ICCV HONOLULU HAWAII
OCT 19-23, 2025

Identity Preserving 3D Head Stylization with Multi-view Score Distillation

Bahri Batuhan Bilecen, Ahmet Berke Gokmen, Furkan Guzelant, Aysegul Dundar

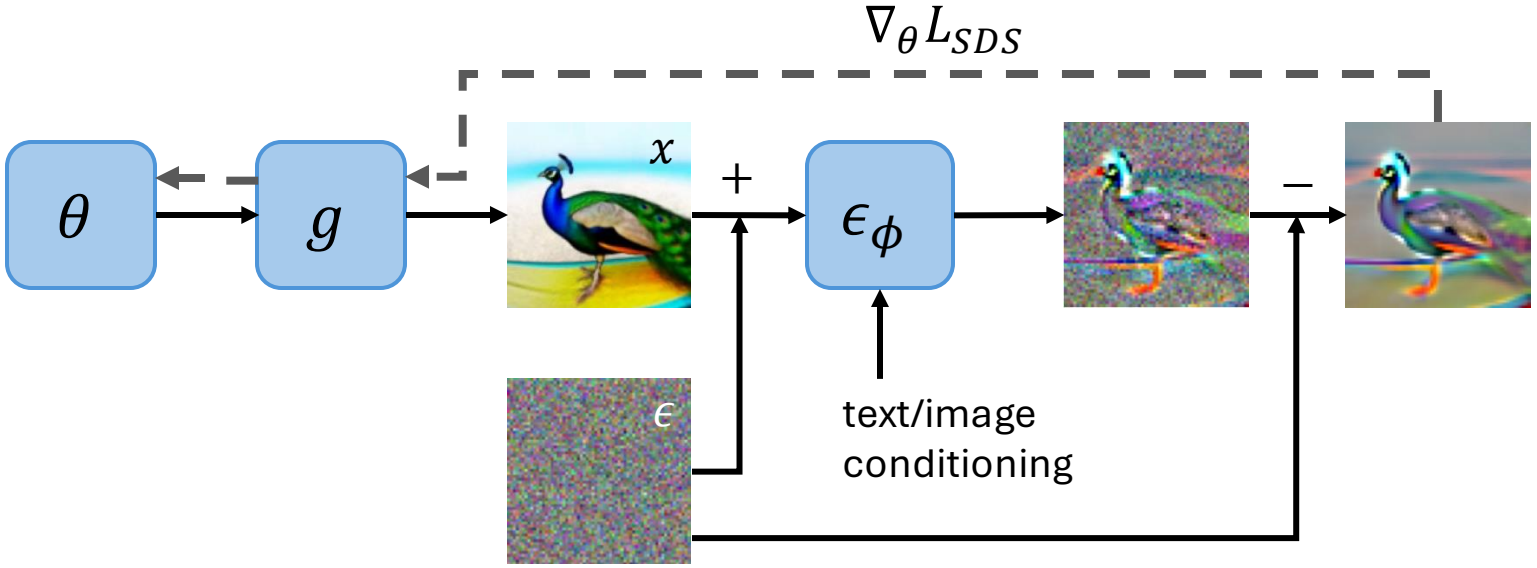
Bilkent University, ETH Zürich, INSAIT, Max Planck Institute

Motivation

3D params θ are rendered with g

Denoiser output

$$\nabla_\theta L_{\text{SDS}}(\phi, x = g(\theta)) \triangleq \mathbb{E}_{t,\epsilon}\left[\omega(t)(\hat{\epsilon}_\phi(x_t; y, t) - \epsilon)\frac{\partial x}{\partial \theta}\right]$$



[1] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "DreamFusion: Text-to-3D using 2D diffusion," in ICLR, 2023.

