebih besar atas elemen visual, tetapi juga dapat mengurangi kebutuhan komputasi dibandingkan dengan pelatihan ulang model besar dari awal (Wallis, n.d.). Oleh karena itu, penerapan LoRA pada *Diffusion models* dapat menjadi solusi ideal dalam menciptakan visualisasi lanskap fantasi berkualitas tinggi, membuka peluang baru dalam industri kreatif, *game*, dan hiburan.

METODE PENELITIAN

Studi Literatur

Studi literatur dilakukan dengan tahapan sebagai berikut ini, pertama pencarian literatur dilakukan melalui jurnal, buku, dan sumber teroercaya lain mengenai *diffusion models*, dengan fokus pada konteks model generatif dan aplikasi *text-to-image*. Selanjutnya, kajian literatur tentang pemetaan deskripsi teks ke gambar dan model-model yang digunakan dalam menciptakan lanskap fantasi. Tahapan berikutnya adalah mengidentifikasi teknik *fine-tuning* yang relevan, serta metode evaluasi yang digunakan

Pengumpulan Dataset

Pada tahap pengumpulan dataset, gambar-gambar lanskap fantasi dikumpulkan dari situs-situs seperti Civitai, Freepik, dan DeviantArt, yang diunduh secara langsung dengan batasan tertentu. Gambar-gambar tersebut beragam dalam hal format dan ukuran. Selanjutnya, dilakukan *captioning* untuk memberikan deskripsi teks pada masing-masing gambar, yang menghasilkan pasangan gambar dan teks secara berurutan. Proses ini penting dalam menyediakan data yang dapat digunakan untuk pelatihan model yang mempelajari hubungan antara gambar dan teks. Setelah gambar dan caption disiapkan, dataset dibagi menjadi dua bagian utama yaitu 80% untuk data pelatihan dan 20% untuk data pengujian yang selanjutnya disimpan dalam *file* berbentuk csv. Penyimpanan dalam format CSV memudahkan penggunaan data di berbagai platform analisis dan memastikan data dapat diakses dengan mudah (Agung et al., 2023). Dataset pelatihan digunakan untuk melatih model dalam mengenali pola antara gambar dan caption, sementara dataset pengujian digunakan untuk mengukur kinerja model dalam memprediksi *caption* berdasarkan gambar baru yang belum dilihat sebelumnya. Pada Tabel 1 menunjukkan jumlah data yang digunakan pada penelitian ini. Terdiri dari dataset pelatihan dan dataset pengujian

Tabel 1.
Pembagian Dataset

No.	Dataset	Jumlah gambar	Jumlah caption
1.	Dataset Pelatihan	40	40
2.	Dataset Pengujian	10	10
	Total	50	50

Preprocessing Dataset

Preprocessing dataset atau dikenal juga sebagai pemrosesan awal data adalah Serangkaian langkah atau proses yang dilakukan pada data sebelum data tersebut

digunakan dalam analisis atau pemodelan (Putra et al., 2024). Tujuan utama dari *Preprocessing dataset* adalah untuk meningkatkan kualitas data, menjamin hasil analisis akurat, dan memperbaiki masalah serta kesalahan yang mungkin muncul pada data mentah. Pada tahap ini dilakukan *preprocessing* data pelatihan yang mencakup data *augmentation* dan *standarization* untuk meningkatkan variasi dan memastikan skala data seragam. Langkah-langkahnya adalah :

- 1) *Resize* : mengubah ukuran gambar ke resolusi tertentu agar seragam
- 2) *Random Resized Crop* : memotong gambar secara acak untuk variasi framing objek
- 3) *Random Horizontal Flip* : membalik gambar secara horizontal dengan probabilitas 50%
- 4) *Color Jitter* : mengubah kecerahan, kontras, dan saturasi gambar secara acak
- 5) *Normalization* : menormalkan nilai piksel ke rentang [-1, 1] dengan mean 0.5 dan standar deviasi 0.5

Proses ini bertujuan untuk meningkatkan generalisasi model dan mempercepat proses pelatihan

