Jurnal Rekayasa Sistem Informasi dan Teknologi
Volume 2, No 3 – Februari 2025
e-ISSN : 3025-888X

GENERATOR GAMBAR KARAKTER JEPANG BERGAYA RETRO MENGGUNAKAN DIFFUSION MODELS DAN DREAMBOOTH

Muhammad Syihab Habibi*[1], Sri Mulyono[2]
Universitas Islam Sultan Agung[1,2]
syihab.demak@gmail.com

Received: 23-01-2025	Revised: 27-01-2025	Approved: 15-02-2025

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan generator gambar karakter Jepang bergaya retro menggunakan diffusion models dengan pendekatan fine-tuning DreamBooth dan regularisasi L2. Model yang digunakan dalam penelitian ini adalah Stable Diffusion v2.1, yang dioptimalkan melalui proses fine-tuning untuk menghasilkan gambar karakter yang sesuai dengan deskripsi teks serta estetika gaya retro. Metode penelitian yang diterapkan mencakup pengumpulan data, pemrosesan dataset, penggunaan alat dan perangkat lunak, pelatihan model generatif, serta evaluasi hasil. Dataset yang digunakan terdiri dari gambar karakter Jepang dengan elemen visual khas seperti ekspresi dinamis, warna pastel, dan pola unik. Dataset dibagi menjadi data pelatihan dan data uji guna mendukung validasi model. Proses pelatihan dilakukan dengan teknik fine-tuning menggunakan DreamBooth serta penerapan regularisasi L2 untuk mengurangi risiko overfitting. Evaluasi model dilakukan dengan menggunakan CLIP Score untuk mengukur kesesuaian gambar yang dihasilkan dengan deskripsi input, serta inspeksi visual untuk memastikan konsistensi atribut visual seperti ekspresi wajah, gaya pakaian, dan elemen dekoratif. Hasil penelitian menunjukkan bahwa fine-tuning dengan DreamBooth dapat meningkatkan kualitas dan akurasi gambar yang dihasilkan. Penggunaan regularisasi L2 membantu mempertahankan variasi visual tanpa mengurangi kesesuaian dengan deskripsi input. Selain itu, eksperimen dengan berbagai parameter seperti inference steps, guidance scale, dan jumlah dataset menunjukkan bahwa peningkatan jumlah dataset serta pengaturan parameter yang tepat dapat menghasilkan gambar dengan kualitas yang lebih baik, lebih stabil, dan lebih sesuai dengan deskripsi yang diberikan.

Kata Kunci: *Diffusion Models, Stable Diffusion, Dreambooth, Fine-Tuning, Regularisasi L2, Generator Gambar, Gaya Retro, CLIP Score*

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (AI) terus mendorong inovasi di berbagai bidang, termasuk seni dan desain digital. Salah satu teknologi terkemuka adalah diffusion models, yang bekerja melalui proses noise dan denoising untuk menghasilkan gambar berkualitas tinggi dari noise acak (Siringo-ringo 2023). Penelitian berjudul "LaDiffGAN: Training GANs with Diffusion Supervision in Latent Spaces" mengusulkan penggabungan kekuatan diffusion models dan GANs untuk tugas penerjemahan gambar yang tidak terawasi, meskipun tantangan dalam kompleksitas pelatihannya masih menjadi hambatan (Liu et al. 2024). Selain itu, penelitian lain berjudul "AnimeDiffusion: Anime Face Line Drawing Colorization via Diffusion Models" menunjukkan potensi diffusion models dalam pewarnaan gambar anime, meskipun fleksibilitasnya masih terbatas pada gaya tertentu (Cao et al. 2023). Teknologi ini menawarkan fleksibilitas tinggi untuk berbagai aplikasi, seperti animasi, desain karakter, dan game. Dalam konteks seni Jepang, gaya visual retro yang terinspirasi oleh estetika tahun 1980-an dan 1990-an menjadi pilihan populer karena daya tarik nostalgia yang dikombinasikan dengan elemen kontemporer (Hakim, Baihaki, and Mohammad Nasihin 2020). Desain ini sering mencakup warna pastel, garis tegas, dan motif khas yang mencerminkan budaya Jepang pada era tersebut. Penelitian berjudul "Personalizing Text-to-Image