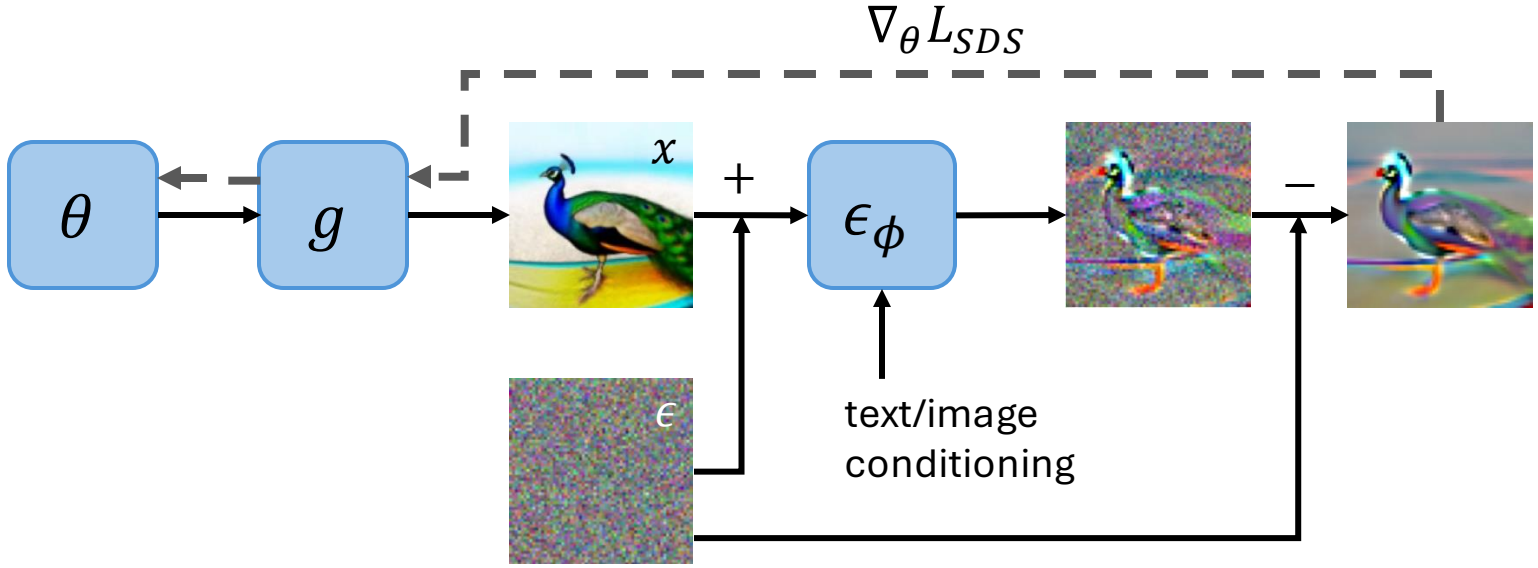
Motivation

3D params θ are rendered with g

When θ are GAN params, **quick mode collapse = ID not preserved**

Denoiser output

$$\nabla_\theta L_{\text{SDS}}(\phi, x = g(\theta)) \triangleq \mathbb{E}_{t,\epsilon}\left[\omega(t)\left(\hat{\epsilon}_\phi(x_t; y, t) - \epsilon\right)\frac{\partial x}{\partial \theta}\right]$$



[1] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "DreamFusion: Text-to-3D using 2D diffusion," in ICLR, 2023.

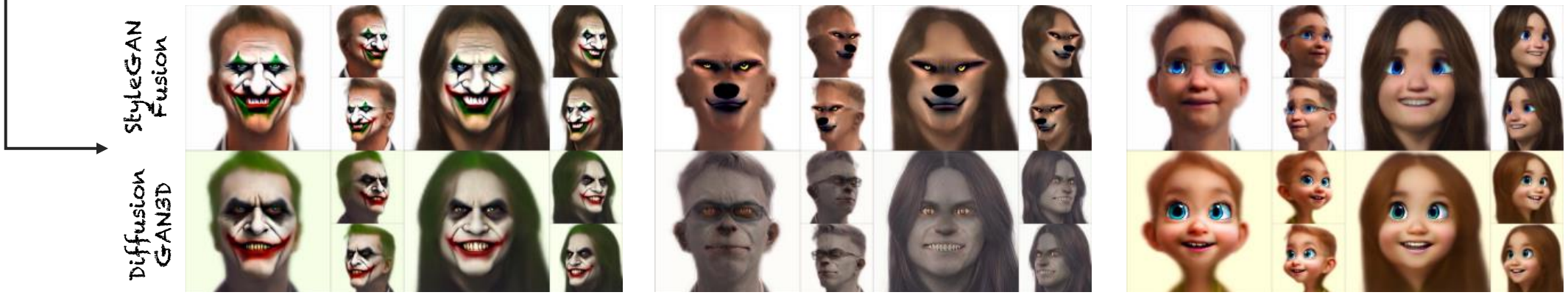
Motivation

3D params θ are rendered with g

When θ are GAN params, **quick mode collapse = ID not preserved**

Denoiser output

$$\nabla_\theta L_{\text{SDS}}(\phi, x = g(\theta)) \triangleq \mathbb{E}_{t,\epsilon}\left[\omega(t)(\hat{\epsilon}_\phi(x_t; y, t) - \epsilon)\frac{\partial x}{\partial \theta}\right]$$

