Gabriele Steidl
Jannis Chemseddine
TU Berlin

Lecture 1: Optimal Transport
Lecture 2: Generative Flows
Lecture 3: Bayesian Inverse Problems
Lecture 4: Experiments

Berlin Mathematics Research Center MATH+
Funded under Germany's Excellence Strategy by Deutsche Forschungsgemeinschaft

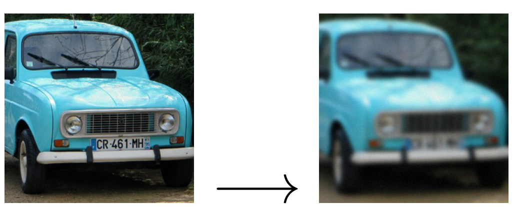
Inverse Problem for a certain class of images:

$$y = \text{noisy}(F(x))$$

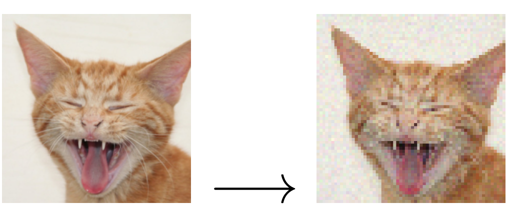
with ill-posed forward operator $F : \mathbb{R}^d \to \mathbb{R}^n$

Typical examples with linear operator $F(x) = Ax$:

- Computerized Tomography (CT)
- Magnetic Resonance Imaging (MRI)
- image restoration (denoising, deblurring, inpainting, superresolution)



Deblurring



Superresolution

Inpainting

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

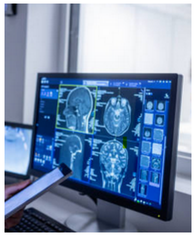
3. Bayesian Inverse Problems

Problem: Well-posedness/conditioned of problem:

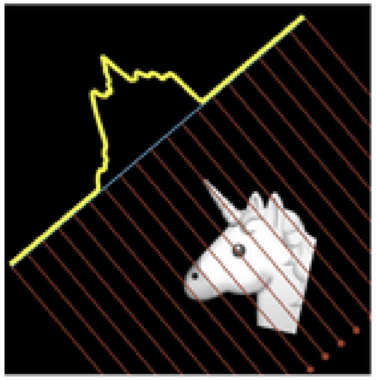
- has solution
- solution is unique
- solution depends continuously/stable on input y

Example: Computer Tomography

$\Longrightarrow$



- Computation of sinograms of Einhorn



History:

- 1917 J. Radon: from sinogram to image (Radon transform)
- 1979 G. Houndsfield and A. Cormack: Nobel Prize in (Computertomographie)

Video: Samuli Siltanen

Variational Problem: find minimizer of

$$\mathcal{J}(x) = \underbrace{\mathcal{D}(F(x), y)}_{\text{data term}} + \lambda \underbrace{\mathcal{R}(x)}_{\text{regularizer, prior}}, \qquad \lambda > 0$$

- data term $\mathcal{D}$ depends usually on the noise model via $-\log p_{Y|X=x}$
- prior includes knowledge on the distribution of p_X

Additive Gaussian noise:

$$y = Ax + \eta, \quad \eta \sim \mathcal{N}(0, \sigma^2 I_d)$$

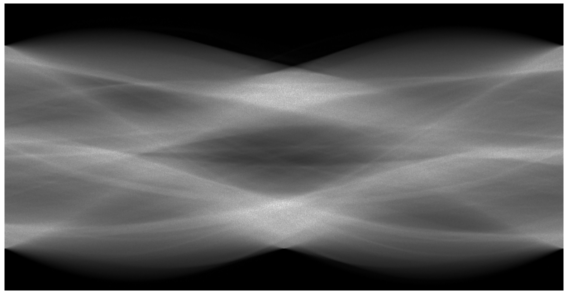
$$\mathcal{D}(F(x), y)) = \frac{1}{2}\|Ax - y\|^2$$

Bayesian Inverse Problems: sample from the posterior $P_{X|Y=y}$, where

$$Y = F(X) + \Xi$$

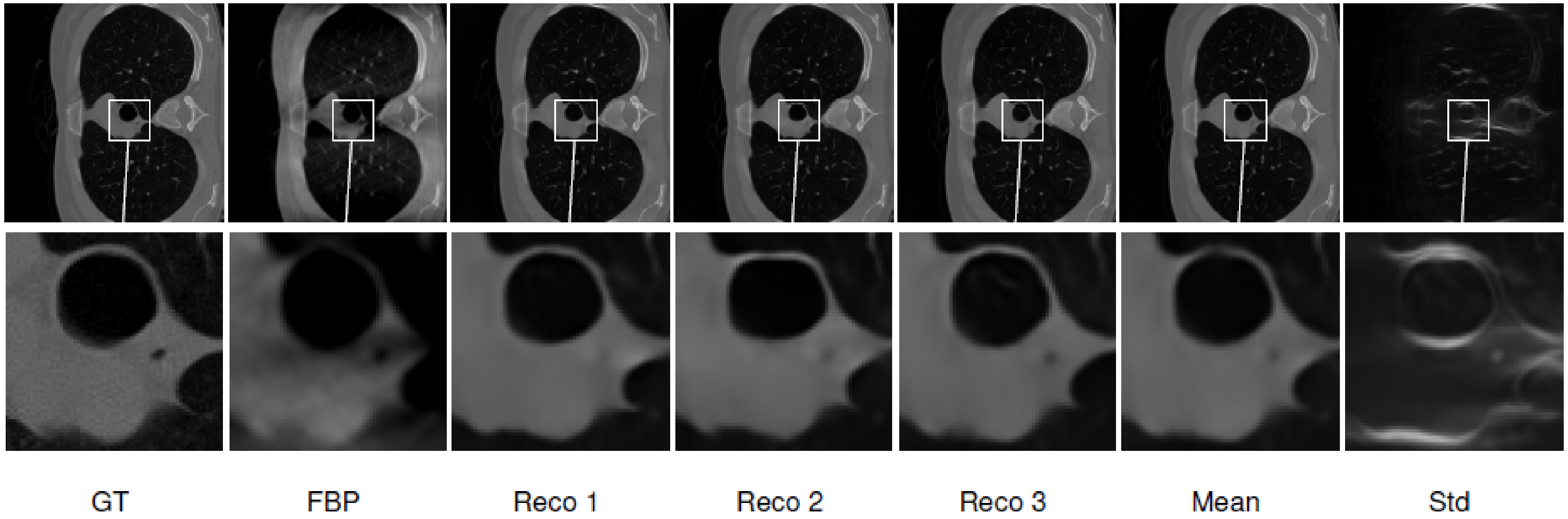
Bayesian law: $p_{X|Y=y}(x) = p_{Y|X=x}p_X(x)/p_Y(y)$

