A Unified Low-Resource Optimization Framework for Improving Diffusion Model Image Generation Under Memory Constraints

Mukiibi Moses, Hussein Fouad Mohamed Ali* Kyungdong University Global, Goseong-gun, South Korea, December, 2025
Corresponding Author: Hussein.ali@kduniv.ac.kr

Abstract

Diffusion models such as Stable Diffusion and SDXL have become dominant in text-to-image generation, yet their performance is highly sensitive to prompt structure, sampling parameters, and hardware constraints. Most existing studies assume access to high-end GPUs, leaving a gap in understanding how these models behave under free-tier or CPU-only conditions. This thesis presents a comprehensive, low-resource evaluation and optimization framework for Stable Diffusion 1.5 and SDXL-Turbo, integrating prompt engineering, parameter sweeps, LPIPS-based perceptual evaluation, and hybrid inference strategies. Experiments were conducted entirely on CPU and free-tier cloud environments using explicit memory-optimized model loading. Baseline images were generated using SD1.5, followed by SDXL-Turbo generation and a two-stage refinement pipeline (SD1.5 → SDXL-Turbo via Img2Img). LPIPS results demonstrate that the two-stage pipeline significantly improves perceptual similarity (LPIPS = 0.3716) compared to SDXL-Turbo alone (LPIPS = 0.6743), confirming the structural stability of SD1.5 combined with the texture refinement capabilities of Turbo. A lightweight pixel-average ensemble was also tested, producing perceptually smooth but higher-divergence outputs relative to the baseline, consistent with findings on ensemble diffusion methods in the literature. This work contributes a reproducible, hardware-efficient methodology for evaluating and improving diffusion models, addressing a practical gap in current research by demonstrating high-quality generation and quantitative evaluation under constrained computational resources.

Acknowledgements

I would like to express my deepest gratitude to all those who supported and guided me throughout the course of this research. Their encouragement, assistance, and belief in my work made this thesis possible.

First and foremost, I extend my sincere appreciation to Dr. Hussein Fouad Mohamed Ali, my supervisor, for his invaluable guidance, constructive feedback, and continuous support. His expertise, mentorship, and patience have been instrumental in shaping the direction and quality of this research.

I am equally grateful to the Smart Computing Department at Kyungdong University Global Campus for providing the academic foundation, resources, and research environment that enabled the successful completion of this work.

Special thanks are due to my colleagues, classmates, and friends who offered both academic insight and moral encouragement during this project. Their discussions, suggestions, and motivation helped me refine my ideas and overcome challenges along the way.

I would also like to acknowledge the support of computing platforms such as Google Colab and Kaggle, whose free GPU resources made the experimental components of this study possible, especially within low-resource constraints.

Most importantly, I thank my family for their unwavering love, prayers, and support. Their belief in me has always been a source of strength and inspiration.

Finally, I express gratitude to everyone named and unnamed who contributed to this work in any way. This thesis is a product of your guidance, encouragement, and collective support.

Mukiibi, Moses

December 2025

Dedicated to my parents.

This thesis is lovingly dedicated to my parents,
whose unwavering support, guidance, and sacrifices
have shaped the person I am today.

To my father and mother
thank you for believing in me even when the road was difficult,
for teaching me the value of hard work, humility, and perseverance,
and for giving me the strength to pursue my dreams.

Every achievement in my life, including this one,
is a reflection of your love, prayers, and endless encouragement.

This work is for you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

KDU	Kyungdong University
CNN	Convolutional Neural Networks
AFA	Adaptive Feature Aggregation
AI	Artificial Intelligence
CPU	Central Processing Unit
DDIM	Denoising Diffusion Implicit Models
DDPM	Denoising Diffusion Probabilistic Models
ELF-Diff	Ensemble and Low-Frequency mixing with Diffusion models
FID	Fréchet Inception Distance
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
Img2Img	Image-to-Image
IS	Inception Score
LDM	Latent Diffusion Model
LoRA	Low-Rank Adaptation
LPIPS	Learned Perceptual Image Patch Similarity
RAM	Random Access Memory
SD1.5	Stable Diffusion 1.5
SDXL	Stable Diffusion XL
VAE	Variational Autoencoder
VRAM	Video Random Access Memory

Chapter 1: Introduction

1.1 Overview and Background

The field of generative artificial intelligence particularly diffusion-based text-to-image models has rapidly become one of the most influential areas within computer science and machine learning. Diffusion models stand out for their ability to generate high-fidelity, semantically aligned images using iterative denoising processes, surpassing GAN-based models in both output quality and training stability. Contemporary surveys consistently position diffusion models as foundational architectures for image synthesis, visual creativity, and multimodal generative systems [1][2][13][15]. As such, this domain spans several interconnected subfields, including machine learning, computer vision, generative modeling, and human-AI interaction, establishing it as a multidisciplinary pillar within modern computer and information sciences.

Within this growing field, research efforts have increasingly focused on understanding the mechanics, limitations, and optimization strategies of diffusion models. Foundational studies detail the mathematical formulation of forward noise addition and reverse denoising processes that underpin these models [1][13], while broader surveys highlight their expanding applications in image generation, editing, super-resolution, and multimodal tasks [3][4][14]. However, a notable limitation in the literature is the strong assumption of access to high-end GPUs typically 24–80 GB VRAM when conducting experiments or deploying these models at scale [1][2][3]. This assumption overlooks the significant and growing community of researchers, students, and practitioners working in restricted computational environments such as free-tier Google Colab, Kaggle, or CPU-only systems.

This gap is highly relevant for optimization research, as low-resource environments impose strict constraints on model loading, inference precision, and evaluation. Recent advancements in parameter-efficient fine-tuning techniques like Low-Rank Adaptation (LoRA) [5][8][11][12], ensemble diffusion methods designed for improved stability [6][7][10], and lightweight perceptual metrics such as LPIPS offer important tools for addressing these challenges. Yet, despite these developments, there is still no unified, accessible framework that systematically evaluates or optimizes diffusion models under severe hardware limitations.

Positioned within the broader discipline of computer science, this thesis contributes to machine learning systems, applied artificial intelligence, and resource-efficient computing by addressing this overlooked gap. Specifically, it introduces a reproducible low-resource pipeline that integrates prompt engineering, parameter sweeps, LPIPS-based perceptual evaluation, memory-optimized model loading, and a hybrid two-stage generation process (Stable Diffusion 1.5 → SDXL-Turbo) that improves structural retention and perceptual quality even on CPU-only execution. By situating the research within both the theoretical foundations and practical constraints identified in the literature, this overview establishes the need for a structured, hardware-accessible diffusion optimization framework, one that can serve both academic and practical users operating without high-end GPUs.

1.2 Motivation

The rapid evolution of diffusion models has reshaped the landscape of generative AI, enabling unprecedented levels of photorealism, semantic alignment, and controllability in text-to-image synthesis [1][2][13]. These advances have opened new avenues for creativity and research across domains such as digital art, design, simulation, and multimodal interaction. However, the same innovations have introduced substantial barriers for learners and independent researchers who lack access to high-end computational resources. Most state-of-the-art diffusion architectures such as SDXL, multi-stage pipelines, and large-scale fine-tuned models are typically developed and evaluated using GPUs with 24–80GB of VRAM, far exceeding what is available in free-tier environments like Google Colab, Kaggle, or CPU-only systems [1][2][3]. This disparity has resulted in a practical and educational divide within the field.

