IMPLEMENTASI LORA UNTUK MEMBUAT GAMBAR TOKOH CERITA ANAK SI KANCIL DAN BUAYA DENGAN *STABLE DIFFUSION*

LAPORAN TUGAS AKHIR

Laporan ini Disusun untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang

DI SUSUN OLEH :
MUHAMMAD SABIL HUDA
32602100087

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

FINAL PROJECT

IMPLEMENTATION OF LORA TO CREATE PICTURES OF CHILDREN'S STORY CHARACTERS SI KANCIL DAN BUAYA USING STABLE DIFFUSION

Proposed to complete the requirement to obtain a bachelor's degree (S-1) at Informatics Engineering Departement of Industrial Technology Faculty Sultan Agung Islamic University

ARRANGED BY :
MUHAMMAD SABIL HUDA
32602100087

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

2025

LEMBAR PENGESAHAN
TUGAS AKHIR

IMPLEMENTASI LORA UNTUK MEMBUAT GAMBAR TOKOH CERITA ANAK SI KANCIL DAN BUAYA DENGAN *STABLE DIFFUSION*

MUHAMMAD SABIL HUDA
32602100087

Telah dipertahankan di depan tim penguji sidang tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : Selasa, 25 Februari 2025

TIM PENGUJI SIDANG TUGAS AKHIR:

Mustafa, ST.MM.M.Kom NIDN. 0623117703 (Penguji 1)	.....	13-03-2025
Badie'ah, S.T. M.Kom NIDN. 0619018701 (Penguji 2)	.....	13-03-2025
Ir. Sri Mulyono, M.Eng NIDN. 0626066601 (Pembimbing)	.....	13-03-2025

Semarang, 25 Februari 2025
Mengetahui,
Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch Taufiq MIT
NIDN. 062203750

NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Muhammad Sabil Huda

NIM : 32602100087

Judul Tugas Akhir : IMPLEMENTASI LORA UNTUK MEMBUAT GAMBAR TOKOH CERITA ANAK SI KANCIL DAN BUAYA DENGAN *STABLE DIFFUSION*

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 apbila 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, 25 Februari 2025

Yang Menyatakan,

Muhammad Sabil Huda

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Sabil Huda

NIM : 32602100087

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal :

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : IMPLEMENTASI LORA UNTUK MEMBUAT GAMBAR TOKOH CERITA ANAK SI KANCIL DAN BUAYA DENGAN *STABLE DIFFUSION.* 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, 25 Februari 2025

Muhammad Sabil Huda

KATA PENGANTAR

Dengan mengucap syukur alhamdulillah atas kehadirat Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul "Implementasi LoRA Untuk Membuat Gambar Tokoh Cerita Anak Si Kancil dan Buaya Dengan *Stable Diffusion*" ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang.

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, S.T., M.T.
3. Dosen pembimbing Ir. Sri Mulyono M. Eng yang telah meluangkan waktu dan memberi ilmu.
4. Orang tua penulis yang telah mengizinkan untuk menyelesaikan laporan ini,
5. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan segala kerendahan hati, penulis menyadari masih terdapat banyak kekurangan dari segi kualitas atau kuantitas maupun dari ilmu pengetahuan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi kesempurnaan laporan ini dan masa mendatang.

Semarang, 25 Februari 2025

Muhammad Sabil Huda

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Kebutuhan visualisasi ilustrasi dalam cerita anak memerlukan pendekatan yang dapat menghasilkan gambar sesuai dengan gaya yang diinginkan. Teknik fine-tuning menggunakan LoRA (Low-Rank Adaptation) pada model Stable Diffusion memungkinkan adaptasi model terhadap dataset spesifik dengan efisiensi parameter yang lebih tinggi dibandingkan metode fine-tuning penuh. Dalam penelitian ini, LoRA diterapkan untuk menghasilkan ilustrasi tokoh cerita anak Si Kancil dan Buaya dengan gaya gambar yang khas, termasuk warna cerah, garis lembut, dan ekspresi karakter yang ramah anak. Sistem berbasis web dikembangkan untuk memungkinkan pengguna memasukkan deskripsi ciri fisik tokoh dan memperoleh ilustrasi tokoh Si Kancik dan Buaya yang sesuai. Hasil penelitian menunjukkan bahwa penerapan LoRA pada Stable Diffusion mampu menghasilkan gambar yang sesuai dengan karakteristik ilustrasi anak, meskipun ada tantangan dalam mempertahankan konsistensi karakter di berbagai tokoh. Penelitian ini berkontribusi pada pengembangan teknologi Text-to-Image Generation dalam industri kreatif, khususnya untuk ilustrasi cerita anak.

Kata Kunci: Stable Diffusion, LoRA Fine-Tuning, Ilustrasi Cerita Anak, Si Kancil dan Buaya, Text-to-Image Generation

ABSTRACT

The need for visualization of illustrations in children's stories requires an approach that can produce images in accordance with the desired style. Fine-tuning techniques using LoRA (Low-Rank Adaptation) on the Stable Diffusion model allow adaptation of the model to specific datasets with higher parameter efficiency than full fine-tuning methods. In this research, LoRA is applied to generate illustrations of the children's story characters The Deer and the Crocodile with a distinctive drawing style, including bright colors, soft lines, and child-friendly character expressions. A web-based system was developed to allow users to input descriptions of the characters' physical characteristics and obtain corresponding illustrations of the deer and crocodile characters. The results show that the application of LoRA to Stable Diffusion is able to produce images that match the characteristics of child illustrations, although there are challenges in maintaining character consistency across different characters. This research contributes to the development of Text-to-Image Generation technology in the creative industry, especially for children's story illustrations.

Keywords: *Stable Diffusion, LoRA Fine-Tuning, Children's Story Illustration, The Deer and the Crocodile, Text-to-Image Generation*

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Cerita anak merupakan bagian dari jenis sastra anak, menurut Saryono dalam Puryanto, sastra anak secara umum meliputi buku bergambar, cerita rakyat, baik berupa cerita binatang, dongeng, legenda, maupun mite, fiksi sejarah, fiksi realistik, fiksi ilmiah, cerita fantasi, dan biografi (Waryantis, 2021).

Buku anak tentunya memiliki kaitan yang erat dengan ilustrasi. Ilustrasi dalam buku anak merupakan salah satu faktor utama dalam penyampaian cerita. Dikutip dari AlHinaai, Waymack dan Hendrickson menemukan bahwa anak lebiih memilih buku dengan lebih dari dua puluh lima persen bagian berupa gambar. Pada umumnya, kebanyakan anak memahami kode piktorial jauh lebih mudah daripada memahami kode verbal karena kode pictorial lebih langsung. Menurut Paris dan Paris, gambar memiliki peran signifkan untukk anak mendapatkan ide dari sebuah teks dan memahaminya. Mereka menyebutkan bahwa anak melihat gambar pada buku terlebih dahulu sebelum membac teks, sehingga mereka mendapat ide utama dari teks melalui pandangan pertama mereka terhadap gambar. Ilustrasi pada buku anak dapat memudahkan anak dalam mencerna topik yang dibawakan buku tersebut, terutama jika topik yang dibawakan merupakan topik berat seperti stress dan depresi (Aqiela & Sihombing, 2023).

Ilustrasi pada buku anak berperan penting dalam menarik perhatian dan mempertahankan atensi tersebut selama membaca. Dikutip dari Al-Hinaai, gambar dapat meningkatkan kenikmatan dari teks dan peran utama dari gambar atau ilustrasi adalah untuk melengkapi buku. Ilustrasi yang menarik dalam buku anak dapat membuat anak bertahan membaca buku tersebut hingga selesai, sehingga keseluruhan cerita dapat tersampaikan sebagai satu informasi yang utuh (Aqiela & Sihombing, 2023).

Pentingnya peran ilustrasi dalam buku anak membuat Bahasa visual pada ilustrasi tersebut menjadi penting dalam penyampaian cerita. Menurut

Hladikova, Bahasa visual pada *picture book* terdiri atas hubungan antara teks dan gambar, pengembangan karakter, *layout*, alur cerita, warna dan gaya ilustrasi (Aqiela & Sihombing, 2023).

LoRA adalah teknik *fine-tuning* yang dikembangkan oleh *Google Research* untuk menghasilkan gambar-gambar yang spesifik menggunakan model berbasis difusi, seperti *Stable Diffusion*. Pengguna dapat membuat model yang menghasilkan gambar berdasarkan karakter, objek, atau gaya tertentu yang dipersonalisasi, dengan melatih model agar mampu "mempelajari" atribut visual unik dari beberapa contoh gambar *input*. Ini menjadikan model dapat menghasilkan gambar baru yang memiliki kemiripan atau fitur khusus dari objek atau karakter tersebut, yang sangat berguna untuk aplikasi seperti avatar, karakter virtual, atau bahkan pembuatan konsep desain (Park dkk., 2023).

Stable Diffusion adalah salah satu implementasi paling sukses dari Diffusion Models. Model ini dirancang untuk menghasilkan gambar yang sangat detail dan realistis berdasarkan teks deskripsi (Ruiz dkk., 2023). Agar lebih ringan dari sisi pemakaian sumber daya, membuatnya lebih cepat di-render pada hardware dengan spesifikasi lebih rendah, serta fokus pada elemen yang sesuai untuk buku anak-anak, seperti gaya ilustrasi yang sederhana dan ramah, maka digunakan *Stable Diffusion v2.1*(Song, 2023).

1.2 Perumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana *diffusion models* dan *LoRA Fine Tuning* dapat digunakan untuk menghasilkan ilustrasi tokoh cerita anak yang berkualitas tinggi dan sesuai dengan narasi teks yang diberikan?

1.3 Pembatasan Masalah

Batasan masalah ini bertujuan untuk memudahkan dan menghindari adanya kegiatan di luar sasaran, sehingga dalam pembuatan laporan ini

perlu ditentukan suatu batasan masalah. Batasan masalah tersebut sebagai berikut :

a. Penelitian ini hanya fokus pada Cerita Anak si Kancil dan Buaya.
b. Penelitian ini hanya fokus pada ilustrasi gambar, tidak termasuk dengan *generate* teks cerita.
c. Penelitian ini hanya menghasilkan beberapa gambar ilustrasi Cerita Anak si Kancil dan Buaya, bukan seluruh buku.
d. *Prompt* yang dimasukkan dalam bahasa inggris dan bukan berasal dari teks cerita maupun sinopsis cerita anak, melainkan *prompt* teks deskripsi gambar yang ingin digenerate.

1.4 Tujuan

Adapun tujuan dari penelitian ini adalah membangun sistem yang dapat menghasilkan ilustrasi tokoh Cerita Anak Si Kancil dan Buaya dengan narasi teks yang diberikan menggunakan *Fine Tuning LoRa* pada *Stable Diffusion V2.1*.

1.5 Manfaat

Tugas akhir ini diharapkan dapat memberikan kontribusi signifikan dalam pemahaman tentang pembuatan ilustrasi tokoh buku anak-anak menggunakan teknologi modern. Serta memberikan kemudahan bagi kreator dalam menghasilkan visualisasi karakter tokoh cerita anak berdasarkan deskripsi teks, akses yang terjangkau untuk menciptakan ilustrasi berbasis AI melalui sistem berbasis web, dengan dukungan personalisasi melalui *fine-tuning* menggunakan LORA.

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 "Implementasi Lora Untuk Membuat Gambar Ilustrasi Tokoh Cerita Anak Si Kancil Dan Buaya Dengan *Stable Diffusion*". 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 dalam memahami permasalahan yang dipilih.

BAB 3 : METODE PENELITIAN

Bab ini menjelaskan proses tahapan- tahapan penelitian dimulai dari analisa kebutuhan sistem, kemudian perancangan sistem hingga selesai dibuat.

BAB 4: HASIL PENELITIAN DAN IMPLEMENTASI SISTEM

Bab ini menjelaskan mengungkapkan hasil penelitian yang berupa pembuatan gambar karakter dari cerita anak berdasarkan deskripsi teks menggunakan *Diffusion Models*.