„Limited Angle" CT



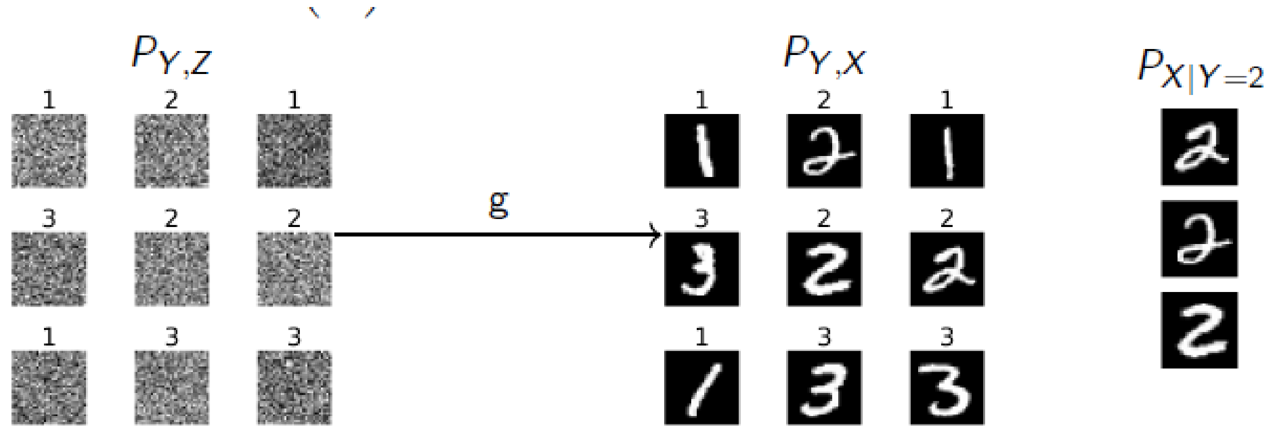
Sinogramm

Bedingte Bildgenerierung: Unsicherheiten in Bildgenerierung werden sichtbar



1. Bayesian OT Flow Matching
2. PnP Flow

Ref: J. Chemseddine, P. Hagemann, G. Steidl, C. Wald,
Conditional Wasserstein distances with applications in Bayesian OT flow matching,
Journal of Machine Learning Research (JMLR) 26 (141), 1-47



We learn $g^\theta : [0,\dots,9] \times \mathbb{R}^d \to [0,\dots,9] \times \mathbb{R}^d$ such that

$$g^\theta_\sharp P_{Y,Z} = P_{Y,X}$$

but we want to sample from the posterior $P_{X|Y=y}$.

"Chain rule" of Kullback-Leibler Divergenz (KL):

$$\mathrm{KL}(P_{Y,X}, P_{Y,Z}) = \mathbb{E}_{y\sim P_Y}\left[\mathrm{KL}(P_{X|Y=y}, P_{Z|Y=y})\right]$$

does not hold true for Wasserstein distances !

Conditional Wasserstein distance

$$W_{p,Y}(P_{Y,X}, P_{Y,Z}) := \Big(\inf_{\alpha \in \Gamma_Y^4} \int_{(A\times B)^2} \|(y_1, x_1) - (y_2, x_2)\|^p \,\mathrm{d}\alpha \Big)^{\frac{1}{p}}$$

with Y-diagonal couplings $\Gamma_Y^4 := \Big\{ \alpha \in \Gamma(P_{Y,X}, P_{Y,Z}) : \pi_\sharp^{1,3}\alpha = \Delta_\sharp P_Y \Big\}$ on the conditional Wasserstein space

$$\mathcal{P}_{p,Y}(A \times B) = \{\mu \in \mathcal{P}_p(A \times B) : \pi_\sharp^A \mu = P_Y\}$$

Then we have the "chain rule":

$$W_{p,Y}^p(P_{Y,X}, P_{Y,Z}) = \mathbb{E}_{y\sim P_Y}\Big[W_p^p(P_{X|Y=y}, P_{Z|Y=y})\Big]$$

Curves in conditional Wasserstein space $\mathcal{P}_{p,Y}$:

- $e_t(y_1, x_1, y_2, x_2) := (1-t)(y_1, x_1) + t(y_2, x_2)$
- $\mu_t = (e_t)_\sharp \alpha$, $\alpha \in \Gamma_Y^4(\mu_0, \mu_1)$ is a curve in $W_{2,Y}$
- $L(\theta) = \mathbb{E}_{t\sim[0,1],(y_1,x_1,y_2,x_2)\sim\alpha}\left[\|u(t, e_t(y_1, x_1, y_2, x_2), \theta) - (0, x_2 - x_1)\|^2\right]$

Refs: PhD Gigli 2008, Kloeckner 2021, Peszek/Poyato 2023, Hosseini et al. 2024, Barboni et al. 2024, Kerrigan et al. 2024