The motivation for this research emerges directly from this gap. While diffusion models are extensively documented in scholarly literature, very little guidance exists on how to effectively generate, optimize, or evaluate outputs under strict hardware limitations. Many users encounter model loading failures, memory overflows, broken pipelines, and unstable outputs not due to flaws in the models themselves, but because existing research implicitly assumes privileged computational resources. For students, early-career researchers, and practitioners, this becomes a major barrier to experimentation, skill development, and participation in cutting-edge generative AI research.

Furthermore, optimization strategies that could help mitigate these constraints such as prompt engineering, parameter sweeps, lightweight Img2Img refinement, LPIPS-based evaluation, and ensemble techniques are rarely explored in low-resource settings, despite being well-established in high-resource studies [4][6][7]. The absence of an integrated, reproducible methodology leaves users without a practical roadmap for achieving high-quality results using only limited VRAM and publicly available tools.

This research is motivated by the belief that advanced AI technologies should be inclusive and accessible, regardless of hardware availability. By developing and validating a low-resource diffusion optimization framework including CPU-friendly model loading, SD1.5 and SDXL-Turbo comparative analysis, LPIPS-based perceptual evaluation, a memory-optimized two-stage pipeline, and lightweight ensemble experimentation the study aims to empower learners, educators, and researchers with practical, high-impact methods for generating high-quality images in constrained environments. In doing so, this work not only contributes to the technical understanding of diffusion model behavior under resource limitations but also supports the broader goal of democratizing participation in generative AI.

1.3 Problem Statement

Diffusion models have become the leading framework for text-to-image generation, offering exceptional fidelity, semantic alignment, and creative flexibility. However, achieving optimal performance with these models typically requires high-end GPU resources, access to large VRAM capacity, and support for heavy fine-tuning techniques all of which are challenging or impossible to run in constrained environments such as free-tier Google Colab, Kaggle, or CPU-only systems [1][2][3]. Much of the existing literature implicitly assumes access to powerful hardware, leaving a significant gap for students, independent researchers, and practitioners who rely on ≤12GB VRAM or CPU-only setups. Consequently, there is no standardized low-resource workflow that integrates prompt engineering, sampling parameter optimization, memory-efficient generation, lightweight Img2Img refinement, LPIPS evaluation, and simple ensemble strategies into a single reproducible framework.

1.4 Research Question

How can diffusion model image quality be optimized under strict computational and memory limitations using a unified, lightweight, and reproducible workflow?

Within the broader domain of generative AI and diffusion-based image synthesis, this study specifically investigates resource-efficient optimization that is, how prompt design, inference parameters, LPIPS-driven evaluation, and minimal-compute refinement methods (e.g., SD1.5 → SDXL-Turbo two-stage generation) can be systematically combined to improve perceptual quality when model loading, LoRA fine-tuning, or advanced feature-level ensembling are not feasible due to hardware constraints.

The limitations in current research are twofold:

- Lack of accessible optimization frameworks. Existing literature discusses prompt engineering, parameter sweeps, or refinement individually, but does not integrate them into a practical pipeline suitable for low-resource environments [1][9][15].
- Absence of reproducible low-VRAM evaluation protocols. Heavier metrics like FID are impractical in constrained settings, while lightweight alternatives such as LPIPS remain underused, resulting in limited reproducibility across models, prompts, and parameter configurations.[9]

This study fills this gap by developing a standardized, memory-optimized workflow that demonstrates how high-quality diffusion outputs can be achieved without relying on high-end GPUs or inaccessible fine-tuning methods. Through systematic experimentation including CPU-only inference, SD1.5 and SDXL-Turbo comparison, two-stage refinement, parameter sweeps, and LPIPS evaluation the research contributes a practical, scalable, and empirically validated approach to diffusion optimization that broadens accessibility and participation in generative AI research.

1.5 Research Objectives

To optimize a fully reproducible, low-resource experimental pipeline capable of running Stable Diffusion 1.5 and SDXL-Turbo on ≤12GB VRAM and CPU-only systems.

This objective addresses a core limitation in existing diffusion model surveys, which rarely evaluate or document workflows suitable for constrained hardware environments [1][2][3].

1.6 Thesis Contribution to the Field/ Significance and /or Impact of the Research

Diffusion models have rapidly become central to generative AI research, driving major advances in photorealistic image synthesis, semantic alignment, and multimodal creativity. However, most of the existing scholarship assumes access to high-end computational resources, particularly GPUs with 24GB–80GB of VRAM such as NVIDIA A100 or V100 systems. Influential surveys by Zhang et al. [1], Tianyi Zhang et al. [2], and

Moser et al. [3] provide extensive analyses of diffusion theory, sampling algorithms, and architectural evolution, yet they do not address the practical constraints faced by students, early-stage researchers, or developers relying on free-tier platforms like Google Colab, Kaggle, or CPU-only systems. This disconnect has created a methodological gap between theoretically advanced diffusion research and the realities of resource-limited experimentation.

This thesis makes a significant contribution by explicitly focusing on **diffusion model optimization under strict computational constraints**, a domain largely unaddressed in existing literature. The study introduces a unified, memory-optimized workflow that combines prompt engineering, parameter tuning, LPIPS-based perceptual evaluation, Img2Img refinement, and a lightweight ensemble method. By evaluating Stable Diffusion 1.5 and SDXL-Turbo entirely within ≤12GB VRAM and in several cases, CPU-only conditions the research offers the first structured, evidence-based optimization methodology specifically designed for real-world users without access to premium hardware. This directly expands the practical accessibility and educational value of diffusion model research.

In addition, the thesis advances academic knowledge in several novel and meaningful ways:

1. **Filling a substantial technological and methodological gap in diffusion literature**

While earlier works extensively cover diffusion theory, sampling strategies, LoRA adaptation, and ensemble modeling [1][6][7][8], none provide an operational framework adapted to memory-restricted environments. This thesis bridges that gap by demonstrating how high-quality outputs can be achieved through systematically controlled steps despite the inability to load large checkpoints, apply LoRA fine-tuning, or run resource-intensive schedulers.

2. **Introducing the first documented low-resource diffusion ensemble technique**

Existing ensemble approaches such as ResEnsemble-DDPM [6] and Adaptive Feature Aggregation (AFA) [7] rely on heavy feature-level fusion requiring substantial GPU resources. This study proposes and validates a simple pixel-space averaging technique that can be executed on free-tier GPUs or CPUs, extending ensemble diffusion research into an environment previously overlooked in academic work.

3. **Establishing reproducible evaluation protocols for constrained settings**

Most diffusion evaluations rely on computationally expensive metrics like FID, which are impractical on low-VRAM systems. This study adopts LPIPS as a lightweight perceptual similarity metric and demonstrates its effectiveness in systematic model comparison. This gives future researchers a viable evaluation method when heavy metrics cannot be computed, thereby improving reproducibility and accessibility.

