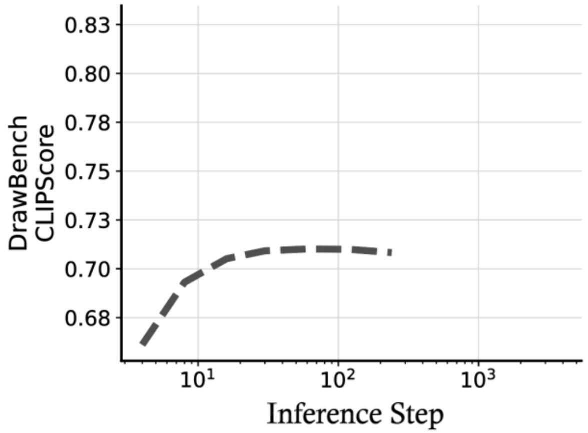
Inference-Time Optimization for Diffusion Models

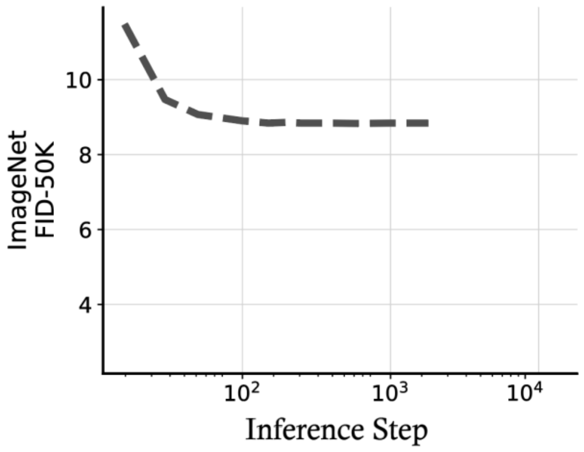
Kaiyuan Deng
Department of Electrical and Computer Engineering

Problem Review

Diffusion models have attracted widespread attention for their ability to generate high-quality images, with many works focusing on improving performance by **increasing** **training data**, **model size**, and the **number of denoising steps**.

Inference-time Optimization (training-free) leverages existing pretrained models for post-processing to quickly enhance generation quality while avoiding the heavy computational cost of retraining. However, current inference methods that **solely rely on increasing denoising steps** often encounter a performance **bottleneck**, where further computational investment does not yield significant gains.





Method- Inference-Time *Initial Noise* Optimization

Instead of merely increasing the number of denoising steps, we can find **better initial noise inputs** by investing additional compute at inference time. The core idea is that by using extra compute to perform a guided search – where verifiers rate the candidates and various search algorithms are used – the overall generation quality can significantly exceed the performance achieved by merely scaling the denoising steps.

