

**PENGEMBANGAN MODEL AI UNTUK MEMBANTU
ILUSTRASI KARAKTER DALAM PANEL KOMIK**

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 : NAVIRA RESTU ANANDA TYANA

NIM : 32602100102

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
2025**

***DEVELOPMENT OF AN AI MODEL TO ASSIST CHARACTER
ILLUSTRATION IN COMIC PANELS***

FINAL PROJECT

*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 : NAVIRA RESTU ANANDA TYANA

NIM : 32602100102

**MAJORING OF INFORMATICS ENGINEERING
INDUSTRIAL TECHNOLOGY FACULTY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG
2025**

LEMBAR PENGESAHAN TUGAS AKHIR
PENGEMBANGAN MODEL AI UNTUK MEMBANTU ILUSTRASI
KARAKTER DALAM PANEL KOMIK

NAVIRA RESTU ANANDA TYANA
32602100102

Telah dipertahankan di depan tim penguji proposal tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 26 Februari 2025

TIM PENGUJI SIDANG TUGAS AKHIR:

Dedy Kurniadi, M.Kom

NIDN. 0622058802

(Penguji 1)

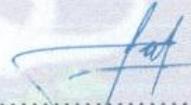


26-02-2025

Sam Farisa, M.Kom

NIDN. 0628028602

(Penguji 2)

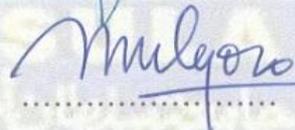


26-02-2025

Ir. Sri Mulyono, M.Eng

NIDN. 0626066601

(Pembimbing)



26-02-2025

Semarang, 26 Februari 2025

Mengetahui,

Kaprodi Teknik Informatika

Universitas Islam Sultan Agung



Moch Taufiq MIT

NIDN. 062203750

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Navira Restu Ananda Tyana

NIM : 32602100102

Judul Tugas Akhir : Pengembangan Model AI Untuk Membantu Ilustrasi
Karakter Dalam Panel Komik

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, 26 Februari 2025

Yang Menyatakan



Navira Restu Ananda Tyana

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Navira Restu Ananda Tyana

NIM : 32602100102

Program Studi : Teknik Informatika

Fakultas : Teknologi Industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : **PENGEMBANGAN MODEL AI UNTUK MEMBANTU ILUSTRASI KARAKTER DALAM PANEL KOMIK** . Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan Agung.

Semarang, 26 Februari 2025

Yang Menyatakan



Navira Restu Ananda Tyana

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 kemudahan dalam menyelesaikan Tugas Akhir yang berjudul Pengembangan Model AI Untuk Membantu Ilustrasi Karakter Dalam Panel Komik. 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.
8. Serta teman terdekat saya atas doa dan dukungan 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,
Penulis



DAFTAR ISI

LEMBAR PENGESAHAN TUGAS AKHIR	Error! Bookmark not defined.
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR....	Error! Bookmark not defined.
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	Error! Bookmark not defined.
KATA PENGANTAR.....	iv
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN.....	xiii
ABSTRAK	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Tugas Akhir.....	5
1.5 Manfaat.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	7
2.1 Tinjauan Pustaka	7
2.2 Dasar Teori	10
2.2.1 Komik.....	10
2.2.2 Komik strip	11
2.2.3 Gaya kartun	12
2.2.4 Text-to-Image.....	13
2.2.5 Base UNET	13
2.2.6 <i>Diffusion Models</i>	16
2.2.7 <i>Stable Diffusion</i>	17
2.2.8 VAE	19

2.2.9	Fine Tune LoRA	20
2.2.10	Data Augmentasi.....	22
2.2.11	Prompt.....	23
2.2.12	Transformer.....	24
2.2.13	CMMD	24
BAB III METODOLOGI PENELITIAN		26
3.1	Studi Literatur.....	26
3.2	Pengumpulan dan Persiapan Dataset.....	27
3.3	Pengolahan Dataset	27
3.4	Pengembangan model.....	28
3.4.1	<i>Stable diffusion 1.5</i>	28
3.4.2	<i>Regularization Fine Tune</i>	28
3.5	Pengujian dan Evaluasi	29
3.5.1	Pengujian.....	29
3.5.2	Evaluasi.....	29
3.6	Alur Kerja Training Sistem	29
3.7	Alur Kerja Sistem User	32
3.8	Perancangan <i>Interface</i>	35
BAB IV HASIL DAN ANALISIS PENELITIAN		37
4.1	Persiapan <i>Dataset</i>	37
4.2	Pengolahan Dataset	37
4.3	Pengembangan Model	43
4.4	Hasil Generate Gambar	49
4.4.1	Penggunaan <i>Stable Diffusion</i>	49
4.4.2	Penggunaan <i>Stable Diffusion</i> dan <i>Negative Prompt</i>	50
4.4.3	Penggunaan Fine Tune.....	51
4.4.4	Penggunaan Fine Tune dan <i>Negative Prompt</i>	51
4.5	Tahap Pengujian	52
4.5.1	Pengujian.....	52
4.5.2	Evaluasi	55
4.6	Perancangan Sistem.....	57

BAB V KESIMPULAN DAN SARAN	60
5.1 Kesimpulan.....	60
5.2 Saran	60
DAFTAR PUSTAKA	
LAMPIRAN	



DAFTAR GAMBAR

Gambar 2. 1 Komik Strip (sumber : <a href="https://www.canva.com/id_id/contoh/EAEITBuNuo0-hitam-dan-putih-
pendidikan-6-panel-komik-strip/">https://www.canva.com/id_id/contoh/EAEITBuNuo0-hitam-dan-putih- pendidikan-6-panel-komik-strip/)	12
Gambar 2. 2 <i>Text to image</i> (sumber : <a href="https://abdulkaderhelwan.medium.com/text-
to-image-generation-model-with-cnn-ca904427d1e7">https://abdulkaderhelwan.medium.com/text- to-image-generation-model-with-cnn-ca904427d1e7)	13
Gambar 2. 3 Arsitektur U-net dasar	15
Gambar 2. 4 <i>Gaussian Noise</i> (J. Rafid Siddiqui n.d.).....	17
Gambar 2. 5 Stable Diffusion Diffusion (Hedlin dkk. 2023).....	18
Gambar 2. 6 <i>Architecture Stable Diffusion</i> (Hedlin dkk. 2023)	18
Gambar 2. 7 Arsitektur VAE	20
Gambar 2. 8 LoRA.....	21
Gambar 2. 9 Data Augmentation (sumber : <a href="https://ubiai.tools/what-are-the-
advantages-anddisadvantages-of-data-augmentation-2023-update/">https://ubiai.tools/what-are-the- advantages-anddisadvantages-of-data-augmentation-2023-update/).....	23
Gambar 3. 1 <i>Wireframe Web</i>	35
Gambar 4. 1 Kumpulan Dataset.....	37
Gambar 4. 2 Kumpulan hasil dtaset yang sudah di <i>captioning</i>	40
Gambar 4. 3 <i>test image</i>	42
Gambar 4. 4 <i>Training Image</i>	43
Gambar 4. 5 <i>train loss</i>	47
Gambar 4. 6 Hasil Validasi step 5.....	48
Gambar 4. 7 Hasil Validasi step 43.....	48
Gambar 4. 8 <i>Ouput Stable Diffusion</i>	49
Gambar 4. 9 Hasil <i>Stable Diffusion</i> dan <i>Negative Prompt</i>	50
Gambar 4. 10 hasil <i>fine tune</i>	51
Gambar 4. 11 Hasil <i>fine-tune</i> dan <i>negative prompt</i>	52
Gambar 4. 12 Kumpulan gambar hasil model.	55
Gambar 4. 13 Hasil Skor CMMD	56
Gambar 4. 14 Bagian atas <i>Interface</i>	57
Gambar 4. 15 Bagian <i>input prompt</i> dan hasil <i>generate</i>	57
Gambar 4. 16 Bagian <i>Input Prompt</i>	58
Gambar 4. 17 Hasil gambar di sistem user	59
Gambar 4. 18 hasil gambar di sistem user (2).....	59

DAFTAR TABEL

Tabel 4. 1 Pembagian Dataset.....	40
Tabel 4. 2 Parameter	44
Tabel 4. 3 Gambar Pengujian.....	53



DAFTAR LAMPIRAN

1. Alur Kerja Training Sistem
2. Alur Kerja Sistem pada User
3. Kode sumber untuk melakukan *captioning*



ABSTRAK

Perkembangan teknologi digital telah memberikan dampak signifikan terhadap industri kreatif, termasuk dalam dunia komik. Dengan hadirnya kecerdasan buatan atau AI, khususnya model difusi seperti *Stable Diffusion*, proses pembuatan komik dapat diotomatisasi dengan menghasilkan gambar dari deskripsi teks. Penelitian ini bertujuan untuk mengembangkan sistem berbasis AI yang mampu mengubah teks menjadi gambar karakter 2D dengan gaya visual kartun dan gothic. Model utama yang digunakan adalah *Stable Diffusion 1.5* dengan metode fine-tuning menggunakan LORA dan data augmentasi untuk meningkatkan kualitas serta konsistensi gambar yang dihasilkan. Sistem yang dikembangkan berbentuk aplikasi berbasis web, di mana pengguna dapat memasukkan deskripsi teks untuk menghasilkan gambar karakter. Penelitian ini berfokus pada visualisasi karakter tanpa elemen latar belakang atau animasi, serta mendukung generasi beberapa karakter dengan gaya yang seragam. Hasil penelitian menunjukkan bahwa penggunaan model difusi memungkinkan pembuatan gambar yang sesuai dengan deskripsi teks, dengan tingkat pemahaman visual yang dapat diterima oleh pembaca. Implementasi teknologi AI dalam pembuatan komik ini memberikan manfaat berupa efisiensi dalam produksi dan fleksibilitas dalam berkarya bagi para komikus. Dengan demikian, penelitian ini berkontribusi dalam mengoptimalkan penggunaan kecerdasan buatan untuk mendukung industri kreatif, khususnya dalam pembuatan komik digital.

Kata kunci: AI, komik digital, *Stable Diffusion*, model difusi, teks ke gambar, *fine-tune*, LORA.

ABSTRACT

The development of digital technology has had a significant impact on the creative industry, including the world of comics. With the presence of artificial intelligence (AI), especially diffusion models such as Stable Diffusion, the comic creation process can be automated by generating images from text descriptions. This research aims to develop an AI-based system capable of converting text into 2D character images with cartoon and gothic visual styles. The main model used is Stable Diffusion 1.5 with fine-tuning methods using LORA and data augmentation to improve the quality and consistency of the resulting images. The developed system is a web-based application, where users can input text descriptions to generate character images. This research focuses on character visualization without background elements or animation, and supports the generation of multiple characters with uniform styles. The results show that the use of a diffusion model enables the generation of images that match the text descriptions, with a level of visual comprehension that is acceptable to readers. The implementation of AI technology in comic creation provides benefits in the form of efficiency in production and flexibility in work for comic artists. Thus, this research contributes to optimizing the use of artificial intelligence to support the creative industry, especially in making digital comics.

Keywords: AI, digital comics, Stable Diffusion, diffusion model, text to image, fine-tune, LORA

BAB I

PENDAHULUAN

1.1 Latar Belakang

Komik adalah salah satu hiburan yang sedang populer saat ini. Komik menjadi salah satu bacaan yang digemari anak-anak dan remaja, tidak sedikit pula orang dewasa yang menyukai bacaan tersebut. Dengan banyaknya penyebaran komik, kedepannya pasti akan menjadi lebih populer lagi. Di dalam komik didominasi oleh gambar-gambar yang seolah-olah berbicara dan gambar tersebut membentuk sebuah narasi dalam cerita dalam komik. Cerita dalam komik bervariasi, seperti komik yang bergenre komedi yang biasanya disukai anak-anak karena kelucuannya, ada pula genre romantis dan drama yang populer dikalangan wanita serta aksi yang lebih populer di kalangan laki-laki, dan lain sebagainya (Pratama, Perdana, dan Shafariya 2022).

Perkembangan teknologi digital telah berkembang pesat dan berdampak luas di berbagai aspek kehidupan masyarakat, termasuk dalam industri kreatif. Salah satu bidang yang sangat diuntungkan dari digitalisasi adalah dunia komik. Teknologi digital tidak hanya mempermudah komunikasi, tetapi juga membuka peluang besar bagi para komikus di Indonesia untuk menjangkau pembaca lebih luas. Para penggemar komik di Indonesia melihat perkembangan komik pada saat ini sebagai kemajuan dari teknologi itu sendiri. Popularitas komik digital atau online terus berkembang dan meningkat sejak awal komik digital muncul dan populer. Data dari [similarweb.com](https://www.similarweb.com), untuk platform webtoon pada periode April hingga Juni 2020, laman tersebut dikunjungi sebanyak 27.08 juta kunjungan di Indonesia, dan mencapai hingga 53.81 juta kunjungan secara global pada bulan Juni 2020. Keberadaan komik digital awalnya dinilai sebagai penggerus komik versi cetak, namun di Indonesia keberadaan komik cetak tetap masih digemari oleh penggemar komik (Putro 2021).

Kecerdasan Buatan atau biasa disebut AI adalah teknologi teknologi baru yang mempelajari dan mengembangkan teori, metode, teknologi, dan sistem aplikasi yang digunakan untuk mensimulasikan, memperluas, dan memperluas kecerdasan manusia. Selain itu, kecerdasan buatan merupakan salah satu cabang dari ilmu komputer yang bertujuan untuk mengungkap esensi kecerdasan dan menghasilkan mesin cerdas baru yang dapat bereaksi dengan cara yang mirip dengan kecerdasan manusia. Dalam konteks ini, kecerdasan buatan memainkan peran penting dalam semua bidang kehidupan. (Xiaojun dkk. 2021)

Dalam eksplorasi pengaturan hubungan “*human versus AI*”, mungkin bidang yang paling sulit dipahami adalah penciptaan dan pemahaman seni. Banyak alternatif menarik yang muncul dikolaborasi antara AI dan seni; namun, pemahaman dan apresiasi seni masih dianggap sebagai kemampuan manusia secara eksklusif. Berakar pada gagasan bahwa keberadaan dan makna seni memang tidak dapat dipisahkan dari interaksi antar manusia, motivasi di balik karya ini adalah untuk mengeksplorasi bagaimana membawa AI ke dalam lingkaran tidak hanya dapat mendorong kemajuan di bidang seni digital dan sejarah seni, tetapi juga menginspirasi perspektif kita tentang masa depan seni. (Cetinic dan She 2022)

Pendekatan terbaru pada generasi gambar adalah pemodelan difusi atau *Model Diffusion* yang dikenal sebagai metode lebih efisien dalam menghasilkan gambar berkualitas tinggi. Perlu pemahaman mendalam tentang matematika untuk memahami teknik ini karena cara kerjanya menambahkan noise secara bertahap pada gambar tanpa mengevaluasi hasil sebelumnya. Setiap pemrosesan gambar dianggap sebagai awal, sehingga sistem tidak bergantung pada gambar sebelumnya, tersisa hanyalah pola *noise* acak yang dapat diolah menjadi gambar realistis. Para peneliti seperti Dhariwal dan Nichol menganggap model difusi sebagai pendekatan yang unggul pada generasi gambar. (Ding dkk. 2021)

Inovasi AI ini berpotensi merevolusi cara komik dibuat, mempermudah para komikus dalam mempercepat produksi tanpa harus mengorbankan

kualitas visualnya. Dengan adanya *diffusion models* salah satu model generatif yang memanfaatkan proses perubahan teks menjadi gambar, pembuatan komik dapat disederhanakan melalui otomatisasi. Proses yang biasanya memerlukan keterampilan ilustrasi manual dan memakan waktu lama kini dapat dipercepat, sehingga komikus memiliki lebih banyak fleksibilitas dalam berkarya.

Adopsi teknologi ini membuka jalan baru bagi komikus untuk menghasilkan karya dengan lebih efisien dan kreatif. Selain itu, teknologi ini juga membantu komikus menghadapi tantangan waktu yang sering kali menjadi hambatan besar. Pembuatan komik yang biasanya memakan waktu lama kini bisa diselesaikan lebih cepat, memungkinkan komikus untuk menjaga kesinambungan penerbitan tanpa mengalami hiatus atau jeda karena kesulitan mengejar tenggat waktu yang telah ditetapkan. Dengan demikian, teknologi AI ini tidak hanya mendukung efisiensi produksi, tetapi juga menjaga kepuasan pembaca yang menantikan komik-komik baru secara teratur.

