

Martin Nicolas Everaert

Marco Bocchio

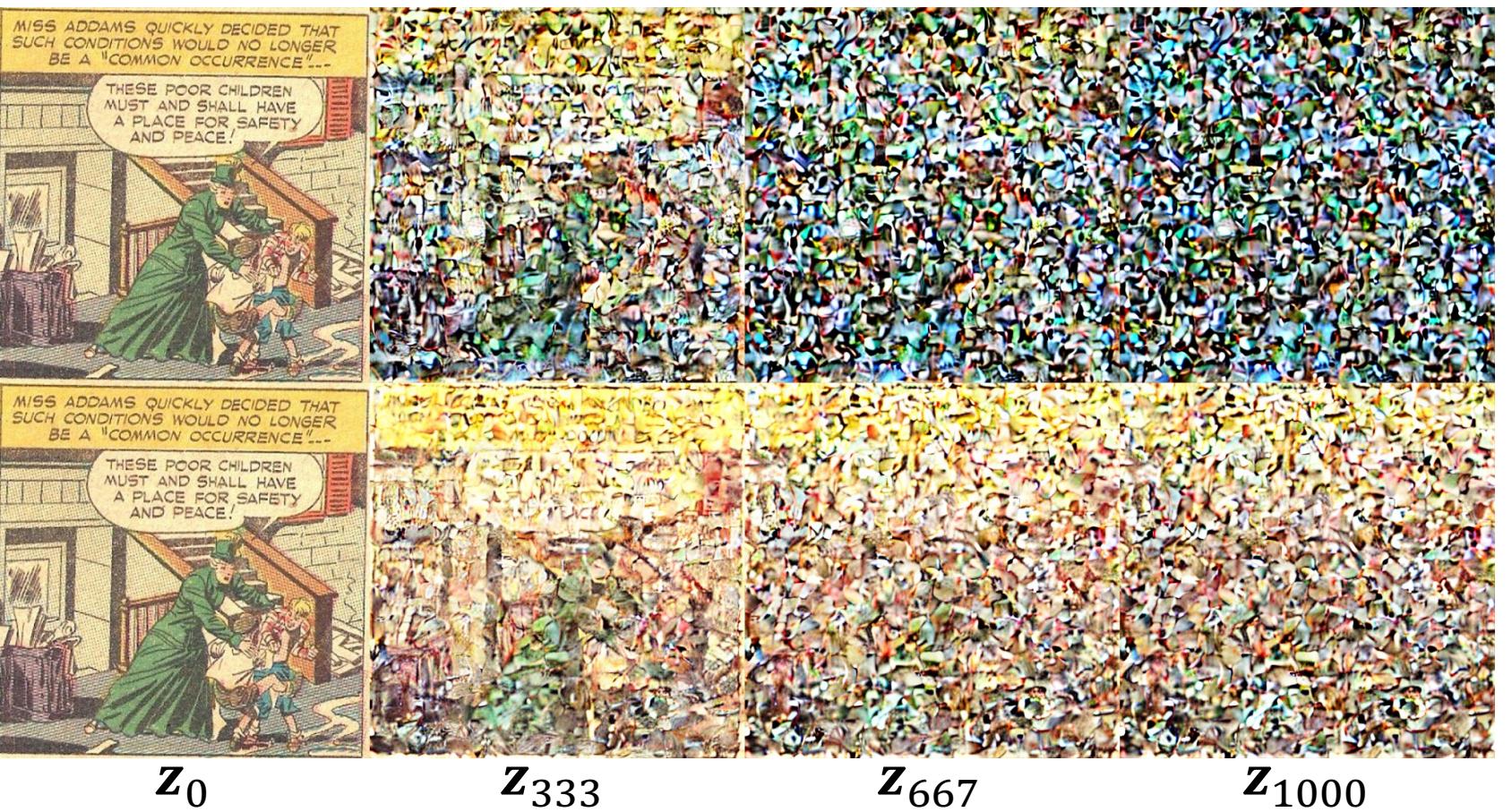
Sami Arpa

Sabine Süsstrunk

Radhakrishna Achanta

Method

We perform **style adaptation** of Stable Diffusion by fine-tuning it with a **style-specific noise distribution** instead of the default $\mathcal{N}(\mathbf{0}_d, \mathbf{I}_{d \times d})$ ^[1,2].

