PENERAPAN *FINE TUNING DREAMBOOTH* DAN *STABLE DIFFUSION* UNTUK PEMBUATAN POSTER FILM ANIMASI

LAPORAN TUGAS AKHIR

Laporan ini diajukan guna memenuhi syarat memperoleh gelar Sarjana (S1) pada Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung.

DI SUSUN OLEH :

NAMA : TIFANA FARISA KARIMA

NIM : 32602100119

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

SEMARANG

2025

FINAL PROJECT

APPLICATION OF FINE TUNING DREAMBOOTH AND STABLE DIFFUSION FOR ANIMATED FILM POSTER CREATION

Submitted to fulfill the requirements to obtain a Bachelor's degree (S1) in the Informatics Engineering Department, Faculty of Industrial Technology, Sultan Agung Islamic University.

ARRANGED BY:

NAME : TIFANA FARISA KARIMA

NIM : 32602100119

MAJORING OF INFORMATICS ENGINEERING

INDUSTRIAL TECHNOLOGY FACULTY

SULTAN AGUNG ISLAMIC UNIVERSITY

SEMARANG

2025

LEMBAR PENGESAHAN

TUGAS AKHIR

PENERAPAN *FINE TUNING DREAMBOOTH* DAN *STABLE DIFFUSION* UNTUK PEMBUATAN POSTER FILM ANIMASI

TIFANA FARISA KARIMA

32602100119

Telah dipertahankan di depan tim penguji proposal tugas akhir

Program Studi Teknik Informatika

Universitas Islam Sultan Agung

Pada tanggal : 24 Februari 2025

TIM PENGUJI SIDANG TUGAS AKHIR:

Sam Farisa Chaerul Haviana, ST, M.Kom NIDN. 0628028602 (Penguji 1)	.....	06-03-2025 .....
Andi Riasyah, ST, M. Kom. NIDN. 0609108802 (Penguji 2)	.....	27-02-2025 .....
Ir. Sri Mulyono, M.Eng NIDN. 0626066601 (Pembimbing)	.....	06-03-2025 .....

Semarang, 24 Februari 2025

Mengetahui,

Kaprodi Teknik Informatika

Universitas Islam Sultan Agung

Moch Taufiq MIT

NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Tifana Farisa Karima

NIM : 32602100119

Judul Tugas Akhir : PENERAPAN *FINE TUNING DREAMBOOTH* DAN *STABLE DIFFUSION* UNTUK PEMBUATAN POSTER FILM ANIMASI

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 06 Maret 2025

Yang Menyatakan

Tifana Farisa Karima

KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah Ta'ala yang telah memberikan penulis rahmat dan karunianya yang luar biasa serta telah memberikan kekuatan serta memberikan memberikan kemudahan dalam menyelesaikan Tugas Akhir yang berjudul Penerapan *Fine Tuning Dreambooth* dan *Stable Diffusion* Untuk Pembuatan Poster Film Animasi. Penyusunan laporan Tugas Akhir ini merupakan salah satu kewajiban untuk memperoleh gelar Sarjana S1 pada Program Studi Teknik Informatika Universitas Islam Sultan Agung Semarang.

Penulis menyadari bahwa selesainya laporan ini tidak lepas dari bimbingan, bantuan, saran serta fasilitas yang diberikan berbagai pihak. Oleh karenanya, pada kesempatan ini dengan segenap rendah hati, tak lupa penulis sampaikan rasa hormat dan terimakasih yang mendalam kepada :

1. Rektor Universitas Islam Sultan Agung Semarang Prof. Dr. H. Gunarto, S.H., M.H.
2. Dekan Fakultas Teknologi Industri Universitas Islam Sultan Agung, Dr. Ir. Novi Marlyana, S.T., M.T., IPU., ASEAN.Eng.
3. Ketua Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung, Moch Taufik, ST, MIT.
4. Koordinator Tugas Akhir Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung, Badieah, ST., M.Kom.
5. Dosen Pembimbing Ir. Sri Mulyono, M.Eng atas waktu yang telah diluangkan dan bimbingan akademis yang telah diberikan hingga selesai nya Tugas Akhir ini.
6. Segenap Dosen Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung yang telah memberikan bimbingan, ilmu dan masukannya.
7. Orang Tua penulis yang terus mendoakan penulis hingga terselesainya laporan ini.

Penulis menyadari bahwa tiada sesuatu hal pun di dunia ini yang sempurna begitu pula dengan laporan Tugas Akhir ini. Oleh karena itu, kritik dan saran dari pembaca sangat penulis butuhkan. Penulis berharap semoga laporan ini bermanfaat bagi semua pihak yang berkepentingan. *Aamiin*

Semarang,

Tifana Farisa Karima

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Pembuatan poster melibatkan banyak hal dan langkah langkah yang membutuhkan pengalaman desain dan kreativitas. Dengan teknologi kecerdasan buatan seperti *Stable Diffusion* dapat digunakan untuk memenuhi kebutuhan otomasi tersebut. Namun, model dasar sering kali tidak spesifik dan tidak konsisten dalam menyertakan elemen tambahan seperti judul. Untuk mengatasi masalah ini *Dreambooth* dapat digunakan untuk meningkatkan kemampuan *Stable Diffusion* dalam memahami prompt yang diberikan. Hasil menunjukkan bahwa model tanpa *Fine Tuning* menghasilkan gambar yang kurang sesuai dengan deskripsi, sementara model yang dilatih dengan *Dreambooth* menghasilkan gambar yang lebih akurat dan detail, serta mampu menyertakan judul. Penggunaan *Negative prompt* juga meningkatkan akurasi dengan mengeliminasi elemen yang tidak diinginkan. Evaluasi menggunakan CLIP *Score* menunjukkan bahwa model yang dilatih dengan *Dreambooth* memiliki nilai sekitar 0,7. Menunjukkan bahwa kombinasi *Fine Tuning Dreambooth* dan *Negative prompt* dapat meningkatkan kualitas generasi poster film.

Kata Kunci: *Stable Diffusion, Dreambooth, Fine Tuning, Negative prompt,* Desain Poster Film

ABSTRACK

Making a poster involves many things and steps that require design experience and creativity. Artificial intelligence technology such as Stable Diffusion can be used to meet these automation needs. However, basic models are often unspecific and inconsistent in including additional elements such as titles. To overcome this problem, Dreambooth can be used to increase Stable Diffusion's ability to understand the prompts given. The results show that the model without Fine Tuning produces images that do not match the description, while the model trained with Dreambooth produces images that are more accurate and detailed, and are able to include titles. The use of Negative prompts also increases accuracy by eliminating unwanted elements. Evaluation using CLIP Score shows that the model trained with Dreambooth has a value of around 0.7. Shows that the combination of Fine Tuning Dreambooth and Negative prompt can improve the quality of movie poster generation.

Keywords: Stable Diffusion, Dreambooth, Fine Tuning, Negative prompt, Movie Poster Design

BAB I
PENDAHULUAN

1.1 Latar Belakang

Karena menggabungkan ekonomi, kreativitas, dan teknologi, industri kreatif saat ini menjadi salah satu sektor dengan pertumbuhan tercepat di dunia. Sektor ini telah berperan penting dalam menciptakan lapangan pekerjaan di banyak negara. Berbagai bidang seperti seni visual, desain, musik, periklanan, dan perfilman berkolaborasi untuk menghasilkan karya yang memiliki nilai estetika tinggi. Namun industri kreatif Indonesia tergolong tertinggal padahal industri kreatif ini memiliki potensi yang cukup besar untuk meningkatkan ekonomi daerah tapi kabar baiknya industri kreatif di Indonesia masih memiliki peluang besar untuk bersaing di pasar global berkat kemajuan teknologi yang ada saat ini (Badriyah dan Lukmandono, 2023).

Poster film adalah komponen penting dalam industri perfilman karena berfungsi sebagai alat promosi yang penting untuk menyampaikan pesan utama, identitas, dan suasana film secara visual. (Munawarah dan Tomi, 2023). Poster juga berfungsi untuk menarik perhatian penonton dan menyampaikan inti cerita dan nuansa film dalam satu perspektif. Poster film animasi merupakan media terdepan dalam menyampaikan informasi film Oleh karena itu, poster film menjadi komponen penting dari strategi pemasaran film.

Dalam konteks film animasi, poster memiliki peran yang lebih spesifik karena perlu menyampaikan nuansa fantasi, karakter yang penuh warna, dan elemen visual yang mencerminkan gaya unik animasi tersebut. Salah satu tantangan saat membuat poster animasi adalah memastikan bahwa karakter yang diilustrasikan sesuai dengan cerita dan estetika visual film animasi itu sendiri namun tetap menarik perhatian audiens secara luas (Effendi, 2023).

Dalam pembuatan poster film animasi cukup membutuhkan waktu yang lama dan proses yang sulit. Desainer harus memperhatikan banyak hal dan pengalaman kreativitas (Lin *dkk.*, 2023). Hal ini memerlukan kerja sama antara desainer, sutradara, produser, dan pihak lain yang terlibat dalam proyek film. Selain itu, desainer sering kali harus melakukan banyak perubahan untuk mencapai hasil yang diinginkan tim kreatif. Dimana bisa menjamin bahwa poster tersebut dapat menyampaikan pesan yang kuat dalam waktu singkat dan menarik perhatian di tengah persaingan pasar yang ketat merupakan tantangan tambahan.

Dalam hal ini, kecerdasan buatan sangat penting untuk mengatasi masalah ini, AI dapat mempercepat proses pembuatan poster film dengan mengotomatisasi tugas tugas yang lebih lama dilakukan sebelumnya, seperti memilih warna, mengatur komposisi, dan memilih elemen visual yang sesuai dengan cerita (Hanifa *dkk.*, 2023).

Diffusion Models telah berkembang menjadi teknologi canggih yang memungkinkan pembuatan poster yang lebih cepat dan efektif berdasarkan deskripsi teks. *Stable Diffusion* adalah salah satu teknologi terdepan di bidang ini, yang menunjukkan kemampuan untuk menyesuaikan gaya visual dengan metode *Fine Tuning*. Dalam pembuatan poster animasi, *Stable Diffusion* memungkinkan penciptaan ilustrasi karakter dan elemen visual yang relevan dengan gaya artistik animasi. Dengan menggunakan *Fine Tuning*, model ini dapat menghasilkan gambar dengan detail dan akurasi yang sesuai dengan tema dan elemen visual khas film animasi, seperti warna yang hidup dan desain karakter yang ekspresif.

Stable Diffusion 2.1 dapat dioptimalkan untuk menghasilkan gambar berkualitas tinggi yang sesuai dengan gaya tertentu dengan menggunakan metode *Fine Tuning*. Namun untuk mencapai hasil yang berkualitas tersebut, penyesuaian keterangan gambar dan prosedur inferensi harus sesuai dengan set pelatihan agar hasil yang dikeluarkan relevan dan akurat. Teknik *Fine Tuning Dreambooth* akan digunakan untuk menyesuaikan model dengan *Dataset* tertentu, dengan fokus pada akurasi menampilkan elemen visual yang

diinginkan. *Stable Diffusion* 2.1 menunjukkan hasil yang lebih baik dalam tugas komposisional sambil mempertahankan kemampuan generatifnya. Dibandingkan versi 1.5 yang membuat versi 2.1 lebih relevan untuk berbagai situasi aplikasi (Krojer *dkk.*, 2023).

1.2 Perumusan Masalah

Rumusan masalah dari penelitian ini dapat dirumuskan dalam satu pertanyaan utama yaitu Bagaimana Menerapkan *Fine Tuning Dreambooth* dan *Stable Diffusion* untuk proses pembuatan poster film animasi berbasis deskripsi teks?

1.3 Pembatasan Masalah

Tujuan pembatasan masalah dibawah ini adalah untuk menghindari kegiatan diluar sasaran, oleh karena itu dalam pembuatan laporan ini perlu ditetapkan batasan masalah tersebut sebagai berikut:

1. Penelitian ini hanya menggunakan *Stable Diffusion* 2.1 yang telah di *Fine Tuning* dengan *Dreambooth* untuk menghasilkan poster film animasi.
2. Sistem dirancang menggunakan *Streamlit* sebagai aplikasi berbasis web, memungkinkan pengguna untuk memasukkan deskripsi teks (*prompt*) melalui antarmuka interaktif yang telah disediakan.
3. Penelitian ini dibatasi pada pembuatan poster film yang hanya menampilkan judul dan object utama.
4. Sistem tidak menyediakan fitur editing gambar secara langsung, pengguna hanya menerima hasil akhir sesuai dengan prompt yang dimasukkan.
5. Sistem hanya mendukung generasi satu poster, tanpa kemampuan untuk menghasilkan beberapa poster dalam satu proses.

1.4 Tujuan

Tujuan dari penelitian ini adalah membangun sistem yang mampu menghasilkan poster film animasi secara otomatis berdasarkan deskripsi teks. Sistem ini dirancang menggunakan metode *Fine Tuning Dreambooth* dan *Stable Diffusion* 2.1 untuk meningkatkan kemampuan model dalam memahami dan menerjemahkan deskripsi teks menjadi visual yang sesuai.

1.5 Manfaat

Penelitian ini bermanfaat untuk industri kreatif karena sistem yang dikembangkan dapat mempercepat proses desain poster film animasi, mengurangi biaya produksi, dan membantu desainer membuat lebih banyak pilihan desain. Studi ini menunjukkan bahwa kecerdasan buatan dapat membantu proses kreatif dengan meningkatkan akurasi generasi gambar berbasis teks dan menjadi dasar untuk pengembangan sistem generatif lainnya di bidang desain dan multimedia.