Namun, panel tetap menjadi elemen penting dalam pengaturan visual komik meskipun AI mempermudah proses ilustrasi. Peletakkan panel memegang peran dalam komik untuk menghubungkan pembaca pada jalan cerita. Maka dari itu dibutuhkan sela di antara panel, atau yang disebut sebagai parit. Parit atau ruang sela inilah yang menimbulkan imajinasi pembaca, dua gambar yang terpisah dalam panel diolah pembaca untuk menjadi sebuah gagasan yang sesuai dengan interpretasi pembaca itu sendiri (Sentosa 2023). Dengan demikian, teknologi AI ini tidak hanya mendukung efisiensi produksi, tetapi juga menjaga kesinambungan penerbitan tanpa mengalami hiatus atau jeda karena kesulitan mengejar tenggat waktu yang telah ditetapkan. Oleh karena itu, meskipun teknologi AI dapat membantu proses produksi, pengaturan dan desain panel tetap menjadi elemen utama dalam memperkuat daya tarik dan penyampaian cerita dalam komik.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana cara agar panel yang dihasilkan sesuai dengan prompt yang dimasukkan ?
2. Bagaimana agar gambar yang dihasilkan konsisten dan mudah dimengerti ?

1.3 Batasan Masalah

Batasan masalah ini bertujuan untuk memudahkan dan menghindari adanya kegiatan diluar sana sasaran, sehingga dalam pembuatan laporan perlu ditentukan batasan masalah. Batasan masalah tersebut sebagai berikut :

1. Penelitian ini hanya menggunakan 1 model dataset yang berupa semua gambar karakter 2D wanita hijab yang sama nantinya akan dikembangkan menjadi sebuah karakter dan beberapa model adegan .
2. *Dataset* yang digunakan adalah kartun wanita hijab tanpa menyertakan elemen animasi atau gerakan.
3. Komik yang digunakan adalah komik berbentuk strip yang terdiri dari 3 panel dengan gaya karakter kartun..
4. Penelitian ini menggunakan metode utama Stable Diffusion versi 1.5 untuk generasi pemodelan karakter. Metode LoRA dan Data Augmentasi digunakan sebagai pendekatan untuk fine tuning.
5. Sistem berbentuk aplikasi berbasis web di mana pengguna hanya dapat memasukkan deskripsi teks (*prompt*) melalui antarmuka yang disediakan. Input lain seperti sketsa, gambar referensi, atau audio tidak termasuk dalam cakupan penelitian.
6. Gambar yang dihasilkan bentuk gambar sederhana yang dipergunakan untuk membantu pemodelan karakter di komik
7. Sistem tidak menyediakan fitur editing gambar secara langsung; pengguna hanya menerima hasil akhir sesuai dengan *prompt* yang dimasukkan.
8. Gambar yang dihasilkan tidak memuat elemen *bubble chat* atau teks percakapan secara otomatis dalam gambar, sehingga proses penambahan teks dialog dilakukan secara manual di luar sistem

1.4 Tujuan Tugas Akhir

Adapun tujuan dari penelitian ini adalah membuat gambar dari 3 *prompt* yang dimasukkan text menjadi 3 panel komik menggunakan *diffusion* model dan menghasilkan gambar sederhana yang dapat dipahami untuk menjadi model karakter di komik. Model AI ini bertujuan untuk membantu ilustrator atau penulis komik dalam menyusun panel komik.

1.5 Manfaat

Manfaat dari penelitian tugas akhir ini adalah :

1. Efisiensi Dalam pembuatan ilustrasi komik, Mengurangi waktu menggambar dan memungkinkan produksi yang lebih cepat.
2. Meningkatkan adopsi AI dalam industri kreatif, Bisa diterapkan di industri komik, animasi, dan digital storytelling.

1.6 Sistematika Penulisan

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

BAB 1 : PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang pemilihan judul tugas akhir “Pembuatan Komik dari Deskripsi Teks Panel menggunakan *Diffusion Models*”. Rumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian, dan sistematika penulisan.

BAB 2 : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini memuat dasar teori yang berfungsi sebagai sumber atau alat dalam memahami permasalahan yang berkaitan dengan implementasi Komik, *Diffusion Model*, *Stable Diffusion*, *Unet*, *LoRA*, *Data Augmentasi*, *Text to Image*.

BAB 3 : METODE PENELITIAN

Bab ini mengungkapkan proses tahapan- tahapan penelitian dimulai dari analisa kebutuhan sistem, kemudian perancangan sistem hingga *prototype* jadi dibuat.

BAB 4 : HASIL PENELITIAN DAN IMPLEMENTASI SISTEM

Pada bab ini penulis mengungkapkan hasil penelitian yakni hasil dari pembuatan *prototype* teks panel pada komik

BAB 5 : KESIMPULAN DAN SARAN

Bab terakhir memuat kesimpulan isi dari keseluruhan uraian bab-bab sebelumnya dan saran-saran dari hasil yang diperoleh dan diharapkan dapat bermanfaat dalam pengembangan selanjutnya



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Dalam beberapa tahun terakhir, aplikasi visi komputer dan teknik pemrosesan gambar telah sangat diuntungkan oleh kemajuan yang dimungkinkan oleh terobosan pembelajaran mendalam. Salah satunya adalah bidang sintesis gambar yang merupakan proses menghasilkan gambar baru dan memanipulasi gambar yang sudah ada. Sintesis gambar adalah tugas yang menarik dan penting karena banyak aplikasi praktis seperti pembuatan seni, pengeditan gambar, realitas virtual, permainan video, dan desain berbantuan komputer (Frolov dkk. 2021).

Otomasi dapat secara progresif terlibat dalam visualisasi visualisasi dan proses narasi untuk membantu pengguna membuat narasi visualisasi naratif dengan lebih mudah, efektif, dan efisien. Kami telah merangkum enam genre visualisasi naratif (yaitu bagan beranotasi, infografis, garis waktu & alur cerita, komik data, *scrollytelling* &, dan *slideshow*, dan video data) berdasarkan penelitian sebelumnya, dan empat jenis alat (yaitu, ruang desain, alat penulisan, alat yang didukung ML / AI alat, alat penghasil ML/AI) berdasarkan tingkat kecerdasan dan otomatisasi alat. Studi ini memungkinkan pengguna untuk memahami elemen desain eksplisit dan implisit dari berbagai narasi visualisasi, memfasilitasi pemilihan alat yang sesuai yang tepat untuk bercerita secara visual. Akan tetapi, survei kami tidak mencakup visualisasi ilmiah. Dalam bidang visualisasi ilmiah, narasi visualisasi telah diterapkan dalam skenario seperti iklim atau narasi kondisi medis (Q. Chen dkk. 2024).

Dalam sistem ini berfokus pada Sintesis gambar dari teks atau *synthesis teks to image*. Sintesis teks-ke-gambar mengacu pada metode komputasi yang menerjemahkan deskripsi tekstual yang ditulis manusia, dalam bentuk kata kunci atau kalimat, ke dalam gambar dengan makna semantik yang serupa dengan teks (Agnese dkk. 2020).

Namun pada kasus-kasus tertentu walaupun menghasilkan gambar-gambar yang luar biasa dengan pendekatan arsitektur komputer vision yang ada tapi tidak ada yang memberikah gambar dengan konsep yang baru dari konteks yang diberikan. Dengan adanya hal tersebut dibutuhkan sebuah model yang dapat mengoptimalkan . Sistem ini memanfaatkan diffusion model, sebuah model difusi teks-ke-gambar yang menggabungkan kekuatan transformator *Language Model* (LM) dengan model difusi dengan ketepatan tinggi untuk menghasilkan tingkat fotorealisme yang belum pernah ada sebelumnya dan tingkat pemahaman bahasa yang mendalam dalam sintesis teks-ke-gambar. Berbeda dengan penelitian sebelumnya yang hanya menggunakan data teks-gambar untuk pelatihan model, temuan utama di balik Imagen adalah bahwa penyematan teks dari LM yang besar, yang telah dilatih sebelumnya pada korpus teks saja, sangat efektif untuk sintesis teks-ke-gambar (Saharia dkk. 2022).

Dari hasil penelitian (Yang dkk. 2023) tentang “*Pixel-Aware Stable Diffusion for Realistic Image Super-resolution and Personalized Stylization*” mendapatkan kesimpulan bahwasanya salah satu dari diffusion model yaitu Stable diffusion memiliki hasil yang memuaskan dapat mencapai gambar yang realistis dan memberikan gambar yang kompleks namun masih memiliki kekurangan sulit untuk mempertahankan konsistensi.

Dari hasil penelitian (Jin dkk. 2024) tentang “*Conditional LoRA Parameter Generation*” mendapatkan kesimpulan bahwasanya untuk pembuatan parameter yang dapat dikontrol dengan performa tinggi, khususnya untuk parameter LoRA. Kami menggunakan *autoencoder* dan model difusi laten bersyarat untuk menangkap distribusi parameter berkinerja tinggi dan melakukan pembangkitan bersyarat, mensintesis seperangkat parameter baru yang disesuaikan dengan kondisi tertentu. Kami menunjukkan bahwa metode kami dapat secara efisien mensintesis parameter model yang baru dan berkualitas tinggi. Distribusi parameter yang dihasilkan oleh COND P-DIFF menunjukkan perbedaan dibandingkan dengan distribusi yang diperoleh melalui metode optimasi konvensional, yang mengindikasikan

tingkat kemampuan generalisasi tertentu. Meskipun demikian, penting untuk menyadari bahwa difusi dalam pembuatan parameter masih belum dieksplorasi secara luas meskipun ada kemajuan yang signifikan dalam bidang sintesis gambar dan video. Dalam karya ini, kami menyajikan metodologi awal untuk difusi parameter bersyarat. Namun, beberapa tantangan masih belum terselesaikan, termasuk mengurangi kebutuhan memori untuk arsitektur model yang besar, meningkatkan generalisasi teknik pembangkitan, dan meningkatkan representasi kondisi dataset. Selain itu, mengintegrasikan grafik pengetahuan dengan difusi bersyarat menawarkan arah yang menjanjikan untuk mengendalikan pembangkitan bersyarat.

Kemudian hasil penelitian dari (Ruiz dkk. 2023) tentang “*DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation*” mendapatkan kesimpulan bahwasanya penggabungan *fine-tuning* dan *Regularization Technique* dapat menanggulangi masalah yang terjadi pada stable diffusion bahkan dapat menyempurnakan gambar memberikan fotorealistik yang sulit dibedakan dengan foto nyata serta kemudahan dalam pengaksesan.

Penambahan *fine-tuning dan regularization technique* menjadi solusi dari kekurangan stable diffusion ini dibuktikan dengan penelitian (Park, Kim, dan Shim 2024) tentang “*TextBoost: Towards One-Shot Personalization of Text-to-Image Models via Fine-tuning Text Encoder*” mendapatkan kesimpulan penyempurnaan penyandi teks dengan metode pelatihan yang inovatif, secara efektif mengurangi overfitting dan memberikan kualitas yang unggul, khususnya dalam kontrol teks serta personalisasi gambar yang berkualitas tinggi.

Kemudian hasil penelitian dari (Jayasumana dkk. 2023) tentang “*Rethinking FID: Towards a Better Evaluation Metric for Image Generation*” mendapatkan kesimpulan para peneliti pembuat gambar untuk memikirkan kembali penggunaan FID sebagai metrik evaluasi utama untuk kualitas gambar. Temuan kami bahwa FID berkorelasi buruk dengan penilai manusia, bahwa FID tidak mencerminkan peningkatan bertahap dari model teks-ke-

gambar yang berulang-ulang dan bahwa FID tidak menangkap distorsi yang jelas menambah kritik yang terus bertambah. Kami khawatir bahwa ketergantungan pada FID dapat menyebabkan peringkat yang cacat di antara metode-metode pembangkitan gambar, dan bahwa ide-ide yang baik dapat ditolak sebelum waktunya. Untuk mengatasi masalah ini, kami mengusulkan CMMD sebagai metrik yang lebih kuat, yang cocok untuk evaluasi model teks-ke-gambar modern.

Dari kelima hasil penelitian di atas penulis dapat mengambil kesimpulan bahwa Stable diffusion dapat dikombinasikan dengan Regularization fine tuning dengan ini penulis mencoba membuat sistem untuk menghasilkan gambar dari teks menggunakan model difusi yaitu stable diffusion yang nantinya data diolah dari dataset yang telah tersedia. Stable diffusion ini memiliki peran memprediksi gambar yang sudah diolah dengan proses encoder, decoder dan pengolahan lain yang mungkin diburuhkan untuk menghasilkan gambar yang beresolusi tinggi. Untuk regularization fine tuning memiliki peran untuk mencegah ketidak konsistensian gambar yang dihasilkan oleh stable diffusion serta agar gambar dihasilkan sesuai dengan data yang tersedia dan teks yang dituliskan pada prompt.

2.2 Dasar Teori

2.2.1 Komik

Komik merupakan kumpulan gambar yang disusun dalam suatu urutan yang terangkai dalam bingkai serta mengungkapkan karakter yang dikemas dalam cerita untuk meningkatkan daya imajinasi pembaca. Komik juga diartikan sebagai bentuk kartun yang memuat karakter untuk memerankan suatu cerita dalam urutan tertentu. Pada umumnya, komik dihubungkan dengan suatu gambar yang dirancang untuk memberikan hiburan kepada pembacanya. Komik memiliki beberapa karakteristik, di antaranya (a) komik biasanya terdiri dari berbagai situasi cerita yang bersambung; (b) bersifat menghibur; (c) apabila komik memiliki perwatakan lain, biasanya dikenal agar kekuatan komik dapat

dihayati; (d) komik memusatkan perhatian di lingkungan sekitar rakyat; (e) pembaca dapat dengan segeramengidentifikasi dirinya melalui perasaan serta tindakan dari perwatakan tokoh utamakarena cerita pada komik mengenai diripribadi; (f) cerita dalam komik ringkas dan menarik perhatian; (g) komik biasanya dilengkapi aksi; (h) pembuatannya lebihhidup dengan pemakaian warna utama secara bebas (Hidayat Agusvian, Asep Sopian, dan Nunung Nursyamsiah 2021).

2.2.2 Komik strip

Komik strip adalah salah satu jenis komik. Komik strip adalah serangkaian gambar yang berurutan dan terdiri dari beberapa panil. Pada tahun 1930-1960 komik strip dimuat pada surat kabar, biasanya mengangkat topik yang sedang hangat dibicarakan. Komik strip pada umumnya terbit secara teratur di media koran, majalah, surat kabar dan internet Komik pada dasarnya berekspresi mengungkapkan keluh kesah, mengungkapkan cinta, informasi, pesan, anekdot, lelucon perihal menarik untuk dilihat dan dibaca. Pada dasarnya komik strip bersifat lebih natural dan bebas. Ciri-ciri komik strip yaitu memiliki sedikit panil, panil yang berurutan tidak satu halaman penuh, memiliki tat letak yang kaku, komposisinya sederhana dan memiliki cerita paling singkat. Komik strip sebuah montase, cara dalam mendeskripsikan satu ide melalui beberapa gambar dan teks berkaitan erat (Prayoga 2021).



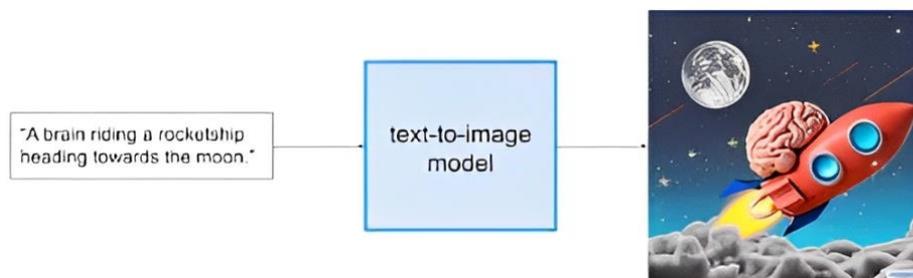
Gambar 2. 1 Komik Strip (sumber : https://www.canva.com/id_id/contoh/EAEITBuNuo0-hitam-dan-putih-pendidikan-6-panel-komik-strip/)

2.2.3 Gaya kartun

Menggambarkan sebuah karakter tentunya dipengaruhi oleh pencipta atau desainer karakter itu sendiri secara umum empat klasifikasi atau aliran gaya gambar utama di dalam menciptakan sebuah gambar karakter baik itu untuk keperluan cerita bergambar, komik maupun karakter animasi. Salah satunya adalah gaya kartun. Gaya kartun pada dasarnya adalah bentuk penegasan melalui penyederhanaan dari usaha pencitraan yang dirancang menyerupai subyek artinya dalam proses kreatif yang dilalui oleh seorang kartunis adalah menggambar suatu objek tanpa harus menekankan detail visual sebagaimana ilustrator gambar realis. Kartunis tidak banyak menghilangkan detail, namun harus lebih memusatkan diri pada detail tertentu. Tahap selanjutnya seorang kartunis perlu penguasaan terhadap beberapa elemen penting dalam visualisasi kartun di antaranya ekspresi, pergerakan, fokus, skala dan teknik pemberian warna serta bayangan (Putri dan Purningsih 2023).

