Tackling the Singularities at the Endpoints of Time Intervals in Diffusion Models

Pengze Zhang[1*] Hubery Yin[1*] Chen Li[2] Xiaohua Xie[1†]

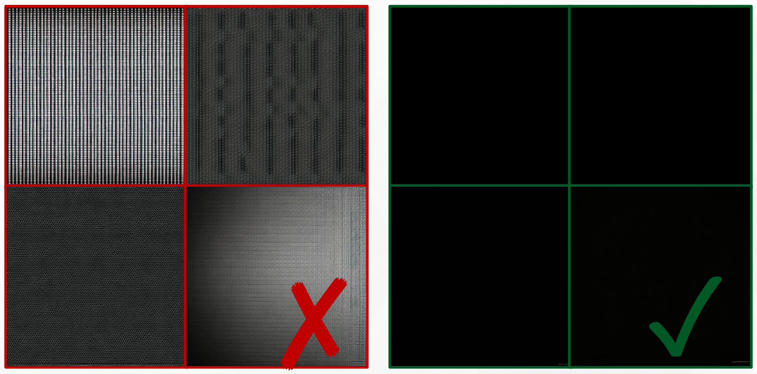
[1]Sun Yat-sen University [2]Wechat, Tencent Inc

CVPR 2024
(Highlight)

Reporter: Pengze Zhang

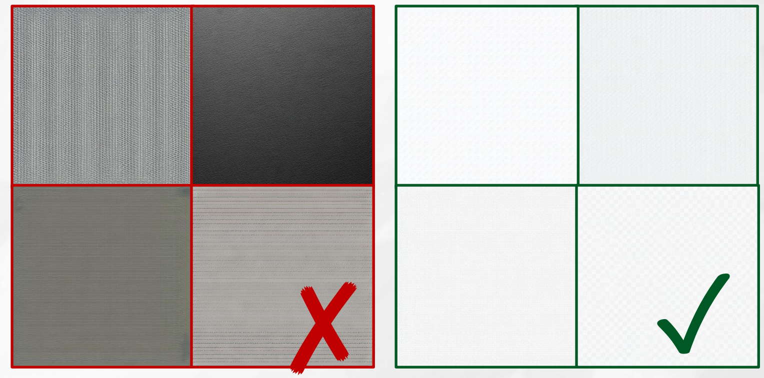
Motivation – Average Brightness Issue

Prompt: Solid **Black** background



Stable Diffusion SingDiffusion

Prompt: Solid **White** background



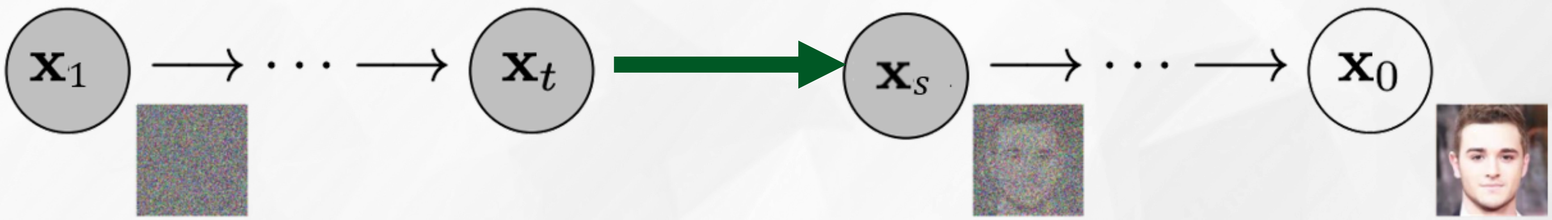
Stable Diffusion SingDiffusion

Prompt: Anne Hathaway, Wedding Dress, **White** Background, Studio

Original Model SingDiffusion (Target)

