Shrinkage Fields for Effective Image Restoration

fast

high quality

Uwe Schmidt and **Stefan Roth**
Department of Computer Science, TU Darmstadt

Image Restoration in Academia

- **Tested on small images** of less than 1 megapixel (MP)
- Restoration typically **takes minutes**

Denoising

481 × 321

Non-blind Deconvolution

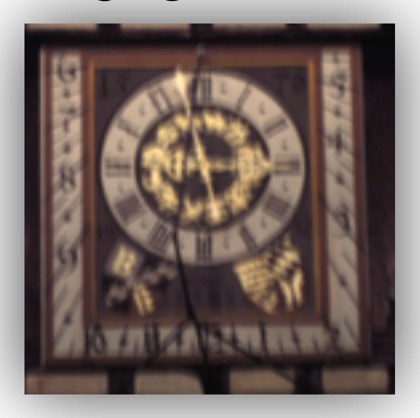
800 × 800

Challenge: Larger Images

- People take **pictures** of **8+ MP nowadays**
- **High-quality methods do not scale**
  - restoration can take hours
- Fast methods
  → lower quality

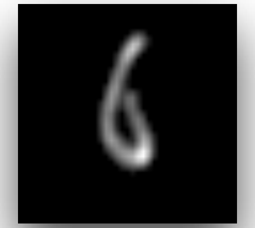
0.2 MP

0.6 MP



Start: MRF-based Image Restoration

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \underbrace{\frac{\lambda}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2}_{\text{Data term}} + \quad \text{MRF prior}$$

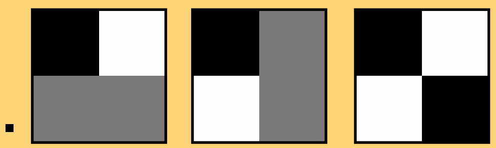


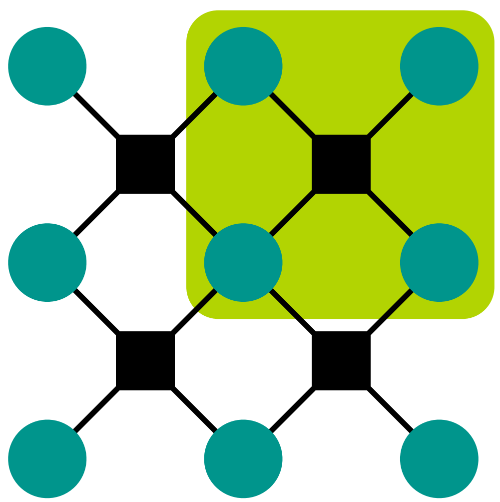
Start: MRF-based Image Restoration

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \overbrace{\frac{\lambda}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2}^{\text{Data term}} + \overbrace{\sum_{c \in \mathcal{C}} \sum_{i=1}^{N} \rho_i(\mathbf{f}_i^{\mathsf{T}} \mathbf{x}_{(c)})}^{\text{MRF prior}}$$

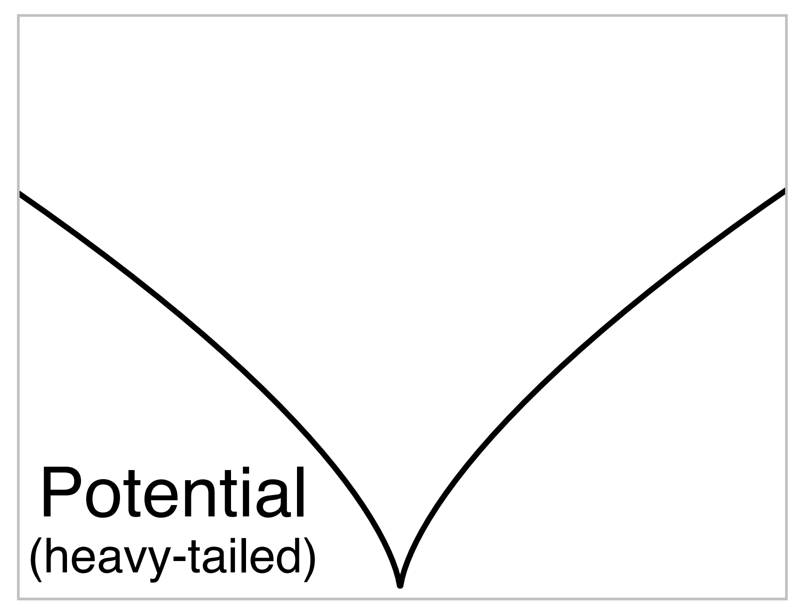
Linear filter

e.g.





Cliques/
factors



[Geman & Reynolds '92; Zhu & Mumford '97, Roth & Black '05]

Half-quadratic MAP Inference

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{\lambda}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \sum_{c \in \mathcal{C}} \sum_{i=1}^{N} \rho_i(\mathbf{f}_i^{\mathsf{T}} \mathbf{x}_{(c)})$$

$$\equiv \arg\min_{\mathbf{x},\mathbf{z}} \frac{\lambda}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \sum_{c \in \mathcal{C}} \sum_{i=1}^{N} \phi_i(\mathbf{f}_i^{\mathsf{T}} \mathbf{x}_{(c)}, z_{ic})$$

- **Half-quadratic inference**
  [Geman et al. '92 '95]

Update
restored image $\mathbf{x}$

Iterate

Update
latent variables $\mathbf{z}$

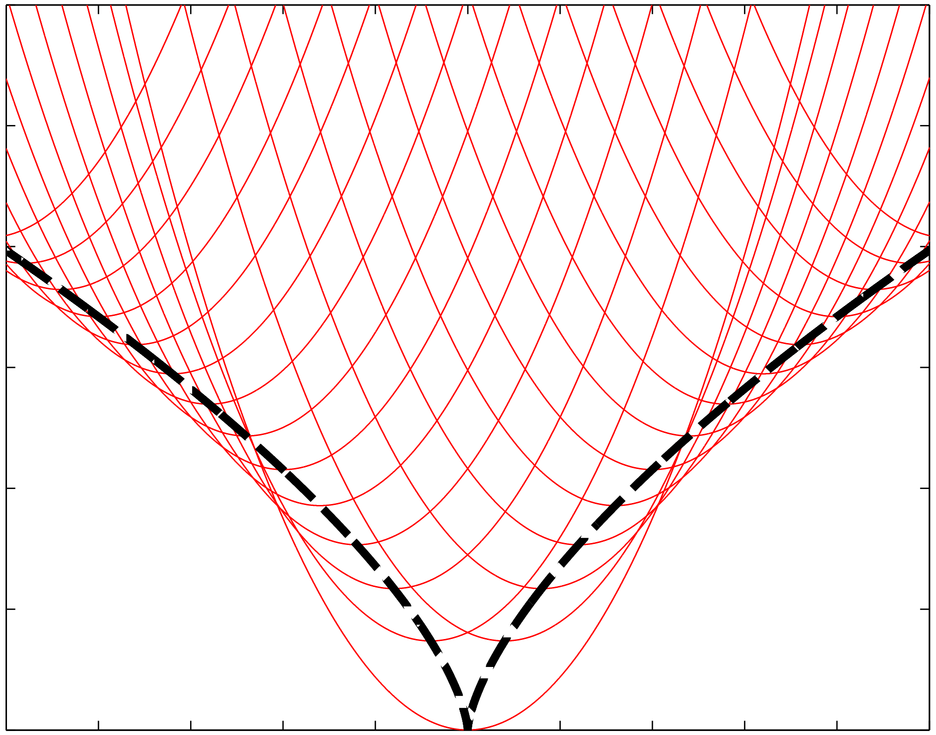
Half-quadratic Additive Form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \; \frac{\lambda}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \sum_{c\in\mathcal{C}}\sum_{i=1}^{N} \rho_i(\mathbf{f}_i^{\mathsf{T}}\mathbf{x}_{(c)})$$

$$\equiv \arg\min_{\substack{\mathbf{x},\mathbf{z}\\ \beta\to\infty}} \; \frac{\lambda}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \sum_{c\in\mathcal{C}}\sum_{i=1}^{N} \rho_i(z_{ic}) + \frac{\beta}{2}\left(\mathbf{f}_i^{\mathsf{T}}\mathbf{x}_{(c)} - z_{ic}\right)^2$$

