Initial noise in generative models

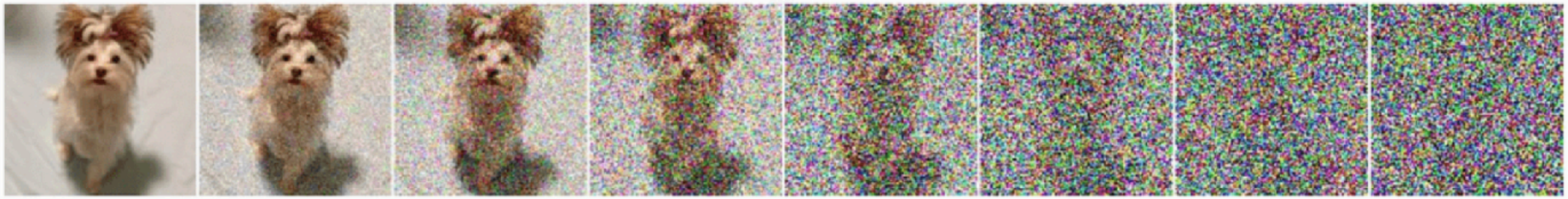
Shuai Ma 2025.8.15

Background

Diffusion Models

Forward SDE (data → noise)

$\mathbf{x}(0)$ — $\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$ → $\mathbf{x}(T)$



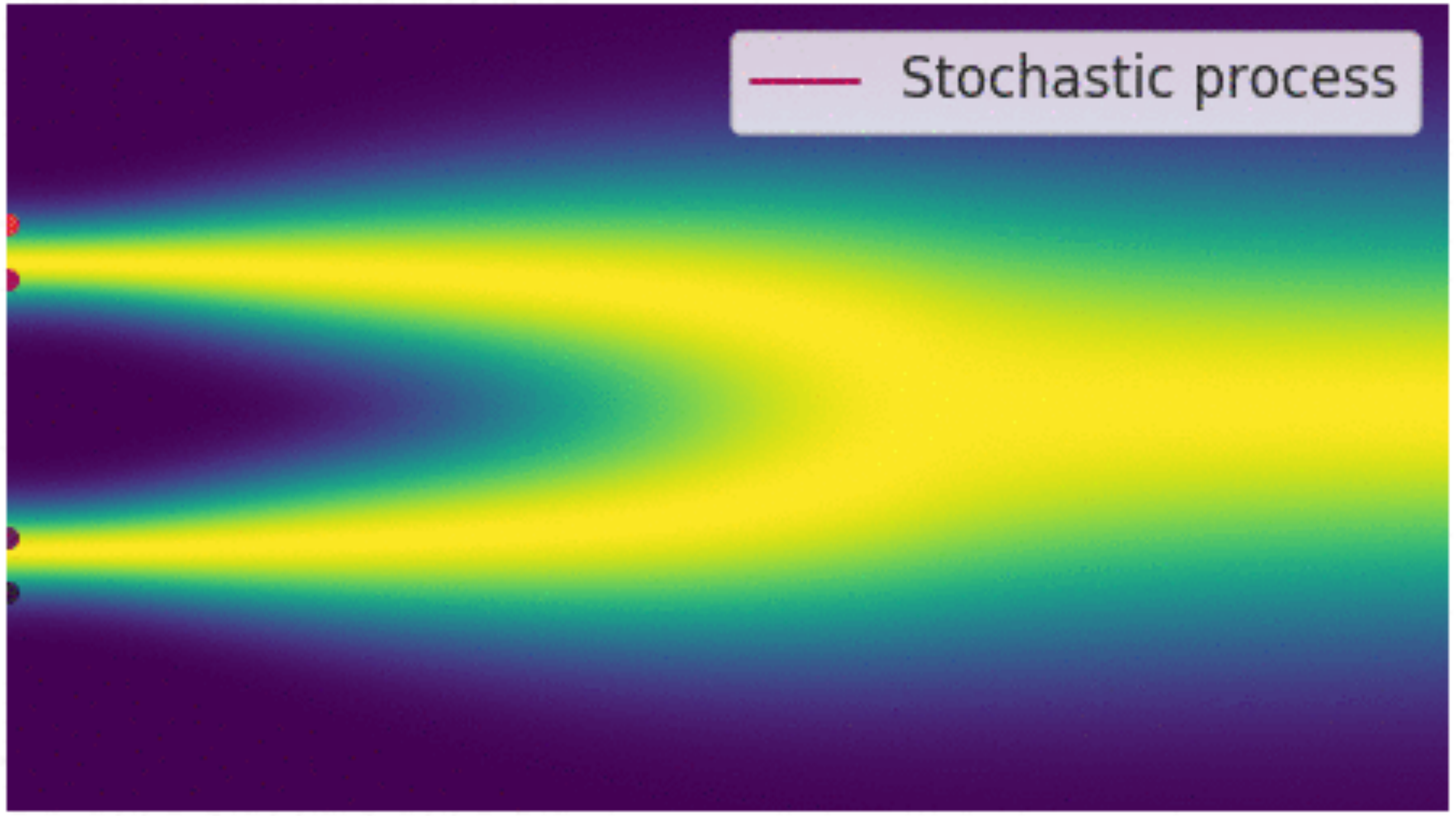
score function

$\mathbf{x}(0)$ ← $\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}$ — $\mathbf{x}(T)$

Reverse SDE (noise → data)

