Deep Optimal Transport: A Practical Algorithm for Photo-realistic Image Restoration

Theo Adrai† Guy Ohayon† Michael Elad† Tomer Michaeli‡

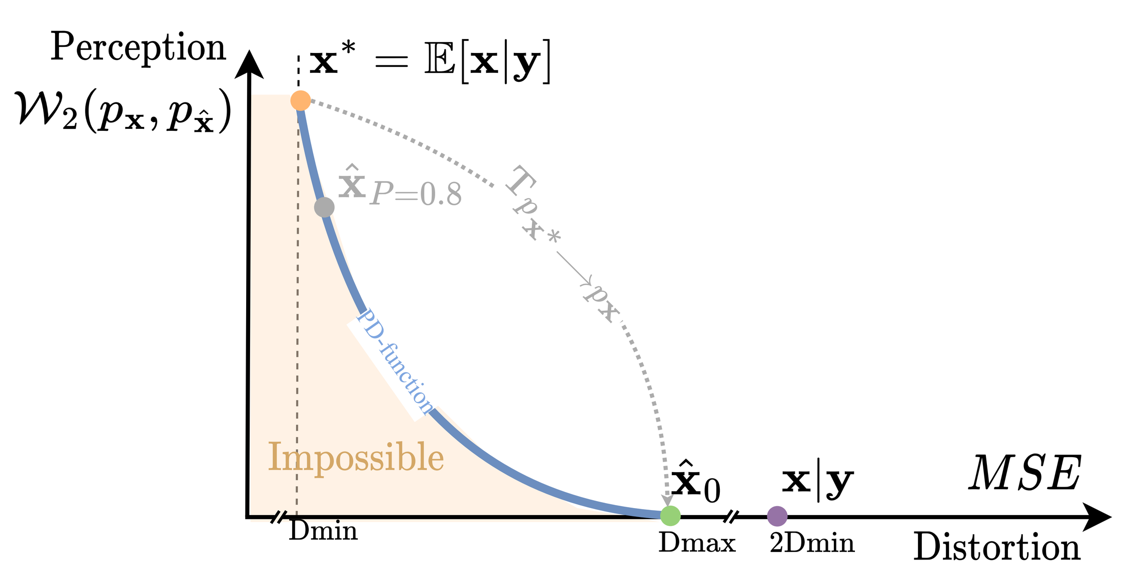
†Faculty of Computer Science ‡Faculty of Electrical & Computer Engineering

1. Contributions

- Improve the visual quality of any image restoration algorithm at inference time.
- Few-shot algorithm working with unpaired images and on any task.
- In some cases, we can improve MSE as well.
- No generative model required.

2. Wasserstein-2 MSE tradeoff

$\mathbf{x} \in \mathbb{R}^n$ represents a source natural image, $\mathbf{y} \in \mathbb{R}^m$ represents its degraded version. [1, 2]