- **Additive half-quadratic form**
  [Wang et al. '08;
  Krishnan & Fergus '09]

Quadratic relaxation



Half-quadratic Additive Form

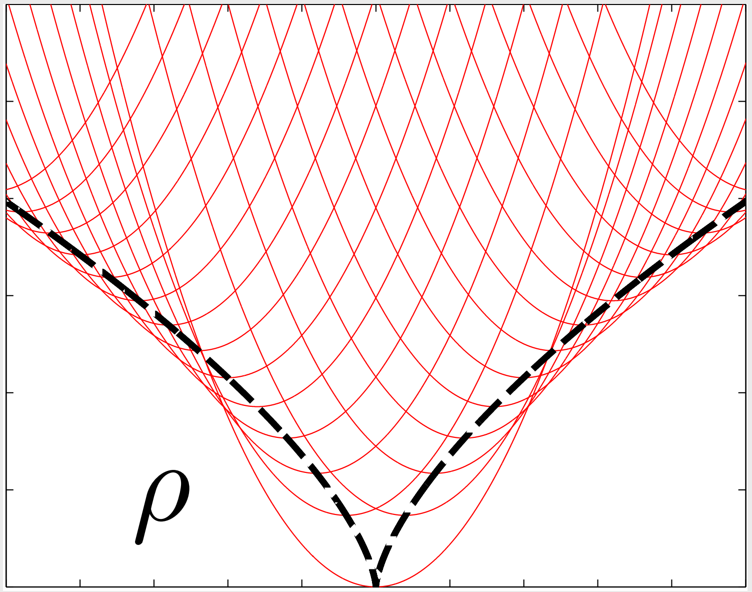
- Computationally appealing

1. Updating latent variables z:

$$z_{ic}^{t+1} \leftarrow f_\beta(v) = \arg\min_z \left(\rho(z) + \frac{\beta}{2}(v-z)^2 \right)$$

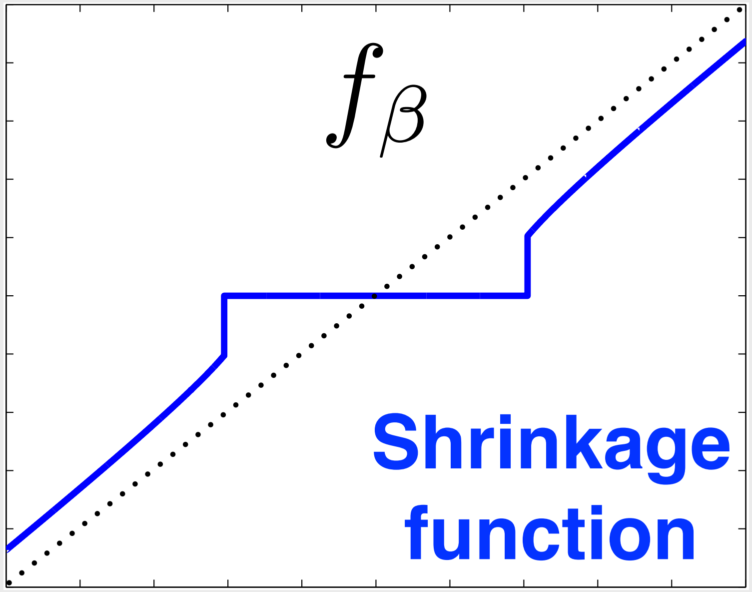
per filter response

$$v = \mathbf{f}_i^\mathsf{T} \mathbf{x}_{(c)}^t$$



$\arg\min$

fast
(1D lookup)



Half-quadratic Additive Form

TECHNISCHE UNIVERSITÄT DARMSTADT

- Computationally appealing

1. Updating latent variables z:

$$z_{ic}^{t+1} \leftarrow f_\beta(v) = \arg\min_z \left(\rho(z) + \frac{\beta}{2}(v - z)^2\right)$$

per filter response

$$v = \mathbf{f}_i^\mathsf{T} \mathbf{x}_{(c)}^t$$

2. Updating the image x:

$$\mathbf{x}^{t+1} \leftarrow \mathbf{\Omega}^{-1} \boldsymbol{\eta}\left(\mathbf{z}^{t+1}\right)$$

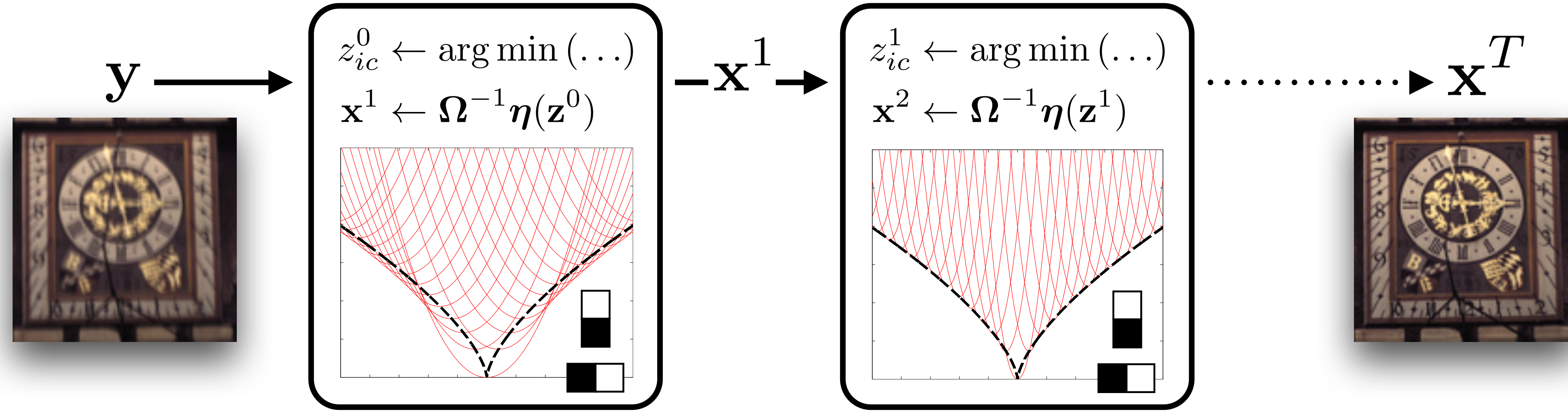
Inference in a Gaussian CRF

- System matrix $\mathbf{\Omega}$ can be diagonalized using **Fourier transform**
- Fast – Complexity $O(n \log n)$

From MRF to Shrinkage Fields

TECHNISCHE
UNIVERSITÄT
DARMSTADT

MRF



Our approach:

- Keep **efficient inference**
- **Richer models**
- **Learning** for high-quality results