Challenge I – Reverse Gaussian Properties

- **Error bound between $p(x_s|x_t)$ and estimated $\widetilde{p}(x_s|x_t)$**



Mixed Gaussian Distribution

$$p(x_s|x_t) = (2\pi\sigma_{s|t}^2)^{-\frac{d}{2}} \sum_{i=1}^{N} \exp\left(-\frac{1}{2\sigma_{s|t}^2}\left(x_s - \frac{\alpha_{t|s}\sigma_s^2 x_t}{\sigma_t^2} - \frac{\alpha_s \sigma_{t|s}^2 y_i}{\sigma_t^2}\right)^2\right)\omega_i(x_t, t)$$

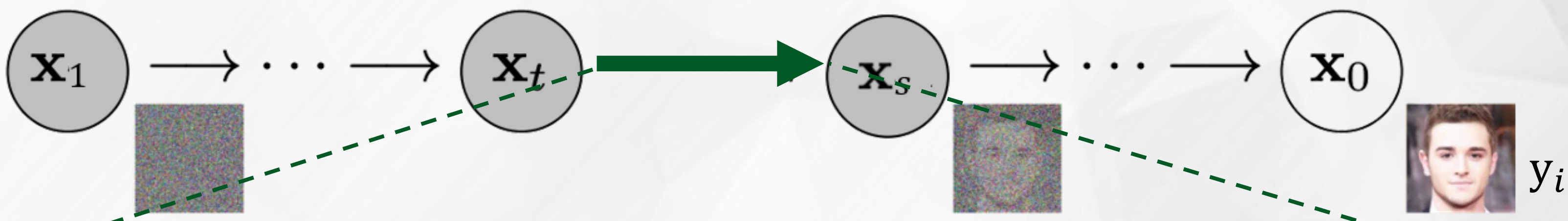
GAP?

$$\widetilde{p}(x_s|x_t) = (2\pi\sigma_{s|t}^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\sigma_{s|t}^2}\left(x_s - \frac{\alpha_{t|s}\sigma_s^2 x_t}{\sigma_t^2} - \frac{\alpha_s \sigma_{t|s}^2 \bar{y}(x_t, t)}{\sigma_t^2}\right)^2\right)$$

Gaussian Distribution

Challenge II – Singularities

- **Singularities at $t = 1$ or $t = 0$ during reverse process.**



$$\tilde{p}(x_s|x_t) = (2\pi\boldsymbol{\sigma}_{s|t}^2)^{-\frac{d}{2}}\exp\left(-\frac{1}{2\sigma_{s|t}^2}\left(x_s - \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2}x_t - \frac{\alpha_s\sigma_{t|s}^2}{\sigma_t^2}\frac{x_t - \sigma_t\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \boldsymbol{t})}{\boldsymbol{\alpha}_t}\right)^2\right)$$

- When $\boldsymbol{t} = \mathbf{1}$, $\boldsymbol{\alpha}_{\mathbf{1}} = 0$, $\tilde{p}(x_s|x_t)$ will encounter a **singularity** divided by 0.
- When $\boldsymbol{s} = \mathbf{0}$, $t \to 0$, $\boldsymbol{\sigma}_{\mathbf{0}|\boldsymbol{t}} = 0$, $\tilde{p}(x_s|x_t)$ will encounter a Gaussian with zero variance, i.e. a **singular** distribution.

Reverse Gaussian Properties

Proposition 1 (Error Bound Estimated by $\boldsymbol{\sigma_{s|t}}$). $\forall s \in (0,1)$, $\exists \tau \in (s,1)$ and $C > 0$, such that $\forall t \in (s,\tau]$, $\int_{\mathbb{R}^d} |\boldsymbol{p}(\boldsymbol{x}_s|\boldsymbol{x}_t) - \tilde{\boldsymbol{p}}(\boldsymbol{x}_s|\boldsymbol{x}_t)| d\,x_s < C\sqrt{\sigma_{s|t}}$.