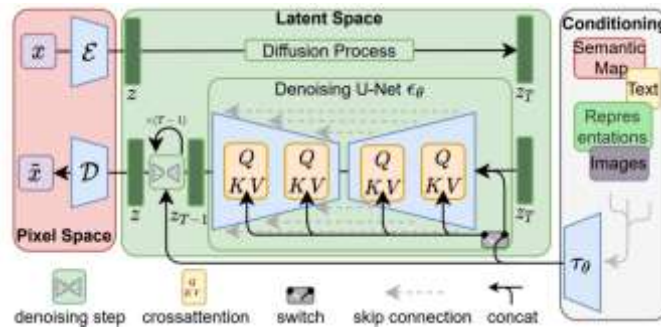
1. Pelatihan Model

Pada penelitian ini, model *Stable Diffusion 2.1* digunakan sebagai dasar pengembangan. *Stable Diffusion* adalah model pembelajaran mesin berbasis latent diffusion, yang mampu menciptakan gambar dengan kualitas tinggi dari deskripsi teks. Untuk meningkatkan spesifikasi visual yang relevan dengan tujuan penelitian, model ini di *fine-tune* menggunakan teknik LoRA.

a. *Stable Diffusion 2.1*

Stable diffusion merupakan salah satu model generatif yang menggunakan arsitektur berbasis *diffusion models* untuk menghasilkan gambar dari deskripsi teks. Model ini bekerja dengan melalui *forward diffusion* dan *reverse diffusion*, yang dimulai dengan gambar *noise* acak dan secara bertahap menyempurnakannya hingga menghasilkan gambar yang sesuai dengan *prompt* yang diberikan. Proses ini tidak hanya menghasilkan gambar yang realistis, tetapi juga memberikan kontrol atas hasil akhir, menjadikannya alat yang sangat bermanfaat dalam berbagai aplikasi industri, seperti permainan, film, dan periklanan (Valvano et al., 2024). Pada versi 2.1, peningkatan akurasi dan fleksibilitas model dalam berbagai domain visual telah dilakukan. *Stable diffusion v2.1* digunakan karena model ini terbuka (*open-source*) dan dapat diakses, berbeda dengan Midjourney yang bersifat tertutup dan hanya tersedia melalui layanan cloud. Dalam pengembangannya, model generatif *Stable diffusion* bekerja pada ruang laten, memungkinkan peningkatan efisiensi komputasi tanpa mengorbankan kualitas gambar (Blattmann et al., 2023).

Pada gambar 1 merupakan arsitektur dari *Stable Diffusion* dimana arsitektur ini menggabungkan teknik *autoencoder* untuk mengompresi data gambar ke representasi laten berdimensi lebih rendah, sehingga mengurangi kebutuhan komputasi. Pada tahap generatif, model memanfaatkan proses *denoising* bertahap dalam ruang laten menggunakan jaringan *U-Net* untuk merekonstruksi gambar secara iteratif. Proses ini dipandu oleh *text encoder* berbasis transformer seperti CLIP yang menghubungkan deskripsi teks dengan distribusi laten, memastikan hasil yang konsisten dengan input deskripsi. Pendekatan ini menghasilkan keseimbangan antara efisiensi dan kualitas gambar yang tinggi, menjadikannya inovasi penting dalam domain *text-to-image generation*



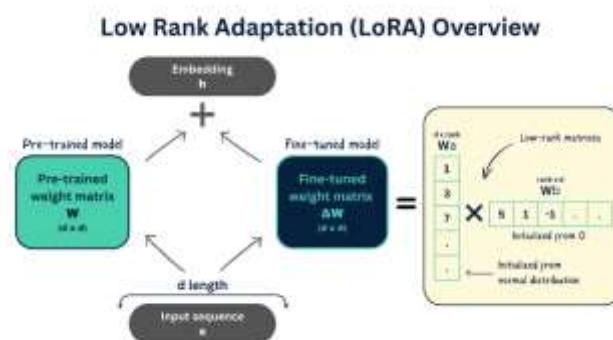
Gambar 1. Arsitektur *Stable Diffusion*

b. *Fine tuning Low-Rank Adaptation (LoRA)*

Dalam *Machine Learning*, *fine-tuning* adalah teknik yang meningkatkan kinerja model dengan menyesuaikan kembali bobot dan bias dari model yang telah dilatih sebelumnya pada tugas baru. Teknik ini biasanya digunakan pada model yang dilatih pada dataset besar sebelum digunakan pada tugas pengenalan citra atau teks yang lebih kecil (Pradana et al., 2024).

