Whitened Score Diffusion: A Structured Prior for Imaging Inverse Problems

Jeffrey Alido[1], Tongyu Li[1], Yu Sun[2], Lei Tian[1]

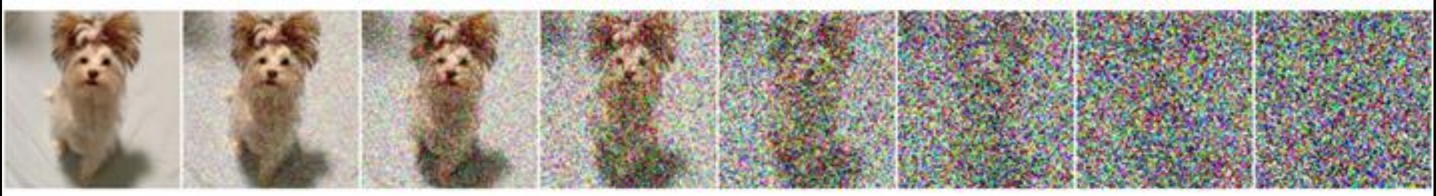
[1]Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA
[2]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21205, USA

WS Diffusion: Background

Diffusion models allow sampling from a distribution given its *score function*

$$\mathbf{dx} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}$$

$p(x_0)$



$p(x_T)$

score function

$$\mathbf{dx} = [\mathbf{f}(\mathbf{x}, t) - \mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)]\mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}$$

Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." arXiv preprint arXiv:2011.13456 (2020).

WS Diffusion: Motivating works

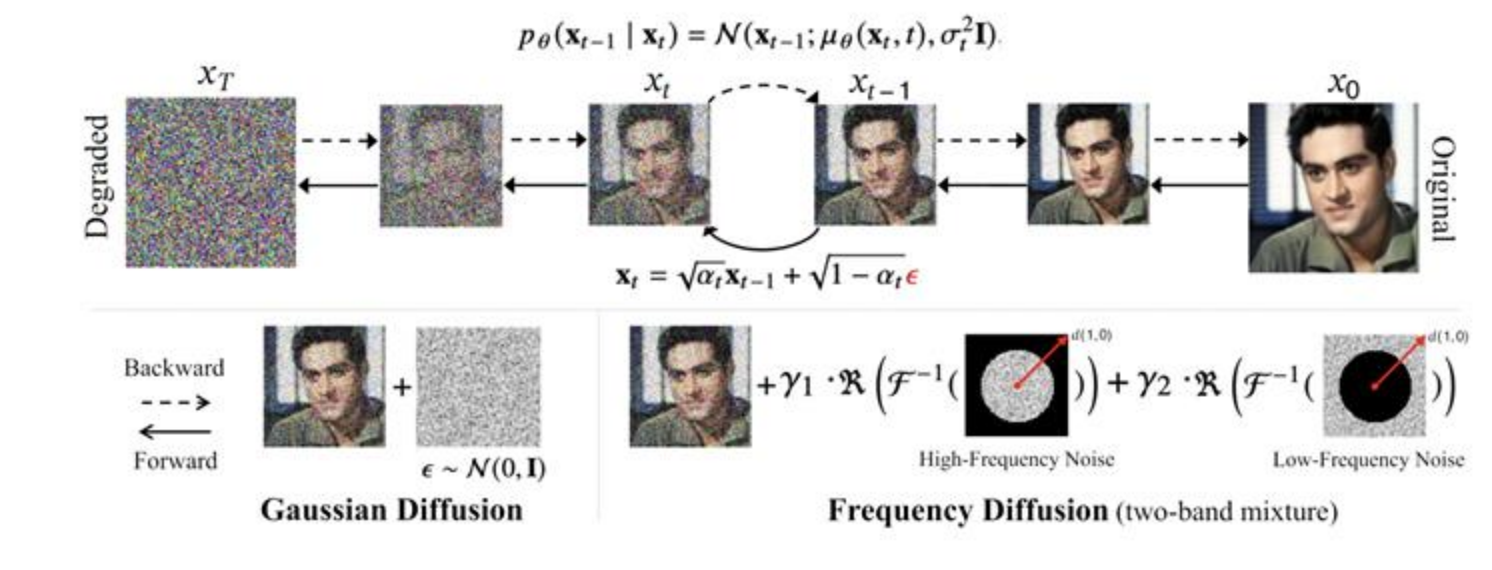


Figure 1: Frequency diffusion under a generalized framework.

Jiralerspong, Thomas, Berton Earnshaw, Jason Hartford, Yoshua Bengio, and Luca Scimeca. "Shaping Inductive Bias in Diffusion Models through Frequency-Based Noise Control." arXiv, March 12, 2025. https://doi.org/10.48550/arXiv.2502.10236.