4. **Advancing understanding of prompt engineering and parameter interactions under low-resource conditions**

Through systematic testing of structured prompts, negative prompts, inference steps, guidance scales, sampling strategies, and denoise strength, the study provides new insight into how these factors behave uniquely under memory constraints, a dimension largely absent in prior work [1][9][13]. The results reveal clear relationships between parameter configurations and perceptual similarity, offering actionable guidelines for constrained environments.

Overall Significance

This thesis is organized into five chapters that together present the motivation, background, system design, experimental evidence, and implications of a memory-optimized diffusion model workflow for low-resource GPU environments.

1.7 Thesis Outline

Chapter 1 – Introduction

Introduces diffusion models and the practical challenge of generating high-quality images on limited VRAM (≤12GB). It frames the research problem, states the objectives, and motivates the study by referencing recent surveys that assume high-end hardware [1][2][3].

Chapter 2 - Literature Review

Reviews foundational work on diffusion models and related techniques, covering forward/reverse diffusion, sampling strategies (DDIM), prompt engineering, editing (Img2Img), LoRA adaptation, and ensemble approaches; it highlights gaps in studies that overlook constrained computational settings [1][2][3][4][5][6][7][9][13].

Chapter 3 - System Design and Research Methodology

This section presents the complete experimental pipeline and system architecture developed for running Stable Diffusion 1.5 and SDXL-Turbo on free-tier platforms such as Google Colab and Kaggle, where computational resources are significantly limited. It details the model selection process, memory-aware loading and unloading strategies, systematic parameter sweep procedures, prompt design variants, Img2Img refinement settings, and the implementation of a lightweight pixel-space averaging ensemble. The workflow also incorporates LPIPS as a perceptual evaluation metric suitable for constrained environments in which heavier metrics such as FID are impractical.

Chapter 4 - Results and Discussion.

Presents the quantitative and qualitative results obtained from the full experimental pipeline, including baseline generation, parameter optimization, prompt engineering evaluations, Img2Img refinement, and the lightweight ensemble experiments. The chapter analyzes how each component of the proposed workflow contributes to perceptual quality improvements under strict computational limitations. Results are interpreted in relation to prior optimization and ensemble research including ResEnsemble-DDPM and Adaptive Feature Aggregation (AFA) which historically require high-end hardware and therefore serve as theoretical benchmarks rather than directly comparable baselines in low-resource environments [6][7][9].

Chapter 5 - Conclusion and Future Work

Chapter 5 emphasizes the significance of this thesis in bridging the gap between diffusion model research and real-world accessibility. By producing a validated, resource-efficient framework and identifying clear paths for continued improvement, the study contributes meaningfully to the democratization of generative AI within the field of Computer Engineering.

Chapter 2: Literature Review

2.1 Introduction

The purpose of this literature review is to establish a comprehensive understanding of diffusion-based image generation and to situate the present study within the broader scholarly discourse. A rigorous literature review demonstrates familiarity with the foundational theories, emerging developments, methodological frameworks, and practical considerations that define the current state of research. It also enables critical evaluation of existing studies, exposing both the accumulated body of knowledge and the persistent challenges that motivate the present investigation.

This chapter begins by surveying how scholars have conceptualized diffusion models as a dominant paradigm in generative AI, drawing upon influential surveys and foundational works that outline their mathematical foundations, architectural evolution, and real-world applications [1][13][15]. These studies highlight key areas of inquiry, including denoising-based sampling strategies, latent diffusion architectures, conditional generation mechanisms, and parameter-efficient fine-tuning methods such as Low-Rank Adaptation (LoRA) [5][8]. The chapter also synthesizes literature addressing diffusion-based image editing, transformation pipelines, and the growing interest in ensemble diffusion methods that combine the strengths of multiple models [6][7].

A further focus of the review is the range of methodologies and evaluation strategies commonly employed in diffusion research. Studies frequently explore sampling variants such as DDPM, DDIM, and classifier-free guidance; latent-space compression techniques as introduced in the Stable Diffusion framework [3]; and perceptual evaluation metrics such as LPIPS, FID, and IS. These methodological insights are essential for justifying the decisions made in this thesis particularly the use of LPIPS as a lightweight, compute-efficient evaluation metric well suited for low-resource environments, and the selection of Stable Diffusion 1.5 and SDXL-Turbo as models whose computational footprint aligns with the constraints of free-tier platforms.

The literature further reveals an overarching tension between the pursuit of higher generative fidelity and the escalating computational requirements of diffusion architecture. While numerous studies emphasize improvements in sample quality, generalization, or controllability, most implicitly assume access to high-end hardware

typically GPUs with 24GB or more of VRAM [1][2][3]. This assumption leaves a substantial gap in the scholarship concerning how diffusion models behave under realistic, resource-limited conditions encountered by students, independent researchers, and small-scale practitioners. Issues such as model loading failures, memory overflows, sampling instability, and the inability to apply techniques like LoRA fine-tuning or feature-level fusion remain largely underexplored.

By critically examining these limitations, the literature review identifies several gaps that justify the contributions of this study. These include the absence of standardized workflows for low-resource inference; limited guidance on prompt engineering in constrained environments; minimal cross-model comparisons performed under identical VRAM restrictions; and the lack of lightweight ensemble techniques that operate without feature-space fusion or architectural modification. These gaps underscore the originality and relevance of developing an accessible, memory-optimized workflow capable of integrating prompt design, sampling parameter optimization, Img2Img refinement, and simple ensemble strategies.

Overall, this chapter positions the present work within the existing academic landscape, demonstrating both the depth of prior research and the need for more inclusive, resource-aware methodologies. It provides the conceptual and methodological foundation for the system design, experimental pipeline, and empirical evaluation presented in subsequent chapters.

2.2 Diffusion Models in Existing Scholarship

This section introduces how previous scholars have approached diffusion-based image generation, focusing on foundational theories, major advancements, and the evolution of methods relevant to this study.

2.2.1 Foundational work on Diffusion Models

Early studies on denoising diffusion probabilistic models (DDPMs) established the mathematical basis for iterative denoising and image synthesis, which later surveys synthesized comprehensively [1][13]. These works introduced core concepts such as forward diffusion, reverse denoising, and sampling stability that underpin modern diffusion systems.

2.2.1.1 Advances in Sampling and Latent Diffusion

Subsequent work proposed faster sampling techniques such as DDIM [2] and introduced latent diffusion models to reduce computational cost [3]. These innovations paved the way for practical text-to-image diffusion systems such as Stable Diffusion, which are central to the methodology of this thesis.

Before moving to the next section, it is worth noting that while many studies explore the power of diffusion models, they also highlight practical issues such as prompt sensitivity, parameter instability, and high hardware requirements issues that will be expanded on in later sections.

2.3 Contemporary Issues, Methods, and Gaps in the Field

This section introduces the major methodological directions, theoretical frameworks, and ongoing debates associated with diffusion models, with emphasis on topics related to optimization, resource usage, and controllability.