Drawback of $W_{p,Y}$: not implemented in standard OT libaries

Y- penalized Wasserstein distance

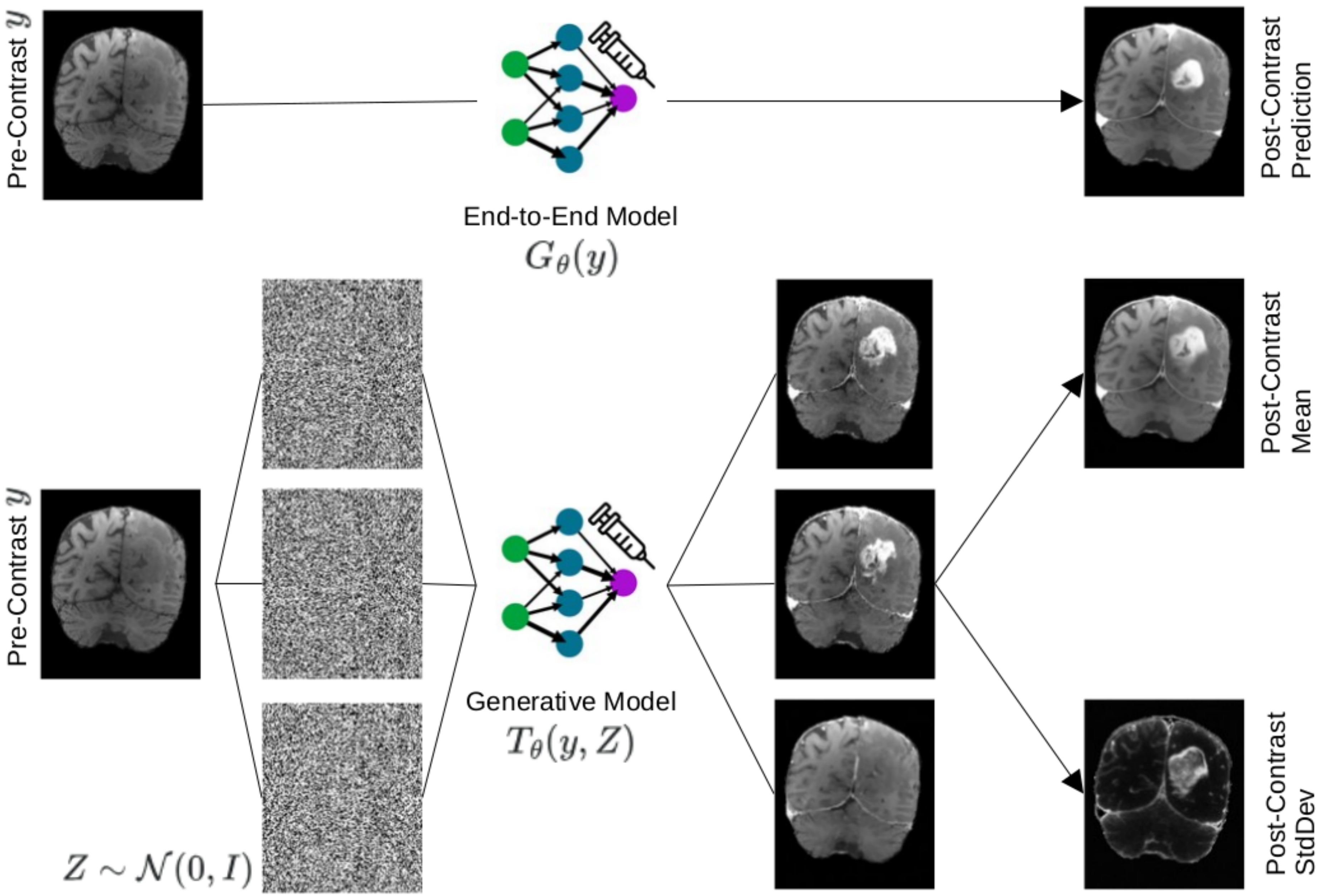
$$W_{p,\beta}(\mu_0,\mu_1)^p := \inf_{\alpha\in\Gamma(\mu_0,\mu_1)} \int_{(A\times B)^2} d_\beta^p((y_1,x_1),(y_2,x_2))\,\mathrm{d}\alpha$$

$$d_\beta^p((y_1,x_1),(y_2,x_2)) := \|x_1-x_2\|^p + \beta\|y_1-y_2\|^p$$

Bayesian OT Flow Matching: Conditional Wasserstein Distances

CIFAR10, $\beta = 1, 3, 20$ and diagonal, $d = 3072$

Bayesian OT Flow Matching: Conditional Wasserstein Distances



Refs: M. Piening, F. Altekrüger, St., E. Hattingen, E. Steidl, Conditional Generative Models for Contrast-Enhanced Synthesis of T1W and T1 Maps in Brain MRI IEEE International Symposium on Biomedical Imaging (ISBI), 2025

1. Bayesian OT Flow Matching
2. PnP Flow

Refs: S. Martin, A. Gagneux, P. Hagemann, G. Steidl,
PnP-Flow: Plug-and-play image restoration with flow matching,
ICLR 2025

2. PnP Forward-Backward Splitting

Variational model: find minimizer of

$$\mathcal{J}(x) = \underbrace{\mathcal{D}(F(x), y)}_{g(x)} + \underbrace{\lambda \mathcal{R}(x)}_{h(x)}, \qquad \lambda > 0$$

- h convex, lsc, $g : \mathbb{R}^d \to \mathbb{R}$ convex, diff. and $\mathrm{Lip} \nabla g = L < \infty$

Fermat's rule:

$$\begin{aligned} 0 &\in \nabla g(\hat{x}) + \partial h(\hat{x}) \\ \hat{x} - \tau \nabla g(\hat{x}) &\in \hat{x} + \tau \partial h(\hat{x}) = (I + \tau \partial h)(\hat{x}) \\ \hat{x} &= (I + \tau \partial h)^{-1} (\hat{x} - \tau \nabla g(\hat{x})) \\ &= \mathrm{prox}_{\tau h} (\hat{x} - \tau \nabla g(\hat{x})) \end{aligned}$$

Forward-Backward Splitting (FBS): (also called proximal gradient algorithm)

$$x^{(r+1)} = \mathrm{prox}_{\tau h} \left(x^{(r)} - \tau \nabla g(x^{(r)}) \right)$$