WS Diffusion

Problem setup

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathrm{b}, \quad \mathrm{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_y)$$

measurement

known imaging system

added noise

$$p(\mathbf{y} \mid \mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x}, \mathbf{\Sigma}_y)$$

likelihood model that relates x to y

we want to to find an x from p(x | y)

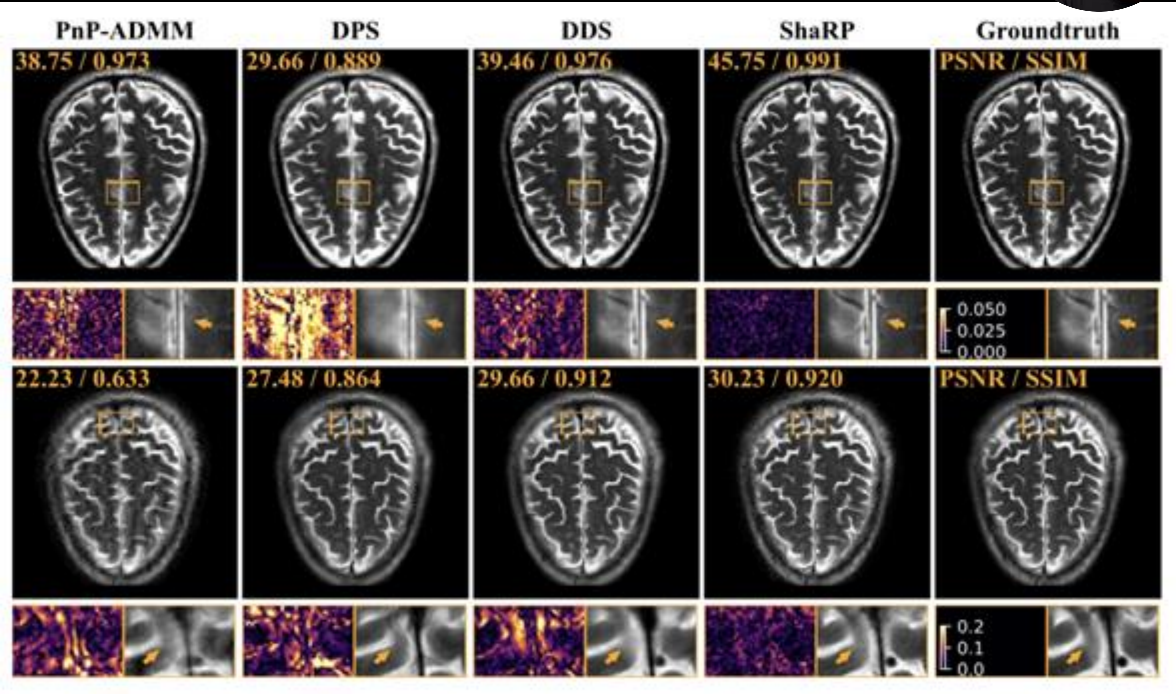
$$p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{x}) p(\mathbf{y} \mid \mathbf{x})$$

prior belief about x to constrain the problem

denoising diffusion model!

WS Diffusion: Motivating works

priors that restore/denoise the image beyond additive white Gaussian noise improves inverse problem solutions



Hu, Yuyang, Albert Peng, Weijie Gan, Peyman Milanfar, Mauricio Delbracio, and Ulugbek S. Kamilov. "Stochastic Deep Restoration Priors for Imaging Inverse Problems," October 4, 2024. https://openreview.net/forum?id=O2aioX2Z2v.

WS Diffusion

change G to a non-diagonal matrix to introduce structure in the noise!

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}$$

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - \mathbf{G}_t \mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)]\mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}$$

learned reverse diffusion process will have spectral inductive bias, and can denoise a wider range of noise beyond AWGN

WS Diffusion: sampling from $p(x_t \mid y)$

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - \mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y})]\mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}$$

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y}) = \underbrace{\overbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)}^{\mathbf{s}_\theta(\mathbf{x}_t, t)}}_{\text{diffusion model}} + \underbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t)}_{\text{known likelihood model}}$$

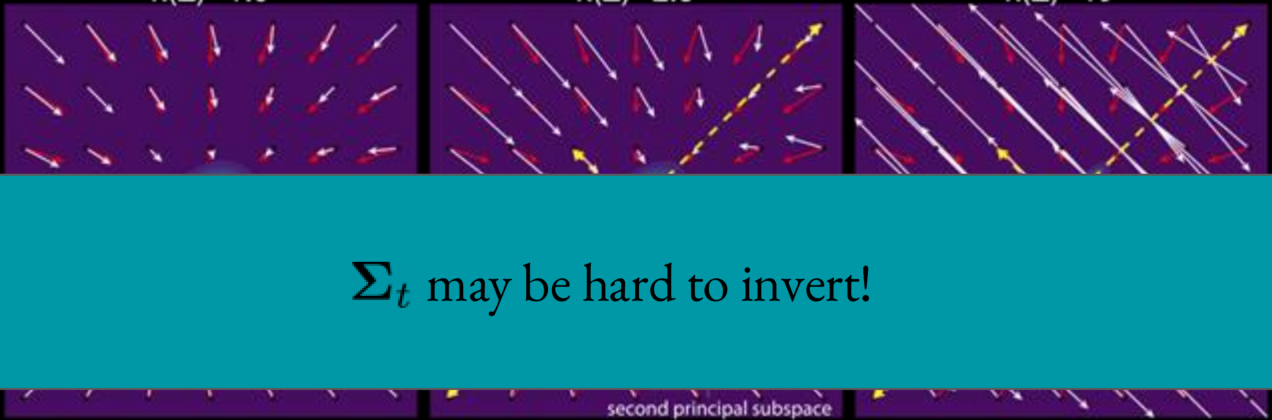
how do we learn this??