Teknik *Fine tuning Low-Rank Adaptation (LoRA)* merupakan teknik *fine tuning* yang digunakan untuk melakukan penyesuaian pada model tanpa perlu melatih ulang seluruh parameter model yang besar. LoRA bekerja dengan menambahkan matriks tingkat rendah pada lapisan tertentu model, sehingga adaptasi terhadap data pelatihan baru menjadi lebih efisien, beban komputasi dan memori yang diperlukan lebih ringan dibandingkan *fine tuning* penuh, serta spesifikasi visual tertentu dapat ditingkatkan sesuai dengan dataset atau kebutuhan aplikasi spesifik.

Pada gambar 2 merupakan arsitektur dari LoRA, pada arsitektur ini, parameter tambahan dengan rank rendah disisipkan ke dalam lapisan model asli, biasanya pada lapisan U-NET untuk menangkap informasi spesifik dari data baru. LoRA bekerja dengan mempertahankan bobot asli sambil menambahkan matriks peringkat rendah yang dapat dilatih, sehingga adaptasi model dapat dilakukan dengan lebih hemat memori dan sumber daya komputasi (Luo et al., 2024).



Gambar 2. Arsitektur *Low-Rank Adaptation (LoRA)*

2. Pengujian dan Evaluasi Model

a. Pengujian

Pada tahap ini dilakukan pengujian untuk membuat gambar dari model yang sudah dilatih, pengujian ini menggunakan kumpulan teks dari gambar pada data pengujian yang belum pernah dilihat model selama proses pelatihan, sehingga menghasilkan jumlah gambar yang sama dengan data pengujian

menggunakan teks yang sama.

b. Evaluasi

Metode Evaluasi yang digunakan pada tahap ini menggunakan kombinasi metrik kuantitatif dan penilaian kualitatif.

1) Metrik Kuantitatif

a) CLIP-MMD (*CLIP Mean Maximum Discrepancy*)

Memanfaatkan model CLIP (*Contrastive Language-Image Pretraining*), CLIP-MMD adalah metrik yang dirancang untuk mengevaluasi kesesuaian semantik antara teks input dan gambar yang dihasilkan. CMMD mengukur jarak antara distribusi embedding gambar yang dihasilkan oleh model dengan distribusi embedding gambar dari dataset referensi (Jayasumana et al., 2023).

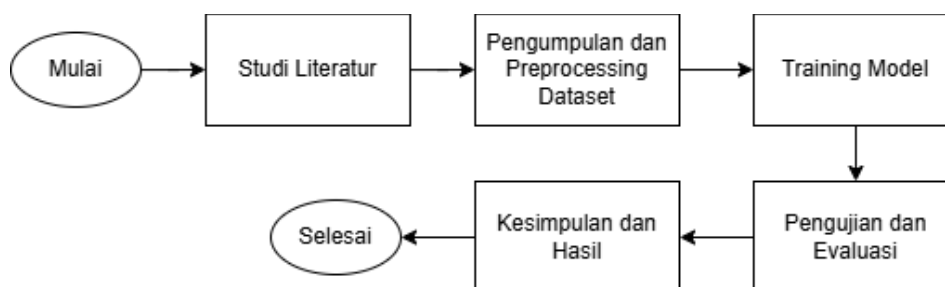
Nilai CLIP-MMD yang lebih rendah menunjukkan bahwa gambar yang dibuat lebih semantik relevan dengan teks yang dimasukkan.

b) FID (*Fréchet Inception Distance*)

FID adalah metrik standar yang digunakan untuk menilai kualitas visual gambar yang dibuat oleh model generatif ini mengukur kesamaan distribusi statistik antara gambar dari data pengujian dan gambar hasil model baru. FID menghitung jarak Frechet antara vektor fitur gambar. Ini diperoleh dengan menggunakan model yang diatur, biasanya *inception-v3*. Nilai FID yang lebih rendah menunjukkan bahwa distribusi gambar yang dihasilkan dari model lebih mirip dengan distribusi gambar aslinya (Yu et al., 2021).

2) Metrik Kualitatif

Hasil gambar dari data pengujian dan hasil gambar dari model yang sudah dilatih akan dibandingkan menggunakan visual mata telanjang untuk membandingkan kualitas visual secara langsung. Perbandingan ini melibatkan analisis kualitas gambar dari segi detail, ketajaman, tekstur, dan keakuratan representasi visual yang sesuai dengan deskripsi teks yang diberikan. Gambar 3 merupakan alur flowchart metode penelitian yang dilakukan :



Gambar 3. Flowchart Metode Penelitian

HASIL DAN PEMBAHASAN

Preprocessing Dataset

Pada tahap *preprocessing*, dataset yang digunakan dalam penelitian ini telah melalui berbagai macam tahapan, seperti pengumpulan data, *captioning* dan pembagian data untuk pelatihan dan pengujian.

