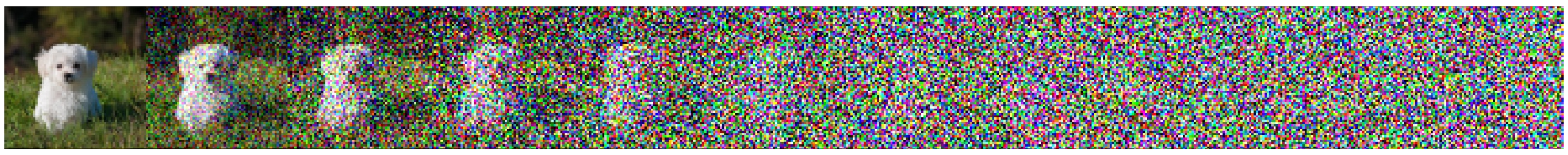
Controlling Style in Diffusion Models though Noise

Martin Nicolas Everaert, Sabine Süsstrunk, Radhakrishna Achanta
Image and Visual Representation Lab - EPFL

Forward Diffusion (Noising)

x_0 $\qquad x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon$ with $\varepsilon \sim \mathcal{N}(0_d, \mathrm{I}_{d \times d})$ $\qquad x_{1000}$



$\hat{x}_0$ $\qquad \hat{x}_t$ $\qquad \hat{x}_{1000}$

Reverse Diffusion (Generation/Sampling)

Image from [2]

Diffusion in Style

Martin Nicolas Everaert
Marco Bocchio
Sami Arpa
Sabine Süsstrunk
Radhakrishna Achanta

ICCV23 PARIS

The initial noise $\hat{x}_{1000}$ affects the style of the generated image $\hat{x}_0$, so adapting it to the style facilitates style adaptation.

We fine-tune Stable Diffusion (SD) [1] with a **style-specific noise distribution** $\mathcal{N}(\mu_{style}, \Sigma_{style})$ instead of the default $\mathcal{N}(0_d, \mathrm{I}_{d \times d})$.

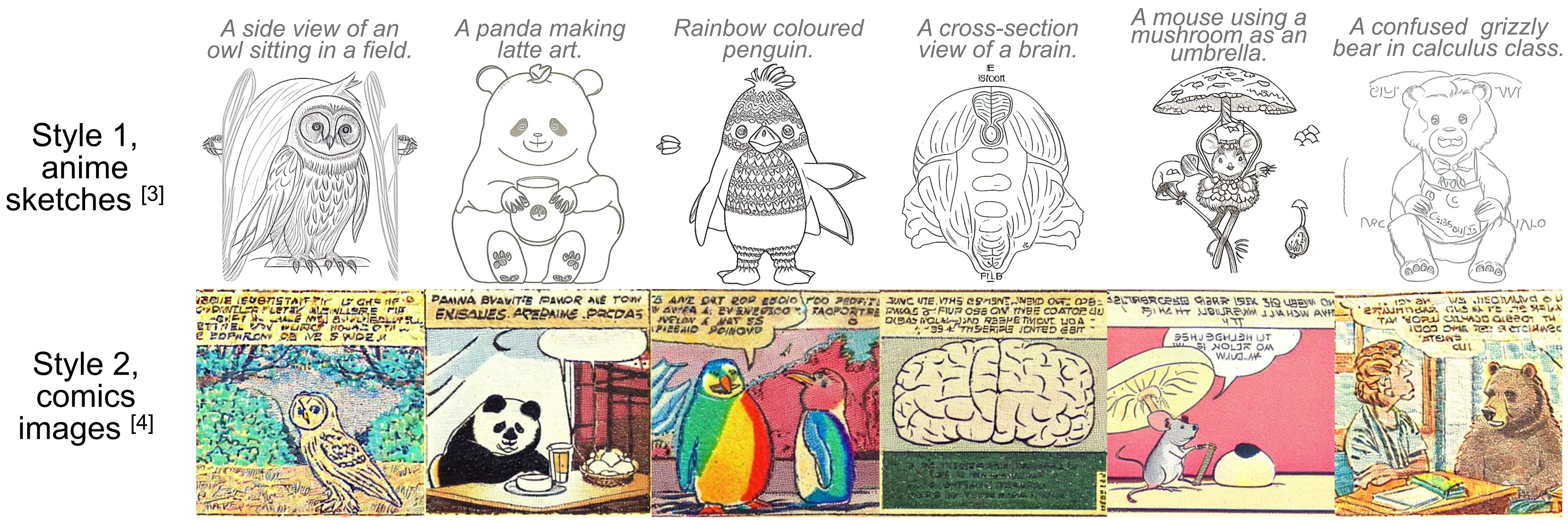
Original diffusion
$\varepsilon \sim \mathcal{N}(0_d, \mathrm{I}_{d \times d})$