WS Diffusion: Denoising score matching

$$\hat{\theta} = \arg\min_{\theta} \mathbb{E}_{t\sim\mathcal{U}(0,1),\mathbf{x}_t\sim p(\mathbf{x}_t|\mathbf{x}_0),\mathbf{x}_0\sim p(\mathbf{x})} \left\{ \|\mathbf{s}_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0)\|_2^2 \right\}$$

known, defined by your custom diffusion process, f, and G

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x_t} \mid \mathbf{x}_0) = \mathbf{\Sigma}_t^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_t) \qquad \mathbf{\Sigma}_t \propto \mathbf{G}_t\mathbf{G}_t^\top$$



$\mathbf{\Sigma}_t$ may be hard to invert!

Vincent, Pascal. "A Connection Between Score Matching and Denoising Autoencoders." *Neural Computation* 23, no. 7 (July 2011): 1661–74. https://doi.org/10.1162/NECO_a_00142.

WS Diffusion: Denoising score matching

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}$$

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - \mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)]\mathrm{d}t + \mathbf{G}_t \mathrm{d}\mathbf{w}$$

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x_t} \mid \mathbf{x}_0) = \boldsymbol{\Sigma}_t^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_t) \qquad \boldsymbol{\Sigma}_t \propto \mathbf{G}_t\mathbf{G}_t^\top$$

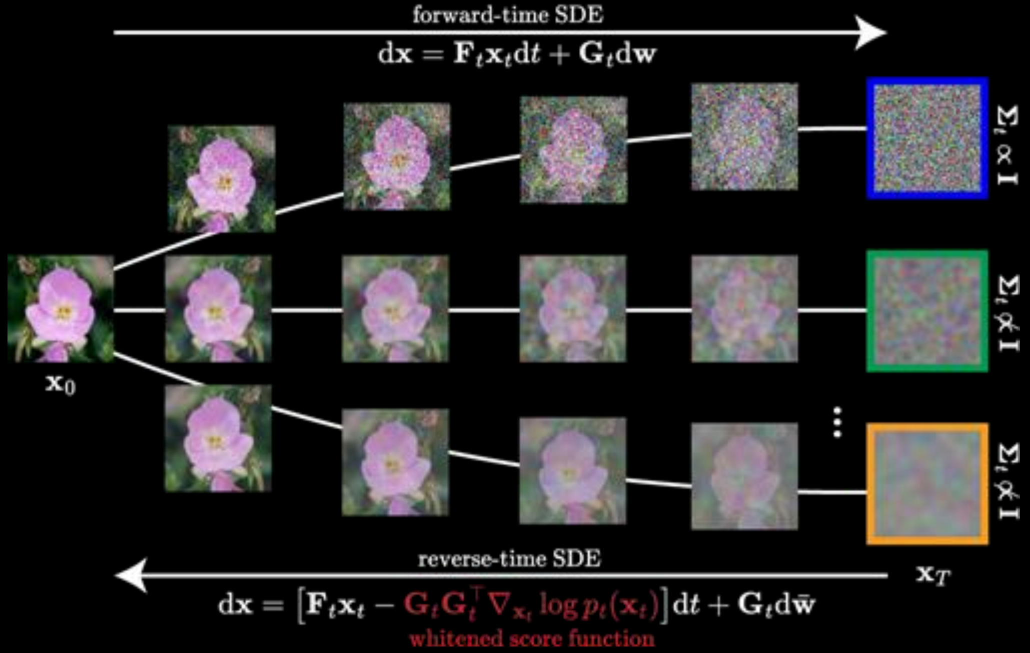
$$\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{x}_0) = \overset{\mathbf{I}}{\cancel{\mathbf{G}_t\mathbf{G}_t^\top \boldsymbol{\Sigma}_t^{-1}}}(\mathbf{x}_t - \boldsymbol{\mu}_t)$$

$$\propto (\mathbf{x}_t - \boldsymbol{\mu}_t)$$

no inversion!