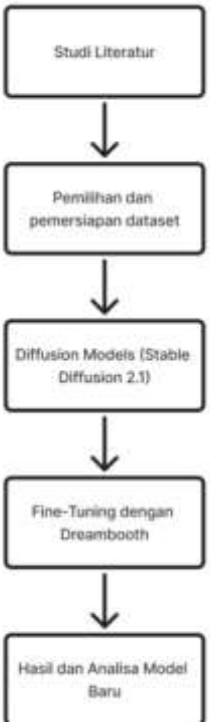
Diffusion Models by Fine-tuning Classification for AI Applications" menunjukkan bahwa fine-tuning menggunakan DreamBooth dapat meningkatkan adaptabilitas diffusion models untuk menghasilkan gambar sesuai kebutuhan tanpa memerlukan pelatihan ulang ekstensif.

Meskipun diffusion models memiliki kemampuan menghasilkan detail visual yang kompleks, tantangan seperti inkonsistensi output dan risiko overfitting selama pelatihan tetap menjadi kendala signifikan (Hendrawati, Wulandari, and Ginantra 2024). Model seperti Stable Diffusion telah banyak digunakan karena kemampuannya menghasilkan gambar berkualitas tinggi, namun keterbatasan seperti ketergantungan pada dataset besar dan proses sampling yang lambat masih menjadi tantangan utama. Oleh karena itu, pengembangan teknik fine-tuning seperti DreamBooth, yang memungkinkan personalisasi model melalui dataset kecil, menjadi solusi potensial. DreamBooth memanfaatkan embedding teks untuk mengintegrasikan atribut visual tertentu ke dalam gambar yang dihasilkan, seperti warna rambut, pakaian, atau elemen dekoratif (Dermawan and Herdianto 2024). Teknik regularisasi L2 membantu menjaga generalisasi model dengan menambahkan penalti terhadap bobot berlebih, sehingga mencegah overfitting dan menghasilkan gambar yang lebih konsisten (Iffah'da 2024)VV.

Secara teori, diffusion models bekerja dengan dua tahap utama: proses destruksi data (forward process) dan proses rekonstruksi (backward process) (Mulyatiningsih 2015) (Asiva Noor Rachmayani 2015). Pada tahap pertama, noise Gaussian ditambahkan ke gambar hingga menjadi tidak dapat dikenali. Selanjutnya, pada tahap kedua, model dilatih untuk secara bertahap menghilangkan noise, sehingga menghasilkan gambar yang realistis. Stable Diffusion, salah satu implementasi diffusion models, menggunakan arsitektur U-Net untuk menangkap fitur spasial dan menghasilkan gambar dengan detail visual tinggi (Aisyiah Rizqy Aulia 2024). U-Net dilengkapi dengan skip connections, yang memungkinkan informasi dari layer dalam untuk ditransfer ke layer dangkal, sehingga meningkatkan akurasi dan stabilitas model. Untuk efisiensi, penelitian berjudul "Parameter-Efficient Fine-tuning for Large Models: A Comprehensive Survey" menunjukkan bagaimana teknik Parameter Efficient Fine-tuning (PEFT) dapat mengurangi kebutuhan sumber daya komputasi. Selain itu, pendekatan seperti Distribution Matching Distillation (DMD) yang diperkenalkan menawarkan solusi fleksibel untuk mempercepat proses pelatihan dalam diffusion models. Penelitian-penelitian ini menunjukkan keberhasilan DreamBooth dalam meningkatkan fleksibilitas model, menggunakan fine-tuning untuk menghasilkan gambar mobil berbasis teks dengan kualitas tinggi meskipun dataset terbatas. Selain itu, penggunaan regularisasi telah terbukti mengurangi overfitting pada model generatif, konteks pelatihan model regresi. Meskipun demikian, studi tentang penggabungan teknik ini untuk menghasilkan gambar karakter bergaya retro masih jarang ditemukan, sehingga membuka peluang penelitian lebih lanjut.

METODE PENELITIAN

Penelitian ini bertujuan untuk mengembangkan generator gambar karakter Jepang bergaya retro menggunakan diffusion models dengan pendekatan fine-tuning DreamBooth dan regularisasi L2. Metode penelitian yang digunakan mencakup pengumpulan data, pemrosesan dataset, penggunaan alat dan perangkat lunak, hingga proses pelatihan model generatif serta evaluasi hasil.