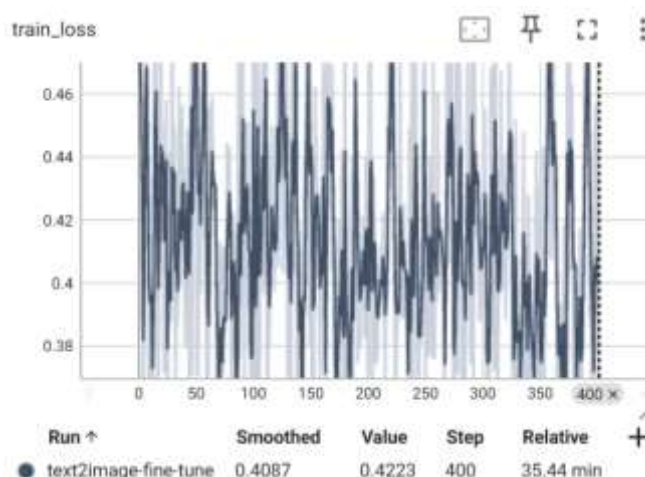
Gambar 4 menunjukkan contoh gambar setelah proses *preprocessing* pada dataset pelatihan. Tujuan dari proses ini adalah untuk meningkatkan kualitas data dan akurasi model selama pelatihan.



Gambar 4. Hasil *Preprocessing Dataset*

Pelatihan Model

Pada tahap *training*, model *Stable Diffusion 2.1* yang telah melalui *fine-tuning* menggunakan teknik LoRA menunjukkan peningkatan kualitas gambar, terutama dalam hal kesesuaian dengan deskripsi teks dan detail visual yang lebih tajam, sesuai dengan tema lanskap fantasi. Proses *fine-tuning* bertujuan untuk meningkatkan kemampuan model dalam menghasilkan gambar yang lebih akurat dan relevan dengan deskripsi teks yang diberikan. Pada gambar 5 Grafik loss dan step digunakan untuk memonitor performa model selama proses pelatihan. Grafik ini menunjukkan bagaimana nilai loss menurun seiring dengan bertambahnya langkah pelatihan meskipun beberapa kali sempat naik pada step tertentu, yang menandakan bahwa model semakin baik dalam memprediksi gambar sesuai dengan deskripsi teks. Nilai loss yang rendah mengindikasikan peningkatan dalam kesesuaian antara output model dan target yang diinginkan (Doshi, 2024).



Gambar 5. Grafik Monitoring *Loss* dan *Step Training*

Beberapa contoh gambar yang dihasilkan oleh model yang telah *difine-tuning* menggunakan LoRA dapat dilihat pada gambar 6. Gambar-gambar ini menunjukkan peningkatan kualitas visual yang signifikan serta kesesuaian deskripsi teks yang lebih

baik, yang mencerminkan efektivitas teknik LoRA dalam meningkatkan hasil akhir model.



Gambar 6. Hasil Generate Model *Fine-tuning*

Pengujian Model

Pengujian dilakukan dengan menghasilkan gambar menggunakan teks dari dataset pengujian yang belum pernah dilihat model selama pelatihan. Setiap teks menghasilkan satu gambar yang kemudian dianalisis untuk menilai kesesuaian semantik dan kualitas visual gambar tersebut. Gambar 7 menunjukkan contoh gambar yang dihasilkan oleh model setelah pengujian menggunakan *prompt* yang sama dengan *dataset* pengujian .