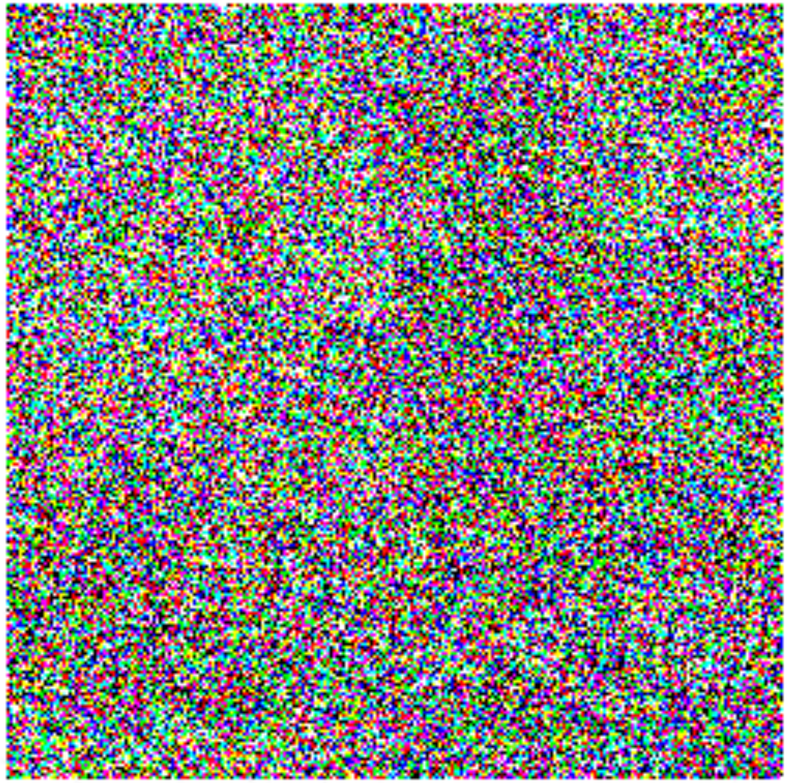
Background

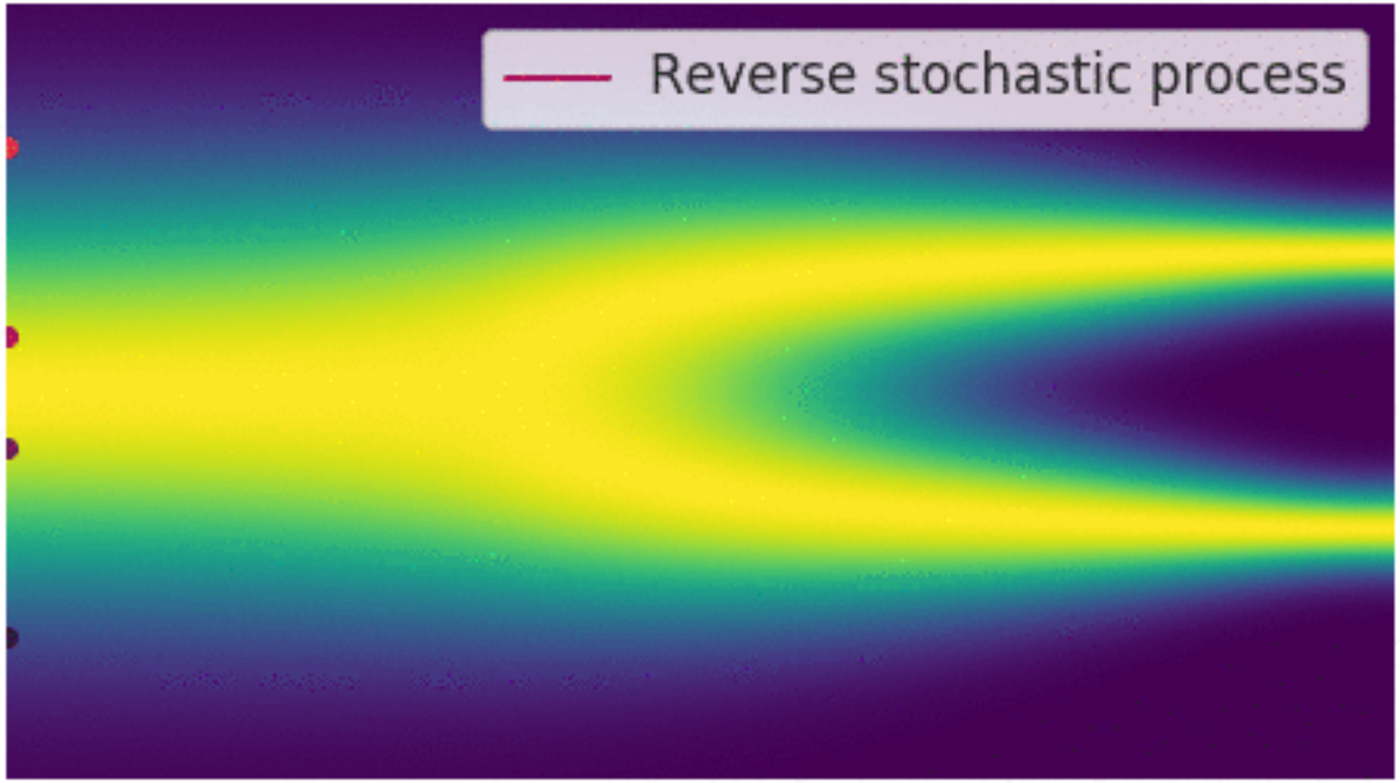
Forward Process



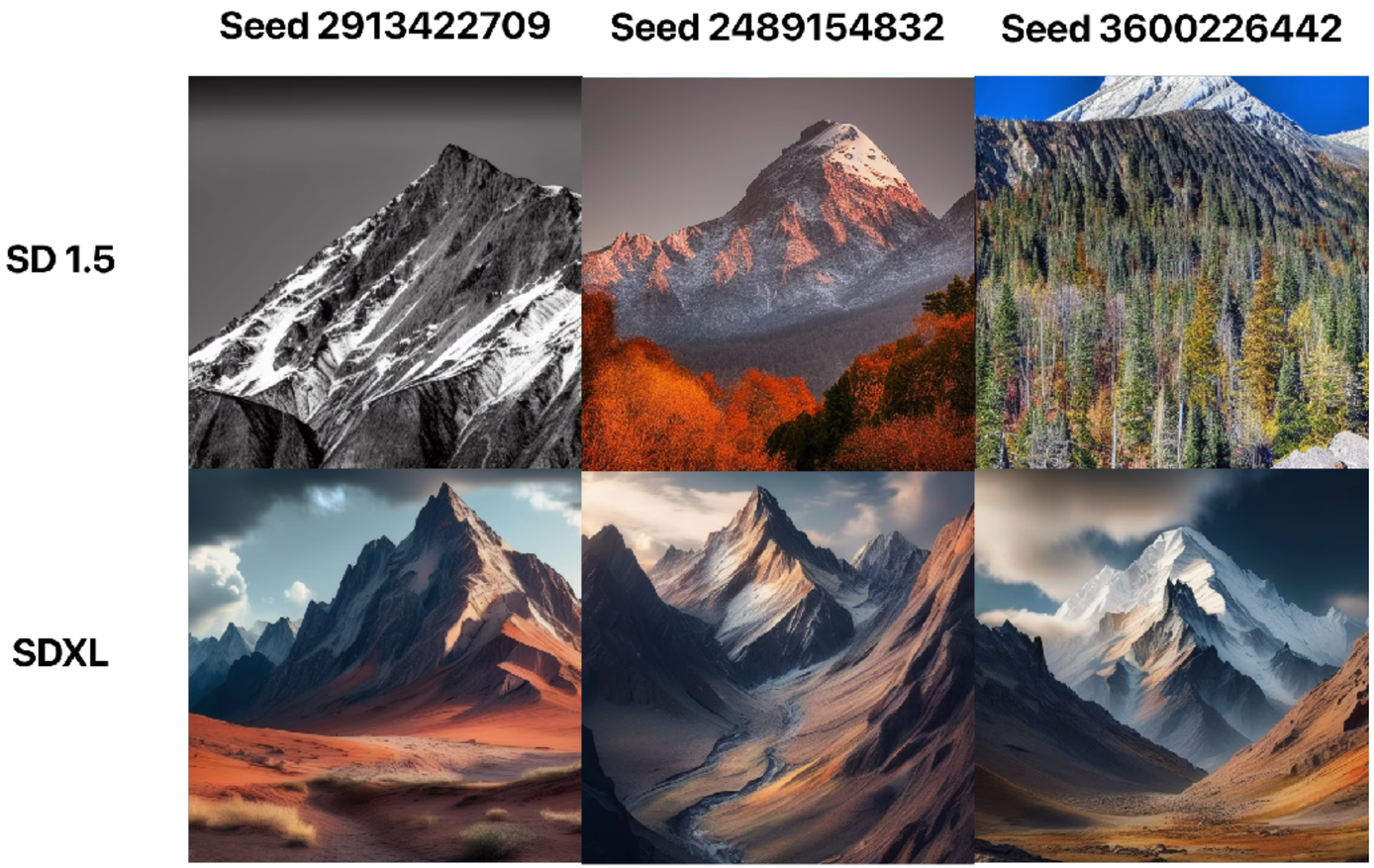
Background

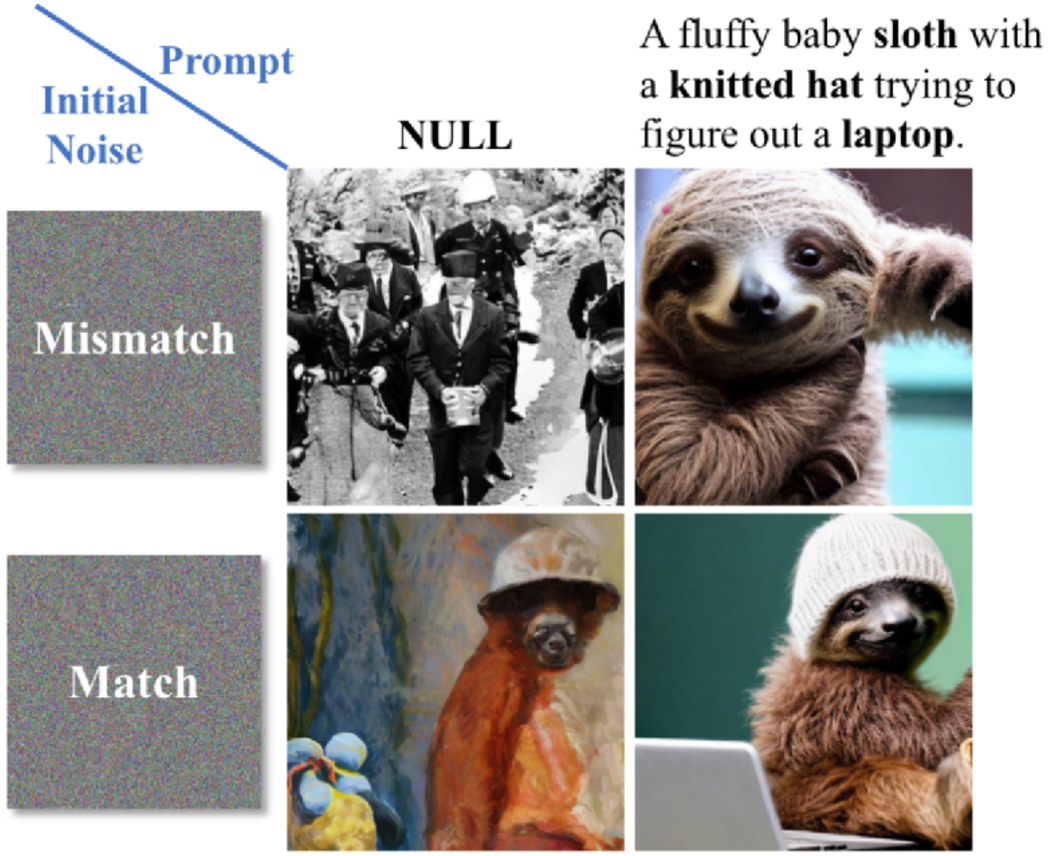
Reverse Process