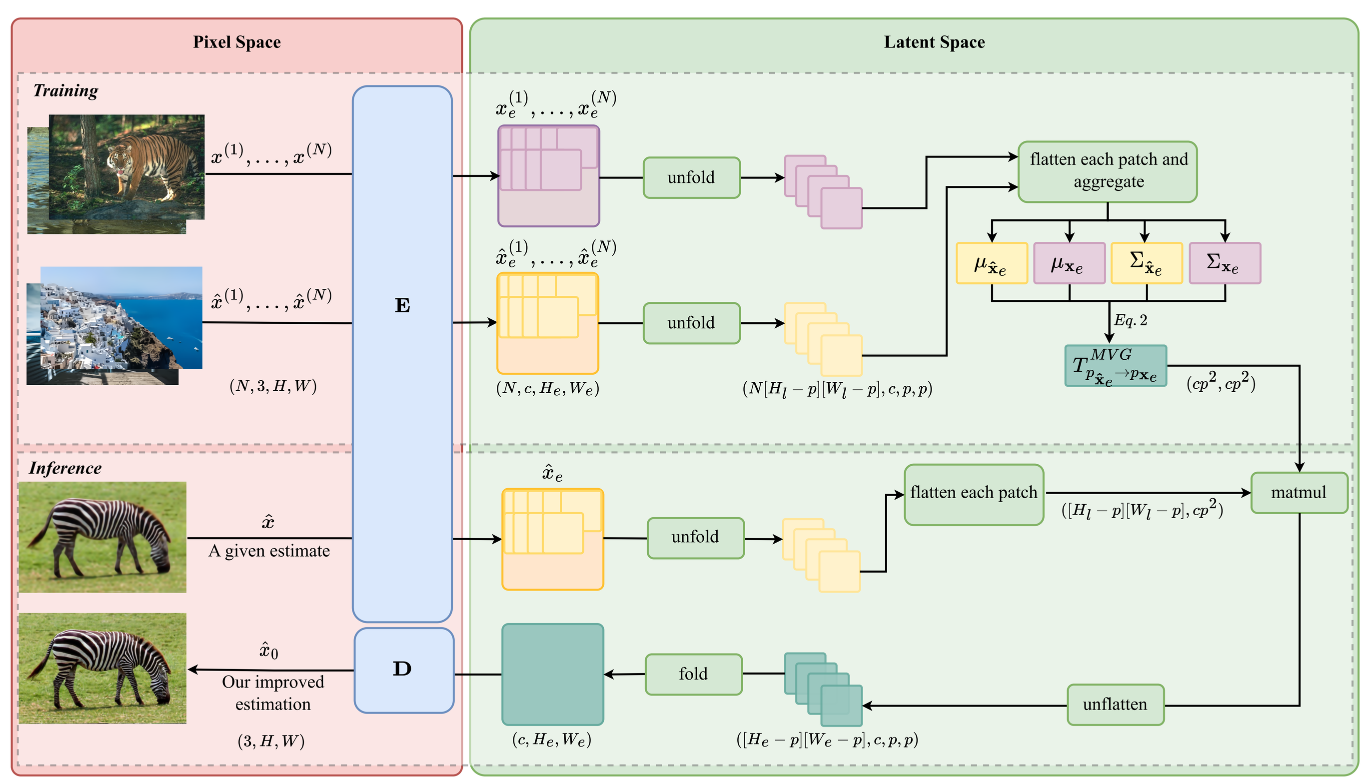
Theorem [2]: $$p_{\hat{\mathbf{x}}_0, \mathbf{x}^*} \in \underset{p_{\mathbf{x}_1,\mathbf{x}_2} \in \Pi(p_{\mathbf{x}}, p_{\mathbf{x}^*})}{\operatorname{argmin}} \mathbb{E}\left[\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2\right] \quad (1)$$

Theorem [2]: $$\hat{\mathbf{x}}_P = (1-\alpha)\hat{\mathbf{x}}_0 + \alpha \mathbf{x}^* \quad (2)$$