BAB 5: KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan dari keseluruhan uraian bab-bab sebelumnya dan saran-saran dari hasil yang diperoleh dan diharapkan dapat bermanfaat dalam penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Perkembangan teknologi di era revolusi industri 4.0 sangat pesat. Mendeteksi gambar palsu menjadi tujuan utama dari visi komuter. Kebutuhan ini menjadi semakin mendesak dengan peningkatan berkelanjutan dari metode sintesis berdasarkan *Generative Adversarial Networks (GAN)*, dan bahkan lebih lagi densgan munculnya metode yang kuat berdasarkan *Diffusion Models (DM)*. Untuk mencapai tujuan ini untuk mendapatkan wawasan tentang fitur gambar mana yang dapat membedakan gambar palsu dengan lebih baik dari gambar asli. Dalam makalah ini Corvi melaporkan studi sistematis nya terhadap sejumlah besar generator gambar dari keluarga yang berbeda, yang bertujuan untuk menemukan karakteristik yang paling relevan secara forensik dari gambar asli dan gambar yang dihasilkan. Eksperimen Corvi memberikan sejumlah hal pengamatan dan menjelaskan beberapa sifat menarik dari gambar sintetis tidak hanya model GAN, tetapi juga model DM dan VQ-GAN (*Vector Quantized Generative Adver- sarial Networks*) memunculkan artefak yang terlihat dalam domain Fourier dan menunjukkan pola reguler yang anomali dalam autokorelasi; ketika set data yang digunakan untuk melatih model tidak memiliki variasi yang cukup, biasnya dapat ditransfer ke gambar yang dihasilkan; gambar sintetis dan gambar asli menunjukkan perbedaan yang signifikan pada konten frekuensi menengah-tinggi yang signifikan, yang dapat diamati pada distribusi daya spektral radial dan sudut (Corvis, n.d.).

Sejak munculnya GAN dan VAE, pembuatan gambar terus berkembang, membuka berbagai aplikasi dunia nyata dengan diperkenalkannya model Stable Diffusion dan DALL-E. Model teks-ke-gambar ini dapat menghasilkan gambar berkualitas tinggi untuk bidang-bidang seperti seni, desain, dan periklanan. Namun, model ini sering menghasilkan gambar yang menyimpang untuk permintaan tertentu. Studi ini mengusulkan sebuah metode untuk mengurangi masalah tersebut dengan menyempurnakan model

Stable Diffusion 3 menggunakan teknik *DreamBooth*. Hasil eksperimen yang menargetkan prompt "berbaring di rumput/jalan" menunjukkan bahwa model yang telah disempurnakan menunjukkan peningkatan kinerja dalam evaluasi visual dan metrik seperti *Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), dan Frechet Inception Distance (FID)*. Survei pengguna juga menunjukkan preferensi yang lebih tinggi untuk model yang disetel model. Penelitian ini diharapkan dapat memberikan kontribusi untuk meningkatkan kepraktisan dan keandalan model teks-ke-gambar (Yoo, 2024).

Kemajuan yang signifikan dalam penerapan model pembuatan teks-ke-gambar model generasi teks-ke-gambar telah mendorong permintaan algoritma adaptasi posthoc yang secara efisien dapat menghapus konsep yang tidak diinginkan (misalnya privasi, hak cipta, dan keamanan) dari model yang telah dilatih sebelumnya dengan pengaruh minimal terhadap sistem pengetahuan yang telah ada yang dipelajari dari pelatihan sebelumnya. Yang sudah ada metode yang ada terutama menggunakan penyesuaian yang tidak diinginkan secara eksplisit konsep yang tidak diinginkan menjadi beberapa alternatif seperti hipernim atau antonimnya. Pada dasarnya, mereka memodifikasi sistem pengetahuan dari model yang telah dilatih dengan mengganti konsep yang tidak diinginkan yang tidak diinginkan menjadi sesuatu yang didefinisikan secara sewenang-wenang oleh pengguna. Lebih jauh lagi, metode-metode ini membutuhkan ratusan langkah optimasi, karena karena mereka hanya mengandalkan *denoising loss* yang digunakan untuk pretraining. Untuk mengatasi tantangan di atas, Zhang mengusulkan *Forget-Me-Not*, solusi yang berpusat pada model dan efisien yang dirancang untuk menghapus identitas, objek, atau gaya dari model teks-ke-gambar yang dikonfigurasi dengan baik hanya dalam waktu 30 detik, tanpa secara signifikan mengganggu kemampuannya untuk menghasilkan konten lain. Berbeda dengan metode yang sudah ada, Zhang memperkenalkan pengarahan ulang perhatian untuk mengarahkan kembali pembuatan model dari konsep yang tidak diinginkan yang dipelajari selama prapelatihan, bukannya menjadi ditentukan oleh pengguna. Selain itu, metode Zhang menawarkan dua

ekstensi praktis penghapusan konten yang berpotensi berbahaya atau NSFW, dan peningkatan akurasi, inklusi, dan keragasman melalui koreksi dan pemisahan konsep (Zhang, 2023).

Irene Rante Lembang pada tahun 2022 melakukan penelitian tentang teknik ilustrasi *digital freehand* dalam pembuatan buku cerita bergambar friends untuk anak usia dini dengan teknik digital, dibuat melalui perangkat digital seperti komputer, laptop, tablet dan lainnya dengan menggunakan software atau perangkat lunak berupa design tools. Berikut beberapa aplikasi desain yang biasa digunakan untuk membuat ilustrasi, yaitu Adobe Creative Suite (Illustrator, Photoshop, Fresco), Corel Draw, Autodesk Sketchbook, Medibang Paint, Procreate, Vectornator, Painter, iBis Paint, iArtbook, dan lain-lain. Dalam pembuatan ilustrasi dengan teknik digital, terdapat dua macam gambar yang dihasilkan, yaitu vektor dan bitmap. Vektor terdiri dari hasil kurva, garis, dan bidang serta memiliki unsur fill dan stroke. Bitmap dibentuk dari sekumpulan titik yang biasa disebut pixel dan memiliki pengaruh terhadap resolusi. Dalam pembuatan buku cerita bergambar untuk anak, *freehand* menjadi teknik yang cocok karena dapat memberikan kesan yang santai dan *fluffy* atau dapat dibilang tidak kaku. Selain memudahkan *illustrator* dalam membuat karya ilustrasi dengan jumlah yang banyak, ilustrasi dengan teknik digital *freehand* juga dapat membuat ilustrasi terlihat lebih menarik, luwes, adaptif, menyenangkan dan ramah terhadap anak-anak (Lembang, 2022).

Dengan tren yang muncul dalam model generatif dan akses publik yang mudah ke model difusi yang telah dilatih sebelumnya pada kumpulan data besar, pengguna dapat menyempurnakan model-model ini untuk menghasilkan gambar wajah pribadi atau dalam konteks baru yang dijelaskan oleh bahasa alami. Penyempurnaan parameter yang efisien (PEFT) seperti *Dreambooth* telah menjadi cara yang paling umum untuk menghemat penggunaan memori dan komputasi di sisi pengguna selama penyempurnaan. Namun, pertanyaan yang wajar adalah apakah gambar pribadi yang digunakan untuk *fine-tuning* akan bocor ke lawan saat berbagi bobot model.

Dalam makalah ini, Yao mempelajari masalah kebocoran privasi dari model difusi yang disetel dengan baik dalam pengaturan praktis, di mana musuh hanya memiliki akses ke bobot model, daripada petunjuk atau gambar yang digunakan untuk penyempurnaan. Yao merancang dan membangun variasi autoencoder jaringan yang mengambil bobot model sebagai masukan dan menghasilkan rekonstruksi gambar pribadi (Yao, 2024).

Model teks-ke-gambar yang besar mencapai lompatan yang luar biasa dalam evolusi AI, memungkinkan kualitas tinggi dan beragam sintesis gambar dari perintah teks yang diberikan. Namun demikian, model-model ini tidak memiliki kemampuan untuk meniru penampilan subjek dalam set referensi yang diberikan dan mensintesis rendisi baru dari mereka dalam konteks yang berbeda. Dalam karya ini, Ruiz menyajikan pendekatan baru untuk personalisasi model difusi teks-ke-gambar. Dengan hanya memberikan beberapa gambar dari sebuah subjek sebagai masukan, Ruiz menyempurnakan model teks-ke-gambar yang telah dilatih sedemikian rupa sehingga model tersebut belajar untuk mengikat pengenal unik dengan subjek tertentu. Setelah subjek disematkan dalam domain keluaran model, pengidentifikasi unik dapat digunakan untuk mensintesis gambar fotorealistik baru dari subjek yang dikontekstualisasikan dalam adegan yang berbeda. Dengan memanfaatkan *prior semantik* yang tertanam dalam model dengan prior spesifik kelas *autogenous* baru kehilangan pelestarian, teknik nya memungkinkan sintesis subjek dalam beragam adegan, pose, pandangan, dan kondisi pencahayaan yang tidak muncul dalam gambar referensi. Ruiz menerapkan teknik nya pada beberapa tugas yang sebelumnya tidak dapat dilakukan, termasuk rekontekstualisasi subjek, sintesis tampilan yang dipandu teks, dan rendering artistik, sambil mempertahankan fitur utama subjek. Ruiz juga menyediakan kumpulan data dan protokol evaluasi baru untuk tugas baru pembuatan berdasarkan subjek ini (Ruiz, 2023).

Model difusi teks-ke-gambar sangat unggul dalam menghasilkan gambar yang beragam, berkualitas tinggi, dan foto yang realistis. Kemajuan ini telah memacu minat yang semakin besar dalam memasukkan identitas spesifik ke

dalam konten yang dihasilkan. Sebagian besar metode saat ini menggunakan pendekatan inversi untuk menyematkan konsep visual target ke dalam ruang penyematan teks menggunakan satu gambar referensi. Namun, wajah yang baru disintesis sangat mirip dengan gambar referensi dalam hal atribut wajah, seperti ekspresi, atau menunjukkan mengurangi kapasitas untuk mempertahankan identitas. Deskripsi teks dimaksudkan untuk memandu atribut wajah yang disintesis wajah mungkin gagal, karena keterikatan yang rumit dari informasi identitas dengan atribut wajah yang tidak relevan dengan identitas yang berasal dari gambar referensi. Untuk mengatasi masalah ini, Li menyajikan penggunaan baru dari ruang penyematan *Style* GAN yang diperluas W+, untuk mencapai pelestarian identitas yang ditingkatkan dan disentanglement untuk model difusi. Dengan menyelaraskan ini ruang laten wajah manusia yang bermakna secara semantik dengan model difusi teks-ke-gambar, Li berhasil mempertahankan tinggi dalam pelestarian identitas, ditambah dengan kapasitas untuk pengeditan semantik. Selain itu, Li mengusulkan tujuan pelatihan baru untuk menyeimbangkan pengaruh dari kondisi kondisi identitas, memastikan bahwa latar belakang yang tidak relevan dengan identitas tetap terpengaruh secara signifikan selama modifikasi atribut wajah modifikasi atribut wajah. Eksperimen ekstensif mengungkapkan bahwa metode Li dengan mahir menghasilkan output teks-ke-gambar yang dipersonalisasi tidak hanya kompatibel dengan deskripsi yang cepat tetapi juga dapat menerima arahan pengeditan *Style* GAN yang umum dalam beragam pengaturan (Li, 2023).

Model difusi *denoising* mewakili topik yang muncul baru-baru ini dalam visi komputer, menunjukkan hasil yang luar biasa dalam bidang pemodelan generatif. Model difusi adalah model generatif dalam yang didasarkan pada dua tahap, tahap difusi maju dan tahap difusi balik. Pada tahap difusi maju, data input secara bertahap diganggu selama beberapa langkah dengan menambahkan *Gaussian noise*. Pada tahap mundur, model ditugaskan untuk memulihkan data input asli dengan belajar untuk secara bertahap membalikkan proses difusi secara bertahap, selangkah demi selangkah.