Observation





The Silent Assistant

NoiseQuery as Implicit Guidance for Goal-Driven Image Generation

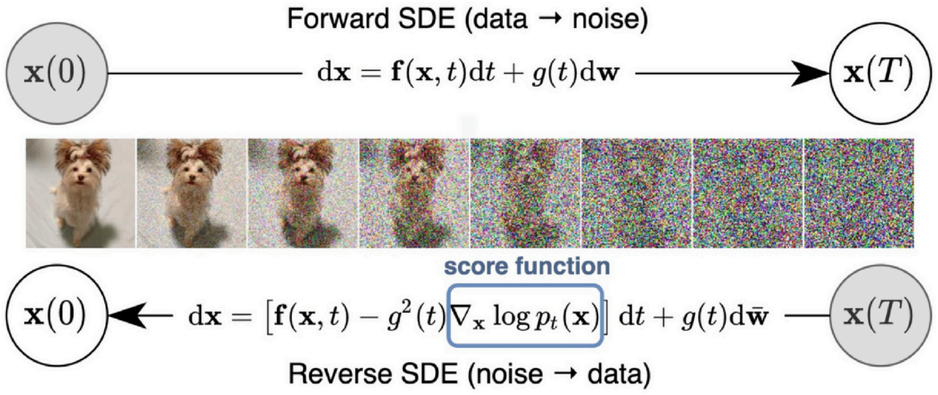
- Cross-model similarities when generating with same initial noise