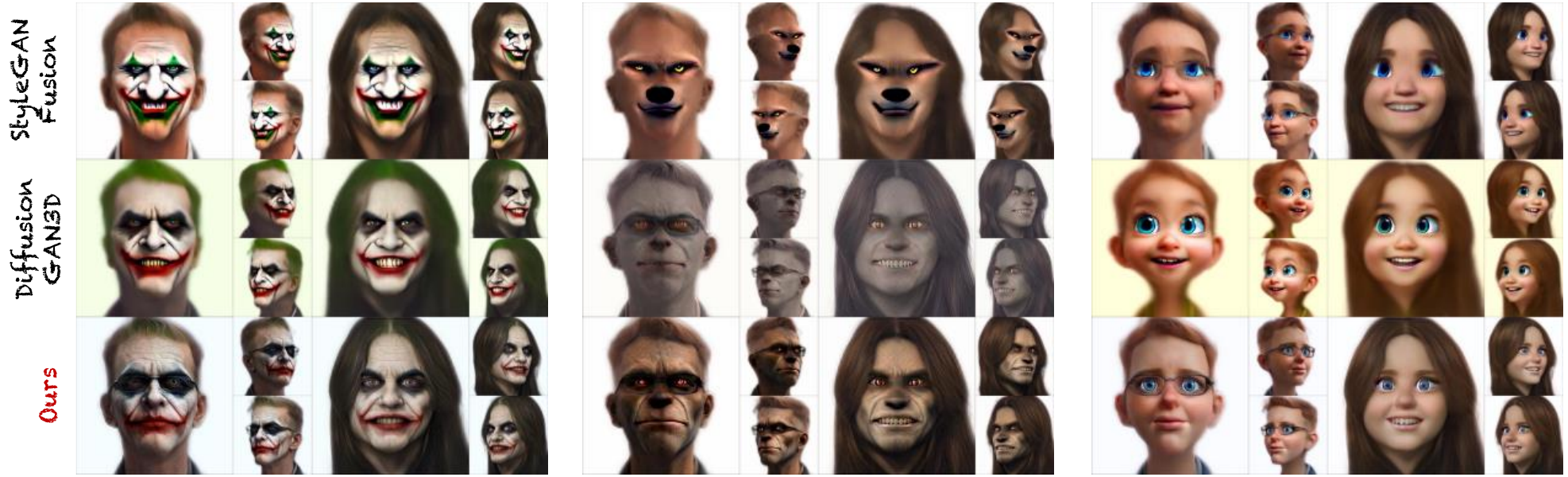


[2] Song et al., "Diffusion guided domain adaptation of image generators", in WACV 2024.
[3] Lei et al., "DiffusionGAN3D: Boosting text-guided 3D generation and domain adaptation by combining 3D GANs and diffusion priors", in CVPR 2024.

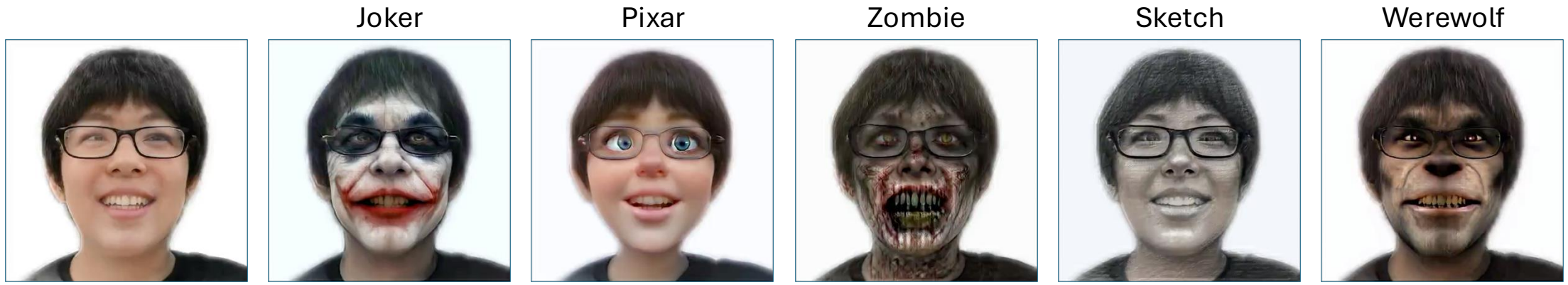
Our approach

In this work, we introduce **log-likelihood distillation (LD), multi-view score distillation** and **score rank weighting to GANs** for domain adaptation, which has significant advantages over SDS.