Gambar 7. Hasil Generate Model *Fine-tuning*

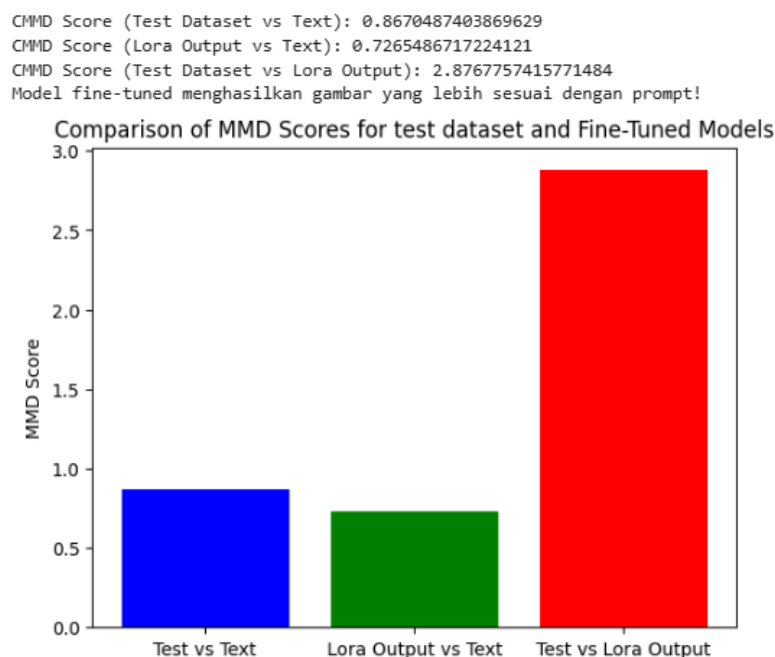
Gambar-gambar ini dianalisis menggunakan metrik kuantitatif berikut :

a. Evaluasi CLIP-MMD

CLIP-MMD digunakan untuk mengevaluasi kesesuaian semantik antara teks input dan gambar yang dihasilkan. Nilai CLIP-MMD yang diperoleh dari pengujian menunjukkan bahwa model menghasilkan gambar yang relevan secara semantik dengan teks input. Nilai CLIP-MMD yang lebih rendah menunjukkan kesesuaian semantik yang lebih baik antara gambar dan teks (Gao et al., 2021).

Gambar 8 menunjukkan perbandingan nilai CLIP-MMD antara data pengujian dan model *fine-tuning* menggunakan LoRA, Hasil analisis CMMD

Scores menunjukkan bahwa gambar dari dataset pengujian yang telah diberi caption memiliki nilai CMMD Score sebesar 0.867 ketika dibandingkan dengan embedding teks, yang mengindikasikan bahwa gambar tersebut sudah cukup sesuai dengan teks yang diberikan. Setelah proses *fine-tuning* menggunakan LoRA, terdapat peningkatan kesesuaian gambar dengan teks, yang tercermin dalam penurunan nilai CMMD Score menjadi 0.726, menandakan bahwa model lebih baik dalam mencocokkan gambar dengan teks. Namun, perbedaan yang cukup besar terjadi antara gambar dari dataset asli dan gambar yang dihasilkan pasca *fine-tuning*, sebagaimana terlihat dari nilai CMMD Score sebesar 2.877, yang menunjukkan adanya perubahan signifikan.



Gambar 8. Skor metrik CLIP-MMD

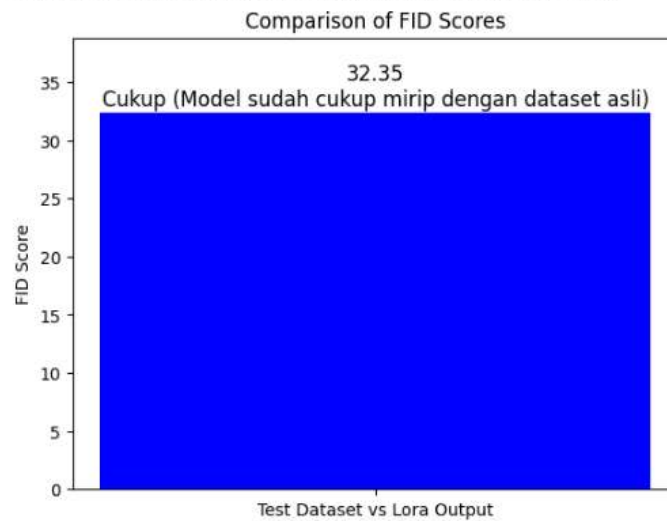
b. Evaluasi FID

FID digunakan untuk mengukur kualitas visual gambar yang dihasilkan oleh model. Nilai FID yang lebih rendah menunjukkan kualitas gambar yang lebih baik. Hasil pengujian menunjukkan bahwa model *fine-tuned* menggunakan LoRA menghasilkan gambar dengan nilai FID lebih rendah dibandingkan model dasar, menandakan kualitas visual yang lebih tinggi.