1.6 Sistematika Penulisan

Untuk mempermudah penulisan tugas akhir ini, penulis membuat suatu sistematika yang terdiri dari:

BAB 1 : PENDAHULUAN

Bagian ini menjelaskan dasar penelitian tentang kemajuan industri kreatif, khususnya pembuatan poster film, dan bagaimana teknologi kecerdasan buatan (AI) memainkan peran penting dalam mempercepat proses kreatif. Latar belakang menjelaskan bagaimana *Diffusion Model*, khususnya *Stable Diffusion*, dapat digunakan untuk membuat poster film berdasarkan deskripsi teks. Selain itu juga membahas metode pengoptimalan *Dreambooth*, yang memungkinkan untuk menyesuaikan model agar dapat menampilkan elemen visual yang diinginkan dengan lebih akurat. Rumusan masalahnya Bagaimana penerapan *Fine Tuning Dreambooth* pada *Stable Diffusion* 2.1 untuk proses pembuatan poster film animasi berbasis deskripsi teks dan juga sejauh mana *Stable Diffusion* 2.1 mendekati keinginan user dalam pembuatan poster film animasi. Bagian ini juga mencakup batasan penelitian untuk memperjelas ruang lingkup, seperti penggunaan model *Stable Diffusion* 2.1 dan keterbatasan input hanya pada deskripsi teks. Tujuan penelitian adalah membangun sistem yang dapat menghasilkan poster film secara otomatis namun juga relevan dengan tema cerita berdasarkan deskripsi teks menggunakan *Fine Tuning Dreambooth* pada *Stable Diffusion* 2.1.

BAB 2 : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini menjelaskan kemajuan dalam *Diffusion Model* serta bagaimana AI digunakan dalam industri kreatif. Penelitian sebelumnya membahas peran *Stable Diffusion* dalam meningkatkan akurasi model. Untuk memberikan gambaran teknis mengenai proses penyesuaian model yang mendukung penelitian ini, teori tambahan tentang *Fine Tuning* dan metode khusus *Dreambooth* juga dijelaskan.

BAB 3 : METODE PENELITIAN

Bab ini menjelaskan tahapan-tahapan yang dilakukan dalam penelitian, mulai dari Studi Literatur, Pengumpulan dan Pengolahan *Dataset*, Penggunaan *Stable Diffusion* 2.1, *Fine Tuning* dengan *Dreambooth* dan Pengujian Model. Proses penelitian ini diuraikan dengan detail, termasuk teknik *Fine Tuning* yang diterapkan pada *Diffusion Models* untuk mendapatkan hasil poster yang sesuai dengan deskripsi.

BAB 4: HASIL PENELITIAN DAN IMPLEMENTASI SISTEM

Bab ini memaparkan hasil implementasi sistem pembuatan poster film berdasarkan deskripsi teks. Hasil dari eksperimen menggunakan *Diffusion Models* ditampilkan dan dianalisis, termasuk kualitas visual poster yang dihasilkan, relevansi dengan deskripsi teks. Selain itu, bab ini juga mencakup pengujian terhadap berbagai deskripsi teks untuk melihat fleksibilitas sistem dalam menghasilkan gaya visual yang beragam.

BAB 5: KESIMPULAN DAN SARAN

Bab terakhir berisi kesimpulan dari penelitian ini, termasuk capaian dalam penggunaan *Stable Diffusion* dan *Dreambooth* untuk menghasilkan poster film berdasarkan deskripsi teks. Selain itu, disampaikan saran-saran untuk pengembangan lebih lanjut, eksplorasi gaya visual yang lebih kompleks.

BAB II
TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Generative Artificial Intelligence (Gen AI) telah mengubah industi kreatif dengan kemampuannya yang menciptakan konten contohnya membuat poster secara otomatis. Dalam penelitian yang berjudul *"A Survey on Generatuve Diffusion Models"* menjelaskan bahwasannya *Diffusion Model* telah berkembang menjadi terobosan besar dalam teknologi *generative,* karena kemampuannya untuk menghasilkan gambar yang realistis dan berkualitas tinggi. Model ini mengatasi beberapa tantangan yang dihadapi oleh model generatif sebelumnya, seperti *Variational AutoEncoders* dan *Generative Adversarial Networks*, dengan menawarkan stabilitas dalam pelatihan dan kualitas generasi yang lebih baik. (Cao *dkk.*, 2024).

Dijelaskan juga dalam penelitian yang berjudul *"Denoising Diffusion Probabilistic Models"* bahwasannya *Diffusion Model* mampu menghasilkan gambar yang konsisten dan sesuai dengan deskripsi Metode ini mengandalkan proses difusi yang stabil, memungkinkan model untuk belajar dari noise secara efisien. Keunggulan utama *Diffusion Model* terletak pada stabilitas dan prediktabilitasnya yang lebih tinggi dibandingkan metode generatif lainnya, seperti GAN. Dengan pendekatan ini, model dapat menghasilkan gambar berkualitas tinggi yang mencerminkan detail yang halus dan kompleksitas. Penelitian ini menyoroti potensi besar *Diffusion Model* dalam aplikasi generatif modern (Ho *dkk.*, 2020).

Dalam penelitian berjudul *"Realistic Noise Synthesis with Diffusion Models"* Model ini bekerja dengan cara mengubah *Noise* acak menjadi gambar yang jelas dan berkualitas melalui proses difusi. Dengan memanfaatkan informasi pengaturan kamera dan konten multiskala, model ini dapat mensintesis *noise* yang selanjutnya digunakan untuk pelatihan model. Proses ini memungkinkan model untuk belajar dari distribusi *noise* yang kompleks dan menghasilkan gambar yang lebih baik, sehingga meningkatkan kualitas hasil denoising secara signifikan. (Wu *dkk.*, 2023).

Salah satu penggunaan menarik dari teknologi ini adalah untuk membuat poster film otomatis, di mana deskripsi teks yang mencakup elemen penting film, seperti karakter, suasana, dan tema, dimasukkan untuk membuat representasi visual. Hal ini mengurangi jumlah waktu yang diperlukan untuk membuat desain awal dan memberi desainer lebih banyak waktu untuk menyempurnakan visual.

Dalam penelitian berjudul *"Usability Analysis Of Stable Diffusion-Based Generative Model For Enriching* Batik Bakaran *Pattern Synthesis" Stable Diffusion* 2.1 mampu menghasilkan gambar yang lebih realistis dibandingkan versi 1.4 dan telah dibuktikan oleh skor Inception *Score* yang lebih tinggi (Septemedi dan Santosa, 2024).

Dalam Penelitian berjudul "A Survey of Diffusion Based Image Generation Models: Issues and Their Solutions", model generatif berbasis difusi yang paling efektif untuk menghasilkan gambar dari deskripsi teks disebut *Stable Diffusion*. Model ini mengubah noise menjadi gambar yang terstruktur secara bertahap melalui dua tahap: maju (pengacakan) dan kembali (pembangunan kembali). Ini memungkinkan model untuk memahami dan meniru detail gambar. Teknik *Fine Tuning* digunakan untuk meningkatkan kinerja model dalam bidang tertentu. Teknik ini memungkinkan model yang telah dilatih sebelumnya disesuaikan dengan *Dataset* tertentu, yang meningkatkan akurasi dan relevansi gambar yang dihasilkan. Pendekatan yang kuat untuk membuat gambar berkualitas tinggi yang sesuai dengan persyaratan khusus, seperti gaya seni atau desain tekstil, dihasilkan melalui kombinasi *Stable Diffusion* dan *Fine Tuning* (Zhang *dkk.*, 2023).

Dalam penelitian berjudul "*Dreambooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation*" menjelaskan bahwa *Fine Tuning Dreambooth* memungkinkan pengguna untuk menghasilkan gambar realistis dari subjek tertentu dengan hanya beberapa gambar referensi (3-5) penyesuaian ini lebih khusus pada elemen gambar tertentu, seperti gaya atau tema visual, dengan hasil yang lebih baik dan akurat (Ruiz dkk., 2023).

Teknik ini menyesuaikan model tanpa menambah latensi inferensi, yang memungkinkan pelatihan yang berpusat pada aspek khusus dari sebuah subjek. Dengan menggunakan *Dreambooth* saat mendesain poster film, model dapat menampilkan fitur khusus film dengan tepat pada visual akhir, yang menghasilkan poster yang lebih realistis dengan biaya yang lebih rendah dan waktu komputasi yang lebih sedikit.

Memodifikasi model difusi dengan *Dreambooth* memiliki potensi besar untuk pembuatan poster otomatis. Dengan menggunakan teknologi ini, pembuat film dan desainer lokal dapat membuat materi promosi yang lebih inovatif dan menarik dengan lebih mudah, *Dreambooth* merupakan fitur yang sangat penting bagi industri kreatif Dimana memiliki kemampuan untuk menyesuaikan fitur khusus seperti gaya visual dan tema tertentu, yang mempercepat proses desain sambil mengurangi waktu dan biaya. Selain itu, teknologi ini meningkatkan variasi gaya dan visual, membuka peluang baru untuk inovasi di pasar lokal, dan membuat industri kreatif Indonesia lebih disukai oleh audiens di seluruh dunia.

2.2 Dasar Teori

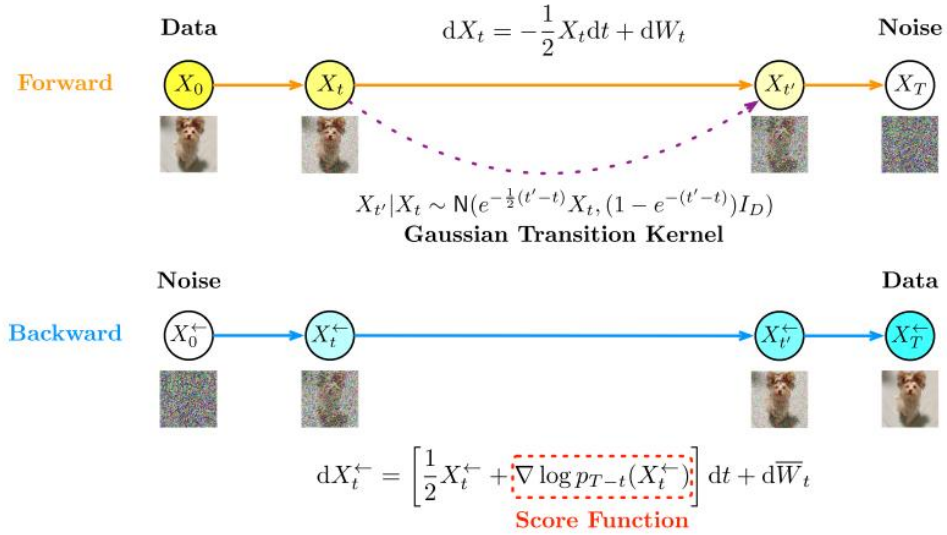
2.2.1 Implementasi AI dalam Industri Kreatif

Sektor kreatif Indonesia telah meningkat secara signifikan berkat pengaruh AI. Produksi konten kreatif seperti seni digital, musik, dan desain grafis dipercepat dengan teknologi AI. Selain itu, AI sangat berguna dalam analisis pasar, di mana ia mengumpulkan dan menganalisis data konsumen untuk menemukan tren dan preferensi yang relevan. Dalam konteks personalisasi layanan, AI mempelajari pola perilaku konsumen untuk memungkinkan pengalaman yang disesuaikan dengan kebutuhan individu. Melalui peningkatan proses desain dan penelitian dan pengembangan, AI juga membantu mengembangkan produk inovatif. Menurut penelitian (Hanifa *dkk.*, 2023) AI dapat meningkatkan kualitas, efisiensi, dan inovasi dalam industri kreatif jika diterapkan dengan benar. Oleh karena itu, jika perusahaan dan pelaku industri kreatif di Indonesia ingin memperoleh keunggulan

kompetitif dan mendorong pertumbuhan industri, mereka harus mempertimbangkan penggunaan AI sebagai strategi.

2.2.2 Diffusion Models

Model generatif *diffusion* telah muncul sebagai salah satu metode terbaru dalam pembelajaran mesin, terutama dalam generasi gambar dan pemrosesan gambar. Kelebihan utama *Diffusion Model* adalah kemampuan mereka untuk menghasilkan sampel dengan kualitas tinggi dan keragaman yang luas. *Diffusion Models* dapat memiliki kualitas output yang lebih baik daripada metode generatif lainnya seperti *Generative Adversarial Networks* (GAN) dan *Variational AutoEncoders* (VAE). Seperti yang ditunjukkan oleh penelitian (Chen dkk., 2024) *Diffusion Models* menunjukkan keberhasilan empiris yang signifikan dalam berbagai aplikasi, seperti visi komputer, audio, dan biologi komputasi. Selain itu, mereka memiliki potensi untuk membangun teori yang lebih mendalam tentang karakteristik statistik dan kemampuan sampling mereka. Hasil dari penelitian (Ahsan dkk., 2024) menjelaskan untuk mengembangkan metodologi yang lebih terarah dapat menggunakan dan meningkatkan Model *Diffusion* dalam aplikasi kreatif dan komersial, terutama dalam pembuatan konten visual yang kreatif.



Gambar 2. 1 *Forward Process/ Backward Process* (Chen *dkk.*, 2024)

Gambar 2.1 menunjukkan dua fase utama dalam model *diffusion*: fase maju (*forward diffusion*) dan fase mundur (*backward diffusion*). Pada fase

maju, data asli secara bertahap dicampur dengan *Noise* Gaussian melalui kernel transisi Gaussian, menghasilkan data yang semakin acak. Sebaliknya, fase mundur dimulai dari data yang penuh *Noise* dan secara bertahap memulihkan data asli yang memandu model untuk membalik proses difusi maju. Pendekatan ini memungkinkan model generatif menghasilkan data berkualitas tinggi dengan memahami struktur statistik data asli.