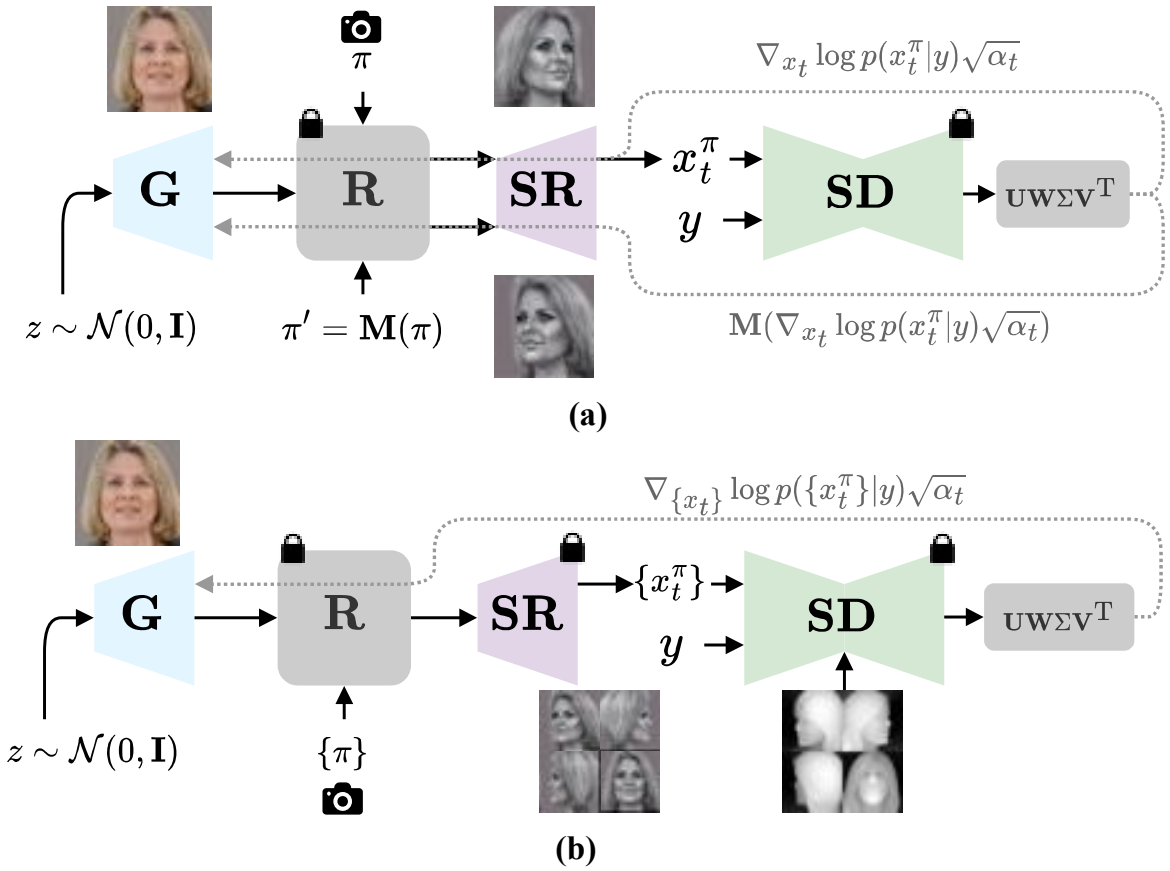
We showcase our results under identity preserving 3D head stylization.



Our approach



Pipeline

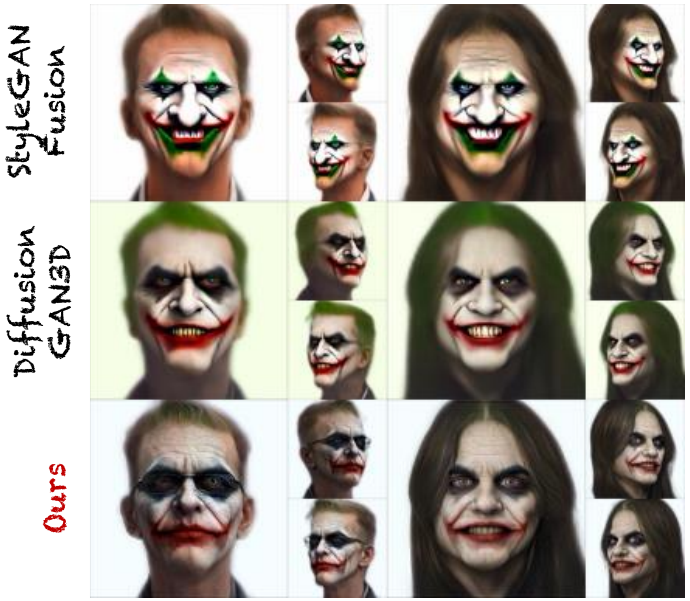


We derive and introduce mirror **(a)** and grid gradients **(b)** to further enforce 3D-consistent and ID-preserving stylization. Furthermore, we omit the gradients through super-resolution layers, and apply score rank re-weighing for eliminating artifacts.

Likelihood distillation instead of SDS

$$\nabla_\theta L_{\text{SDS}}(\phi, x = g(\theta)) \triangleq \mathbb{E}_{t,\epsilon}\left[\omega(t)(\hat{\epsilon}_\phi(x_t; y, t) - \epsilon)\frac{\partial x}{\partial \theta}\right]$$

$$\nabla_\theta \mathcal{L}_{\text{LD}} = -\mathbb{E}_{\pi, x_t}\left\{\nabla_{x_t} \log p(x_t^\pi | y) \frac{\partial x_t^\pi}{\partial x_0^\pi}\frac{\partial x_0^\pi}{\partial \theta}\right\},$$



Subtracting ground-truth noise gives us a low-variance guidance signal, which avoids divergence but **collapses GAN training into a single mode (i.e., identically stylized faces for different identities).**

This approach can be generalized to other feedforward large reconstruction models, to create a stylized generator.