Gambar 9 menyajikan nilai FID untuk model dasar dan model *fine-tuned*, hasil interpretasi menunjukkan bahwa nilai FID yang lebih rendah mencerminkan kualitas *fine-tuning* yang lebih baik, karena distribusi fitur gambar yang dihasilkan semakin mendekati dataset asli. Meskipun nilai FID sebesar 32,35 masih terbilang cukup tinggi, hal ini menunjukkan bahwa *fine-tuning* dengan LoRA telah berhasil menghasilkan gambar yang cukup mirip dengan dataset awal, meskipun terdapat beberapa perbedaan yang menunjukkan kreativitas dan keberagaman. Faktor-faktor yang berkontribusi terhadap nilai FID ini termasuk kemungkinan bahwa *fine-tuning* dapat lebih ditingkatkan dengan data tambahan untuk kesesuaian yang lebih optimal, namun secara keseluruhan, model LoRA mampu menghasilkan gambar yang sangat

mendekati karakteristik dataset, dengan variasi yang memberikan kekayaan pada hasil akhirnya.

FID Score (Test Dataset vs Lora Output): 32.35
Interpretasi: Cukup (Model sudah cukup mirip dengan dataset asli)



Gambar 9. Skor Metrik FID

Berdasarkan hasil evaluasi CLIP-MMD dan FID, dapat disimpulkan bahwa model *Stable Diffusion* 2.1 yang di-*fine-tune* menggunakan teknik LoRA berhasil meningkatkan kualitas gambar secara visual dan semantik. Perbandingan hasil pengujian menunjukkan bahwa *fine-tuning* model dengan LoRA menghasilkan gambar yang lebih sesuai dengan teks deskripsi serta memiliki kualitas visual yang lebih baik, seperti yang terlihat pada Gambar 7, dibandingkan dengan penelitian lain yang menggunakan model *Stable Diffusion* tanpa *fine-tuning*, penelitian ini menunjukkan kelebihan dalam hal efisiensi komputasi dan peningkatan kualitas gambar dengan penerapan LoRA, yang lebih ringan dibandingkan dengan *fine-tuning* penuh. Hal ini membuat penelitian ini lebih efisien dalam hal penggunaan memori dan waktu pelatihan. Keunikan dari penelitian ini terletak pada penerapan teknik *Low-Rank Adaptation* (LoRA) untuk *fine-tuning* model *Stable Diffusion*, yang memungkinkan adaptasi model yang lebih efisien tanpa mengorbankan kualitas hasil. Penggunaan CLIP-MMD untuk menilai kesesuaian semantik dan FID untuk kualitas visual juga memberikan pendekatan yang lebih komprehensif dalam mengevaluasi model. Penelitian ini memberikan kontribusi dalam memperkenalkan metode *fine-tuning* yang lebih efisien dalam menghasilkan gambar berkualitas tinggi dari teks, khususnya dalam konteks lanskap fantasi, yang memiliki potensi besar untuk aplikasi dalam seni digital, desain grafis, dan pembuatan konten kreatif, di mana penciptaan visual yang imajinatif dan detail sangat dihargai.

KESIMPULAN

Bahwa penerapan teknik *Low-Rank Adaptation* (LoRA) untuk *fine-tuning* model *Stable Diffusion* 2.1 berhasil meningkatkan kualitas gambar, baik dari segi kesesuaian semantik dengan teks maupun kualitas visual. Proses *preprocessing* yang mencakup pengumpulan data, *captioning*, serta pembagian data untuk pelatihan dan pengujian telah dilakukan dengan cermat untuk mempersiapkan dataset, guna mendukung akurasi model. Teknik *fine-tuning* dengan LoRA menunjukkan hasil yang signifikan, tercermin dari penurunan nilai CLIP-MMD, yang menunjukkan kesesuaian semantik yang lebih baik antara gambar dan teks. Selain itu, penurunan nilai FID juga

menandakan adanya peningkatan dalam kualitas visual gambar. Meskipun nilai FID masih relatif tinggi, gambar yang dihasilkan memiliki kemiripan yang cukup dekat dengan dataset asli, dengan variasi yang memberikan dimensi tambahan pada hasil akhir. Keunikan dari penelitian ini terletak pada efisiensi penggunaan LoRA dalam fine-tuning, yang memungkinkan peningkatan kualitas gambar tanpa membebani sumber daya komputasi secara berlebihan. Selain itu, penerapan metode ini dalam konteks pembuatan gambar lanskap fantasi memberikan kontribusi signifikan terhadap bidang seni digital, desain grafis, dan pembuatan konten kreatif.