2.2.4 Text-to-Image

Pembuatan teks-ke-gambar mengacu pada proses pembuatan konten visual berdasarkan deskripsi tekstual menggunakan kecerdasan buatan. Teknologi ini mengubah kata dan frasa menjadi gambar, memungkinkan bentuk baru dari ekspresi kreatif dan penceritaan visual yang menggabungkan bahasa dengan seni visual. Teknologi ini telah menjadi terkenal di berbagai bidang, termasuk game dan periklanan, yang memungkinkan pembuatan prototipe dan eksplorasi konsep visual secara cepat. Pembuatan teks-ke-gambar menggunakan algoritme yang rumit, termasuk jaringan saraf dan model pembelajaran mendalam, untuk menginterpretasikan dan memvisualisasikan detail dari teks. Sistem ini dilatih dengan kumpulan data besar pasangan gambar-teks, yang memungkinkannya mempelajari hubungan antara deskripsi tekstual dan representasi visual yang sesuai. Ketika pengguna memasukkan deskripsi, AI memprosesnya untuk menghasilkan gambar yang mencerminkan konsep yang dijelaskan (Fiveable 2024).

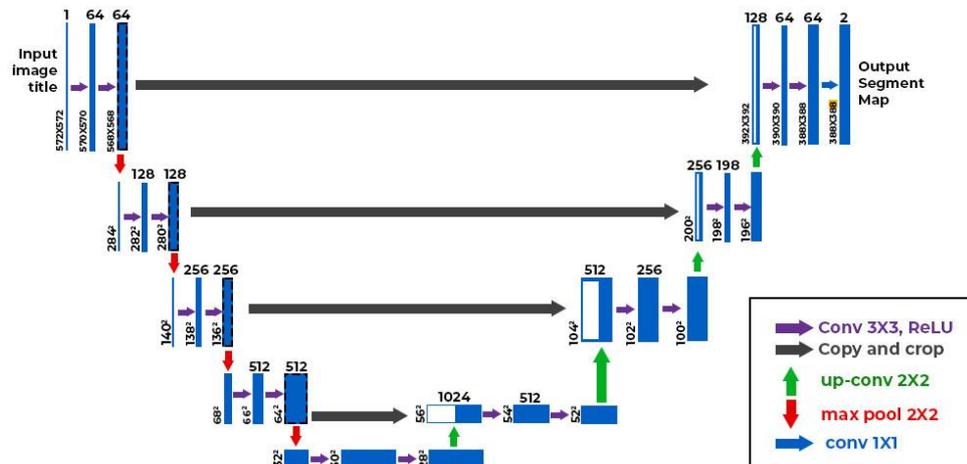


Gambar 2. 2 *Text to image* (sumber : <https://abdulkaderhelwan.medium.com/text-to-image-generation-model-with-cnn-ca904427d1e7>)

2.2.5 Base UNET

U-Net Tradisional. U-Net tradisional adalah arsitektur jaringan dua dimensi. U-Net memodifikasi dan memperluas *Fully Convolutional Network* (FCN), membuatnya bekerja dengan sedikit gambar pelatihan dan menghasilkan segmentasi yang lebih akurat. Ide utamanya adalah mengganti jaringan penyusutan umum dengan lapisan berurutan dan operasi pengumpulan terkait dengan operator downsampling, yang dilengkapi dengan operator upsampling. Oleh karena itu, resolusi keluaran dimunculkan

oleh lapisan-lapisan ini. Resolusi tinggi dari jalur yang dikontrak digabungkan dengan keluaran yang di-upampling untuk lokalisasi. Oleh karena itu, lapisan konvolusional sekuensial dapat mempelajari fitur-fitur yang bagus dan menghasilkan segmentasi yang lebih akurat. Modifikasi penting dalam arsitektur U-Net terletak pada bagian upsampling, di mana terdapat banyak sekali sejumlah saluran fitur yang memungkinkan jaringan menyebarkan data kontekstual ke lapisan resolusi lebih tinggi. Oleh karena itu, jalur ekspansi kira-kira simetris dengan jalur kontraksi, membentuk struktur berbentuk U. Jaringan menerapkan bagian efektif dari setiap konvolusi—peta segmentasi hanya berisi piksel, dan konteks lengkap piksel dapat diperoleh pada gambar masukan. Metode ini memungkinkan segmentasi yang mulus dalam pencitraan berukuran besar menggunakan strategi ubin tumpang tindih yang penting, yang tanpanya resolusi akan dibatasi oleh memori GPU. CNN tradisional biasanya dihubungkan ke beberapa lapisan yang terhubung sepenuhnya setelah konvolusi dan peta fitur yang dihasilkan oleh lapisan konvolusional dipetakan menjadi vektor fitur dengan panjang tetap untuk klasifikasi tingkat gambar. Namun, struktur FCN yang ditingkatkan mengidentifikasi gambar pada tingkat piksel, sehingga memfasilitasi tugas segmentasi dalam pencitraan pada tingkat semantik. U-Net dapat diterapkan pada segmentasi karena ukuran pengukuran gambar medis yang besar. Tidak mungkin memasukkan gambar medis berukuran besar ke dalam jaringan ketika gambar tersebut disegmentasi dan harus dipotong menjadi potongan-potongan kecil. Strategi penggarapan yang tumpang tindih cocok untuk pemotongan potongan kecil menggunakan U-Net karena struktur jaringannya (Yin dkk. 2022).



Gambar 2. 3 Arsitektur U-net dasar

Gambar diatas adalah *base* Unet ,panah mewakili operasi yang berbeda, kotak biru mewakili peta fitur pada setiap lapisan, dan kotak abu-abu mewakili peta fitur yang dipotong dari jalur kontrak. Angka pada kotak biru seperti angka 1-64-64 adalah jumlah kernel yang dilewati gambar saat proses konvolusi dan angka 512x512 atau 284² adalah ukuran dari file gambar. Konvolusi sendiri adalah operasi matematika yang digunakan dalam pemrosesan citra dan jaringan saraf atau CNN untuk mengekstrak fitur dari gambar. Seperti yang telah disebutkan sebelumnya, jaringan U-net dapat dibagi menjadi dua bagian: Bagian pertama adalah jalur kontrak yang menggunakan arsitektur CNN pada umumnya. Setiap blok dalam jalur kontrak terdiri dari dua konvolusi 3x3 yang berurutan diikuti oleh unit aktivasi ReLU dan lapisan max-pooling. Berikut aktivasi menggunakan ReLU :

$$ReLU(x) = \max(0, x)$$

Keterangan :

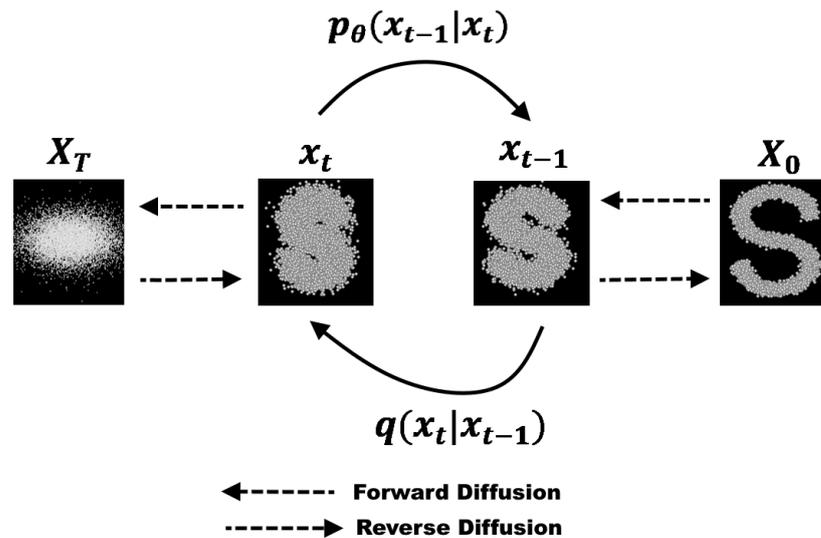
1. ReLU : Rectified Linear Unit (nama fungsi aktivasi)
2. x : Nilai input ke fungsi aktivasi (hasil dari kombinasi linier bobot dan bias dari lapisan sebelumnya)
3. $\max(0, x)$: Fungsi maksimum yang memilih nilai terbesar antara 0 dan x :
 - Jika $x < 0$, maka output = 0
 - Jika $x \geq 0$, maka output = x

- Jika $x > 0$, maka output = 0

Susunan ini diulang beberapa kali. Kebaruan dari u-net ada pada jalur ekspansif di mana pada setiap tahap, peta fitur di-upsampling menggunakan konvolusi naik 2x2. Kemudian, peta fitur dari lapisan yang sesuai di jalur kontraksi dipotong dan digabungkan ke peta fitur yang telah di-upsampling. Hal ini diikuti dengan dua konvolusi 3x3 berturut-turut dan aktivasi ReLU. Pada tahap akhir, konvolusi 1x1 tambahan diterapkan untuk mengurangi peta fitur ke jumlah saluran yang diperlukan dan menghasilkan gambar yang tersegmentasi. Pemotongan ini diperlukan karena fitur piksel di bagian tepi memiliki informasi kontekstual yang paling sedikit dan oleh karena itu perlu dibuang. Hal ini menghasilkan jaringan yang menyerupai bentuk u dan yang lebih penting, menyebarkan informasi kontekstual di sepanjang jaringan yang memungkinkannya untuk menyegmentasikan objek di suatu area dengan menggunakan konteks dari area yang tumpang tindih yang lebih besar (Siddique dkk. 2021).

2.2.6 Diffusion Models

Model probabilistik difusi(yang akan kita sebut “model difusi” untuk singkatnya) adalah rantai Markov berparameter yang dilatih menggunakan inferensi variasi untuk menghasilkan sampel yang cocok dengan data setelah waktu tertentu. Transisi dari rantai ini dipelajari untuk membalikkan proses difusi, yang merupakan rantai Markov yang secara bertahap menambahkan noise ke data ke arah yang berlawanan dengan arah pengambilan sampel hingga sinyal dihancurkan. Ketika difusi terdiri dari sejumlah kecil noise Gaussian, maka cukup memadai untuk mengatur transisi rantai pengambilan sampel menjadi bersyarat Gaussian juga, memungkinkan parameterisasi jaringan saraf yang sangat sederhana (S. Chen dkk. 2023).



Gambar 2. 4 *Gaussian Noise* (J. Rafid Siddiqui n.d.)

Keterangan :

X_0 = Data Asli (Gambar Awal)

X_{t-1} = Data pada langkah sebelumnya

X_t = Data pada langkah saat ini

X_T = Data akhir yang penuh noise (distribusi acak)

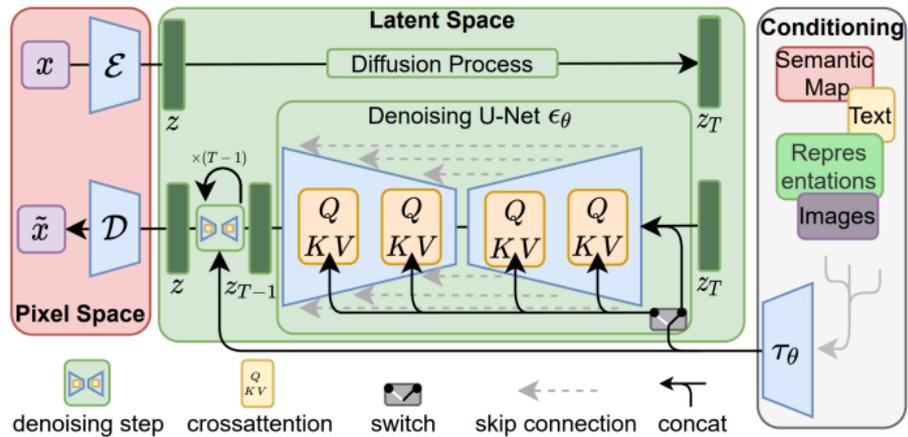
$Q(x_t|x_{t-1})$ = Proses Forward diffusion

$p_{\theta}(x_t|x_{t-1})$ = Proses reverse diffusion (diatur oleh parameter model θ)

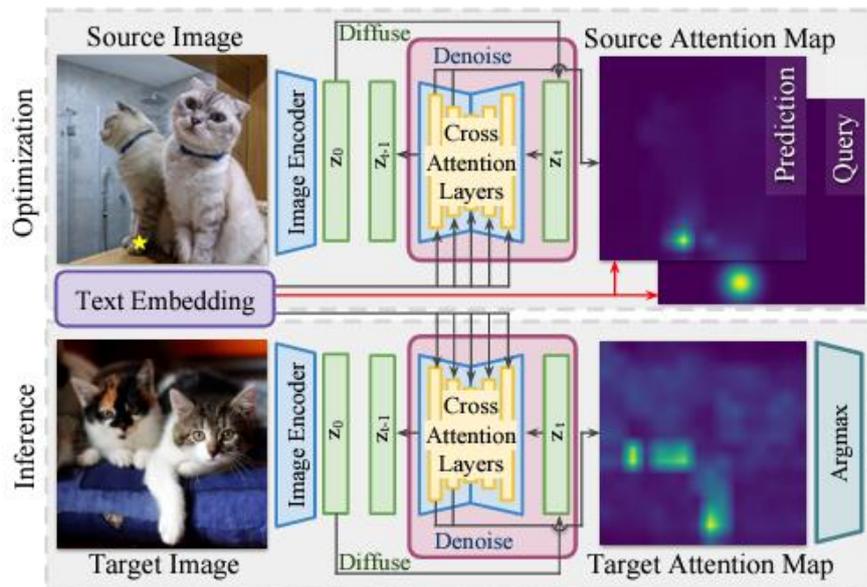
2.2.7 *Stable Diffusion*

Stable Diffusion adalah model teks-ke-gambar, dirilis pada tahun 2022, yang menggunakan teknik pembelajaran mendalam yang disebut difusi laten untuk menghasilkan gambar berdasarkan deskripsi teks. Tidak seperti beberapa model *Text-to-Image* sebelumnya, kode dan bobot model *Stable Diffusion* adalah tersedia untuk umum dan dapat dijalankan pada sebagian besar perangkat keras konsumen. Untuk menghasilkan gambar, *Stable Diffusion* menggunakan CLIP untuk memproyeksikan perintah teks ke dalam ruang penyematan teks-gambar bersama, dan memilih gambar yang kasar dan berisik yang secara semantik dekat dengan prompt input. Gambar ini kemudian dikenakan metode *denoising* berdasarkan laten model difusi laten untuk menghasilkan gambar akhir. Selain prompt teks, skrip pembuatan

Teks-ke-Gambar dalam *Stable Diffusion* memungkinkan pengguna untuk memasukkan berbagai parameter seperti jenis pengambilan sampel, dimensi gambar keluaran, dan nilai seed (Hedlin dkk. 2023).



Gambar 2. 5 Stable Diffusion Diffusion (Hedlin dkk. 2023)



Gambar 2. 6 Architecture Stable Diffusion (Hedlin dkk. 2023)

Difusi stabil: sebuah tonggak kerja pada ruang laten. Kerangka kerja representatif yang melatih model difusi pada ruang laten adalah Stable Diffusion, yang merupakan versi yang ditingkatkan dari *Latent Diffusion*

Model (LDM). Mengikuti Dall-E yang mengadopsi VQ-VAE untuk mempelajari buku kode visual, *Stable diffusion* menggunakan VQ-GAN untuk representasi laten pada tahap pertama. Khususnya, VQ-GAN meningkatkan VQ-VAE dengan menambahkan tujuan yang berlawanan untuk meningkatkan kealamian gambar yang disintesis. Dengan VAE yang telah dilatih sebelumnya, difusi stabil membalikkan proses difusi maju yang mengganggu ruang laten dengan noise. Difusi stabil juga memperkenalkan perhatian silang sebagai pengkondisian tujuan umum untuk berbagai sinyal kondisi seperti teks. Hasil eksperimen dalam menyoroti bahwa melakukan pemodelan difusi pada ruang laten secara signifikan mengungguli pemodelan pada ruang piksel dalam hal pengurangan kompleksitas dan pelestarian detail. Pendekatan serupa juga telah diselidiki dalam VQ-difusi dengan strategi difusi mask-lalu-ganti. Menyerupai temuan dalam metode ruang-piksel, panduan bebas-pengklasifikasi juga secara signifikan meningkatkan model difusi teks-ke-gambar dalam ruang laten (Kandwal dan Nehra 2024).