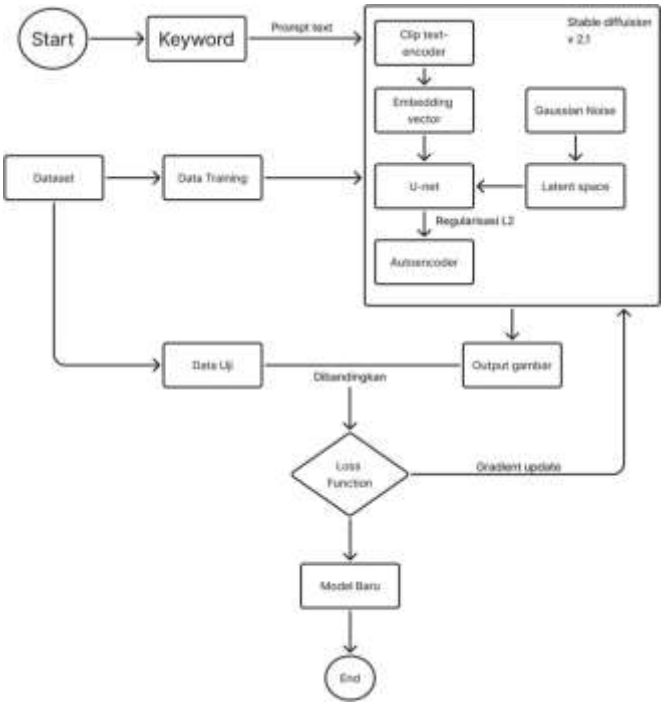


Gambar 1. Alur Metode Penelitian

Penelitian ini fokus pada pengembangan sistem generasi gambar berbasis teks dengan gaya retro. Karakter Jepang dipilih sebagai objek utama karena memiliki elemen visual yang khas, seperti ekspresi dinamis, warna pastel, dan pola unik. Model yang digunakan adalah Stable Diffusion v2.1, dengan proses fine- tuning untuk menghasilkan gambar karakter yang sesuai deskripsi teks dan sesuai dengan estetika gaya retro.

Rancangan Kegiatan

Rancangan kegiatan penelitian ini dimulai dengan studi literatur untuk menganalisis teknologi diffusion models, teknik fine-tuning, regularisasi, dan penerapan DreamBooth. Selanjutnya, kegiatan dilanjutkan dengan pengumpulan dan pengolahan dataset, di mana dataset yang berisi gambar karakter Jepang dengan gaya retro dipilih dari sumber terpercaya. Dataset dibagi menjadi dua kategori utama, yaitu data pelatihan dan data uji, guna mendukung proses pelatihan dan validasi model.



Gambar 2. Alur Traning Sistem

Pada tahap pelatihan model, Stable Diffusion v2.1 digunakan sebagai model utama dan dilakukan fine-tuning menggunakan metode DreamBooth untuk menyesuaikan model dengan dataset yang spesifik. Selain itu, regularisasi L2 diterapkan untuk mengurangi risiko overfitting, sehingga model dapat menghasilkan gambar yang konsisten dan berkualitas tinggi. Terakhir, dilakukan evaluasi dan validasi dengan mengukur kualitas gambar yang dihasilkan menggunakan metrik fidelity dan memeriksa kesesuaian visual antara gambar hasil generasi dengan deskripsi teks yang diberikan. Analisa Kebutuhan Sistem yaitu :

Perangkat keras:

- Komputer atau laptop dengan spesifikasi minimum RAM 8 GB, prosesor dual-core, dan koneksi internet yang stabil.
- Akses GPU/TPU (misalnya Tesla T4 atau K80) melalui platform cloud seperti Google Colab untuk mempercepat proses pelatihan model. Perangkat lunak:
- Google Colab: Platform untuk pelatihan model dengan akses ke GPU/TPU.
- Library Python:
  PyTorch untuk membangun dan melatih model diffusion. Transformers untuk encoding teks menggunakan CLIP Text Encoder. Diffusers untuk implementasi dan fine-tuning DreamBooth.
  Matplotlib dan NumPy untuk visualisasi dan analisis data. Pandas untuk pengorganisasian metadata dataset.