DAFTAR PUSTAKA

- Afshar, S. V., Eshaghi, S., & Kim, I. (2022). Pattern Analysis of Virtual Landscape within Educational Games. *Journal of Digital Landscape Architecture*, 2022(7), 435–442. <https://doi.org/10.14627/537724042>
- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 05, 97–100.
- Awards, T. G. (2024). *Game of the Year*. <https://Thegameawards.com/>.
- Bahia, A. B. (2023). Game Design At the Art Museum: the Nubla Case of Educathysen. *Heranca - History, Heritage and Culture Journal*, 6(1), 53–68. <https://doi.org/10.29073/heranca.v6i1.681>
- Blattmann, A., Dockhorn, T., Kulal, S., Mendelevitch, D., Kilian, M., Lorenz, D., Levi, Y., English, Z., Voleti, V., Letts, A., Jampani, V., & Rombach, R. (2023). *Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets*.
- Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, I. S. K. (2024). A Survey of Text-to-Image Diffusion Models in Generative AI. *Proceedings of the 14th International Conference on Cloud Computing, Data Science and Engineering, Confluence 2024*, 14(8), 73–78. <https://doi.org/10.1109/Confluence60223.2024.10463372>
- Doshi, A. (2024). The Loss in AI Translation. *Computational Linguistics, August*, 0–26.
- Gao, R., Liu, F., Zhang, J., Han, B., Liu, T., Niu, G., & Sugiyama, M. (2021). Maximum Mean Discrepancy Test is Aware of Adversarial Attacks. *Proceedings of Machine Learning Research*, 139(Mmd), 3564–3575.
- Howarth, J. (2024). *How Many Gamers Are There? (New 2024 Statistics)*. EXploding Topics.
- 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>
- Jin, Z., & Song, Z. (2023). *Generating coherent comic with rich story using ChatGPT and Stable Diffusion*.
- Kwon, Y., Wu, E., Wu, K., & Zou, J. (2024). DataInfl: Efficiently Estimating Data Influence in Lora-Tuned LLMs and Diffusion Models. *12th International Conference on Learning Representations, ICLR 2024*.
- Luo, Z., Xu, X., Liu, F., Koh, Y. S., Wang, D., & Zhang, J. (2024). *Privacy-Preserving Low-Rank Adaptation for Latent Diffusion Models*. 1–19.
- Mehrafrooz, B. (2024). *10 Most Common Challenges of Designing Great Game Environments*. Pixune.Com.
- Peng, Y. (2024). A Comparative Analysis Between GAN and Diffusion Models in Image Generation. *Transactions on Computer Science and Intelligent Systems Research*, 5(D), 189–195. <https://doi.org/10.62051/0f1va465>

- 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>
- Purwanto, H., Fandi, F., & Imaniawan, D. (2024). *Klasifikasi Tingkat Retensi Pemain Video Game Online*. 4(2), 75–82.
- Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, R., & Efrizoni, L. (2024). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 273–281. <https://doi.org/10.57152/malcom.v4i1.1085>
- Valvano, G., Agostino, A., De Magistris, G., Graziano, A., & Veneri, G. (2024). Controllable Image Synthesis of Industrial Data using Stable Diffusion. *Proceedings - 2024 IEEE Winter Conference on Applications of Computer Vision, WACV 2024*, 5342–5351. <https://doi.org/10.1109/WACV57701.2024.00527>
- Wallis, P. (n.d.). *LORA: Low-Rank Adaptation Of Large Language Models*. 1–26.
- Wang, W., Sun, Y., Yang, Z., Hu, Z., Tan, Z., & Yang, Y. (2024). *Replication in Visual Diffusion Models: A Survey and Outlook*. 1–20.
- Yu, Y., Zhang, W., & Deng, Y. (2021). Frechet inception distance (fid) for evaluating gans. *Researchgate.Net, September*, 1–7.
- Zhou, Y., Zhang, R., Chen, C., Li, C., Tensmeyer, C., Yu, T., Gu, J., & Xu, J. (n.d.). *Towards Language-Free Training for Text-to-Image Generation*. 17907–17917.