Proposition 2 (Error Bound Estimated by $\boldsymbol{\alpha_s}$). $\exists \nu \in (0,1)$ and $C > 0$, such that $\forall \nu \le s < t \le 1$, $\int_{\mathbb{R}^d} |\boldsymbol{p}(\boldsymbol{x}_s|\boldsymbol{x}_t) - \tilde{\boldsymbol{p}}(\boldsymbol{x}_s|\boldsymbol{x}_t)| d\,x_s < C\sqrt{\alpha_s}$.

Mixed Gaussian Distribution

$$\boldsymbol{p}(\boldsymbol{x}_s|\boldsymbol{x}_t) = (2\pi\sigma_{s|t}^2)^{-\frac{d}{2}} \sum_{i=1}^{N} \exp\left(-\frac{1}{2\sigma_{s|t}^2}\left(x_s - \frac{\alpha_{t|s}\sigma_s^2 x_t}{\sigma_t^2} - \frac{\alpha_s \sigma_{t|s}^2 y_i}{\sigma_t^2}\right)^2\right)\omega_i(x_t, t)$$

Error Bound

$$\tilde{\boldsymbol{p}}(\boldsymbol{x}_s|\boldsymbol{x}_t) = (2\pi\sigma_{s|t}^2)^{-\frac{d}{2}} \exp\left(-\frac{1}{2\sigma_{s|t}^2}\left(x_s - \frac{\alpha_{t|s}\sigma_s^2 x_t}{\sigma_t^2} - \frac{\alpha_s \sigma_{t|s}^2 \bar{y}(x_t, t)}{\sigma_t^2}\right)^2\right)$$

Gaussian Distribution

Sampling at $t = 1$

- Sampling process:

$$x_s = \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2} x_t + \frac{\alpha_s \sigma_{t|s}^2}{\sigma_t^2} \frac{x_t - \sigma_t \epsilon_\theta(x_t, t)}{\boxed{\boldsymbol{\alpha_t}}} + \sigma_{s|t} z_t$$

$t = 1$, $\boldsymbol{\alpha_t} = 0$, resulting in a division-by-zero singularity.

- Singularity is removable:

$$\lim_{t \to 1^-} \frac{x_t - \sigma_t \epsilon_\theta(x_t, t)}{\alpha_t} = \lim_{t \to 1^-} \bar{y}(x_t, t) = \frac{1}{N} \sum_i y_i$$

Replace ϵ-prediction with **x-prediction** at $t = 1$, i.e., estimating $\boldsymbol{\bar{y}(x_1, 1)}$.