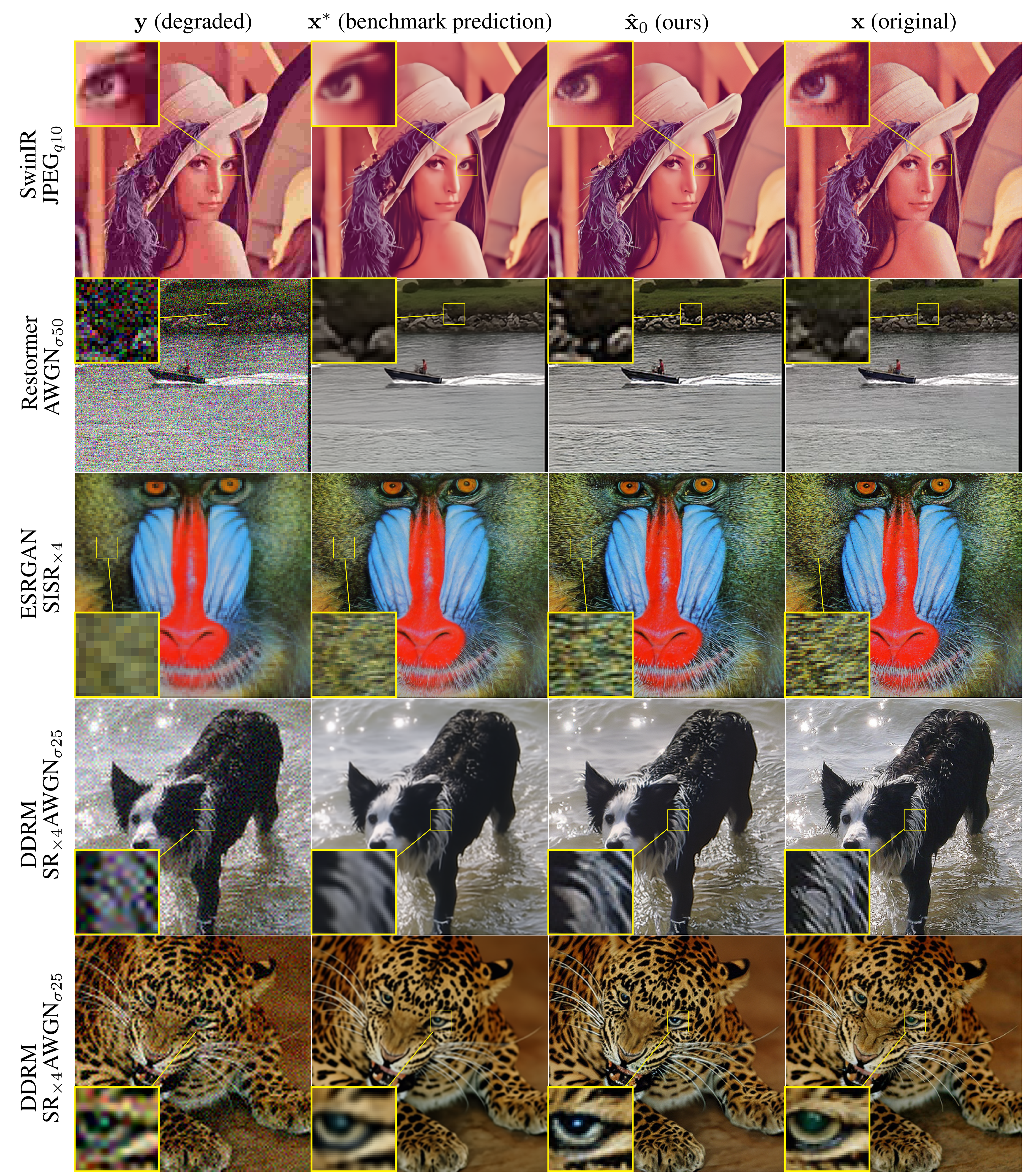
3. Method

- We perform optimal transport in the latent space of a VAE;
- We assume the distribution in the latent space is Gaussian, and thus use the formula for Gaussian transport:

$$T^{\text{MVG}}_{p_{\mathbf{x}_1} \longrightarrow p_{\mathbf{x}_2}}(x_1) = \Sigma_{\mathbf{x}_1}^{-\frac{1}{2}} \left(\Sigma_{\mathbf{x}_1}^{\frac{1}{2}} \Sigma_{\mathbf{x}_2} \Sigma_{\mathbf{x}_1}^{\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_{\mathbf{x}_1}^{-\frac{1}{2}} \cdot (x_1 - \mu_{\mathbf{x}_1}) + \mu_{\mathbf{x}_2}.$$