Dataset: Kumpulan gambar karakter Jepang yang dikumpulkan dari beberapa sumber dan diklasifikasikan berdasarkan atribut visual, seperti pose, gaya, dan elemen karakter.

Variabel Penelitian

- Kualitas gambar: Diukur berdasarkan tingkat detail visual, kesesuaian gaya retro, dan keselarasan dengan deskripsi teks.

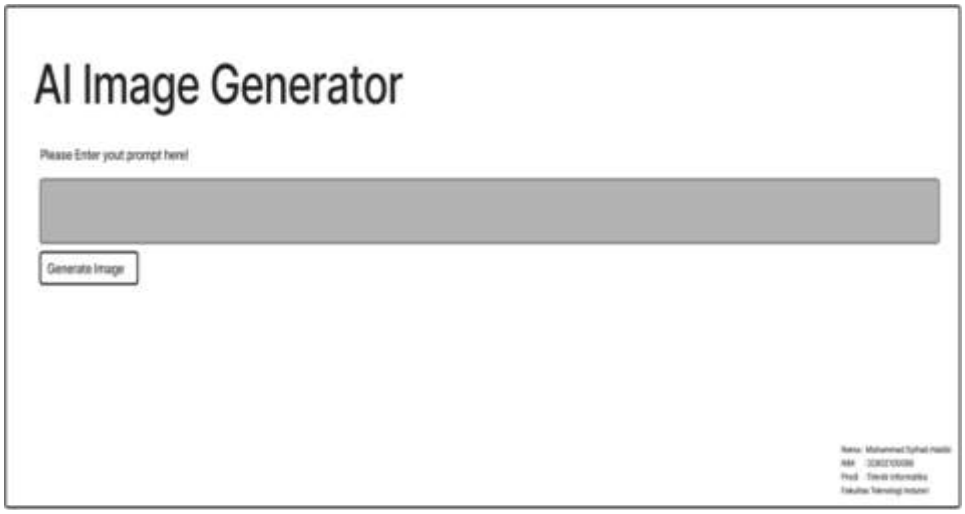
- Generalisasi model: Dihitung melalui evaluasi terhadap data uji untuk memastikan kemampuan model menghasilkan gambar yang konsisten dengan variasi yang memadai.
- Overfitting: Diidentifikasi melalui perbandingan antara performa model pada data pelatihan dan data uji.

Teknik Analisis Data

- Evaluasi fidelity gambar: Menggunakan CLIP score untuk mengukur kesesuaian antara gambar referensi dan gambar hasil generasi.
- Validasi visual: Melalui inspeksi manual terhadap atribut seperti ekspresi wajah, gaya pakaian, dan elemen dekoratif pada gambar.
- Analisis loss function: Menggunakan regularisasi L2 untuk memastikan model tetap stabil dan mengurangi risiko overfitting selama pelatihan.

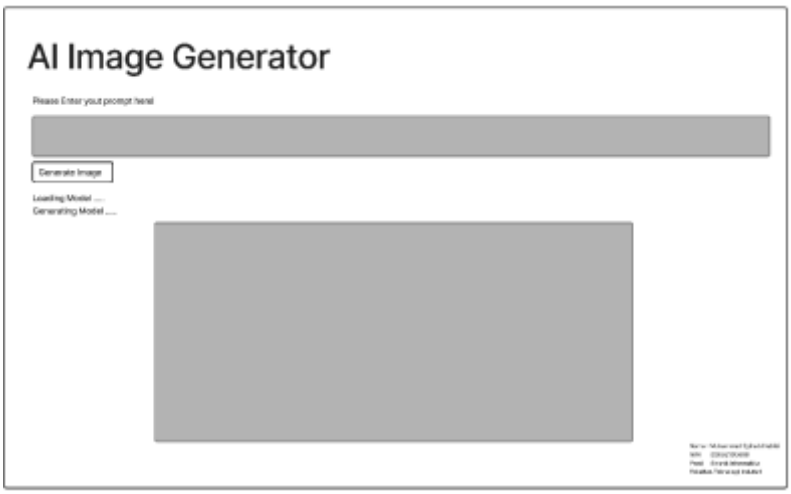
Perancangan Sitem Antar Muka

Antarmuka pengguna dirancang dengan pendekatan sederhana yang menekankan fungsi utama dan kemudahan navigasi. Halaman awal menampilkan kolom input persegi panjang berlatarkan abu-abu muda dengan placeholder bertuliskan "Please Enter your prompt here!" untuk memandu pengguna memasukkan deskripsi gambar.