Convergence for $\tau \in (0, \frac{2}{L})$:

$$\mathcal{J}\left(x^{(r)}\right) - \mathcal{J}(\hat{x}) = \mathcal{O}\left(1/r\right).$$

2. PnP Forward-Backward Splitting

From FBS to Plug-and-Play FBS:

Algorithm 1 FBS

Init: $x_0 \in \mathbb{R}^d$, $\gamma > 0$

Iterations: For $n = 0, 1, \ldots$

$z_n = x_n - \gamma \nabla (\mathcal{D} \circ F)(x_n)$

$x_{n+1} = \mathrm{prox}_{\gamma \mathcal{R}}(z_{n+1})$

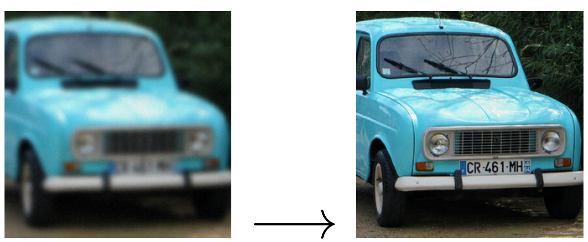
Algorithm 2 PnP-FBS

Init: $x_0 \in \mathbb{R}^d$, $\gamma > 0$

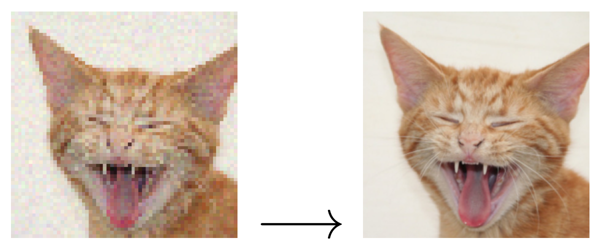
Iterations: For $n = 0, 1, \ldots$

$z_n = x_n - \gamma \nabla (\mathcal{D} \circ F)(x_n)$

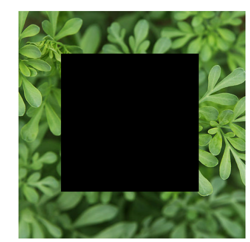
$x_{n+1} = D(z_n)$



Deblurring



Superresolution



Inpainting

Algorithm 3 PnP-Flow Matching

Input: Pre-trained network v^θ, $T_t^\theta = \mathrm{Id} + (1-t)v_t^\theta$, $(t_n)_n$ with $t_n = n/N$ and adaptive stepsizes $(\gamma_n)_{1\le n\le N}$.

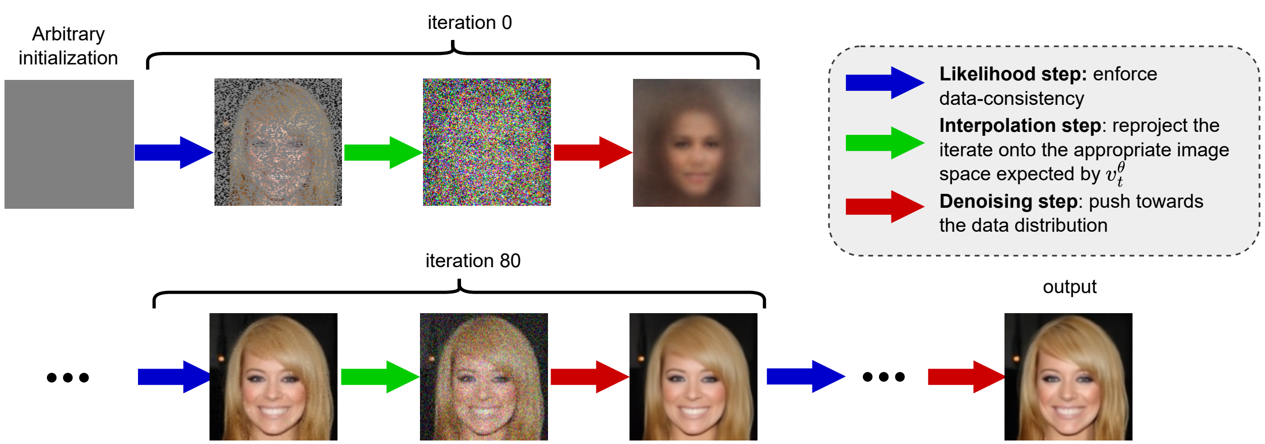
Initialize: $x_0 \in \mathbb{R}^d$.

Iteration For $n = 0, 1, \dots, N$

$z_n = x_n - \gamma_n \nabla(\mathcal{D} \circ F)(x_n)$ Gradient step

$\tilde{z}_n = (1-t_n)\varepsilon + t_n z_n$, $\varepsilon \sim \mu_0$ Interpolation step

$x_{n+1} = T_{t_n}^\theta(\tilde{z}_n)$ PnP denoising step



Theorem. Assume that the vector field v^θ is continuous. Let the time sequence $(t_n)_{n\in\mathbb{N}}$ satisfy $\sum_{n=0}^{\infty}(1-t_n) < +\infty$ and let $\gamma_n := 1 - t_n$, $n \in \mathbb{N}$. If the sequence $(x_n)_{n\in\mathbb{N}}$ obtained by the algorithm is bounded, then it converges.

Figure 11: Comparison of restoration methods on AFHQ-Cat: denoising (1st row), Gaussian deblurring (2nd row), super-resolution (3rd row), free-form inpainting (4th row).