Extending LD to mirror gradients and grid denoising

We divide LD gradient chain into two where poses π and π' do not match:

$$\nabla_\theta \mathcal{L}_{\text{LD}} = -\mathbb{E}_{\pi,x_t}\left\{\nabla_{x_t}\log p(x_t^\pi|y)\frac{\partial x_t^\pi}{\partial x_0^\pi}\frac{\partial x_0^\pi}{\partial \theta} + \sum_{\pi\neq\pi'}\frac{\partial \log p(x_t^\pi|y)}{\partial x_0^{\pi'}}\frac{\partial x_0^{\pi'}}{\partial \theta}\right\}$$

We assume that when input x_0^π is mirrored with flip operator **M**, it yields the same result with the yaw-symmetric pose π':

$$\nabla_\theta \mathcal{L}_{\text{LD}} = -\mathbb{E}_{\pi,x_t}\left\{\nabla_{x_t}\log p(x_t^\pi|y)\sqrt{\bar{\alpha}_t}\left(\frac{\partial x_0^\pi}{\partial \theta} + \mathbf{M}\frac{\partial x_0^{\pi'}}{\partial \theta}\right)\right\}$$

We utilize the same score estimation for mirror poses but also mirror the gradients.

In grid denoising, we create a 2x2 image grid and feed it to the denoiser as a whole:

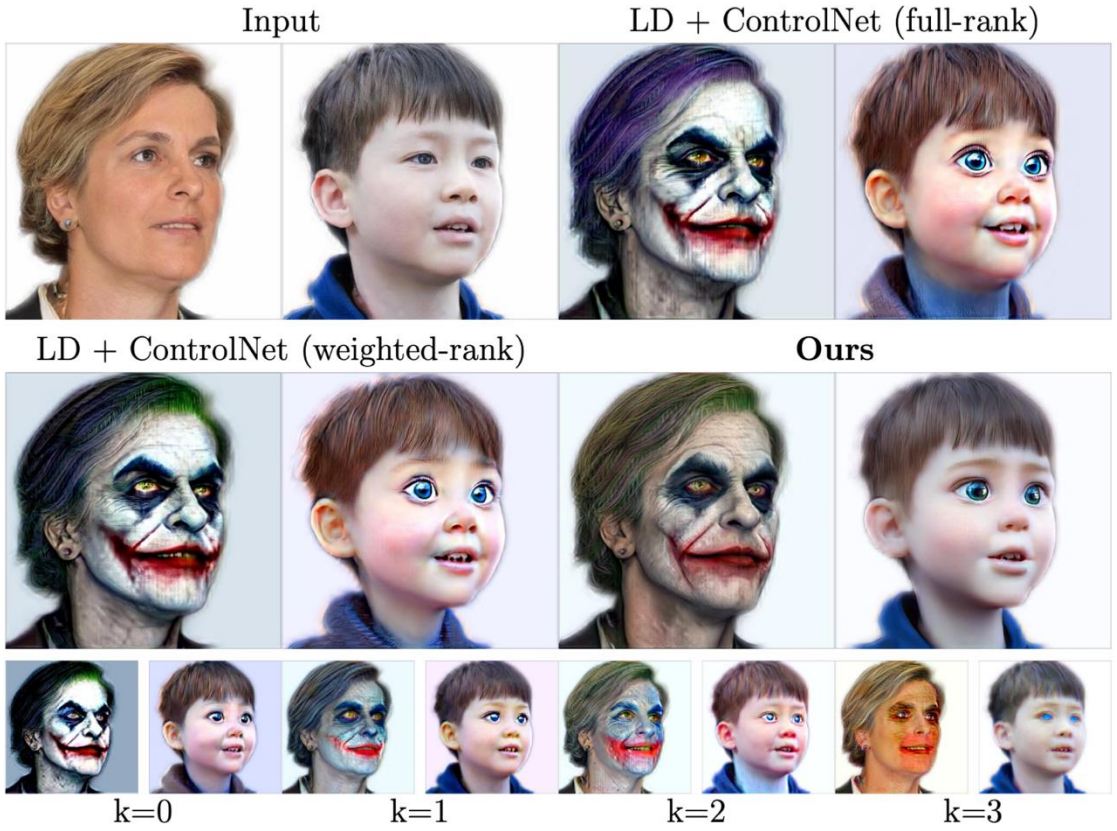
$$\nabla_\theta \mathcal{L}_{\text{LD}_g} = -\mathbb{E}_{\pi,\{x_t\}}\left\{\nabla_{\{x_t\}}\log p(\{x_t^\pi\}|y)\frac{\partial\{x_t^\pi\}}{\partial\{x_0^\pi\}}\frac{\partial\{x_0^\pi\}}{\partial \theta}\right\},$$

Now, denoiser can implicitly correlate between different renders of θ.