Gambar 3. Rancangan Tampilan Awal

Di bawahnya, terdapat tombol "Generate Image" dengan desain bersih— berlatar putih dan teks hitam—yang mudah dikenali dan diakses. Setelah tombol diaktifkan, antarmuka memberikan umpan balik berupa status proses seperti "Loading Model..." dan "Generating Model..." yang muncul di bawah tombol. Pada bagian bawah halaman, area besar berbentuk persegi panjang dengan latar abu-abu disiapkan untuk menampilkan hasil gambar yang dihasilkan, memastikan pengguna dapat langsung memusatkan perhatian pada hasil akhir.



Gambar 4. Rancangan Tampilan Saat Gambar Berhasil di-generate

HASIL PENELITIAN DAN PEMBAHASAN

Dataset dikumpulkan dalam format gambar standar (.png, .jpg, .jpeg) dan diberi nama ulang untuk memastikan konsistensi. Model Stable Diffusion v2.1 digunakan sebagai model utama dalam generasi gambar, tetapi inkonsistensi output pada atribut visual spesifik diatasi melalui fine-tuning menggunakan metode DreamBooth. Evaluasi awal menggunakan CLIP Score dilakukan untuk menilai kesesuaian gambar yang dihasilkan dengan deskripsi input.

Tabel 1.
Penggunaan Prompt Terhadap Clip Score

Tahap	Prompt yang digunakan	Hasil Generasi	Rata-Rata CLIP Score
Generasi awal	*"Retro anime style, a man with glasses"*		0.32
Prompt Indonesia	"Gambar anime seorang laki-laki dengan gaya retro"		0.28
Negative Prompt	*"bad anatomy, ugly face, low quality, NSFW Content, Harsh lines, overly sharp edges, or digital precission, modern, futuristic"*		0.33

Spesifik Prompt	*"Retro style, anime, a man with glasses, the color are earthy and natural with soft gradients, giving a vintage film-like quality to the atmosphere"*		0.69

Model Stable Diffusion v2.1 digunakan sebagai model utama dalam generasi gambar. Seperti yang ditunjukkan pada table 1 Meskipun model ini sudah dilengkapi dengan kemampuan untuk menangani input berbasis teks, beberapa inkonsistensi output terlihat pada atribut visual tertentu, seperti warna dan elemen desain. Untuk mengatasi masalah ini, dilakukan fine-tuning menggunakan metode DreamBooth, yang memungkinkan penyesuaian lebih lanjut pada model agar dapat menghasilkan gambar dengan kualitas lebih tinggi dan kesesuaian yang lebih baik dengan deskripsi yang diberikan. Dilakukan pengaturan parameter seperti seed, inference steps, dan guidance scale juga dieksplorasi untuk melihat bagaimana parameter-parameter ini mempengaruhi kualitas gambar yang dihasilkan. Penggunaan seed yang konsisten memastikan bahwa gambar yang dihasilkan tetap serupa meskipun dilakukan pada waktu yang berbeda. Penyesuaian pada inference steps dan guidance scale menghasilkan gambar dengan detail yang lebih baik seperti yang ditunjukkan pada table 2, meskipun waktu pelatihan meningkat dengan parameter yang lebih tinggi.