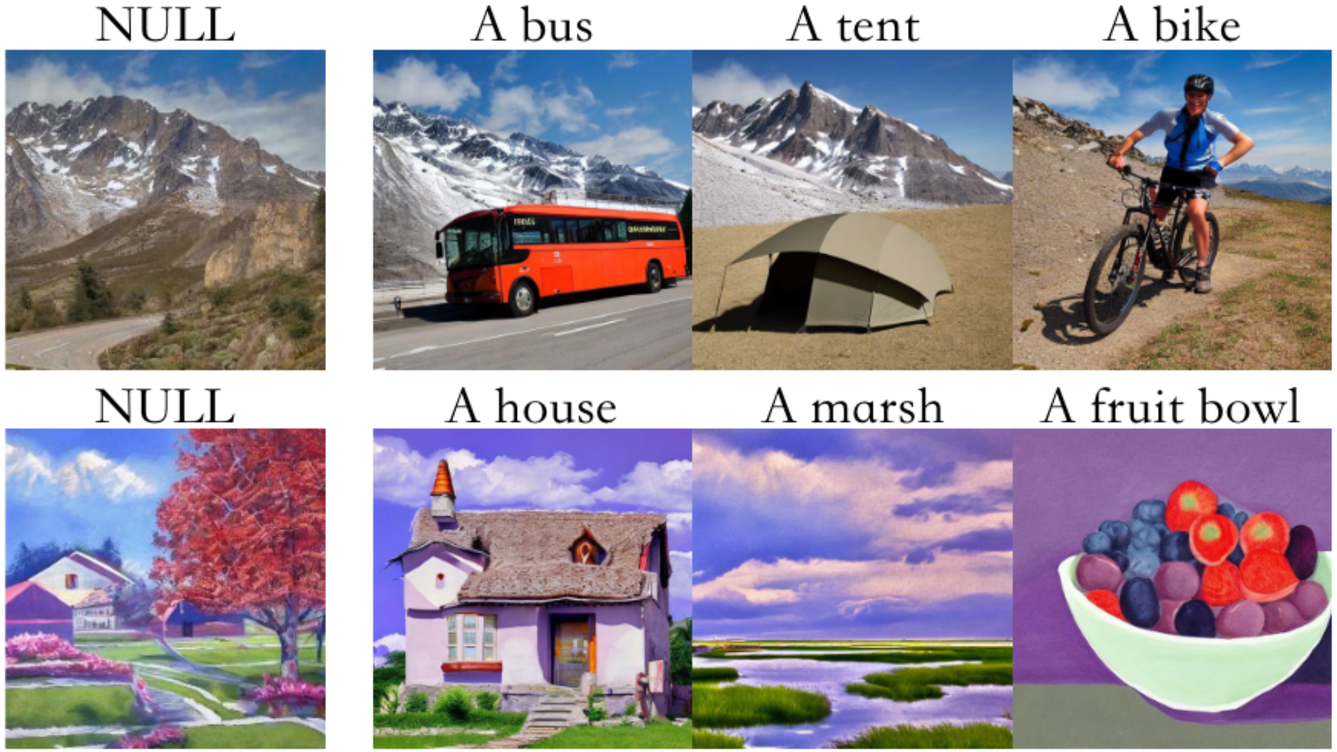
The Silent Assistant

NoiseQuery as Implicit Guidance for Goal-Driven Image Generation

- Similar colors, texture, sharpness
- Implicitly encodes clues

$$x_T = 0.068265 \cdot x_0 + 0.997667 \cdot \epsilon$$



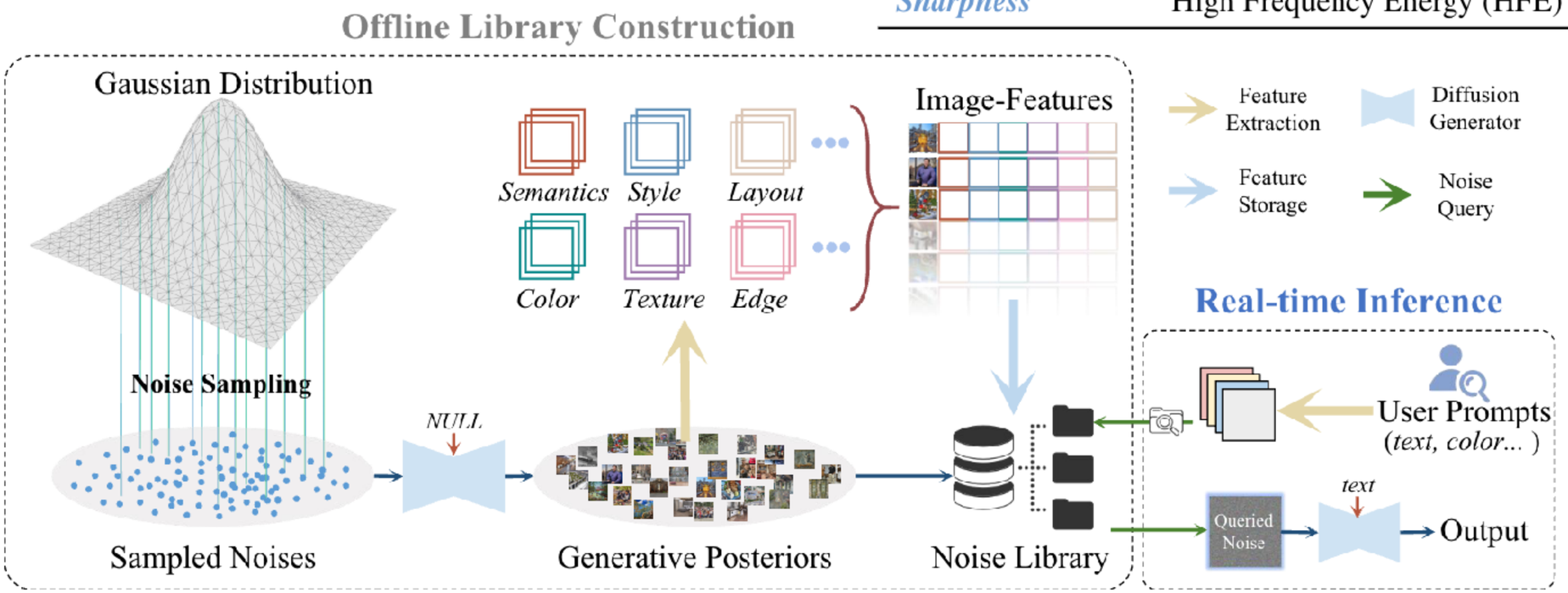


The Silent Assistant

NoiseQuery as Implicit Guidance for Goal-Driven Image Generation

- 100k random noise samples

Generation Goals	Feature Type	Match Function
Semantics	CLIP [42], BLIP [29]	Cosine Similarity
Style	Gram Matrix [16]	MSE
Color	RGB, HSV, LAB	Absolute Difference
Texture	GLCM [19]	Euclidean Distance
Shape	Hu Moments [22]	Euclidean Distance
Sharpness	High Frequency Energy (HFE)	Absolute Difference