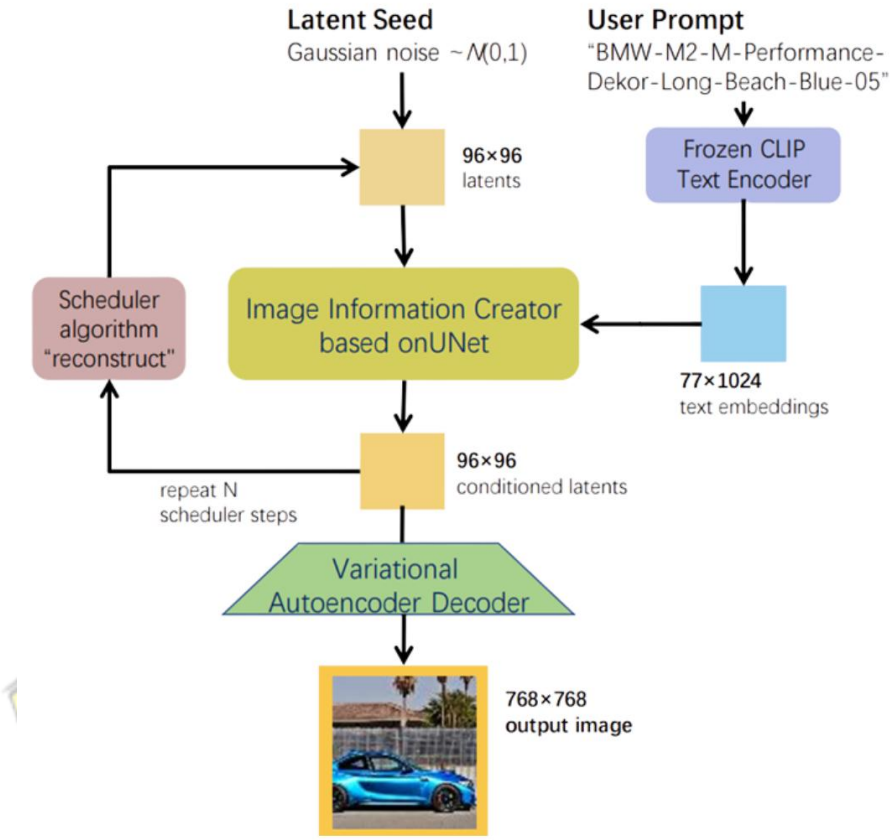
2.2.3 *Stable Diffusion*

Stable Diffusion adalah model generatif berbasis Diffusion Model yang dikembangkan oleh Stability AI. Model ini dirancang untuk menghasilkan gambar berkualitas tinggi dari deskripsi teks. Model ini bekerja dengan mengubah noise menjadi gambar terstruktur melalui dua fase: fase forward, di mana noise ditambahkan ke gambar, dan fase backward, di mana model membangun kembali gambar dari noise tersebut. *Stable Diffusion* menggunakan pendekatan *Latent Diffusion Model* (LDM), yang menggabungkan *autoEncoder* dengan model difusi yang dilatih dalam ruang *laten autoEncoder*. Selama pelatihan, gambar dikodekan melalui *Encoder* menjadi *representasi laten*. Proses ini dapat dirangkum dalam rumus berikut:

$$X' = D(UNet(E(X), f(P))) \quad (1)$$

Penjelasan :

- X : Gambar input dengan ukuran $H \times W \times 3$.
- $E(X)$: *Encoder* mengubah gambar menjadi representasi laten Z dengan ukuran $\frac{H}{8}$ x $\frac{W}{8}$ x 4.
- $f(P)$: Text *Encoder* OpenCLIP-ViT/H mengubah prompt teks P menjadi representasi vektor T.
- *U-Net*(Z,T) : *U-Net* memproses Z dengan kontrol dari T melalui mekanisme *cross attention.*
- $D(\cdot)$ *: Decoder* mengubah hasil dari UNet kembali menjadi gambar output X'.



Gambar 2. 2 Diagram *Stable Diffusion* (Gao *dkk*., 2024)

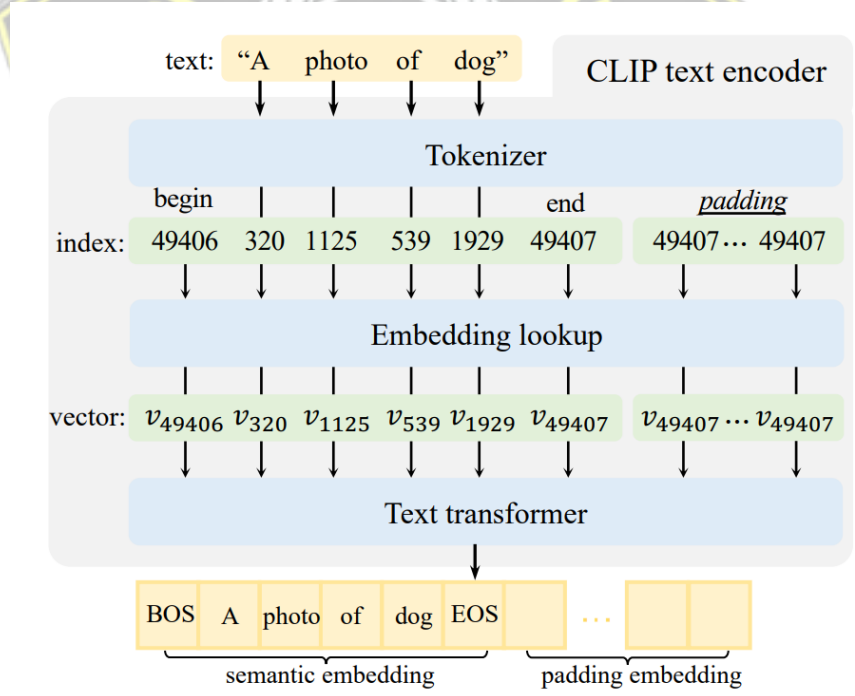
Gambar 2.2 adalah Diagram *Stable Diffusion* yang memulai prosesnya dengan *Latent Seed*, yaitu noise acak yang diambil dari distribusi. Pada saat yang sama, prompt teks yang diberikan oleh pengguna dikodekan menggunakan *Frozen CLIP Text Encoder*, menghasilkan representasi teks dalam bentuk *text embeddings*. Kedua informasi ini dikombinasikan dalam *Image Information Creator berbasis U-Net*, yang bertugas merekonstruksi gambar dari noise secara bertahap melalui serangkaian langkah yang dikontrol oleh *scheduler algorithm*. Setelah proses difusi selesai, hasilnya diterjemahkan kembali ke ruang piksel oleh *Variational Autoencoder Decoder (VAE Decoder)*, menghasilkan gambar akhir dengan resolusi tinggi.

Versi *Stable Diffusion* 2.1 menawarkan peningkatan kualitas gambar dengan resolusi dan detail yang lebih baik. Model ini dilatih menggunakan *text Encoder* baru yang dikembangkan oleh LAION, memberikan rentang ekspresi yang lebih luas dibandingkan dengan versi sebelumnya. *Stable*

Diffusion mendukung *Fine Tuning*, memungkinkan pengguna menyesuaikan model dengan *Dataset* spesifik untuk meningkatkan akurasi dan relevansi gambar yang dihasilkan. Dengan melakukan *Fine Tuning*, model dapat lebih responsif terhadap variasi dalam prompt teks dan menghasilkan output yang lebih sesuai dengan kebutuhan spesifik pengguna.

2.2.4 *Text To Image Generation*

Text to Image generation merupakan cabang *artificial intelligence* yang bertujuan mengonversi deskripsi teks menjadi gambar yang akurat dan sesuai. Proses ini mencakup beberapa komponen utama, seperti mengubah teks menjadi representasi vektor yang dapat dipahami model, serta memanfaatkan teknik generatif seperti GAN (*Generative Adversarial Networks*) atau *Diffusion Models* untuk menghasilkan gambar dari representasi tersebut (Tao *dkk.*, 2022)



Gambar 2. 3 *Text Encoder* pada CLIP (Yu *dkk.*, 2024)

Gambar 2.3 adalah *Text Encoder* pada CLIP untuk menilai kualitas gambar yang dihasilkan berdasarkan teks input, digunakan CLIP *Score*. Skor ini dihitung menggunakan kesamaan kosinus (*cosine similarity*) antara *embedding teks* dan embedding gambar dari model yang telah di *Fine Tuning*,

seperti yang ditampilkan pada Gambar 2.3. Secara umum, CLIP Score dapat dihitung dengan rumus berikut:

:

$$S = \frac{E_t . E_i}{||E_t||. ||E_i||}$$

S adalah Clip *Score*, nilai kesesuaian antara teks dan gambar, (2)

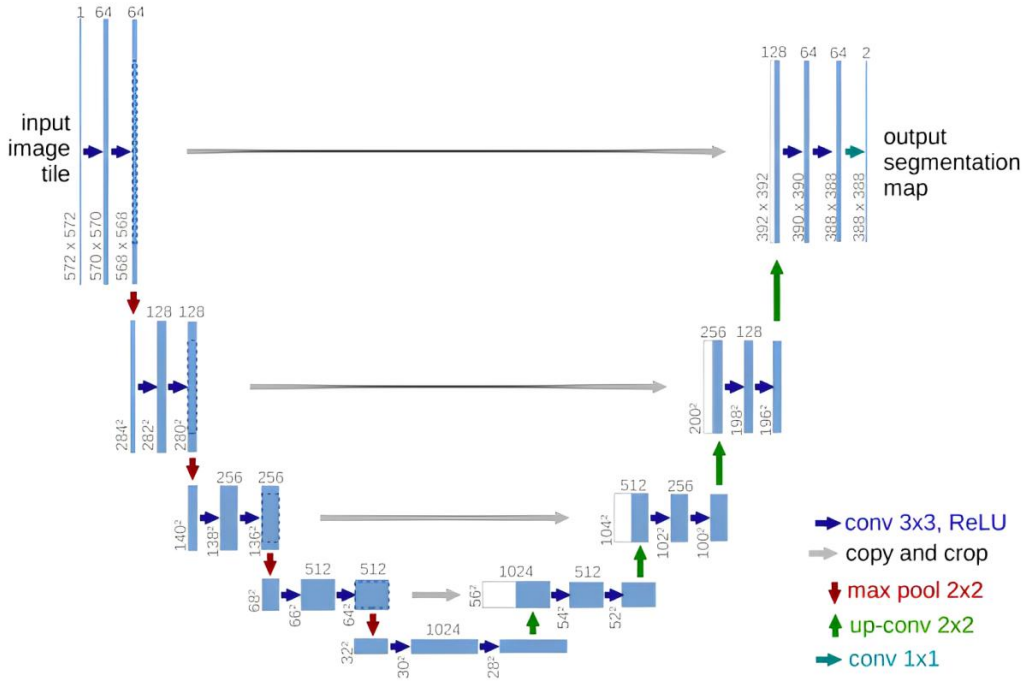
E_t adalah vector embedding dari teks (prompt input)

E_i adalah vector embedding dari gambar hasil dari model yang sudah di finetune

CLIP *Score* merupakan metrik evaluasi yang krusial dalam *Text-to-Image Generation*, karena mampu mengukur sejauh mana model memahami dan mengonversi deskripsi teks menjadi elemen visual yang sesuai. Nilai CLIP *Score* yang lebih tinggi menunjukkan tingkat kesesuaian yang lebih baik antara teks dan gambar yang dihasilkan, sehingga model menjadi lebih andal dalam menghasilkan gambar yang sesuai dengan berbagai variasi deskripsi teks (Li *dkk.*, 2023).

2.2.5 *U-Net*

U-Net adalah arsitektur *neural* yang fleksibel untuk berbagai tugas pemrosesan sinyal kontinu pada grid, seperti segmentasi gambar. *U-Net* menunjukkan kerangka desain dan analisis yang lebih komprehensif, yang mencakup peran khusus *Encoder* dan *Decoder* dalam skala resolusi tinggi, serta hubungan mereka ke *ResNets* melalui prakondisi. Metode ini memasukkan *Multi-ResNets*, versi *U-Net* dengan *Encoder* berbasis wavelet tanpa parameter yang dapat dipelajari. Metode ini memungkinkan segmentasi gambar, dan model generatif seperti model difusi dengan kinerja yang lebih baik. Hasil penelitian (Williams dkk., 2023) menunjukkan bahwa *U-Net* dengan *average pooling* yang memanfaatkan *Noise* frekuensi tinggi dengan baik. Ini memungkinkan pengembangan arsitektur *neural* yang lebih diskalabel dan alami di berbagai bidang aplikasi.



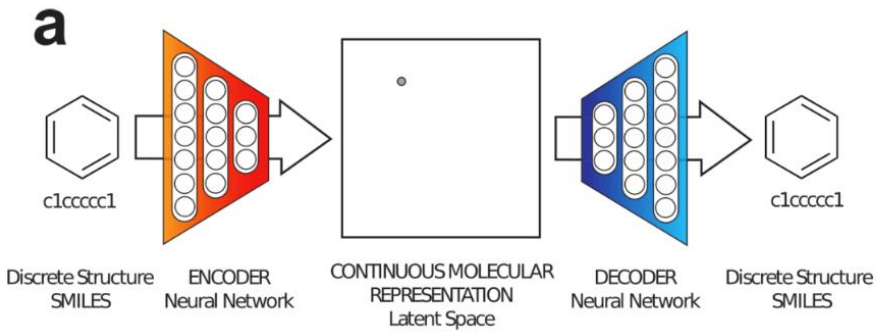
Gambar 2. 4 Arsitektur *U-Net* (Ronneberger *dkk.*, 2021)

Gambar 2.4 menunjukkan arsitektur *U-Net* yang digunakan untuk segmentasi citra, terdiri dari dua bagian utama: *Encoder* dan *Decoder*. *Encoder* berfungsi mengekstraksi fitur dari citra input melalui beberapa blok convolutional 3x3 dengan aktivasi ReLU (biru), diikuti oleh max pooling 2x2 (merah) untuk mengurangi dimensi spasial sambil meningkatkan jumlah fitur. Setelah mencapai titik terdalam, *Decoder* merekonstruksi citra menggunakan *up-convolution* 2x2 (hijau) untuk memperbesar dimensi spasial, serta menggabungkan fitur dari *Encoder* melalui operasi copy and crop (abu-abu) guna mempertahankan informasi spasial. Akhirnya, lapisan convolution 1x1 (hijau) menghasilkan peta segmentasi dengan jumlah channel sesuai jumlah kelas. Struktur simetris ini menjadikan *U-Net* sangat efektif untuk segmentasi citra medis dan tugas serupa yang memerlukan detail spasial yang akurat.