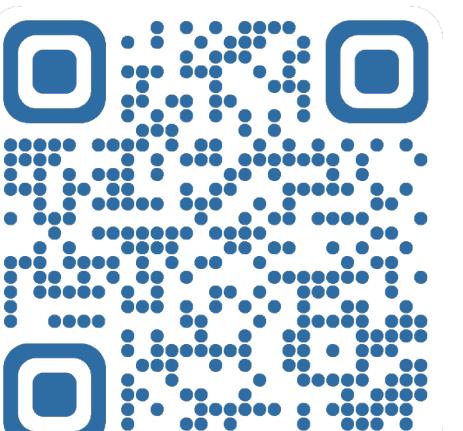
Original diffusion
 $\mathcal{N}(\mathbf{0}_d, \mathbf{I}_{d \times d})$ 

We compute the style-specific noise parameters μ_{style} and Σ_{style} from a **small set of images of the desired style**.

Apart from the style-specific noise distribution $\mathcal{N}(\mu_{\text{style}}, \Sigma_{\text{style}})$, the fine-tuned model **can be used like Stable Diffusion**.

Intuition

The initial latent tensor $\hat{\mathbf{z}}_{1000}$ affects images composition and style, so adapting it to the style facilitates style adaptation.

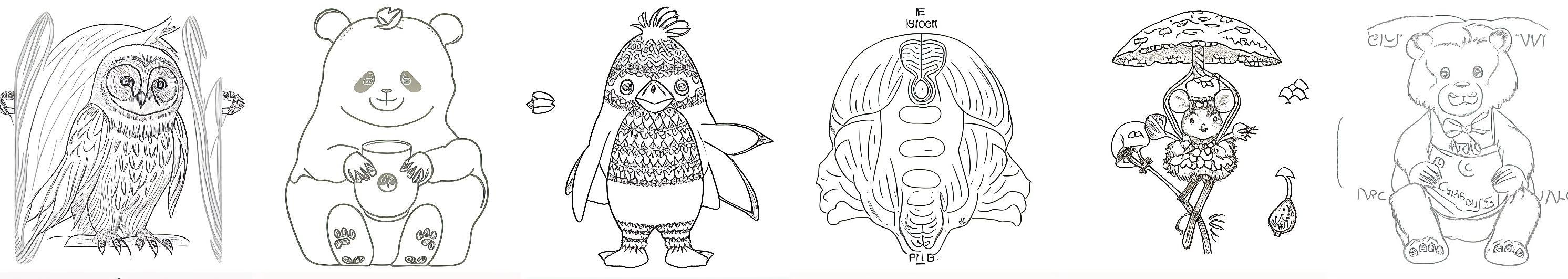
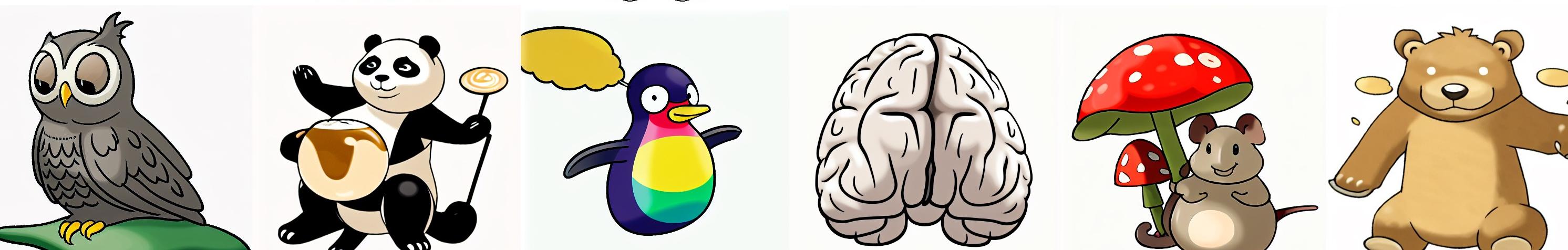
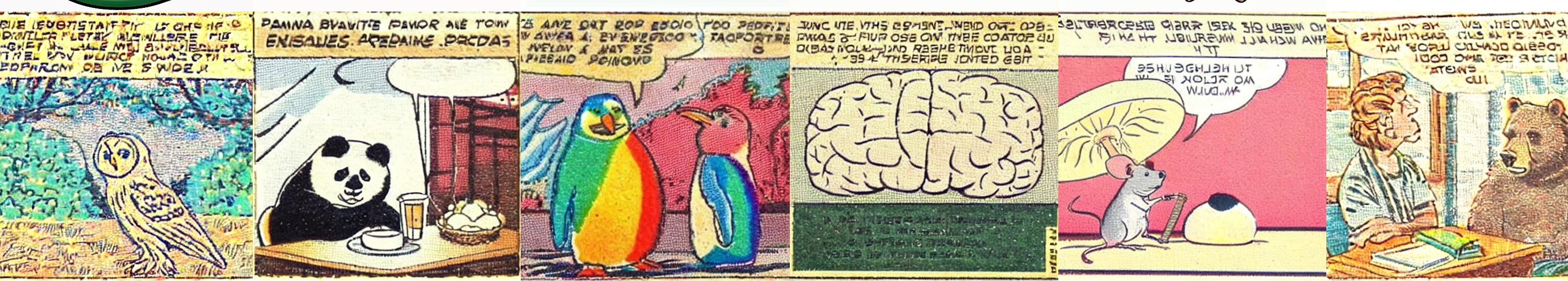