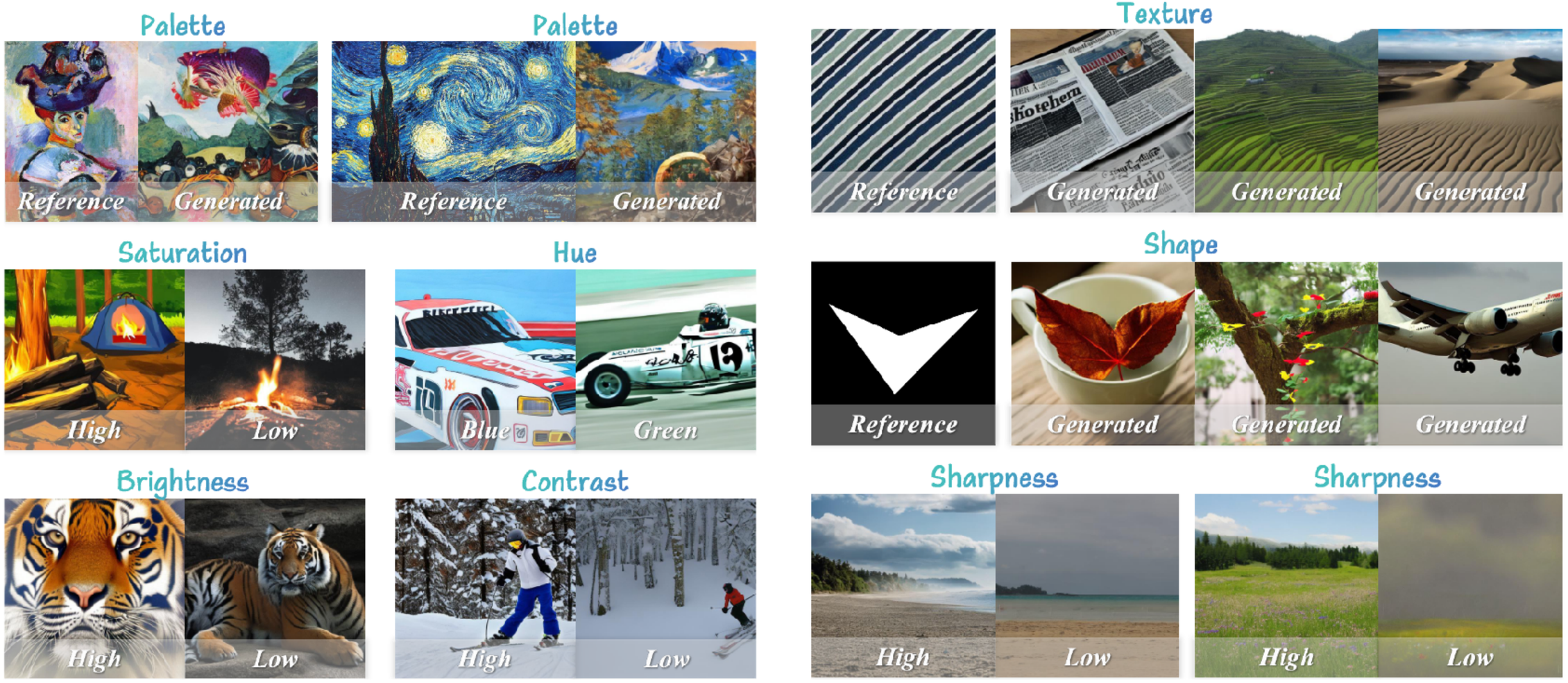
The Silent Assistant

NoiseQuery as Implicit Guidance for Goal-Driven Image Generation

Base Model	Method	DrawBench [47]				MSCOCO [33]				Time Cost
		ImageReward	PickScore	HPS v2	CLIPScore	ImageReward	PickScore	HPS v2	CLIPScore	
SD 1.5	Base Model	0.04	21.11	24.57	30.90	0.15	21.41	25.65	31.08	1.334 s
	+ NoiseQuery	**0.08**	**21.16**	**25.02**	**31.41**	**0.27**	**21.48**	**26.07**	**31.47**	1.336 s
	+ Diffusion-DPO [53]	0.09	21.29	25.02	31.19	0.25	21.64	26.31	31.26	1.350 s
	+ Diffusion-DPO [53] + NoiseQuery	**0.17**	**21.33**	**25.25**	**31.41**	**0.35**	**21.68**	**26.60**	**31.55**	1.352 s
SD 2.1	Base Model	0.12	21.33	24.93	31.13	0.36	21.72	26.58	31.40	1.301 s
	+ NoiseQuery	**0.26**	**21.46**	**25.39**	**31.68**	**0.44**	**21.76**	**26.82**	**31.50**	1.303 s
	+ CFG++ [10]	0.12	21.33	24.83	31.13	0.37	21.72	26.66	31.31	3.724 s