2.2.8 VAE

Variational Autoencoder (VAE) adalah sebuah *autoencoder* khusus yang didasarkan pada inferensi *variational* Bayes yang awalnya diusulkan oleh Kingma dan Welling, Doersch. Tujuan dari VAE adalah untuk dapat mempelajari distribusi data pelatihan sehingga dengan mengambil sampel darinya, kita dapat menghasilkan data baru, karena data pelatihan belum tentu memiliki distribusi matematika yang terdefinisi dengan baik, kami memaksa distribusi *output encoder* (dikenal sebagai ruang laten) untuk mengikuti distribusi yang diketahui, misalnya distribusi normal. Gambar 2 menunjukkan arsitektur VAE yang memiliki encoder dan jaringan inferensi variasi, diikuti oleh *decoder* yang mengambil sampel dari ruang laten untuk menghasilkan output. Perbedaan utama antara AE dan VAE adalah AE mempelajari representasi terkompresi dari input, dan dekompresinya untuk mencocokkan input yang diberikan. Sebaliknya, VAE adalah model Bayesian yang mempelajari representasi terkompresi dari AE, dan membangun parameter yang mewakili distribusi probabilitas data. Model ini dapat

mengambil sampel dari distribusi ini dan menghasilkan sampel data input baru. Oleh karena itu, VAE adalah model generatif, di mana AE yang hanya melakukan rekonstruksi tidak memiliki interpretasi generatif yang jelas. (Weidner, 2020)

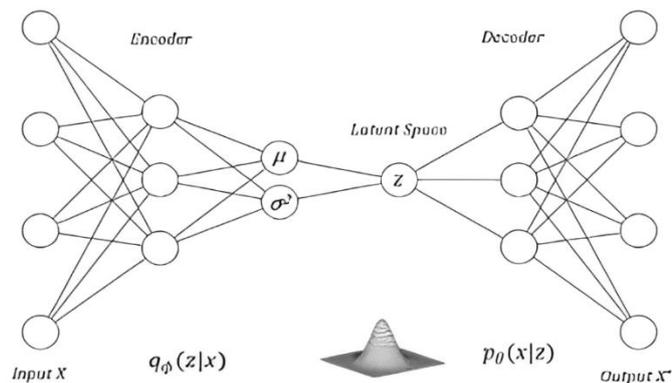


FIGURE 2. Architecture of Variational Autoencoder (VAE).

Gambar 2. 7 Arsitektur VAE

Keterangan :

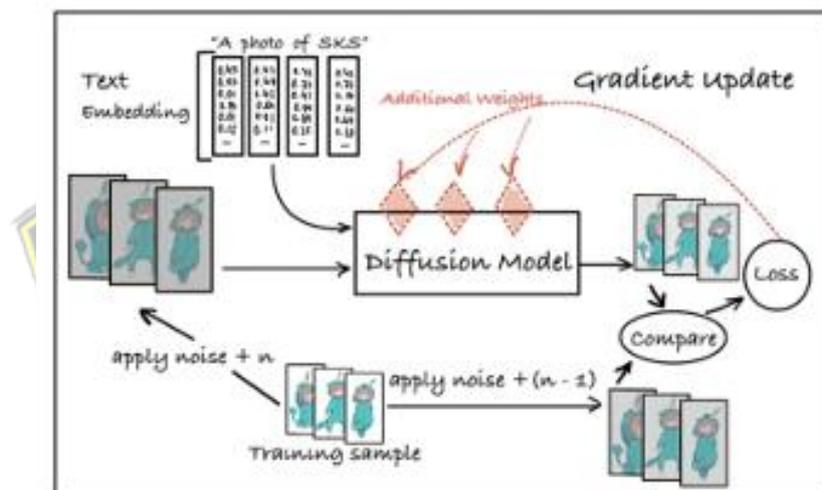
1. $q_{\phi}(z|x)$: Distribusi probabilistik dari vektor laten z berdasarkan input x
2. Mean (μ) : Nilai rata-rata distribusi laten
3. Variance (σ^2) : Variasi atau sebaran distribusi laten
4. z : Vektor laten
5. $p_{\theta}(x|z)$: Distribusi probabilistik dari output yang direkonstruksi

2.2.9 Fine Tune LoRA

Fine Tuning dalam *deep learning* adalah suatu bentuk pembelajaran transfer. Hal ini melibatkan pengambilan model yang telah dilatih sebelumnya, yang telah dilatih pada kumpulan data yang besar untuk tugas umum seperti pengenalan gambar atau pemahaman bahasa alami, dan membuat penyesuaian kecil pada parameter internalnya. Tujuannya adalah untuk mengoptimalkan kinerja model pada tugas baru yang terkait tanpa memulai proses pelatihan dari awal (Amanatullah 2023).

Low-Rank Adaptation (LoRA) adalah teknik yang digunakan untuk menyempurnakan model Difusi Stabil, khususnya *stable-diffusion-v1.5* (.).

Mirip dengan *DreamBooth*, LoRA menawarkan solusi yang lebih efisien dengan membuat dan melatih hanya beberapa lapisan baru alih-alih menyesuaikan seluruh gradien model. Diimplementasikan dalam lapisan lintas-perhatian yang menghubungkan data gambar dengan petunjuk tekstual, LoRA memungkinkan model difusi untuk mengenali kata-kata baru sebagai konsep yang unik. Penyempurnaan ini meningkatkan efektivitas model tanpa mengubah struktur fundamentalnya atau memerlukan pelatihan ulang yang lengkap semua bobot untuk setiap konsep baru (Beltr dkk. 2024).



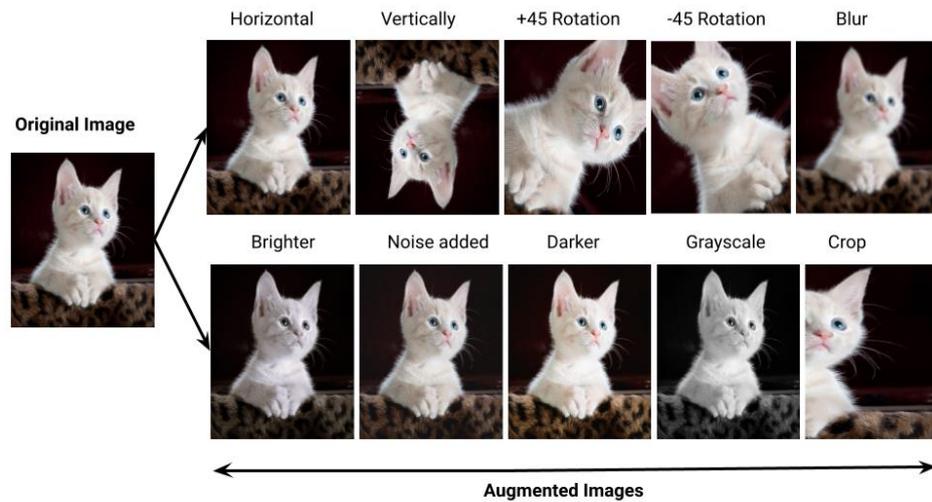
Gambar 2. 8 LoRA

Gambar tersebut adalah proses pada *fine tune* LoRA. Proses dimulai dengan *text embedding*, di mana teks seperti "A photo of SKS" diubah menjadi representasi vektor melalui model teks (misalnya CLIP). Vektor ini berfungsi sebagai panduan untuk menghasilkan gambar yang sesuai dengan deskripsi teks. Diffusion Model atau (*Pre-trained Model*), model difusi yang sudah dilatih sebelumnya digunakan sebagai dasar. LoRA memungkinkan penambahan bobot tambahan (*Additional Weights*) pada model yang sudah dilatih sebelumnya, tanpa perlu menyesuaikan seluruh parameter model. Bobot tambahan ini dioptimalkan untuk meningkatkan keselarasan antara gambar yang dihasilkan dengan deskripsi teks. Proses difusi melibatkan menambahkan noise secara bertahap pada gambar latih dan model difusi

kemudian berusaha menghilangkan noise berdasarkan panduan dari text embedding. Tujuannya adalah agar gambar yang dihasilkan semakin mirip dengan deskripsi teks. Perhitungan *Loss* dan Pembaruan Gradien atau *Gradient Update* setelah gambar dihasilkan, hasilnya dibandingkan dengan gambar latih. Perbedaan hasil dihitung sebagai *loss*. Gradien pembaruan (*gradient update*) hanya diterapkan pada bobot LoRA untuk meminimalkan loss, tanpa memperbarui seluruh parameter model utama. Proses Difusi (*Noise Injection* dan *Denoising*) Setelah beberapa iterasi *fine-tuning*, model akan mampu menghasilkan gambar yang lebih akurat dan konsisten dengan deskripsi teks. Karena LoRA hanya memperbarui bobot tambahan, model hasil *fine-tuning* tetap efisien dalam hal ukuran dan kecepatan inferensi.

2.2.10 Data Augmentasi

Augmentasi data memainkan peran yang sangat diperlukan dalam model pembelajaran mendalam, karena memaksa model untuk mempelajari fitur invarian, dan dengan demikian membantu generalisasi. Augmentasi data diterapkan di banyak bidang tugas penglihatan termasuk pengenalan objek, persepsi video, dan segmentasi semantik. Selain mempelajari fitur invarian, data augmentasi data juga memiliki aplikasi spesifik lainnya dalam pembelajaran mendalam. Sebagai contoh, pelatihan adversarial memanfaatkan augmentasi data untuk membuat sampel adversarial dan dengan demikian meningkatkan ketahanan adversarial model. Dalam hal augmentasi data dalam deteksi objek, mengusulkan pembelajaran penguatan untuk mencari kebijakan augmentasi yang kompleks termasuk transformasi warna transformasi warna, transformasi geometris, dan transformasi kotak pembatas. CutMix mengusulkan strategi pencampuran spasial untuk meningkatkan deteksi hasil. *RandAugment* menggunakan ruang pencarian yang diperkecil untuk mencari kebijakan augmentasi klasifikasi dan deteksi (Li dkk. 2024).



Gambar 2. 9 Data Augmentation (sumber :<https://ubiai.tools/what-are-the-advantages-anddisadvantages-of-data-augmentation-2023-update/>)

2.2.11 Prompt

Prompt Engineering adalah pendekatan untuk mengadaptasi model besar yang telah dilatih sebelumnya, yang juga dikenal sebagai model dasar, untuk tugas-tugas baru dengan menambah input model dengan petunjuk khusus tugas. Secara khusus, input model ditambah dengan bagian tambahan, yang disebut prompt, yang dapat berupa instruksi bahasa alami yang dibuat secara manual, instruksi bahasa alami yang dibuat secara otomatis, atau representasi vektor yang dibuat secara otomatis. Instruksi bahasa alami juga disebut sebagai prompt diskrit atau prompt keras, sedangkan representasi vektor disebut prompt kontinu atau prompt lunak. . Ide utama dari prompt engineering adalah memberikan petunjuk bersama dengan masukan untuk memandu model yang telah dilatih sebelumnya untuk menyelesaikan tugas baru dengan menggunakan pengetahuan yang ada. Jika petunjuknya berupa bahasa alami yang dapat ditafsirkan oleh manusia (hard prompt), penelitian terkait telah disebut sebagai *InContext Learning*, yang memungkinkan model untuk belajar dari instruksi tugas, demonstrasi dengan beberapa contoh, atau informasi pendukung dalam konteks. Selain itu, petunjuk dapat berupa representasi vektor kontinu (petunjuk lunak). Pekerjaan terkait telah disebut sebagai *Prompt-Tuning*, yang mengoptimalkan petunjuk secara langsung di

ruang penyematan model. Dibandingkan dengan paradigma tradisional, teknik prompt memiliki banyak keuntungan. Pertama, dibutuhkan sedikit data berlabel untuk mengadaptasi model yang telah dilatih sebelumnya ke tugas-tugas baru, yang sangat mengurangi upaya pengawasan manusia dan sumber daya komputasi untuk finetuning. Kedua, *prompt engineering* memungkinkan model yang telah dilatih sebelumnya untuk melakukan prediksi pada tugas-tugas baru hanya berdasarkan prompt tanpa memperbarui parameter model apa pun, sehingga memungkinkan untuk melayani tugas-tugas hilir dalam skala besar dengan menggunakan model yang sama (Gudk. 2023).

2.2.12 Transformer

Transformer adalah model pembelajaran mendalam yang didasarkan pada mekanisme perhatian, yang diterapkan secara luas dalam tugas-tugas seperti pemrosesan bahasa alami dan analisis deret waktu. Keunggulan utamanya adalah kemampuannya untuk menangkap ketergantungan jarak jauh dan kemampuan komputasi paralelnya yang kuat dalam memproses data urutan. Oleh karena itu, Transformer diperkenalkan ke dalam jaringan denoising untuk memproses data deret waktu, menangkap ketergantungan yang rumit dan meningkatkan kemampuan model selama proses *denoising* terbalik dari model difusi. Transformer digunakan untuk mengkodekan data input dan melakukan pemrosesan bersyarat dalam proses denoising model difusi (Modeling dan Knowledge 2025).

2.2.13 CMMD

CLIP-MMD digunakan untuk menghitung kesesuaian gambar yang dihasilkan terhadap deskripsi teks. Metriks evaluasi ini memanfaatkan embedding dari CLIP, model yang sudah dilatih untuk memetakan teks dan gambar sehingga dapat memahami hubungan kesesuaian antara teks dan gambar. Penggabungan embedding dari CLIP menggunakan konsep *Maximum Mean Discrepancy*. Cara kerja pada sistem yaitu, folder hasil generate model baru dihitung CLIP MMD menggunakan deskripsi teks atau prompt yang digunakan saat menghasilkan gambar. Kemudian dibandingkan

dengan gambar yang ada pada dataset pengujian. CLIP-MMD dengan nilai yang rendah menunjukkan bahwa gambar yang dihasilkan mendekati distribusi gambar referensi (dataset pengujian) yang menunjukkan bahwa kualitas gambar baik, teks deskripsi sesuai (Jayasumana dkk. 2023).

Untuk dua distribusi probabilitas didefinisikan oleh :

$$dist_{MMD}^2(P, Q) = \mathbb{E}_{x,x'}[k(x, x')] + \mathbb{E}_{y,y'}[k(y, y')] - 2\mathbb{E}_{x,y}[k(x, y)]$$

Keterangan :

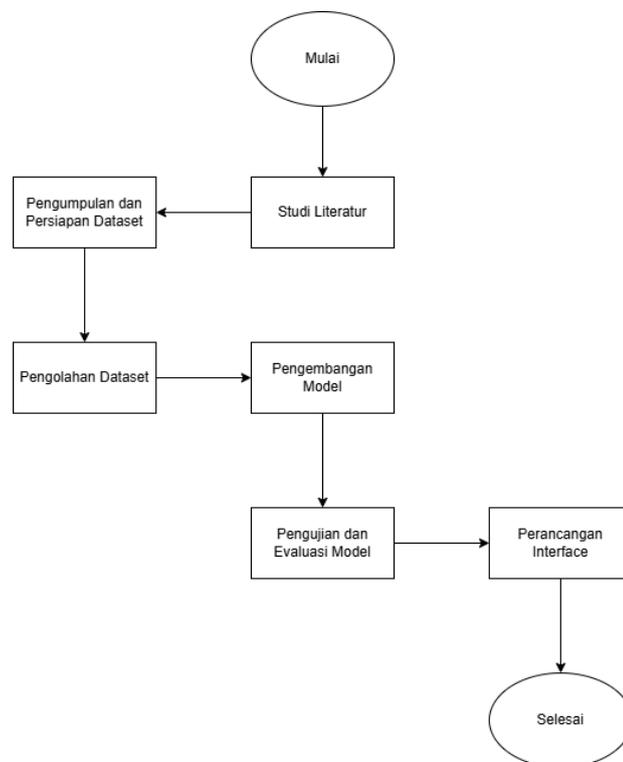
1. x dan x' adalah sampel dari distribusi pertama P (embedding gambar referensi)
2. y dan y' adalah sampel dari distribusi kedua Q (embedding gambar hasil model)
3. $k(x,y)$ adalah kernel positif definit (biasanya kernel RBF Gaussian)



BAB III

METODOLOGI PENELITIAN

Penelitian ini menggunakan metode atau algoritma yang disebut *diffusion models* dan LoRA. Dalam penelitian ini ada kedua algoritma tersebut digabungkan untuk *generate image* model ai untuk komik sesuai dengan prompt yang dimasukan . Langkah langkah yang harus dilakukan dalam penelitian ini adalah sebagai berikut :



Gambar 7. 1 Metode Penelitian

3.1 Studi Literatur

Dalam penelitian ini akan dilakukan tinjauan terhadap beberapa e-book, makalah, jurnal, tesis dan skripsi terdahulu akan diulas selain mengunjungi berbagai situs web. Tujuan dari tinjauan ini adalah untuk mempelajari teori dibalik konsep sistem *diffusion model*, *stable diffusion*, LoRA.