Project website
<https://lrvl.github.io/diffusion-in-style/>

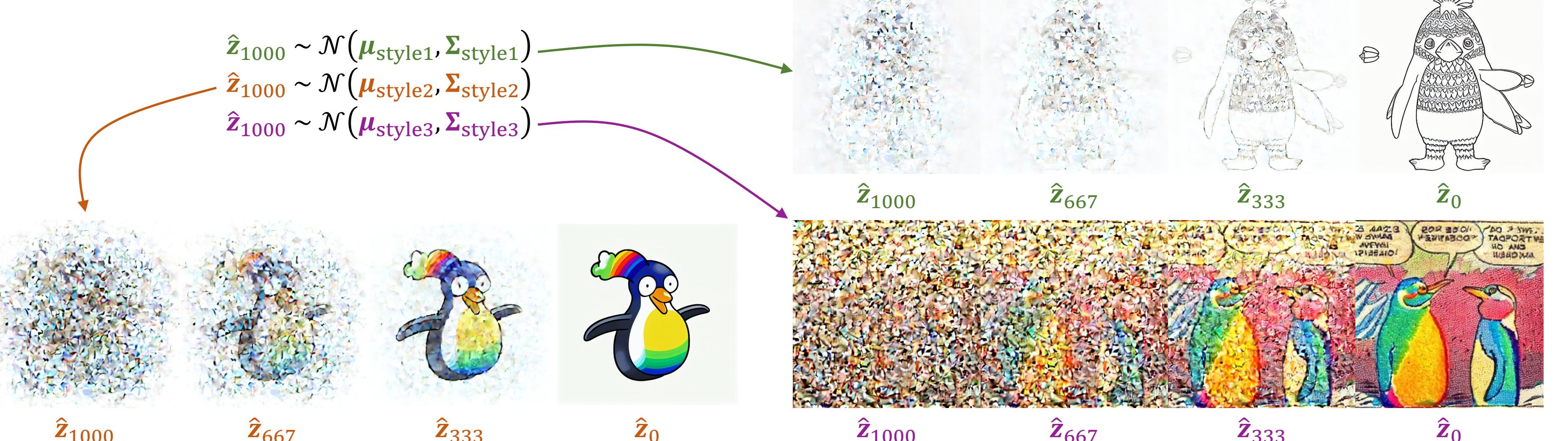
Acknowledgement:
This work is supported by Innosuisse grant 48552.1 IP-ICT.

Results

We use our approach to fine-tune Stable Diffusion v1.5^[1] to different styles, such as **anime sketches**, **few-shot Pokemon images**, and **comics images**.