	+ CFG++ [10] + NoiseQuery	**0.27**	**21.43**	**25.55**	**31.61**	**0.47**	**21.76**	**26.97**	**31.67**	3.726 s
SD-Turbo	Base Model	0.26	21.78	25.23	31.29	0.47	22.07	26.22	31.51	0.072 s
	+ NoiseQuery	**0.41**	**21.87**	**25.66**	**31.58**	**0.50**	**22.17**	**26.82**	**31.76**	0.074 s
	+ ReNO [11]	1.67	23.40	32.48	32.55	-	-	-	-	23.56 s
	+ ReNO [11] + NoiseQuery	**1.71**	**23.52**	**32.92**	**32.78**	-	-	-	-	23.56 s
PixArt-α	Base Model	0.70	22.08	28.27	30.83	0.78	22.24	29.33	31.48	4.327 s
	+ NoiseQuery	**0.82**	**22.11**	**28.45**	**31.27**	**0.79**	**22.33**	**29.56**	**31.64**	4.328 s
	+ LaVi-Bridge [67]	0.63	22.08	28.35	30.92	0.75	22.31	29.49	31.86	5.092 s
	+ LaVi-Bridge [67] + NoiseQuery	**0.72**	**22.24**	**28.61**	**31.35**	**0.78**	**22.35**	**29.67**	**32.01**	5.094 s

The Silent Assistant

NoiseQuery as Implicit Guidance for Goal-Driven Image Generation



(a) Color properties guidance.

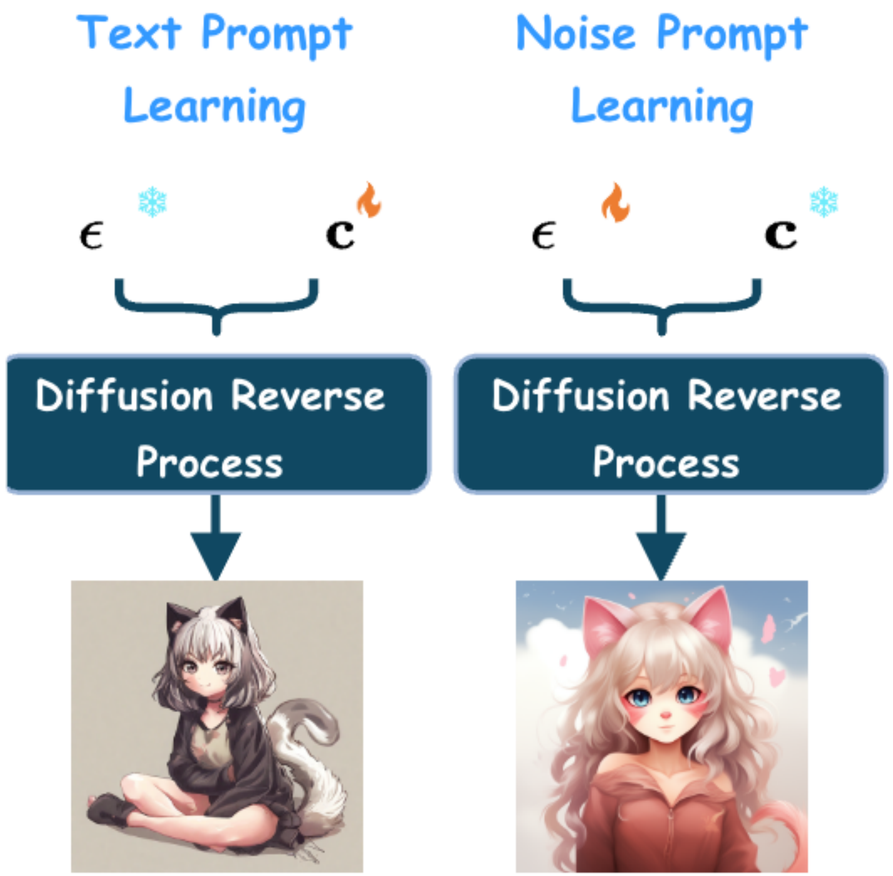
(b) Structural features guidance.