3.2 Pengumpulan dan Persiapan Dataset

Pada tahap ini sebagai persiapan awal dan pengembangan model adalah mengumpulkan data. Data yang digunakan adalah menggunakan dataset pada website penyedia layanan dataset terbuka yang diambil adalah data data gambar tidak bentuk, GIF, video, ataupun model 3D namun 2D diantaranya dari penelusuran Google, Civit AI, Kaggle, Lexica dan pixabay. Website tersebut bebas penggunaan dan terhindar dari hak cipta ketika digunakan.

3.3 Pengolahan Dataset

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

a. *Resize*

Resize adalah mengubah ukuran gambar yang ada secara keseluruhan, baik itu memperbesar atau memperkecil tanpa memotong. Tujuan dari adanya *Resize* adalah untuk menyamakan resolusi dataset.

b. *Cropping*

Cropping adalah proses memotong bagian tertentu pada gambar. Tujuan dari proses ini adalah untuk mengkonsentrasikan perhatian pada bagian tertentu dari gambar dan menghilangkan bagian yang tidak penting.

c. *Color jitter*

Color jitter adalah proses mengubah kecerahan, kontras, dan saturasi gambar secara acak. Bertujuan untuk membantu model menangani gambar dari berbagai kondisi pencahayaan.

d. *Rotation dan flipping* gambar

Rotation dan flipping adalah melakukan pemutaran dan pembalikan pada gambar dataset yang ada. Jenis pembalikan yang terjadi pada proses ini yaitu *horizontal flip*, *vertical flip*, dan *diagonal flip*. Bertujuan agar dataset lebih bervariasi.

3.4 Pengembangan model

3.4.1 *Stable diffusion 1.5*

Pada proses pengembangan model, *Stable diffusion* versi 1.5 berperan sebagai model dasar untuk menghasilkan model visual komik. Dalam penelitian ini, *Stable diffusion* versi 1.5 tidak secara langsung digunakan dengan bentuk aslinya. Namun, nantinya model *Stable diffusion* akan ditingkatkan menggunakan *fine tuning* dengan teknik LoRA. Dengan begitu akan menghasilkan model baru yang dapat menghasilkan model visual komik dengan gaya yang sesuai dengan dataset yang sudah dikumpulkan.

3.4.2 *Regularization Fine Tune*

Pada tahapan ini dilakukan proses *fine-tuning* atau training terhadap model *Stable Diffusion* versi 1.5 agar model dapat menghasilkan gambar dengan gaya dan karakteristik tertentu berdasarkan dataset yang ada. Teknik *fine-tuning* yang digunakan adalah LoRA (*Low Rank Adaptation*) dan Data Augmentasi.

Dengan *fine-tuning* LoRA, model dapat menyesuaikan tanpa harus melatih ulang seluruh parameter model, melainkan hanya melakukan pelatihan pada beberapa layer tertentu yang relevan. Dengan menerapkan *fine-tuning* menggunakan LoRA, diharapkan model dapat menghasilkan visual komik dengan gaya yang lebih spesifik dan sesuai dengan kebutuhan pengguna sambil tetap menjaga efisiensi proses pelatihan dan inferensi.

Selain itu, Data Augmentasi diterapkan untuk meningkatkan kualitas dan variasi data latih. Beberapa teknik augmentasi yang digunakan dalam proses ini meliputi:

1. Rotasi dan Flip Acak

Memungkinkan model belajar dari berbagai sudut gambar tanpa kehilangan makna.

2. *Crop* dan *Resize* Acak

Membantu model memahami berbagai komposisi gambar dengan lebih baik.

Dengan kombinasi LoRA dan Data Augmentasi, proses *fine-tuning* menjadi lebih optimal dalam menyesuaikan model dengan gaya visual komik

yang diinginkan, sekaligus memastikan bahwa model dapat menggeneralisasi dengan baik terhadap variasi input yang lebih luas.

3.5 Pengujian dan Evaluasi

3.5.1 Pengujian

Setelah melalui tahap pengembangan, selanjutnya adalah tahapan pengujian. Dimana pada tahap ini sistem akan menghasilkan gambar dari model yang sudah dilatih. Untuk menghasilkan gambar, prompt yang digunakan berupa teks yang sama dengan *caption* dari data pengujian. Alasan penggunaan teks yang 20 sama adalah untuk memudahkan proses perbandingan subjektif antara gambar hasil model baru dan gambar data pengujian.

3.5.2 Evaluasi

Evaluasi kualitas gambar menggunakan CLIP-MMD (*CLIP Maximum Mean Discrepancy*) bertujuan untuk mengukur tingkat kesesuaian gambar yang dihasilkan terhadap deskripsi teks. CLIP (*Contrastive Language Image Pretraining*) adalah model yang dilatih untuk mengenali hubungan antara teks dan gambar dalam satu ruang representasi. CLIP-MMD (*CLIP Maximum Mean Discrepancy*) dihitung dengan mengukur jarak antara dua distribusi vektor fitur yang dihasilkan oleh CLIP untuk gambar dan teks. Nilai CLIP-MMD yang lebih rendah menunjukkan bahwa distribusi gambar hasil model mendekati distribusi gambar nyata berarti kualitas gambar tinggi. Jika nilai CMMD tinggi berarti gambar hasil model memiliki distribusi yang jauh dari data nyata kualitas gambar rendah.

3.6 Alur Kerja Training Sistem

Alur sistem ini bertujuan untuk menghasilkan panel komik yang sesuai dengan deskripsi teks dan gaya visual dari komik tersebut.

Pendekatan yang digunakan dengan memanfaatkan model pembelajaran mesin berbasis diffusion model, yaitu stable diffusion 1.5. Model ini telah difine tune menggunakan LoRA dan Data Augmentasi agar mampu menghasilkan gambar yang lebih spesifik dan sesuai dengan gaya visual komik diinginkan. Alur kerja training dapat dilihat pada **Lampiran 1**.

Pada gambar yang ditunjukkan pada lampiran 1 adalah alur kerja *stable diffusion* yang dikombinasikan dengan Lora serta Data Augmentation, pada alur tersebut memiliki tahap yang untuk menghasilkan generasi gambar. Berikut adalah alur kerja training menggunakan Stable Diffusion v1.5 dan fine tuning LoRA yang ditunjukkan pada gambar 8.1

1. *Dataset* atau *Training sample*

Dataset atau *training sample* ini menyediakan koleksi gambar yang digunakan sebagai dasar pelatihan model atau pretrained model. *Dataset* ini mencakup gambar yang sesuai dengan domain atau gaya yang relevan untuk kebutuhan proyek misalnya gaya atau estetika yang diinginkan dan memiliki variasi yang cukup dalam pose, ekspresi dan komposisi untuk mencegah bias model misalnya ilustrasi komik, karakter, atau latar belakang. *Dataset* ini biasanya dilengkapi sebuah label, deskripsi teks atau anotasi yang menghubungkan visual dengan teks selama pelatihan (*text-to-image-task*). Ukuran dataset dengan besar dan kualitas tinggi agar model dapat belajar dengan baik.

2. *Input text*

Input text dalam *stable diffusion* berfungsi sebagai prompt yaitu deskripsi yang memberikan arahan kepada model tentang gambar yang ingin dihasilkan dapat memberikan detail spesifik dalam prompt untuk mempengaruhi hasil gambar.

3. *Stable diffusion*

Pada bagian *stable diffusion* ini memilih tahap yaitu *image information creation*, *apply noise + n*, *apply noise -n*, *processed image tensor*, *data augmentation*, *image decoder*. Berikut penjelasannya :

1) *Text Encoder & Frozen CLIP*

Text Encoder menggunakan model CLIP (*Contrastive Language-Image Pretraining*) untuk mengonversi teks menjadi representasi numerik atau token embeddings. CLIP berfungsi sebagai penghubung antara deskripsi teks dan elemen visual.

2) *Token Embeddings*

Token embeddings adalah representasi vektor dari input teks yang digunakan sebagai masukan ke *Diffusion Model*. *Embeddings* ini memberikan konteks kepada model tentang detail dan gaya gambar yang harus dihasilkan.

3) *Image information creation*

a) *Diffusion Model*

Proses inti *Stable Diffusion* dimulai di sini . Input ke *Diffusion Model* adalah gambar yang sudah diberi *noise* (*noise image*) dan *token embeddings*. Prosesnya yaitu Model mengambil *noise image* sebagai masukan dan Model dilatih untuk memprediksi dan menghapus *noise*, seperti berikut :

- *Apply Noise + $\backslash(n\backslash)$ (Forward Diffusion)*

Forward diffusion digunakan selama pelatihan, di mana *noise* ditambahkan secara bertahap ke gambar asli untuk menciptakan *noise image* . Proses ini digunakan untuk mengajari model memahami distribusi *noise*.

- *Apply Noise + $\backslash(n-1\backslash)$ (Reverse Diffusion)*

Reverse diffusion digunakan dalam proses generasi gambar, di mana model secara iteratif menghapus *noise* dari *noise image*. Proses ini bertujuan untuk menghasilkan gambar final berdasarkan *embeddings* teks (*model output image*).

b.) U-Net (*Neural Network dalam Diffusion Model*)

Bagian utama dari *Diffusion Model* adalah arsitektur U-Net . Fungsi U-Net adalah memproses *noise image* berdasarkan *embeddings* teks dan memprediksi *noise* yang perlu dihapus di setiap langkah.U-Net membandingkan gambar hasil saat ini dengan *sample calibration image* untuk memastikan kualitasnya (mengurangi perbedaan gambar).

4) *Processed Image Tensor*

Processed Image Tensor berisi *Noise Image & Remove Noise*. *Noise Image* adalah Gambar yang telah diberi *noise* sepenuhnya dalam *forward diffusion*. Dalam *reverse diffusion*, *noise image* diproses secara bertahap untuk menghasilkan *denoised image* (gambar bersih).

5). Data Augmentasi

Setelah gambar di-*denoise*, data augmentasi diterapkan untuk meningkatkan kualitas atau menyesuaikan hasil sesuai kebutuhan proyek. Augmentasi juga membantu menciptakan variasi gambar untuk memperkaya dataset.

6). *Image Decoder*

Denoised image yang telah selesai diterjemahkan oleh decoder menjadi gambar dalam format yang dapat dilihat. *Decoder* bertanggung jawab untuk mengubah tensor menjadi gambar akhir yang lebih realistis.

4. *Generated Image*

Gambar akhir dihasilkan berdasarkan input teks awal. Ini adalah hasil yang sesuai untuk digunakan sebagai panel komik dengan detail yang telah *define-tune*.

3.7 Alur Kerja Sistem User

Setelah memahami alur sistem training secara teknis, alur kerja sistem user yaitu sistem yang digunakan user agar dapat menghasilkan panel komik dari deskripsi teks. Alur kerja sistem user dapat dilihat pada **Lampiran 2**.

Pada lampiran 2 mengilustrasikan alur kerja sistem pembangkitan gambar AI yang menggunakan *Stable Diffusion*. Ini dimulai dengan input pengguna melalui antarmuka Streamlit, memvalidasi *prompt*, dan kemudian memprosesnya menggunakan model *Stable Model Difusi Stabil*. Sistem mengoptimalkan model, memindahkannya ke GPU, dan menghasilkan gambar berdasarkan *prompt* menggunakan *seed* acak. Terakhir, sistem menampilkan gambar yang dihasilkan kepada pengguna. Langkah-langkah yang terlibat :

1. Pengguna mengetikkan *prompt*

Dengan menggunakan antarmuka Streamlit, pengguna mengetikkan “Sebuah Apel” ke dalam bidang teks. Sebuah string Python dibuat dari string mentah ini.

2. Antarmuka Streamlit

Masukan diproses oleh kerangka kerja Streamlit. Dengan menghapus semua yang berpotensi berbahaya karakter atau skrip yang berpotensi berbahaya, ini dapat membersihkan input. Setelah itu, prompt disiapkan untuk pemrosesan selanjutnya, mungkin dengan menambahkan metadata seperti stempel waktu atau ID pengguna, atau dengan mengenkapsulasinya dalam struktur data.

3. Menerima *Prompt*

Prompt yang telah diproses dikirimkan ke sistem *backend*. Apakah *Prompt Valid*? *Prompt* diperiksa oleh fungsi validasi. Fungsi ini dapat memverifikasi minimum dan maksimum durasi. Cari kata atau ekspresi yang dilarang. Pastikan tidak ada informasi pribadi. Periksa untuk menentukan. Apakah lolos dan mengembalikan nilai benar untuk kueri “Sebuah apel”.

4. Memuat *Stable Diffusion*

Model `runwayml/stable-diffusion-v1-5` dimuat oleh sistem. Hal ini membutuhkan mengambil bobot dan arsitektur model dari tempat penyimpanan. Menyiapkan model dalam memori Mengkonfigurasi lapisan internal dan parameter persiapan model untuk mode inferensi

5. Menggunakan Transformasi untuk Memproses Teks

Perintah ditangani menggunakan bahasa alami. Membagi “Sebuah Apel” menjadi token dikenal sebagai tokenisasi. Token diterjemahkan ke representasi numerik selama penyandian. Memastikan memastikan bahwa input memiliki panjang yang tepat untuk model dikenal sebagai padding. Memasukkan token unik yang mirip dengan [SEP] dan [CLS] untuk model yang didasarkan pada BERT Convolutional neural networks (CNN) dengan beberapa lapisan konvolusi, pooling dan unit non-linear telah

menunjukkan keberhasilan yang cukup besar dalam tugas-tugas visi komputer.

6. *Optimize Model*

Sebuah perpustakaan yang mempercepat dan mengoptimalkan model menggunakan berbagai metode untuk ketepatan Model pelatihan disusun berdasarkan perangkat keras yang tersedia, merakit pemrosesan data terbaik dan memuat saluran pipa

7. CUDA untuk memindahkan model ke GPU

Model dan bobotnya disalin ke memori GPU RAM GPU dialokasikan dengan menjalankan fungsi CUDA. GPU menyalin parameter model dari CPU. Grafik komputasi adalah dikonfigurasi untuk pemanfaatan GPU.

8. Menetapkan Generator Acak

Menginisialisasi generator angka acak, sebuah seed ditetapkan, mungkin menggunakan waktu saat ini atau kuantitas yang telah ditentukan sebelumnya. Keadaan acak disiapkan untuk memastikan pengulangan. Generator ini akan memengaruhi difusi proses difusi dan menyuntikkan ketidakpastian yang disengaja ke dalam proses pembuatan visual.

9. Gunakan Diffuser untuk Menghasilkan Gambar

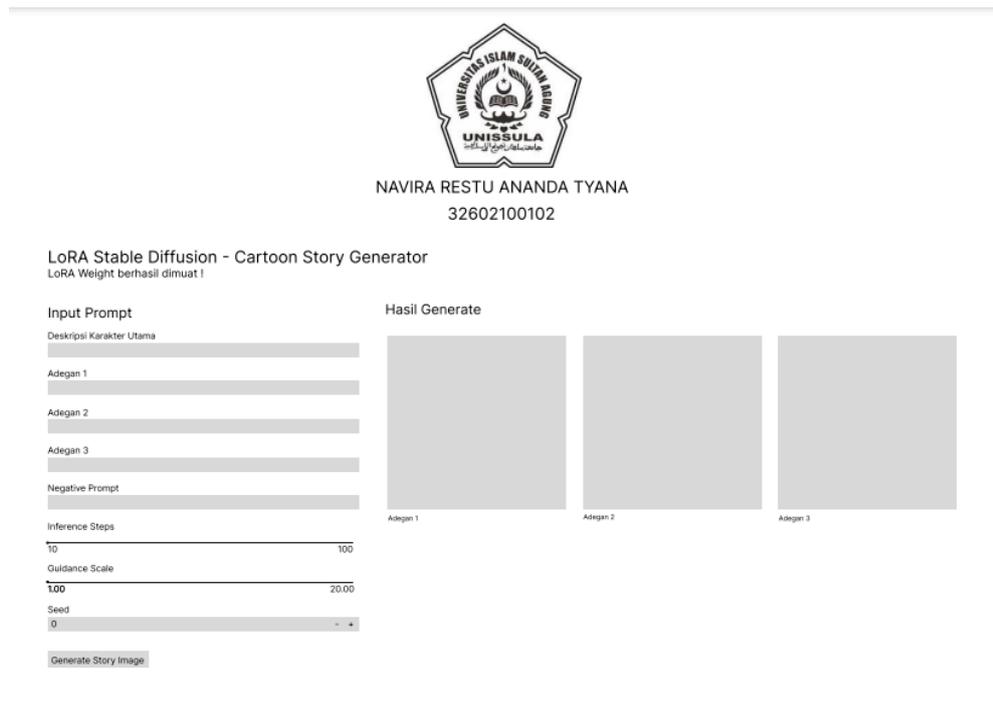
Proses pembuatan gambar kode berlangsung. Kebisingan awal Distribusi noise awal diatur oleh prompt yang dikodekan. Secara iteratif, proses difusi mengubah noise menjadi gambar. Model memprediksi dan menghilangkan noise pada setiap tahap. Metode ini diulangi beberapa kali.

