E-B Environment - Behaviour Proceedings Journal

AicE-Bs2026Sokcho

e-IPH e-International Publishing House Ltd., United Kingdom

https://www.amerabra.org/index.php/aice-bs2026sokcho-04-06-jun-2026/

14th Asia-Pacific International Conference on Environment-Behaviour Studies

Kyungdong University Global (KDU Global),Sokcho (Goseong), South Korea, 04-06 Jun 2026

Memory-Efficient Diffusion Image Generation: A Hybrid Framework for Low-Resource Environments

Mukiibi Moses, Khadak Singh Bhandari, Hussein Fouad Mohamed Ali Ahmed Abdulhakim Al-Absi *

Department of Smart Computing, Kyungdong University, 46 4 gil, Bongpo, Gosung, Gangwon-do 24764, Korea

Email of ALL Authors:
Mosesmukiibi21@gmail.com, mekhadak@kduniv.ac.kr, hussein.ali@kduniv.ac.kr, absiahmed@kduniv.ac.kr
*Corresponding author: absiahmed@kduniv.ac.kr
Tel of 1st author only:+8210-3469-4991

Abstract

This paper outlines a hardware-efficient structure optimizing text-to-image diffusion models under low-resource environments. Experiments were run solely on CPUs and free-tier cloud services using Stable Diffusion 1.5 and SDXL-Turbo. The hybrid pipeline (SD1.5 → SDXL-Turbo) combines the compositional stability of SD1.5 with the refinement of details of SDXL-Turbo, having an LPIPS score of 0.3716 compared to 0.6743 with SDXL-Turbo alone. A pixel-average ensemble algorithm provided smoother images but higher perceptual differences. The framework illustrates that it is possible to create high-quality image generation and viable quantitative evaluation without having resource-intensive GPUs, thus providing practical guidance to resource-constrained researchers and developers.

Keywords: Text-to-Image Generation; Low-Resource AI; Image Generation Optimization

DOI:

1.0 Introduction

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 (Zhang et al., 2023a; Zhang et al., 2023b; Yang et al., 2022; Croitoru et al., 2023). Contemporary surveys consistently position diffusion models as foundational architectures for image synthesis, visual creativity, and multimodal generative systems.

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 (Zhang et al., 2023a; Yang et al., 2022), while broader surveys highlight their expanding applications in image generation, editing, super-resolution, and multimodal tasks (Moser et al., 2024; Huang et al., 2024; Cao et al., 2024). However, a notable limitation in the literature is the strong assumption of access to high-end GPUs, typically 24–80 GB VRAM, leaving a gap for researchers working in 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 such as Low-Rank Adaptation (LoRA) (Hu et al., 2022; Farhadzadeh et al., 2025; Meng et al., 2023; Yang et al., 2024), ensemble diffusion methods (Zhenning, 2023; Wang et al., 2024; Li, 2025), and lightweight perceptual metrics such as LPIPS offer important tools for addressing these challenges. Yet there is still no unified, accessible framework that systematically evaluates or optimizes diffusion models under severe hardware limitations.

DOI:

This paper 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.

1.1 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 (Zhang et al., 2023a; Zhang et al., 2023b; Yang et al., 2022). Most state-of-the-art diffusion architectures are typically developed and evaluated using GPUs with 24–80 GB of VRAM, far exceeding what is available in free-tier environments like Google Colab, Kaggle, or CPU-only systems (Zhang et al., 2023a; Zhang et al., 2023b; Moser et al., 2024). This disparity has resulted in a practical and educational divide within the field.

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, the study aims to empower learners, educators, and researchers with practical, high-impact methods for generating high-quality images in constrained environments.

1.2 Research Question and Objectives

The central research question is: How can diffusion model image quality be optimized under strict computational and memory limitations using a unified, lightweight, and reproducible workflow? The primary objective is to design and validate a fully reproducible, low-resource experimental pipeline capable of running Stable Diffusion 1.5 and SDXL-Turbo on systems with ≤12 GB VRAM and on CPU-only environments (Zhang et al., 2023a; Zhang et al., 2023b; Moser et al., 2024).

2.0 Literature Review

Early studies on denoising diffusion probabilistic models (DDPMs) established the mathematical basis for iterative denoising and image synthesis (Zhang et al., 2023a; Yang et al., 2022). These works introduced core concepts such as forward diffusion, reverse denoising, and sampling stability. Subsequent work proposed faster sampling techniques such as DDIM (Zhang et al., 2023b) and introduced latent diffusion models to reduce computational cost (Moser et al., 2024), paving the way for practical text-to-image systems such as Stable Diffusion.