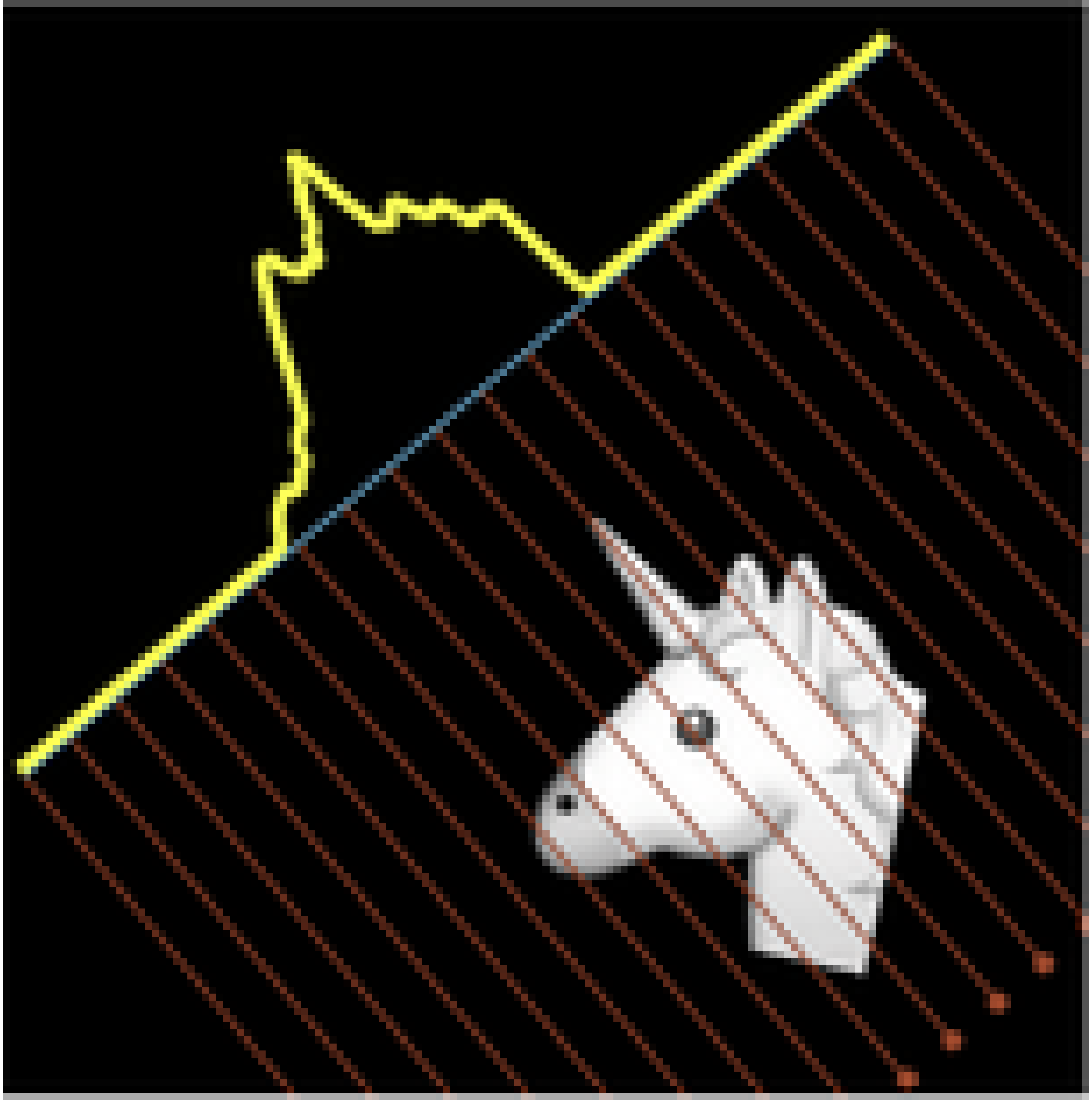
Berlin Mathematics Research Center

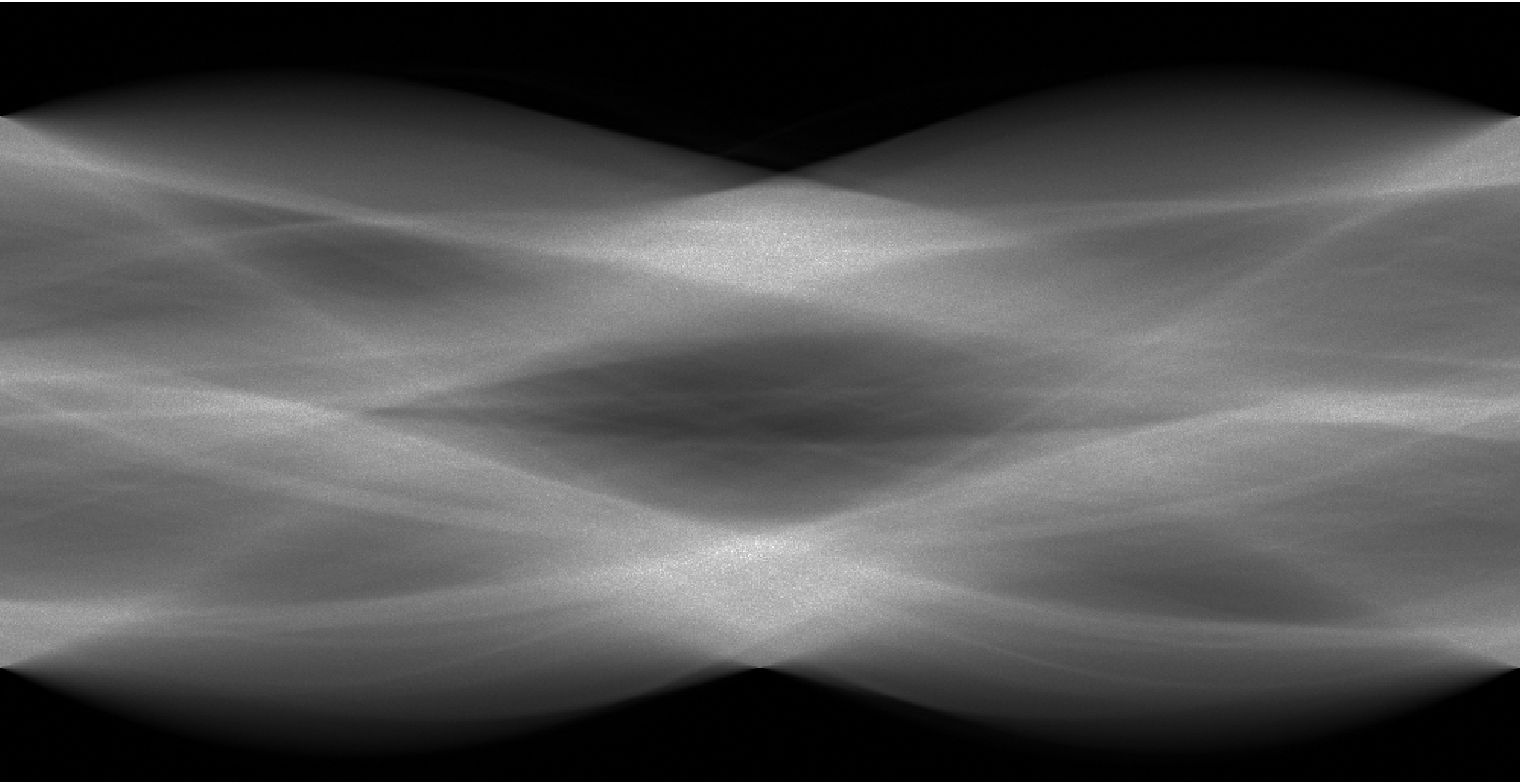