10. Menampilkan gambar

Data foto yang sudah diproses ditampilkan. Setelah itu, data piksel yang belum diproses diformat. Streamlit dapat ditampilkan (PNG, misalnya). Pasca-pemrosesan atau penyesuaian warna dapat digunakan. Foto telah diskalakan agar sesuai dengan antarmuka pengguna. Untuk merender gambar di browser pengguna, fungsi tampilan gambar Streamlit digunakan (Prasad dkk. 2024)

3.8 Perancangan *Interface*

Pada tahap ini penulis merancang *user interface* untuk penggunaan sistem. Tujuan adanya user interface dari sistem adalah untuk mempermudah pengguna jika ingin menghasilkan model visual dari sistem ini, yaitu sistem dengan model *Stable diffusion* versi 1.5 yang telah *difine-tuning* menggunakan teknik LoRA.



Gambar 3.1 Wireframe Web

Gambar diatas merupakan wireframe dari sistem generate model visual komik. Pada saat menggunakan sistem, pengguna harus memasukkan beberapa parameter seperti :

1. *Prompt* : berisikan deskripsi teks yang menggambarkan karakter utama yang ingin dihasilkan serta adegan 1,2,3 yang berisi alur cerita karakter utama.
2. *Jumlah inference step*: digunakan untuk menentukan jumlah iterasi dalam proses menghasilkan gambar. Semakin tinggi inference step maka gambar akan semakin detail namun prosesnya akan membutuhkan waktu yang lebih lama.

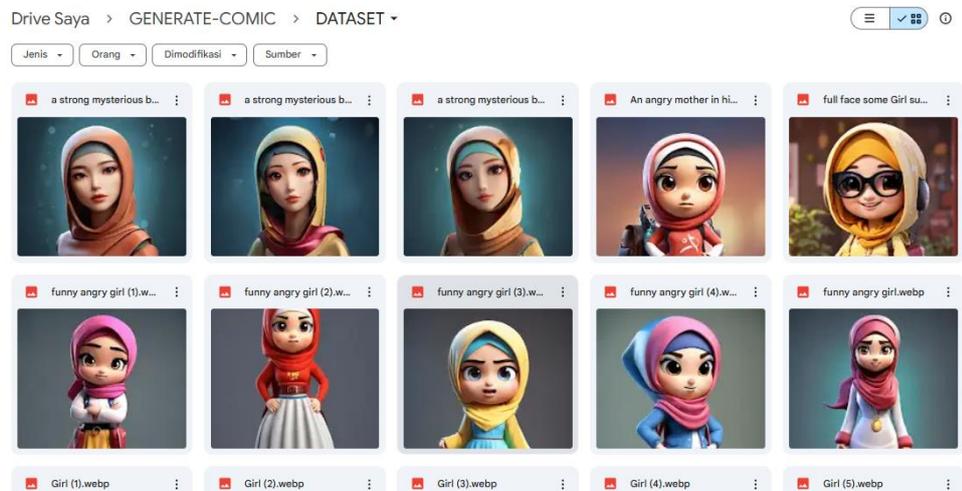
3. *Guidance scale*: digunakan untuk mengontrol seberapa kuat gambar harus mengikuti prompt.
4. *Seed*: angka acak untuk memastikan replikasi gambar yang sama.



BAB IV HASIL DAN ANALISIS PENELITIAN

4.1 Persiapan Dataset

Dataset yang berhasil dikumpulkan oleh penulis berupa gambar wanita hijab bergaya kartun yang diperoleh diberbagai platform. Kemudian dilakukan beberapa proses sebelum digunakan ke tahap selanjutnya. Proses tersebut berupa captioning dan pembagian dataset.



Gambar 4. 1 Kumpulan Dataset

Gambar diatas merupakan hasil dari pengumpulan dataset yang belum dibagi menjadi data pelatihan dan data pengujian. Dataset tersebut tersimpan pada folder dengan nama dataset dan berisikan 52 gambar yang terdiri dari gambar gadis yang berhijab yang bersumber dari beberapa platform.

4.2 Pengolahan Dataset

Dataset yang berhasil dimasukan ke folder selanjutnya dilakukan preprocessing untuk meningkatkan variasi dataset dan mempermudah sistem untuk mengenali dataset karena data bersih dan konsisten.

Berikut kode yang digunakan :

```
from google.colab import drive
import os
from PIL import Image
from tqdm import tqdm
import torch
from transformers import BlipProcessor,
BlipForConditionalGeneration
```

```

# Mount Google Drive
drive.mount('/content/drive')

# Direktori input dan output
INPUT_DIR = "/content/drive/MyDrive/GENERATE-
COMIC/DATASET" # Dataset asli
OUTPUT_DIR = "/content/drive/MyDrive/GENERATE-
COMIC/DATASET-RESULT" # Dataset hasil preprocessing

# Pastikan folder output ada
os.makedirs(OUTPUT_DIR, exist_ok=True)

# Ukuran target
TARGET_SIZE = (512, 512)

# Muat model BLIP untuk image captioning
processor =
BlipProcessor.from_pretrained("Salesforce/blip-image-
captioning-base")
model =
BlipForConditionalGeneration.from_pretrained("Salesforce/b
lip-image-captioning-base")

# Fungsi untuk memproses gambar dan menghasilkan deskripsi
def preprocess_images_and_generate_caption(input_dir,
output_dir, target_size):
    for img_file in tqdm(os.listdir(input_dir),
desc="Processing Images"):
        input_path = os.path.join(input_dir, img_file)

        # Pastikan hanya memproses file gambar
        if not img_file.lower().endswith(('.png', '.jpg',
'.jpeg', '.bmp', '.gif', '.webp')):
            continue

        try:
            # Buka gambar
            with Image.open(input_path) as img:
                # Konversi ke RGB (jika grayscale atau
lainnya)
                img = img.convert("RGB")

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

                # Ubah ekstensi output menjadi .jpg
                output_filename =
os.path.splitext(img_file)[0] + ".png"
                output_path = os.path.join(output_dir,
output_filename)

                # Simpan gambar yang sudah diresize dalam
format JPG

```

```

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

        # Proses gambar untuk menghasilkan caption
        inputs = processor(images=img,
return_tensors="pt")
        out = model.generate(**inputs)
        caption = processor.decode(out[0],
skip_special_tokens=True)

        # Simpan caption dalam file .txt dengan
nama yang sama
        caption_filename =
os.path.splitext(img_file)[0] + ".txt"
        caption_path = os.path.join(output_dir,
caption_filename)
        with open(caption_path, 'w') as
caption_file:
            caption_file.write(caption)

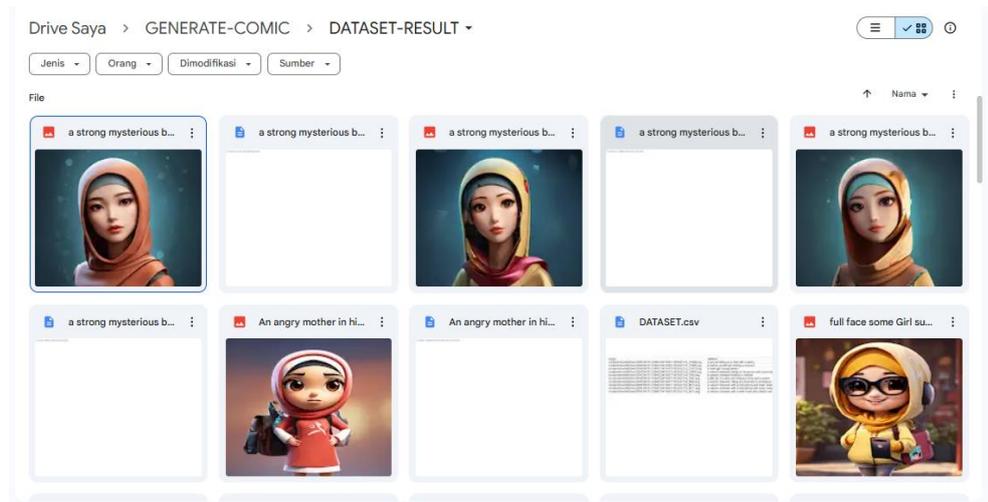
    except Exception as e:
        print(f"Error processing {img_file}: {e}")

# Jalankan preprocessing dan generate caption
preprocess_images_and_generate_caption(INPUT_DIR,
OUTPUT_DIR, TARGET_SIZE)

```

Kode di atas merupakan kode untuk tahapan pengolahan dataset. Pengolahan dataset tersebut meliputi perubahan ukuran dan perubahan tipe file.

Kode untuk melakukan captioning terlihat pada **Lampiran 3. Captioning** pada dataset, yaitu menambahkan deskripsi teks pada gambar berdasarkan elemen-elemen yang terdapat di dalamnya. Dalam proses ini, penulis memanfaatkan model *Bootstrapping Language Image Pretraining* (BLIP) untuk membantu menghasilkan deskripsi yang sesuai. BLIP berfungsi dalam menciptakan teks deskriptif serta mencocokkan gambar dengan deskripsinya.



Gambar 4. 2 Kumpulan hasil dataset yang sudah di *captioning*

Gambar diatas merupakan hasil dari proses *captioning* dari dataset awal. Pada folder dataset result berisi 52 gambar dan 52 teks deskripsi dalam bentuk txt. Selain itu, pada folder tersebut dataset yang ada sudah dipasangkan dengan teks deskripsi.

Tabel 4. 1 Pembagian Dataset

No	Dataset	Jumlah	Jumlah Caption
1	Data Pelatihan	41	41
2	Data Pengujian	11	11
Total		52	52

Tabel diatas menunjukkan jumlah pembagian dataset. Terdapat data pelatihan, yang digunakan untuk melatih model. Kemudian data pengujian yang digunakan untuk pengujian hasil gambar model setelah *fine-tuning*.

Kode Sumber :

```
import os
import shutil
from sklearn.model_selection import train_test_split

# Directories for images and captions
data_folder = "/content/drive/MyDrive/GENERATE-COMIC/DATASET-RESULT"
train_folder = "/content/drive/MyDrive/GENERATE-COMIC/TRAINING IMAGES"
test_folder = "/content/drive/MyDrive/GENERATE-COMIC/TEST IMAGES"
```

```

# Periksa isi folder data
file_list = [f for f in os.listdir(data_folder) if
f.lower().endswith(('.png', '.jpg', '.jpeg'))]

# Cek apakah file gambar ditemukan
print(f"Jumlah gambar yang ditemukan: {len(file_list)}")
print(f"Daftar file gambar: {file_list}")

# Split into training and testing sets
train_files, test_files = train_test_split(file_list,
test_size=0.2, random_state=42)

# Cek jumlah file setelah split
print(f"Jumlah file training: {len(train_files)}")
print(f"Jumlah file testing: {len(test_files)}")

# Fungsi untuk menyalin gambar dan caption
def copy_files(file_name, source_folder,
destination_folder):
    image_path = os.path.join(source_folder, file_name)
    text_path = os.path.splitext(image_path)[0] + '.txt' #
Mengasumsikan file caption memiliki nama yang sama dengan
gambar dan ekstensi .txt

    # Periksa apakah file gambar ada
    if os.path.exists(image_path):
        # Salin file gambar
        shutil.copy(image_path,
os.path.join(destination_folder, file_name))
    else:
        print(f"Warning: Gambar tidak ditemukan untuk
{file_name}")

    # Salin file caption jika ada
    if os.path.exists(text_path):
        shutil.copy(text_path,
os.path.join(destination_folder,
os.path.basename(text_path)))
    else:
        print(f"Warning: Caption file tidak ditemukan untuk
{file_name}")

# Pastikan folder tujuan ada
os.makedirs(train_folder, exist_ok=True)
os.makedirs(test_folder, exist_ok=True)

```

```

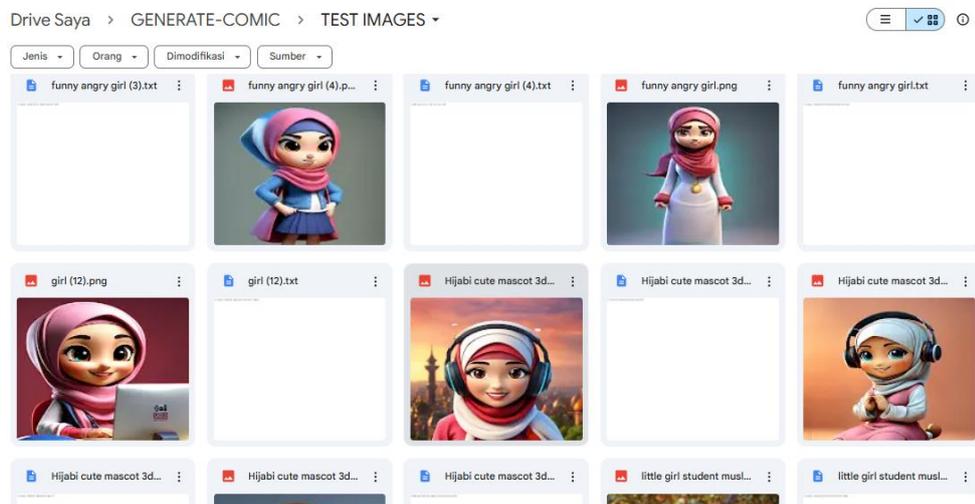
# Salin file untuk training dan testing
print("Menyalin file untuk training...")
for file_name in train_files:
    copy_files(file_name, data_folder, train_folder)

print("Menyalin file untuk testing...")
for file_name in test_files:
    copy_files(file_name, data_folder, test_folder)

print("Dataset berhasil dibagi menjadi training dan
testing, termasuk file caption!")

```

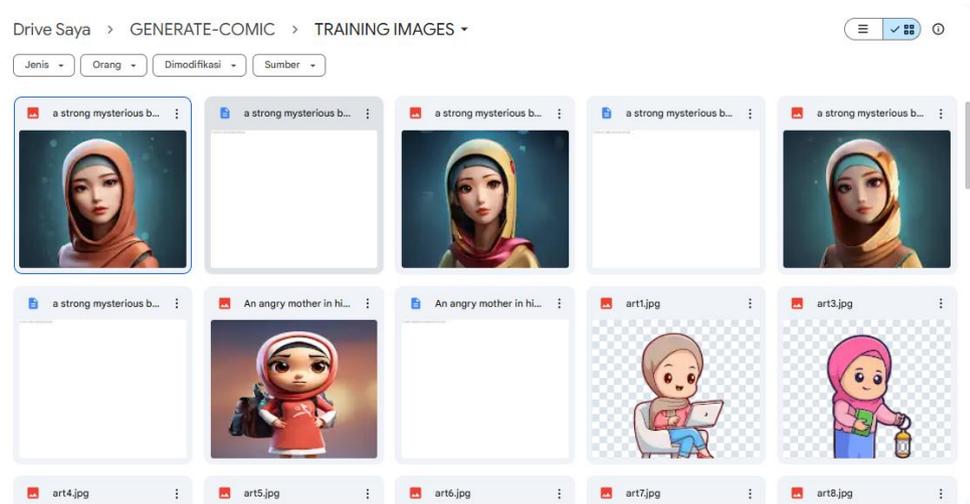
Kode di atas digunakan untuk membagi dataset yang telah diberi deskripsi atau caption menjadi dua bagian: data pelatihan dan data pengujian. Pembagian dilakukan dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Selain itu, data hasil pembagian ini juga disimpan dalam format CSV untuk memudahkan pemrosesan lebih lanjut.



Gambar 4. 3 test image

Gambar diatas merupakan hasil dari pembagian dataset, yaitu data pengujian. Dataset tersebut tersimpan pada folder dengan nama *TEST IMAGES*. Didalamnya terdapat 11 data pengujian yang nantinya akan digunakan pada proses pengujian terhadap hasil gambar model setelah *fine-*

tuning.



Gambar 4. 4 Training Image

Gambar diatas merupakan hasil dari pembagian dataset, yaitu data pelatihan. Dataset tersebut tersimpan pada folder dengan nama *TRAINING IMAGES*. Didalamnya terdapat 41 data pelatihan yang nantinya akan digunakan pada proses training model.

4.3 Pengembangan Model

Pada tahap ini penulis telah melakukan *fine-tuning* pada model *Stable diffusion*. Tujuan dilakukannya proses ini adalah untuk melatih agar model dapat menghasilkan gambar yang sesuai dengan dataset. Adapun penulis melakukan uji coba terhadap 500 iterasi training pada model *Stable diffusion*.