Golden Noise for Diffusion Models

A Learning Framework

- Text Prompt Learning & Noise Prompt Learning

$$\phi^* = \arg\min_\phi \mathbb{E}_{(\mathbf{x}_{T_i}, \mathbf{x}'_{T_i}, \mathbf{c}_i)\sim\mathcal{D}}[\ell(\phi(\mathbf{x}_{T_i}, \mathbf{c}_i), \mathbf{x}'_{T_i})].$$



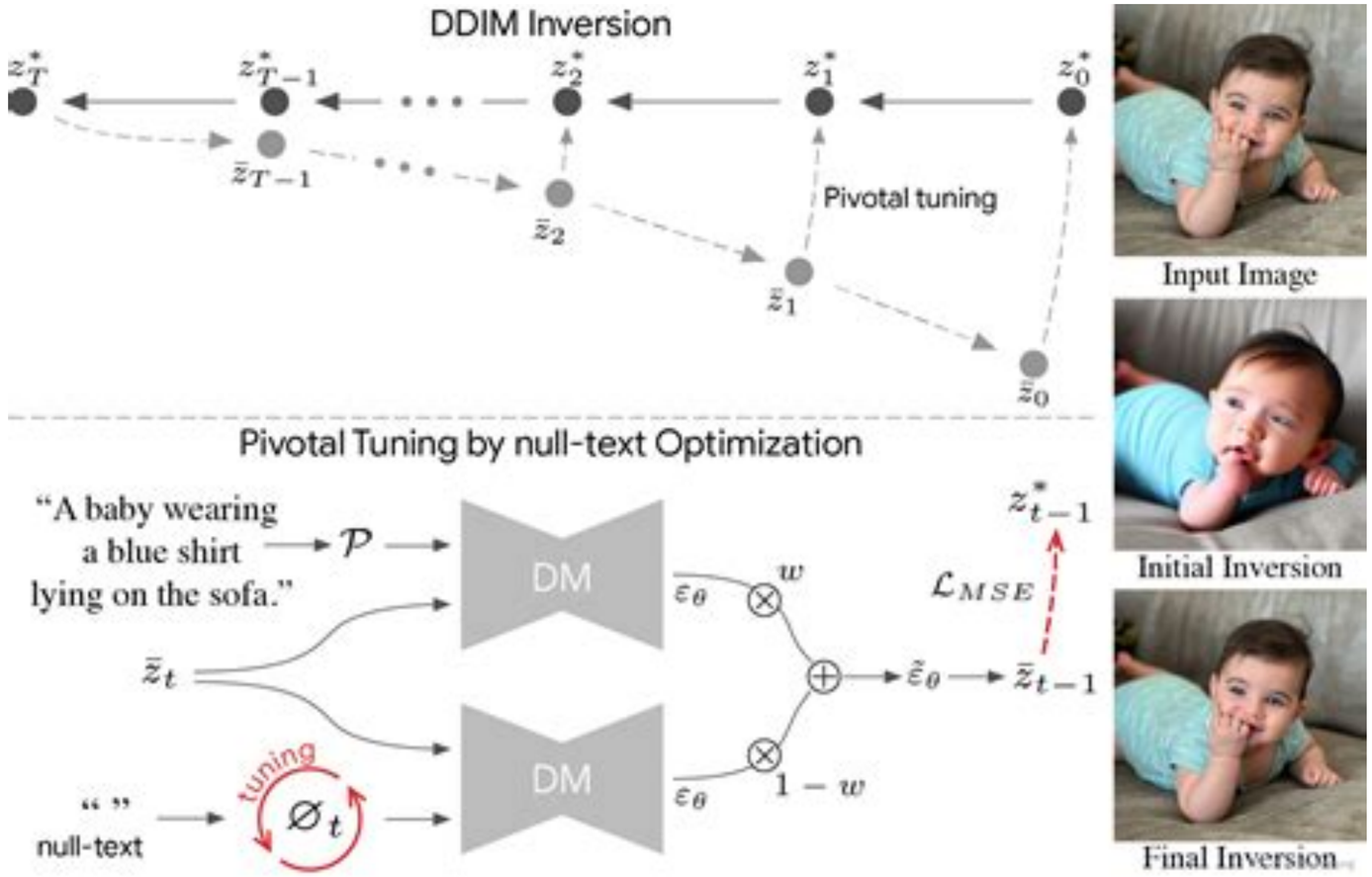
Golden Noise for Diffusion Models

A Learning Framework

- DDIM Inversion

$$\mathbf{x}_{t-1} = \mathrm{DDIM}(\mathbf{x}_t) = \alpha_{t-1}\left(\frac{\mathbf{x}_t - \sigma_t \epsilon_\theta(\mathbf{x}_t, t)}{\alpha_t}\right) + \sigma_{t-1}\epsilon_\theta(\mathbf{x}_t, t) \quad (1)$$

$$\mathbf{x}_t = \text{DDIM-Inversion}(\mathbf{x}_{t-1}) = \frac{\alpha_t}{\alpha_{t-1}}\mathbf{x}_{t-1} + \left(\sigma_t - \frac{\alpha_t}{\alpha_{t-1}}\sigma_{t-1}\right)\epsilon_\theta(\mathbf{x}_t, t) \quad (2)$$