Sampling at $t = 0$

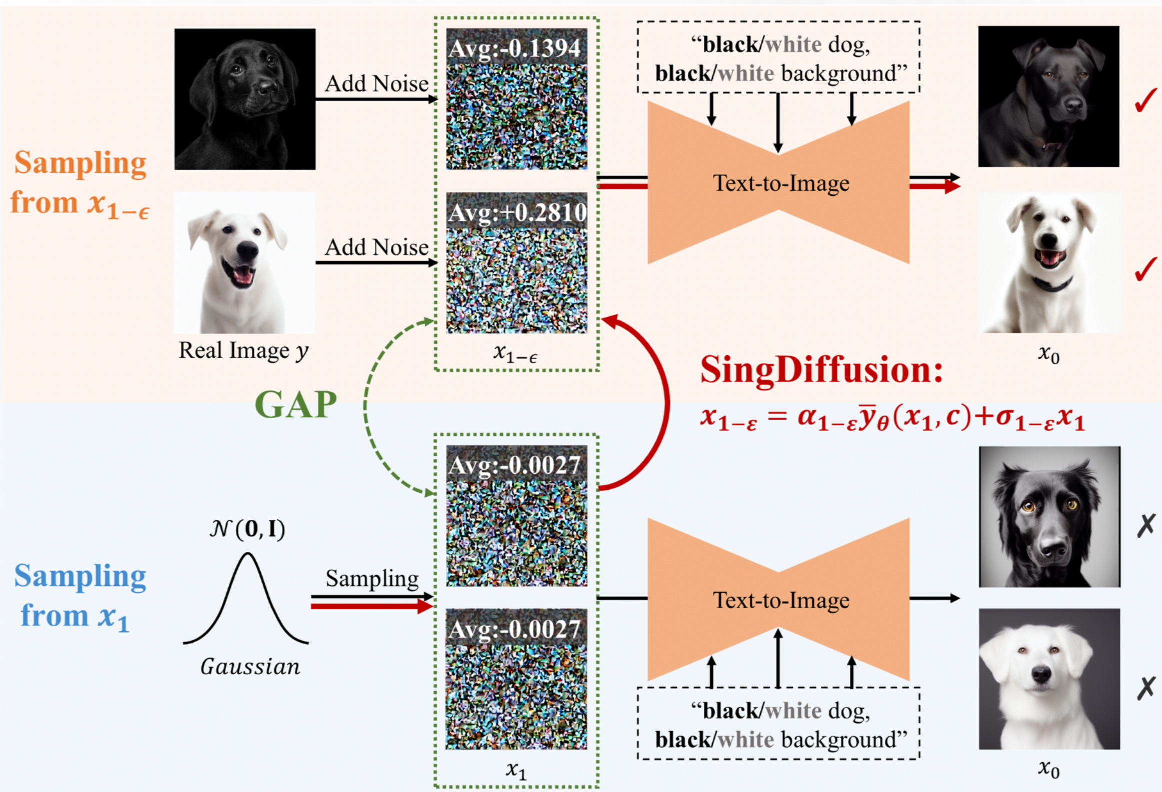
- When $s = 0$ and t is small, $\tilde{p}(x_0|x_t)$ degenerates into a **singular distribution**, a Gaussian with zero variance:

$$\tilde{p}(x_0|x_t) = \delta(x_0 - y_{j_0})$$

Where $j_0 = \arg\min_{\mathrm{j}} |x_t - \alpha_t y_j|$.

- This singularity directs the sampling process to converge at the **correct point** $y_{j_0} = \bar{y}(x_0, 0)$, which is **no need to avoid**.

SingDiffusion



Algorithm 1 Training

repeat
  $x_0, c \sim p(x_0, c)$, $x_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
  Take gradient descent step on $\nabla_\theta \left\| \bar{y}_\theta(x_1, c) - x_0 \right\|^2$
until converged

Algorithm 2 Sampling

$x_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
$\varepsilon = 1/T$
$x_{1-\varepsilon} = \alpha_{1-\varepsilon}\bar{y}_\theta(x_1, c) + \sigma_{1-\varepsilon}x_1$
for $t = 1 - \varepsilon, \ldots, \varepsilon$ **do**
  Calculate $x_{t-\varepsilon}$ using existing sampling algorithms
end for
return x_0