TECHNISCHE UNIVERSITÄT DARMSTADT

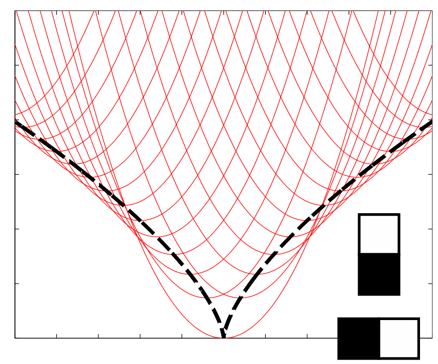
From MRF to Shrinkage Fields

MRF

$\mathbf{y} \longrightarrow$

$z_{ic}^{0} \leftarrow \arg\min\,(\ldots)$

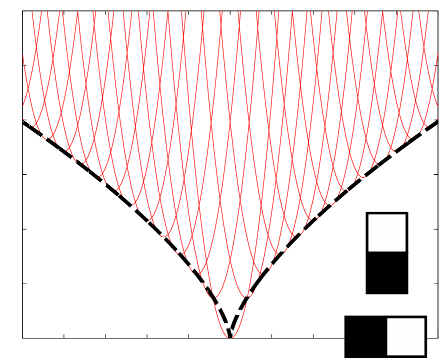
$\mathbf{x}^{1} \leftarrow \boldsymbol{\Omega}^{-1}\boldsymbol{\eta}(\mathbf{z}^{0})$



$\longrightarrow \mathbf{x}^{1} \longrightarrow$

$z_{ic}^{1} \leftarrow \arg\min\,(\ldots)$

$\mathbf{x}^{2} \leftarrow \boldsymbol{\Omega}^{-1}\boldsymbol{\eta}(\mathbf{z}^{1})$



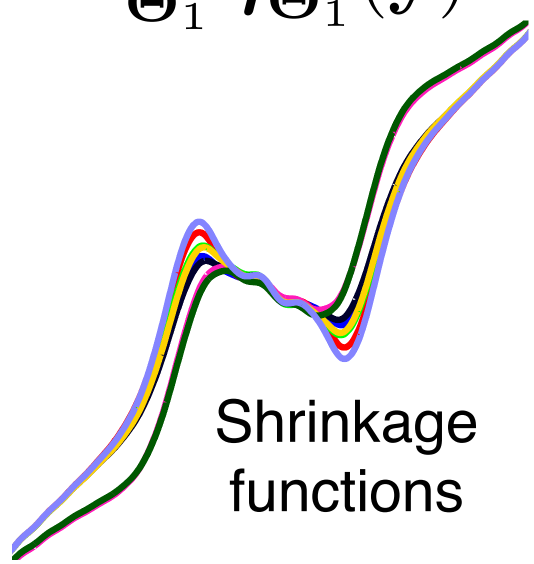
$\cdots\cdots\cdots\blacktriangleright\ \mathbf{x}^{T}$

Replace potential with shrinkage function

CSF

$\mathbf{y} \longrightarrow$

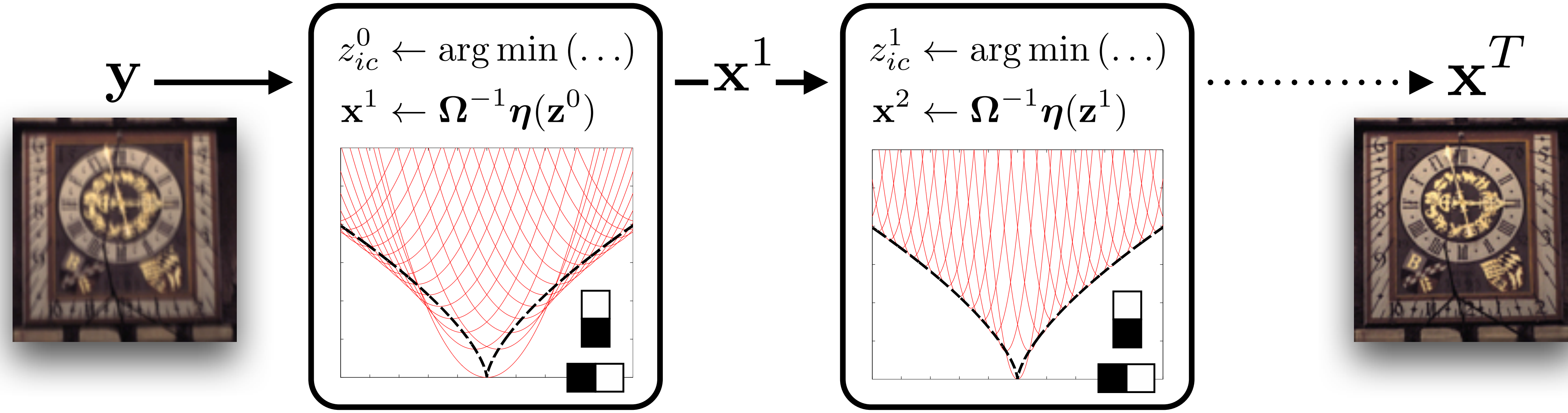
$\boldsymbol{\Omega}_{\boldsymbol{\Theta}_1}^{-1}\boldsymbol{\eta}_{\boldsymbol{\Theta}_1}(\mathbf{y})$