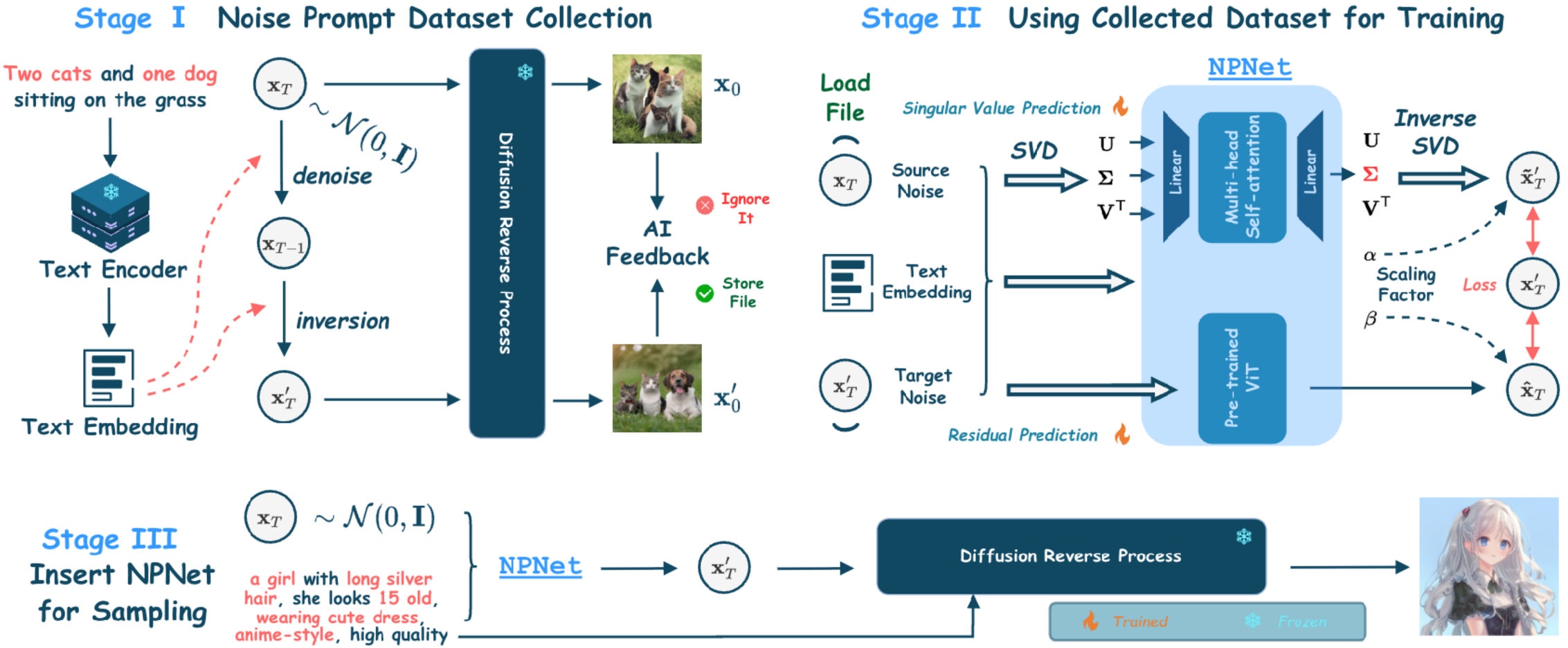


Golden Noise for Diffusion Models

A Learning Framework

$$\mathbf{e} = \sigma(\mathbf{x}_T, \mathcal{E}(\mathbf{c})) \quad \hat{\mathbf{x}}_T = \varphi'(\psi(\varphi(\mathbf{x}_T + \mathbf{e}))$$

$$\mathcal{L}_{\text{MSE}} = \text{MSE}(\mathbf{x}'_T, \mathbf{x}'_{T_{pred}}),$$
$$\text{where } \mathbf{x}'_{T_{pred}} = \tilde{\mathbf{x}}'_T + \beta\hat{\mathbf{x}}_T,$$



Golden Noise for Diffusion Models

A Learning Framework

Golden Noise for Diffusion Models

A Learning Framework

Golden Noise for Diffusion Models

A Learning Framework

Model	Dataset	Method	PickScore (↑)	HPSv2 (↑)	AES (↑)	ImageReward (↑)	CLIPScore (↑)	MPS(%) (↑)
SDXL	Pick-a-Pic	Standard	21.69	28.48	6.0373	58.01	0.8204	-
		Inversion [1]	21.71	28.57	6.0503	63.27	0.8250	51.41
		NPNet (ours)	21.86	28.68	6.0540	65.01	0.8408	52.14
	DrawBench	Standard	22.31	26.72	5.5952	62.21	0.8077	-
		Inversion	22.37	26.91	5.6017	67.09	0.8081	51.98
		NPNet (ours)	22.38	27.14	5.6034	70.67	0.8153	53.70
	HPD	Standard	22.88	29.71	5.9985	96.63	0.8734	-
		Inversion	22.89	29.78	5.9948	97.39	0.8708	53.03
		NPNet (ours)	22.94	29.88	5.9922	98.81	0.8813	56.02
DreamShaper-xl-v2-turbo	Pick-a-Pic	Standard	22.41	32.12	6.0161	98.09	0.8267	-
		Inversion	22.40	32.03	6.0236	100.97	0.8277	49.14
		NPNet (ours)	22.73	32.69	6.0646	106.74	0.8958	52.34
	DrawBench	Standard	22.98	30.39	5.6735	98.84	0.8186	-
		Inversion	22.94	30.10	5.6852	96.74	0.8189	46.62
		NPNet (ours)	23.11	30.78	5.7005	108.14	0.8224	53.53
	HPD	Standard	23.68	30.96	6.1408	129.89	0.8868	-
		Inversion	23.67	31.00	6.0811	131.80	0.8912	46.94
		NPNet (ours)	23.70	34.08	6.1283	135.98	0.8942	52.49