Experiments – Average Brightness Issue

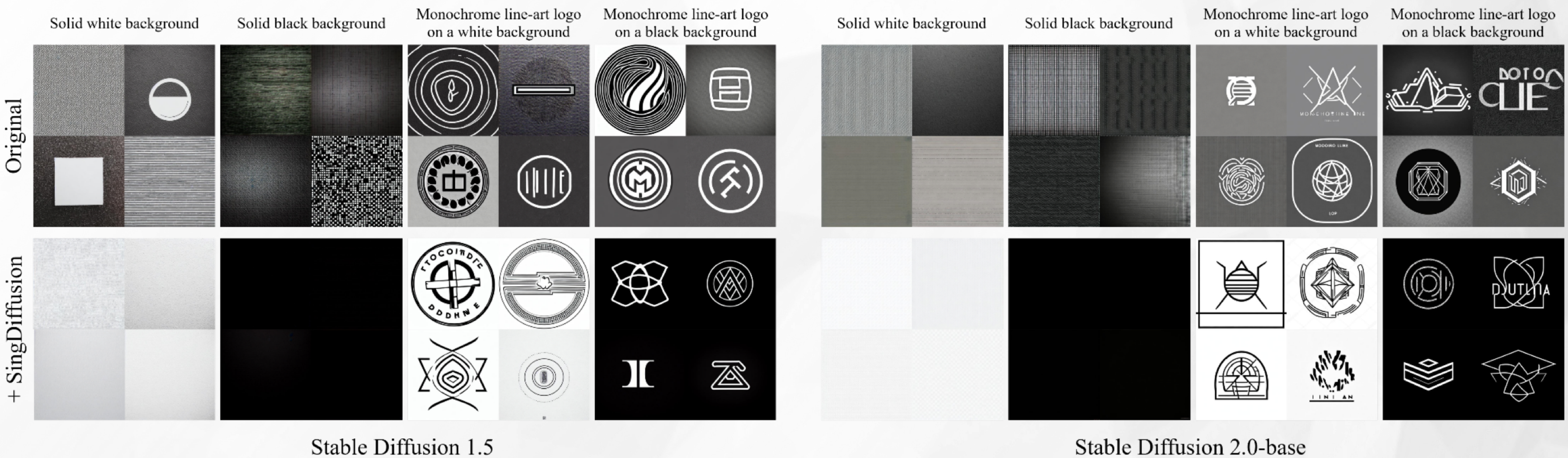


Figure 3. Comparison of stable diffusion models and SingDiffusion on average brightness issue.

Table 1. Comparison of average brightness.

Model	"Solid **white** background"	"Solid **black** background"	"Monochrome line-art logo on a **white** background"	"Monochrome line-art logo on a **black** background"
SD-1.5	141.43	83.09	137.95	113.66
+ Ours	**212.59**	**3.04**	**223.68**	**11.52**
SD-2.0-base	150.52	99.67	136.13	104.45
+ Ours	**227.43**	**0.29**	**228.68**	**10.87**

Experiments - FID v.s. CLIP

Table 2. Comparison of stable diffusion model and SingDiffusion on FID score and CLIP score without classifier guidance.

Model	SD-1.5		SD-2.0-base	
	FID ↓	CLIP ↑	FID ↓	CLIP ↑
Original	31.86	26.70	25.17	27.48
+ SingDiffusion	**21.09**	**27.71**	**18.01**	**28.23**

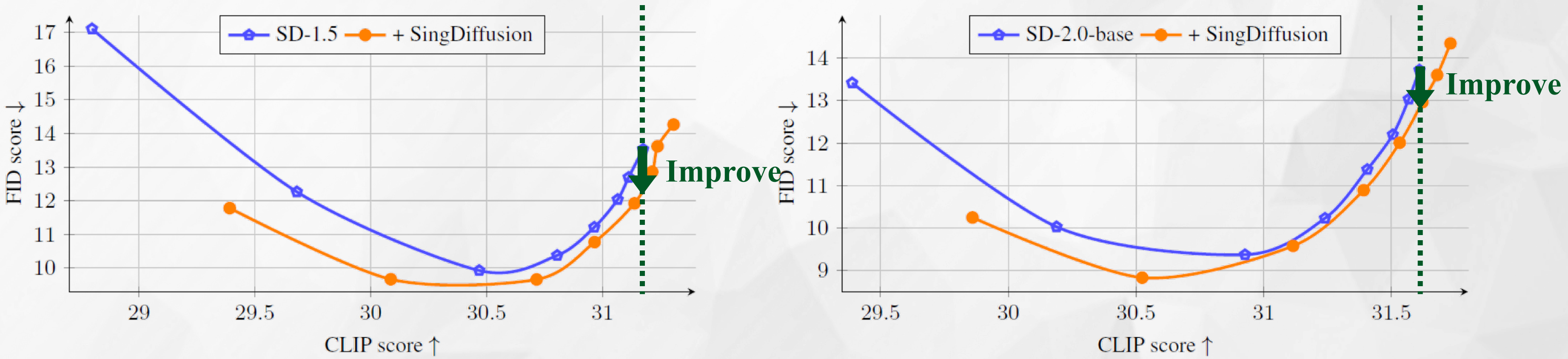


Figure 4. Comparison of Pareto curves between SingDiffusion, SD-1.5, and SD-2.0-base on 30k COCO images, across various guidance scales in [1.5, 2, 3, 4, 5, 6, 7, 8].