Model difusi dihargai secara luas untuk kualitas dan keragaman sampel yang dihasilkan, meskipun beban komputasi yang diketahui, yaitu kecepatan rendah karena banyaknya langkah yang terlibat selama pengambilan sampel. Dalam survei ini, Croitoru menyediakan tinjauan komprehensif artikel tentang model difusi denoising yang diterapkan dalam penglihatan, yang terdiri dari teoritis dan praktis kontribusi teoretis dan praktis di lapangan. Pertama, Croitoru mengidentifikasi dan menyajikan tiga kerangka kerja pemodelan difusi generik, yang didasarkan pada denoising model probabilistik difusi, jaringan skor yang dikondisikan dengan noise, dan persamaan diferensial stokastik. Croitoru membahas lebih lanjut hubungan antara model difusi dan model generatif dalam lainnya, termasuk penyandi otomatis variasional, jaringan permusuhan generatif, model berbasis energi, model *autoregresif*, dan aliran normalisasi. Kemudian, Croitoru memperkenalkan kategorisasi multi-perspektif difusi model yang diterapkan dalam visi komputer. Akhirnya, Croitoru menggambarkan keterbatasan model difusi saat ini dan membayangkan beberapa hal yang menarik menarik untuk penelitian di masa depan (Croitoru, 2023).

2.2 Dasar Teori

2.2.1 *Text to Image*

Sintesis *text-to-image* 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. Pada penelitian sebelumnya, sintesis gambar mengandalkan analisis korelasi kata dengan gambar yang dikombinasikan dengan metode yang diawasi untuk menemukan keselarasan terbaik dari konten visual yang sesuai dengan teks. Kemajuan terbaru dalam *deep learning (DL)* telah membawa seperangkat metode pembelajaran mendalam tanpa pengawasan, khususnya model genetika mendalam yang mampu menghasilkan gambar visual yang realistis dengan menggunakan model jaringan saraf yang terlatih dengan baik. Perubahan arah dari pendekatan berbasis visi komputer ke

metode yang digerakkan oleh kecerdasan buatan (AI) memicu minat yang kuat dalam industri, seperti realitas virtual, permainan rekreasi & profesional (*eSports*), dan desain berbantuan komputer, untuk secara otomatis menghasilkan gambar yang menarik dari deskripsi bahasa alami berbasis teks (Agnese, 2019).

Sistem ini menggunakan korelasi antara kata kunci atau *keyphrase* dan gambar serta mengidentifikasi unit tesks yang informatif dan dapat digambarkan, kemudian mencari bagian gambar yang paling mungkin dikondisikan pada teks, dan pada akhirnya mengoptimalkan tata letak gambar yang dikondisikan pada bagian teks dan gambar (Agnese, 2019).

Proses pembelajaran yang diawasi bertujuan untuk mempelajari model generatif berlapis untuk menghasilkan konten visual. Karena pembelajaran disesuaikan/dikondisikan oleh atribut yang diberikan, *model generatif Attribute2Image* dapat menghasilkan gambar dengan atribut yang berbeda, seperti warna rambut, usia (Agnese, 2019).

Sintesis teks-ke-gambar berbasis GAN menggabungkan pembelajaran diskriminatif dan generatif untuk melatih jaringan saraf yang menghasilkan gambar yang dihasilkan secara semantik mirip dengan sampel pelatihan atau disesuaikan dengan subset gambar pelatihan (yaitu *output* yang dikondisikan). $\phi()$ adalah fungsi penyematan fitur, yang mengubah teks sebagai vektor fitur. z adalah vektor laten yang mengikuti distribusi normal dengan rata-rata nol. $\hat{x} = G(z, \phi(t)$ menunjukkan gambar sintetis yang dihasilkan dari generator, menggunakan vektor laten z dan fitur teks $\phi(t)$ sebagai *input*. $D(\hat{x}, \phi(t))$ menyatakan prediksi diskriminator berdasarkan input $\hat{x}$ gambar yang dihasilkan dan informasi teks $\phi(t)$ dari gambar yang dihasilkan (Agnese, 2019).

2.2.2 *Stable Diffusion*

Stable Diffusion adalah salah satu implementasi paling sukses dari Diffusion Models. Model ini dirancang untuk menghasilkan gambar yang sangat detail dan realistis berdasarkan teks deskripsi (Ruiz dkk., 2023). Agar

lebih ringan dari sisi pemakaian sumber daya, membuatnya lebih cepat di-render pada hardware dengan spesifikasi lebih rendah, serta fokus pada elemen yang sesuai untuk buku anak-anak, seperti gaya ilustrasi yang sederhana dan ramah, maka digunakan *Stable Diffusion v2.1*(Song, 2023).

Gaussian Blur: Yang menarik, *Gaussian blur* tampaknya memiliki pengaruh positif pada kinerja RECCE kinerja RECCE, terutama untuk metode *Stable Diffusion v1.5*, yang mengalami peningkatan substansial akurasi. Hal ini mungkin mengisyaratkan karakteristik inheren tertentu dari deepfake ini menjadi lebih jelas atau terdeteksi dengan blur (Song, 2023).

2.2.3 *Fine Tuning*

Fine-tuning adalah proses penyesuaian model yang telah dilatih sebelumnya agar lebih cocok untuk tugas spesifik dengan menggunakan *Dataset* yang lebih relevan. Salah satu metode yang digunakan adalah *fine-tuning*, di mana model generatif diadaptasi secara khusus menggunakan set data kecil yang terkait erat dengan subjek tertentu (Ruiz, 2023). Misalnya, dalam model LoRA, *model diffusion fine-tuned* dengan data gambar dari subjek tertentu, memungkinkan pembuatan gambar karakter spesifik dari deskripsi teks dengan akurasi tinggi.

LoRA adalah teknik *fine-tuning* yang dikembangkan oleh *Google Research* untuk menghasilkan gambar-gambar yang spesifik menggunakan model berbasis difusi, seperti *Stable Diffusion*. Pengguna dapat membuat model yang menghasilkan gambar berdasarkan karakter, objek, atau gaya tertentu yang dipersonalisasi, dengan melatih model agar mampu "mempelajari" atribut visual unik dari beberapa contoh gambar *input*. Ini menjadikan model dapat menghasilkan gambar baru yang memiliki kemiripan atau fitur khusus dari objek atau karakter tersebut, yang sangat berguna untuk aplikasi seperti avatar, karakter virtual, atau bahkan pembuatan konsep desain (Park dkk., 2023).

Penyempurnaan parameter yang efisien (PEFT) seperti *Low Rank Adaptation (LoRA)* telah menjadi cara yang paling umum untuk menghemat penggunaan memori dan komputasi di sisi pengguna selama penyempurnaan.

Namun, pertanyaan yang wajar adalah apakah gambar pribadi yang digunakan untuk fine-tuning akan bocor ke lawan saat berbagi bobot model. Dalam makalah ini, kami mempelajari masalah kebocoran privasi dari model difusi yang disetel dengan baik dalam pengaturan praktis, di mana musuh hanya memiliki akses ke bobot model, daripada petunjuk atau gambar yang digunakan untuk penyempurnaan. Kami merancang dan membangun variasi autoencoder jaringan yang mengambil bobot model sebagai masukan dan menghasilkan rekonstruksi gambar pribadi (Yao, 2024).

Gambar 2. 1 Flowchart fine tuning LoRA

Pelatihan dalam *Dreambooth* dimulai dengan mengumpulkan gambar-gambar dari objek yang diinginkan. Model dasar, seperti *Stable Diffusion*, kemudian di-*fine-tune* dengan gambar-gambar ini, untuk menjaga agar model tetap mempertahankan kemampuan umum yang ada dalam model awalnya sambil belajar fitur spesifik dari objek yang ingin direpresentasikan. *Dreambooth* mempertahankan kemampuan untuk menghasilkan gambar dengan resolusi tinggi dan detail yang tajam, namun tetap menjaga konteks visual yang mungkin berbeda dari gambar asli. Keunggulan *Dreambooth* terletak pada fleksibilitas dan kemampuannya untuk mempelajari detail yang presisi dari data pelatihan terbatas, memungkinkan kreasi visual yang tak terbatas dan unik untuk setiap pengguna (Luo, 2023).

Dreambooth lebih cocok untuk personalisasi objek atau karakter baru, sementara LoRA lebih sering digunakan untuk modifikasi gaya atau perubahan kecil dalam model. LoRA lebih cocok untuk skenario multi-adaptasi, di mana ingin melatih beberapa konsep secara modular, tetapi jika hanya satu konsep spesifik yang ingin dipelajari secara mendalam, *Dreambooth* memberikan hasil yang lebih (Luo, 2023).

2.2.4 *Autoencoders* (AEs)

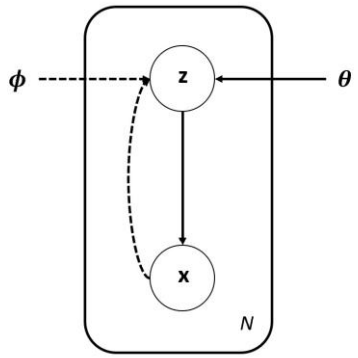
Autoencoders (AEs) adalah jenis jaringan saraf yang dirancang untuk melakukan kompresi data, seperti gambar, dengan cara mengonversi *input* menjadi representasi laten yang lebih sederhana dan terkompresi, kemudian mendekodekannya kembali untuk menghasilkan *output* yang sedekat mungkin dengan *input* asli (Chen & Guo, 2023). Proses ini dimulai dengan *encoder*, yang berfungsi untuk mengubah *input* menjadi vektor laten berdimensi lebih rendah menggunakan lapisan konvolusi dan *pooling*, seperti yang ditunjukkan pada gambar 2. 6. Setelah itu, *decoder* berfungsi untuk melakukan operasi seperti dekonvolusi dan upsampling guna merekonstruksi *input* dari vektor laten yang telah dikompresi. Keduanya dikendalikan oleh fungsi kehilangan *Mean Squared Error* (MSE), yang mengukur perbedaan antara *input* asli dan hasil rekonstruksi pada tingkat piksel.

Gambar 2. 2 Arsitektur Auto-*encoder* (Berahmand dkk., 2024)

2.2.5 *Variational Autoencoders* (VAE)

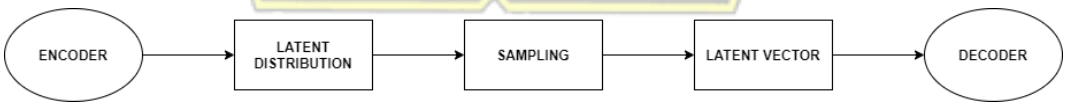
Peningkatan besar dalam kemampuan representasi autoencoder telah dicapai oleh model *Variational Autoencoders (VAE)*. Mengikuti *Variational Bayes (VB) Inference*, VAE adalah model generatif yang mencoba menggambarkan pembangkitan data melalui distribusi probabilistik. Secara khusus, diberikan dataset yang diamati X = {xi} iN = 1 dari sampel V i.i.d, Bank mengasumsikan model generatif untuk setiap datum xi yang dikondisikan pada variabel laten acak yang tidak teramati xi, di mana θ adalah parameter yang mengatur distribusi generatif. Model generatif ini juga setara dengan dekoder probabilistik. Secara simetris, Bank mengasumsikan perkiraan distribusi posterior atas variabel laten zi yang diberikan datum xi yang dilambangkan dengan pengenalan, yang setara dengan penyandi

probabilistik dan diatur oleh parameter ϕ. Terakhir, Bank mengasumsikan distribusi prior untuk variabel laten zi yang dilambangkan dengan $p\theta$ (zi). Gambar 3 menggambarkan hubungan yang dijelaskan di atas. Parameter θ dan ϕ tidak diketahui dan perlu dipelajari dari data. Variabel laten yang diamati zi dapat diinterpretasikan sebagai kode yang diberikan oleh model pengenalan $q\phi$ (z|x) (Bank, 2023).



Gambar 6. 1 Representasi grafis dari VAE

Encoder mengembalikan mean dan deviasi standar untuk setiap input; kemudian, vektor laten diambil sampelnya dari distribusi ini & dikirim ke decoder utk merekonstruksi masukan. Pelatihan sekarang dikendalikan oleh kerugian rekonstruksi & hilangnya kesamaan, yang merupakan perbedaan KL antara distribusi ruang laten & Gaussian standar. VAE dilatih tidak hanya untuk merekonstruksi gambar, tetapi juga untuk menghasilkan vektor laten dari distribusi normal (Bank, 2023).