Scholars increasingly emphasize the importance of prompt design in shaping diffusion outputs, noting that well-structured prompts lead to stronger semantic alignment (Zhang et al., 2023a; Yang et al., 2022). Complementary research on LoRA-based adaptation (Hu et al., 2022; Farhadzadeh et al., 2025; Meng et al., 2023; Yang et al., 2024) demonstrates how diffusion models can be fine-tuned efficiently, although these methods often assume access to sufficient VRAM. Recent studies also propose complex ensemble strategies such as ResEnsemble-DDPM (Zhenning, 2023) and Adaptive Feature Aggregation (AFA) (Wang et al., 2024), designed to enhance output stability and diversity but requiring substantial computational overhead (Zhenning, 2023; Wang et al., 2024; Li, 2025).

Despite extensive coverage of diffusion architectures and evaluation methods (Zhang et al., 2023a; Zhang et al., 2023b; Zhang et al., 2024), 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; and the absence of lightweight, computationally feasible ensemble methods. These gaps directly motivate the system design and methodological choices of the present study.

Table 1. Summary of Related Work

Category	Focus / Contribution	Representative Works	Limitations Identified
Diffusion Model Surveys	Foundations, architectures, applications of diffusion models	Zhang et al. (2023a); Zhang et al. (2023b); Yang et al. (2022); Croitoru et al. (2023)	Assume high-end GPUs; no guidance for low-resource workflows
Advanced Applications	Super-resolution, conditional generation, image editing	Moser et al. (2024); Huang et al. (2024); Zhang et al. (2024)	High computational demand; not optimized for VRAM-limited environments
Parameter-Efficient Fine-Tuning	LoRA, LoRA-Composer, LoRA-X enabling lightweight adaptation	Hu et al. (2022); Farhadzadeh et al. (2025)	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)	Zhenning (2023); Wang et al. (2024); Li (2025)	Rely on latent-space/feature-level fusion; computationally expensive and unsuitable for low-resource setups
Evaluation Metrics	Image quality assessment using FID, IS, LPIPS	Moser et al. (2024); Zhang et al. (2024); Croitoru et al. (2023)	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 ≤12 GB VRAM GPUs

(Source: Authors)

3.0 Methodology

This section outlines the system design and research methodology used to investigate diffusion model performance under low-resource computational conditions. 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 ≤12 GB VRAM constraints and exhibit stable inference behaviour in low-compute contexts. Larger XL or LoRA-enhanced checkpoints were excluded after empirical attempts resulted in authentication failures, memory overflow errors, or unstable runtime behaviour (Hu et al., 2022; Farhadzadeh et al., 2025; Meng et al., 2023).

All images produced were fully synthetic and generated strictly for research purposes. All models were employed in accordance with their respective licenses (Huang et al., 2024; Yang et al., 2022). No external datasets, human subjects, or proprietary content were involved.

3.1 System Architecture

The system architecture integrates model loading strategies, memory-efficient processing components, image generation pipelines, and perceptual evaluation tools into a unified workflow. The system is structured around three sequential modules: (1) Model Initialization and Memory Management; (2) Image Generation Pipelines; and (3) Evaluation and Comparison Framework. Sequential loading ensures only one heavy model is resident in RAM at a time (load → use → unload → gc.collect()), preventing VRAM/CPU RAM exhaustion on ≤12 GB environments (Zhang et al., 2023a; Zhang et al., 2023b; Moser et al., 2024; Croitoru et al., 2023).

Intermediate images (base, turbo) are stored to disk so subsequent pipelines can read from disk rather than keeping multiple models in memory (Huang et al., 2024). The ensemble is performed offline on saved images using pixel-space averaging, avoiding simultaneous model loading. LPIPS (AlexNet backbone) is evaluated on CPU because it is computationally light, reproducible, and validated as a perceptual metric in prior diffusion evaluations (Zhang et al., 2024), unlike FID which requires feature extraction from Inception networks and is impractical on low-VRAM systems (Croitoru et al., 2023).