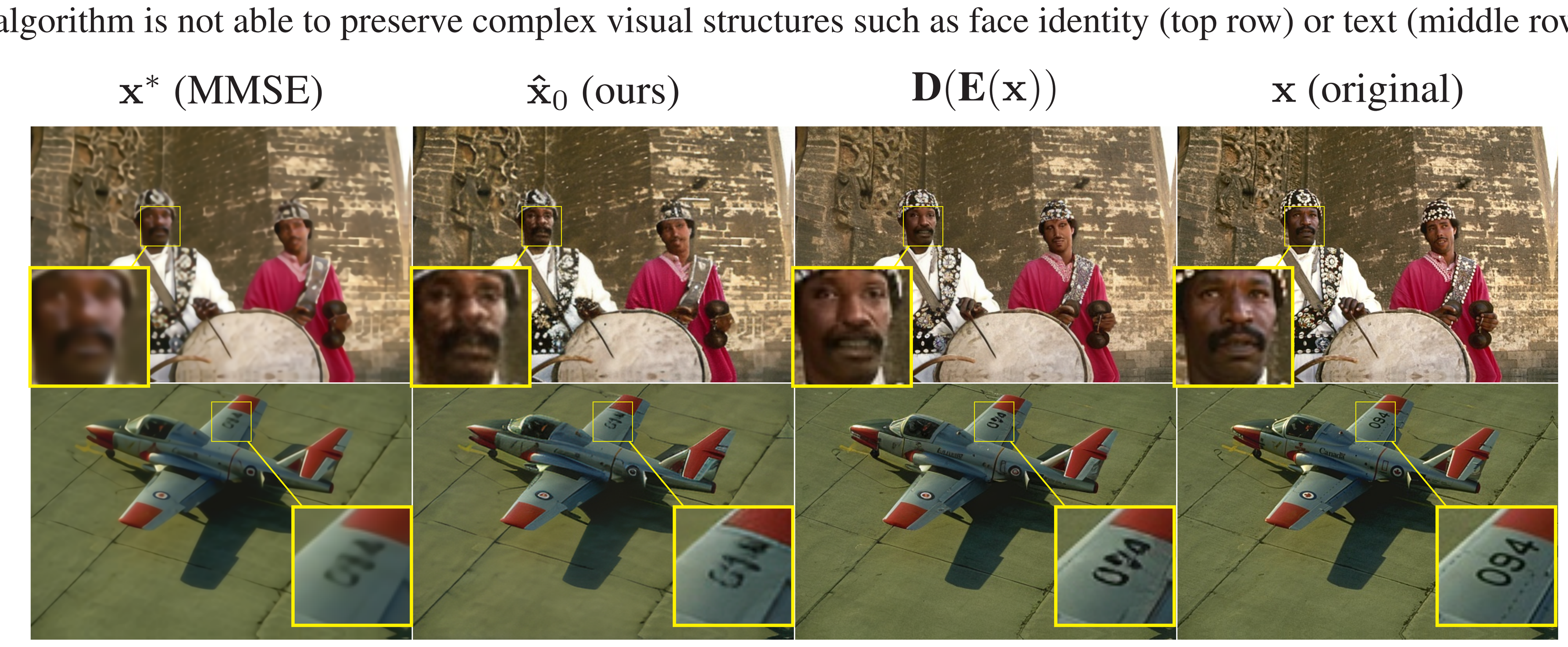


4. Visual examples



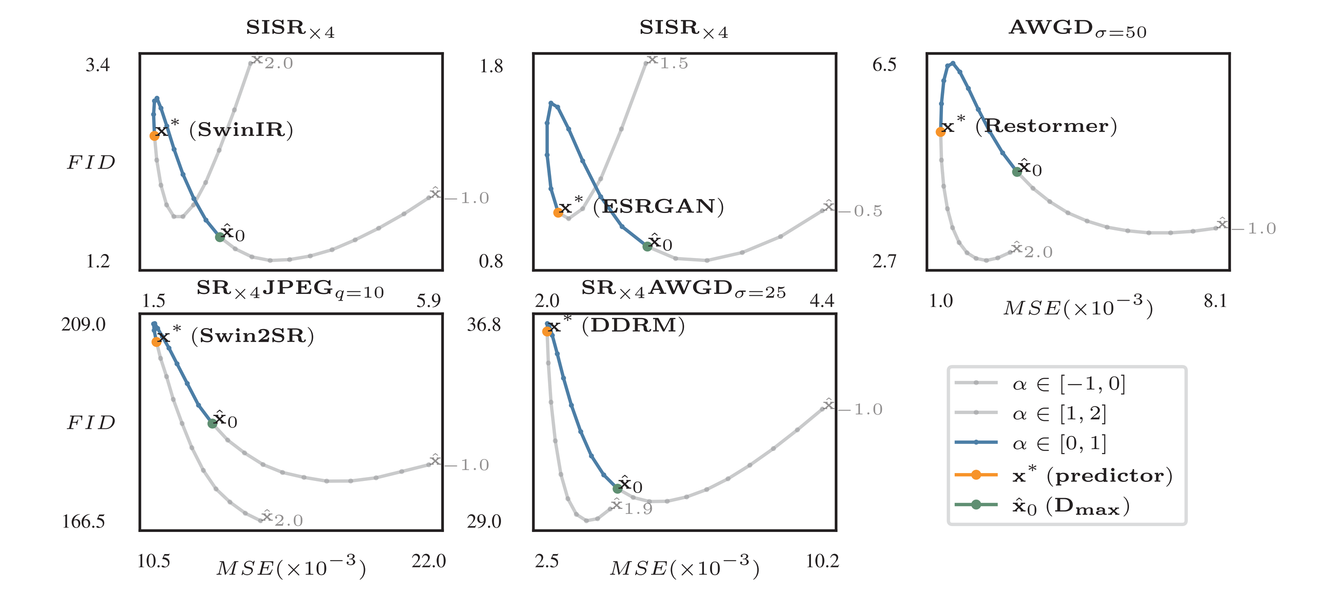
5. Limitations

- Our method's reconstruction capabilities are bounded by that of the VAE.
- Our algorithm is not able to preserve complex visual structures such as face identity (top row) or text (middle row).



6. Trading perception and distortion using out-of-the-box predictors

- We interpolate a given predictor (orange) and our improved $\mathbf{D_{max}}$ estimation (green).
- Using eq. (2) with $\alpha \in [0, 1]$ we approximate the PD FID-MSE function (blue curve).
- With $\alpha \in [-1, 0] \cup [1, 2]$ we extrapolate outside of the PD curve (light gray), beyond the theory-inspired area, to further improve performance.