MATH+

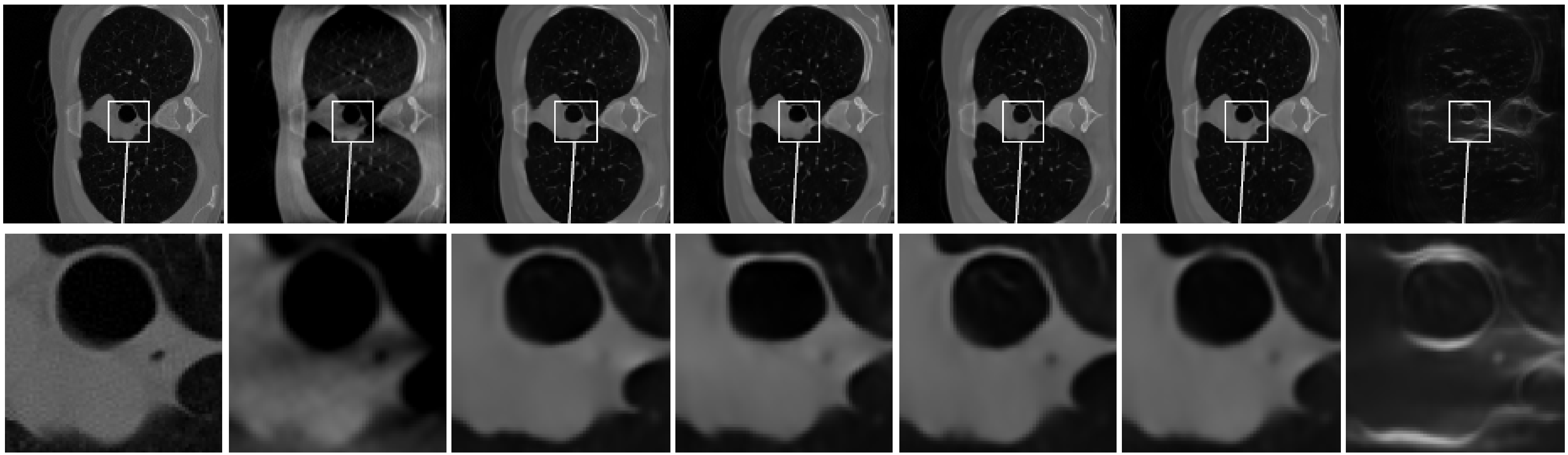
Funded under Germany's Excellence Strategy by

