A Generative Perspective on MRFs in Low-Level Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

Uwe Schmidt

Qi Gao

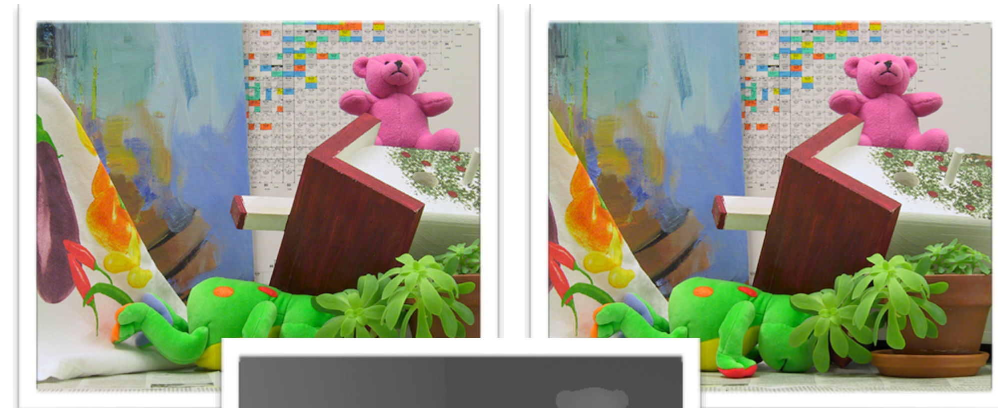
Stefan Roth

Department of Computer Science
TU Darmstadt

Low-Level Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

Super-Resolution

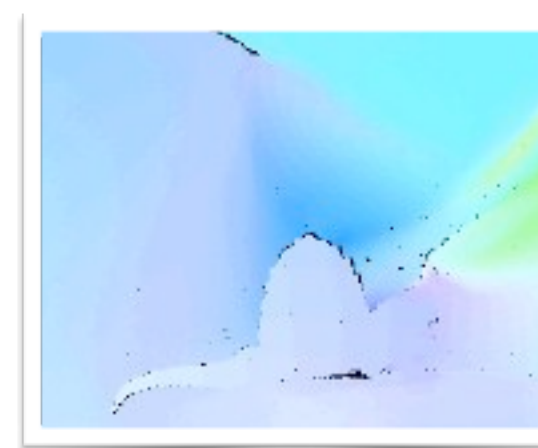


Stereo

Discriminative
✔ performance
✘ versatility

Generative
✔ versatility
✘ learning

Image Restoration

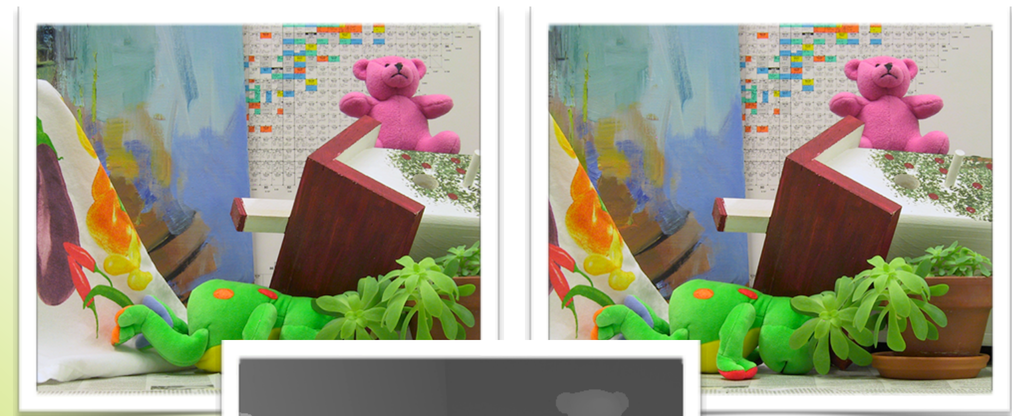


Optical Flow

Low-Level Vision

TECHNISCHE UNIVERSITÄT DARMSTADT

Super-Resolution



Stereo

Generative
MRF
Priors

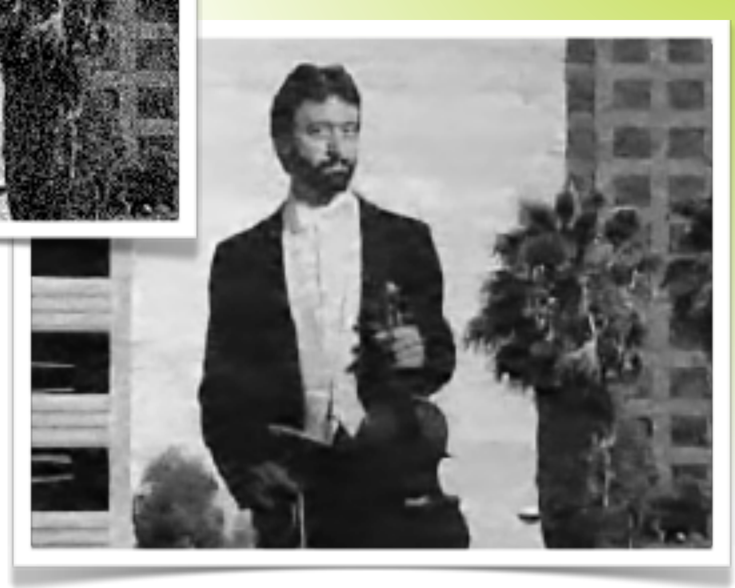