$\cdots\cdots\ \mathbf{x}^{1}\ \cdots\blacktriangleright$

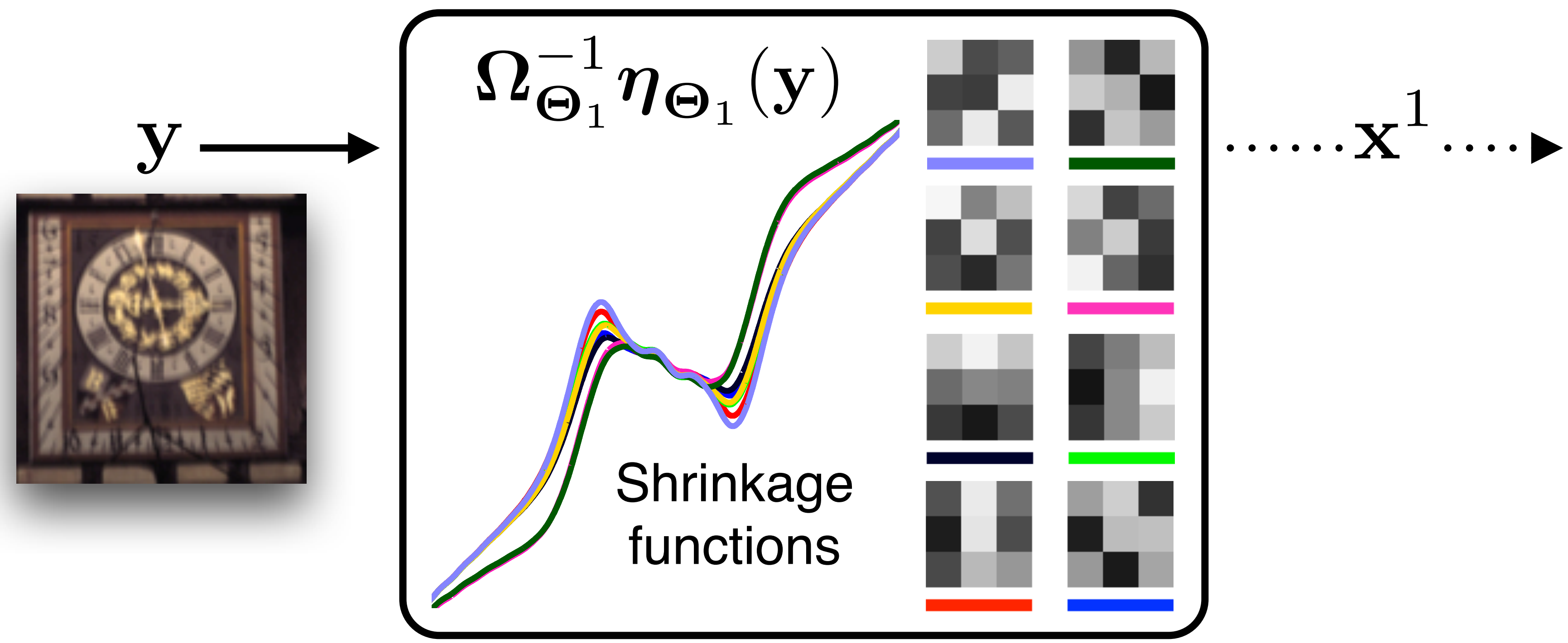
From MRF to Shrinkage Fields

MRF



Loss-based training of filters and shrinkage functions

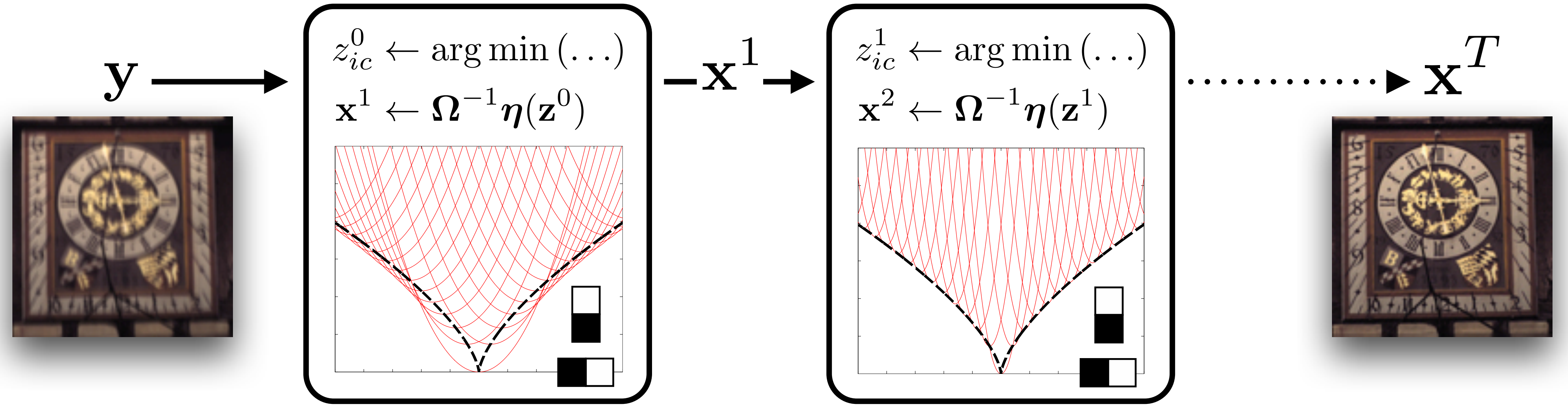
CSF



From MRF to Shrinkage Fields

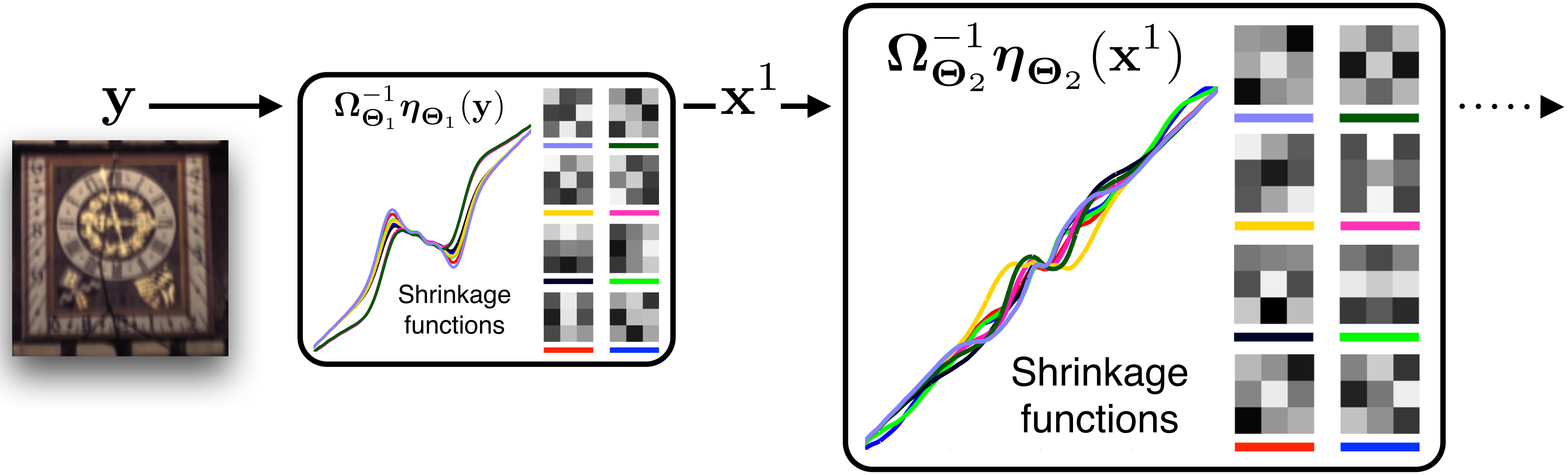
TECHNISCHE UNIVERSITÄT DARMSTADT

MRF



Custom parameters per stage

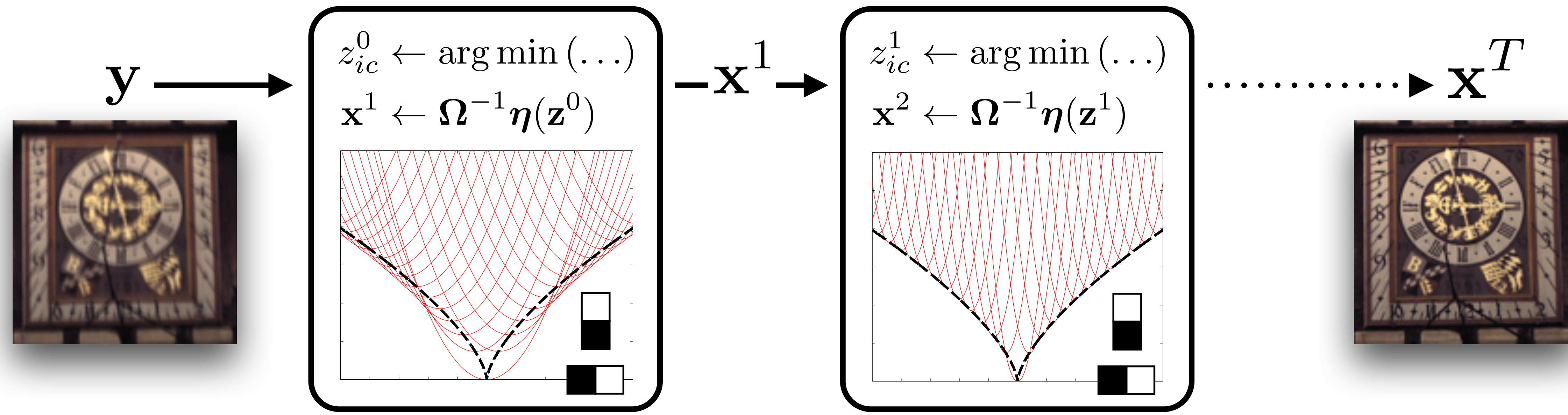
CSF



From MRF to Shrinkage Fields

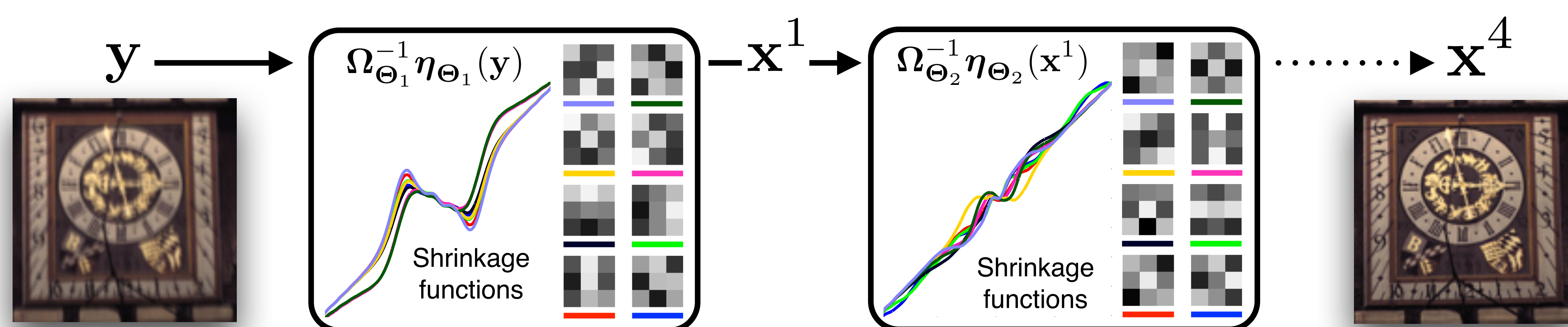
TECHNISCHE
UNIVERSITÄT
DARMSTADT

MRF



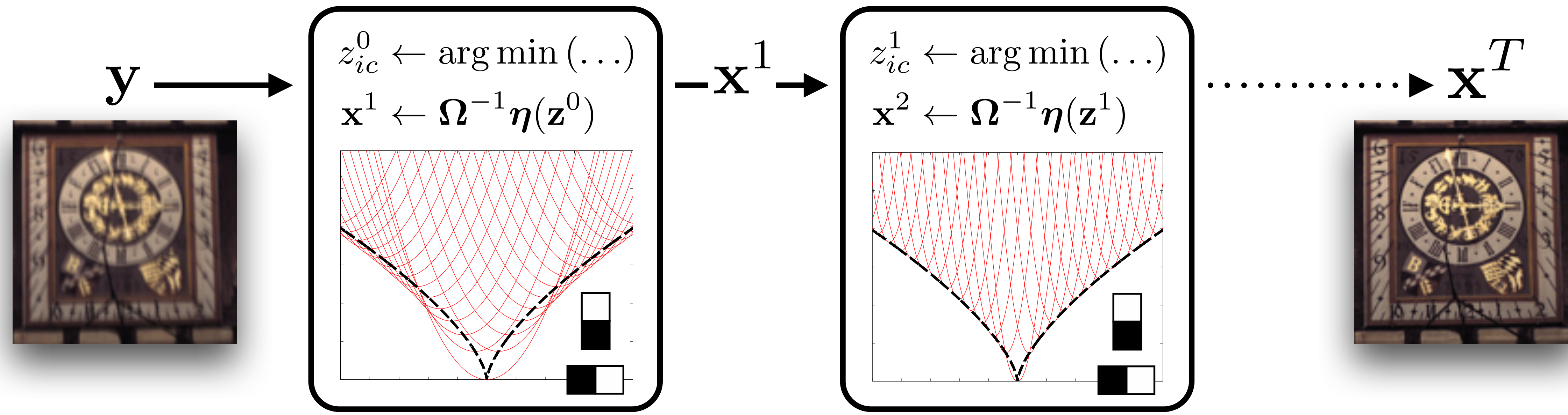
Fixed, small number of stages

CSF

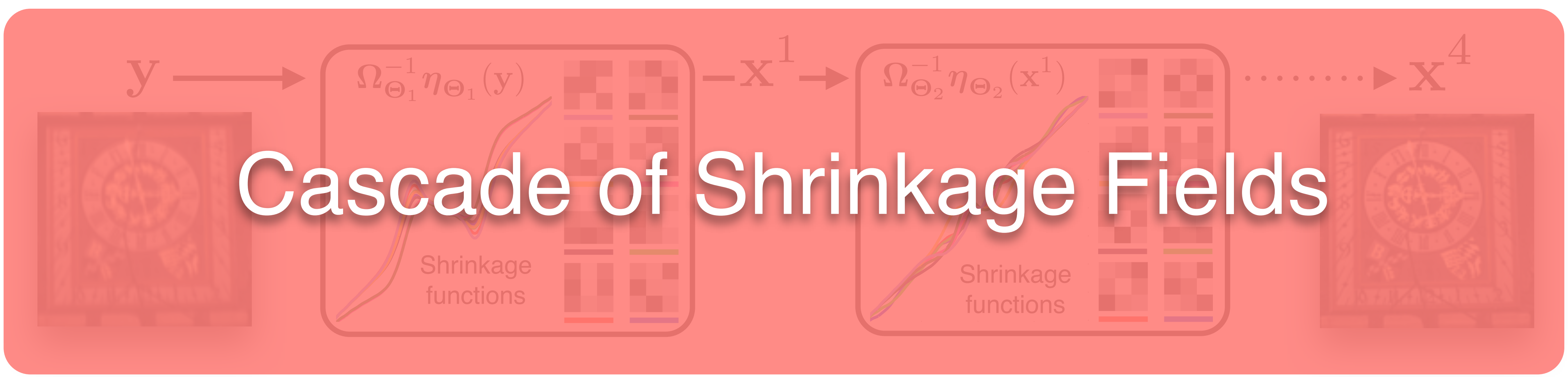


From MRF to Shrinkage Fields

MRF



CSF

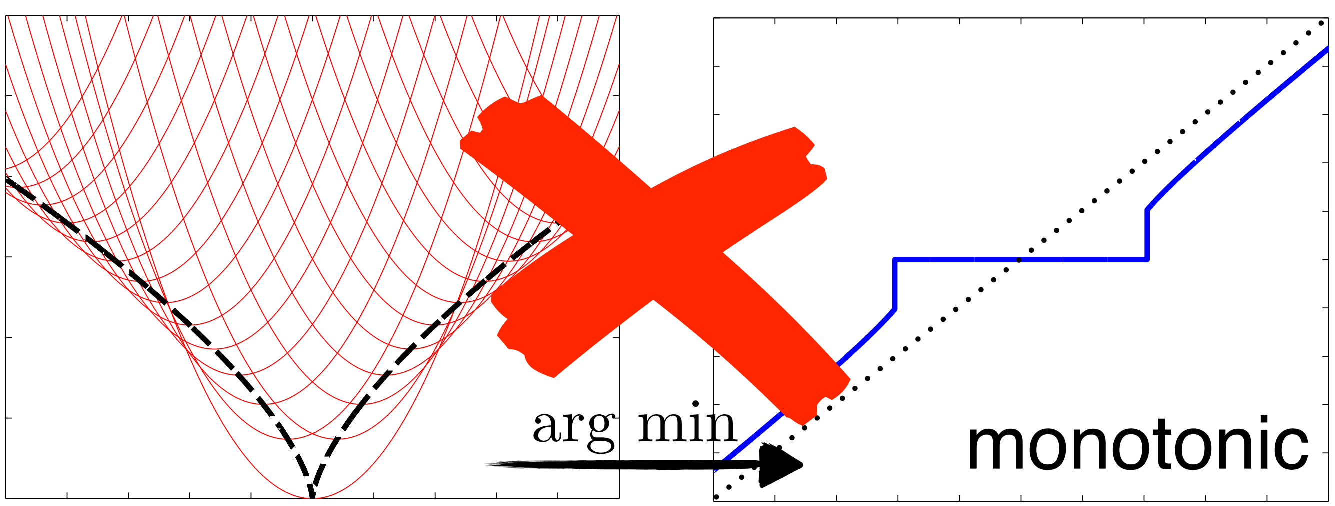


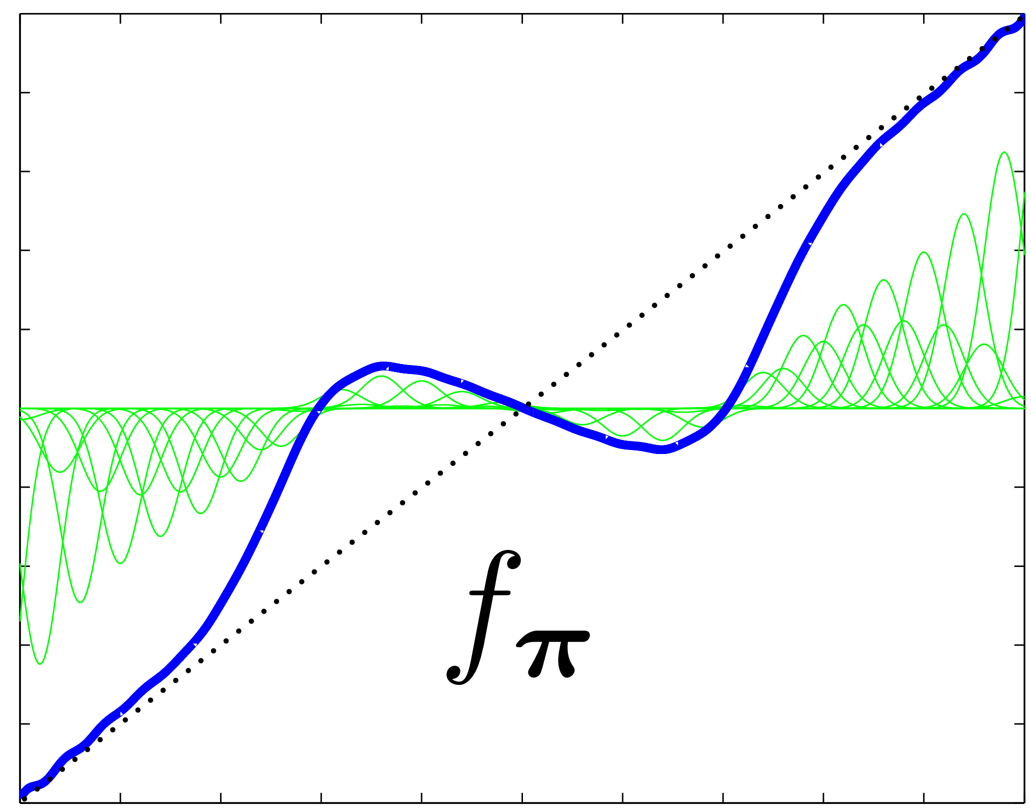
Shrinkage Fields: Details

TECHNISCHE UNIVERSITÄT DARMSTADT