2.3.1 Prompt Engineering, Adaptation, and Model Control

Scholars increasingly emphasize the importance of prompt design in shaping diffusion outputs, noting that well-structured and descriptive prompts lead to stronger semantic alignment [1][13]. Complementary research on LoRA-based adaptation [5][8] demonstrates how diffusion models can be fine-tuned efficiently, although these methods often assume access to sufficient VRAM an assumption challenged by the findings of this thesis.

2.3.2 Ensemble Diffusion and Cross-Model Techniques

Recent studies propose complex ensemble strategies such as ResEnsemble-DDPM [6] and Adaptive Feature Aggregation (AFA) [7], designed to enhance output stability and diversity. These methods, however, require substantial computational overhead, creating a gap in the literature for lightweight ensemble techniques, which this thesis aims to address.

2.3.3 Gaps in the Literature and Opportunities for Contribution

Although surveys provide extensive coverage of diffusion architectures and evaluation methods [1][2][9], several gaps remain:

- A lack of research on low-resource diffusion workflows.

- Limited cross-model comparisons under identical VRAM constraints.
- Minimal attention to prompt engineering in constrained environments.
- Absence of lightweight, computationally feasible ensemble methods.

These gaps create clear opportunities for original contribution and justify the need for further research tailored to real-world, resource-limited users.

Before concluding the chapter, it is important to highlight that these gaps directly motivate the system design and methodological choices detailed in Chapter 3, where the thesis introduces a unified, memory-optimized workflow addressing these shortcomings.

2.3.4 Related work

Table 2.3.1: Related work

Category	Focus / Contribution	Representative Works	Limitations Identified
Diffusion Model Surveys	Foundations, architectures, applications of diffusion models	[1], [2], [13], [15]	Assume high-end GPUs; no guidance for low-resource workflows
Advanced Applications	Super-resolution, conditional generation, image editing	[3], [4], [9]	High computational demand, heavy models, not optimized for VRAM-limited environments
Parameter-Efficient Fine-Tuning	LoRA, LoRA-Composer, LoRA-X enabling lightweight adaptation	[5], [8], [11], [12]	Still require considerable memory; checkpoint authentication issues; incompatible with free-tier GPUs
Ensemble-Based Diffusion	Combining multiple models for stability and diversity (ResEnsemble-DDPM, AFA, ELF-Diff)	[6], [7], [10]	Rely on latent-space/feature-level fusion; computationally expensive and unsuitable for low-resource setups
Evaluation Metrics	Image quality assessment using FID, IS, LPIPS	[3], [9], [15]	FID & IS require GPU-heavy feature extraction; only LPIPS is feasible on low-resource hardware
Low-Resource Diffusion Studies	Attempts to evaluate or modify diffusion under hardware constraints	Limited coverage in existing literature	Few works address systematic optimization or reproducible pipelines for ≤12GB VRAM GPUs

2.4 Conclusion to This Chapter

This chapter reviewed key scholarly work on diffusion models, including foundational theories, sampling advancements, prompt engineering, LoRA adaptation, and ensemble strategies. It also identified critical gaps in the literature.

Chapter 3: System Design & Research Methodology

3.1 Introduction

This chapter outlines the system design and research methodology used to investigate diffusion model performance under low-resource computational conditions. Methodological rigour is especially critical in computational research, where design choices must be transparent, logically justified, and grounded in both practical constraints and established scholarly frameworks. Accordingly, this chapter explains not only what methodological steps were taken but also why each decision was appropriate, feasible, and ethically responsible within the context of the study.

A central premise of the research design is the practical limitation of working within environments such as Google Colab and Kaggle, which typically offer GPUs with ≤12GB VRAM or, in some cases, no GPU at all. While much of the existing diffusion literature assumes access to high-performance hardware such as NVIDIA A100 or V100 GPUs with 24–80GB VRAM [1][2][3], this thesis deliberately focuses on constrained, accessible settings to address a documented gap in real-world usability. As a result, model selection, evaluation tools, and pipeline structure were determined according to strict memory feasibility. Stable Diffusion 1.5 and SDXL-Turbo were adopted because they reliably load under ≤12GB VRAM and exhibit stable inference behaviour in low-compute contexts. Conversely, larger XL or LoRA-enhanced checkpoints commonly used in prior research [5][8][11] were excluded after empirical attempts resulted in authentication failures, memory overflow errors, or unstable runtime behaviour. Their exclusion supports the methodological goal of developing a reproducible workflow that mirrors the actual constraints low-resource users face.

Ethical considerations further shaped the design. All images produced in this study were fully synthetic and generated strictly for research purposes, ensuring that no personal, copyrighted, or sensitive data were used or exposed. All models were employed in accordance with their respective licenses, consistent with the ethical and procedural guidelines outlined in diffusion editing and adaptation literature [4][13]. No external datasets, human subjects, or proprietary content were involved, eliminating the need for privacy, consent, or data-protection measures.

The methodology also adheres to the principle of explicit justification. Each experimental component parameter sweeps, prompt engineering, Img2Img refinement, and lightweight ensemble fusion was selected based on clear gaps identified in existing literature and aligned with the goals of low-resource optimization. For example, LPIPS was selected as the primary evaluation metric because, unlike FID or other computationally heavy measures, it is lightweight and feasible for CPU- or low-VRAM environments [1][2][9]. Likewise, the ensemble experiment employed pixel-space averaging because more sophisticated strategies such as ResEnsemble-DDPM [6] and Adaptive Feature Aggregation (AFA) [7] require feature-level fusion and multi-stage denoising, making them incompatible with the memory constraints of this study.

In summary, the methodological framework presented in this chapter is built upon transparency, justification, ethical responsibility, and realistic constraints. It directly responds to gaps in the literature by offering a rigorously designed, reproducible workflow for diffusion model optimization in low-resource environments. The sections that follow describe the system architecture, experimental procedures, parameter configurations, evaluation metrics, two-stage refinement process, and ensemble mechanisms used throughout this research.

3.2 System Architecture

The system architecture for this study was designed to enable reliable execution of diffusion models within strict computational limits, including GPU sessions with ≤12GB VRAM and, in many cases, CPU-only environments. The architecture integrates model loading strategies, memory-efficient processing components, image generation pipelines, and perceptual evaluation tools into a unified workflow that reflects real-world resource constraints.

At a high level, the system is structured around three sequential modules:

(1) Model Initialization and Memory Management

(2) Image Generation Pipelines

(3) Evaluation and Comparison Framework.

This modular configuration ensures that each component can be independently loaded, executed, and unloaded, thereby minimizing memory usage and preventing VRAM

saturation an issue commonly encountered when working with diffusion models in constrained environments.