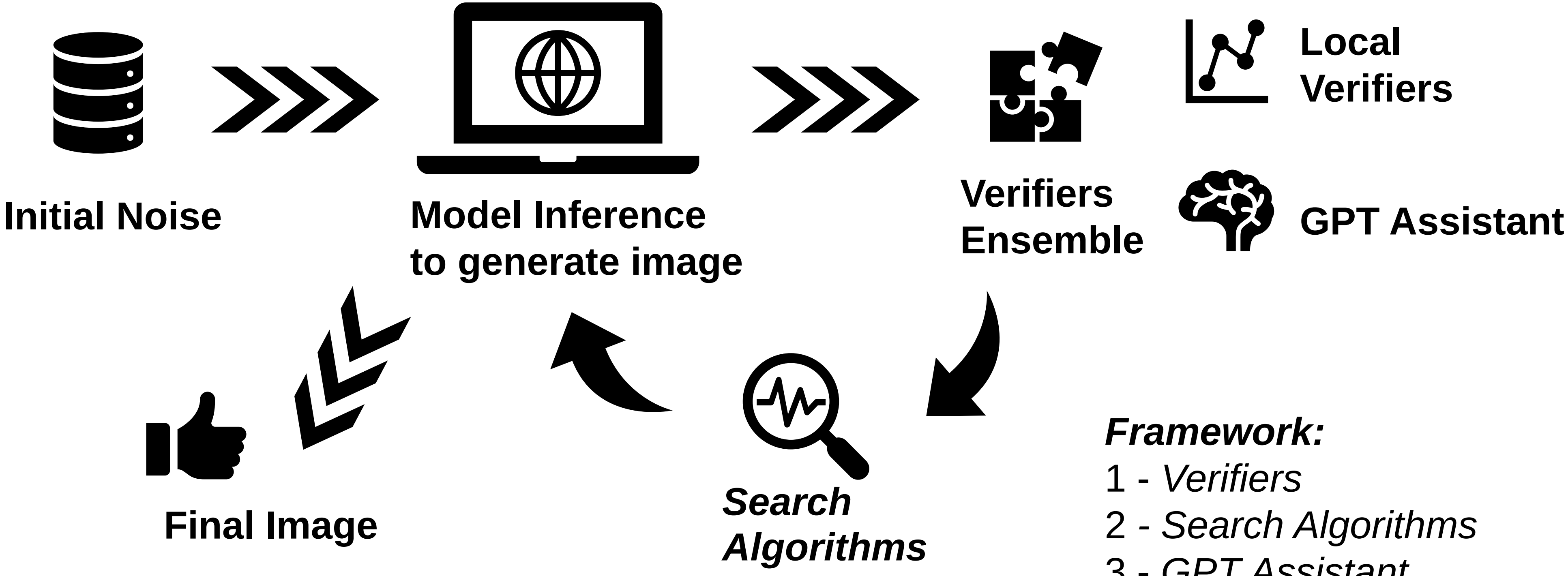
Framework:

1 - *Verifiers*: to measure the quality of noises/images

2 - *Search Algorithms*: to explore the noise space for better initial noise candidates.

3 - *GPT Assistant* : to provide a human-like score for noises/images

Method- Inference-Time *Initial Noise* Optimization

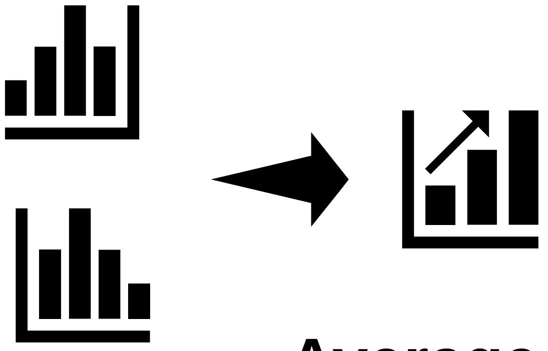


Framework:
1 - *Verifiers*
2 - *Search Algorithms*
3 - *GPT Assistant*

Method- Inference-Time *Initial Noise* Optimization

Verifiers Ensemble

- Aesthetic Score - Beauty
- CLIP Score – Consistency
- ImageReward Score - Holistic Quality
- GPT Assistant Score - Human-like Insight



Average Rank

Method- Inference-Time *Initial Noise* Optimization

Search over Paths
By introducing perturbations along the sampling trajectory and evaluating the intermediate state images using verifiers, this method helps uncover potential improvements across the whole generation process.

Zero-Order Search
to generate multiple perturbed candidates within the local neighborhood of the current noise, then select the locally optimal candidate based on verifier scores.

Experiment Result

Model = SDXL-base-1.0   Step = 30

Base

Proposed Method

1.An off-road vehicle and a bike in the forest

2.a beautiful landscape painting

Experiment Result

Model = SDXL-base-1.0 Step = 30

Base

Proposed Method

3.knight in armor, standing on a cliff, looking at the sunset

4.a church in the forest

Experiment Result

Method	Aesthetic Score (↑)	CLIP Score (↑)	ImageReward Score(↑)	GPT Score(↑)
Random noise	5.38	0.64	0.91	84
Ensemble + ZO	5.72	0.69	0.96	87
Ensemble + Paths	5.99	0.70	0.95	86
Ensemble + both	6.31	0.82	1.22	89

Table 1: Statistical Result

Thank you