- Replace potentials with shrinkage functions
  - **simplifies learning** (closed-form gradients)
  - modeled as mixture of Gaussian RBF kernels
  - not limited to monotonic functions

$$f_{\boldsymbol{\pi}}(v) = \sum_{j=1}^{M} \pi_j \exp\left(-\frac{\gamma}{2}(v-\mu_j)^2\right)$$





Shrinkage Fields: Details

TECHNISCHE UNIVERSITÄT DARMSTADT

Algorithm

$$\mathbf{x}^0 \leftarrow \mathbf{y}$$

$$\textbf{for } t \leftarrow 1 \textbf{ to } T \textbf{ do } \mathbf{x}^t \leftarrow g_{\mathbf{\Theta}_t}(\mathbf{x}^{t-1})$$

$$\text{with} \quad g_{\mathbf{\Theta}}(\mathbf{x}) = \mathcal{F}^{-1}\left[\frac{\mathcal{F}\left(\lambda \mathbf{K}^\mathsf{T}\mathbf{y} + \sum_{i=1}^N \mathbf{F}_i^\mathsf{T} f_{\boldsymbol{\pi}_i}(\mathbf{F}_i\mathbf{x})\right)}{\lambda \check{\mathbf{K}}^* \circ \check{\mathbf{K}} + \sum_{i=1}^N \check{\mathbf{F}}_i^* \circ \check{\mathbf{F}}_i}\right]$$

$$= \mathbf{\Omega}^{-1}\boldsymbol{\eta}$$

fast Gaussian CRF inference

Convolution
Fourier
Shrinkage

- Standard gradient-based training
  - **discriminative training** with loss function (here, PSNR)
  - **greedy** stage-by-stage **or joint training**

Related Work

- Discriminative wavelet shrinkage [Hel-Or & Shaked '08]
  - learned piece-wise linear shrinkage functions
  - no random field
- Active Random Field (ARF) [Barbu '09]
  - combine model and optimization, but learning cumbersome
  - only local inference via gradient descent
  - very fast runtime, but limited image quality
- Cascade of Regression Tree Fields (RTF) [Schmidt et al. '13]
  - cascade of Gaussian CRFs
  - conceptually motivated by half-quadratic inference
  - slower iterative inference without runtime guarantees

Experiments

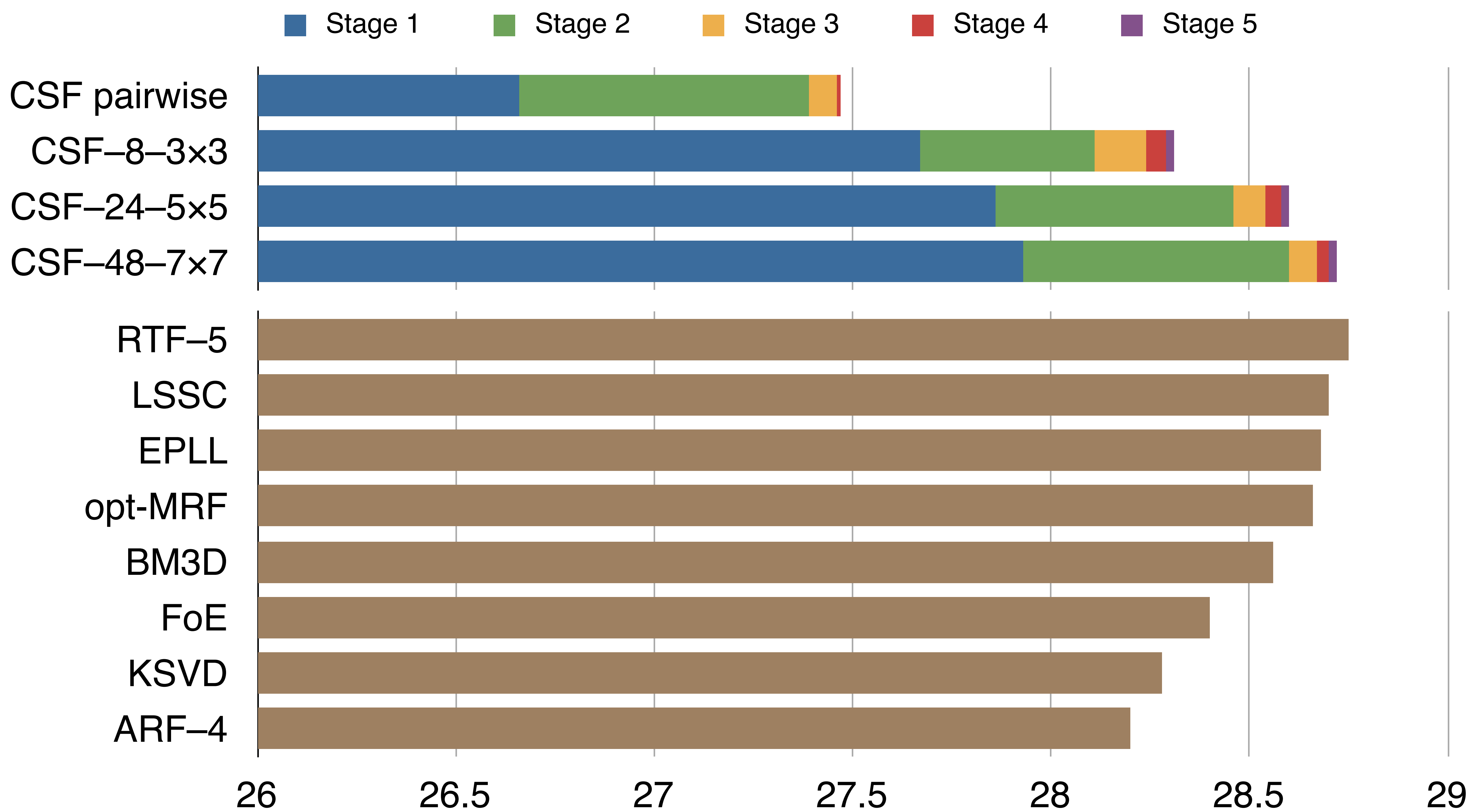
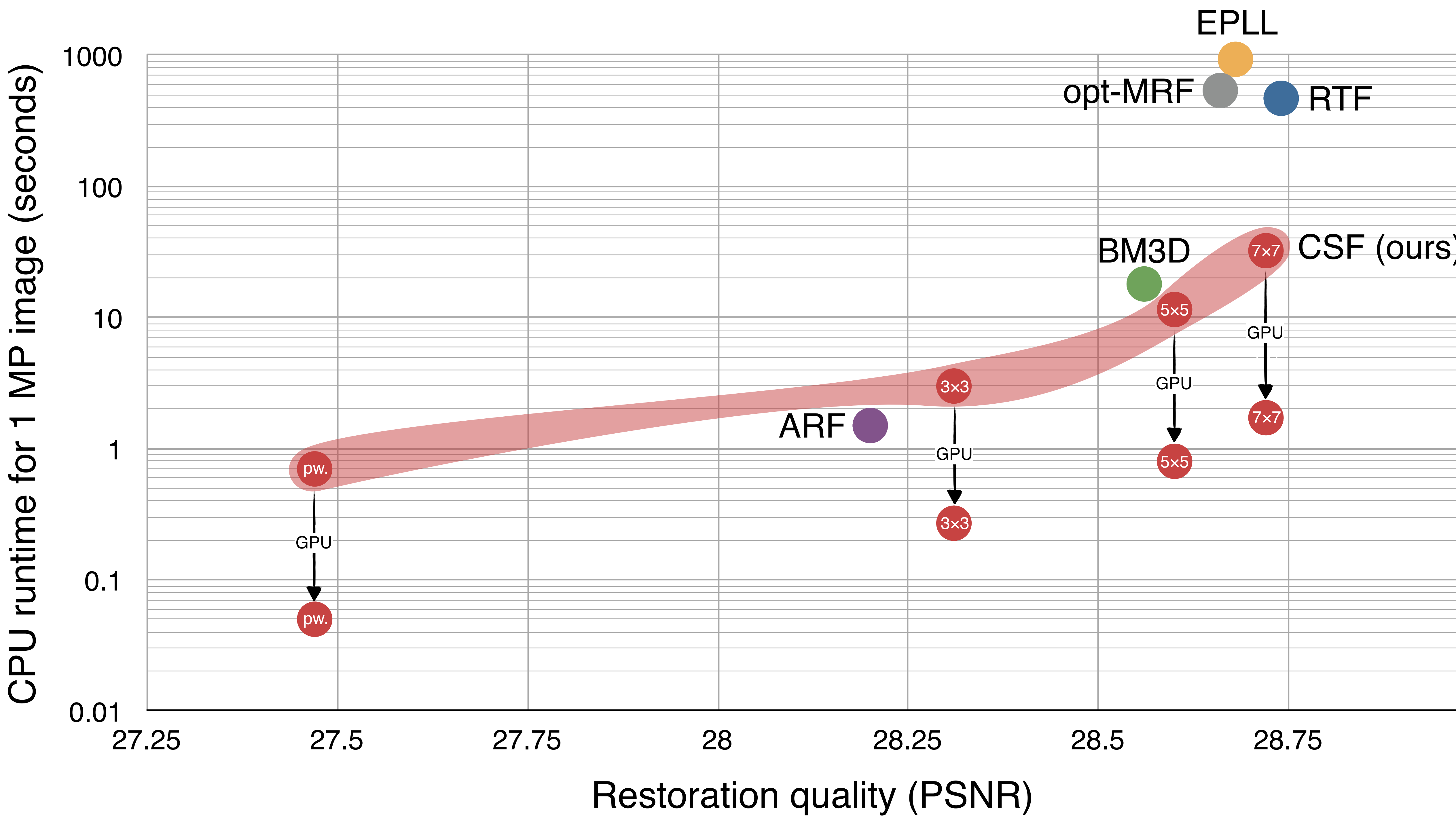


Image Denoising. Average PSNR (dB) on 68 test images ($\sigma = 25$)

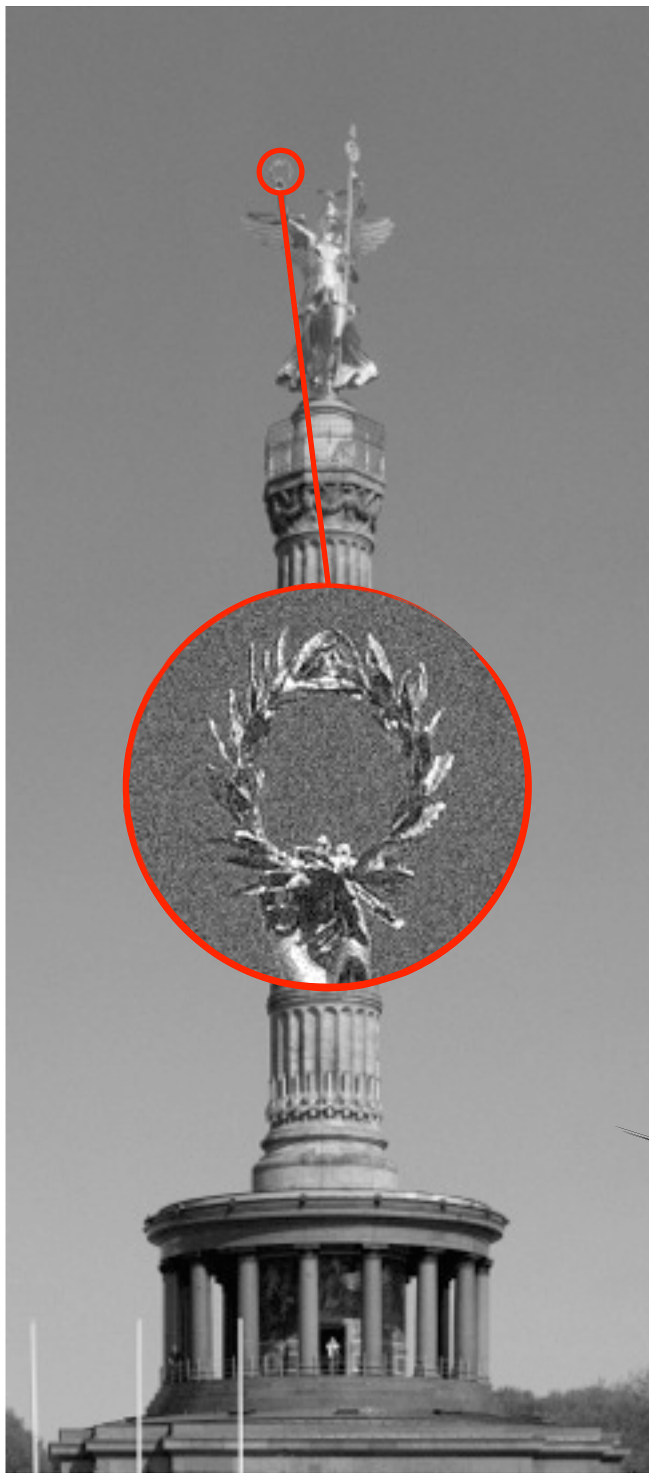
Restoration Quality vs. Runtime

TECHNISCHE UNIVERSITÄT DARMSTADT



Denoising a 16 Megapixel Image

TECHNISCHE UNIVERSITÄT DARMSTADT



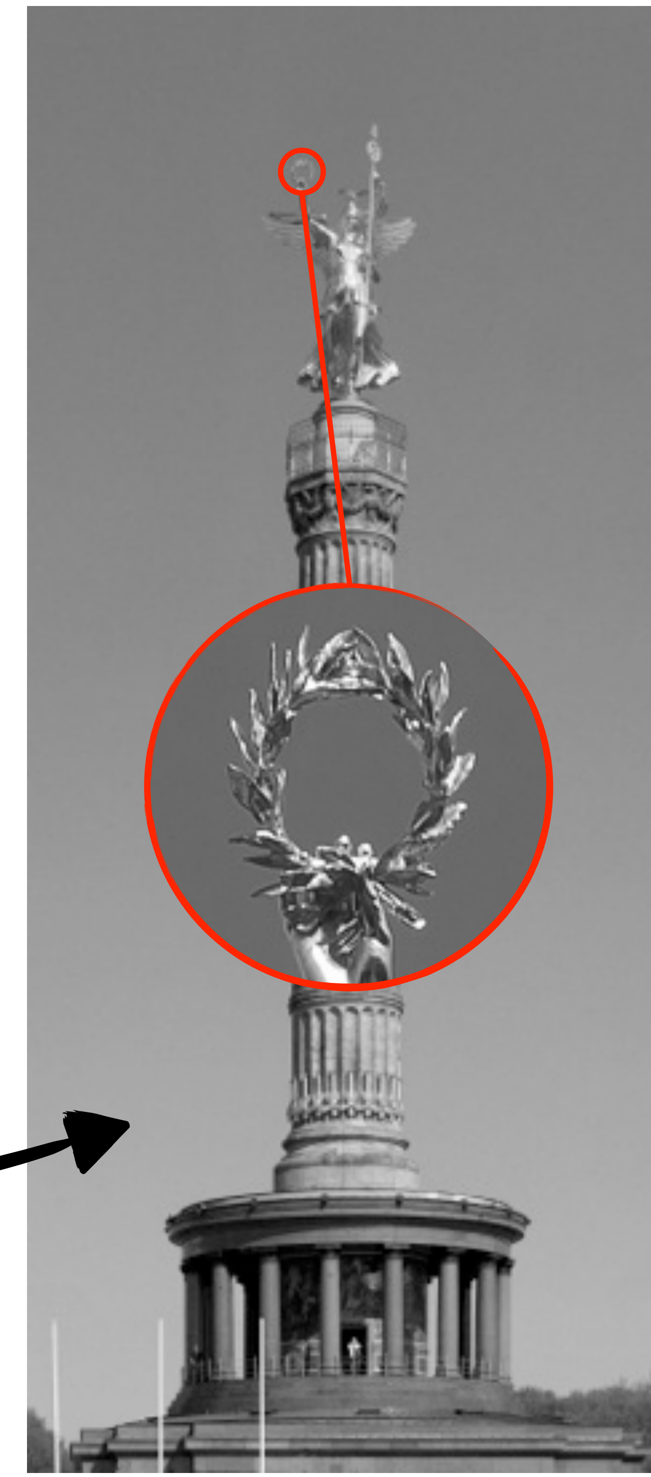
Model:

CSF–48–7×7 (4 stages)
(our best model)

Runtimes (Matlab code):

~ 3 min
(CPU, multithreaded)

~ **42 sec**
(GPU)



BY-SA Thomas Wolf, www.foto-tw.de

Denoising a 16 Megapixel Image

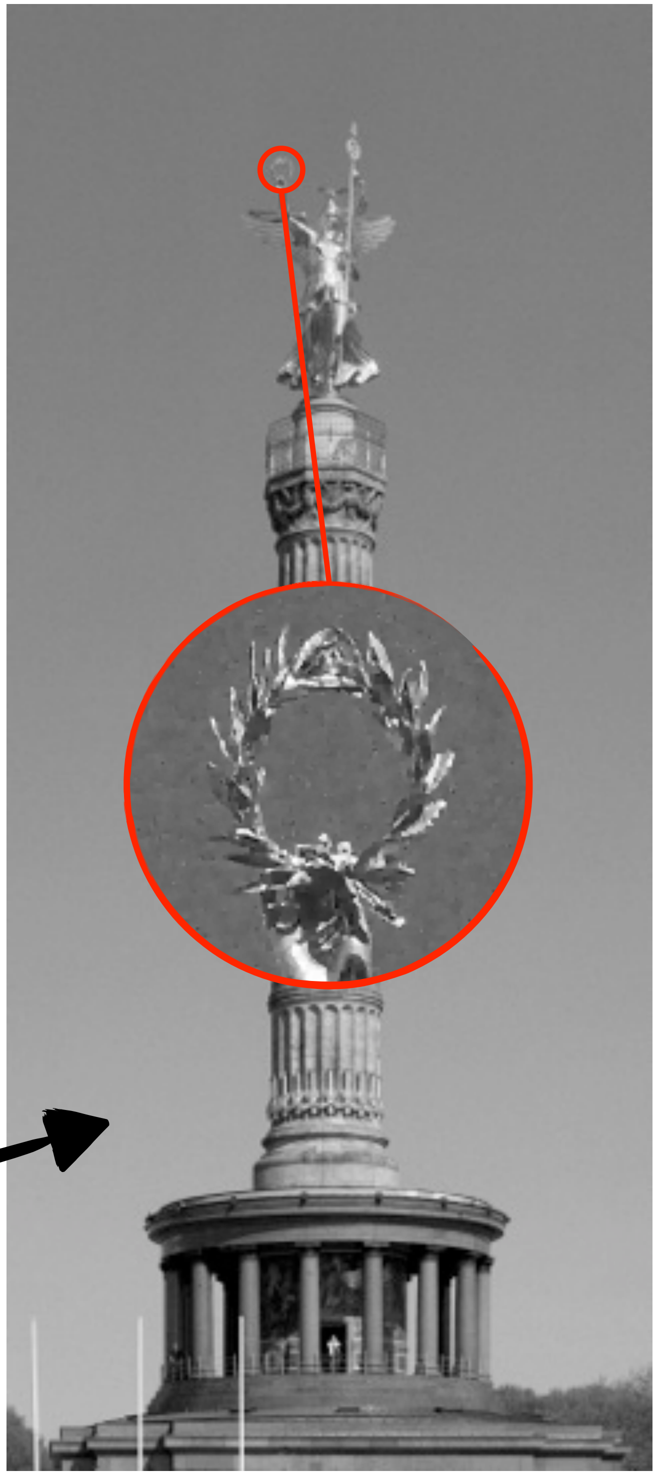
Model:

CSF pairwise (4 stages)
(our fastest model)

Runtimes (Matlab code):

~ 10.5 sec
(CPU, multithreaded)

~ **1.5 sec**
(GPU)



Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Integrated** random-field **model and optimization** approach
  - extends the additive form of half-quadratic inference
- Uses **shrinkage** functions **instead of potentials**
  - increases model flexibility
  - enables easy learning with standard gradient-based methods
- **Scalable** to megapixel images
- **High restoration quality** through **loss-based training**
  - learning of model and optimization parameters
- MATLAB **code available**: `http://goo.gl/w6Z4mm`