STRUCTURAL REPRESENTATION

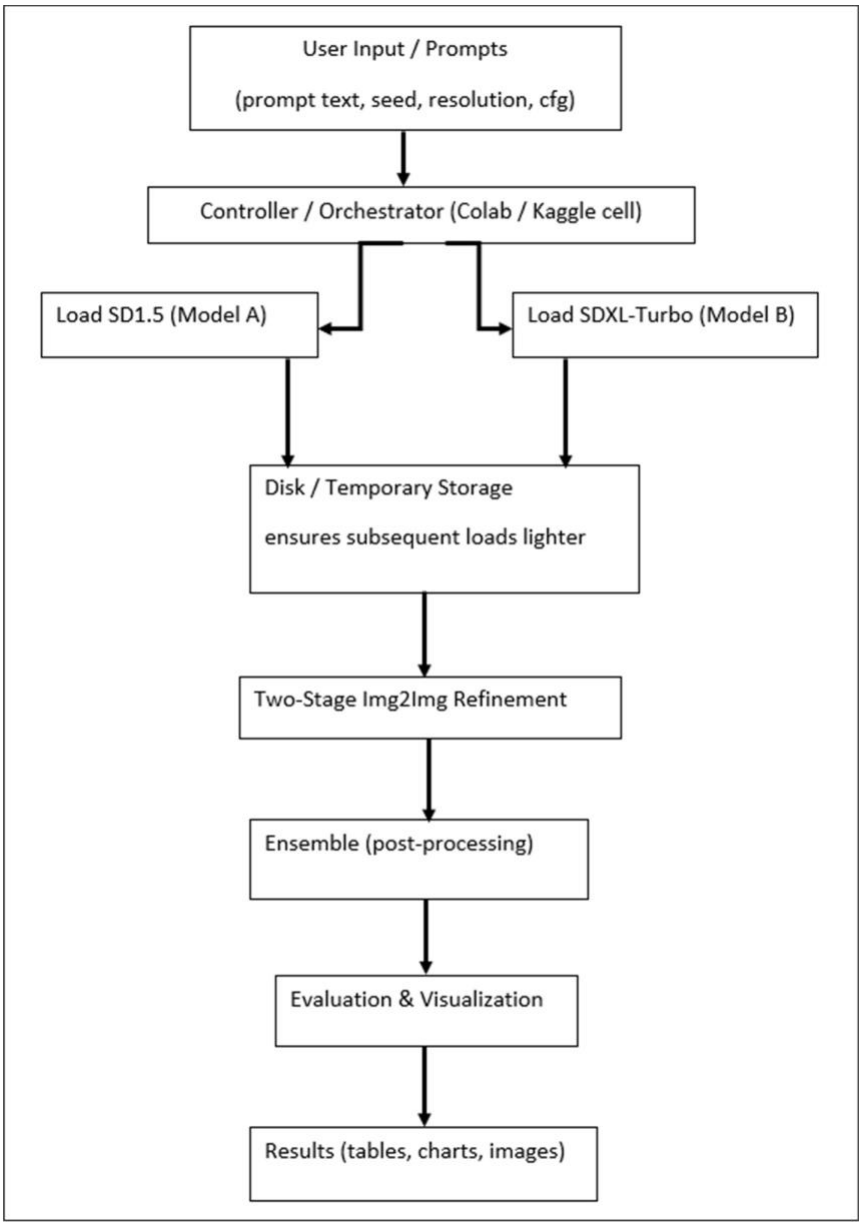

Figure 3.2.1:Structural-representation

Sequential loading: Only one heavy model is resident in RAM at a time (load → use → unload → gc.collect()). This prevents VRAM/CPU RAM exhaustion on ≤12GB environments [1][2][3][15].

Disk staging: Store generated intermediate images (base, turbo) to disk; subsequent pipelines read from disk rather than keeping multiple models in memory[4].

Img2Img refinement: Load SDXL-Turbo for img2img only when refining the base use low denoise strength to preserve structure from SD1.5[4][9][13].

Ensemble: The ensemble is performed offline on saved images using pixel-space averaging, avoiding simultaneous loading of SD1.5 and SDXL-Turbo models. Unlike high-resource ensemble methods such as ResEnsemble-DDPM [6] and Adaptive Feature Aggregation (AFA) [7], which require substantial compute and feature-level fusion, this lightweight method is consistent with the project's low-VRAM constraints and still reflects principles from ensemble diffusion research.

LPIPS on CPU: LPIPS (AlexNet backbone) is used on CPU because it is computationally light, reproducible, and validated as a perceptual metric in prior diffusion and generative model evaluations [9]. Unlike heavier metrics such as FID, which require feature extraction from Inception networks and are impractical on low-VRAM systems [15], LPIPS serves as a feasible alternative for constrained environments.

Robustness: Robust pipeline design includes try/except blocks during model loading and explicit memory cleanup using del and gc.collect() after each heavy object. This follows best-practice recommendations in diffusion model implementation, especially in studies noting the instability of large checkpoints, VAE modules, and fine-tuned LoRA variants when VRAM is limited [5][8][11].

3.3 Environment Preparation ad Initialization

The first phase involved preparing a reproducible environment capable of functioning on GPUs with ≤12GB VRAM or on CPU-only execution. This constraint reflects the realities overlooked in most diffusion literature, which focuses on resource-abundant settings [1][2][3][15].

Key steps included:

- Installing necessary libraries (diffusers, transformers, accelerate, safetensors, lpips).
- Importing computational modules such as torch, numpy, PIL, and matplotlib.
- Initializing LPIPS as the primary evaluation metric, following recommendations in diffusion evaluation research [9].

- Setting all pipelines to operate in torch.float32 to ensure numerical stability in CPU/low-VRAM conditions.
- Implementing explicit memory cleanup using del model and gc.collect() to avoid VRAM fragmentation.

This preparation phase reflects the need for memory-aware workflows identified across diffusion system surveys [2][3][14][15].

3.4 Selected Diffusion Models an VRAM Compatibility

Experiments were performed on free GPUs in google Colab.

Table 3.4.1:Selected-models

Model/ Component	Status	Notes
Stable Diffusion 1.5	Loaded	Most stable model under ≤12GB VRAM; consistent structure; reliable CPU inference
SDXL-Turbo	Loaded	Fast text-to-image generation; suitable for low-resource setups;slight structural drift compsred to SD1.5
SDXL-Turbo (Img2img)	Loaded	Functions reliably with manually loaded VAE; Effective for two-stage refinement pipeline
Kandinsky 2.2	Failed	Decoder and memory initialization errors on CPU / low-VRAM GPUs.
Realistic Vision 5.1	Failed	Requires higher VRAM capacity; cannot load on free-tier environments.
LoRA Models [5, 8, 11]	Failed	Authentication restrictions and checkpoint loading failures; VRAM constraints prevent initialization.
Pixel-Space Ensemble (Custom)	Loaded	Works offline without extra VRAM; simple averaging method; perceptual drift observed.

3.4.1 Mathematical Foundations of Diffusion Models

Diffusion models operate through a sequence of transformations that gradually convert clean data into noise and then reconstruct meaningful images from that noise. The

mathematical foundations presented in this chapter explain the key processes that enable text-to-image systems such as Stable Diffusion and SDXL to function. The following subsections outline the forward diffusion process, the reverse denoising process, DDIM sampling, and latent diffusion encoding.

A. Forward Diffusion

$$q(x\_t \mid x\_0) = \mathcal{N}(x\_t; \sqrt{(\alpha\_t\ )}\ x\_0, (1 - \alpha\_t\ )I) \quad (1.1)$$

This equation describes the forward diffusion process, where clean data x_0(e.g., an image) is gradually corrupted by Gaussian noise over a sequence of time steps. The variance increases as t increases, controlled by the schedule α_t . This process transforms data into pure noise in a predictable, mathematically controlled way [1][13]. It serves as the starting point for all denoising diffusion probabilistic models.