DFG Deutsche Forschungsgemeinschaft

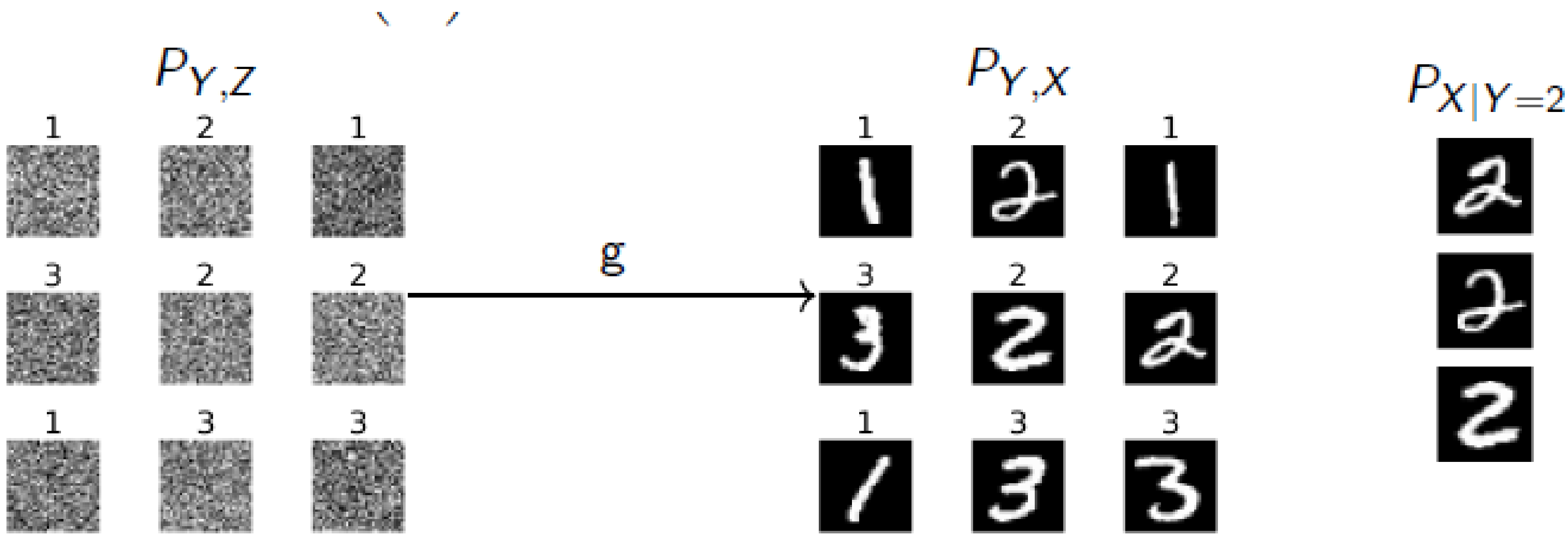

CR-461-MH



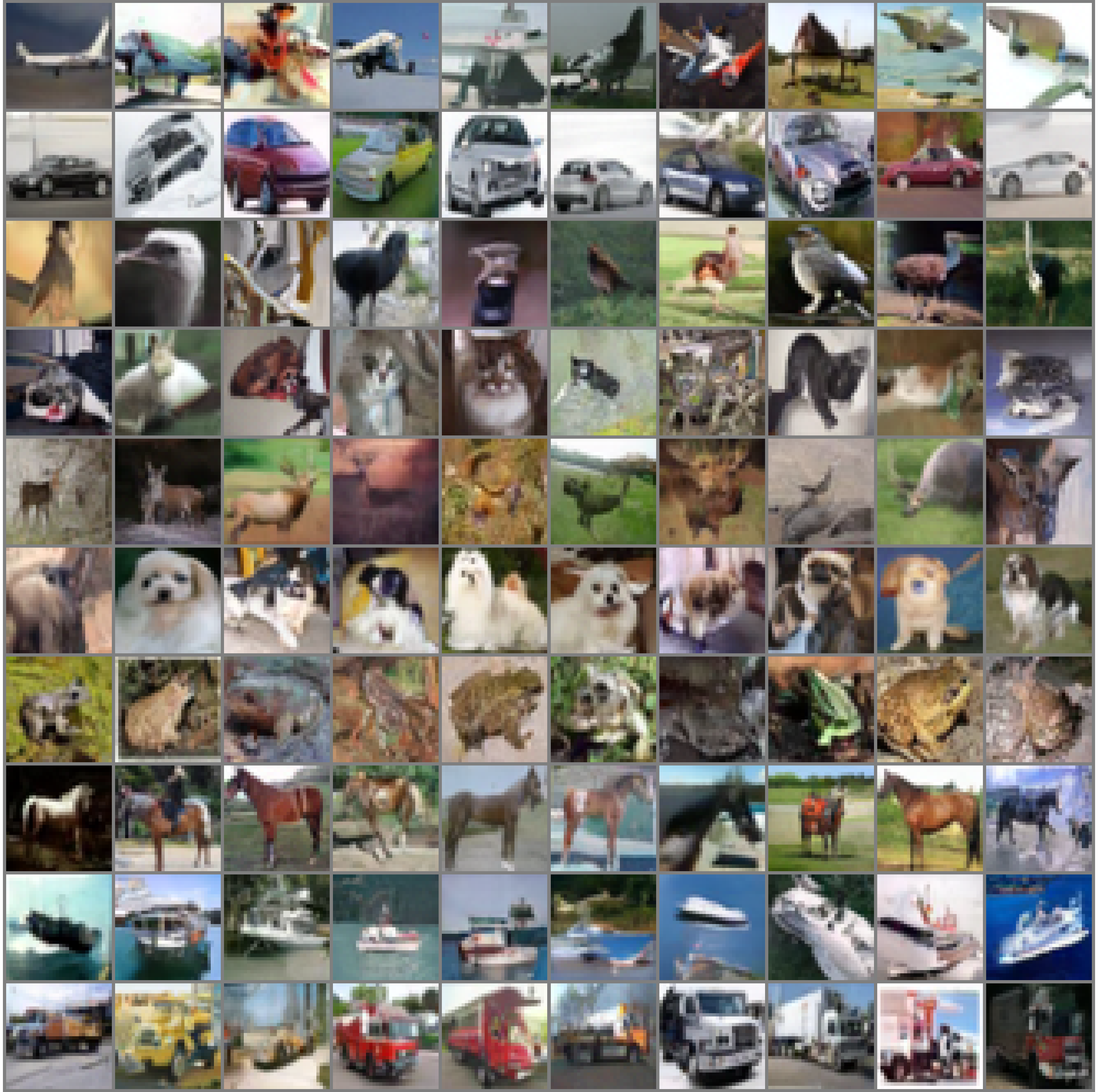


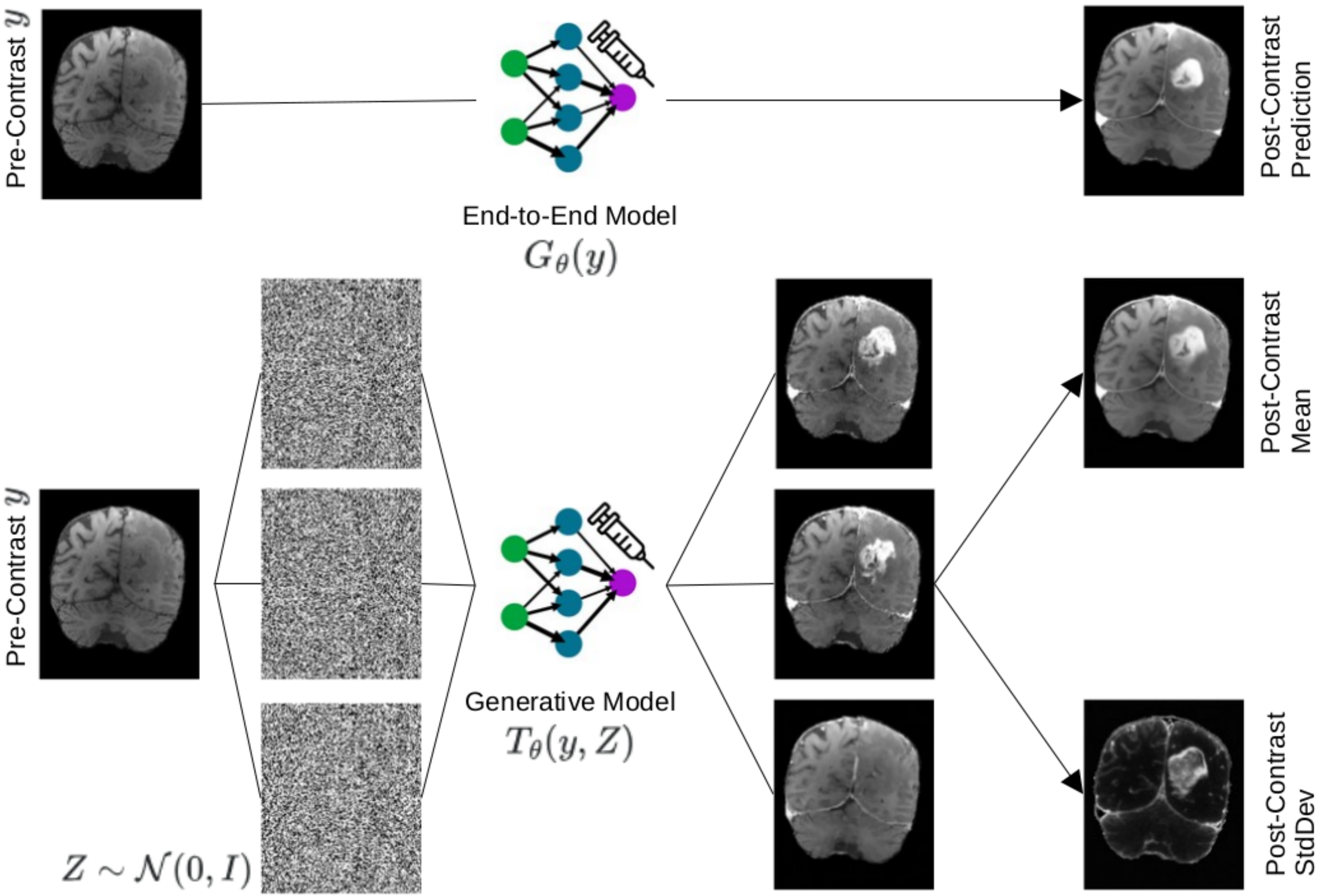


GT FBP Reco 1 Reco 2 Reco 3 Mean Std



$P_{Y,Z}$
1
2
1
3
2
2
1
3
3
g
$P_{Y,X}$
1
2
1
3
2
2
1
3
3
$P_{X|Y=2}$




Pre-Contrast y
End-to-End Model
$G_\theta(y)$
Post-Contrast Prediction
Pre-Contrast y
$Z \sim \mathcal{N}(0, I)$
Generative Model
$T_\theta(y, Z)$
Post-Contrast Mean
Post-Contrast StdDev