WS Diffusion: training the model

Loss function: $$\mathbb{E}_{t\sim U(0,1),\mathbf{x}_t\sim p(\mathbf{x}_t|\mathbf{x}_0),\mathbf{x}_0\sim p(\mathbf{x})}\left\{\left\|\mathbf{n}_\theta(\mathbf{x}_t,t)-\mathbf{G}_t\mathbf{G}_t^\top\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t\mid\mathbf{x}_0)\right\|_2^2\right\}$$



WS Diffusion: sampling from $p(x_t \mid y)$

$$d\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - \mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y})]dt + \mathbf{G}_t d\mathbf{w}$$

conventional:

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y}) = \underbrace{\overbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)}^{\mathbf{s}_\theta(\mathbf{x}_t, t)}}_{\text{diffusion model}} + \underbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t)}_{\text{known likelihood model}}$$

G is a hyperparameter manually set

Whitened score:

$$\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y}) = \underbrace{\overbrace{\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)}^{\mathbf{n}_\theta(\mathbf{x}_t, t)}}_{\text{WS diffusion model}} + \underbrace{\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t)}_{\text{preconditioned likelihood model}}$$

G is absorbed in the WS diffusion model, don't need to specify

WS Diffusion: sampling from $p(x_t \mid y)$

$$\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t)$$

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) \approx \boldsymbol{\Sigma}_y^{-1} \nabla_{\mathbf{x}_t} \mathbf{r}(\mathbf{x}_t)$$

potentially also difficult to invert with structured noise

noising process in the diffusion model is designed to match the noise in inverse problems!

$$\boldsymbol{\Sigma}_\mathbf{y} \propto \mathbf{G}_t\mathbf{G}_t^\top$$

$$\mathbf{I}$$

$$\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t)$$

$$\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t) = \mathbf{G}_t\mathbf{G}_t^\top \boldsymbol{\Sigma}_\mathbf{y}^{-1} \nabla_{\mathbf{x}_t} \mathbf{r}(\mathbf{x}_t)$$

$$\propto \nabla_{\mathbf{x}_t} \mathbf{r}(\mathbf{x}_t)$$

WS Diffusion: sampling from p(x_t | y)

$\nabla_{\mathbf{x}_t} \mathbf{r}(\mathbf{x}_t)$ approximation by Jalal *et al.*

$$
\begin{aligned}
\mathbf{G}_t\mathbf{G}_t^\top \nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) &\approx \mathbf{G}_t\mathbf{G}_t^\top \overbrace{\nabla_{\mathbf{x}_t} \left\{-\tfrac{1}{2}(\mathbf{y} - \mathbf{A}\mathbf{x}_t)^\top \boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}_t)\right\}} \\
&= \mathbf{G}_t\mathbf{G}_t^\top \boldsymbol{\Sigma}_y^{-1} \left\{\mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_t)\right\} \\
&\propto \mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_t) \qquad \boldsymbol{\Sigma}_{\mathbf{y}} \propto \mathbf{G}_t\mathbf{G}_t^\top
\end{aligned}
$$

Jalal, Ajil, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G. Dimakis, and Jonathan I. Tamir. "Robust Compressed Sensing MRI with Deep Generative Priors." arXiv, December 6, 2021. https://doi.org/10.48550/arXiv.2108.01368.

WS Diffusion: Final algorithm

$$\mathrm{d}\mathbf{x}_t = \left[\mathbf{F}_t\mathbf{x}_t - \frac{1}{2}\mathbf{G}_t\mathbf{G}_t^\top\left(\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t}\log p_t(\mathbf{y}\mid\mathbf{x}_t)\right)\right]\mathrm{d}t$$

WS DM

likelihood approx

discretize to sample x from p(x | y)

Algorithm 1 WS diffusion priors for imaging inverse problems

Require: T, $\mathbf{A}$, $\mathbf{y}$, $\{\beta_t\}_{t=0}^T$, $\mathbf{n}_\theta$

1: Initialize $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

2: **for** $t = T$ to 0 **do**

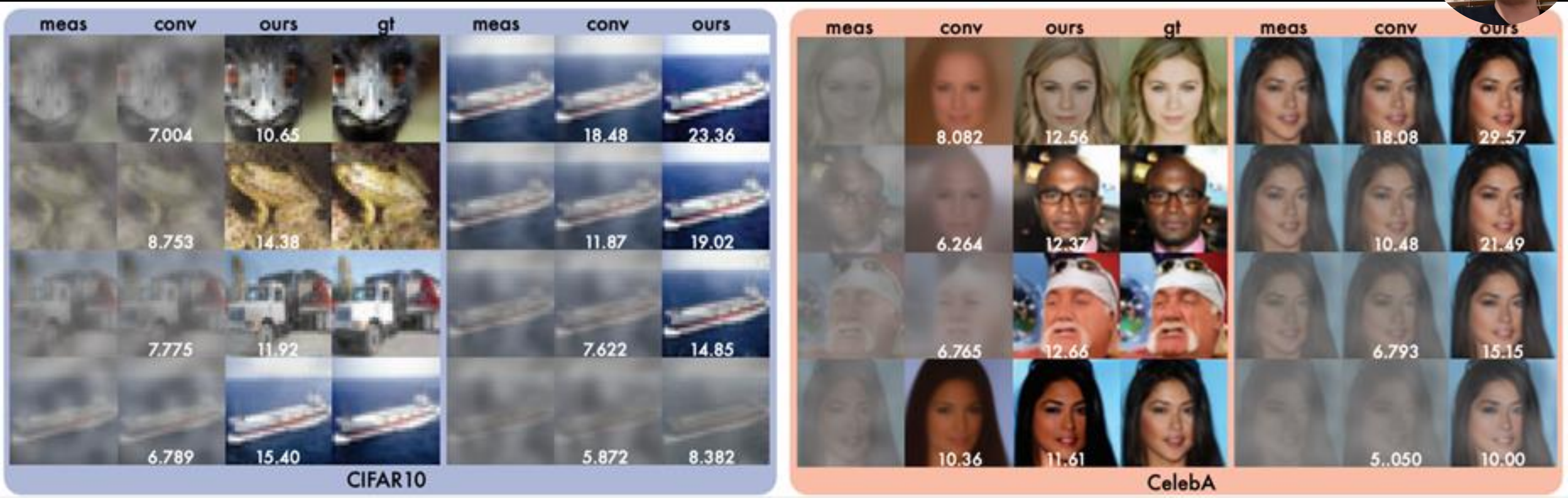
3: $\quad \mathbf{x}'_t \leftarrow (2 - \sqrt{1 - \beta_t\Delta t})\mathbf{x}_t + \frac{\mathbf{n}_\theta(\mathbf{x}_t, t)\Delta t}{2}$