2.2.6 *Variational AutoEncoders*

Variational AutoEncoders (VAE) merupakan model generatif yang memanfaatkan pendekatan Bayesian untuk belajar merepresentasikan data dalam ruang laten berdimensi rendah dan menghasilkan data baru yang realistis. Dalam arsitektur VAE, *Encoder* bertugas mengembalikan nilai rata-

rata dan deviasi standar untuk setiap input, yang kemudian digunakan untuk mengambil sampel vektor laten dari distribusi probabilitas tersebut. Vektor laten ini dikirim ke *Decoder* untuk merekonstruksi kembali masukan awal. Selain untuk merekonstruksi data, VAE juga dirancang untuk menghasilkan vektor laten yang mengikuti distribusi normal, sehingga memungkinkan penciptaan data baru yang serupa dengan data pelatihan.



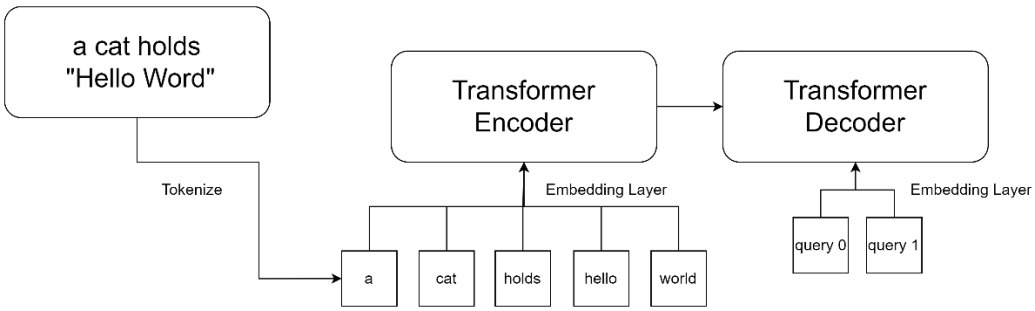
Gambar 2. 5 Arsitektur Dasar VAE (Kingma dan Welling, 2019)

Gambar 2.5 menunjukkan ilustrasi arsitektur dasar VAE, di mana *Encoder* mengubah data diskrit menjadi representasi kontinu dalam ruang laten, yang kemudian dapat digunakan untuk merekonstruksi kembali data awal melalui *Decoder*. Dalam konteks *text to image generation*, VAE diterapkan untuk mempelajari hubungan antara representasi teks dan gambar dalam ruang laten secara *variational*. Selain itu, ruang laten yang dihasilkan oleh VAE juga dapat digunakan untuk eksplorasi dan optimasi properti data, seperti dalam penelitian desain molekul generatif, di mana representasi laten memungkinkan pencarian molekul baru yang memiliki sifat optimal. Penelitian ini menunjukkan bahwa VAE dapat digunakan secara efektif untuk menghasilkan gambar dari teks meskipun tidak ada pasangan teks gambar yang tersedia selama pelatihan (Kang *dkk.*, 2023).

2.2.7 *Transformer*

Pada tahap awal pembuatan poster film berbasis AI, diperlukan proses generasi tata letak (layout generation) yang menentukan posisi elemen-elemen dalam gambar, seperti teks dan objek utama. Salah satu pendekatan

yang digunakan adalah dengan model berbasis *Transformer***,** yang mampu memahami hubungan antara teks dan tata letak dalam suatu desain.



Gambar 2. 6 *Layout Generation*

Gambar 2.6 ini menunjukkan alur kerja Stage *Layout Generation***,** di mana sebuah kalimat seperti *"a cat holds 'Hello World'"* ditokenisasi dan diproses melalui Transformer Encoder. Encoder ini mengubah teks menjadi representasi numerik yang dapat dipahami oleh model. Selanjutnya, Transformer Decoder menghasilkan prediksi tata letak, termasuk posisi dan bounding box untuk teks yang akan ditampilkan dalam poster. Hasilnya kemudian dirender menjadi gambar dan diikuti dengan pembuatan masking layout**,** yang menentukan area teks dalam desain akhir.

Setelah proses layout generation selesai, hasil tata letak yang dihasilkan dapat digunakan sebagai dasar untuk tahap selanjutnya**,** yaitu pembuatan visual poster**.** Model generatif seperti *Stable Diffusion* atau *Dreambooth* kemudian dapat memanfaatkan hasil layout ini untuk menghasilkan gambar sesuai dengan teks dan objek utama yang telah ditentukan sebelumnya. Dengan pendekatan ini, poster film yang dihasilkan tidak hanya mempertimbangkan kualitas visual tetapi juga memastikan elemen teks tertata dengan baik dan mudah dibaca.

2.2.8 *Fine Tuning Dreambooth*

Dreambooth memungkinkan pengguna untuk membuat representasi yang sangat spesifik dari suatu subjek hanya dengan beberapa gambar referensi. Proses ini bekerja dengan mengaitkan subjek yang diberikan dengan pengenal khusus, yang memungkinkan model *Diffusion* menghasilkan gambar yang

sangat realistis. Menariknya, metode ini hanya memerlukan tiga hingga lima gambar subjek, sehingga sangat praktis dan dapat diterapkan dalam berbagai situasi(Raj *dkk.*, 2023).

Selain memungkinkan penyesuaian hasil *Diffusion*, *DreamBooth* juga memastikan bahwa kualitas dan konsistensi visual gambar yang dihasilkan tetap terjaga. Teknik *Fine Tuning DreamBooth* sangat efektif dalam berbagai bidang kreatif, seperti desain grafis dan pembuatan poster film. Metode ini dapat menghasilkan citra baru yang sulit dibedakan dari citra awal. Dengan menggunakan *Fine Tuning DreamBooth*, pengguna dapat membuat variasi gambar yang beragam dan berkualitas tinggi (Ruiz dkk., 2023)

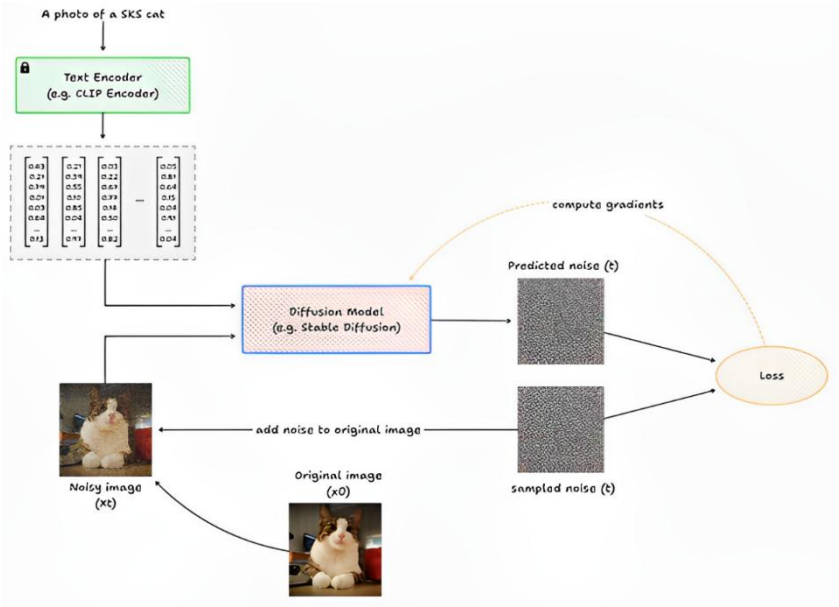
Rumus *Loss Function* digunakan untuk melatih model *Dreambooth*.

$$L = E_x, \epsilon \sim N(0, \mathrm{I}), t[\| \epsilon - \epsilon_\emptyset(x_t, t, c) \|^2] \quad (3)$$

Penjelasan :

- ϵ : noise yang ditambahkan.
- $\epsilon_\emptyset(x_t, t, c)$: noise yang diprediksi oleh model berdasarkan gambar (x_t), waktu difusi (t), dan kondisi teks (c, seperti nama objek atau gaya).

Dreambooth menggunakan fungsi loss untuk melatih model diffusion agar dapat mempelajari gaya atau konsep tertentu dari *Dataset* kecil. Fungsi ini menghitung selisih antara noise yang ditambahkan ke data gambar dan noise yang diprediksi oleh model berdasarkan kondisi tertentu, seperti deskripsi teks. Dengan pendekatan ini, model dapat memahami karakteristik spesifik dari *Dataset* sambil tetap mempertahankan kemampuan generatif aslinya. Pendekatan ini memastikan model mampu menghasilkan gambar baru yang konsisten dengan gaya atau konsep yang diinginkan, menjadikannya efektif untuk personalisasi kreatif.



Gambar 2. 7 Diagram Proses Pelatihan *Dreambooth*

Gambar 2.7 adalah Proses *Fine Tuning DreamBooth* dilakukan dengan mengambil gambar asli yang mengandung konsep baru, kemudian menambahkan *noise* secara bertahap untuk menghasilkan gambar yang terdegradasi. Gambar yang telah diberi noise ini kemudian dimasukkan ke dalam model *Diffusion* untuk memprediksi noise yang terdapat dalam gambar tersebut.

Perbandingan antara *noise* yang diprediksi oleh model dan *noise* yang sebenarnya telah ditambahkan ke gambar asli digunakan untuk menghitung fungsi *loss*. Hasil *loss* ini kemudian digunakan untuk menghitung gradien dan memperbarui parameter model *Diffusion* agar semakin akurat dalam memahami konsep yang diajarkan. Input yang diberikan ke model *Diffusion* terdiri dari gambar yang telah diberi noise, timestep yang telah disampel, serta teks prompt yang merepresentasikan konsep baru, misalnya: *"A picture of a SKS cat"*.

Dalam tahap awal, *embedding* teks yang mengandung konsep baru mungkin belum dikenali oleh model. Oleh karena itu, model *Diffusion* pada awalnya akan kurang akurat dalam menghilangkan noise. Namun, seiring dengan proses pelatihan yang dilakukan berulang kali, model akan semakin memahami konsep yang diajarkan dengan menghubungkan gambar yang

diberikan dengan token khusus dalam teks. Hal ini memungkinkan model untuk menghasilkan gambar dengan konsep yang lebih akurat sesuai dengan karakteristik yang diinginkan serta lebih efisien dalam waktu pelatihan meskipun memerlukan lebih banyak memori (Qowy *dkk.*, 2024)

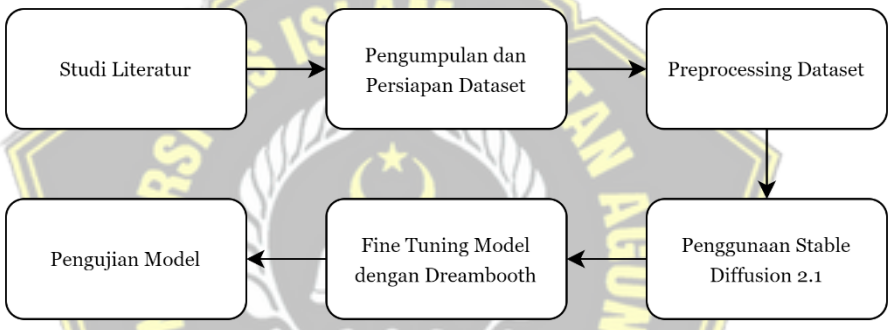
2.2.9 *Negative Prompt*

Negative prompt merupakan teknik dalam model generatif bersyarat seperti *Stable Diffusion* yang memungkinkan pengguna menentukan elemen yang harus dihindari dalam hasil yang dihasilkan. Studi oleh (Ban *dkk.*, 2024) bahwa *negative prompt* bekerja melalui dua mekanisme utama, yaitu *Delayed Effect* dan *Deletion Through Neutralization*. Efek tertunda terjadi karena *negative prompt* baru mulai mempengaruhi gambar setelah *positive prompt* menghasilkan objek yang sesuai, sedangkan mekanisme netralisasi bekerja dengan membatalkan noise positif yang merepresentasikan objek yang ingin dihapus. Selain itu, penelitian ini mengidentifikasi fenomena *Reverse Activation***,** di mana penerapan *negative prompt* terlalu awal justru dapat menyebabkan objek yang ingin dihapus malah muncul dalam hasil akhir. Untuk mengatasi hal ini, penelitian ini mengusulkan teknik *Controllable Inpainting***,** yaitu penerapan negative prompt setelah titik kritis dalam proses *reverse diffusion* guna meningkatkan efektivitas penghapusan objek tanpa mengubah latar belakang secara drastis. Studi ini memberikan wawasan mendalam tentang cara kerja *negative prompt* serta optimalisasinya dalam aplikasi generatif berbasis teks ke gambar.

BAB III
METODE PENELITIAN

3.1 Metode Penelitian

Pada tahap pelatihan, penulis akan membuat generator gambar dengan menggunakan *Stable Diffusion* v2.1 dan menerapkan *Fine Tuning* melalui metode *Dreambooth*. Kemudian, pada tahap pengembangan aplikasi berbasis web, penulis akan menggunakan Streamlit untuk membangun antarmuka yang memungkinkan pengguna memasukkan deskripsi karakter secara langsung dan mendapatkan hasil visualisasi gambar sesuai dengan parameter yang telah ditetapkan.



Gambar 3. 1 Tahapan Penelitian

Pada Gambar 3.1 yaitu tahap penelitian dimana peneliti melakukan dua tahap, yaitu tahap menggunakan *Stable Diffusion* dan tahap yang menerapkan Teknik *Fine Tuning Dreambooth*. Berikut adalah tahapan dalam penelitian ini :

3.1.1 Studi Literatur

Pada tahap ini, dilakukan peninjauan dan analisis terhadap sumber-sumber yang relevan. Peninjauan literatur mencakup berbagai aspek dari Diffusion Models, termasuk *Stable Diffusion, text-to-image generation,* serta komponen utama seperti *U-Net, VAE, dan Transformer*. Selain itu, ditinjau pula teknik *Fine Tuning Dreambooth,* yang memungkinkan model menghasilkan gambar dengan karakteristik yang lebih spesifik berdasarkan data yang diberikan.