Gambar 6. 2 Arsitektur Variational Autoencoders (VAE)

2.2.6 U-Net

Segmentasi gambar medis otomatis adalah topik penting dalam domain medis dan secara berturut-turut menjadi mitra penting dalam paradigma diagnosis berbantuan komputer. U-Net adalah arsitektur segmentasi gambar yang paling luas karena fleksibilitasnya, desain modular yang dioptimalkan, dan keberhasilannya dalam semua modalitas gambar medis. Selama

bertahun-tahun, model U-Net mendapatkan perhatian yang luar biasa dari para peneliti akademis dan industri. Beberapa perluasan jaringan ini telah diusulkan untuk mengatasi skala dan kompleksitas yang diciptakan oleh tugas-tugas medis. Mengatasi kekurangan model U-Net yang naif adalah langkah terdepan bagi vendor untuk memanfaatkan model varian U-Net yang tepat untuk bisnis mereka. Memiliki ringkasan varian yang berbeda di satu tempat akan memudahkan para pembangun untuk mengidentifikasi penelitian ssyang relevan. Selain itu, bagi para peneliti ML, hal ini akan membantu mereka memahami tantangan tugas biologis yang menjadi tantangan model (Azad, 2024).

2.2.7 *Parameter Seed*

Seed merupakan angka yang digunakan untuk menginisialisasi pembuatan angka acak. Ini digunakan untuk membuat inisial kebisingan laten. Dengan menanam benih, kita bisa memperbanyak hasilnya. Dalam implementasinya, Difusi Stabil menghasilkan tensor acak di ruang laten (Shen, 2024).

Kita mengendalikan ini tensor dengan mengatur seed generator nomor acak. Jika kita menyetel benih ke nilai tertentu, kita akan selalu melakukannya dapatkan tensor acak yang sama. Ingatlah bahwa AI menghasilkan keseluruhan gambar dari noise - seperti mengambil beberapa pensil warna bersama-sama & mengecat titik-titik kecil pada lembaran tersebut, hingga akhirnya menghasilkan gambar yang diinginkan (Shen, 2024).

2.2.8 Parameter CFG

Baru-baru ini, generasi yang dipandu teks dalam model difusi telah mencapai tingkat yang belum pernah terjadi sebelumnya, seperti Dall E-3. Kekuatan generatif ini berasal dari tiga aspek.

1. Pertama, untuk merepresentasikan teks yang tidak terstruktur, penyematan bahasa ekspresif digunakan untuk menanamkan setiap token dalam teks

yang diberikan, seperti CLIP dalam *Stable Diffusion*, dan T5, dan dalam Imagen.

2. Kedua, untuk memfasilitasi interaksi antara informasi teks dan gambar, model difusi biasanya meningkatkan tulang punggung jaringan, seperti U-net tulang punggung U-net, dengan mekanisme perhatian silang. Ini mekanisme ini melibatkan pemanfaatan penyematan gambar sebagai kueri dan penyematan kunci dan nilai yang berasal dari teks (Shen dkk., 2024).
3. Ketiga, *Classifier-Free Guidance (CFG)* baru-baru ini secara luas dilibatkan sebagai teknik yang ringan dan kuat teknik yang ringan dan kuat untuk mendorong kepatuhan terhadap permintaan teks dalam beberapa generasi. Alih-alih melatih pengklasifikasi tambahan, CFG menggabungkan estimasi skor dari model difusi dengan atau dengan atau tanpa prompt bersyarat. Beberapa karya lain lebih lanjut memisahkan prompt menjadi beberapa konsep dan menghasilkan gambar dengan menggabungkan satu set model difusi dengan masing-masing model difusi yang dikondisikan pada komponen konsep tertentu (Shen dkk., 2024).

2.2.9 Cerita Anak "Si Kancil dan Buaya"

Kisah Si Kancil dan Buaya merupakan bagian dari folklore atau cerita rakyat yang telah diwariskan secara turun-temurun di berbagai budaya di Asia Tenggara, khususnya di Indonesia, Malaysia, dan sekitarnya. Cerita ini termasuk dalam kumpulan dongeng Si Kancil, yang menggambarkan kecerdikan hewan kecil dalam menghadapi bahaya atau lawan yang lebih besar dan kuat (La Bania & Milawaty, 2019).

Kisah Si Kancil dan Buaya pertama kali berkembang secara lisan di masyarakat Nusantara sebelum akhirnya ditulis dalam berbagai bentuk sastra, seperti naskah kuno, buku cerita anak, dan komik. Dalam tradisi lisan, cerita ini sering diceritakan oleh orang tua kepada anak-anak sebagai pengajaran moral. Kancil sendiri adalah simbol kepintaran, sementara buaya sering digambarkan sebagai hewan yang mudah tertipu dan rakus (La Bania & Milawaty, 2019).

Seiring waktu, kisah ini menjadi alat pendidikan karakter bagi anak-anak. Pesan utama dari cerita ini adalah bahwa kecerdikan dan akal dapat mengalahkan kekuatan fisik. Namun, di beberapa versi lain, ada juga pesan tentang kejujuran dan akibat dari menipu, karena dalam beberapa cerita lain Kancil akhirnya mendapatkan balasan atas kelicikannya. Sosok Kancil dapat dipahami sebagai tokoh dengan karakter yang penuh dengan resistensi terhadap mereka yang lebih kuat darinya. Dirinya yang lemah malah menjadi sosok "pahlawan" yang menyelamatkan kelompoknya atau dirinya sendiri dari represi golongan pemilik kuasa. Di sini, dapat dilihat bahwa cerita Kancil merupakan representasi dari pertentangan antarkelas, yakni kelompok atas dan kelompok bawah, kelompok penguasa dengan kelompok tanpa kuasa seperti rakyat jelata. Di sisi lain, cerita Kancil juga dipahami sebagai sebuah penggambaran ekspresi dari kondisi psikologi kelompok rakyat jelata dalam upaya melakukan perlawanan terhadap mereka yang melakukan represi, seperti para penjajah. Dalam hal ini, cerita Kancil dapat dianggap sebagai salah satu arkeologi sejarah usaha perjuangan rakyat jelata melawan ketimpangan dan ketidaksetaraan dari golongan para penguasa, di mana Kancil dikonstruksi sebagai simbol dari perlawanan terhadap tindakan represi, ketidaksetaran, atau pun kesenjangan dalam struktur hubungan sosial antara golongan kelas atas dan kelas bawah (La Bania & Milawaty, 2019).

Kini, cerita Si Kancil dan Buaya telah banyak diadaptasi ke dalam buku bergambar, pertunjukan wayang, film animasi, dan media digital lainnya. Meski mengalami berbagai modifikasi, esensi dari kisah ini tetap bertahan sebagai salah satu dongeng paling dikenal di Nusantara (La Bania & Milawaty, 2019).

2.2.10 *Conditional Maximum Mean Discrepancy (CMMD)*

Conditional Maximum Mean Discrepancy (CMMD) adalah perpanjangan dari konsep *Maximum Mean Discrepancy (MMD)* yang digunakan untuk mengukur perbedaan antara dua distribusi probabilitas dalam ruang fitur dengan mempertimbangkan kondisi tertentu (Ren dkk., 2021).

MMD adalah metrik yang digunakan dalam statistik dan pembelajaran mesin untuk membandingkan dua distribusi probabilitas (P) dan (Q) dalam ruang fitur. MMD mengukur seberapa jauh ekspektasi fungsi-fungsi dalam Ruang Reproduksi Kernel Hilbert (RKHS) terhadap kedua distribusi tersebut (Ren dkk., 2021).

CMMD memperluas MMD dengan mempertimbangkan distribusi bersyarat. Misalkan kita memiliki variabel acak (X) dan (Y), CMMD mengukur perbedaan antara distribusi bersyarat (P(X | Y)) dan (Q(X | Y)), bukan hanya distribusi marginal (Ren dkk., 2021).

Misalkan kita memiliki dua dataset sampel:

X = { $x_1, x_2, ..., x_m$ }) dari distribusi (P)

Y = { $y_1, y_2, ..., y_n$ }) dari distribusi (Q)

Maka rumus CMMD :

$$\text{CMMD} = \mathrm{E}_{X,X'}[k(X,X')] + \mathrm{E}_{Y,Y'}[k(Y,Y')] - 2\mathrm{E}_{X,Y}[k(X,Y)]$$

$k(X,X')$ merupakan Fungsi Radial Basis Function (RBF) Variabel X

$k(Y,Y')$ merupakan Fungsi Radial Basis Function (RBF) Variabel Y

Aplikasi CMMD :

- *Domain Adaptation*: Digunakan untuk mengukur perbedaan distribusi bersyarat antara data sumber dan target dalam transfer learning.
- *Causal Inference*: Membantu dalam analisis sebab-akibat dengan membandingkan distribusi bersyarat dari variabel sebelum dan sesudah perlakuan (*treatment*).
- *Fairness in Machine Learning*: Digunakan untuk mengevaluasi bias dalam model AI berdasarkan distribusi bersyarat dari variabel sensitif.

Keunggulan CMMD dibanding MMD

- Memperhitungkan struktur kondisi dalam data, sehingga lebih informatif dibanding MMD standar.
- Lebih cocok untuk masalah yang melibatkan perubahan distribusi bersyarat, seperti domain adaptation dan fairness.

CMMD adalah alat yang kuat dalam analisis distribusi probabilitas bersyarat, terutama dalam *machine learning* dan statistik (Ren dkk., 2021).

BAB III
METODE PENELITIAN

3.1 Metode Penelitian

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

Gambar 3. 1 Tahapan Penelitian

Keluaran dari tugas akhir ini adalah sistem yang dapat menghasilkan gambar karakter dari cerita anak “Kancil dan Buaya” berdasarkan deskripsi teks yang diberikan oleh. Dengan menggunakan pendekatan *Diffusion model*, mana memanfaatkan proses transformasi iteratif dari *noise* acak menjadi gambar yang lebih realistis berdasarkan pola pelatihan model (*training*).

3.1.1 Studi Pustaka dan Pengumpulan Data

Sebelum memulai proyek pembuatan ilustrasi buku anak-anak, penting untuk melakukan studi pustaka. Ini melibatkan penelusuran berbagai sumber terkait teknik ilustrasi, psikologi warna, dan tema yang relevan dengan cerita yang akan diilustrasikan. Selain itu, kumpulkan data mengenai audiens target, termasuk preferensi visual dan gaya bercerita yang menarik bagi anak-anak. Informasi ini akan menjadi dasar yang kuat dalam merancang ilustrasi yang tidak hanya menarik tetapi juga edukatif.

3.1.2 Menyiapkan Lingkungan Kerja

Setelah studi pustaka, langkah berikutnya adalah menyiapkan lingkungan kerja yang mendukung kreativitas. Pastikan perangkat keras seperti komputer

dan tablet grafis berfungsi dengan baik. Instal perangkat lunak yang diperlukan untuk model diffusion, seperti aplikasi pengolah gambar dan perangkat lunak AI yang akan digunakan untuk menghasilkan ilustrasi. Ciptakan suasana kerja yang nyaman dengan pencahayaan yang baik dan minim gangguan, sehingga tim bisa fokus dalam berkreasi.

3.1.3 Memasukkan Prompt Deskripsi

Dalam sebuah cerita, tokoh memiliki peran penting dalam membangun alur dan memberikan kesan mendalam bagi pembaca atau pendengar. Untuk menghidupkan karakter dalam cerita *Si Kancil dan Buaya*, deskripsi tokoh harus disusun dengan jelas dan menarik. Pada bab ini, akan dibahas cara memasukkan prompt deskripsi ciri fisik tokoh secara efektif agar cerita lebih hidup dan dapat dipahami dengan baik.

Si Kancil adalah tokoh utama dalam cerita yang digambarkan sebagai hewan kecil, lincah, dan cerdik. Ia memiliki tubuh ramping dengan bulu cokelat kemerahan yang membuatnya mudah berbaur dengan lingkungan hutan. Kakinya kecil tetapi kuat, memungkinkan ia bergerak cepat dan melompat dengan lincah. Telinganya tegak dan sensitif terhadap suara di sekitarnya. Matanya tajam dan selalu terlihat penuh kecerdikan, mencerminkan kepintarannya dalam menghadapi tantangan. Ekornya pendek dan sering kali bergerak saat ia merasa waspada.