Tabel 2.
Penyesuaian Parameter

Inference Step	
Guidance Scale (CFG)	

Nilai Inference Step yang lebih tinggi menghasilkan gambar dengan detail lebih baik, namun memerlukan waktu lebih lama. Sedangkan untuk Guidance Scale semakin kecil nilai CFG, gambar yang dihasilkan lebih bebas dan kreatif. Sebaliknya, nilai CFG yang lebih besar akan menghasilkan gambar yang lebih sesuai dengan deskripsi yang diberikan. Hasil evaluasi berdasarkan jumlah dataset yang digunakan menunjukkan bahwa jumlah dataset yang lebih besar meningkatkan kualitas output. Pengujian dilakukan dengan tiga jumlah dataset: 10, 20, dan 30 gambar. Dengan 10 gambar, hasilnya kurang konsisten, dan beberapa elemen latar belakang yang tidak relevan muncul. Namun, ketika jumlah dataset ditingkatkan menjadi 20 dan 30, gambar yang dihasilkan lebih stabil, dengan elemen-elemen visual yang lebih sesuai dan risiko overfitting yang lebih rendah.

Tabel 3.
Pengaruh Penggunaan Dataset

Dataset 10	
Dataset 30	
Dataset 30	
Dataset 50	

Seperti yang ditunjukkan pada table 3 ada eksperimen dengan dataset 10, overfitting terlihat dari munculnya elemen latar belakang yang tidak diinginkan karena model cenderung "menghapal" pola tertentu alih-alih memahami konteks deskripsi prompt secara menyeluruh. Dengan peningkatan jumlah dataset menjadi 20, 30 dan 50, model memiliki lebih banyak variasi data untuk dipelajari, yang membantu mengurangi pengaruh overfitting. Pengujian ini juga menunjukkan bahwa dengan 30 dataset, gambar yang dihasilkan memiliki kualitas yang lebih tinggi, lebih stabil, dan lebih sesuai dengan deskripsi. Jumlah dataset yang lebih banyak memungkinkan model untuk belajar variasi visual dengan lebih baik, mengurangi ketergantungan pada pola tertentu dari data pelatihan. Hal ini mengarah pada hasil yang lebih konsisten dan berkualitas tinggi, sehingga penggunaan dataset yang lebih banyak terbukti lebih efektif.

Tabel 4.
Variasi Max Train

Max Train	Ukuran Output	Dataset	Checkpoint
200	4 GB	15	-
2000	25 GB	15	500
1000	1.6 GB	10	-
1000	4.6 GB	30	-

Dengan 10 dataset, ukuran output sekitar 1.6 GB, sementara dengan 30 dataset, ukuran output meningkat menjadi 4.5 GB. Meskipun ukuran file lebih besar, hasil gambar yang dihasilkan lebih berkualitas dan lebih kompleks. Peningkatan jumlah dataset memperkaya variasi data yang dipelajari oleh model, meningkatkan kemampuan model dalam menghasilkan gambar yang lebih beragam dan realistis. Meskipun pelatihan dengan 2000 langkah dan output yang lebih besar menghasilkan hasil yang lebih kompleks, eksperimen dengan 1000 langkah dan ukuran output yang lebih besar serta dataset yang lebih banyak menghasilkan model yang stabil dan efisien, menjadikannya pilihan terbaik dalam eksperimen ini Penggunaan regularisasi L2 dalam eksperimen ini terbukti efektif dalam mempertahankan variasi visual tanpa mengurangi kesesuaian gambar dengan deskripsi yang diberikan. Seperti yang digunakan pada table 5, sebelum menggunakan regularisasi L2, gambar yang dihasilkan cenderung terfokus pada pola-pola dalam dataset, menghasilkan kesamaan yang tinggi dengan gambar referensi, tetapi dengan beberapa perbedaan visual yang jelas.

Tabel 5.
Penggunaan Regularisasi

Kondisi	Prompt yang digunakan	Hasil Generasi	Rata-Rata CLIP Score
Sebelum menggunakan regularisai	*"1980s-inspired anime style, a cat in market"*		0.73
Sesudah menggunakan regularisai	*"1980s-inspired anime style, a cat in market"*		0.75

Sebelum menggunakan regularisai	*"1980s-inspired anime style, a cat in market"*		0.68
Sesudah menggunakan regularisai	*"1980s-inspired anime style, a cat in market"*		0.70

Setelah penerapan L2, gambar yang dihasilkan menjadi lebih variatif dan tidak terpaku pada pola tertentu dalam dataset, dengan hasil yang lebih mendekati gambar referensi, baik dalam detail maupun keseluruhan visual, menunjukkan bahwa L2 berhasil meningkatkan fleksibilitas model dan kesesuaiannya dengan prompt. Sistem generasi gambar ini diimplementasikan dalam bentuk web menggunakan framework Streamlit untuk memudahkan pengguna menghasilkan gambar karakter Jepang bergaya retro berdasarkan deskripsi teks.