3.1.2 Pengumpulan dan Persiapan *Dataset*

Pada tahap ini adalah pencarian dan pengumpulan *Dataset* poster minimalist yang hanya menampilkan judul tanpa elemen visual yang kompleks. Poster jenis ini umumnya menonjolkan tipografi dan desain teks yang sederhana namun menarik. Tujuan pengumpulan *Dataset* ini adalah untuk melatih model agar mampu menghasilkan desain poster yang fokus pada judul.

3.1.3 Preprocessing *Dataset*

Peneliti melakukan preprocessing *Dataset* untuk mempersiapkannya dalam pelatihan model menggunakan *Dreambooth*. Setelah *Dataset* terkumpul, dilakukan proses pengolahan dengan langkah langkah berikut :

- Penamaan *Dataset*, yaitu memberikan nama file yang konsisten dan deskriptif untuk setiap gambar agar mudah diidentifikasi dan diorganisir. Penamaan yang terstruktur membantu dalam proses pelabelan dan meminimalkan kesalahan saat pengolahan data.
- Mengubah resolusi gambar menjadi 512x512 piksel. Resolusi ini dipilih karena kompatibel dengan arsitektur model generatif yang digunakan, seperti *Stable Diffusion*, yang umumnya dioptimalkan untuk resolusi tersebut. Dengan menyamakan resolusi, model dapat mempelajari fitur visual dengan lebih konsisten, sehingga meningkatkan akurasi hasil generasi gambar.
- Pemberian captioning pada setiap gambar untuk memberikan konteks teks yang sesuai dengan konten visual. Caption ini digunakan sebagai prompt teks saat pelatihan model, sehingga model dapat memahami hubungan antara teks dan gambar. Caption ditulis secara ringkas namun tetap deskriptif untuk membantu model menghasilkan gambar yang relevan dengan prompt yang diberikan.

Dalam penelitian ini, peneliti menggunakan 50 dataset poster film yang hanya menampilkan judul dan objek utama. Dataset tersebut dikumpulkan dari berbagai platform online.

Setelah proses penamaan, pengubahan resolusi, dan penambahan caption selesai, caption dan informasi *Dataset* digabungkan menjadi satu file CSV. CSV ini berisi kolom yang terstruktur dengan rapi, seperti nama file gambar dan caption yang sesuai, sehingga memudahkan integrasi dengan *Dreambooth* saat pelatihan. Format CSV dipilih karena sederhana dan kompatibel dengan banyak framework machine learning. Dengan *Dataset* yang sudah terorganisir dan terdokumentasi dalam satu file CSV, proses pelatihan model menjadi lebih efisien dan terstruktur.

3.1.4 Penggunaan *Stable Diffusion* 2.1

Pada tahap ini, model akan dipersiapkan dengan konfigurasi awal untuk menghasilkan gambar berdasarkan deskripsi teks.

- *U-Net* Tunggal dengan *Cross-Attention*
  - Menggunakan arsitektur *U-Net* yang dioptimalkan dengan mekanisme *cross-attention* untuk meningkatkan kualitas gambar.
  - *U-Net* bekerja dalam ruang laten untuk mengurangi kompleksitas komputasi tanpa mengorbankan detail.
  - Proses *denoising* dilakukan secara bertahap dalam iterasi, menghasilkan gambar dengan resolusi tinggi dan detail yang lebih baik.
- CLIP *Encoder* (OpenCLIP-ViT/H)
  - Text *Encoder* terbaru yang lebih akurat dalam memahami deskripsi teks.
  - Peningkatan pada text *Encoder* memungkinkan model merespons variasi prompt dengan lebih baik dan menghasilkan gambar yang lebih sesuai dengan instruksi pengguna.
- *Latent Seed Initialization*
  - Model tetap menggunakan ruang laten dengan *Gaussian Noise*, tetapi dengan distribusi noise yang lebih stabil untuk hasil yang lebih realistis.
- VAE Auto*Encoder* (VAE) *Decoder*
  - VAE *Decoder* mengubah representasi laten menjadi gambar akhir dengan warna dan detail yang lebih akurat.

- *Stable Diffusion* 2.1 mendukung resolusi gambar hingga 768×768 piksel secara default (lebih tinggi dari 512×512 pada versi 1.5), memberikan hasil yang lebih tajam dan lebih kaya detail.

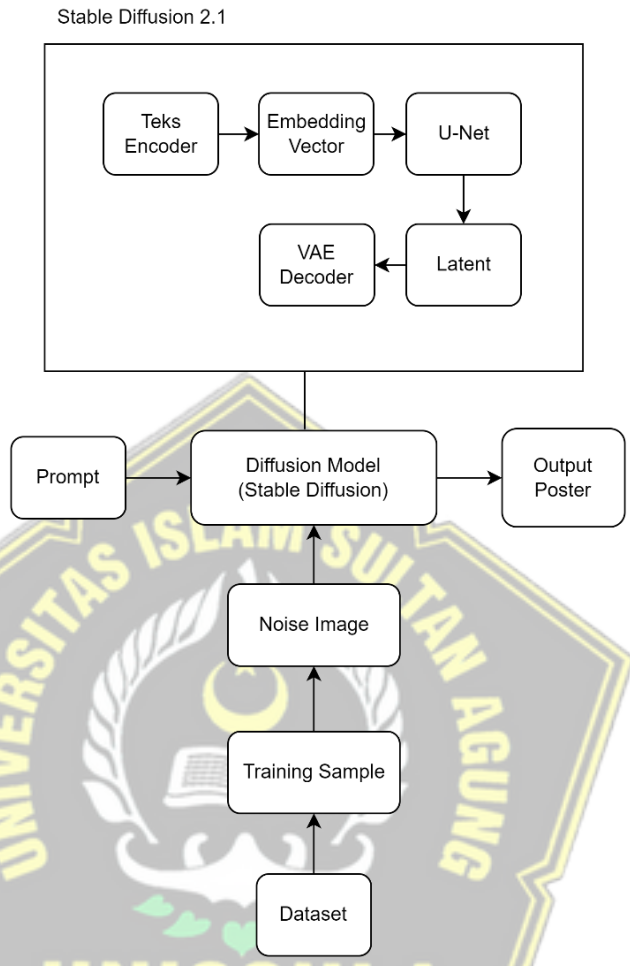
3.1.5 *Fine Tuning* Model dengan *Dreambooth*

Pada tahap ini, model yang telah dibangun melalui pelatihan dasar dapat ditingkatkan lagi melalui *Fine Tuning* menggunakan *Dreambooth*. Proses ini memberi *Dreambooth* kemampuan tambahan untuk mengidentifikasi karakteristik khusus yang diinginkan, seperti gaya atau elemen tertentu yang lebih spesifik. Dengan cara ini, model dapat lebih disesuaikan untuk menghasilkan gambar yang lebih dekat dengan deskripsi yang diinginkan, termasuk penyesuaian karakter, ekspresi.

3.1.6 Pengujian Model

Evaluasi kualitas gambar dilakukan menggunakan CLIP *Score*, yang mengukur kesesuaian antara gambar yang dihasilkan dengan deskripsi teks yang sesuai. CLIP (*Contrastive Language Image Pretraining)* adalah model yang dilatih untuk memahami hubungan antara teks dan gambar dalam ruang representasi bersama. CLIP *Score* dihitung berdasarkan kesamaan kosinus antara vektor fitur gambar dan vektor fitur teks deskriptif. Skor ini berada dalam rentang 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan bahwa gambar yang dihasilkan lebih relevan dengan teks deskripsi.

3.2 Alur Kerja Training Sistem



Gambar 3. 2 Alur Kerja Training Sistem

Alur dari sistem ini adalah untuk menghasilkan poster film animasi yang sesuai dengan deskripsi teks dan gaya visual dari film tersebut. Pendekatan yang digunakan adalah memanfaatkan model pembelajaran mesin berbasis *Diffusion Model*, yaitu *Stable Diffusion* 2.1. Dengan memanfaatkan model *Stable Diffusion* 2.1 yang telah di *Fine Tuning* menggunakan *Dreambooth*, sistem ini diharapkan dapat menghasilkan poster film yang sesuai. Berikut adalah alur kerja training menggunakan *Stable Diffusion* XL dan *Fine Tuning Dreambooth* yang ditunjukkan pada gambar 3.2:

1. Persiapan Training Sample:

   Mengumpulkan *Dataset* yang terdiri dari gambar-gambar terkait dengan film animasi, seperti karakter utama, latar, dan elemen visual khas.

Pastikan *Dataset* ini mencakup variasi yang cukup agar model dapat belajar dengan baik. *Dataset* ini nantinya akan digunakan untuk proses *Fine Tuning* model *Stable Diffusion* 2.1 menggunakan teknik *Dreambooth*.

2. Transformasi *Noise* Gambar:

   Selain *Dataset* gambar asli, proses *Fine Tuning* juga membutuhkan versi gambar yang telah diberikan *Noise*. *Noise* gambar ini dibuat dengan menambahkan gangguan acak pada *Dataset* gambar asli. Penambahan *Noise* ini bertujuan untuk melatih model agar dapat menghilangkan *Noise* dan menghasilkan gambar yang lebih realistis. *Transformasi Noise* gambar ini dilakukan sebagai bagian dari algoritma *Diffusion* yang digunakan oleh *Stable Diffusion* 2.1.

3. *Fine Tuning Stable Diffusion* 2.1 dengan *Dreambooth*:

   Model *Stable Diffusion* 2.1 merupakan model generatif yang telah dilatih pada *Dataset* gambar umum, sehingga memiliki kemampuan untuk menghasilkan gambar yang realistis. Namun, untuk menghasilkan poster film animasi yang spesifik, perlu dilakukan *Fine Tuning* menggunakan *Dreambooth*. *Dreambooth* adalah teknik *Fine Tuning* yang memungkinkan model untuk belajar dari *Dataset* yang relatif kecil, namun tetap dapat menghasilkan gambar yang sesuai dengan domain atau konsep tertentu. Dalam proses *Fine Tuning* ini, model *Stable Diffusion* 2.1 akan diajarkan untuk menghasilkan Poster Film. *Dataset* gambar asli dan gambar dengan *Noise* akan digunakan sebagai input untuk proses *Fine Tuning*.

4. Proses Pembuatan Poster:

   Setelah *Fine Tuning*, model *Stable Diffusion* 2.1 yang telah disesuaikan siap digunakan untuk membuat poster film animasi.

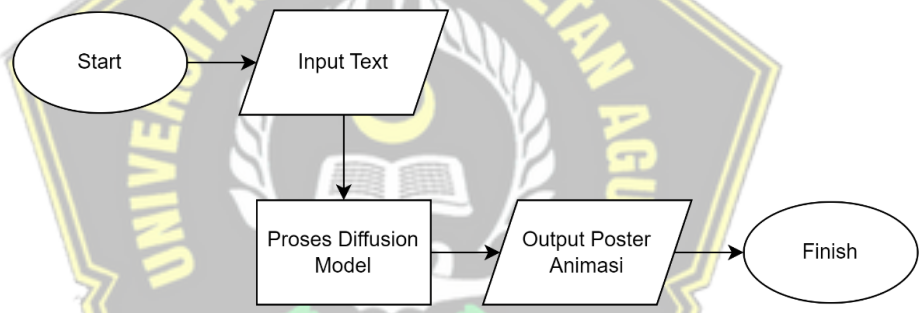
   Pengguna dapat memasukkan deskripsi teks. Teks deskripsi akan diproses melalui *Text Encoder* untuk diubah menjadi *representasi numerik* (*vektor embedding*). Hasilnya akan menjadi input bagi model *U-Net* dasar untuk menghasilkan representasi laten. Representasi laten tersebut selanjutnya

akan diproses oleh *Diffusion Model* (*Stable Diffusion* 2.1) untuk menghasilkan gambar poster film animasi.

Dengan alur kerja ini, sistem dapat memanfaatkan kekuatan model pembelajaran mesin, khususnya *Stable Diffusion* 2.1 yang telah di *Fine Tuning* menggunakan *Dreambooth*, untuk menghasilkan poster film animasi yang sesuai dengan gaya visual dan karakteristik film tersebut. Proses evaluasi dan iterasi dapat dilakukan untuk menyempurnakan hasil sesuai dengan kebutuhan.

3.3 Alur Kerja User

Setelah memahami alur sistem secara teknis, berikut adalah alur yang akan dilalui oleh pengguna dalam menggunakan system ini agar dapat menghasilkan poster dari deskripsi teks.



Gambar 3. 3 Alur Kerja User

Gambar 3.3 adalah alur pengguna untuk menggunakan sistem pembuatan poster film animasi berbasis *Stable Diffusion* XL dan *Dreambooth*:

1. Start
   - Proses dimulai
2. Input Text Prompt
   - Pengguna memberikan deskripsi teks atau "prompt" yang berisi informasi visual mengenai poster animasi yang akan dihasilkan.
3. Proses *Diffusion Model*
   - Model yang telah melewati proses *Fine Tuning* akan diproses dengan teks yang diinpukan dan menghasilkan output.

4. Output Poster Animasi
   - Model menghasilkan poster animasi berdasarkan input teks dan hasil proses *Fine Tuning*.

3.4 Analisis Kebutuhan Sistem

Untuk memastikan bahwa sistem yang dibangun dapat memenuhi tujuan penelitian, yaitu menghasilkan gambar poster film animasi berdasarkan deskripsi teks, analisis kebutuhan sistem mencakup perangkat lunak dan perangkat keras, serta kebutuhan fungsional dan non-fungsional sistem.