Prompt deskripsi ciri fisik untuk Si Kancil:

"Bayangkan seekor kancil kecil yang lincah dengan bulu cokelat kemerahan. Matanya tajam dan penuh kecerdikan, seolah selalu memiliki rencana cerdik dalam pikirannya. Telinganya tegak, menangkap setiap suara di sekelilingnya, sementara kakinya yang kecil tetapi kuat membantunya berlari cepat melewati semak-semak dan melompati rintangan dengan gesit."

Buaya dalam cerita ini berperan sebagai tokoh yang tertipu oleh kecerdikan Si Kancil. Ia digambarkan sebagai makhluk besar dan kuat, dengan sisik keras berwarna hijau gelap yang membuatnya tampak menakutkan. Tubuhnya panjang dan berat, dengan ekor yang kuat yang dapat menghantam air dengan keras. Matanya kecil tetapi tajam, dengan kelopak

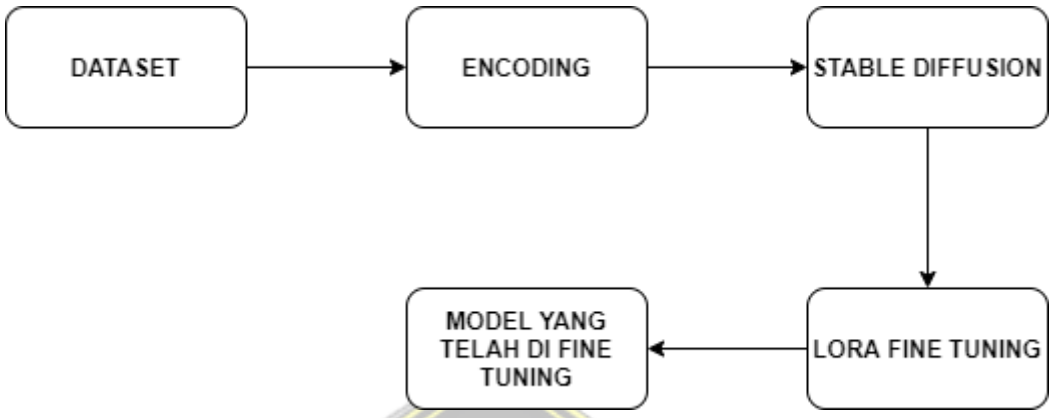
tebal yang melindunginya dari air dan cahaya terang. Rahangnya lebar dengan deretan gigi tajam yang mampu merobek mangsanya dalam sekejap. Kakinya pendek tetapi kokoh, memungkinkannya merayap dengan tenang di tepi sungai sebelum menerkam mangsanya.

Prompt deskripsi ciri fisik untuk Buaya:

"Di tepi sungai yang deras, seekor buaya raksasa berjemur di atas pasir. Tubuhnya besar dan kuat, dengan sisik hijau gelap yang berkilauan terkena sinar matahari. Rahangnya lebar dengan deretan gigi tajam yang siap merobek mangsa dalam sekali gigitan. Matanya kecil tetapi penuh kewaspadaan, sementara ekornya yang panjang dan kuat bergerak perlahan di air, siap menghantam kapan saja."

Untuk menghidupkan cerita, *prompt* deskripsi ciri fisik tokoh harus dimasukkan secara alami ke dalam narasi. Kombinasikan deskripsi tokoh dengan lingkungan sekitarnya untuk menciptakan gambaran yang lebih imajinatif.

3.1.4 *Fine Tuning*



Gambar 3. 2 Alur Kerja *Training* Sistem

Berikut adalah alur kerja *training* seperti yang ditunjukkan pada Gambar 3. 2 :

1. Mencari dan Menyiapkan Dataset

   Dataset yang baik sangat penting untuk mendapatkan hasil yang sesuai dengan karakter Si Kancil dan Buaya.

   - Menentukan Kriteria Dataset

     Gambar harus sesuai dengan gaya ilustrasi anak-anak (misalnya kartun atau lukisan digital). Resolusi tinggi untuk menghindari *noise* saat pelatihan. Variasi pose dan ekspresi tokoh agar model dapat memahami karakter dengan baik.

   - Mencari Sumber Dataset

     Google Images & Pinterest Cari ilustrasi yang sesuai (perhatikan hak cipta).

   - *Preprocessing Dataset*

     Hapus gambar yang buram, memiliki *watermark*, atau tidak sesuai dengan konsep cerita. Sesuaikan ukuran gambar (misalnya 512x512 px) untuk optimalisasi *training*.

   Setelah dataset siap, tahap berikutnya adalah mengubahnya ke dalam format yang bisa dipahami oleh model.

2. Encoding
   - *Captioning* (Pemberian Label pada Gambar)

     Setiap gambar perlu diberi deskripsi teks yang menjelaskan kontennya. Bisa dilakukan dengan *BLIP (Bootstrapped Language Image Pretraining)*.

     Contoh caption:

     "A cartoon-style illustration of Si Kancil standing near the river, looking at the crocodile."

     "Si Kancil is jumping over the crocodiles in a colorful children's book illustration."
   - *Tokenization*

     Teks deskripsi dikonversi menjadi token agar bisa digunakan dalam training. Biasanya menggunakan *tokenizer* dari model CLIP yang ada di *Stable Diffusion*.
   - *Latent Encoding* (Mengubah Gambar ke Format *Latent Space*)

     Stable Diffusion tidak langsung bekerja dengan gambar biasa, tapi mengubahnya ke dalam *latent space* menggunakan *variational autoencoder (VAE).* Ini membantu model memahami fitur visual dalam bentuk *compressed representation*.
   - Tahap *Preprocessing Text* dengan *Encoder*

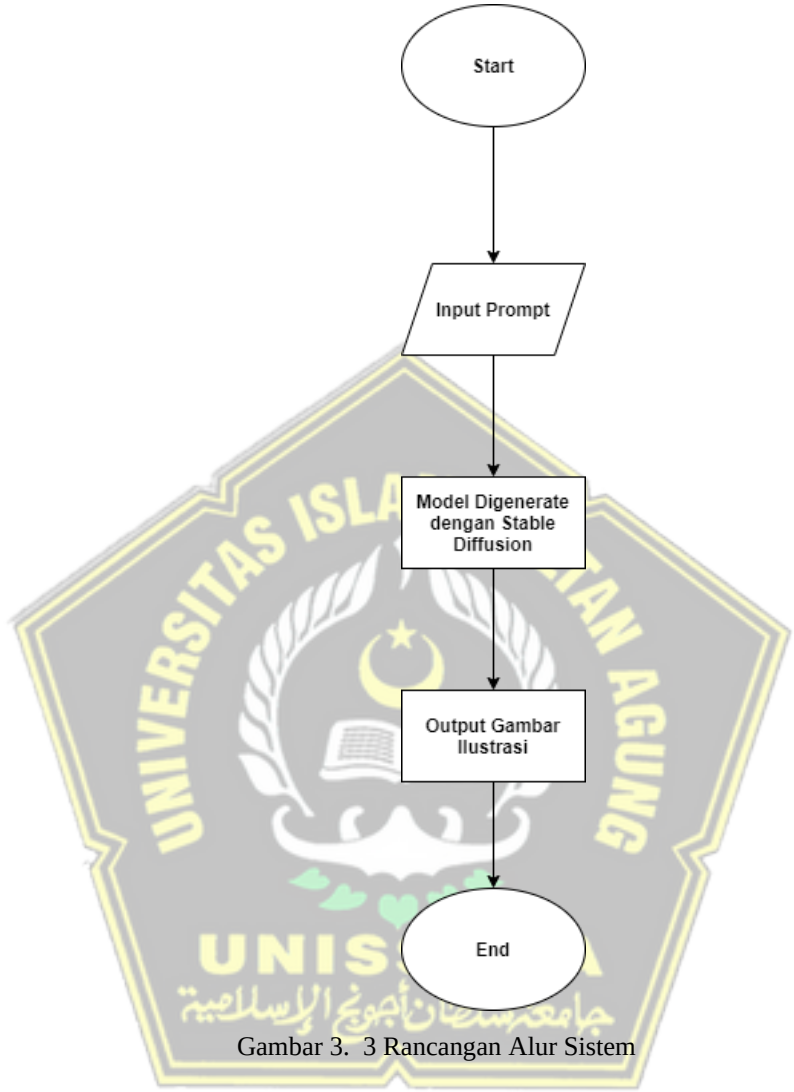
     Text/Keyword Encoding : Deskripsi teks yang merepresentasikan karakter diproses menggunakan CLIP *Text Encoder*, menghasilkan representasi vektor teks (*embedding* vector). *Embedding teks* ini digunakan sebagai *prompt*, yang menjadi acuan bagi model untuk memahami elemen visual karakter. *Library Transformers* juga digunakan untuk memastikan kompatibilitas *embedding* dengan arsitektur *Stable Diffusion*.
3. Tahapan di Dalam *Diffusion Models* (*Stable Diffusion* v2.1)
   - *Gaussian Noise* Addition : *Noise Gaussian* ditambahkan iteratif di *latent space*, memastikan gambar acak sepenuhnya sebelum di-de*noise*.

U-Net Architecture : Proses *denoising* terjadi di *latent space*, dengan U-Net bertugas menghapus *noise* secara bertahap sesuai informasi teks.

- Variational Auto*encoder Decoder* : Setelah proses *denoising* selesai, representasi laten diterjemahkan kembali menjadi gambar melalui *decoder autoencoder*, menghasilkan *output* gambar akhir.
- *Iterasi Fine-tuning* : Jika hasil belum optimal, pelatihan dilanjutkan dengan menyesuaikan parameter seperti *learning rate* atau *dataset* tambahan.

4. Tahap Pelatihan dengan *Fine-Tuning LORA*
    - Pembelajaran Berbasis Gambar Referensi : Gambar-gambar referensi dari *dataset* digunakan untuk mengaitkan deskripsi teks dengan karakter visual tertentu.
    - *Forward Pass* : Model menerima *text embedding* dan gambar referensi untuk menghasilkan latent representation. Gambar awal ini mengandung *noise* yang secara bertahap akan dihilangkan selama pelatihan.
    - Perhitungan *Loss* : Menggunakan fungsi *loss*, seperti *mean squared error* (MSE), untuk menghitung kesalahan antara gambar yang dihasilkan dan gambar referensi.
    - *Backward Pass* : *Loss* yang dihitung digunakan untuk memperbarui bobot jaringan neural.

3.2 Perancangan Alur Sistem



Gambar 3. 3 Rancangan Alur Sistem

1. *Input Prompt*

   Gunakan prompt yang sesuai untuk menghasilkan ilustrasi. Misalnya, "Ilustrasi Si Kancil berpetualang di hutan." Proses ini akan menghasilkan gambar berdasarkan *prompt* yang diberikan.

2. Generate Gambar Ilustrasi

   Sistem akan memproses deskripsi teks melalui *Text Encoder* menjadi *Text Embedding*, dan *U-Net* untuk menghasilkan representasi laten. Model *Stable Diffusion v1.5* akan memproses representasi laten dan menambahkan *Noise* sesuai dengan algoritma yang telah dilatih.

Akhirnya, sistem akan menghasilkan gambar poster berdasarkan input yang diberikan.

3. *Output*

   Gambar Ilustrasi yang telah dihasilkan akan menjadi *output* akhir dari sistem.

3.3 Analisis Kebutuhan Sistem

Analisis kebutuhan sistem dilakukan untuk memastikan sistem yang dibangun dapat memenuhi tujuan penelitian, mencakup kebutuhan *library*.

Tabel 3. 1 Tabel *Library*

Library	**Deskripsi**	**Fungsi**
Pytorch	*Library deep learning* fleksibel untuk membangun, melatih, dan menguji model pembelajaran mendalam.	• Implementasi model *Stable Diffusion* 1.5. • Mendukung backpropagation untuk pelatihan model. • Memanfaatkan GPU untuk efisiensi.
Transformers	*Library* untuk pemrosesan teks dan *Encoding* representasi vektor.	• *Encoding* deskripsi teks menggunakan CLIP Text *Encoder*. • Integrasi dengan Diffusers untuk proses teks ke gambar.
Diffusers	*Library* khusus untuk *Diffusion Models*, termasuk *Stable Diffusion*.	• Pipeline untuk proses *denoising*. • Mendukung *fine-tuning Dreambooth*. • Akses ke pre-trained model *Stable Diffusion*.