Our style-adapted diffusion
$\varepsilon \sim \mathcal{N}(\mu_{style}, \Sigma_{style})$

x_0 $\quad x_{333}$ $\quad x_{667}$ $\quad x_{1000}$

We compute the style-specific noise parameters μ_{style} and Σ_{style} from **a small set of images of the desired style**. We use the fine-tuned model to denoise the initial noise $\hat{x}_{1000} \sim \mathcal{N}(\mu_{style}, \Sigma_{style})$.

We use our approach to fine-tune SD 1.5 [1] to different styles, *e.g.* anime sketches, or comics images.



$\hat{x}_{1000}$ $\quad \hat{x}_{667}$ $\quad \hat{x}_{333}$ $\quad \hat{x}_0$ $\quad \hat{x}_{1000}$ $\quad \hat{x}_{667}$ $\quad \hat{x}_{333}$ $\quad \hat{x}_0$

Exploiting the Signal-Leak Bias in Diffusion Models

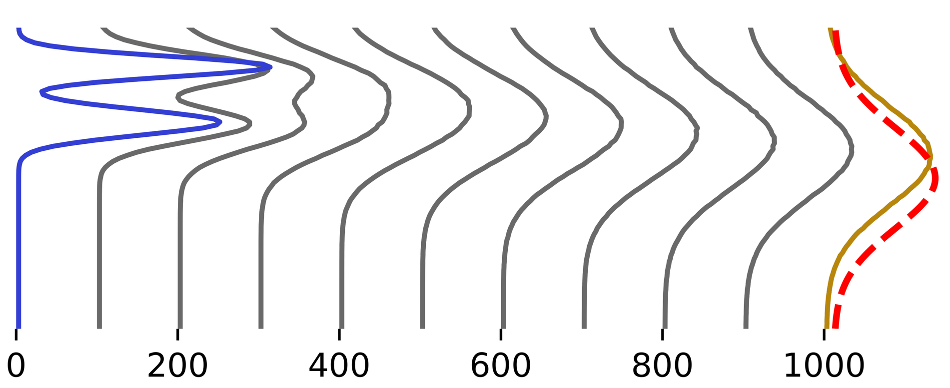
Martin Nicolas Everaert
Athanasios Fitsios
Marco Bocchio
Sami Arpa
Sabine Süsstrunk
Radhakrishna Achanta

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

Diffusion models never fully corrupt images during training [5,6]:

$x_{1000} = \sqrt{\bar{\alpha}_{1000}}\, x_0 + \sqrt{1-\bar{\alpha}_{1000}}\, \varepsilon$ with $x_0 \sim p(x_0)$ and $\varepsilon \sim \mathcal{N}(0_d, \mathrm{I}_{d \times d})$
$\approx 0.068\, x_0 \; + 0.998\, \varepsilon$

However, the process of **generating images starts with pure noise** $\hat{x}_{1000} \sim \mathcal{N}(0_d, \mathrm{I}_{d \times d})$, oblivious of the **signal leak** $\sqrt{\bar{\alpha}_{1000}}\, x_0$ present in x_{1000} during training, **creating a bias**.