- Perangkat Keras
  a. Komputer atau Laptop: Harus memiliki prosesor dual-core, seperti Intel i5 atau AMD Ryzen 5. Sistem operasi (Windows, macOS, atau Linux) yang mendukung browser web.
  b. Akses Internet Stabil: Anda harus dapat mengakses Google Colab dan mengunggah dan mengunduh model dan data.
- Perangkat Lunak
  a. Google Colab : Akses GPU atau TPU di Google Colab (misalnya, Tesla T4 atau K80) untuk mendukung pelatihan dan inferensi model.
  b. *Library*

Tabel 3. 1 Tabel *Library*

Library	**Deskripsi**	**Fungsi Dalam Penelitian**
PyTorch	*PyTorch* adalah *Library deep learning* berbasis Python yang digunakan untuk membangun, melatih, dan menjalankan model pembelajaran mesin. Itu mendukung komputasi tensor dan peningkatan kecepatan	• Digunakan sebagai framework utama untuk menjalankan *Stable Diffusion* 2.1 dan *Dreambooth*. • Memfasilitasi *Fine Tuning* model dengan mengoptimalkan bobot menggunakan gradient descent.

	GPU, menjadikannya pilihan utama untuk deep learning dan termasuk dalam model difusi seperti *Stable Diffusion*.	• Mendukung komputasi berbasis GPU, mempercepat proses training dan inferensi. • Menyediakan *automatic differentiation (autograd)* untuk pelatihan model.
*Transformer*s	*Library* dari Hugging Face yang menyediakan model berbasis arsitektur *Transformer*, seperti CLIP, yang digunakan untuk evaluasi teks-gambar dalam penelitian ini.	• Digunakan untuk evaluasi hasil poster film animasi dengan CLIP *Score*. • CLIP *(Contrastive Language-Image Pretraining)* mencocokkan deskripsi teks dengan gambar untuk menilai seberapa relevan gambar yang dihasilkan oleh *Stable Diffusion*.
Diffusers	Diffusers adalah *Library* dari Hugging Face yang dirancang untuk bekerja dengan denoising diffusion models, termasuk *Stable Diffusion* 2.1.	• Digunakan untuk menghasilkan poster film dari deskripsi teks menggunakan *Stable Diffusion* 2.1. • Mendukung text-to-image generation, memungkinkan konversi

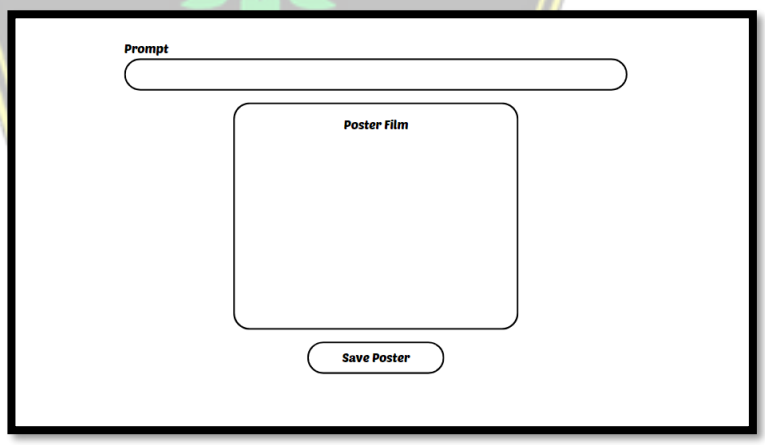
		teks deskripsi film menjadi gambar poster. • Memfasilitasi *Fine Tuning Dreambooth*, yang digunakan untuk menyesuaikan model *Stable Diffusion* agar menghasilkan poster dengan karakteristik tertentu. • Menyediakan pipeline inference untuk mempercepat proses generasi gambar.
Pandas	Pandas adalah *Library* Python yang digunakan untuk manipulasi dan analisis data dalam format tabel (DataFrame).	• Menyimpan dan mengelola *Dataset* deskripsi teks dan hasil poster yang dihasilkan. • Membantu dalam penganalisisan hasil evaluasi seperti menyimpan skor CLIP untuk setiap poster. • Memfasilitasi pengolahan data latih sebelum digunakan untuk *Fine Tuning Dreambooth*.
Numpy	*Library* Python untuk komputasi numerik	• Digunakan untuk komputasi tensor dan

	yang mendukung operasi array multidimensi dan aljabar linear.	matriks dalam training model diffusion. • Membantu dalam pengolahan data gambar yang dihasilkan oleh *Stable Diffusion*. • Dipakai dalam perhitungan CLIP *Score*, karena banyak operasi vektor yang digunakan dalam evaluasi kesamaan teks-gambar.
Matplotlib	*Library* Python untuk visualisasi data, terutama dalam bentuk grafik dan gambar.	• Digunakan untuk menampilkan dan membandingkan hasil poster yang dihasilkan oleh model. • Membantu dalam visualisasi skor CLIP, misalnya dengan membuat grafik hubungan antara deskripsi teks dan kualitas poster. • Memfasilitasi analisis visual terhadap performa *Fine Tuning Dreambooth*.

3.5 Perancangan User Interface

Gambar 3. 4 Tampilan Awal Sistem

Untuk AI Image Generator, User Interface ini dirancang secara minimalis dengan fokus pada kemudahan penggunaan dan fungsionalitas. Pada tampilan awal, halaman dimulai dengan judul "Poster Film Animasi" di bagian atas sebagai penanda utama bahwa ini akan membuat poster dari deskripsi teks. Di bawahnya, terdapat kolom input prompt berbentuk persegi panjang dengan latar, di mana pengguna dapat memasukkan prompt atau deskripsi gambar yang ingin dihasilkan. Tepat di bawah kolom input, terdapat tombol "*Generate*" yang mudah dikenali untuk memungkinkan pengguna membuat poster.



Gambar 3. 5 Tampilan saat generasi poster

Setelah tombol "*Generate*" ditekan, antarmuka akan menampilan hasil Poster di bagian bawah halaman, terdapat area besar berbentuk persegi panjang yang berfungsi sebagai tempat output poster ditampilkan setelah proses selesai. Fokus utama elemen ini adalah Output Poster Film Animasi.

BAB IV
HASIL DAN ANALISIS PENELITIAN

4.1 Preprocessing *Dataset*

4.1.1 Mount Goggle Drive

Proses Mounting Google Drive di Google Colab menggunakan Python. Langkah ini sangat penting sebelum melakukan preprocessing *Stable Diffusion* 2.1 dan *Fine Tuning* menggunakan *Dreambooth*, karena memungkinkan akses langsung ke file dan *Dataset* yang tersimpan di Google Drive. Dengan menjalankan kode berikut Google Colab akan meminta izin autentikasi pengguna untuk menghubungkan penyimpanan cloud mereka ke lingkungan Colab, sehingga semua file dapat diakses melalui direktori `/content/drive`.

Setelah proses mounting berhasil, pengguna dapat membaca, menyimpan, atau memodifikasi *Dataset* yang diperlukan untuk pelatihan model. Ini sangat berguna dalam pelatihan yang memerlukan penyimpanan besar, seperti pelatihan model *Stable Diffusion* dan *Dreambooth*. Dengan menyimpan *Dataset* dan model di Google Drive, pengguna dapat dengan mudah melanjutkan tanpa kehilangan data, sekaligus menghemat ruang penyimpanan di Colab yang terbatas.

4.1.2 Penyesuaian *Dataset*

Dalam tahap awal penelitian ini, dilakukan proses preprocessing untuk memastikan data gambar yang digunakan sesuai dengan kebutuhan model. Proses ini mencakup beberapa langkah penting, yaitu penamaan ulang file agar lebih terstruktur, mengubah resolusi gambar agar seragam dan optimal untuk pelatihan model, serta menambahkan caption yang berfungsi sebagai deskripsi teks dari setiap gambar. Untuk mempermudah dan mengotomatisasi proses ini, dibuat sebuah fungsi khusus yang menangani setiap tahap tersebut secara sistematis.

```
  def preprocess_images_and_Generate_caption(input_dir,
output_dir, target_size):
      poster_counter = 1
```

```
  for img_file in tqdm(os.listdir(input_dir),
desc="Processing Images"):
      input_path = os.path.join(input_dir, img_file)

  if not img_file.lower().endswith((".png", ".jpg", ".jpeg",
".bmp", ".gif")):
      continue
      try:

  with Image.open(input_path) as img:

      img = img.convert("RGB")

      img_resized = img.resize(target_size,
Image.Resampling.LANCZOS)

      output_filename = f"pop art{poster_counter}.png"
      output_path = os.path.join(output_dir, output_filename)

      img_resized.save(output_path, format="PNG", quality=95)

      inputs = processor(images=img, return_tensors="pt")
      out = model.Generate(**inputs)
      caption = processor.decode(out[0],
skip_special_tokens=True)

      caption_filename = f"poster{poster_counter}.txt"
      caption_path = os.path.join(output_dir,
caption_filename)
      with open(caption_path, "w") as caption_file:
      caption_file.write(caption)

       poster_counter += 1

       except Exception as e:
         print(f"Error processing {img_file}: {e}")
```

Source Code diatas adalah Fungsi untuk memproses gambar yang sekaligus menghasilkan deskripsi, Fungsi ini bertugas untuk memproses gambar dalam sebuah direktori dengan beberapa langkah utama: mengganti nama file, mengubah ukuran gambar, dan menghasilkan caption otomatis menggunakan model BLIP. Proses dimulai dengan membaca daftar file dalam direktori input, kemudian setiap gambar yang valid (dengan format .png, .jpg, .jpeg, .bmp, atau .gif) akan diproses lebih lanjut. Untuk memastikan semua gambar dalam format yang seragam, gambar dikonversi ke mode RGB, sehingga kompatibel dengan berbagai aplikasi yang membutuhkan standar warna tertentu.

Setelah dikonversi, gambar akan diubah ukurannya ke dimensi target yang telah ditentukan menggunakan metode LANCZOS, yang merupakan teknik resampling berkualitas tinggi untuk mempertahankan detail gambar. File yang telah diproses kemudian disimpan dalam format PNG di direktori output dengan nama yang telah disesuaikan, seperti poster1.png, poster2.png, dan seterusnya. Dengan cara ini, setiap gambar memiliki struktur penamaan yang lebih rapi dan mudah diorganisir.

Selain pemrosesan gambar, fungsi ini juga menggunakan BLIP (*Bootstrapped Language-Image Pretraining*) untuk menghasilkan deskripsi otomatis dari setiap gambar. Model BLIP menerima input gambar yang sudah diproses, kemudian menghasilkan teks deskriptif yang menggambarkan isi gambar tersebut. Hasil deskripsi ini disimpan dalam file .txt dengan nama yang sesuai dengan gambar, seperti poster1.txt, poster2.txt, dan seterusnya, sehingga setiap gambar memiliki pasangan file teks yang berisi caption-nya.

Untuk menghindari kegagalan dalam pemrosesan seluruh *Dataset*, fungsi ini dilengkapi dengan error handling. Jika terjadi kesalahan dalam membaca atau memproses sebuah gambar, pesan error akan ditampilkan, tetapi proses tetap berlanjut untuk gambar lainnya. Dengan pendekatan ini, fungsi dapat berjalan secara efisien tanpa harus terhenti hanya karena satu file mengalami masalah, sehingga memastikan *Dataset* tetap terolah dengan baik.

Selanjutnya kode fungsi itu dijalankan dengan memanggil `preprocess_images_and_Generate_caption()` dan juga dengan memasukkan tiga parameter utama: `INPUT_DIR` sebagai lokasi gambar asli, `OUTPUT_DIR` sebagai tempat menyimpan hasil *preprocessing*, dan `TARGET_SIZE` sebagai ukuran target gambar. Fungsi ini akan membaca semua gambar dalam direktori input, mengubah ukurannya sesuai spesifikasi, serta menghasilkan deskripsi otomatis menggunakan model BLIP.

Setiap gambar yang telah diproses akan disimpan dengan nama berurutan di direktori output, baik dalam bentuk file gambar maupun file teks yang berisi caption yang dihasilkan. Dengan menjalankan kode ini, seluruh proses preprocessing dan pembuatan caption berlangsung secara otomatis, memastikan *Dataset* siap digunakan untuk analisis lebih lanjut atau pelatihan model kecerdasan buatan.

4.2 *Stable Diffusion* 2.1

4.2.1 Pipeline For Image Generation

Sebelum menjalankan proses inferensi dengan *Stable Diffusion*, langkah pertama yang perlu dilakukan adalah menginisialisasi model yang akan digunakan. Dalam penelitian ini, digunakan model *Stable Diffusion* v2.1 yang dikembangkan oleh Stability AI.

Setelah model diinisialisasi, langkah berikutnya adalah memuat pipeline dari model yang telah dipilih serta mengonfigurasi scheduler untuk mengontrol proses difusi. Dalam penelitian ini, digunakan `DPMSolverMultistepScheduler`, yang dikenal mampu meningkatkan efisiensi dan kualitas hasil gambar. Model kemudian dipindahkan ke perangkat GPU (CUDA) agar proses inferensi berjalan lebih cepat dan optimal.

Dengan pipeline yang telah dikonfigurasi dan dijalankan di GPU, kini model siap digunakan untuk menghasilkan gambar berdasarkan deskripsi teks yang diberikan. Langkah selanjutnya adalah memberikan prompt teks dan menjalankan proses inferensi untuk menghasilkan gambar yang sesuai dengan deskripsi.

4.3 *Fine Tuning Dreambooth*

4.3.1 Training

```
!python3 /content/drive/MyDrive/scripts/train_dreambooth.py \
--pretrained_model_name_or_path="stabilityai/stable-diffusion-2-1" \
--output_dir="/content/drive/MyDrive/Fine Tuned Model" \
--csv_file="/content/drive/MyDrive/Dataset/Dataset.csv" \
--revision="main" \
--seed=777 \
--resolution=512 \
--train_batch_size=1 \
--train_text_encoder \
--mixed_precision="fp16" \
--use_8bit_adam \
--gradient_accumulation_steps=1 \
--learning_rate=1e-6 \
--lr_scheduler="constant" \
--lr_warmup_steps=80 \
--max_train_steps=1000
```

Gambar 4. 1 Menjalakan *Training Dreambooth*

Pada Gambar 4.1 adalah Kode yang menjalankan skrip pelatihan *Dreambooth* yang ada di direktori Google Drive menggunakan *Stable Diffusion* 2.1 dengan berbagai parameter yang disesuaikan

- `--pretrained_model_name_or_path` : Menggunakan model dasar *Stable Diffusion* 2.1 dari Stability AI.
- `--output_dir` : Model yang sudah di-finetune akan disimpan di folder Fine Tuned Model.
- `--csv_file` : *Dataset* yang digunakan untuk pelatihan diambil dari file CSV yang berlokasi di *Dataset*.csv.
- `--revision` : Menggunakan versi utama (main) dari model yang diunduh.
- `--seed` : Seed 777 digunakan agar hasil pelatihan tetap konsisten.
- `--resolution` : Melatih model dengan ukuran gambar 512x512.
- `--train_batch_size` : Menggunakan batch size 1, artinya model akan diproses satu per satu per iterasi.
- `--train_text_Encoder` : Melatih text *Encoder* agar dapat lebih memahami deskripsi teks.