Matplotlib	*Library* visualisasi data.	• Menampilkan grafik evaluasi *fidelity* dan personalisasi. • Visualisasi metrik pelatihan, seperti *loss* function.
NumPy	*Library* untuk komputasi numerik.	• Normalisasi data numerik. • Operasi pada array dan matriks untuk pemrosesan gambar dan analisis data.
Pandas	*Library* manipulasi data berbasis tabel.	• Pengorganisasian metadata *dataset* • Menyajikan hasil evaluasi fidelitas dalam tabel.

3.4 Perancangan User Interface

Perancangan *User Interface* untuk *AI Image Generator* dirancang dengan pendekatan minimalis yang mengutamakan fungsionalitas utama dan kemudahan penggunaan. Pada tampilan awal, halaman dimulai dengan judul yang terletak di bagian atas sebagai penanda utama. Di bawahnya, terdapat kolom *input* berbentuk persegi panjang dengan latar abu-abu muda, yang berfungsi sebagai tempat pengguna memasukkan *prompt* atau deskripsi gambar yang ingin dihasilkan. Untuk memberikan panduan, kolom ini dilengkapi dengan *placeholder* bertuliskan "*Please Enter your prompt here!"*. Tepat di bawah kolom *input*, terdapat tombol bertuliskan "*Generate Image*" dengan desain sederhana latar putih dan teks hitam yang mudah diidentifikasi dan diakses pengguna.

Gambar 3. 4 Tampilan Saat Generasi Gambar

Setelah tombol "*Generate Image*" ditekan, antarmuka menampilkan status proses seperti "*Loading Model..."* dan "*Generating Model..."* di bawah tombol untuk memberikan feedback kepada pengguna mengenai status generasi gambar. Di bagian bawah halaman, terdapat area besar berbentuk persegi panjang dengan latar abu-abu yang berfungsi sebagai tempat *output* gambar ditampilkan setelah proses selesai. Elemen ini dirancang secara sentral untuk memprioritaskan perhatian pengguna terhadap hasil yang dihasilkan.

BAB IV
HASIL DAN ANALISIS PENELITIAN

4.1 Pemersiapan Model dan Dataset

Dalam proses *fine-tuning model*, tahap awal yang dilakukan adalah menyiapkan dataset yang sesuai dengan tokoh dan suasana dalam cerita Si Kancil dan Buaya. Dataset yang digunakan terdiri dari berbagai ilustrasi bertema fauna, hutan, sungai, serta ekspresi karakter yang sesuai dengan cerita anak-anak. Untuk memastikan konsistensi gaya gambar, gambar dalam dataset dipilih dengan mempertimbangkan keseragaman warna, detail karakter, serta komposisi lingkungan.

4.1.1 Seleksi dan Pengolahan Dataset

Tahap *preprocessing dataset* dilakukan untuk memastikan kualitas gambar tetap tinggi dan sesuai dengan kebutuhan model. Langkah yang dilakukan meliputi *resize* gambar agar sesuai dengan resolusi yang optimal untuk *Stable Diffusion 2.1*, normalisasi warna untuk menjaga keseimbangan visual. Dataset yang digunakan dalam *fine-tuning* dikumpulkan dari berbagai sumber dan dikurasi untuk memastikan bahwa gambar memiliki gaya visual yang konsisten. Beberapa kriteria seleksi dataset meliputi:

- Kesesuaian dengan cerita anak-anak, di mana gambar harus memiliki warna cerah, garis lembut, dan kesan yang ramah untuk anak-anak.
- Keseragaman desain karakter, memastikan bahwa Si Kancil dan Buaya memiliki bentuk yang konsisten di setiap ilustrasi.
- Keberagaman sudut pandang, agar model dapat mengenali karakter dalam berbagai pose dan ekspresi, termasuk saat Si Kancil tersenyum, berpikir, atau tertawa, serta Buaya dalam posisi diam, menyelam, atau marah.

Gambar 4. 1 Contoh *Dataset*

Setelah seleksi, dilakukan preprocessing dengan menyesuaikan resolusi gambar, menormalkan warna, serta mengonversi gambar ke format *latent space* menggunakan *Variational Autoencoder (VAE)* agar model lebih mudah memahami pola gambar yang akan dihasilkan.

4.1.2 Konfigurasi Model *Fine-Tuning*

Fine-tuning dilakukan menggunakan *Stable Diffusion 2.1* sebagai model dasar dengan penerapan *LoRA* untuk memungkinkan pelatihan yang lebih efisien. Sebagai model dasar, *Stable Diffusion 2.1* dipilih karena memiliki resolusi gambar yang lebih tinggi dibandingkan versi sebelumnya, yang memungkinkan detail lebih tajam dan realistis dalam ilustrasi. *Parameter fine-tuning* yang digunakan dalam pelatihan LoRA meliputi *Rank (r)* untuk menentukan jumlah parameter yang diperbarui, *Learning Rate* untuk mengontrol seberapa cepat model belajar dari dataset, serta penggunaan

Optimizer AdamW yang dinilai mampu mempercepat proses konvergensi model. Selain itu, eksperimen dilakukan dengan variasi *epoch* (5, 10, dan 20) serta *batch size* (4, 8, dan 16) guna mengevaluasi dampaknya terhadap kualitas gambar yang dihasilkan.

Beberapa parameter yang digunakan dalam pelatihan meliputi:

```
!accelerate launch /content/drive/MyDrive/scripts/train_text_to_image_lora_csv.py \
  --pretrained_model_name_or_path="stabilityai/stable-diffusion-2-1" \
  --train_data_dir="/content/drive/MyDrive/train_dataset/train_dataset.csv" \
  --image_column="image" \
  --caption_column="captions" \
  --dataloader_num_workers=8 \
  --resolution=512 \
  --random_flip \
  --train_batch_size=1 \
  --gradient_accumulation_steps=4 \
  --max_train_steps=600 \
  --learning_rate=2e-5 \
  --max_grad_norm=1 \
  --lr_scheduler="cosine_with_restarts" \
  --lr_warmup_steps=100 \
  --output_dir="/content/drive/MyDrive/lora_outputs" \
  --checkpointing_steps=100 \
  --validation_prompt="a cartoon illustration of The spotted antlered deer has an appearance
  --seed=777 \
  --rank=16 \
  --report_to tensorboard \
  --mixed_precision fp16 \
  --snr_gamma=5.0
```

Gambar 4. 2 Kode Untuk Konfigurasi Model

- *Rank (r)* = 16 → Menentukan jumlah parameter yang diperbarui dalam LoRA. *LoRA rank* (semakin tinggi, semakin kompleks adaptasi LoRA).
- *Learning Rate* = 2e-5 → Mengontrol seberapa cepat model belajar dari dataset. *Learning rate* 2×10^{-5}.
- *Seed*=777 → Mengatur *seed* untuk memastikan hasil yang dapat direproduksi.
- *Mixed_precision fp16* → Menggunakan *mixed precision (FP16)* untuk mempercepat pelatihan di GPU.
- SNR_Gamma=5.0 → Parameter SNR (*Signal-to-Noise Ratio*) yang dapat membantu stabilisasi pelatihan.

4.2 Implementasi *Fine-Tuning LoRA*

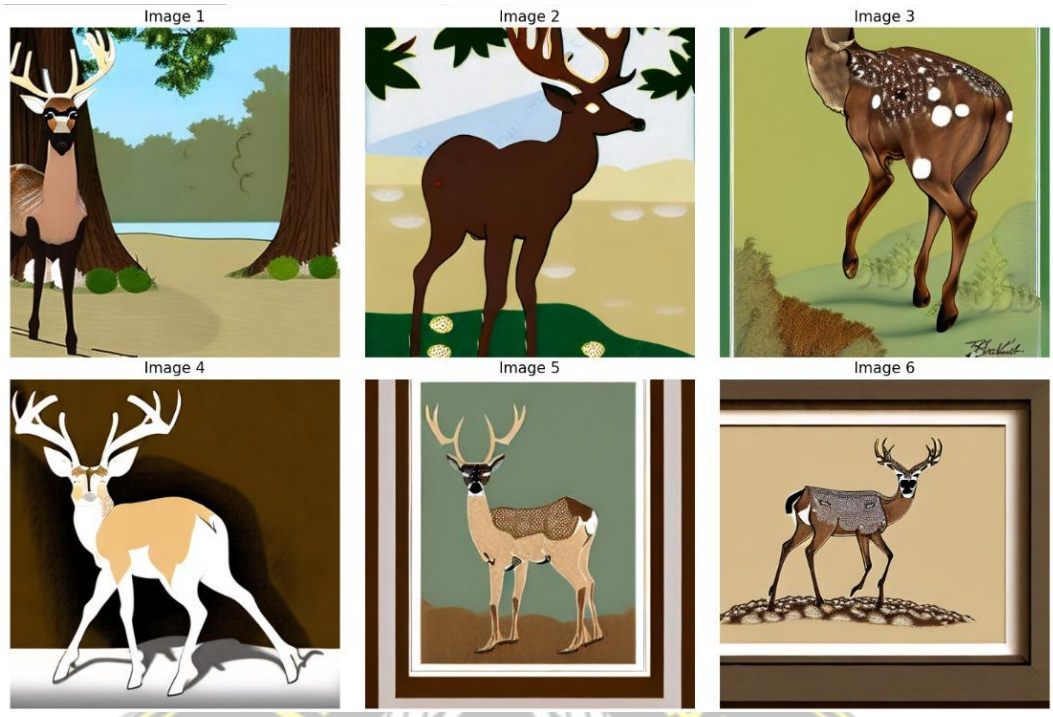
Fine-tuning menggunakan LoRA dilakukan untuk memungkinkan model *Stable Diffusion 2.1* memahami dan mereplikasi gaya ilustrasi yang sesuai dengan cerita Si Kancil dan Buaya. Proses ini dilakukan dengan melatih model menggunakan subset parameter yang relevan, sehingga *fine-tuning* dapat dilakukan dengan konsumsi memori yang lebih rendah dibandingkan metode seperti *DreamBooth*.

Selama proses *fine-tuning*, model diuji dengan beberapa konfigurasi parameter untuk menilai pengaruh jumlah *epoch* dan *batch size* terhadap hasil gambar yang dihasilkan. Setelah proses pelatihan selesai, model yang telah dioptimalkan diintegrasikan kembali ke dalam *pipeline Stable Diffusion 2.1* untuk dibandingkan dengan model sebelum *fine-tuning*. Tujuan utama dari pengujian ini adalah memastikan bahwa model dapat menghasilkan gambar Si Kancil dan Buaya yang sesuai dengan gaya ilustrasi anak-anak, dengan komposisi yang sejalan dengan cerita.

4.2.1 Proses Pelatihan Model

Model dilatih dengan memasukkan dataset yang telah diproses dan menggunakan *Loss Function Mean Squared Error (MSE)* untuk menilai perbedaan antara gambar asli dan gambar yang dihasilkan selama pelatihan. Selama proses pelatihan, model diuji dengan berbagai prompt deskriptif yang menggambarkan ciri fisik tokoh dalam cerita Si Kancil dan Buaya, seperti:

"A spotted antlered deer has an appearance characterized by light brown fur dotted with distinctive white spots. These spots are most prominent along its back and flanks, creating a beautiful contrast to its darker coat color. Male deer, have large branched antlers, the structure of which resembles smooth tree branches. The legs are long and slender. Big, dark doe eyes."