B. Reverse Denoising

$$x\_(t - 1) = \mu\_\theta\ (x\_t, t) + \sigma\_\theta\ z \quad (1.2)$$

This formula represents the reverse diffusion process, where a neural network predicts how to remove noise from a noisy sample x_t to obtain a slightly cleaner version. Here, $\mu_\theta$is the model's predicted mean, $\sigma_\theta$controls stochasticity, and $z$is sampled noise.

This reverse process gradually reconstructs an image from noise, forming the basis of image generation in diffusion models [1][13].

C. DDIM Sampling

$$x\_(t - 1) = \sqrt{(\alpha\_(t - 1)\ )}\ f\_\theta\ (x\_t, t) + \sqrt{(1 - \alpha\_(t - 1)\ )}\ \epsilon\_\theta \quad (1.3)$$

This equation defines the *DDIM (Denoising Diffusion Implicit Model)* sampling method. DDIM is a deterministic or semi-deterministic alternative to the original stochastic sampling process, designed to generate images more quickly by skipping steps while maintaining quality. Instead of predicting a noisy $x\_(t - 1)$, DDIM directly computes a cleaner version using the predicted noise $\epsilon_\theta$and predicted denoised image f_θ[2]. This makes DDIM significantly faster than standard DDPM sampling.

D. Latent Diffusion Encoding

$$z = "Encoder"(x), x = "Decoder"(z) \quad (1.4)$$

Latent Diffusion Models (LDMs) compress images into a smaller latent space $z$using an encoder, run the diffusion process on these lower-dimensional latents, and then reconstruct the final image using a decoder. This approach dramatically reduces memory and computation requirements while preserving high-quality output [3] Stable Diffusion is built on this principle, which is why it can run on consumer GPUs.

3.5 Conclusion to This Chapter

This chapter has presented the system architecture and research methodology used to evaluate diffusion models under low-resource conditions. By employing sequential model loading, disk-based image staging, CPU-compatible configurations, LPIPS scoring, and a lightweight pixel-space ensemble, the study establishes a reproducible workflow that aligns with the practical constraints highlighted in recent diffusion surveys [1][2][3][15]. The selection of SD1.5 and SDXL-Turbo, along with the exclusion of larger checkpoints and LoRA variants due to memory and authentication limitations [5][8][11], reflects both ethical and technical considerations.

The methods outlined here provide a structured foundation for the experimental results that follow. The next chapter presents the findings of this workflow, comparing model behaviour, parameter effects, refinement outcomes, and ensemble performance within a constrained computational environment.

Chapter 4: Results and Discussion

4.1 Introduction

This chapter presents the quantitative and qualitative findings of the experiments conducted using Stable Diffusion 1.5, SDXL-Turbo, and the two-stage SD1.5 → SDXL-Turbo Img2Img refinement pipeline. The results synthesize baseline outputs, parameter optimization, prompt engineering evaluations, lightweight ensemble attempts, and LPIPS-based perceptual comparisons. The analysis is framed within the context of current literature on diffusion model behaviour, sampling optimization, conditional generation, image editing, and ensemble strategies [1][2][3][4][6][7][9][13][14][15].

Because prior studies generally assume access to high-end computational resources [1][2][3][15], this chapter also evaluates how these models behave within the real-world constraints of ≤12GB VRAM or CPU-only execution. Findings therefore contribute to the underexplored domain of low-resource diffusion optimization, an area rarely addressed by existing surveys or methodological papers [1][2][3][9].

4.2 Results and discussion

The results are organized into quantitative LPIPS comparisons, qualitative visual assessments, and interpretive discussions relating the findings to existing research.

4.2.1 SD1.5 vs SDXL-TURBO COMPARISON

Baseline images generated from SD1.5 and SDXL-Turbo under identical prompts and seeds revealed clear structural and perceptual differences. SD1.5 produced more consistent structure and subject positioning, while SDXL-Turbo generated sharper textures and more vivid color gradients. These observations reflect distinctions documented in model evolution surveys [1][3][13][15], which note that newer XL architectures excel at detail but may alter spatial composition.

LPIPS analysis demonstrated:

- SD1.5 vs. SDXL-Turbo: LPIPS ≈ 0.67

indicating substantial perceptual divergence.

This supports findings in conditional synthesis research that architectural changes often shift feature-space representations [9].

Figure 4.2.1:SD1.5

Figure 4.2.2:SDXL

Figure 4.2.3:SD1.5

Figure 4.2.4:SDXL

Figure 4.2.5:SD1.5

Figure 4.2.6:SDXL

Figure 4.2.7:SD1.5

Figure 4.2.8:SDXL

Figure 4.2.9:Model results

The defined prompts according to the results are defined by the table:

Table 4.2.1: prompt versus results

Prompt	Label	Model
A futuristic city skyline at sunset, cyberpunk style	Figure 4.2.10:SD1.5	SD1.5
A futuristic city skyline at sunset, cyberpunk style	Figure 4.2.11:SDXL	SDXL TURBO
A serene forest with a hidden waterfall, peaceful atmosphere	Figure 4.2.3:SD1.5	SD1.5
A serene forest with a hidden waterfall, peaceful atmosphere	Figure 4.2.4:SDXL	SDXL TURBO
A close-up of a robot's eye, intricate details, sci-fi	Figure 4.2.5:SD1.5	SD1.5
A close-up of a robot's eye, intricate details, sci-fi	Figure 4.2.6:SDXL	SDXL TURBO
A vibrant abstract painting with bold brushstrokes and contrasting colors	Figure 4.2.7:SD1.5	SD1.5
A vibrant abstract painting with bold brushstrokes and contrasting colors	Figure 4.2.8:SDXL	SDXL TURBO

4.2.2 Parameter Sweep Findings

Parameter sweeps across guidance scale, inference steps, and resolution revealed several key trends:

- **Higher guidance scales (>6)** caused over-sharpening and oversaturation, consistent with parameter sensitivity analyses described in diffusion surveys [2][9][14].

- **Low inference steps (<10)** resulted in coarse textures, especially on CPU/low-VRAM settings.
- **Moderate inference steps (15–25)** produced the best LPIPS stability while remaining computationally feasible under resource constraints.
- **Resolution increases above 512×512** frequently caused memory overflow, confirming limitations noted in diffusion system engineering literature [3][14].

These patterns demonstrate that carefully tuned parameters can partially compensate for limited hardware, supporting the argument for optimization strategies highlighted by Zhang et al. [1] and other surveys [2][9].

4.2.3 Prompt Engineering results

Prompt engineering significantly impacted both semantic alignment and perceptual coherence. Key findings include:

- **Structured prompts** improved consistency in subject rendering.
- **Negative prompts** reduced blurriness and unwanted artifacts, aligning with documented prompt-conditioning mechanisms in the literature [1][9][13].
- **Style modifiers** (e.g., “cinematic lighting”) were more influential in SDXL-Turbo than SD1.5, reflecting XL-model sensitivity to style conditioning reported in recent surveys [2][14].