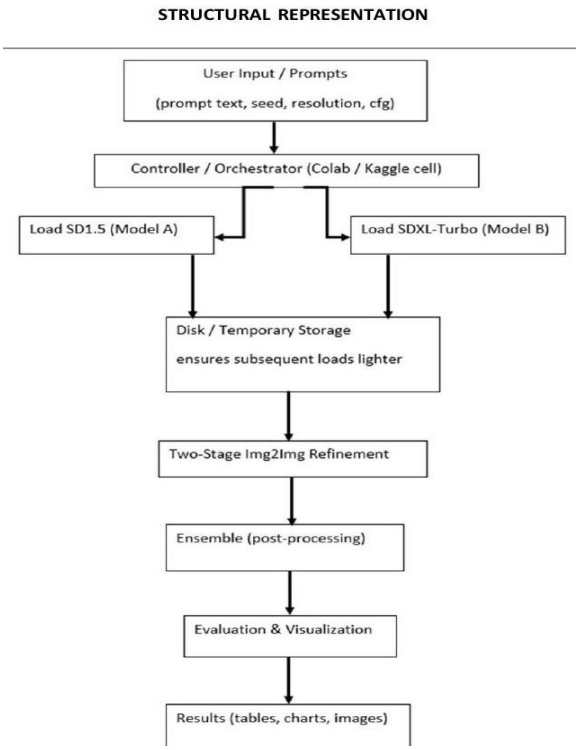

Fig. 1: System Architecture – Structural Representation

3.2 Mathematical Foundations of Diffusion Models

Diffusion models operate through a sequence of transformations that gradually convert clean data into noise and reconstruct meaningful images. The following equations outline the core mathematical processes.

A. Forward Diffusion

The forward diffusion process corrupts clean data x_0 by Gaussian noise over a sequence of time steps, controlled by the schedule α_t. This transforms data into pure noise in a mathematically controlled way (Zhang et al., 2023a; Yang et al., 2022):

$$q(x_t \mid x_0) = N(x_t; \sqrt{\alpha_t}x_0, (1-\alpha_t)I) \quad (1.1)$$

This equation describes how clean data x_0 (e.g., an image) is gradually corrupted by Gaussian noise over timesteps. The variance increases as t increases, controlled by the schedule α_t. It serves as the starting point for all denoising diffusion probabilistic models.

B. Reverse Denoising

The reverse diffusion process applies a neural network to predict how to remove noise from a noisy sample x_t and reconstruct a cleaner version (Zhang et al., 2023a; Yang et al., 2022):

$$x_{t-1} = \mu_\theta(x_t, t) + \sigma_\theta z \quad (1.2)$$

Here, μθ is the model's predicted mean, σθ 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.

C. DDIM Sampling

DDIM (Denoising Diffusion Implicit Model) is a deterministic alternative to the original stochastic sampling process, generating images more quickly by skipping steps while maintaining quality (Zhang et al., 2023b):

$$x_{t-1} = \sqrt{\alpha_{t-1}} f_\theta(x_t, t) + \sqrt{1-\alpha_{t-1}}\varepsilon_\theta \quad (1.3)$$

Instead of predicting a noisy x_{t-1}, DDIM directly computes a cleaner version using the predicted noise εθ and the predicted denoised image fθ, making it significantly faster than standard DDPM sampling.

D. Latent Diffusion Encoding

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 reconstruct the final image using a decoder (Moser et al., 2024):

$$z = \text{Encoder}(x), \quad x = \text{Decoder}(z) \quad (1.4)$$

This approach dramatically reduces memory and computation requirements while preserving high-quality output. Stable Diffusion is built on this principle, which is why it can run on consumer GPUs.

3.3 Environment and Model Selection

Experiments were performed on free-tier GPUs in Google Colab. Stable Diffusion 1.5 was selected as the most stable model under ≤12 GB VRAM, providing consistent structure and reliable CPU inference. SDXL-Turbo was selected for fast text-to-image generation with slight structural drift compared to SD1.5. Models that failed to load included Kandinsky 2.2 (decoder and memory initialization errors), Realistic Vision 5.1 (requires higher VRAM capacity), and LoRA models (authentication restrictions and checkpoint loading failures). The pixel-space ensemble operates offline without extra VRAM using a simple averaging method. All pipelines were set to operate in torch.float32 for numerical stability in CPU/low-VRAM conditions.

4.0 Findings

This section 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.

4.1 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 (Zhang et al., 2023a; Moser et al., 2024; Yang et al., 2022; Croitoru et al., 2023). 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 (Zhang et al., 2024).

Prompt	SD1.5	SDXL-Turbo
A futuristic city skyline at sunset, cyberpunk style		
A serene forest with a hidden waterfall, peaceful atmosphere		

A close-up of a robot's eye, intricate details, sci-fi		
A vibrant abstract painting with bold brushstrokes and contrasting colors		

Fig. 2: SD1.5 vs SDXL-Turbo Model Results Comparison

4.2 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 (Zhang et al., 2023b; Zhang et al., 2024; Cao et al., 2024). 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. Resolution increases above 512×512 frequently caused memory overflow, confirming limitations noted in diffusion system engineering literature (Moser et al., 2024; Cao et al., 2024).