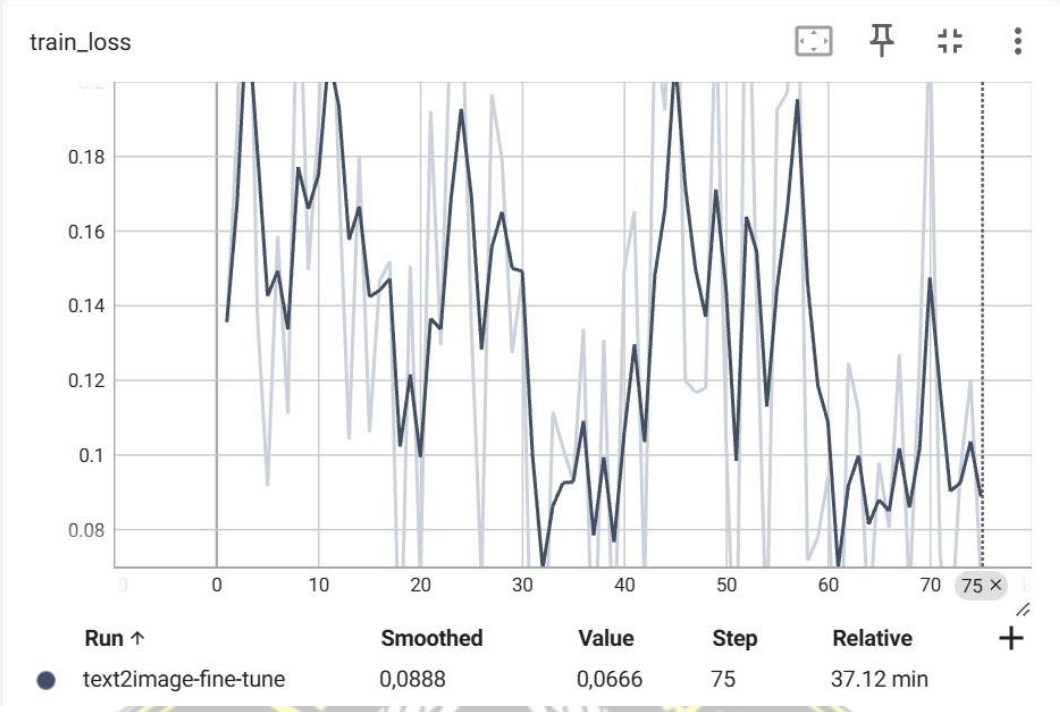


Gambar 4. 3 Hasil Generalisasi Awal

Setelah pelatihan selesai, model diintegrasikan ke dalam *pipeline Stable Diffusion 2.1* untuk diuji dan dibandingkan dengan model sebelum *fine-tuning*.

4.2.2 Evaluasi Ilustrasi yang Dihasilkan

Evaluasi dilakukan dengan melihat apakah gambar yang dihasilkan sesuai dengan cerita dan mempertahankan konsistensi karakter, kesesuaian ekspresi, serta akurasi latar belakang.



Gambar 4. 4 Grafik Loss Model

Gambar 4.4 menunjukkan hasil *Loss* rata-rata skor sekitar 0.1, menunjukkan perbaikan dalam pemahaman dan akurasi visualisasi sehingga dapat menghasilkan gambar berkualitas tinggi.

4.3 Generasi Illustrasi Tokoh Cerita "Kancil dan Buaya"

4.3.1 Tokoh 1 - Si Kancil

Teks *Prompt* Deskripsi:

"A spotted antlered deer has an appearance characterized by light brown fur dotted with distinctive white spots. These spots are most prominent along its back and flanks, creating a beautiful contrast to its darker coat color. Male

deer, have large branched antlers, the structure of which resembles smooth tree branches. The legs are long and slender. Big, dark doe eyes."

Gambar 4. 5 Gambar Hasil Generasi Sebelum Fine Tuning

Gambar 4. 6 Gambar Hasil Generasi Setelah Fine Tuning

4.3.2 Tokoh 2 - Buaya

Teks *Prompt* Deskripsi:

"A crocodiles are reptiles with sturdy, elongated bodies covered in tough, scaly skin that ranges from olive green in color. Its head is large and triangular, equipped with a wide, gaping mouth containing sharp, pointed

teeth. His eyes were fixed high above his head. The crocodile's muscular tail is long and thick. Its limbs are strong and clawed."

Gambar 4. 7 Gambar Hasil Generasi Sebelum Fine Tuning

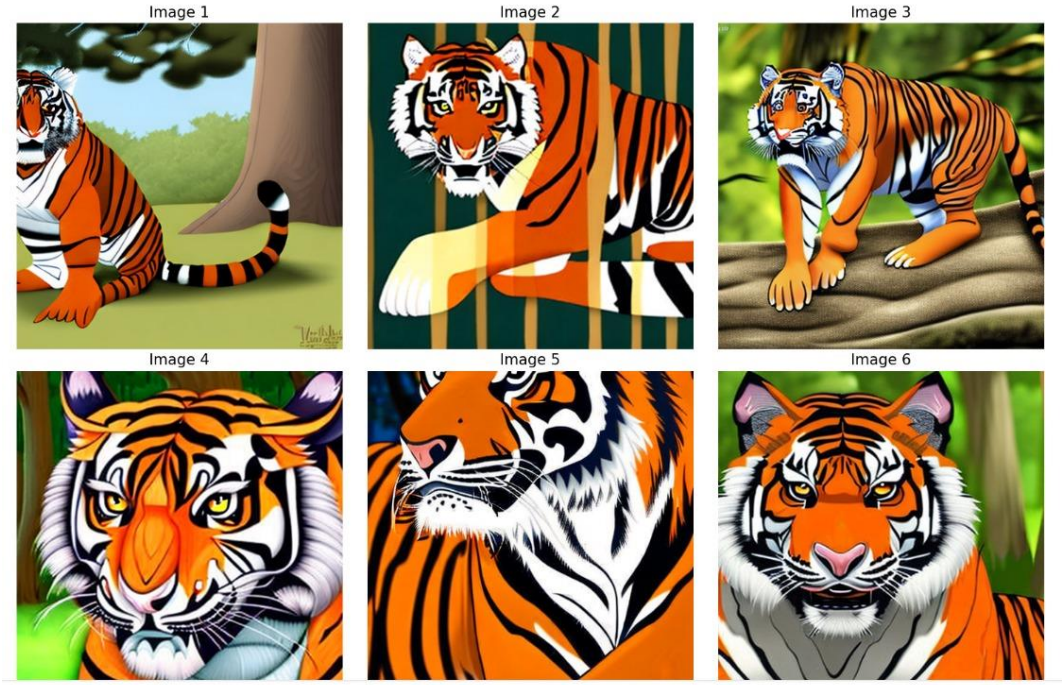
Gambar 4. 8 Gambar Hasil Generasi Setelah Fine Tuning

4.3.3 Tokoh 3 - Harimau

Teks *Prompt* Deskripsi:

"A tiger is a large cat with a muscular body, large head and long tail. It's coat is orange-brown with vertical black stripes that are unique to each individual, serving as camouflage in its natural habitat Tigers have sharp eyes with round pupils strong legs with sharp retractable claws. Background forest."

Gambar 4. 9 Gambar Hasil Generasi Sebelum Fine Tuning

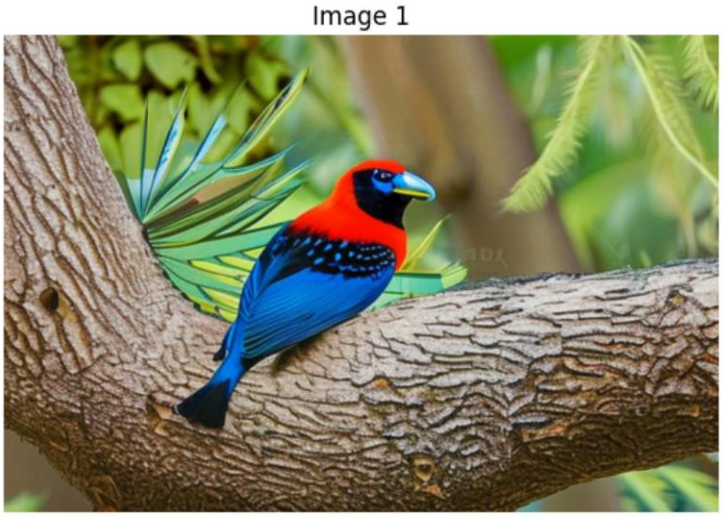


Gambar 4. 10 Gambar Hasil Generasi Setelah Fine Tuning

4.3.4 Tokoh 4 - Burung

Teks *Prompt* Deskripsi:

"A Birds have generally slender bodies covered in feathers that maintain body temperature and aid in flight They have a pair of wings although not all birds are capable of flight Bird beaks vary in shape and size according to the type of food. Background forest."



Gambar 4. 11 Gambar Hasil Generasi Sebelum Fine Tuning

Gambar 4. 12 Gambar Hasil Generasi Setelah Fine Tuning

4.4 Evaluasi Kinerja Model

Evaluasi model dilakukan dengan membandingkan kualitas gambar sebelum dan sesudah *fine-tuning*. Salah satu aspek utama yang diperiksa adalah kesesuaian visual dengan cerita, yakni apakah ilustrasi yang dihasilkan mampu menangkap prompt deskripsi tokoh dalam Si Kancil dan Buaya. Sebelum *fine-tuning*, gambar yang dihasilkan sering kali tidak sesuai dengan tema cerita, seperti desain karakter yang kurang konsisten atau latar belakang yang tidak menggambarkan suasana hutan dan sungai dengan baik. Setelah fine-tuning menggunakan dataset yang lebih spesifik, gambar yang dihasilkan menunjukkan peningkatan dalam hal detail karakter, warna, dan suasana cerita.

```
CMMD Score (Pretrained): 0.7167680859565735
CMMD Score (Fine-tuned): 0.5824565887451172
Model fine-tuned menghasilkan gambar yang lebih sesuai dengan prompt!
```

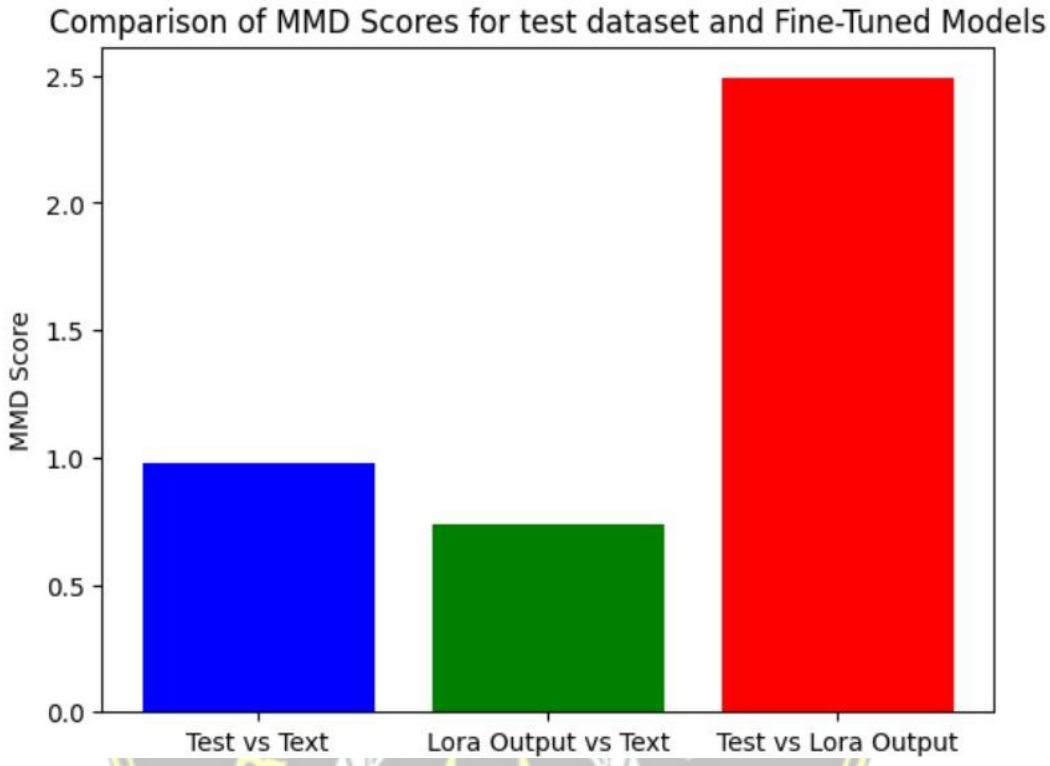
Gambar 4. 13 Gambar CMMD Score Pretained dan Fine-Tuned