Rank-weighing with SVD

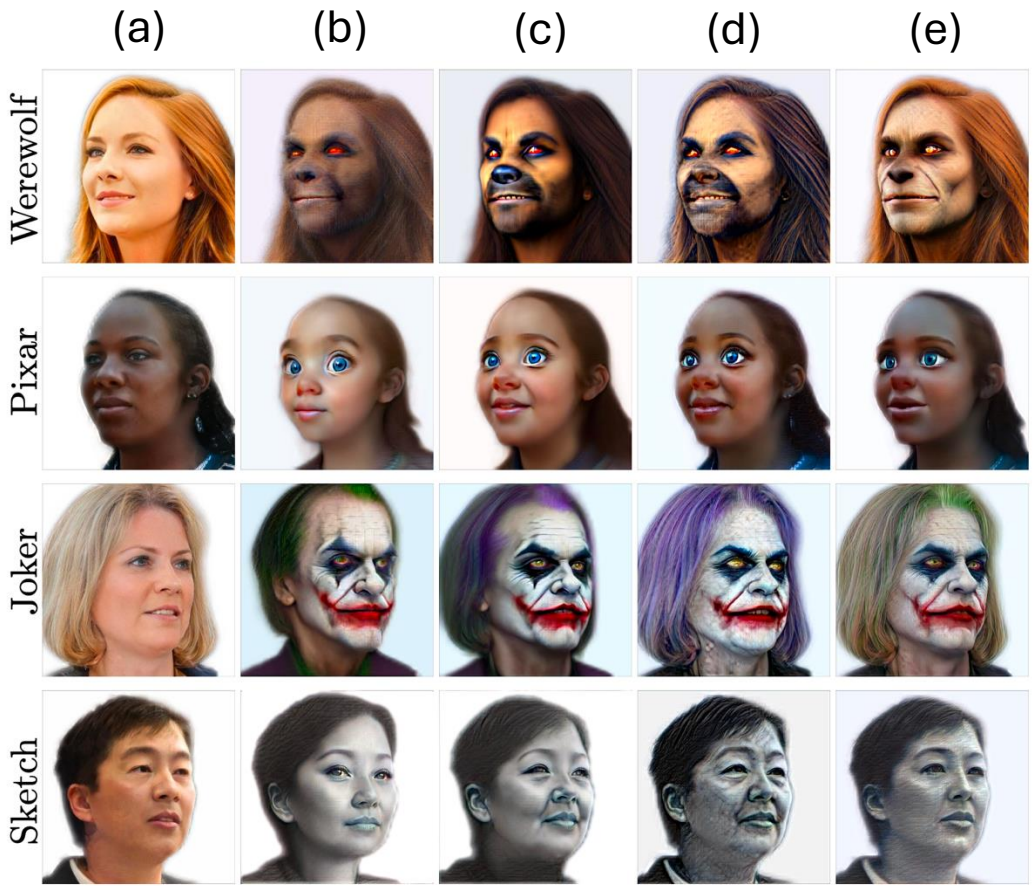
We perform SVD in the latent space channel (4-dim) of the denoiser, and weigh the matrices:

$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}} = \mathrm{SVD}(\nabla_\theta \log p(x_0^\pi|y)),$$
$$\nabla_\theta \log \tilde{p}(x_0^\pi|y) = \mathbf{U}\mathbf{W}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}},$$
$$\mathbf{W} = \mathrm{diag}(1, 0.75, 0.5, 0.25)$$