4: $\quad \mathbf{x}_{t-1} \leftarrow \mathbf{x}'_t - \lambda_t \frac{\beta_t\mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}_t)}{2}$

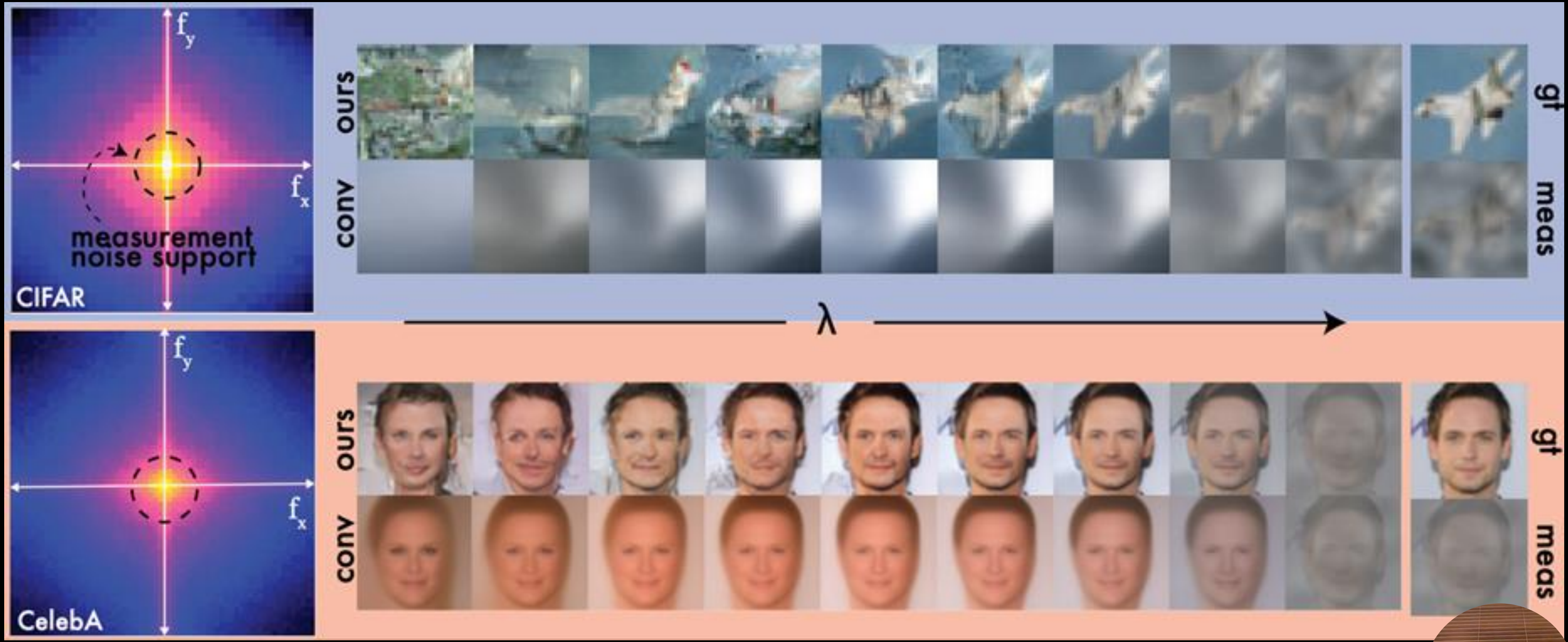
5: **end for**

6: **return** $\mathbf{x}_0$

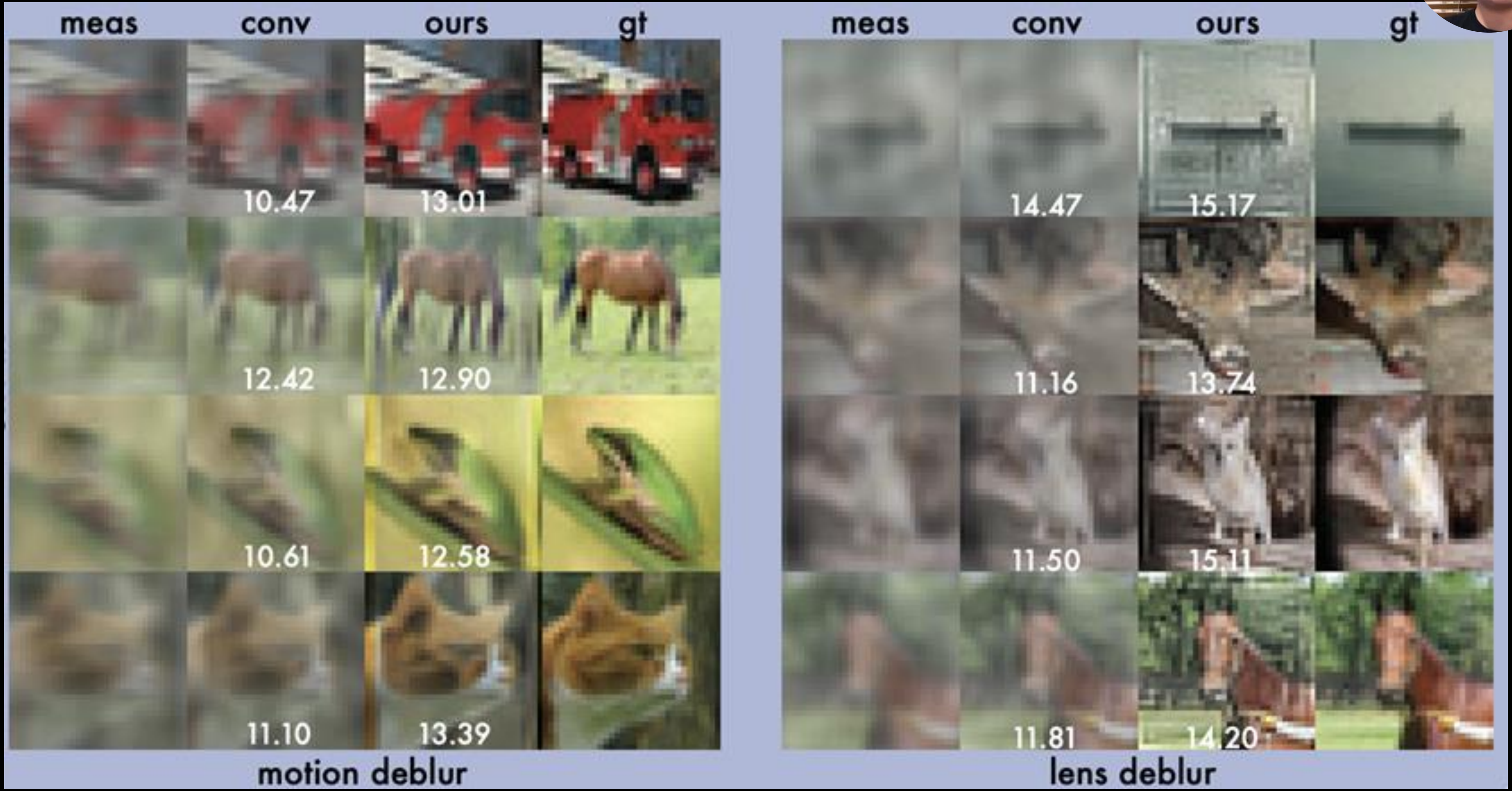