Gambar 4.13 menunjukkan penelitian ini berhasil menghasilkan gambar yang sesuai dengan *prompt text* yang kita berikan dibuktikan dengan adanya

penurunan skor *Conditional Maximum Mean Discrepancy (CMMD)* saat *pretrained* dibandingkan sesudah *fine tuning* yaitu yang semula sebesar 0,7 saat *pretrained* kemudian turun menjadi 0,5 saat *fine tuning.* Hal tersebut menunjukkan model yang telah di *fine tuned* menghasilkan gambar yang lebih sesuai dengan *prompt.*

Selain evaluasi visual, model juga diuji untuk melihat konsistensi karakter antar ilustrasi. Dalam cerita anak-anak, penting untuk memastikan bahwa karakter utama memiliki bentuk dan warna yang tetap di setiap adegan. Pengujian ini dilakukan dengan menghasilkan beberapa ilustrasi yang menggambarkan berbagai momen dalam cerita, kemudian membandingkan apakah model berhasil mempertahankan desain karakter Si Kancil dan Buaya di setiap gambar. Model yang telah di-*fine-tune* dengan LoRA menunjukkan peningkatan dalam aspek ini, dengan bentuk dan ekspresi karakter yang lebih seragam dibandingkan model sebelum *fine-tuning.*

Analisis tambahan dilakukan menggunakan histogram warna, untuk melihat apakah model yang telah dioptimalkan mampu mempertahankan palet warna yang seragam dalam seluruh ilustrasi cerita. Hasil menunjukkan bahwa model yang telah di-*fine-tune* memiliki distribusi warna yang lebih konsisten, dengan dominasi warna hijau dan biru untuk hutan dan sungai, serta warna cokelat dan oranye untuk karakter Si Kancil. Hal ini menunjukkan bahwa model telah berhasil menyesuaikan diri dengan tema visual yang diinginkan.



Gambar 4. 14 Grafik Perbandingan *MMD Scores* Dari *Test Dataset* dan *Model Fine Tuning*

4.5 Analisis Hasil dan Tantangan Implementasi

Hasil *fine-tuning* menunjukkan bahwa model yang telah dioptimalkan menggunakan LoRA dapat menghasilkan ilustrasi yang lebih sesuai dengan cerita Si Kancil dan Buayas. Detail karakter menjadi lebih jelas, warna lebih harmonis, dan latar belakang lebih mendukung narasi cerita. Selain itu, model yang telah di-*fine-tune* menunjukkan konsistensi karakter antar ilustrasi, yang sangat penting dalam buku cerita anak-anak.

Namun, terdapat beberapa tantangan selama implementasi fine-tuning ini. Salah satu kendala yang muncul adalah distorsi pada gambar tertentu, terutama ketika jumlah *epoch* terlalu tinggi. Distorsi ini terlihat dalam beberapa hasil gambar, di mana karakter mengalami deformasi atau latar belakang terlihat tidak natural. Hal ini menunjukkan bahwa meskipun *fine-tuning* meningkatkan kualitas gambar, jumlah *epoch* yang terlalu tinggi dapat menyebabkan *overfitting* pada dataset latihannya.

Dibandingkan dengan metode *fine-tuning* lainnya seperti *DreamBooth*, LoRA memiliki beberapa keunggulan dan keterbatasan. Dari segi efisiensi, LoRA lebih cepat karena hanya melatih sebagian parameter model, memungkinkan *fine-tuning* dilakukan dengan konsumsi memori yang lebih rendah. Namun, dari segi kontrol terhadap model, *DreamBooth* masih lebih unggul dalam menjaga detail karakter secara menyeluruh. Oleh karena itu, pemilihan metode *fine-tuning* harus disesuaikan dengan kebutuhan proyek, apakah lebih mengutamakan efisiensi atau kualitas hasil akhir.

4.6 Hasil Implementasi Menggunakan *Streamlit*

Aplikasi *AI Image Generator* berhasil diimplementasikan menggunakan framework Streamlit, yang dirancang untuk membangun antarmuka pengguna berbasis web secara sederhana dan interaktif. *Framework* ini memungkinkan pengintegrasian model AI dengan antarmuka pengguna secara mulus, sehingga pengguna dapat menghasilkan gambar berdasarkan deskripsi teks yang dimasukkan.

Pada tampilan utama aplikasi, judul "*AI Image Generator*" ditampilkan di bagian atas sebagai identitas utama aplikasi. Di bawahnya, terdapat kolom *input* berbentuk persegi panjang yang memungkinkan pengguna memasukkan deskripsi atau *prompt*. Kolom ini dilengkapi dengan placeholder bertuliskan "*Please Enter your prompt here!*" untuk memandu pengguna. Tombol "*Generate Image*" ditempatkan di bawah kolom *input*,

dengan desain minimalis berupa latar putih dan teks hitam, membuatnya mudah diakses.

Gambar 4. 15 Generasi Gambar Menggunakan Streamlit

Setelah pengguna menekan tombol "*Generate Image*", aplikasi akan memberikan umpan balik berupa teks status seperti "*Loading Model...*" dan "*Generating Model...*" yang muncul tepat di bawah tombol. Feedback ini dirancang untuk memberi kejelasan bahwa proses sedang berlangsung. Setelah proses selesai, gambar hasil generasi ditampilkan pada area besar di bagian bawah halaman, dengan latar abu-abu yang dirancang agar hasil terlihat jelas.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini telah berhasil menunjukkan bahwa deskripsi karakter berupa teks dapat diterjemahkan menjadi gambar bergaya kartun anak-anak menggunakan *Diffusion Models*. Selain itu, dengan menerapkan *prompt* yang lebih spesifik dan *negative prompt*, model mampu meningkatkan fidelity hasil visualisasi. Teknik ini membantu memastikan gambar yang dihasilkan lebih akurat dan mendekati interpretasi yang diinginkan pengguna, sekaligus mengurangi elemen yang tidak diinginkan. Hasil *Loss* menunjukkan rata-rata skor sekitar 0.1, menunjukkan perbaikan dalam pemahaman dan akurasi visualisasi sehingga dapat menghasilkan gambar berkualitas tinggi. Penelitian ini juga berhasil menghasilkan gambar yang sesuai dengan *prompt text* yang kita berikan dibuktikan dengan adanya penurunan skor *Conditional Maximum Mean Discrepancy (CMMD)* saat *pretrained* dibandingkan sesudah *fine tuning* yaitu yang semula sebesar 0,7 saat *pretrained* kemudian turun menjadi 0,5 saat *fine tuning*. Hal tersebut menunjukkan model yang telah di *fine tuned* menghasilkan gambar yang lebih sesuai dengan *prompt.*

5.2 Saran

Penelitian selanjutnya disarankan untuk mengeksplorasi penggabungan berbagai gaya seni guna menghasilkan visualisasi yang lebih beragam. Penggunaan *dataset* yang lebih luas dan mencakup elemen lingkungan atau latar juga dapat meningkatkan kemampuan personalisasi model. Selain itu, implementasi sistem dengan dukungan bahasa lokal, seperti Bahasa Indonesia, dapat menjadi langkah penting untuk meningkatkan relevansi dan aksesibilitas model dalam konteks lokal. Untuk meningkatkan efisiensi generasi gambar, penelitian berikutnya dapat memfokuskan pada optimasi arsitektur model atau penggunaan perangkat keras yang lebih canggih.

DAFTAR PUSTAKA

Agnese, J., Herrera, J., Tao, H., Zhu, X., Oct, C. V, Komputer, T., Komputer, I., & Atlantic, F. (2019). *Survei dan Taksonomi Jaringan Saraf Tiruan untuk Sintesis Teks-ke-Gambar*. *2016*.

Aqiela, H. S., & Sihombing, R. M. (2023). Analisis Ilustrasi Buku Anak sebagai Media Edukasi Stress dan Depresi kepada Anak. *IMATYPE: Journal of Graphic Design Studies*, *2*(2), 79. https://doi.org/10.37312/imatype.v2i2.7457

Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2024). Medical Image Segmentation Review: The Success of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–38. https://doi.org/10.1109/TPAMI.2024.3435571

Bank, D., Koenigstein, N., & Giryes, R. (2023). Autoencoders. *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook, Third Edition*, 353–374. https://doi.org/10.1007/978-3-031-24628-9_16

Berahmand, K., Daneshfar, F., Salehi, E. S., Li, Y., & Xu, Y. (2024). Autoencoders and their applications in machine learning: a survey. In *Artificial Intelligence Review* (Vol. 57, Nomor 2). Springer Netherlands. https://doi.org/10.1007/s10462-023-10662-6

Chen, S., & Guo, W. (2023). Auto-encoders in deep learning—a review with new perspectives. *Mathematics*, *11*(8), 1777.

Corvi, R., Cozzolino, D., Poggi, G., Nagano, K., & Verdoliva, L. (n.d.). *Sifat gambar sintetis yang menarik : dari jaringan permusuhan generatif ke model difusi*.

Croitoru, F. A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion Models in Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(9), 10850–10869. https://doi.org/10.1109/TPAMI.2023.3261988

La Bania, & Milawaty. (2019). The Value of Lower Class Resistance in Indonesia in the Mousedeer Story. *Mozaik Humaniora*, *19*(2), 172–180. https://e-

journal.unair.ac.id/MOZAIK/article/view/12488

Lembang, I. R., Ade, D., Riyadhi, N., Tiyas, M., & Dk, M. (2022). Teknik Ilustrasi Digital Freehand Dalam Pembuatan Buku Cerita Bergambar “Friends” Untuk Anak Usia Dini. *Prosiding Seminar Nasional Tetamekraf*, *1*(2), 46.

Li, X., Hou, X., & Loy, C. C. (2023). *When StyleGAN Meets Stable Diffusion: a $\mathscr{W}_+$ Adapter for Personalized Image Generation*. 2187–2196. https://doi.org/10.1109/CVPR52733.2024.00213

Luo, S. (2023). *Technical Report LCM-LORA: A UNIVERSAL STABLE-DIFFUSION ACCELERATION MODULE*. 1–7.

Park, J., Ko, B., & Jang, H. (2023). StyleBoost: A Study of Personalizing Text-to-Image Generation in Any Style using DreamBooth. *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)*, 93–98.

Ren, C. X., Ge, P., Dai, D. Q., & Yan, H. (2021). Learning Kernel for Conditional Moment-Matching Discrepancy-Based Image Classification. *IEEE Transactions on Cybernetics*, *51*(4), 2006–2018. https://doi.org/10.1109/TCYB.2019.2916198

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). *DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation*. 22500–22510. https://doi.org/10.1109/cvpr52729.2023.02155

Shen, D., Song, G., Xue, Z., Wang, F.-Y., & Liu, Y. (2024). *Rethinking the Spatial Inconsistency in Classifier-Free Diffusion Guidance*. 9370–9379. http://arxiv.org/abs/2404.05384

Song, H., Huang, S., Dong, Y., & Tu, W.-W. (2023). *Robustness and Generalizability of Deepfake Detection: A Study with Diffusion Models*. 1–8. http://arxiv.org/abs/2309.02218

Waryanti, E., Puspitoningrum, E., Violita, D. A., & Muarifin, M. (2021). Struktur Cerita Anak Dalam Cerita Rakyat Timun Mas dan Buto Ijo Dalam Saluran Youtube Riri Cerita Anak Interaktif (Kajian Sastra Anak). *Prosiding SEMDIKJAR (Seminar Nasional Pendidikan dan Pembelajaran)*, *4*, 12–29.

Yao, D. (2024). *Risks When Sharing LoRA Fine-Tuned Diffusion Model Weights*.

1–37. http://arxiv.org/abs/2409.08482

Yoo, H. (2024). *Fine Tuning Text-to-Image Diffusion Models for Correcting Anomalous Images*. http://arxiv.org/abs/2409.16174

Zhang, E., Wang, K., Xu, X., Wang, Z., & Shi, H. (2023). *Forget-Me-Not: Learning to Forget in Text-to-Image Diffusion Models*. 1755–1764. http://arxiv.org/abs/2303.17591