Experiments - Plug-and-Play

Figure 1. Our method can be trained once and seamlessly integrated into different pre-trained models in a plug-and-play fashion.

Experiments – Integrate with ControlNet

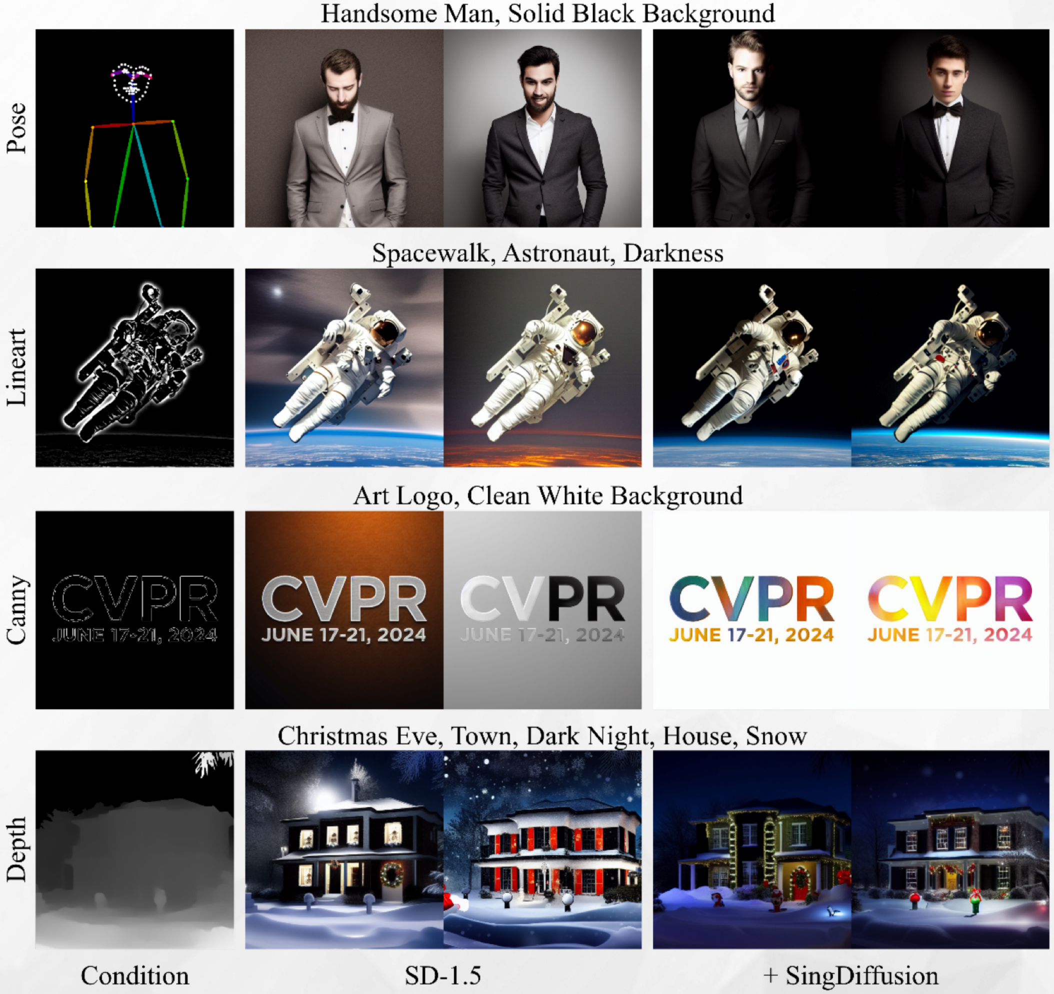


Figure 8. SingDiffusion integrates seamlessly with ControlNet.

THANKS

https://pangzecheung.github.io/SingDiffusion/