$\hat{x}_{1000} \sim \mathcal{N}(0_d, \mathrm{I}_{d \times d})$
$\neq p(x_{1000})$

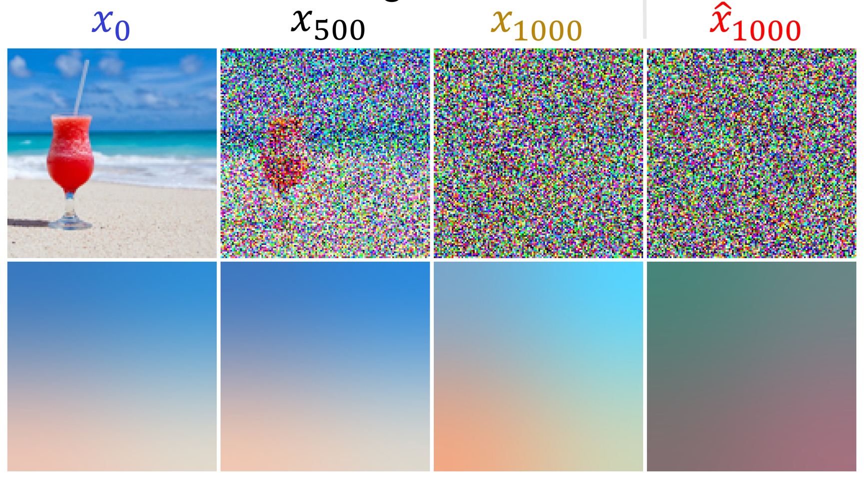
The diffusion model uses the signal-leak $\sqrt{\bar{\alpha}_{1000}}\, x_0$ in x_{1000} to deduce the **low-frequency information** about x_0. Using $\hat{x}_{1000} \sim \mathcal{N}(0_d, \mathrm{I}_{d \times d})$ **biases** the low-frequency components towards **medium values**.

Training $\qquad$ Inference

x_0 $\quad x_{500}$ $\quad x_{1000}$ $\quad \hat{x}_{1000}$

x_t

3 lowest frequency components $(/\sqrt{\bar{\alpha}_t})$



Instead of retraining or fine-tuning [5,6,A] to remove this bias, we exploit it to our advantage by **including a signal-leak** $\sqrt{\bar{\alpha}_{1000}}\, \tilde{x}$ in $\hat{x}_{1000}$ **at inference time**, starting generating images from:

$$\hat{x}_{1000} = \sqrt{\bar{\alpha}_{1000}}\, \tilde{x} + \sqrt{1-\bar{\alpha}_{1000}}\, \varepsilon \quad \text{with } \tilde{x} \sim q(\tilde{x}) \text{ and } \varepsilon \sim \mathcal{N}(0_d, \mathrm{I}_{d \times d})$$

With $q(\tilde{x}) = \mathcal{N}(\mu_{style}, \Sigma_{style})$, we exploit the bias to generate images $\hat{x}_0$ in the style we want:

At inference time, we can control the low-frequency components of the generated images $\hat{x}_0$ by setting the desired ones (here, the mean color) in $\tilde{x}$:

ICCP2024 Poster

References:

[A] Everaert et al. "Diffusion in style." ***ICCV*** **2023.**
[B] Everaert et al. "Exploiting the signal-leak bias in diffusion models." ***WACV*** **2024.**
[1] Rombach et al. "High-resolution image synthesis with latent diffusion models." *CVPR* 2022.
[2] Nichol and Dhariwal. "Improved denoising diffusion probabilistic models." *ICML* 2021.
[3] Taebum. "Anime Sketch Colorization dataset." Kaggle dataset. 2018.
[4] Simon and Kirby. "48 Famous Americans." 1947.
[5] Guttenberg. "Diffusion with Offset Noise." 2023
[6] Lin et al. "Common Diffusion Noise Schedules and Sample Steps are Flawed." *WACV* 2024.
[7] Karan. “line-art” model. via huggingface.co/sd-concepts-library. 2022.
[8] MatAIart. “nasa space” model. Via huggingface.co/sd-dreambooth-library. 2022

Diffusion in Style

Signal-Leak Bias

Acknowledgement:
The 2 works were supported by Innosuisse grant 48552.1 IP-ICT.

EPFL