WS Diffusion: results on dehazing



WS Diffusion: noise Fourier support overlaps with image's support

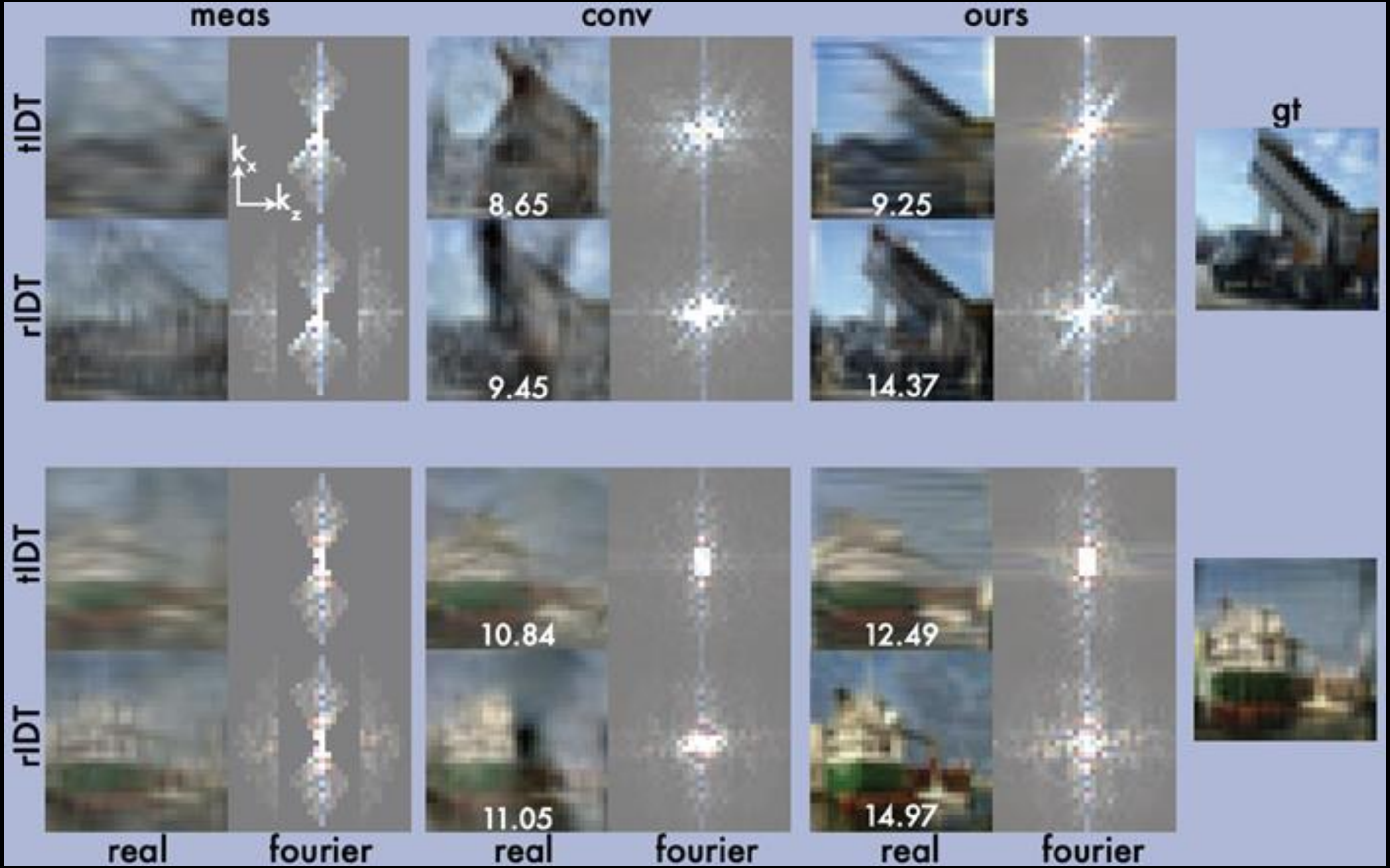


WS Diffusion: results on deblurring + dehazing

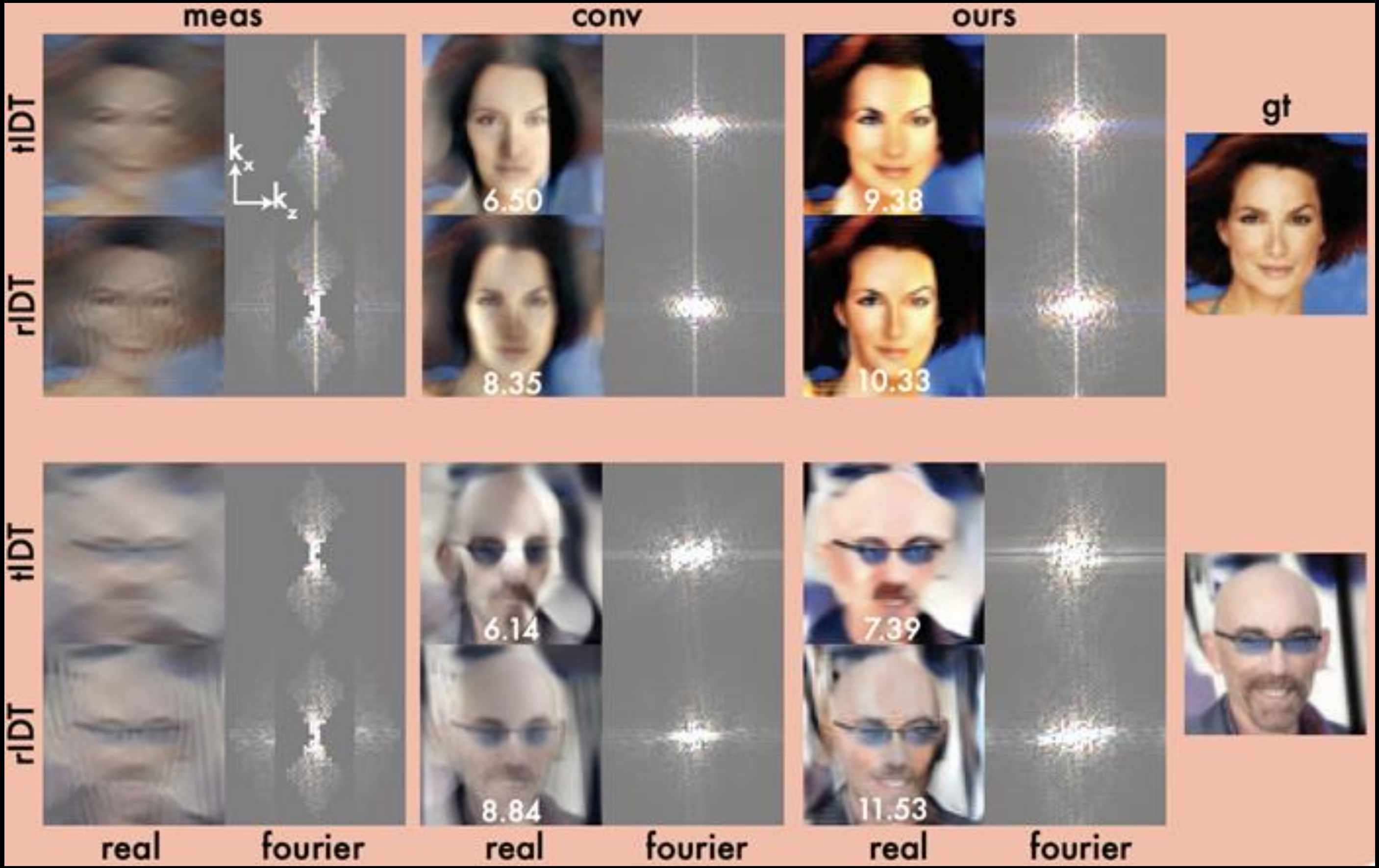


WS Diffusion: results on deblurring + dehazing