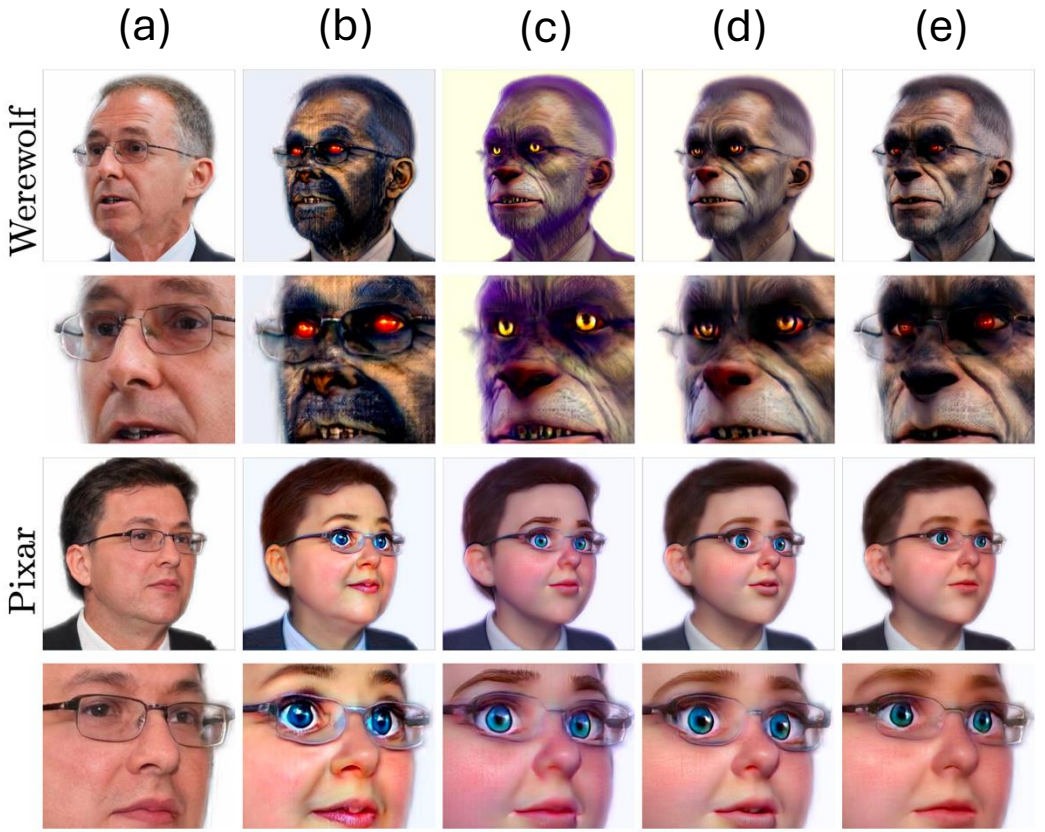


Qualitative ablation on rank-weighing.

Qualitative results

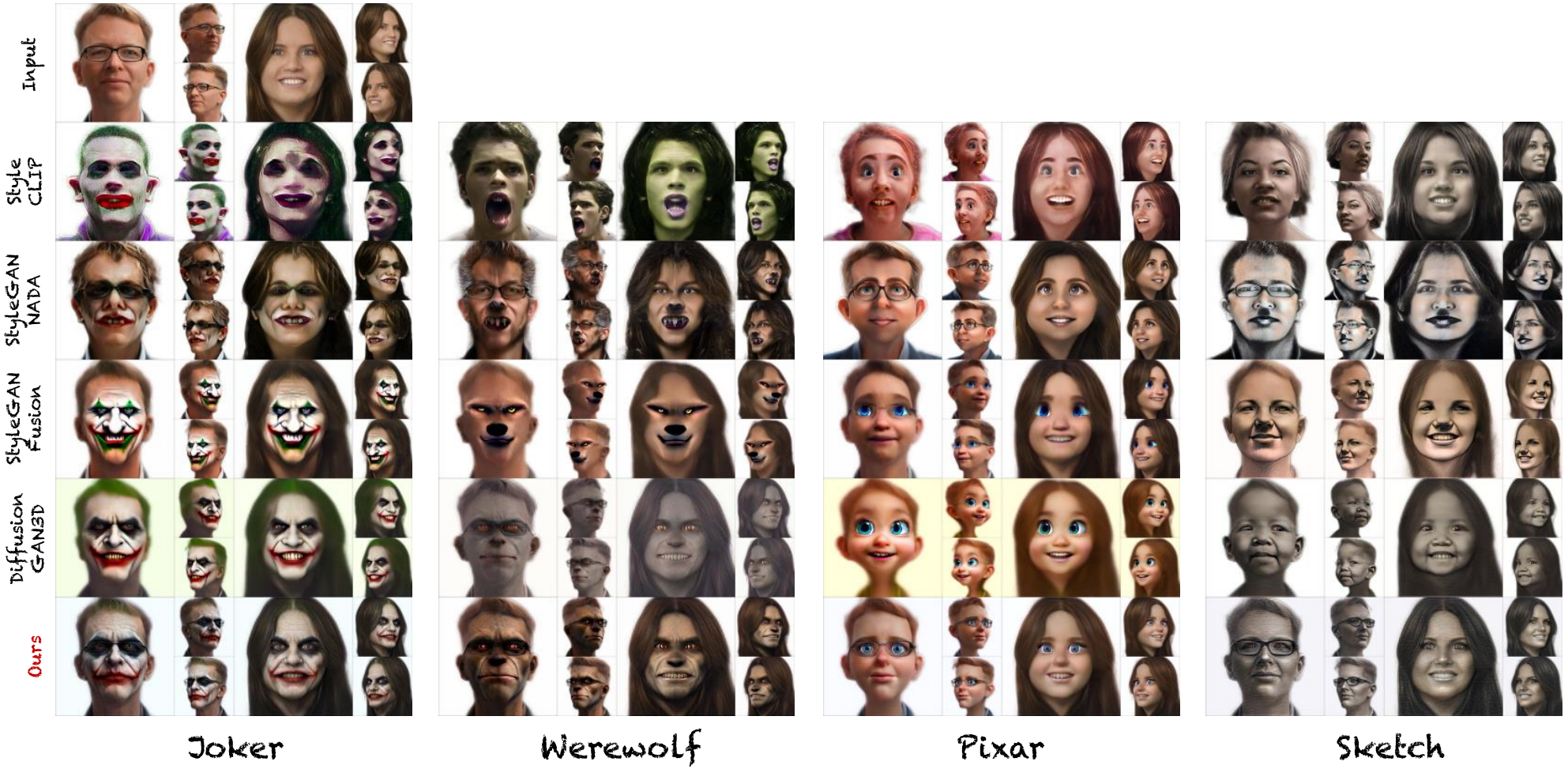


Ablation on the proposed components.
Input **(a)**, SDS **(b)**, SDS+ControlNet **(c)**, LD+ControlNet **(d)**, ours **(e)**.