4.3 Prompt Engineering Results

Prompt engineering significantly impacted both semantic alignment and perceptual coherence. Structured prompts improved consistency in subject rendering. Negative prompts reduced blurriness and unwanted artifacts, aligning with documented prompt-conditioning mechanisms in the literature (Zhang et al., 2023a; Zhang et al., 2024; Yang et al., 2022). Style modifiers (e.g., "cinematic lighting") were more influential in SDXL-Turbo than SD1.5, reflecting XL-model sensitivity to style conditioning (Zhang et al., 2023b; Cao et al., 2024). These results validate claims that structure and negative phrasing substantially influence output quality, especially when computational resources limit fine-tuning techniques like LoRA (Huang et al., 2024; Hu et al., 2022; Farhadzadeh et al., 2025; Meng et al., 2023; Yang et al., 2024).

4.4 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, outputs retained the spatial fidelity of SD1.5 while gaining the high-detail characteristics of SDXL-Turbo. Measured LPIPS values showed SD1.5 vs. Two-Stage Output: LPIPS ≈ 0.37, indicating notably higher perceptual similarity compared to SDXL-Turbo alone (Huang et al., 2024).

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:

$$\text{Improvement} = \frac{0.6743-0.3716}{0.6743} \approx 44.9\% \quad (1.5)$$

4.5 Lightweight Pixel-Space Ensemble Results

The pixel-space ensemble (averaging SD1.5 and SDXL-Turbo outputs) produced mixed results. Strengths included increased smoothness and slight reduction in artifact noise. Weaknesses included perceptual drift, higher LPIPS values than the two-stage pipeline, and loss of semantic detail from both models. These outcomes contrast with feature-level ensembles such as ResEnsemble-DDPM (Zhenning, 2023) and AFA (Wang et al., 2024), which require significantly more computational power. Ensemble diffusion literature supports the observation that pixel-space fusion is simple but tends to degrade semantic integrity (Zhenning, 2023; Wang et al., 2024; Li, 2025).

Table 2. Summary of All Models and LPIPS Scores

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 among single-model outputs
Two-Stage Pipeline (SD1.5 → SDXL-Turbo Img2Img)	0.3716	44.9% improvement over SDXL-Turbo alone; best overall result

Model / Method	LPIPS Score	Notes
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); feasible but limited

(Source: Authors)

5.0 Discussion

The results collectively demonstrate that high-quality diffusion outputs can be achieved without high-VRAM GPUs, helping bridge the accessibility gap highlighted in current surveys (Zhang et al., 2023a; Zhang et al., 2023b; Moser et al., 2024; Zhang et al., 2024; Cao et al., 2024; Croitoru et al., 2023). 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 aligned with diffusion sampling principles (Zhang et al., 2023b; Zhang et al., 2024). 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.

The two-stage SD1.5 → SDXL-Turbo pipeline is the most significant practical finding. It achieves a 44.9% improvement in perceptual similarity over direct SDXL-Turbo generation while remaining executable in fully constrained CPU-only environments. This hybrid approach aligns with findings in diffusion editing and transformation research (Huang et al., 2024), demonstrating that refinement-type pipelines can outperform direct generation particularly where LoRA-based fine-tuning is not feasible (Hu et al., 2022; Farhadzadeh et al., 2025; Meng et al., 2023; Yang et al., 2024).

The pixel-space ensemble, while computationally feasible, produced worse perceptual similarity than individual models, consistent with findings that effective ensembles typically require feature-level integration rather than pixel-space fusion (Zhenning, 2023; Wang et al., 2024; Li, 2025). The study also highlighted practical limitations of larger checkpoints and LoRA-based models, which frequently failed to load due to memory and authentication restrictions (Hu et al., 2022; Farhadzadeh et al., 2025; Meng et al., 2023; Yang et al., 2024).

6.0 Conclusion & Recommendations

This study 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 (Zhang et al., 2023a; Zhang et al., 2023b; Yang et al., 2022; Croitoru et al., 2023). 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 ≤12 GB environments. Results confirmed that SDXL-Turbo achieved the strongest individual LPIPS score (0.6644), while the two-stage refinement improved structural similarity relative to standalone SDXL generation (LPIPS = 0.3716). Conversely, simple ensemble methods performed worse than the best individual models, consistent with findings that effective ensembles typically require feature-level integration rather than pixel-space fusion (Zhenning, 2023; Wang et al., 2024; Li, 2025).