- `--mixed_precision` : Menggunakan Floating Point 16-bit (fp16) untuk mempercepat komputasi.
- `--use_8bit_adam` : Menggunakan 8-bit Adam optimizer, yang lebih hemat memori.
- `--gradient_accumulation_steps` : Menentukan jumlah iterasi sebelum update parameter. Di sini, 1 berarti parameter diperbarui setiap iterasi.
- `--learning_rate` : Learning rate kecil (0.000001) untuk memastikan model belajar secara bertahap.
- `--lr_scheduler` : Menggunakan constant learning rate, tanpa penurunan secara bertahap.
- `--lr_warmup_step` : Mengatur warmup 80 langkah pertama sebelum model mulai belajar secara optimal.
- `--max_train_steps` : Model akan dilatih selama 1.000 langkah.

4.3.2 Pipeline *Dreambooth*

Gambar 4. 2 Memuat Model *Fine Tuning*

Gambar 4.2 adalah kode yang bertujuan untuk memuat model yang telah di *fine-tune* menggunakan *Dreambooth* di Google Colab. Pertama, variabel `fine_tuned_model_dir2` menyimpan path direktori tempat model yang sudah di-fine-tune disimpan, yaitu di Google Drive pada folder "Fine Tuned Model/1000". Kemudian, fungsi `load_model()` digunakan untuk memuat model dari direktori tersebut ke dalam variabel `pipeline2`. Model ini di-load ke perangkat GPU dengan menyertakan parameter `device="cuda"`, yang mempercepat proses komputasi menggunakan CUDA dari NVIDIA. Setelah itu, terdapat output yang menunjukkan bahwa proses loading model sedang berlangsung, dengan indikator progres mencapai 100%. Proses ini

memerlukan waktu sekitar 1 menit 36 detik dengan kecepatan 13,65 iterasi per detik.

4.4 Hasil *Generate* Poster Film Animasi

4.4.1 Penggunaan *Stable Diffusion*

Hasil generasi gambar menggunakan *Stable Diffusion* 2.1 dengan prompt: *"animated movie poster 'rocket' only one spaceship is prominently displayed in the center"*. Model ini digunakan untuk menghasilkan ilustrasi bergaya poster animasi dengan fokus utama pada sebuah pesawat luar angkasa di Tengah.

Gambar 4. 3 *Output Stable Diffusion*

Gambar 4.3 menunjukkan bahwa model telah berhasil menghasilkan ilustrasi yang sesuai dengan prompt yang diberikan. Namun, berdasarkan evaluasi menggunakan CLIP, gambar ini memperoleh nila**i** 0,35.

Nilai itu menunjukkan bahwa meskipun gambar sudah cukup menggambarkan tema yang diminta, terdapat beberapa aspek yang mungkin tidak sepenuhnya sesuai, seperti jumlah objek utama yang lebih dari satu atau elemen tambahan yang kurang mendukung prompt secara optimal.

4.4.2 Penggunaan *Stable Diffusion* dan *Negative prompt*

Hasil generasi menggunakan *Stable Diffusion* 2.1 dengan prompt yang sama seperti sebelumnya *"animated movie poster 'rocket' only one spaceship*

is prominently displayed in the center", namun dengan tambahan *negative prompt***.**

Negative prompt digunakan untuk menghindari elemen-elemen yang tidak diinginkan dalam gambar, sehingga hasil yang dihasilkan lebih sesuai dengan ekspektasi. Model ini mencoba mengurangi objek-objek tambahan yang dapat mengganggu komposisi utama atau menyebabkan hasil yang tidak realistis.

Gambar 4. 4 *Output* SD dan *Neg Prompt*

Gambar 4.4 menunjukkan adanya perubahan dalam komposisi dibandingkan dengan gambar sebelumnya. Dengan penerapan *negative prompt*, model berhasil menyaring beberapa elemen yang tidak diinginkan, namun masih terdapat beberapa aspek yang kurang sesuai, seperti teks yang dihasilkan secara tidak akurat.

Dari evaluasi menggunakan CLIP *Score*, gambar ini memperoleh nilai 0,38. Nilai ini sedikit lebih tinggi dibandingkan dengan sebelumnya (0,35), yang menunjukkan bahwa model sedikit lebih sesuai dengan prompt yang diberikan. Namun, peningkatan ini masih relatif kecil, sehingga dapat disimpulkan bahwa *negative prompt* membantu dalam beberapa aspek, tetapi tidak sepenuhnya menghilangkan ketidaksesuaian dalam gambar.

4.4.3 Penggunaan *Fine Tuning*

Hasil generasi menggunakan *Stable Diffusion* 2.1**,** namun kali ini dengan model yang telah di-*Fine Tuning***.** Proses *Fine Tuning* dilakukan untuk meningkatkan kesesuaian antara prompt dan output gambar dengan lebih optimal, sehingga menghasilkan ilustrasi yang lebih akurat dan sesuai dengan ekspektasi. Prompt yang digunakan tetap sama: *"animated movie poster 'rocket' only one spaceship is prominently displayed in the center"*. Dengan model yang telah dioptimalkan hasil generasi gambar semakin mendekati prompt yang diberikan, baik dalam aspek komposisi, detail, maupun kualitas visual secara keseluruhan**.**

Gambar 4. 5 Output *Fine Tuning*

Gambar 4.5 menunjukkan perubahan yang baik dibandingkan dengan model sebelumnya. Detail dalam ilustrasi menjadi lebih jelas, dan komposisi tampak lebih mendekati konsep poster animasi dengan fokus utama pada sebuah roket di tengah.

Dari evaluasi menggunakan CLIP *Score*, gambar ini memperoleh nilai 0,61, yang menunjukkan peningkatan yang cukup besar dibandingkan nilai sebelumnya (0,35 dan 0,38). Hal ini mengindikasikan bahwa model fine-tuned lebih mampu menangkap elemen-elemen yang sesuai dengan deskripsi prompt.

Namun, meskipun hasilnya lebih baik, masih terdapat beberapa elemen tambahan yang tidak sepenuhnya diharapkan, seperti munculnya objek-objek yang tidak relevan (misalnya elemen asing yang tidak terkait dengan roket atau luar angkasa). Ini menunjukkan bahwa meskipun *Fine Tuning* meningkatkan akurasi, tetap diperlukan penyempurnaan lebih lanjut untuk mencapai hasil yang benar-benar optimal.

4.4.4 Penggunaan *Fine Tuning* dan *Negative prompt*

Hasil generasi menggunakan *Stable Diffusion* 2.1 dengan model yang telah di *Fine Tuning* serta tambahan *negative prompt*. *Fine Tuning* diterapkan untuk meningkatkan kesesuaian gambar dengan prompt yang diberikan, memastikan bahwa elemen utama seperti roket di tengah lebih dominan dan memiliki detail yang lebih baik. Sementara itu, *negative prompt* digunakan untuk menghindari munculnya objek yang tidak relevan atau mengganggu komposisi utama. Dengan kombinasi ini, diharapkan hasil yang dihasilkan lebih akurat, bersih, dan sesuai dengan konsep poster animasi bertema roket.

Gambar 4. 6 Output *Fine Tuning* menggunakan Neg Prompt

Gambar 4.6 menunjukkan peningkatan dibandingkan eksperimen sebelumnya. Komposisi lebih rapi dengan roket utama yang jelas di tengah, serta elemen pendukung yang lebih tertata. Dari evaluasi gambar ini memperoleh nilai 0,77, meningkat dari hasil sebelumnya (0,35, 0,38, dan

0,61). Hal ini menunjukkan bahwa kombinasi *Fine Tuning* dan *negative prompt* secara efektif meningkatkan akurasi serta estetika visual, menghasilkan gambar yang lebih sesuai dengan prompt. Meskipun masih ada beberapa kekurangan kecil, seperti teks yang belum sepenuhnya terbaca dengan jelas, hasil ini menunjukkan kualitas yang jauh lebih baik dibandingkan model yang belum dioptimalkan.

4.5 Pengujian Model

4.5.1 CLIP *Score*

CLIP *Score* adalah metrik yang digunakan untuk mengukur tingkat kesesuaian antara teks (prompt) dan gambar yang dihasilkan oleh model generatif. Semakin tinggi nilai CLIP *Score*, semakin baik gambar yang dihasilkan mencerminkan deskripsi yang diberikan dalam prompt. Dalam eksperimen ini, beberapa model digunakan dengan pendekatan yang berbeda, termasuk penggunaan *Fine Tuning* serta *negative prompt*, untuk meningkatkan akurasi dan kualitas gambar. Dari hasil yang diperoleh, terdapat perbedaan signifikan dalam struktur dan komposisi gambar seiring dengan meningkatnya CLIP *Score*.

Gambar 4. 7 Output dan nilai CLIP

Gambar pertama memiliki nilai sebesar 0.35 dan menunjukkan beberapa pesawat luar angkasa yang tersebar di luar angkasa tanpa adanya fokus utama pada satu objek. Desainnya terlihat lebih abstrak, dengan berbagai elemen seperti stasiun luar angkasa, roket kecil, dan objek lain yang tidak berhubungan secara langsung dengan inti prompt. Kekurangan utama dari gambar ini adalah ketidakmampuannya untuk memberikan fokus pada satu pesawat luar angkasa utama, sehingga kurang sesuai dengan permintaan

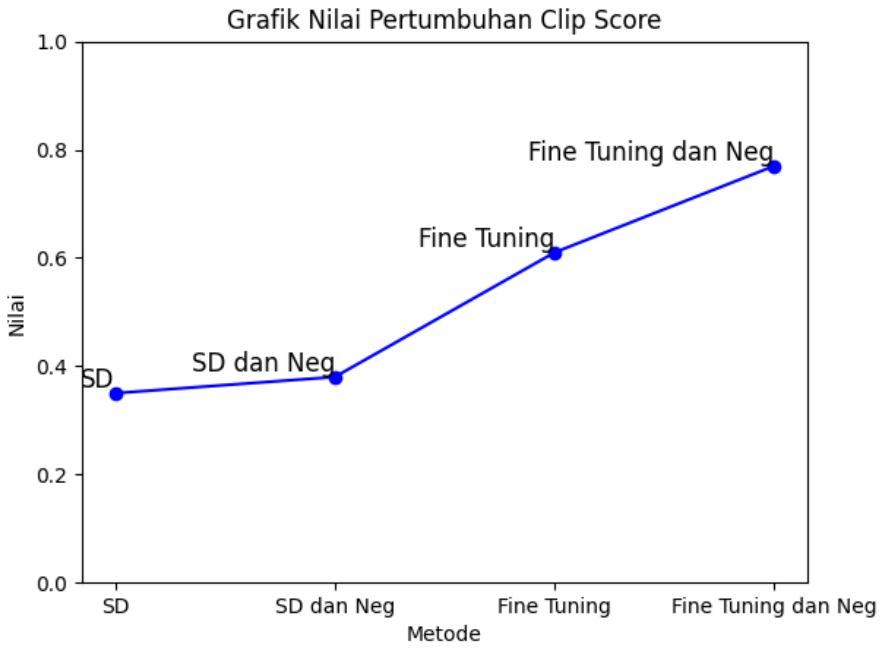
prompt. Oleh karena itu, skor CLIP yang diberikan cukup rendah karena tingkat kesesuaian gambar dengan teks tidak optimal.

Pada gambar kedua memiliki nilai yang sedikit meningkat menjadi 0.38. Meskipun masih memiliki beberapa objek tambahan yang tersebar di luar angkasa, gambar ini mulai menunjukkan struktur yang lebih terorganisir dibandingkan gambar pertama. Perbedaan yang terlihat adalah adanya roket utama yang lebih jelas, tetapi masih ada elemen tambahan yang mengganggu, seperti teks yang kurang dapat dibaca dan berbagai objek yang tidak berhubungan langsung dengan prompt. Hal ini menyebabkan kenaikan skor yang sangat kecil karena meskipun terdapat peningkatan dalam struktur gambar, kesesuaian dengan prompt masih belum optimal.

Peningkatan yang lebih terlihat pada gambar ketiga yang memiliki nilai sebesar 0.61. Model yang telah melalui proses *Fine Tuning* mulai menunjukkan hasil yang lebih baik dengan menampilkan satu pesawat luar angkasa utama yang lebih jelas di bagian tengah gambar. Namun, masih terdapat beberapa elemen yang tidak relevan, seperti keberadaan objek berbentuk wajah berbulu di latar belakang, yang membuat gambar ini tidak sepenuhnya sesuai dengan prompt. Meskipun begitu, peningkatan kesesuaian gambar dengan teks dapat terlihat secara jelas dibandingkan dengan dua gambar sebelumnya, yang menandakan bahwa *Fine Tuning* membantu dalam meningkatkan akurasi generasi gambar.

Gambar keempat, dengan nilai tertinggi yaitu 0.77, menunjukkan hasil terbaik dari seluruh eksperimen ini. Dengan menggunakan kombinasi *Fine Tuning* dan *negative prompt*, model mampu menghasilkan gambar dengan komposisi yang lebih rapi dan sesuai dengan prompt. Satu pesawat luar angkasa utama ditampilkan secara dominan di bagian tengah, dengan desain yang lebih realistis dan struktur yang lebih jelas. Latar belakangnya juga lebih mendukung tema luar angkasa, tanpa adanya objek yang tidak relevan seperti pada gambar ketiga. Selain itu, teks dalam gambar lebih mudah dibaca, yang menambah kesesuaian antara gambar dan deskripsi yang diberikan. Dari hasil ini dapat disimpulkan bahwa *Fine Tuning* yang lebih lanjut, ditambah dengan

penggunaan *negative prompt*, dapat secara signifikan meningkatkan kualitas gambar yang dihasilkan serta kesesuaiannya dengan prompt awal.