CR-461-MH

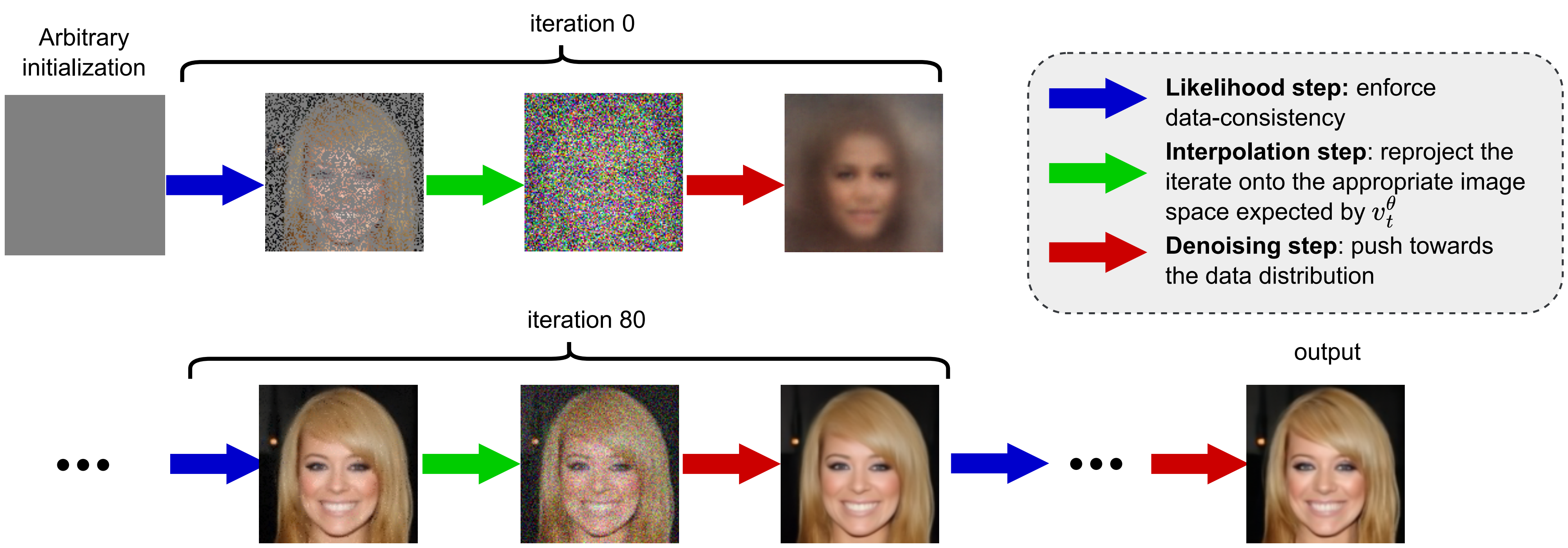

Arbitrary initialization
iteration 0
Likelihood step: enforce data-consistency
Interpolation step: reproject the iterate onto the appropriate image space expected by v_t^θ
Denoising step: push towards the data distribution
iteration 80
output

Figure 11: Comparison of restoration methods on AFHQ-Cat: denoising (1st row), Gaussian deblurring (2nd row), super-resolution (3rd row), free-form inpainting (4th row).