Berikut kode yang digunakan :

```
# Step 4: Run the training script
!accelerate launch /content/drive/MyDrive/GENERATE-COMIC/train_text_to_image_lora_csv.py \
  --pretrained_model_name_or_path="runwayml/stable-diffusion-v1-5" \
  --train_data_dir="/content/drive/MyDrive/GENERATE-COMIC/TRAINING IMAGES/train.csv" \
  --image_column="image" \
  --caption_column="captions" \
  --dataloading_workers=8 \
  --resolution=512 \
  --center_crop \
  --random_flip \
  --train_batch_size=1 \
```

```

--gradient_accumulation_steps=4 \
--max_train_steps=500 \
--learning_rate=1e-04 \
--max_grad_norm=1 \
--lr_scheduler="cosine" \
--lr_warmup_steps=0 \
--output_dir="/content/drive/MyDrive/GENERATE-COMIC/LORA-
OUTPUT" \
--checkpointing_steps=50 \
--validation_prompt="A cartoon of A cute young girl
wearing a blue hijab, smiling warmly while holding a gift
box, detailed 3D rendering, soft lighting, warm night
atmosphere, bokeh lights in the background, ultra-realistic
textures, vibrant colors, highly detailed, Pixar-style,
cinematic shot" \
--seed=1337

```

Kode diatas menunjukkan parameter yang digunakan saat pelatihan model. Parameter tersebut berfungsi untuk mengontrol proses pelatihan model dan berpengaruh pada kesesuaian bobot terhadap data. Berikut adalah beberapa parameter yang digunakan :

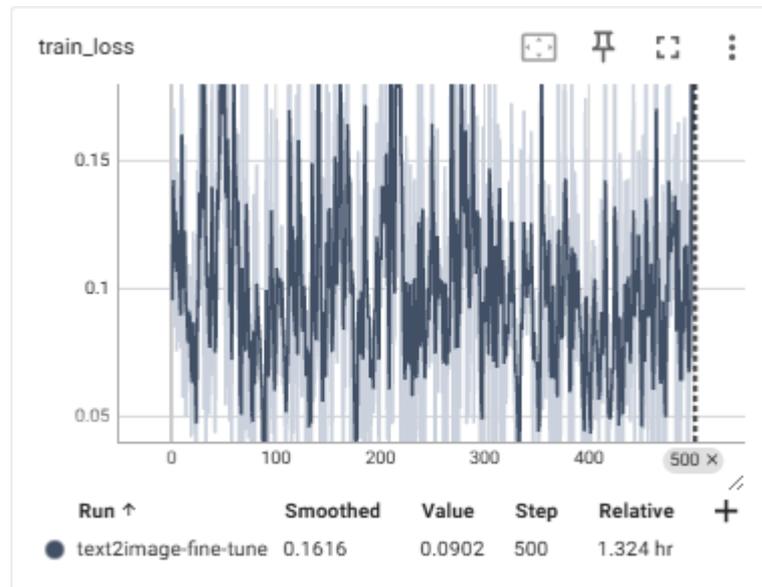
Tabel 4. 2 Parameter

Parameter	Keterangan
--pretrained_model_name_or_path	Model dasar yang digunakan sebagai pretrained model. Di sini menggunakan Stable Diffusion v1-5 dari runwayml.
--train_data_dir	Path ke file CSV yang berisi daftar gambar dan caption yang digunakan untuk pelatihan.
--image_column	Nama kolom dalam CSV yang berisi path gambar.
--caption_column	Nama kolom dalam CSV yang berisi deskripsi atau caption dari gambar.
--dataloaders_num_workers	Jumlah worker yang digunakan untuk memproses data dalam

	batch saat training. Semakin tinggi, semakin cepat loading datanya (disarankan 4 atau 8 tergantung kapasitas CPU).
<code>--resolution</code>	Resolusi gambar yang akan digunakan saat pelatihan (di sini 512x512).
<code>--center_crop</code>	Mengaktifkan crop di tengah gambar agar sesuai dengan resolusi.
<code>--random_flip</code>	Mengaktifkan augmentasi random flip horizontal untuk meningkatkan variasi data.
<code>--train_batch_size</code>	Jumlah gambar dalam satu batch pelatihan. Batch size kecil (1) digunakan agar lebih hemat VRAM.
<code>--gradient_accumulation_steps</code>	Mengakumulasi gradien selama 4 langkah sebelum melakukan update model, sehingga simulasi batch lebih besar tanpa memakan terlalu banyak memori GPU.
<code>--max_train_steps</code>	Jumlah total langkah pelatihan yang akan dilakukan (500 step).
<code>--learning_rate</code>	Nilai learning rate (kecepatan belajar) model, diatur $1e-4$ (0.0001) agar model belajar secara stabil.

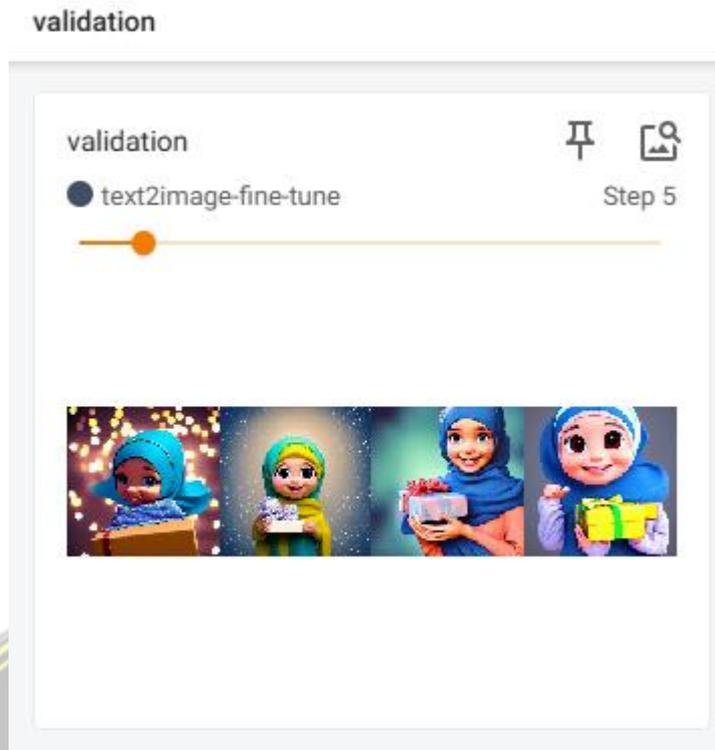
<code>--max_grad_norm</code>	Batas maksimum normalisasi gradien untuk menghindari eksplosif gradient.
<code>--lr_scheduler</code>	Jenis penjadwalan learning rate, di sini menggunakan cosine yang secara bertahap menurunkan learning rate.
<code>--lr_warmup_steps</code>	Jumlah langkah pemanasan (warmup) sebelum penyesuaian learning rate.
<code>--output_dir</code>	Direktori tempat menyimpan model hasil pelatihan dan checkpoint.
<code>--checkpointing_steps</code>	Setiap 50 langkah akan menyimpan checkpoint (snapshot dari model).
<code>--validation_prompt</code>	Prompt validasi untuk mengecek kualitas model setelah pelatihan dengan deskripsi spesifik.
<code>--seed</code>	Seed random agar hasil pelatihan bisa direproduksi dengan hasil yang sama (1337 digunakan untuk konsistensi).

Ketika melakukan proses pelatihan, penulis memantau *loss* yang ada selama proses pelatihan berlangsung melalui grafik yang berhasil dikirimkan melalui *tensorboard*.



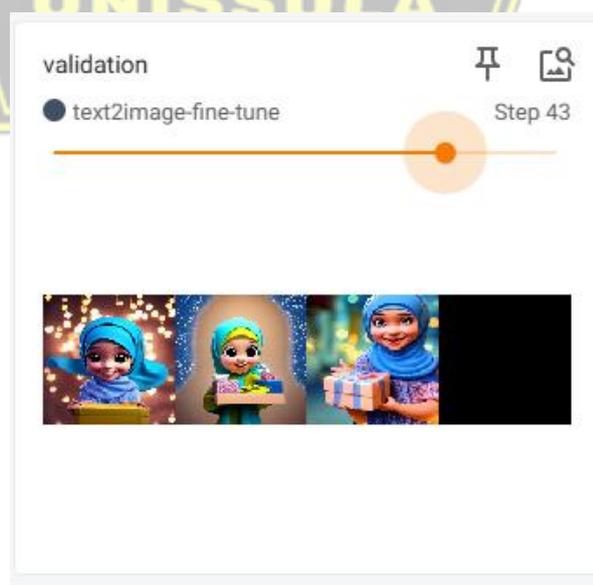
Gambar 4. 5 *train loss*

Gambar diatas menunjukkan grafik *loss* pada saat proses *training* berlangsung. Berdasarkan grafik tersebut, terlihat bahwa nilai *loss* mengalami penurunan secara bertahap beriringan dengan bertambahnya langkah iterasi. Hal tersebut menunjukkan bahwa proses *training* model berhasil mempelajari data yang ada. Namun juga terdapat kondisi dimana *loss* mengalami fluktuasi, dimana pada proses *training* nilai *loss* naik dan turun secara drastis. Ini menunjukkan adanya ketidakstabilan dalam proses model mempelajari data. Secara keseluruhan, grafik tersebut menunjukkan bahwa model belajar dengan baik, tetapi masih mengalami *fluktuasi loss* yang dapat mengakibatkan terhambatnya konvergensi model. Selain grafik *loss*, terdapat juga validasi *prompt* yang dihasilkan selama proses pelatihan berlangsung pada setiap iterasinya. Validasi *prompt* digunakan untuk mengevaluasi sejauh mana model mampu memahami dan menghasilkan *output* yang sesuai dengan *prompt* yang diberikan. Selama proses pelatihan, validasi *prompt* dilakukan pada setiap iterasi untuk mengukur performa model dalam menerjemahkan *prompt* menjadi *output* yang diharapkan.



Gambar 4. 6 Hasil Validasi step 5

Hasil validasi menunjukkan bahwa model telah berhasil menghasilkan gambar namun masih belum merepresentasikan karakter dengan baik. jika dilihat dengan subjektif gambar tersebut masih ada yang kurang sempurna dalam membentuk gambar.



Gambar 4. 7 Hasil Validasi step 43

Gambar diatas merupakan hasil validasi gambar step 43 dari model yang dilatih menggunakan teknik LoRA fine-tuning. Gambar tersebut menggunakan prompt “*A cartoon of A cute young girl wearing a blue hijab, smiling warmly while holding a gift box, soft lighting, warm night atmosphere, bokeh lights in the background, ultra-realistic textures, vibrant colors, highly detailed, cinematic shot*”. Secara umum, model berhasil menghasilkan suatu karakter dengan bentuk dan warna yang cukup konsisten.

4.4 Hasil Generate Gambar

4.4.1 Penggunaan Stable Diffusion

Berikut merupakan hasil generasi gambar menggunakan Stable Diffusion 1.5 dengan *prompt*: “ *A cute young girl wearing a blue hijab, smiling warmly while holding a gift box*”. Model ini digunakan untuk menghasilkan ilustrasi karakter anak perempuan menggunakan hijab warna biru , senyum hangat sambil memegang kotak hadiah.



Gambar 4. 8 Ouput Stable Diffusion

Gambar 4.8 menunjukkan bahwa model telah berhasil menghasilkan ilustrasi yang sesuai dengan prompt yang diberikan. Gambar sudah cukup

menggambarkan tema yang diminta, terdapat beberapa aspek yang mungkin tidak sepenuhnya sesuai, seperti objek utama masih kaku serta bagian tubuh masih kurang jelas.

4.4.2 Penggunaan *Stable Diffusion* dan *Negative Prompt*

Gambar berikut merupakan hasil generasi menggunakan *Stable Diffusion 1.5* dengan *prompt* yang sama seperti sebelumnya: "" *A cute young girl wearing a blue hijab, smiling warmly while holding a gift box*"", 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. 9 Hasil *Stable Diffusion* dan *Negative Prompt*

Gambar 4.9 menunjukkan bahwa model telah berhasil menghasilkan ilustrasi yang sesuai dengan *prompt* yang diberikan. Gambar sudah cukup

menggambarkan tema yang diminta dan gambar sudah tidak kaku namun proporsi tubuh masih kurang.

4.4.3 Penggunaan Fine Tune

Gambar berikut merupakan hasil generasi menggunakan *Stable Diffusion* 1.5, 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: "A cute young girl wearing a blue hijab, smiling warmly while holding a gift box". 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. 10 hasil *fine tune*

Gambar 4.10 menunjukkan perubahan yang baik dibandingkan dengan model sebelumnya. Detail dalam ilustrasi menjadi lebih jelas, dan komposisi tampak lebih mendekati konsep karakter kartun dengan fokus utama pada gadis hijab namun kartun yang digunakan semi realis terkesan 3D.

4.4.4 Penggunaan Fine Tune dan Negative Prompt

Gambar berikut merupakan hasil generasi menggunakan *Stable Diffusion* 1.5 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

gadis hijab gaun biru 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 karakter komik.



Gambar 4. 11 Hasil *fine-tune* dan *negative prompt*

Gambar 4.11 menunjukkan peningkatan dibandingkan eksperimen sebelumnya. Komposisi lebih rapi dengan karakter utama yang jelas serta sudah sesuai dengan style yang diinginkan yaitu kartun, serta elemen pendukung yang lebih tertata.

4.5 Tahap Pengujian

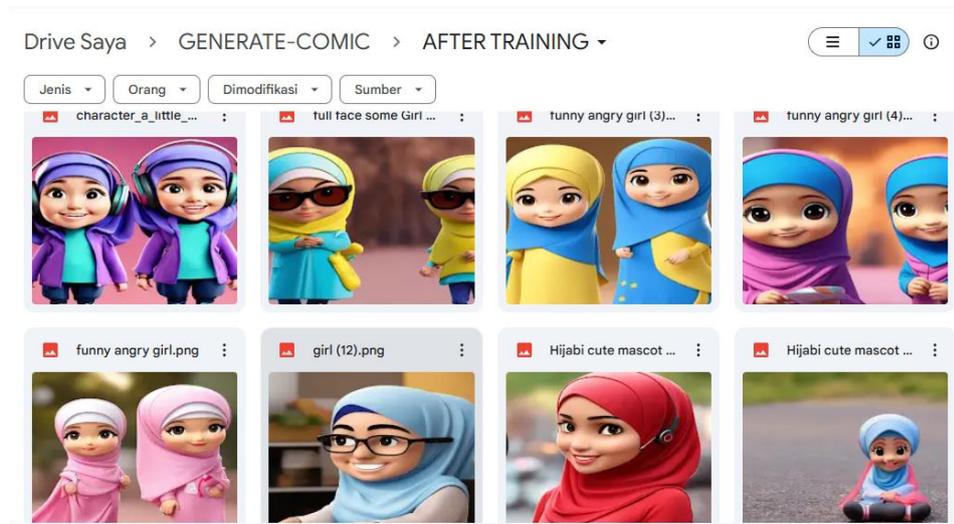
4.5.1 Pengujian

Pada tahap pengujian, model akan melakukan generate gambar berdasarkan dataset pengujian. Model akan menggunakan teks deskripsi yang sama dengan dataset pengujian dan menghasilkan gambar yang sesuai dengan deskripsi teks. Berikut adalah teks deskripsi atau caption pada data pengujian :

Tabel 4. 3 Gambar Pengujian

Gambar	Nama File	Prompt
	full face some Girl super hiro.png	a little girl in a yellow jacket and sunglasses
	funny angry girl (3).png	a cartoon character in a yellow and blue dress
	funny angry girl (4).png	a little girl in a blue coat and pink scarf
	funny angry girl.png	a cartoon character in a white dress and pink scarf
	girl (12).png	a cartoon character sitting at a desk with a laptop

	Hijabi cute mascot 3d wears headset (1).png	a woman in headphones and a red shirt
	Hijabi cute mascot 3d wears headset (5).png	a cartoon character sitting on the ground
	Hijabi cute mascot 3d wears headset (6).png	a little girl wearing headphones and a purple jacket
	little girl student muslim.png	a cartoon character in a blue coat and pink pants
	little hijab girl (2).png	a little girl dressed as a superhero
	muslim superhero.png	a cartoon character holding a box in a street



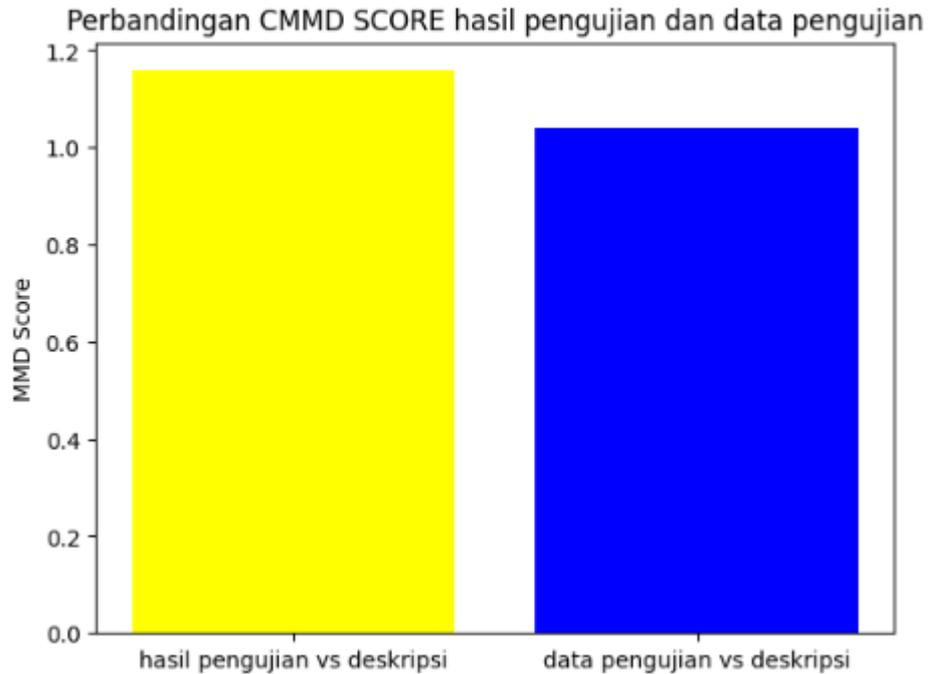
Gambar 4. 12 Kumpulan gambar hasil model.