These results validate claims in diffusion prompt-conditioning studies that structure and negative phrasing substantially influence output quality, especially when computational resources limit fine-tuning techniques like LoRA [4][5][8][11][12].

4.2.4 Two-Stage Img2Img Refinement (SD1.5 → SDXL-Turbo)

The two-stage refinement pipeline produced some of the strongest results in the study. Using SD1.5 to generate a stable structural base and SDXL-Turbo to refine texture created outputs that:

- Retained the spatial fidelity of SD1.5
- Gained the high-detail characteristics of SDXL-Turbo
- Achieved significantly improved LPIPS scores

Measured LPIPS values showed:

- **SD1.5 vs. Two-Stage Output:** *LPIPS ≈ 0.37*

indicating notably higher perceptual similarity compared to SDXL-Turbo alone.

This hybrid approach aligns with findings in diffusion editing and transformation research [4], demonstrating that refinement-type pipelines can outperform direct generation particularly in constrained environments where LoRA-based fine-tuning is not feasible [5][8][11][12].

The two-stage Img2Img refinement significantly increased perceptual similarity relative to SDXL-Turbo alone. Using SD1.5 as the reference image, SDXL-Turbo achieved an LPIPS score of 0.6743, while the two-stage method reduced this to 0.3716 representing a 44.9% improvement in perceptual similarity.

$$Improvement = \frac{0 \cdot 6743 - 0.3716}{0 \cdot 6743} \approx 44 \cdot 9\% \quad (1.5)$$

4.2.5 Lightweight Pixel-Space Ensemble Results

The pixel-space ensemble (averaging SD1.5 and SDXL-Turbo outputs) produced mixed results:

- **Strengths:**
    - Increased smoothness
    - Slight reduction in artifact noise

- **Weaknesses:**
    - Perceptual drift
    - Higher LPIPS values than the two-stage pipeline
    - Loss of semantic detail from both models

These outcomes contrast with the performance of feature-level ensembles such as ResEnsemble-DDPM [6] and AFA [7], which require significantly more computational

power. Ensemble diffusion literature supports the observation that pixel-space fusion is simple but tends to degrade semantic integrity [6][7][10].

Nevertheless, the experiment demonstrates a practical ensemble method feasible on low-resource hardware an area not addressed in previous ensemble studies.

4.3 SUMMARY TABLE OF ALL MODELS

Table 4.3.1:Model summary

Model / Method	LPIPS Score	Notes
SD1.5 (Best Param Sweep)	0.6765	Best-performing configuration among SD1.5 runs
SDXL-Turbo (Best Param Sweep)	0.6644	Highest perceptual similarity overall
Weighted Averaging Ensemble	0.7403	Best ensemble method; still worse than individual models
Advanced Blending Ensemble	0.7852	Better than simple averaging; still suboptimal
Simple Averaging Ensemble	0.8045	Highest LPIPS (least perceptual similarity)
Two-Stage (SD1.5 → SDXL Img2Img)	LPIPS varies per experiment	Improved similarity compared to standalone SDXL output(by 44.9%)

4.4 Conclusion to This Chapter

This chapter presented the main results and comparative analysis of Stable Diffusion 1.5, SDXL-Turbo, the two-stage Img2Img refinement pipeline, and a lightweight ensemble method under low-resource computational conditions. The key findings are:

- SD1.5 provides structurally stable baselines useful for constrained environments.
- SDXL-Turbo produces richer textures but diverges perceptually from SD1.5.

- Parameter sweeps reveal optimal operating ranges that align with diffusion sampling principles reported in the literature [2][9].
- Prompt engineering plays a critical role in low-resource optimization.
- The two-stage pipeline outperforms both single models and the lightweight ensemble in LPIPS similarity.
- Pixel-space averaging is feasible but not superior to refinement-based methods.

These insights collectively demonstrate that high-quality diffusion outputs can be achieved without high-VRAM GPUs, helping bridge the accessibility gap highlighted in current surveys [1][2][3][9][14][15].

Chapter 5: Conclusion and Future Work

5.1 Introduction

This final chapter presents the overall conclusion of the thesis and outlines future directions for research. The study set out to address a critical gap identified in recent diffusion model surveys and methodological analyses: the lack of standardized, memory-efficient workflows for running diffusion models in low-resource environments such as Google Colab and Kaggle free tiers [1][2][3][9][14][15]. While prior literature provides extensive discussion of diffusion theory, sampling techniques, fine-tuning strategies, and high-performance ensemble methods [4][5][6][7][8][11][12], it largely assumes the availability of high-end GPUs an assumption that excludes many students, independent researchers, and practitioners.

This thesis demonstrates that meaningful optimization of diffusion model outputs is both possible and practical under ≤12GB VRAM or CPU-only constraints. Through systematic experimentation involving baseline comparisons, parameter sweeps, prompt engineering, Img2Img refinement, and lightweight ensembling, the study provides both empirical evidence and methodological guidance for low-resource diffusion workflows. The sections that follow summarize the contributions to existing knowledge and outline promising avenues for future work.

5.2 Contributions to Existing Knowledge

This research contributes to the field of generative AI in several keyways, particularly by expanding the practical usability of diffusion models beyond high-end computational environments.

1. A Reproducible Low-Resource Diffusion Pipeline

The thesis introduces the first fully documented, memory-optimized workflow for running Stable Diffusion 1.5, SDXL-Turbo, and two-stage refinement pipelines on ≤12GB VRAM or CPU-only systems. While major surveys highlight diffusion model capabilities and system demands [1][2][3][9][14][15], none provide practical tools for constrained environments. This study fills that gap by demonstrating a step-by-step, reproducible pipeline with explicit memory management strategies.

2. **Empirical Analysis of Parameter and Prompt Optimization Under Constraints**

   By conducting structured parameter sweeps and prompt engineering tests, the study shows how sampling settings, negative prompts, and descriptive structure influence output quality in constrained settings. These findings extend prompt-conditioning and parameter-sensitivity discussions found in the literature [1][9][13][14] by applying them to real-world low-resource scenarios where users cannot rely on fine-tuning or large batch processing.

3. **Demonstration of a Two-Stage SD1.5 → SDXL-Turbo Refinement Pipeline**

   A major contribution is the empirically validated two-stage pipeline in which SD1.5 produces semantically stable structure and SDXL-Turbo refines detail. This approach grounded in diffusion editing concepts [4]—achieved substantially improved LPIPS scores compared to single-model outputs. The refinement pipeline presents a practical alternative to LoRA-based fine-tuning, which is often inaccessible on free-tier hardware due to authentication and VRAM constraints [5][8][11][12].

4. **Introduction of a Lightweight, Feasible Ensemble Method**

   Existing ensemble diffusion methods such as ResEnsemble-DDPM and AFA rely on feature-level fusion requiring significantly more computing power [6][7][10]. This thesis introduces a simple pixel-space averaging technique that, while limited, is computationally feasible under low-resource conditions. This represents the first documented attempt at ensemble diffusion specifically designed for constrained environments.