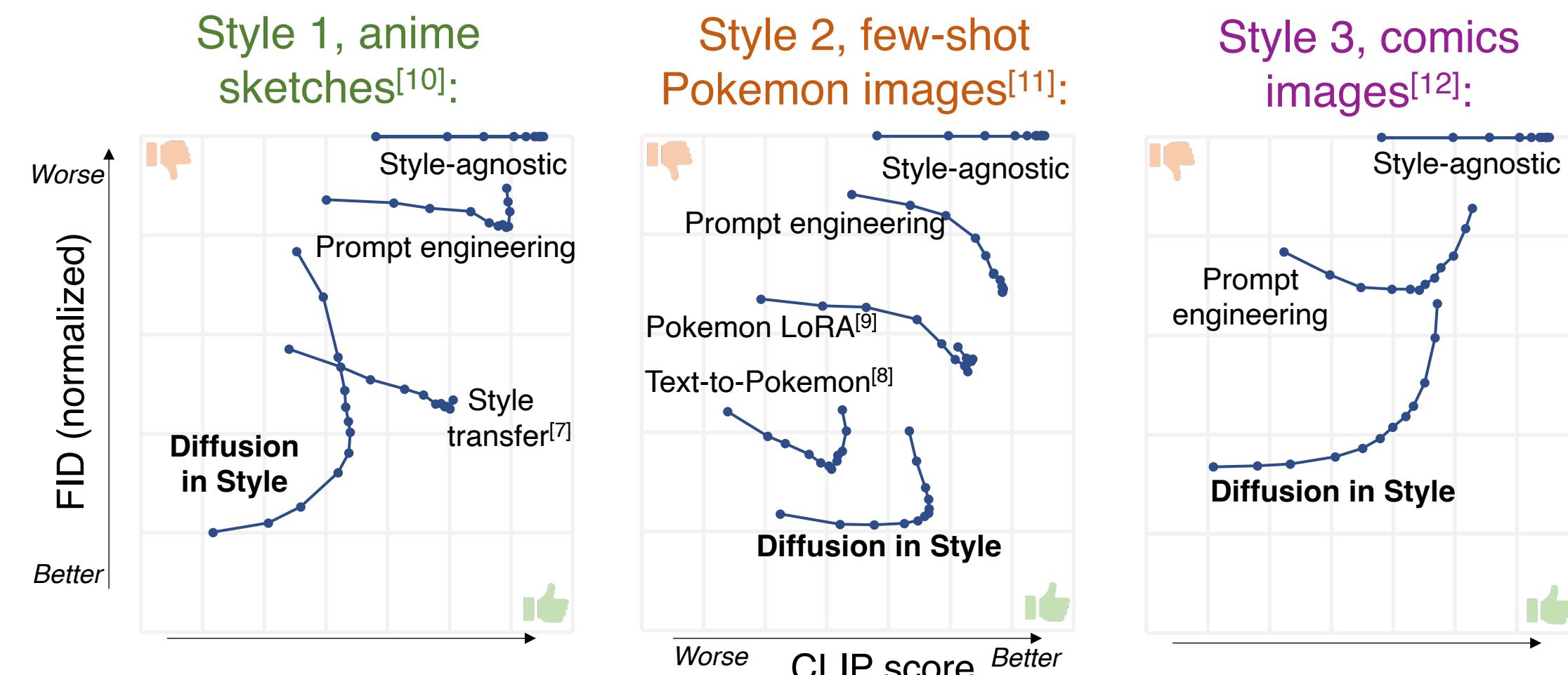
Style 1, anime sketches^[10]:Style 2, few-shot Pokemon images^[11]:Style 3, comics images^[12]:

We sample the initial latent tensor $\hat{\mathbf{z}}_{1000}$ from the style-specific noise distribution and use the fine-tuned U-Net to iteratively denoise it.