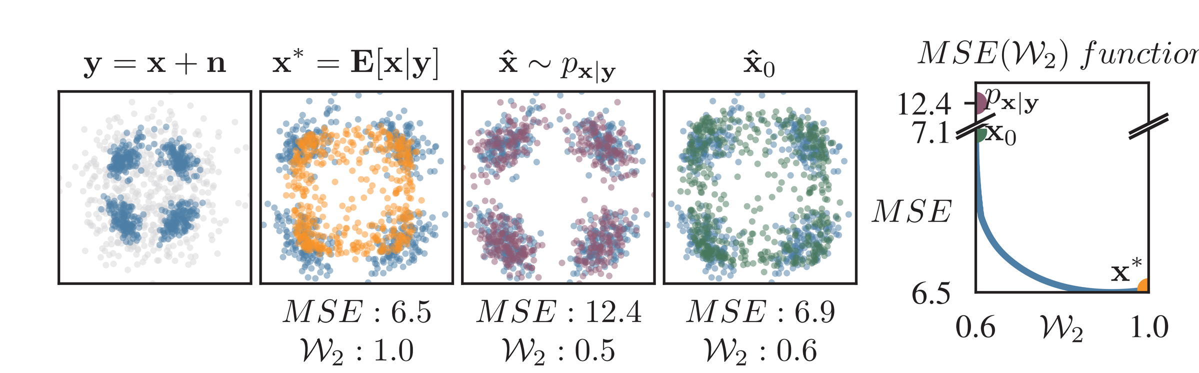


7. Quantitative analysis

Task	Signal	Distortion: PSNR ↑	Distortion: SSIM ↑	Distortion: LPIPS ↓	Perception: FID ↓	Perception: IS ↑	Perception: KID×10^3 ↓
	$\mathbf{x}$	∞	1	0	0	240.53±4.42	0
	$\mathbf{D}(\mathbf{E}(\mathbf{x}))$	27.10	0.81	0.13	0.24	234.71±4.04	0.02±0.07
SISR$_{\times 4}$	SwinIR [Liang, ICCV 2021]	28.10	**0.84**	**0.24**	2.54	201.52±4.85	1.24±0.24
	$\hat{\mathbf{x}}_{0.9}$	**28.15**	**0.84**	**0.24**	2.80	198.69±2.97	1.38±0.24
	$\hat{\mathbf{x}}_{-0.2}$	25.08	0.77	0.25	**1.19**	**216.74**±4.26	**0.38**±0.89
	$\hat{\mathbf{x}}_0$	25.48	0.78	0.23	1.39	214.63±5.50	0.69±0.23
JPEG$_{q=10}$	SwinIR [Liang, ICCV 2021]	**29.68**	**0.86**	**0.30**	8.95	161.73±3.36	6.52±0.77
	$\hat{\mathbf{x}}_{1.1}$	29.58	**0.86**	**0.30**	8.36	166.50±3.12	6.08±0.75
	$\hat{\mathbf{x}}_{-0.2}$	23.74	0.76	0.31	**7.56**	**166.65**±3.58	**5.68**±0.83
	$\hat{\mathbf{x}}_0$	24.84	0.78	0.30	**8.14**	**163.14**±3.93	**6.15**±0.77
AWGN$_{\sigma=50}$	Restormer [Zamir, CVPR 2022]	**30.18**	**0.86**	0.26	5.21	178.62±2.83	3.29±0.56
	$\hat{\mathbf{x}}_{1.1}$	30.09	**0.86**	**0.25**	4.63	183.36±3.20	2.61±1.53
	$\hat{\mathbf{x}}_{1.7}$	27.26	0.82	**0.25**	**2.73**	**198.93**±5.13	**1.76**±1.58
	$\hat{\mathbf{x}}_0$	25.31	0.78	0.27	4.42	**182.86**±2.21	2.93±1.62
SR$_{\times 4}$JPEG$_{q=10}$	Swin2SR [Conde, ECCV 2022]	19.75	**0.55**	0.53	205.00	5.95±0.49	40.68±3.34
	$\hat{\mathbf{x}}_{0.8}$	**19.81**	**0.55**	0.53	209.82	5.91±0.69	43.28±3.86
	$\hat{\mathbf{x}}_{1.9}$	18.44	0.49	**0.51**	**168.12**	6.36±0.69	**19.95**±2.84
	$\hat{\mathbf{x}}_0$	18.45	0.48	**0.51**	183.80	**6.55**±0.61	29.07±3.58
SISR$_{\times 4}$	ESRGAN [Wang, ECCV 2018]	26.77	0.80	**0.21**	1.06	221.68±3.06	0.43±0.14
	$\hat{\mathbf{x}}_{0.7}$	**27.00**	**0.81**	**0.21**	1.51	215.87±3.64	0.56±0.21
	$\hat{\mathbf{x}}_{-0.2}$	24.84	0.74	0.23	**0.80**	**221.89**±2.53	**0.30**±0.20
	$\hat{\mathbf{x}}_0$	25.33	0.74	0.22	0.89	220.96±3.19	0.34±0.18
SR$_{\times 4}$AWGN$_{\sigma=50}$	DDRM [Kawar, NIPS 2022]	**26.10**	**0.75**	0.34	36.44	43.52±3.33	5.09
	$\hat{\mathbf{x}}_{1.2}$	25.91	**0.75**	**0.33**	33.68	44.90±4.06	3.88
	$\hat{\mathbf{x}}_{1.7}$	24.48	0.70	0.35	**29.05**	**47.91**±2.69	**1.47**
	$\hat{\mathbf{x}}_0$	23.19	0.69	0.35	29.71	46.36±4.18	1.91

8. Numerical Experiement

- The MMSE (orange) has the best MSE but the worst perceptual index $\mathcal{W}_2$.
- The Posterior (purple) has the best perception but half of the optimal MSE.
- The $\mathbf{D_{max}}$ estimator ($\hat{\mathbf{x}}_0$, green) maintains the MSE of $\mathbf{x}^*$ while attaining a perceptual quality close to $\mathbf{x}|\mathbf{y}$.



References

[1] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[2] D. Freirich, T. Michaeli, and R. Meir, "A theory of the distortion-perception tradeoff in wasserstein space," in *Advances in Neural Information Processing Systems*, 2021.

The Henry and Marilyn Taub Faculty of Computer Science

TECHNION Israel Institute of Technology

The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING