ICML
International Conference
On Machine Learning

Washington
University in St. Louis

UC RIVERSIDE

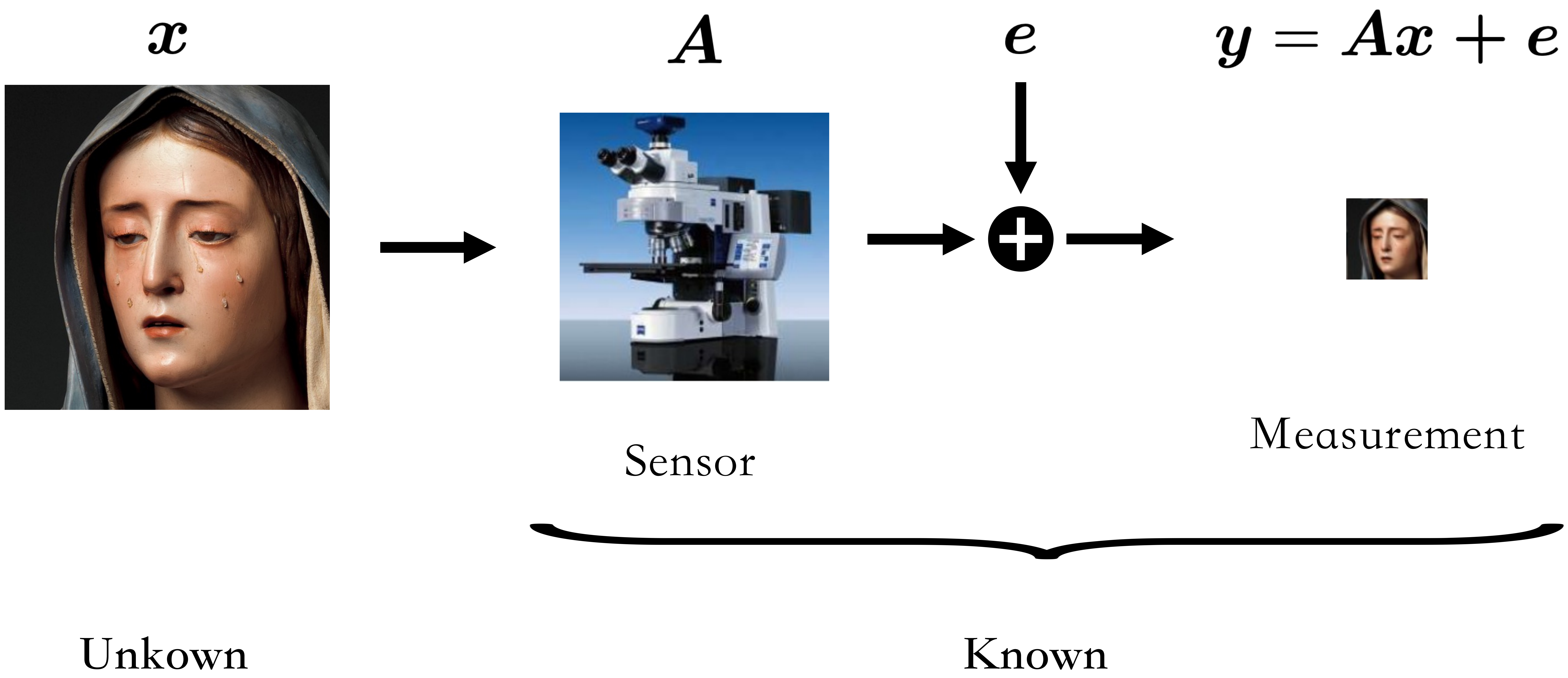
Prior Mismatch and Adaptation in PnP-ADMM with a Nonconvex Convergence Analysis

Shirin Shoushtari*, Jiaming Liu*, Edward P. Chandler, M. Salman Asif, Ulugbek S. Kamilov

ICML 2024

Many **computational imaging** problems can be viewed as inverse problems

Many computational imaging problems can be viewed as inverse problems



Many computational imaging problems can be viewed as inverse problems

Forward Problem: generate $\boldsymbol{y}$ from $\boldsymbol{x}$

$\boldsymbol{x}$

$\boldsymbol{A}$

$\boldsymbol{e}$

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$$

Many computational imaging problems can be viewed as inverse problems

$\boldsymbol{x}$

$\boldsymbol{A}$

$\boldsymbol{e}$

$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$

Inverse Problem: generate $\boldsymbol{x}$ from $\boldsymbol{y}$

Imaging inverse problems are challenging problems!

Washington University in St. Louis

- Solution is not unique
- Data is noisy
- Signals can be high-dimensional

Inverse problems can be solved with model-based optimization approach

Washington University in St. Louis

Inverse problems can be solved with model-based optimization approach

Washington University in St. Louis

- Image recovery can be formulated as an optimization task

Regularizer

$$\widehat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^n} \{g(\boldsymbol{x}) + h(\boldsymbol{x})\}$$

Data-fidelity

Inverse problems can be solved with model-based optimization approach

Washington University in St. Louis

- Image recovery can be formulated as an optimization task

Regularizer

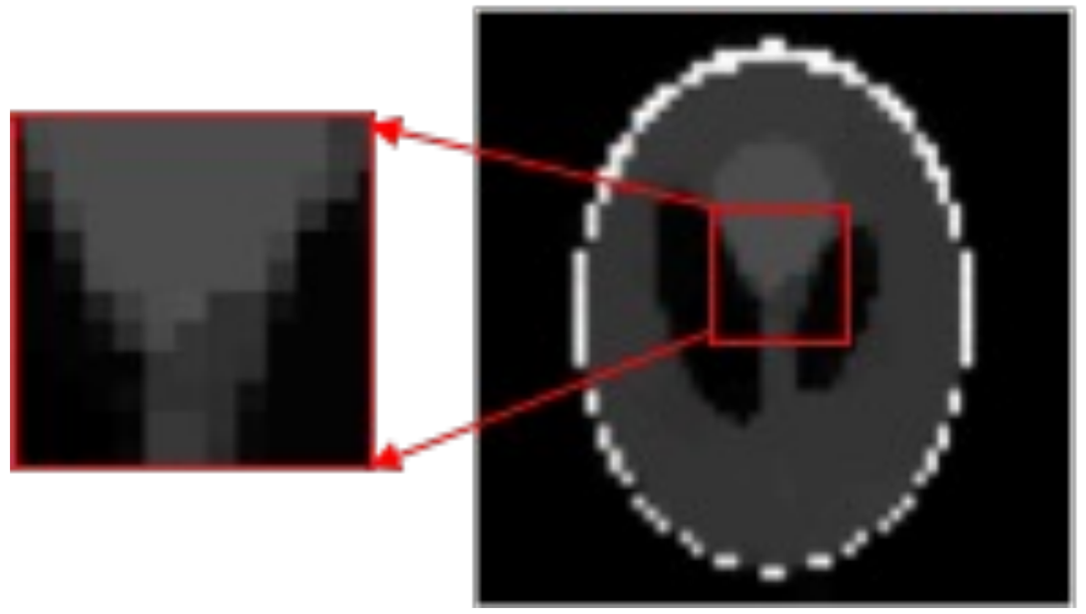
$$\widehat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{R}^n}\{g(\boldsymbol{x}) + h(\boldsymbol{x})\}$$

Data-fidelity

$$\widehat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{R}^n}\{\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \tau\|D\boldsymbol{x}\|_1\}$$

Total Variation

Regularized by TV



ADMM is an optimization algorithm designed to minimize the sum of two functions

Washington University in St. Louis

ADMM is an optimization algorithm designed to minimize the **sum of two functions**

- Introduce variable $\boldsymbol{z} = \boldsymbol{x}$

- Introduce variable $\boldsymbol{s}$ ⟶ Lagrange multiplier

- Form the augmented Lagrangian $\phi(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{s}) = g(\boldsymbol{x}) + h(\boldsymbol{z}) + \frac{1}{\gamma}\boldsymbol{s}^{\mathsf{T}}(\boldsymbol{x} - \boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2$

ADMM is an optimization algorithm designed to minimize the sum of two functions

Minimization with respect to g

ADMM is an optimization algorithm designed to minimize the sum of two functions

Washington University in St. Louis

(ADMM)

Minimization with respect to g → $\boldsymbol{x}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{z}^{k-1} - \boldsymbol{s}^{k-1})$

$$\mathsf{prox}_{\gamma g}(\boldsymbol{v}) = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\{\frac{1}{2\gamma}\|\boldsymbol{v}-\boldsymbol{x}\|_2^2 + g(\boldsymbol{x})\}$$

ADMM is an optimization algorithm designed to minimize the sum of two functions

Washington University in St. Louis

(ADMM)

Minimization with respect to g → $\boldsymbol{x}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{z}^{k-1} - \boldsymbol{s}^{k-1})$

Minimization with respect to h

ADMM is an optimization algorithm designed to minimize the sum of two functions

Washington University in St. Louis

(ADMM)

Minimization with respect to g	→	$\boldsymbol{x}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{z}^{k-1} - \boldsymbol{s}^{k-1})$
Minimization with respect to h	→	$\boldsymbol{z}^k \leftarrow \mathsf{prox}_{\gamma h}(\boldsymbol{x}^k + \boldsymbol{s}^{k-1})$

$$\mathsf{prox}_{\gamma h}(\boldsymbol{v}) = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\mathsf{arg\,min}}\{\frac{1}{2\gamma}\|\boldsymbol{v}-\boldsymbol{x}\|_2^2 + h(\boldsymbol{x})\}$$

ADMM is an optimization algorithm designed to minimize the sum of two functions

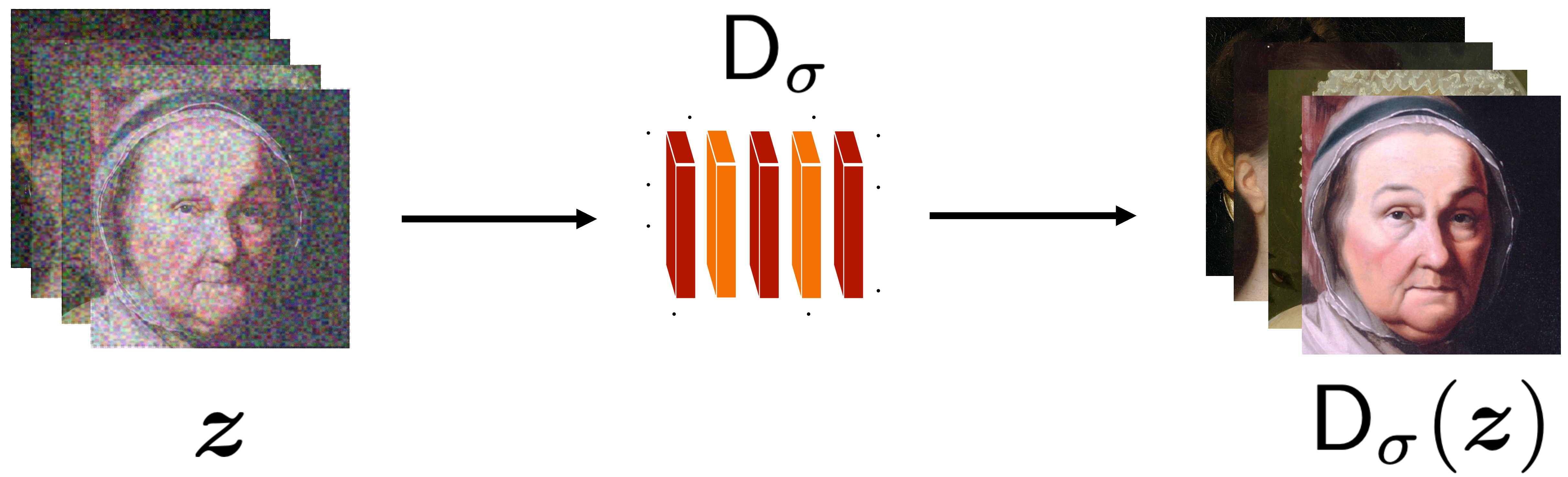
(ADMM)

Minimization with respect to g	→	$\boldsymbol{x}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{z}^{k-1} - \boldsymbol{s}^{k-1})$
Minimization with respect to h	→	$\boldsymbol{z}^k \leftarrow \mathsf{prox}_{\gamma h}(\boldsymbol{x}^k + \boldsymbol{s}^{k-1})$
Update Variables	→	$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + \boldsymbol{x}^k - \boldsymbol{z}^k$

PnP methods are deep model-based architectures

Learned model: Pre-trained **image denoising** neural networks

D_σ : more noisy image $\longrightarrow$ less noisy image



PnP methods are deep model-based architectures

Washington University in St. Louis

(ADMM)

$$\boldsymbol{x}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{z}^{k-1} - \boldsymbol{s}^{k-1})$$

$$\boldsymbol{z}^k \leftarrow \mathsf{prox}_{\gamma h}(\boldsymbol{x}^k + \boldsymbol{s}^{k-1})$$

$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + \boldsymbol{x}^k - \boldsymbol{z}^k$$

$\longrightarrow$

(PnP-ADMM)

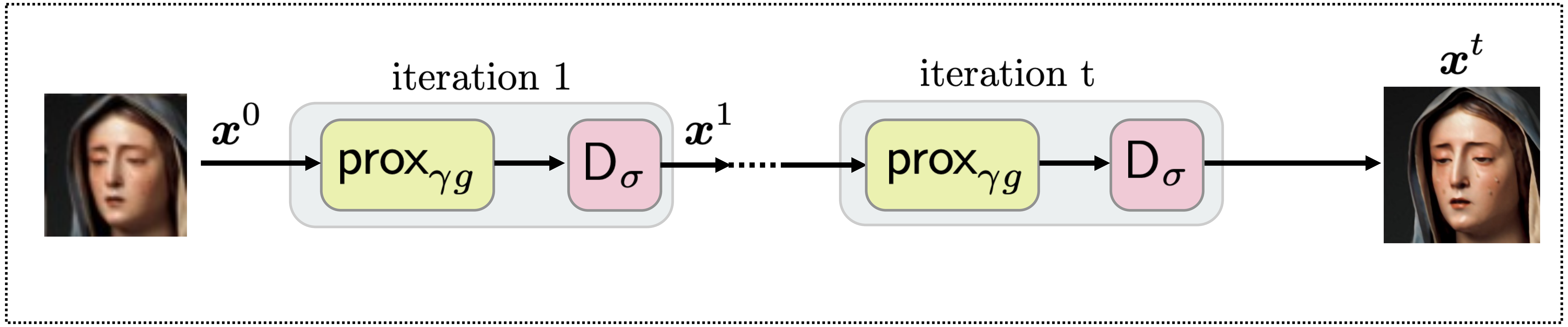
$$\boldsymbol{x}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{z}^{k-1} - \boldsymbol{s}^{k-1})$$

$$\boldsymbol{z}^k \leftarrow \mathsf{D}_{\sigma}(\boldsymbol{x}^k + \boldsymbol{s}^{k-1})$$

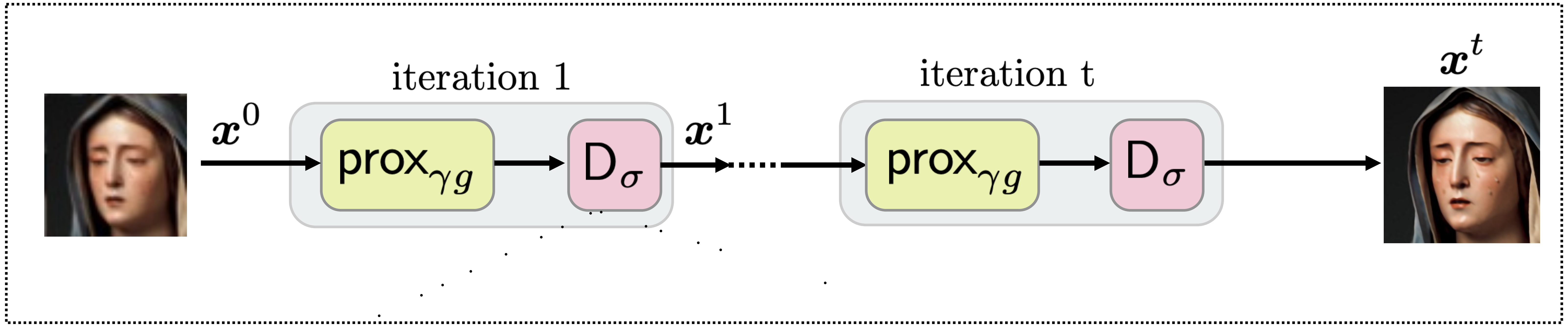
$$\boldsymbol{s}^k \leftarrow \boldsymbol{s}^{k-1} + \boldsymbol{x}^k - \boldsymbol{z}^k$$

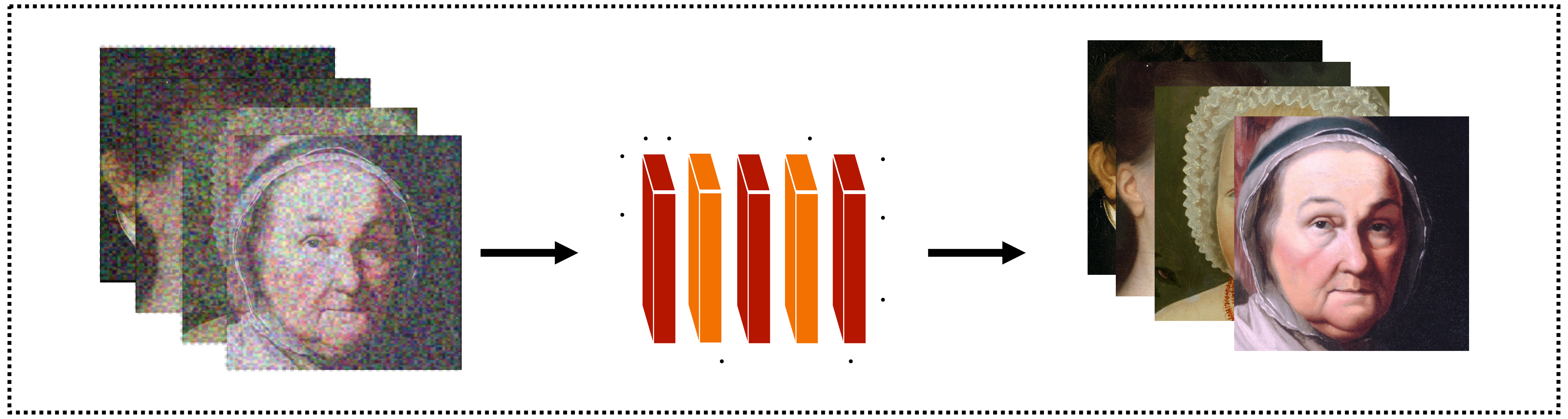
Shift distribution results in the mismatched PnP priors

Shift distribution results in the mismatched PnP priors



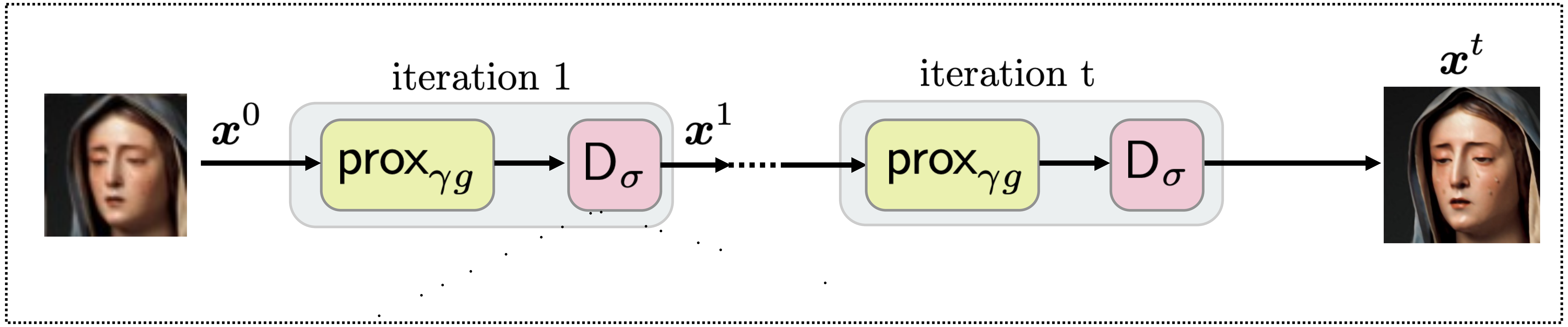
Shift distribution results in the mismatched PnP priors

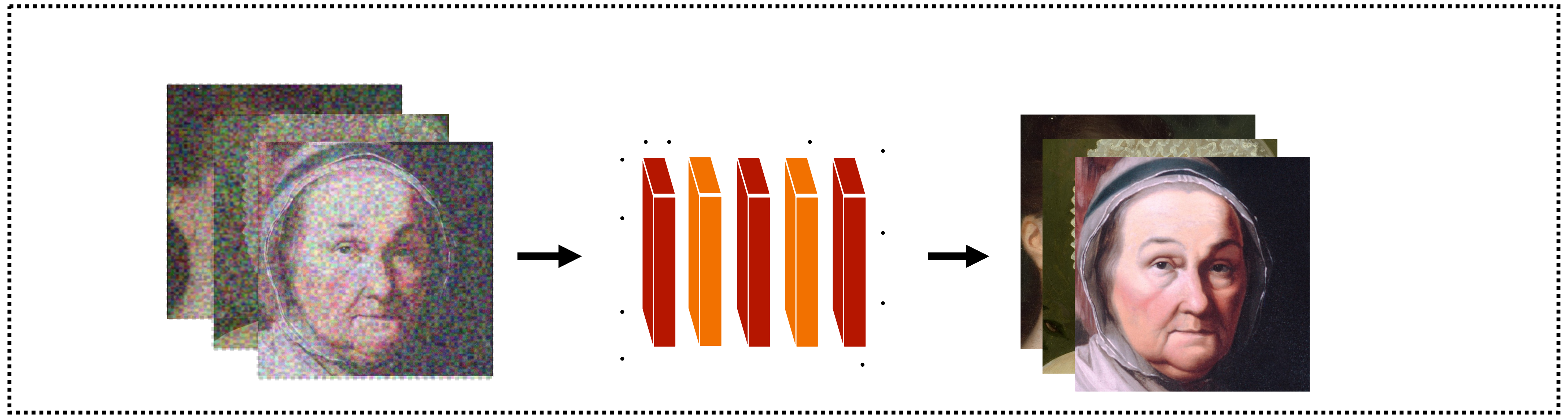




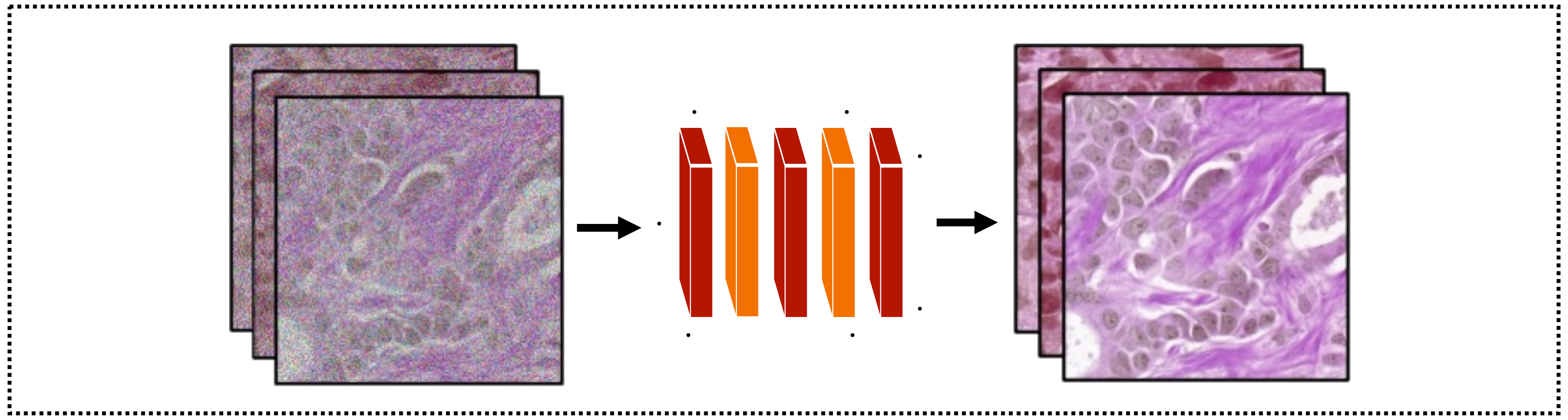
MetFaces dataset

Shift distribution results in the mismatched PnP priors





MetFaces dataset



BreCaHAD dataset

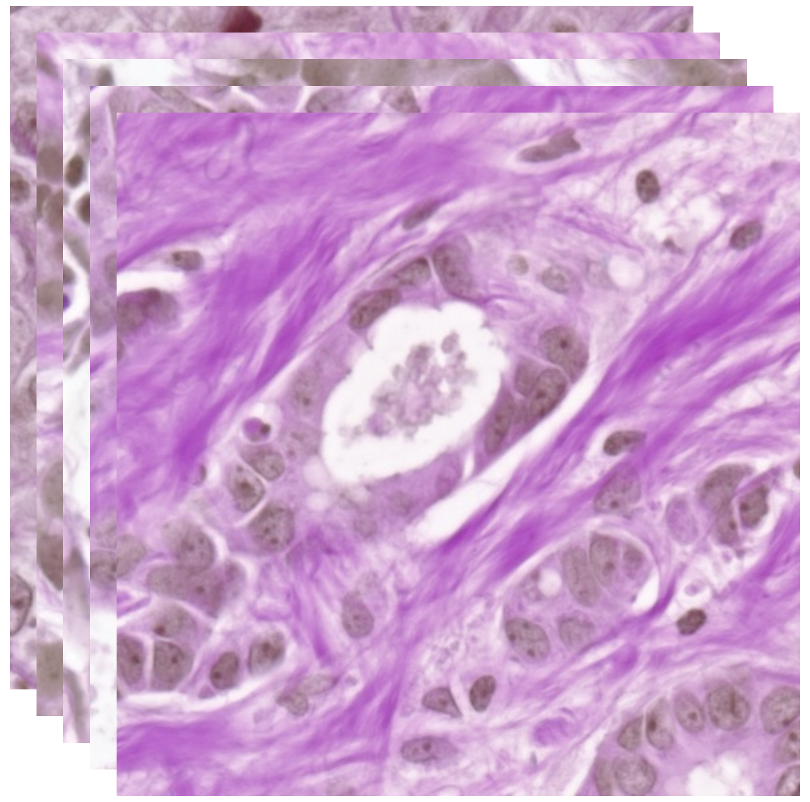
Shift distribution results in the mismatched PnP priors

MetFaces

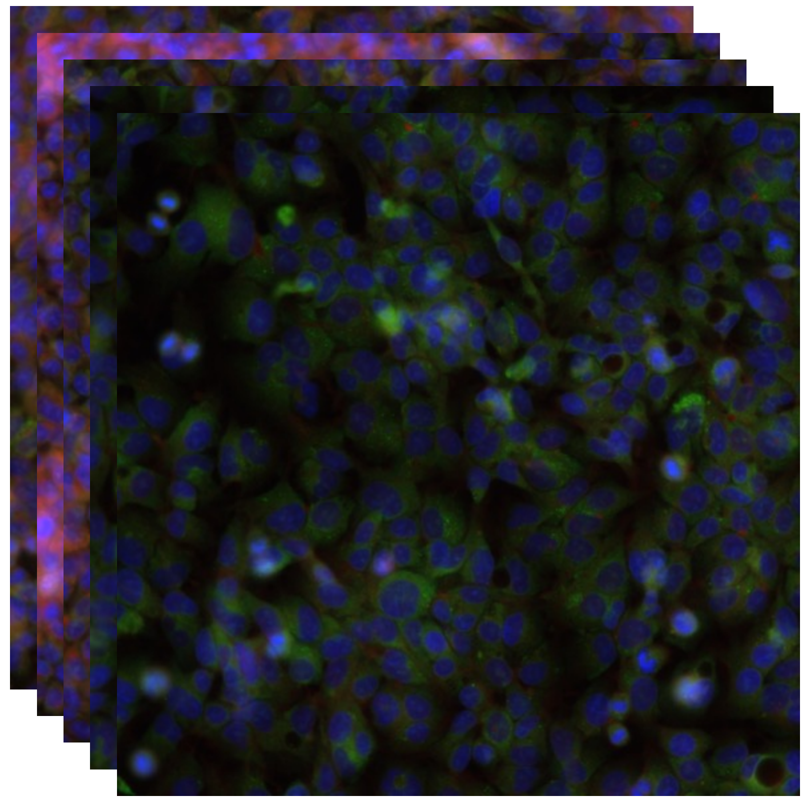
CelebA

AFHQ

BreCaHAD

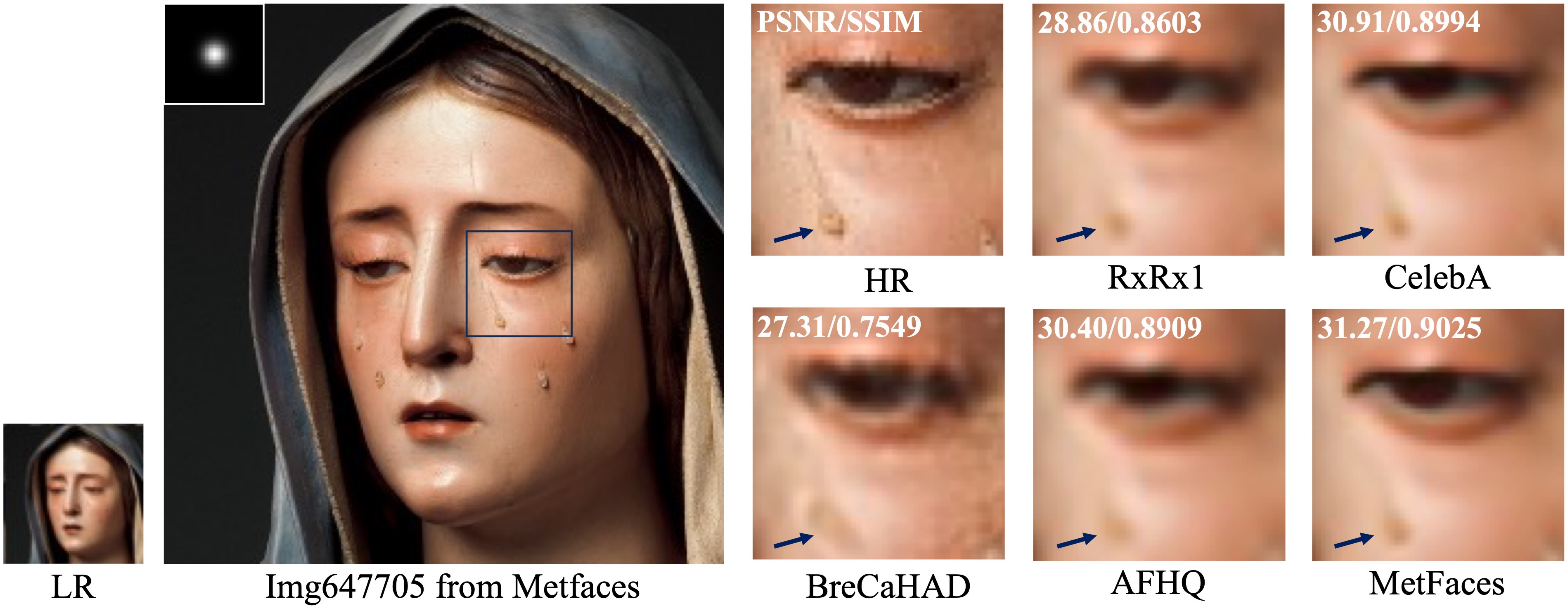


RxRx1



Mismatched priors result in suboptimal performance in PnP

Washington University in St. Louis



PnP-ADMM theoretically converges for nonconvex data term and expansive MMSE denoisers

- Suppose the mismatched denoiser is trained with the MSE loss

$$\widehat{\mathsf{D}}_\sigma(\boldsymbol{z}) = \underset{\widehat{\mathsf{D}}}{\arg\min}\, \mathbb{E}[\|\boldsymbol{x} - \widehat{\mathsf{D}}(\boldsymbol{z})\|_2^2]$$

MMSE denoiser for $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$

PnP-ADMM theoretically converges for nonconvex data term and expansive MMSE denoisers

Washington University in St. Louis

- Suppose the mismatched denoiser is trained with the MSE loss

$$\widehat{\mathsf{D}}_\sigma(\boldsymbol{z}) = \underset{\widehat{\mathsf{D}}}{\arg\min}\, \mathbb{E}[\|\boldsymbol{x} - \widehat{\mathsf{D}}(\boldsymbol{z})\|_2^2]$$

MMSE denoiser for $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$

- Assume that mismatched denoiser $\widehat{\mathsf{D}}_\sigma$ and target denoiser D_σ have a bounded distance

$$\|\widehat{\mathsf{D}}_\sigma(\boldsymbol{x}^k) - \mathsf{D}_\sigma(\boldsymbol{x}^k)\|_2 \leq \delta_k$$

PnP-ADMM theoretically converges for nonconvex data term and expansive MMSE denoisers

Washington University in St. Louis

- Suppose the denoiser is trained with the MSE loss

$$\widehat{\mathsf{D}}_\sigma(\boldsymbol{z}) = \mathop{\mathsf{arg\,min}}_{\widehat{\mathsf{D}}} \mathbb{E}[\|\boldsymbol{x} - \widehat{\mathsf{D}}(\boldsymbol{z})\|_2^2]$$

MMSE denoiser for $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{e}$

- Assume that mismatched denoiser $\widehat{\mathsf{D}}_\sigma$ and target denoiser D_σ have a bounded distance

$$\|\widehat{\mathsf{D}}_\sigma(\boldsymbol{x}^k) - \mathsf{D}_\sigma(\boldsymbol{x}^k)\|_2 \leq \delta_k$$

- Define the error $\varepsilon_k := \max\{\delta_k, \delta_k^2\}$

PnP-ADMM theoretically converges for nonconvex data term and expansive MMSE denoisers

Run PnP-ADMM with L-smooth regularization term, mismatched MMSE denoiser $\widehat{\mathsf{D}}_\sigma$ with per-iteration ε_k error, and penalty parameter $0 < \gamma \leq (4L)$ for $t \geq 1$ iterations. There exits $f = g + h$

$$\frac{1}{t}\sum_{k=1}^{t} \|\nabla f(\boldsymbol{x}^k)\|_2^2 \leq \frac{A_1}{t}(\phi(\boldsymbol{x}^0, \boldsymbol{z}^0, \boldsymbol{s}^0) - \phi^*) + A_2\overline{\varepsilon}_t$$

- $\overline{\varepsilon}_t = (1/t)(\varepsilon_1 + \cdots + \varepsilon_t)$
- If ε_k is summable, then $\nabla f(\boldsymbol{x}^t) \to \mathbf{0}$

PnP-ADMM theoretically converges for nonconvex data term and expansive MMSE denoisers

Washington University in St. Louis

- Suppose the denoiser is trained with the MSE loss without any mismatch

$$\mathsf{D}_\sigma(\boldsymbol{z}) = \underset{\mathsf{D}}{\arg\min}\, \mathbb{E}[\|\boldsymbol{x} - \mathsf{D}(\boldsymbol{z})\|_2^2]$$

Run PnP-ADMM with L-smooth regularization term, target MMSE denoiser D_σ, and penalty parameter $0 < \gamma \leq (4L)$ for $t \geq 1$ iterations. Then we have

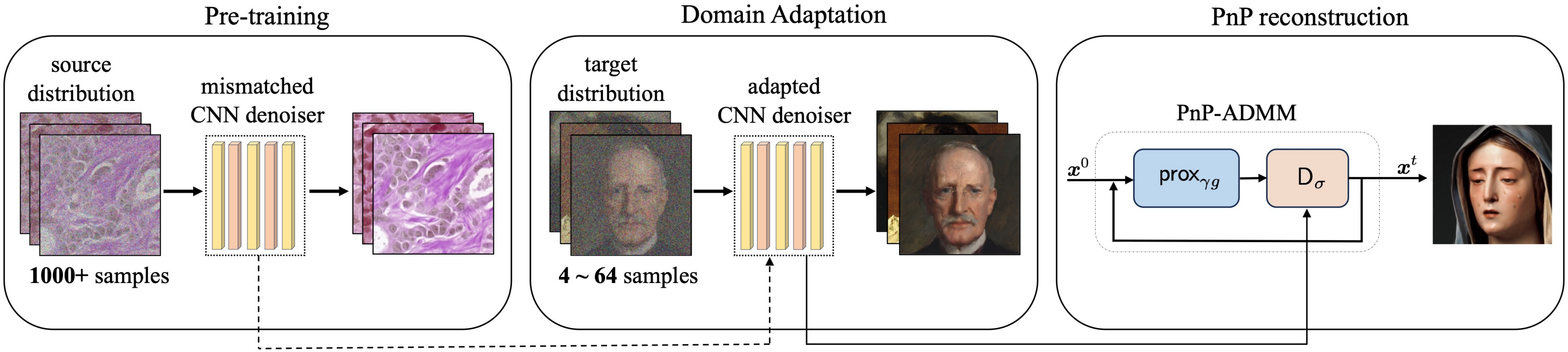
$$\frac{1}{t}\sum_{k=1}^{t} \|\nabla f(\boldsymbol{x}^k)\|_2^2 \leq \frac{C}{t}(\phi(\boldsymbol{x}^0, \boldsymbol{z}^0, \boldsymbol{s}^0) - \phi^*)$$

Run PnP-ADMM with L-smooth regularization term, mismatched MMSE denoiser $\widehat{\mathsf{D}}_\sigma$ with per-iteration ε_k error, and penalty parameter $0 < \gamma \leq (4L)$ for $t \geq 1$ iterations. Then we have

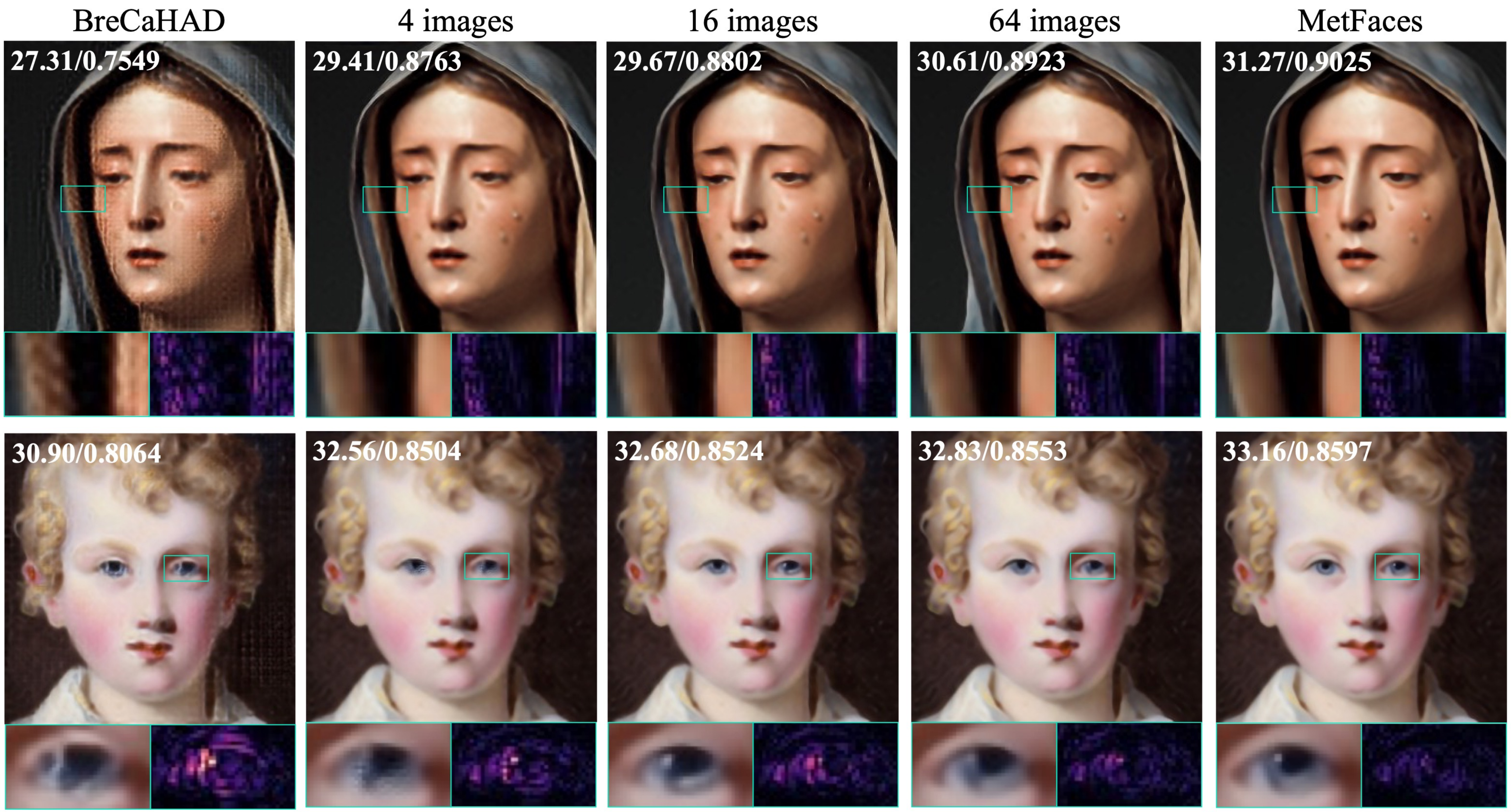
$$\frac{1}{t}\sum_{k=1}^{t} \|\nabla f(\boldsymbol{x}^k)\|_2^2 \leq \frac{A_1}{t}(\phi(\boldsymbol{x}^0, \boldsymbol{z}^0, \boldsymbol{s}^0) - \phi^*) + A_2\overline{\varepsilon}_t$$

Domain Adaptation is an strategy to close the gap that arises with shift distribution in PnP priors

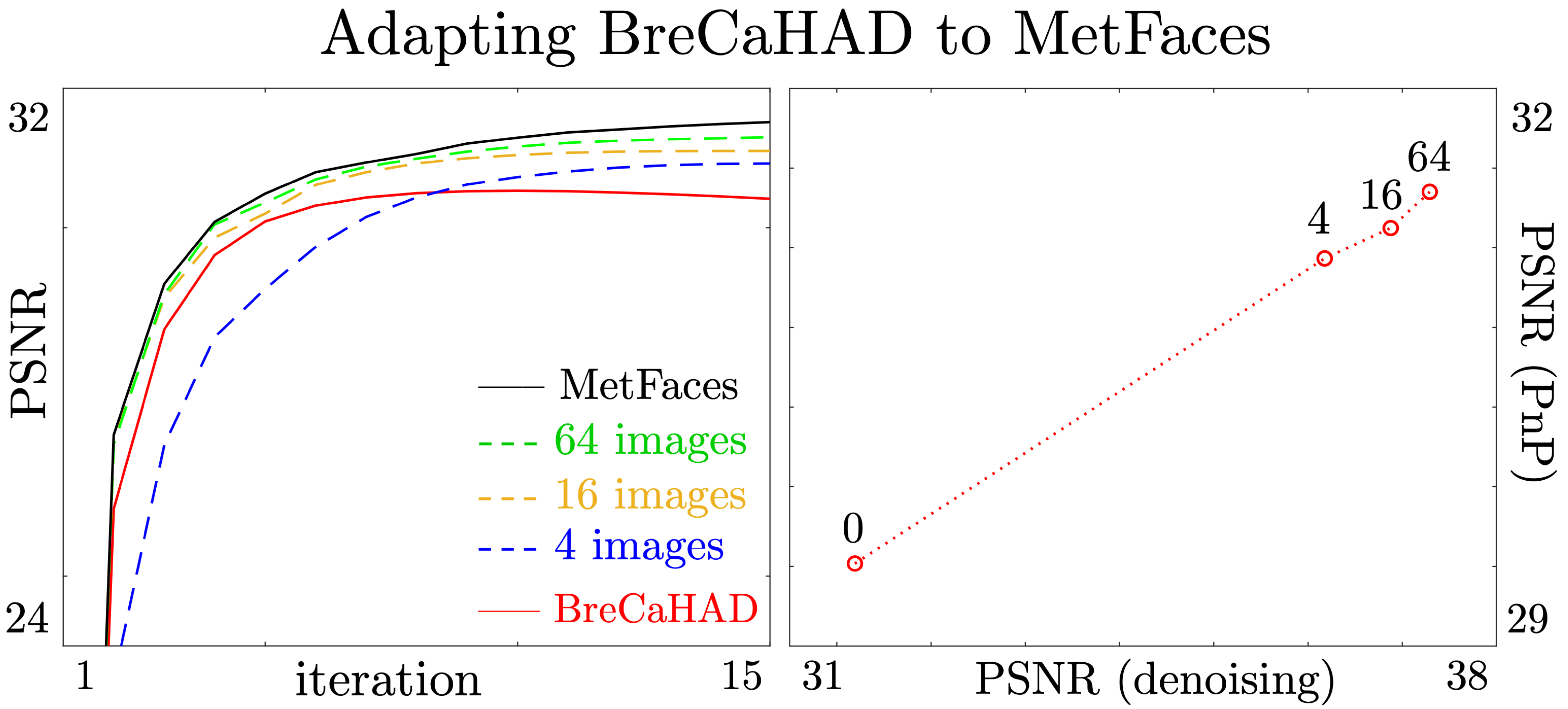
Domain Adaptation is an strategy to close the gap that arises with shift distribution in PnP priors



Domain Adaptation is an strategy to close the gap that arises with Distribution shift in PnP priors



Domain Adaptation is an strategy to close the gap that arises with Distribution shift in PnP priors



To conclude

Washington University in St. Louis

- Influence of mismatched denoisers in PnP-ADMM can be evaluated theoretically and numerically
- Domain adaptation is able to reduce the effect of distribution mismatch in PnP methods