Gambar diatas adalah kumpulan gambar yang dihasilkan model menggunakan prompt yang sama dengan gambar uji. Gambar menunjukkan hasil pengujian menggunakan model baru tanpa menggunakan *negative prompt*. Secara visual, gambar yang dihasilkan sudah jelas dan dapat menyerupai karakter dengan gaya kartun. Ini menunjukkan model mampu menghasilkan karakter sesuai dengan deskripsi.

4.5.2 Evaluasi

Pada bagian ini, gambar yang dihasilkan oleh model pada tahap pengujian dilakukan evaluasi. Evaluasi berfungsi untuk menilai seberapa sesuai dan bagus kualitas dari gambar yang dihasilkan model. Menggunakan CLIP-MMD (CLIP Maximum Mean Discrepancy) sebagai metrik pengujian dengan hasil skor yang diperoleh ketika dilakukan evaluasi menggunakan CLIP-MMD direpresentasikan pada diagram dibawah.

→ CMMD Score (Test Dataset vs Text): 1.1575220823287964
CMMD Score (Lora Output vs Text): 1.0400726795196533
Model fine-tuned menghasilkan gambar yang lebih sesuai dengan prompt!



Gambar 4. 13 Hasil Skor CMMD

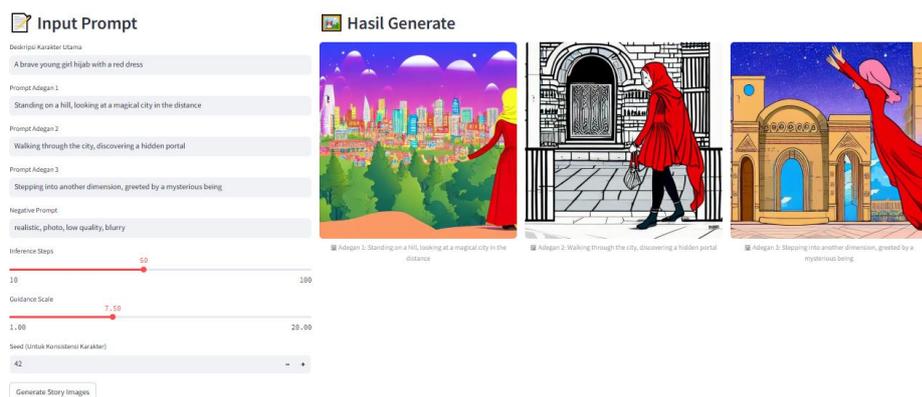
Gambar diatas menunjukkan score yang diperoleh CLIP-MMD dalam membandingkan kesesuaian gambar dengan teks deskripsi yang dihasilkan oleh model baru dan data uji. Dataset pengujian mendapatkan *score* 1.1575220823287964 sedangkan model baru menghasilkan *score* 1.0400726795196533. Hal ini menunjukkan bahwa gambar yang dihasilkan oleh model baru menghasilkan gambar yang lebih sesuai dengan teks deskripsi, meskipun hanya memiliki perbedaan yang tipis. Karena sesuai dengan cara kerja CMMD jika hasil *embedding text* dan gambar hasil model dibanding *embedding text* dan dataset uji mengalami penurunan dapat dikatakan hasil model sudah sesuai dengan deskripsi teks dan model belajar dengan baik.

4.6 Perancangan Sistem

Berikut adalah tampilan website dari sistem generate model visual yang siap digunakan oleh pengguna. Pengguna dapat melakukan generate gambar pada website streamlit yang telah dibuat.



Gambar 4. 14 Bagian atas *Interface*



Gambar 4. 15 Bagian *input prompt* dan hasil *generate*

Gambar diatas merupakan tampilan website, pengguna dapat memasukkan beberapa parameter seperti prompt, jumlah gambar, seed, jumlah inference step, Guidance scale. Hasil Generate akan muncul disampingnya.

Input Prompt

Deskripsi Karakter Utama

A brave young girl hijab with a red dress

Prompt Adegan 1

Standing on a hill, looking at a magical city in the distance

Prompt Adegan 2

Walking through the city, discovering a hidden portal

Prompt Adegan 3

Stepping into another dimension, greeted by a mysterious being

Negative Prompt

realistic, photo, low quality, blurry

Inference Steps



Guidance Scale



Seed (Untuk Konsistensi Karakter)

42

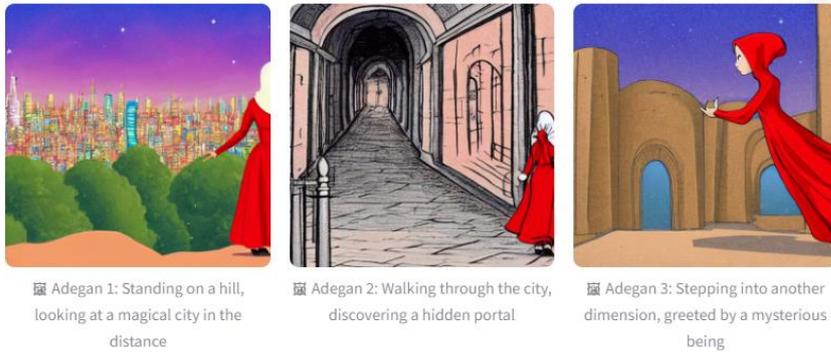
- +

Generate Story Images

Gambar 4. 16 Bagian *Input Prompt*

Gambar diatas adalah tempat measukan *prompt* , deskripsi karakter utama, adegan 1 ,adegan 2, dan adegan 3 sebagai alur cerita.

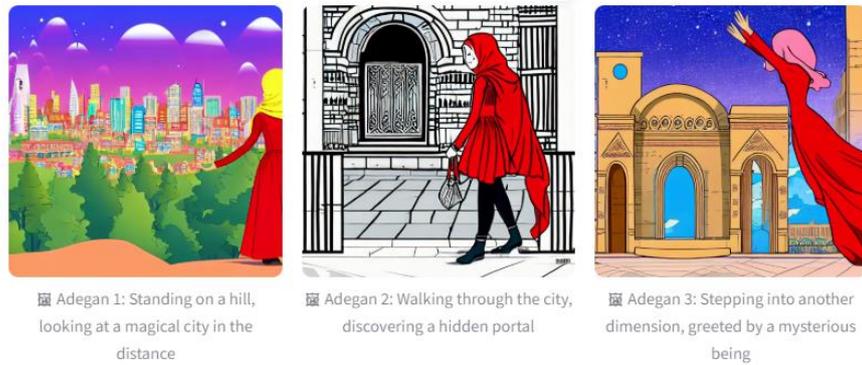
Hasil Generate



Gambar 4. 17 Hasil gambar di sistem user

Gambar diatas adalah gambar *hasil generate* yang digunakan pada sistem *user* berbentuk 3 panel. 3 adegan tanpa ditambah dengan *negative prompt*.

Hasil Generate



Gambar 4. 18 hasil gambar di sistem user (2)

Gambar diatas adalah gambar *hasil generate* yang digunakan pada sistem *user* berbentuk 3 panel. 3 adegan tanpa ditambah dengan *negative prompt*.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini menunjukkan bahwa gambar yang dihasilkan oleh model fine-tuned memiliki kualitas yang jelas dan sesuai dengan deskripsi teks. Hal ini dibuktikan melalui hasil uji metrik menggunakan CMMD, di mana skor perbandingan antara dataset uji dan hasil model fine-tuned mengalami penurunan, yang mengindikasikan peningkatan kesesuaian antara gambar dan deskripsi teks.

Penggunaan empat prompt yang terdiri dari deskripsi karakter utama, adegan 1, adegan 2, dan adegan 3 berhasil menghasilkan tiga gambar yang membentuk rangkaian panel komik. Ketiga gambar tersebut saling terhubung, menciptakan alur cerita yang konsisten dan sesuai dengan deskripsi teks yang dimasukkan. Meskipun demikian, masih terdapat kekurangan dalam mempertahankan konsistensi bentuk karakter utama di setiap panel, yang menjadi area potensial untuk pengembangan lebih lanjut.

5.2 Saran

Untuk penelitian selanjutnya, disarankan untuk berfokus pada mempertahankan konsistensi karakter utama dalam hasil gambar panel. Hal ini bertujuan agar setiap panel dalam komik memiliki kesinambungan visual yang kuat, sehingga pembaca dapat dengan mudah mengenali karakter utama dan mengikuti alur cerita dengan lebih baik. Selain itu, optimasi lebih lanjut juga diperlukan untuk menambahkan teks yang mendukung kelengkapan cerita, seperti teks percakapan dan narasi. Penambahan teks ini diharapkan dapat meningkatkan kedalaman cerita dan memperkuat koneksi emosional antara pembaca dan karakter, sehingga pengalaman membaca menjadi lebih imersif dan menyenangkan.

DAFTAR PUSTAKA

- Agnese, Jorge, Jonathan Herrera, Haicheng Tao, dan Xingquan Zhu. 2020. "A survey and taxonomy of adversarial neural networks for text-to-image synthesis." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10(4).
- Amanatullah. 2023. "Fine-Tuning the Model: What, Why, and How." *Medium*. <https://medium.com/@amanatulla1606/fine-tuning-the-model-what-why-and-how-e7fa52bc8ddf>.
- Beltr, Natalia, Arianna Michelangelo, Hannes Mueller, dan Cerquides Bueno. 2024. "Master ' s Degree in Data Science for Decision Making How to generate new versions of an original character ? An application of LoRA and DreamBooth Fine-Tuning of Diffusion Models " .” (July).
- Cetinic, Eva, dan James She. 2022. "Understanding and Creating Art with AI: Review and Outlook." *ACM Transactions on Multimedia Computing, Communications and Applications* 18(2).
- Chen, Qing, Shixiong Cao, Jiazhe Wang, dan Nan Cao. 2024. "How Does Automation Shape the Process of Narrative Visualization: A Survey of Tools." *IEEE Transactions on Visualization and Computer Graphics* 30(8): 4429–48.
- Chen, Shoufa, Peize Sun, Yibing Song, dan Ping Luo. 2023. "DiffusionDet: Diffusion Model for Object Detection." *Proceedings of the IEEE International Conference on Computer Vision: 19773–86*.
- Ding, Ming dkk. 2021. "CogView: Mastering Text-to-Image Generation via Transformers." *Advances in Neural Information Processing Systems* 24(NeurIPS): 19822–35.
- Fiveable. 2024. "Text-to-image generation." <https://library.fiveable.me/keywords/art-and-artificial-intelligence/text-to-image-generation>.
- Frolov, Stanislav dkk. 2021. "Adversarial text-to-image synthesis: A review." *Neural Networks* 144: 187–209.
- Gu, Jindong dkk. 2023. "A Systematic Survey of Prompt Engineering on Vision-Language Foundation Models." : 1–21. <http://arxiv.org/abs/2307.12980>.
- Hedlin, Eric dkk. 2023. "Unsupervised Semantic Correspondence Using Stable Diffusion." *Advances in Neural Information Processing Systems* 36(NeurIPS): 1–14.
- Hidayat Agusvian, Asep Sopian, dan Nunung Nursyamsiah. 2021. "Development of Comic in Qiroah Learning Introduction Material for VII Grade at Mts Muallimin NW Pancor / Pengembangan Media Komik Pada Pembelajaran Qiroah Materi Perkenalan Kelas VII Mts Muallimin NW Pancor." *al*

Mahāra: Jurnal Pendidikan Bahasa Arab 7(1): 45–63.

- J. Rafid Siddiqui, PhD. “Diffusion Models Made Easy.” *Data Science*.
<https://towardsdatascience.com/diffusion-models-made-easy-8414298ce4da/#bypass>.
- Jayasumana, Sadeep dkk. 2023. “Rethinking FID: Towards a Better Evaluation Metric for Image Generation.” : 9307–15. <http://arxiv.org/abs/2401.09603>.
- Jin, Xiaolong dkk. 2024. “Conditional LoRA Parameter Generation.” : 1–18.
<http://arxiv.org/abs/2408.01415>.
- Kandwal, Siddharth, dan Vibha Nehra. 2024. “A Survey of Text-to-Image Diffusion Models in Generative AI.” *Proceedings of the 14th International Conference on Cloud Computing, Data Science and Engineering, Confluence 2024* 14(8): 73–78.
- Li, Yuhang dkk. 2024. “A Simple Background Augmentation Method for Object Detection with Diffusion Model.” <http://arxiv.org/abs/2408.00350>.
- Modeling, Diffusion, dan Transformer Knowledge. 2025. “Diffusion Modeling and Transformer Knowledge ’ LIIXVLRQ 0RGHOLQJ DQG 7UDQVIRUPHU . QRZOHGJH Distillation Techniques for Time Series Imputation ’ LVWLOODWLRQ 7HFKQLTXHV IRU 7LPH 6HULHV , PSXWDWLRQ.” : 3103–8.
- Park, NaHyeon, Kunhee Kim, dan Hyunjung Shim. 2024. “TextBoost: Towards One-Shot Personalization of Text-to-Image Models via Fine-tuning Text Encoder.” <http://arxiv.org/abs/2409.08248>.
- Prasad, Aravindra, K. Rohitaksha, C. B. Abhilash, dan Shashank Dhananjaya. 2024. “Stable Diffusion Image Processing.” *Library Progress International* 44(3): 5917–25.
- Pratama, Sandi Destian, Bayu Bambang Perdana, dan Sellya Shafariya. 2022. “Analisis Tata Letak Panel Komik Cetak Dan Komik Online.” *Wacadesain* 3(1): 38–47.
- Prayoga, Dwiki Setya. 2021. “Teknik Membuat Komik Strip Digital.” *Jurnal Desain Komunikasi Visual Asia* 4(2): 87–97.
- Putri, Martha Tisna Ginanjar, dan Iis Purnengsih. 2023. “Gaya Visual Nusantara dalam Media Cover Komik Sangkuriang Karya R. A. Kosasih.” *Jurnal Desain* 10(3): 605.
- Putro, Dimas. 2021. “Perkembangan Tren Membaca Komik Pada Era Digital Di Indonesia.” *SOURCE : Jurnal Ilmu Komunikasi* 7(2): 115.
- Ruiz, Nataniel dkk. 2023. “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation.” : 22500–510.
- Saharia, Chitwan dkk. 2022. “Photorealistic Text-to-Image Diffusion Models with

Deep Language Understanding.” *Advances in Neural Information Processing Systems* 35(NeurIPS).

Sentosa, Debora Audrey. 2023. “Perancangan Komik Digital Scrambled Egg sebagai Media Edukasi Keluarga Single Parent.” *Citradirga - Jurnal Desain Komunikasi Visual dan Intermedia* 3(02): 33–41.

Siddique, Nahian, Sidike Paheding, Colin P. Elkin, dan Vijay Devabhaktuni. 2021. “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications.” *IEEE Access* 9(January 2021): 82031–57.

Wei, Ruoqi dkk. 2020. “Variations in Variational Autoencoders - A Comparative Evaluation.” *IEEE Access* 8: 153651–70.

Xiaojun, Li dkk. 2021. “Study on the Application Fields and Development Prospects of Artificial Intelligence.” *Proceedings - 2021 2nd International Conference on Artificial Intelligence and Education, ICAIE 2021*: 101–5.

Yang, Tao dkk. 2023. “Pixel-Aware Stable Diffusion for Realistic Image Super-resolution and Personalized Stylization.” : 1–24.
<http://arxiv.org/abs/2308.14469>.

Yin, Xiao Xia dkk. 2022. “U-Net-Based Medical Image Segmentation.” *Journal of Healthcare Engineering* 2022.