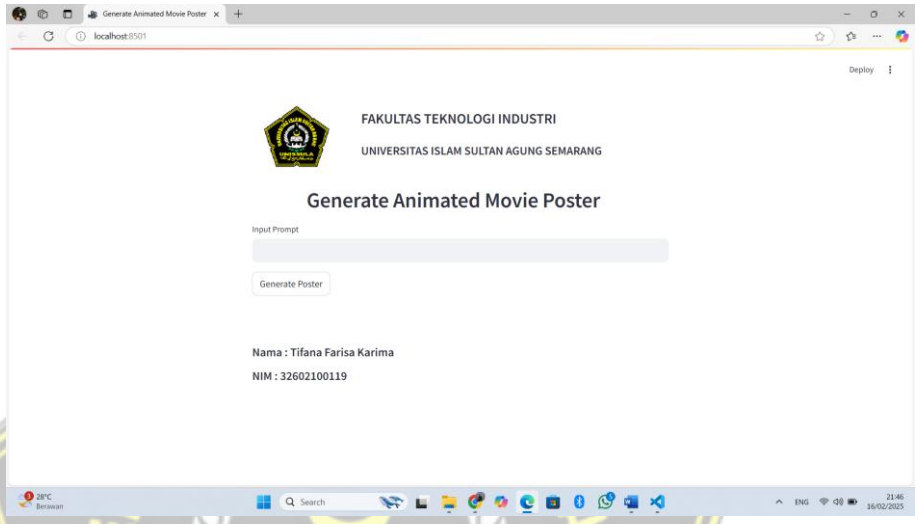
Gambar 4. 8 Grafik Nilai Pertumbuhan Clip

Gambar 4.8 menunjukkan pertumbuhan nilai CLIP *Score* berdasarkan metode yang digunakan dalam proses generasi gambar. Dari grafik, terlihat bahwa metode awal (SD) memiliki nilai CLIP *Score* yang rendah. Setelah menambahkan *negative prompt* (SD dan Neg), terjadi sedikit peningkatan. Namun, peningkatan yang lebih terjadi setelah model mengalami *Fine Tuning*, yang menunjukkan bahwa pelatihan lebih lanjut dapat meningkatkan kesesuaian gambar dengan prompt. Puncak dari peningkatan terjadi ketika *Fine Tuning* dikombinasikan dengan *negative prompt* (*Fine Tuning* dan Neg), menghasilkan CLIP *Score* tertinggi, yang menandakan bahwa metode ini paling efektif dalam meningkatkan kualitas dan akurasi gambar sesuai dengan deskripsi yang diberikan.

4.6 Hasil Menggunakan Streamlit

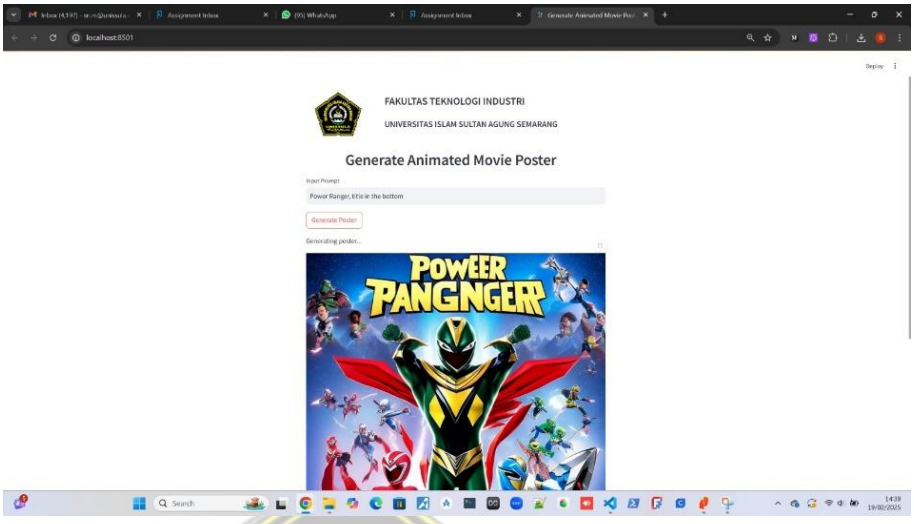
Aplikasi *Generate* Animated Movie Poster dikembangkan menggunakan Streamlit untuk memungkinkan pengguna menghasilkan poster film animasi secara otomatis berdasarkan deskripsi teks. Saat pertama kali dibuka, aplikasi

menampilkan antarmuka awal yang terdiri dari kolom input teks untuk memasukkan deskripsi poster, serta tombol "*Generate* Poster" untuk memulai proses pembuatan gambar. Pada tahap ini, belum ada gambar yang ditampilkan, hanya tampilan sederhana yang menunggu interaksi dari pengguna.



Gambar 4. 9 Tampilan Streamlit Awal

Gambar 4.9 adalah tampilan awal streamlit. Setelah pengguna memasukkan deskripsi dan menekan tombol "*Generate* Poster"**,** aplikasi mulai memproses input menggunakan model AI yang telah di-fine-tune dengan *Stable Diffusion* dan *Dreambooth***.** Selama proses ini berlangsung, Streamlit menampilkan indikator pemrosesan untuk memberi tahu pengguna bahwa sistem sedang bekerja. Proses ini memerlukan waktu beberapa detik sebelum akhirnya menghasilkan gambar poster sesuai dengan deskripsi yang diberikan.



Gambar 4. 10 Tampilan Setelah *Generate* Poster

Pada gambar 4.10 adalah tampilan streamlit setelah *Generate* poster, aplikasi secara otomatis memperbarui antarmuka dengan menampilkan hasil poster yang telah dibuat**.** Pengguna dapat melihat poster film animasi yang dihasilkan langsung di halaman aplikasi. Selain itu, tersedia fitur penyimpanan gambar**,** di mana pengguna dapat mengunduh hasil poster dengan menekan tombol "*Save Poster*"**.**

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil menerapkan *Fine Tuning Dreambooth* dan *Stable Diffusion* 2.1 untuk menghasilkan poster film animasi dari deskripsi teks. Hasilnya menunjukkan bahwa *Fine Tuning* dan *Negative Prompt* meningkatkan kualitas visual dan akurasi gambar.

Model dasar *Stable Diffusion* memiliki CLIP *Score* rendah (0.35 - 0.38), menunjukkan keterbatasan dalam memahami deskripsi teks. Penggunaan *negative prompt* meningkatkan akurasi visual dengan CLIP *Score* 0.61, tetapi masih menghasilkan elemen yang kurang relevan. *Fine Tuning Dreambooth* meningkatkan CLIP *Score* lebih lanjut, dan kombinasi *Fine Tuning* dengan *negative prompt* menghasilkan skor (0.77), dengan komposisi gambar yang lebih sesuai dan fokus pada satu objek utama.

Penelitian ini membuktikan bahwa *Fine Tuning Dreambooth* dan *Stable Diffusion* dapat meningkatkan kualitas generasi gambar untuk pembuatan poster film animasi secara otomatis. Hal ini menunjukkan potensi besar dalam industri kreatif untuk mempercepat desain dan memberikan lebih banyak opsi bagi kreator.

5.2 Saran

Penelitian selanjutnya disarankan untuk berfokus pada penyempurnaan elemen tulisan pada poster, khususnya dalam meningkatkan keterbacaan judul. Meskipun penggunaan *Negative prompt* telah diterapkan, hasil generasi masih menunjukkan bahwa teks pada poster terkadang sulit dibaca. Oleh karena itu, diperlukan optimasi lebih lanjut, baik melalui teknik *Fine Tuning* yang lebih spesifik, penggunaan model generatif terbaru, maupun pendekatan tambahan seperti *post processing* teks, agar judul yang dihasilkan lebih jelas, estetis, dan sesuai dengan desain poster film animasi yang diinginkan.

DAFTAR PUSTAKA

Ahsan, M.M. *dkk.* (2024) "A Comprehensive Survey on Diffusion Models and Their Applications." Tersedia pada: http://arxiv.org/abs/2408.10207.

Badriyah dan Lukmandono (2023) "Prioritas Pengembangan Industri Kreatif Melalui Pendekatan Location Quotient dan Location Modeling," *Prosiding SENASTITAN: Seminar Nasional Teknologi Industri Berkelanjutan*, 3(0), hal. 128–137. Tersedia pada: https://ejournal.itats.ac.id/senastitan/article/view/4248.

Ban, Y. *dkk.* (2024) "Understanding the Impact of Negative Prompts: When and How Do They Take Effect?" Tersedia pada: http://arxiv.org/abs/2406.02965.

Cao, H. *dkk.* (2024) "A Survey on Generative Diffusion Models," *IEEE Transactions on Knowledge and Data Engineering*, 36(7), hal. 2814–2830. Tersedia pada: https://doi.org/10.1109/TKDE.2024.3361474.

Chen, M. *dkk.* (2024) "An Overview of Diffusion Models: Applications, Guided Generation, Statistical Rates and Optimization," hal. 1–39. Tersedia pada: http://arxiv.org/abs/2404.07771.

Effendi, F.P. (2023) "Analisis Semiotika Pada Poster Animasi Disney 'Luca,'" *Professional: Jurnal Komunikasi dan Administrasi Publik*, 10(1), hal. 335–346. Tersedia pada: https://doi.org/10.37676/professional.v10i1.3939.

Gao, Z. *dkk.* (2024) "Dependability Evaluation of Stable Diffusion with Soft Errors on the Model Parameters," *Proceedings of the IEEE Conference on Nanotechnology*, hal. 442–447. Tersedia pada: https://doi.org/10.1109/NANO61778.2024.10628863.

Hanifa *dkk.* (2023) "Peran AI terhadap kinerja industri kreatif Indonesia," *Nucl. Phys.*, 13(1), hal. 104–116.

Ho, J. *dkk.* (2020) "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, 2020-Decem(NeurIPS 2020), hal. 1–25.

Kang, M. *dkk.* (2023) "Variational Distribution Learning for Unsupervised Text-to-Image Generation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2023-June, hal. 23380–23389. Tersedia pada: https://doi.org/10.1109/CVPR52729.2023.02239.

Kingma, D.P. dan Welling, M. (2019) "An introduction to variational

autoencoders," *Foundations and Trends in Machine Learning*, 12(4), hal. 307–392. Tersedia pada: https://doi.org/10.1561/2200000056.

Krojer, B. *dkk.* (2023) "Are Diffusion Models Vision-And-Language Reasoners?," *Advances in Neural Information Processing Systems*, 36(NeurIPS), hal. 1–21.

Li, J.S. *dkk.* (2023) "Augmenters at SemEval-2023 Task 1: Enhancing CLIP in Handling Compositionality and Ambiguity for Zero-Shot Visual WSD through Prompt Augmentation and Text-To-Image Diffusion," *17th International Workshop on Semantic Evaluation, SemEval 2023 - Proceedings of the Workshop*, hal. 44–49. Tersedia pada: https://doi.org/10.18653/v1/2023.semeval-1.5.

Lin, J. *dkk.* (2023) "AutoPoster: A Highly Automatic and Content-aware Design System for Advertising Poster Generation," *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, hal. 1250–1260. Tersedia pada: https://doi.org/10.1145/3581783.3611930.

Munawarah, P.A. dan Tomi, M. (2023) "Analisis Semiotika Poster Film Dilan 1990," *Jurnal Cahaya Mandalika ISSN*, 4(3), hal. 356–367.

Qowy, A.B. dan Dkk (2024) "The Comparison of the Effectiveness and Efficiency of Fine-Tuning Models on Stable Diffusion in Creating Concept Art," *Jurnal Teknik Informatika*, 17(1), hal. 21–29. Tersedia pada: https://doi.org/10.15408/jti.v17i1.37942.

Raj, A. *dkk.* (2023) "DreamBooth3D: Subject-Driven Text-to-3D Generation," *Proceedings of the IEEE International Conference on Computer Vision*, hal. 2349–2359. Tersedia pada: https://doi.org/10.1109/ICCV51070.2023.00223.

Ronneberger *dkk.* (2021) "U-Net: Convolutional Networks for Biomedical Image Segmentation," hal. 1–8.

Ruiz, N. *dkk.* (2023) "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation," hal. 22500–22510. Tersedia pada: https://doi.org/10.1109/cvpr52729.2023.02155.

Septemedi, K.A. dan Santosa, Y.P. (2024) "Usability Analysis of Stable Diffusion-Based Generative Model for Enriching Batik Bakaran Pattern Synthesis," *Proxies : Jurnal Informatika*, 7(2), hal. 128–146. Tersedia pada: https://doi.org/10.24167/proxies.v7i2.12472.

Tao, M. *dkk.* (2022) "DF-GAN: A Simple and Effective Baseline for Text-to-Image Synthesis," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022-June, hal. 16494–16504. Tersedia pada: https://doi.org/10.1109/CVPR52688.2022.01602.

Williams, C. *dkk.* (2023) "A Unified Framework for U-Net Design and Analysis," *Advances in Neural Information Processing Systems*, 36(NeurIPS), hal. 1–38.

Wu, Q. *dkk.* (2023) "Realistic Noise Synthesis with Diffusion Models," 0. Tersedia pada: http://arxiv.org/abs/2305.14022.

Yu, H. *dkk.* (2024) "Uncovering the Text Embedding in Text-to-Image Diffusion Models." Tersedia pada: http://arxiv.org/abs/2404.01154.

Zhang, T. *dkk.* (2023) "A Survey of Diffusion Based Image Generation Models: Issues and Their Solutions," (1). Tersedia pada: http://arxiv.org/abs/2308.13142.