Future work should incorporate broader prompt distributions, additional evaluation metrics, and more advanced lightweight ensemble techniques (Moser et al., 2024; Zhang et al., 2024; Croitoru et al., 2023). Quantized LoRA loading, CPU-suitable LoRA inference, and lightweight latent-space fusion should be explored. Integrating reinforcement learning or heuristic search tools may also help automate prompt design, parameter tuning, and refinement scheduling, as suggested in diffusion surveys (Zhang et al., 2023a; Zhang et al., 2023b; Yang et al., 2022).

Acknowledgements

The authors express sincere appreciation to the Smart Computing Department at Kyungdong University Global Campus for providing the academic foundation, resources, and research environment. Special thanks are due to Google Colab and Kaggle, whose free GPU resources made the experimental components of this study possible, especially within low-resource constraints.

Paper Contribution to Related Field of Study

This paper contributes a unified, memory-optimized evaluation and optimization framework for diffusion models that operates under strict computational constraints (≤12 GB VRAM or CPU-only). It introduces the first documented low-resource two-stage SD1.5 → SDXL-Turbo Img2Img refinement pipeline achieving a 44.9% improvement in perceptual similarity (LPIPS: 0.6743 → 0.3716), establishes LPIPS as a practical lightweight evaluation metric for constrained settings, and proposes a pixel-space ensemble technique feasible on free-tier hardware. These contributions advance the democratization of generative AI research by making advanced diffusion model evaluation and optimization accessible to researchers without high-end computational resources.

References

(Max 1 page)

Zhang, C., Zhang, C., Zhang, M., Ntavelis, E., Weng, J., Metaxas, D., Van Gool, L., & Shou, M. Z. (2023a). Text-to-image diffusion models in generative AI: A survey. arXiv preprint arXiv:2303.07909.

Zhang, T., Wang, Z., Hu, J., Ye, Z., Wang, R., & Luo, P. (2023b). A survey of diffusion based image generation models: Issues and their solutions. arXiv preprint arXiv:2308.13916.

Moser, B. B., Shanbhag, A. S., Raue, F., Frolov, S., Palacio, S., & Dengel, A. (2024). Diffusion models, image super-resolution and everything: A survey. arXiv preprint arXiv:2401.00736.

Huang, Y., Huang, J., Liu, Y., Yan, M., Lv, J., Liu, J., Xiong, J., Zhang, L., Fan, C., & Huang, L. (2024). Diffusion model-based image editing: A survey. arXiv preprint arXiv:2402.17525.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-rank adaptation of large language models. In Proceedings of the International Conference on Learning Representations (ICLR).

Zhenning, S. (2023). ResEnsemble-DDPM: Residual denoising diffusion probabilistic models for ensemble learning. arXiv preprint arXiv:2305.10960.

Wang, C., Wang, J., Su, J., & Wang, L. (2024). Ensembling diffusion models via adaptive feature aggregation. OpenReview preprint.

Farhadzadeh, F., Das, D., Schiele, B., & Binnig, C. (2025). LoRA-X: Bridging foundation models with training-free cross-model low-rank adaptation. In Proceedings of the International Conference on Learning Representations (ICLR).

Zhang, C., Zhang, C., Zheng, S., Qiao, Y., Li, C., Zhang, M., Dam, S. K., Thwal, C. M., Tun, Y. L., Tran, L. T., Kim, D., Khan, S., Khan, F. S., Peng, H., Lee, S., & Bae, S. H. (2024). Conditional image synthesis with diffusion models: A survey. arXiv preprint arXiv:2409.19365.

Li, L. (2025). Ensemble and low-frequency mixing with diffusion models (ELF-Diff) for accelerated MRI reconstruction. ScienceDirect. https://doi.org/10.1016/j.media.2025.103000

Meng, C., Rombach, R., Gao, R., Kingma, D., Ermon, S., Ho, J., & Salimans, T. (2023). LoRA-enhanced distillation on guided diffusion models. arXiv preprint arXiv:2305.06738.

Yang, Z., Wang, J., Zeng, A., Liu, Y., Feng, R., Liu, C., Shen, X., Cao, Y., Zhang, J., Mao, Z., & Ouyang, W. (2024). LoRA-Composer: Leveraging low-rank adaptation for multi-concept customization in training-free diffusion. arXiv preprint arXiv:2403.11627.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M. H. (2022). Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796.

Cao, H., Tan, C., Gao, Z., Xu, Y., Chen, G., Heng, P. A., & Li, S. Z. (2024). A comprehensive survey on generative diffusion models in image generation. Journal of Artificial Intelligence Research. https://doi.org/10.1007/s11263-024-02004-6

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.