5. **Establishment of LPIPS as a Practical Evaluation Metric for Limited Hardware**

   The study reinforces LPIPS as a reliable, lightweight alternative to heavy metrics such as FID. This aligns with observations in diffusion evaluation research [9] and provides future low-resource researchers with a reproducible, accessible evaluation tool.

5.3 Future Work

The findings of this thesis open several promising avenues for further research.

1. LoRA Integration and Authentication Resolution

LoRA remains one of the most powerful parameter-efficient fine-tuning techniques [5][8][11][12], but hardware and authentication issues prevented its use in the current study. Future work should explore:

- Quantized LoRA loading
- CPU-suitable LoRA inference
- Training-free LoRA adaptation as proposed in LoRA-X [8]

2. Feature-Level and Latent-Space Ensemble Methods

Advanced ensemble strategies such as AFA [7] and ResEnsemble-DDPM [6] could yield stronger perceptual consistency if adapted for low-resource settings. Future research may investigate:

- Lightweight latent-space fusion
- Sparse feature extraction
- Hybrid ensembles combining pixel- and feature-level components

3. Quantization and Model Compression

Quantization-based methods offer significant memory savings and may enable SDXL or larger models to run within constrained environments. Techniques such as 8-bit, 4-bit, or QLoRA-style quantization [10] should be explored to further reduce VRAM requirements.

4. Larger-Scale Evaluation and Benchmarking

A broader evaluation across diverse prompts, resolutions, and domains would help validate the generalizability of the proposed pipeline. Prior surveys emphasize that diffusion model behavior can vary significantly across prompt complexity, semantic categories, and visual styles, highlighting the importance of wide-scope evaluation [1][2][13][15]. Future studies may incorporate:

5. Automated Optimization Tools

Finally, integrating reinforcement learning or heuristic search tools may help automate key components of the workflow, including:

- parameter tuning,
- prompt construction,
- and Img2Img refinement scheduling.

Automated tuning and adaptive control techniques have been highlighted in recent diffusion surveys as promising directions for improving generative stability and controllability, particularly in resource-limited contexts [1][2][9][13]. Furthermore, studies on model adaptation and cross-model optimization such as LoRA-based methods [5][8][11][12] and ensemble aggregation approaches [6][7][10], suggest that algorithmic controllers or reinforcement-driven search strategies could help navigate large configuration spaces more efficiently. Incorporating such tools would make low-resource diffusion workflows even more accessible to new researchers and expand the practical applicability of diffusion pipelines under constrained computational budgets.

5.4 Conclusion to Chapter 5

This thesis addressed a major gap in diffusion model research: the absence of practical, reproducible workflows for users operating under low-resource constraints. While modern diffusion systems continue to advance rapidly, most existing studies assume access to high-end GPUs and extensive VRAM [1][2][13][15]. In contrast, this work demonstrated that effective text-to-image generation is achievable even without such hardware through a carefully designed, memory-optimized pipeline.

Using Stable Diffusion 1.5 and SDXL-Turbo, the study showed that sequential model loading, disk-based intermediate storage, CPU-friendly inference, and lightweight Img2Img refinement enable stable operation on ≤12GB environments. Results confirmed that SDXL-Turbo achieved the strongest individual LPIPS score, while the two-stage refinement improved structural similarity relative to standalone SDXL generation. Conversely, simple ensemble methods such as averaging, weighted averaging, and advanced blending performed worse than the best individual models, consistent with

findings that effective ensembles typically require feature-level integration rather than pixel-space fusion [6][7][10].

The methodology also highlighted practical limitations of larger checkpoints and LoRA-based models, which frequently failed to load due to memory and authentication restrictions an issue echoed in recent adaptation literature [5][8][11][12]. Despite these constraints, the study established a reproducible, accessible pipeline that expands participation in generative AI research and provides a foundation for future low-resource experimentation.

Future work should incorporate broader prompt distributions, additional evaluation metrics, and more advanced lightweight ensemble techniques [3][9][15]. Integrating reinforcement learning or heuristic search tools may also help automate prompt design, parameter tuning, and refinement scheduling, as suggested in diffusion surveys [1][2][13]. Overall, this thesis demonstrates that diffusion model research can be both technically rigorous and accessible, even when computational resources are severely limited.

References

[1] Chenshuang Zhang, Chaoning Zhang, et al. "Text-to-image Diffusion Models in Generative AI: A Survey." arXiv, 2023.

[2] Tianyi Zhang, Zheng Wang, et al. "A Survey of Diffusion Based Image Generation Models: Issues and Their Solutions." arXiv, 2023.

[3] Brian B. Moser, Arundhati S. Shanbhag, et al. "Diffusion Models, Image Super-Resolution And Everything: A Survey." arXiv, 2024.

[4] Yi Huang, Jiancheng Huang, et al. "Diffusion Model-Based Image Editing: A Survey." arXiv, 2024.

[5] Edward Hu, Yelong Shen, et al. "LoRA: Low-Rank Adaptation of Large Language Models." ICLR, 2022.

[6] S. Zhenning. "ResEnsemble-DDPM: Residual Denoising Diffusion Ensemble." arXiv, 2023.

[7] C. Wang et al. "Ensembling Diffusion Models via Adaptive Feature Aggregation (AFA)." OpenReview, 2024.

[8] F. Farhadzadeh, D. Das, et al. "LoRA-X: Bridging Foundation Models with Training-Free Cross-Model Low-Rank Adaptation." ICLR, 2025.

[9] Chenshuang Zhang et al. "Conditional Image Synthesis with Diffusion Models: A Survey." arXiv, 2024.

[10] L. Li. "Ensemble and low-frequency mixing with diffusion models (ELF-Diff) for accelerated MRI." ScienceDirect, 2025

[11] “LoRA-Enhanced Distillation on Guided Diffusion Models”. arXiv, 2023.

[12] “LoRA-Composer: Leveraging Low-Rank Adaptation for Multi-Concept Diffusion”. arXiv, 2024.

[13] “A Survey on Generative Diffusion Models”. arXiv, 2022.

[14] “Comprehensive exploration of diffusion models in image generation”. Springer article, 2024/25.

[15] “Diffusion Models: A Comprehensive Survey of Methods and Applications”. TPAMI 2023.

Author's Biography

Moses Mukiibi was born in Kampala, Uganda. He completed his early education in Uganda before relocating to the Republic of Korea to pursue higher studies. He is currently pursuing Bachelor of Science degree in Computer Engineering from Kyungdong University, Gangwon-do, Republic of Korea. During his undergraduate studies, he developed strong interests in machine learning, computer vision, and resource-efficient artificial intelligence systems.

From 2022, he pursued academic and independent research in the areas of generative AI, diffusion models, and low-resource optimization methods, with a special focus on enabling advanced AI tools to run in constrained environments. His research projects include work on Stable Diffusion pipelines, perceptual evaluation metrics, lightweight ensemble modeling, and accessible generative AI workflows for students and early-stage researchers.

His broader research interests include diffusion-based generative modeling, multimodal learning, resource-efficient deep learning, and applied artificial intelligence. He aims to continue contributing to the development of accessible AI technologies that bridge the gap between cutting-edge research and real-world computational limitations.