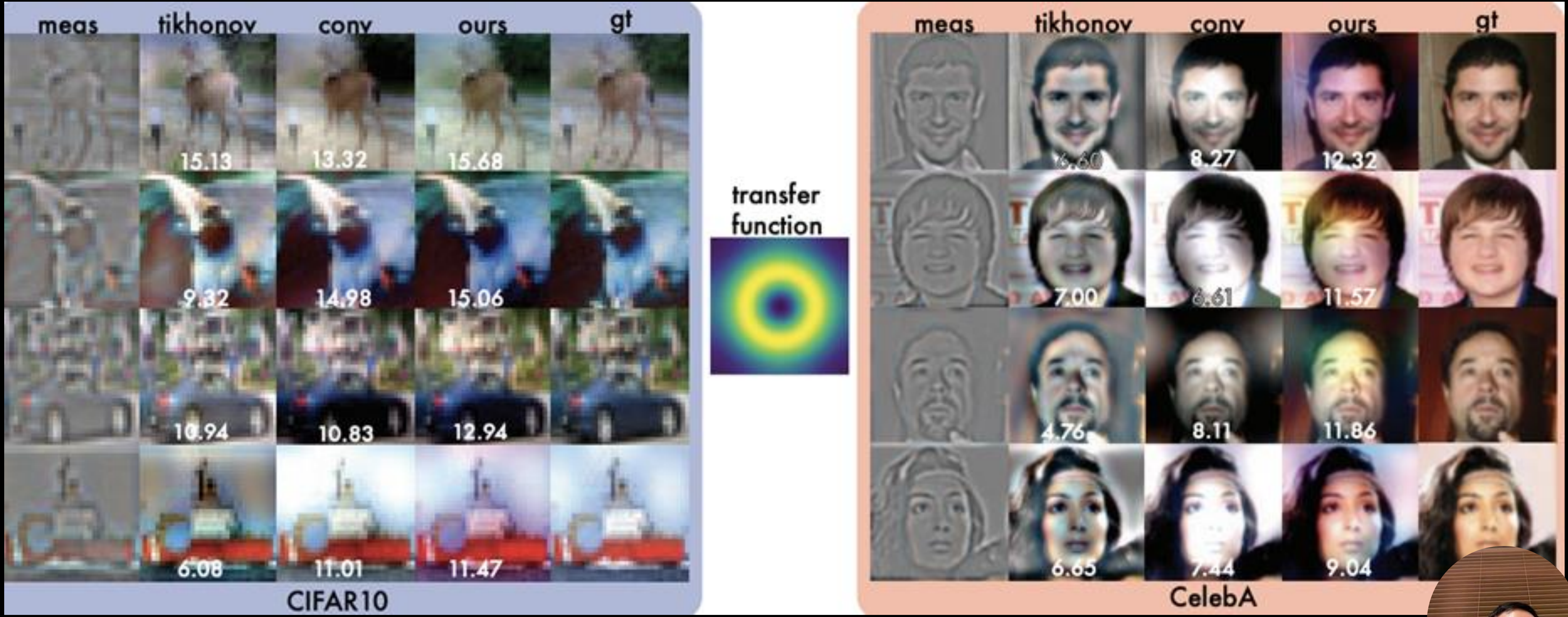
WS Diffusion: results on intensity diffraction tomography (IDT)



WS Diffusion: results on intensity diffraction tomography (IDT)



WS Diffusion: results on Laplace imaging (transport of intensity, TIE)



WS Diffusion: Discussion

- Denoising WS diffusion model prior removes structured noise better than conventional score-based diffusion models
- Consistently better peak signal-to-noise ratio (PSNR) compared to conventional diffusion model priors
- Very sensitive to regularization parameter λ
- Time-complexity is too slow