Diffusion-Based vs. Non-Diffusion Trajectory Guidance in Long-Horizon Robotic Tasks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the impact of using different diffusion-based trajectory guidance mechanisms (e.g., latent diffusion vs. pixel-space diffusion) on long-horizon task success rates in RoboBench, and how does. 11 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Text-driven Visual Synthesis with Latent Diffusion Prior. Research question: What is the impact of using different diffusion-based trajectory guidance mechanisms (e.g., latent diffusion vs. pixel-space diffusion) on long-horizon task success rates in RoboBench, and how does this compare to non-diffusion-based approaches?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

15 papers retrieved. 11 claims extracted; 1 independently verified. Quality review score: 4.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Using the proposed method produces a significantly better FID score and a competitive CLIP score compared to StyleGANFus	×	0.05
The proposed method produces more detailed results than the latent diffusion-guided baselines and the CLIP-guided method	×	0.08
The feature matching loss LFM is inspired by the GAN discriminator feature matching loss proposed in pix2pixHD.	×	0.08
The feature matching loss LFM operates similarly to the latent score distillation loss LLSD, but its signal is amplified	×	0.08
The proposed method uses feature matching and KL losses to reintroduce the decoder into the optimization procedure.	×	0.06
The proposed method uses a pretrained (and fixed) decoder for feature matching, not an additional trainable discriminato	×	0.04
The proposed method uses latent code with added noise residual and the clean latent as the decoder input, as opposed to	×	0.04
The direct gradient computation involves calculating the UNet Jacobian, which is computationally expensive.	×	0.01
Diffusion models have recently shown great potential in synthesis tasks, achieving competitive or even better performanc	×	0.07
The score distillation loss is computed in the latent space, termed 'latent score distillation (LSD)'.	×	0.05
The proposed Feature Matching Loss (FM) loss uses features in multiple decoder layers of the latent diffusion model to a	✓	0.16

References

- http://arxiv.org/abs/2602.06413v1
- http://arxiv.org/abs/2403.18760v2
- http://arxiv.org/abs/2302.08510v2