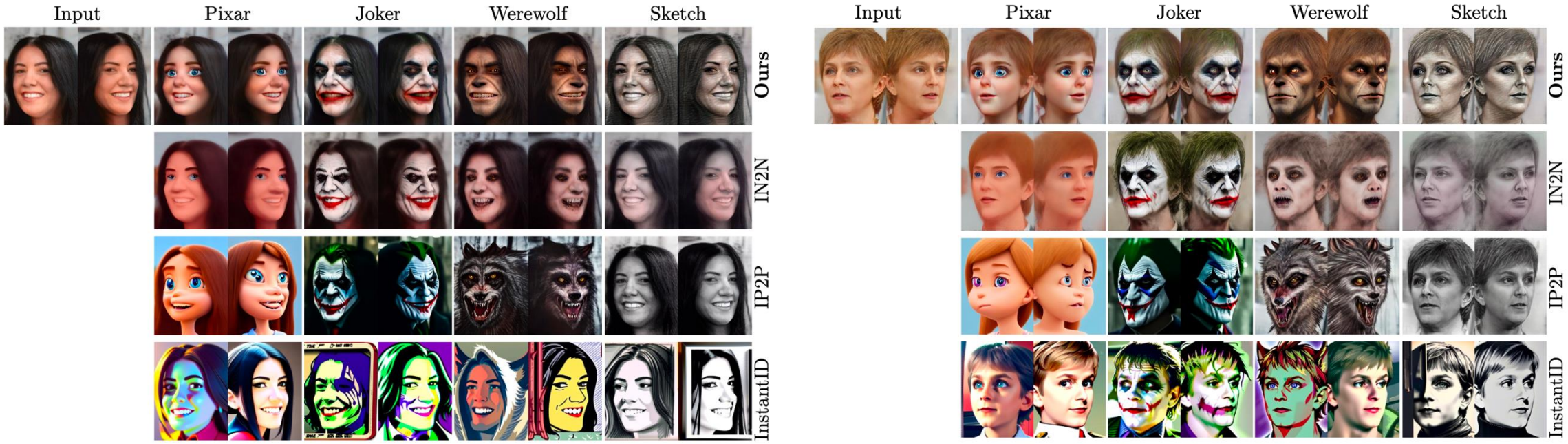
Ablation on the proposed components.
Input **(a)**, LD+ControlNet+rank weighing **(b)**, grid after SR **(c)**, grid before SR **(d)**, ours **(e)**.

Qualitative results



Qualitative results of our head stylization in comparison to previous state-of-the-art.

Qualitative results



Qualitative results of our head stylization in comparison to previous state-of-the-art.

Quantitative results

Quantitative results of our head stylization in comparison to previous state-of-the-art.

		Pixar				Joker				Werewolf				Sketch				Statue			
		FID	CLIP	ID	$\Delta\mathcal{D}$	FID	CLIP	ID	$\Delta\mathcal{D}$	FID	CLIP	ID	$\Delta\mathcal{D}$	FID	CLIP	ID	$\Delta\mathcal{D}$	FID	CLIP	ID	$\Delta\mathcal{D}$
2D	InstructPix2Pix	144.4	0.82	0.40	0.024	116.8	0.90	0.44	0.023	178.1	0.82	0.28	0.013	89.5	0.65	0.24	0.002	60.9	0.91	0.39	0.011
	InstantID	160.8	0.68	0.59	0.050	170.3	0.69	0.47	0.066	183.3	0.58	0.53	0.065	162.7	0.65	0.59	0.047	160.4	0.71	0.56	0.038
3D domain apt.	StyleCLIP	118.2	0.77	0.52	0.022	97.6	0.75	0.38	0.031	248.8	0.63	0.42	0.039	103.7	0.75	0.54	0.026	181.6	0.63	0.61	0.012
	StyleGAN-NADA	81.1	0.81	0.61	0.013	116.3	0.81	0.50	0.011	212.9	0.75	0.45	0.012	99.7	0.71	0.51	0.034	154.5	0.76	0.36	0.039
	StyleGANFusion	168.0	0.76	0.60	0.003	119.3	0.80	0.41	0.007	203.9	0.70	0.41	0.012	99.0	0.74	0.53	0.012	85.5	0.83	0.47	0.007
	DiffusionGAN3D	189.3	0.82	0.46	0.024	110.7	0.86	0.47	0.009	132.3	0.81	0.68	0.002	159.4	0.71	0.48	0.016	166.2	0.82	0.43	0.009
	Ours	77.6	0.86	0.69	0.014	67.7	0.89	0.56	0.003	99.7	0.85	0.56	0.002	91.6	0.77	0.75	0.005	144.5	0.82	0.55	0.006

Thank you!

Project page, preprint, and code