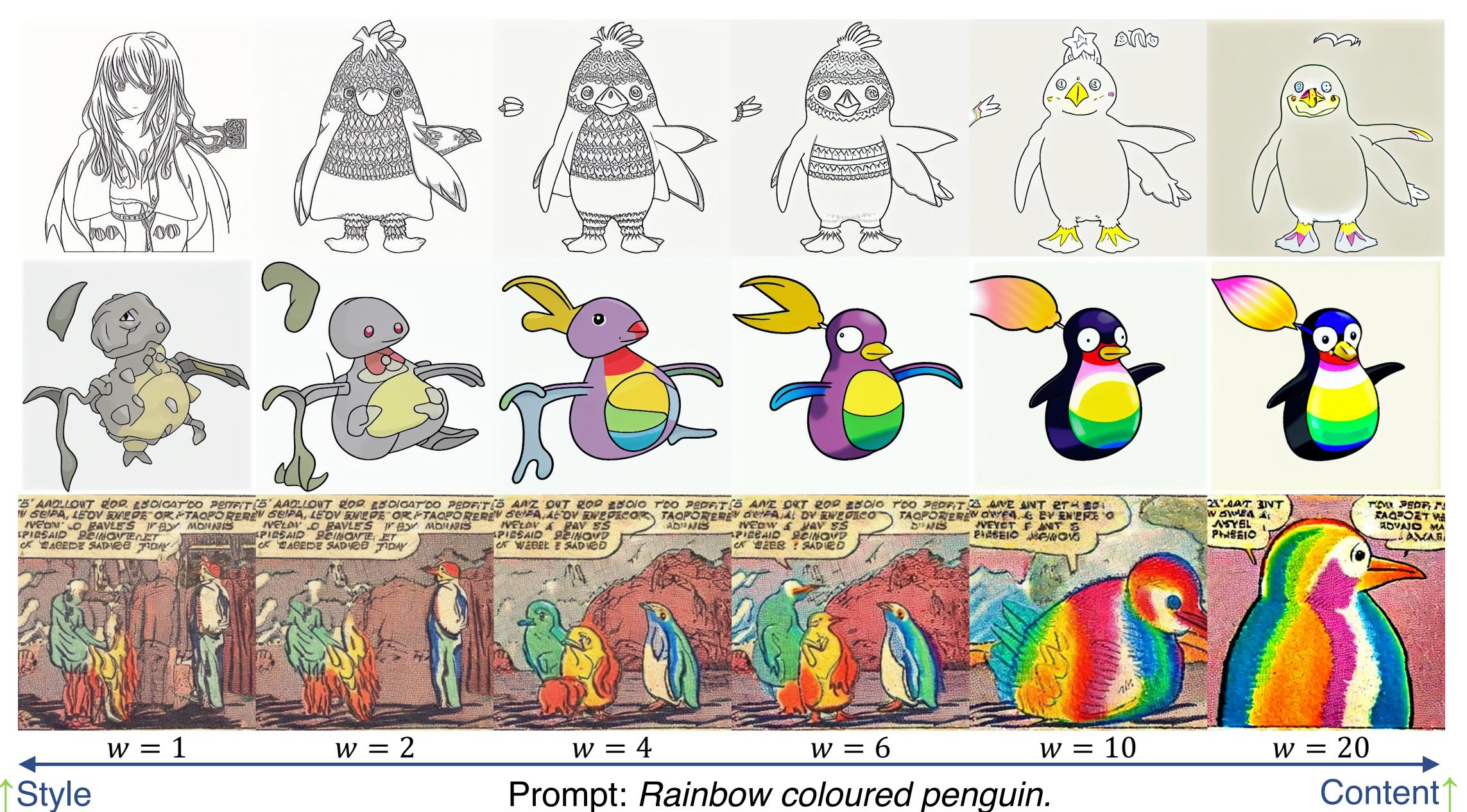


Evaluation

Evaluating **CLIP**^[3] and **FID**^[4,5] scores on a range of guidance weights^[6] w , our method **outperforms** prompt engineering, style transfer^[7], and fine-tuning without noise distribution change^[8,9].



The J-shape of the curves indicates a **trade-off between style and content**.



References

- [1] Rombach et al. CVPR 2022
- [2] Ho et al. NeurIPS 2020
- [3] Radford et al. PMLR 2021
- [4] Heusel et al. NIPS 2017
- [5] Wright et al. GCPR 2022
- [6] Taebum, Anime Sketch Colorization dataset (2018), <https://www.kaggle.com/tktaebum/anime-sketch-colorization-pair>
- [7] Chan et al. CVPR 2022
- [8] Lambda Labs. Text-to-Pokemon model (2022), <https://huggingface.co/lambda-labs/st-d-pokemon-diffusers>
- [9] Liu et al. ICLR 2021, <https://huggingface.co/datasets/huggan/few-shot-pokemon>
- [10] Ho et al. NeurIPS Workshop 2021
- [11] Sayak. Pokemon LoRA model (2023), <https://huggingface.co/sayakpaul/st-d-model-finetuned-lora-t4>
- [12] Simon & Kirby. 48 Famous Americans (1947), <https://digitalcomicmuseum.com/index.php?did=24742>