Gambar 5. Hasil Generalisasi Dengan Streamlit

Seperti yang ditunjukkan pada gambar 5 hasil tidak hanya mencakup karakter itu sendiri, tetapi juga dapat memuat aksi atau kegiatan yang sedang dilakukan oleh karakter, seperti naik sepeda atau melakukan kegiatan lain yang disesuaikan dengan deskripsi. Selain itu, sistem juga memungkinkan pengguna untuk menghasilkan latar belakang yang sesuai dengan cerita atau konteks, seperti pasar atau desa, memberikan gambaran yang lebih hidup dan kontekstual.

KESIMPULAN

Penelitian ini berhasil membangun generator gambar karakter Jepang bergaya retro menggunakan diffusion models dengan fine-tuning DreamBooth, menghasilkan gambar realistis yang sesuai dengan deskripsi pengguna. Pemrosesan data dan dataset yang memadai memainkan peran penting dalam menjaga konsistensi dan variasi gaya, sementara regularisasi L2 mencegah overfitting, memastikan kualitas visual yang stabil. Penelitian selanjutnya dapat mengeksplorasi gaya visual lain, memperluas dataset, dan mengintegrasikan fitur interaktif untuk meningkatkan fleksibilitas. Selain itu, optimasi

waktu inferensi melalui distilasi model atau arsitektur efisien dapat mempercepat generasi gambar tanpa mengurangi kualitas.

DAFTAR PUSTAKA

Aisyiah Rizqy Aulia, Mahameru Rosy Rochmatullah. 2024. "The Use of the Audit Tool and Linked Archive System (Atlas) By Public Accounting Firm (Paf) Auditors in Indonesia: An Extended Technology Acceptance Model (Tam) Analysis Penggunaan." *COSTING:Journal of Economic, Business and Accounting* 7(2015).

Asiva Noor Rachmayani. 2015. *Model Pembelajaran Addie Integrasi Pedati Di Smk Pgri Karisma Bangsa Sebagai Pengganti Praktek Kerja Lapangan Dimasa Pandemi Covid-19*.

Cao, Yu et al. 2023. "AnimeDiffusion: Anime Face Line Drawing Colorization via Diffusion Models." 14(8): 1–14. http://arxiv.org/abs/2303.11137.

Dermawan, Raja Diky, and Herdianto. 2024. "Meningkatkan Kinerja Output ChatGPT Melalui Teknik Prompt Engineering Yang Dapat Dikustomisasi." *Journal Of Social Science Research* 4(1): 10646–64.

Hakim, Nazar Yepta, Baihaki, and Mohammad Nasihin. 2020. "Konsep Dan Visual Artistik Dalam Pengembangan Gim."

Hendrawati, Theresia, Dewa Ayu Putri Wulandari, and Ni Luh Wiwik Sri Rahayu Ginantra. 2024. "Penerapan Metode Stable Diffusion Dengan Fine Tuning Untuk Pola Endek Bali." *Jurnal Teknologi Informasi Komunikasi (e-Journal)* 10: 8–10.

Iffah'da, Annisa Nurba. 2024. "Penerapan Multiplanar Reconstruction Pada Arsitektur U-Resnet Dan Mobilenet Dalam Proses Segmentasi Hati Citra Tiga Dimensi Hasil Ct Scan." *Αγαη* 15(1): 37–48.

Liu, Xuhui et al. 2024. "LaDiffGAN: Training GANs with Diffusion Supervision in Latent Spaces." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* 1(d): 1115–25.

Mulyatiningsih, Endang. 2015. "Pengembangan Model Pembelajaran." *Islamic Education Journal*: 35,110,114,120,121.

Siringo-ringo, M M. 2023. "Peran Sektor Teknologi Dalam Mendorong Inovasi Dan Pertumbuhan Ekonomi Di Tahun 2023." *Circle Archive* 1(2): 1–12. http://circle-archive.com/index.php/carc/article/view/44%0Ahttps://circle-archive.com/index.php/carc/article/download/44/44.