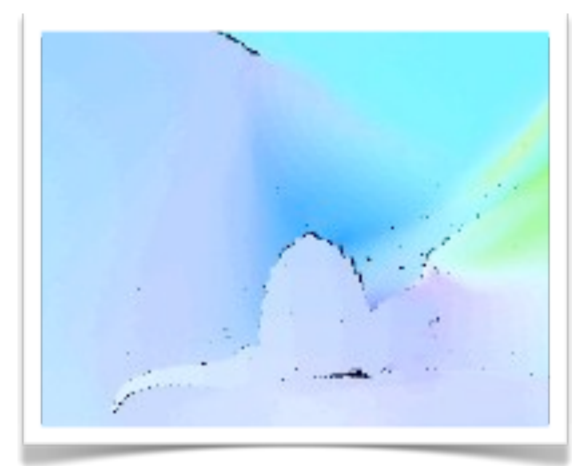


Image Restoration



Optical Flow

Common MRF Evaluation

MRF prior $p(\mathbf{x})$ • Application likelihood $p(\mathbf{y}|\mathbf{x})$ → Posterior $p(\mathbf{x}|\mathbf{y})$

$\mathbf{y}$

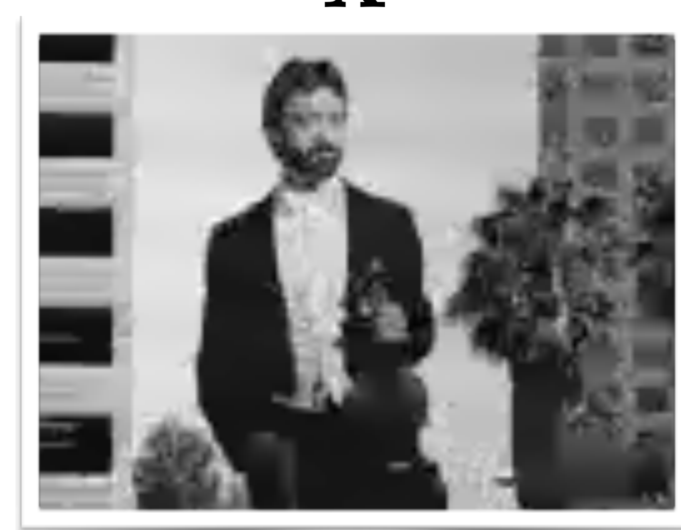
generate/ measure

Ground truth

Indirect model evaluation

MAP estimation (Gradient methods, Graph cuts, ...)

$\hat{\mathbf{x}}$



MAP estimate

Compare

→ Measure PSNR, ...

Desirable MRF Evaluation

- Purpose of MRF priors
  - Model statistical properties of natural images and scenes

➡ Evaluate generative properties [Zhu & Mumford '97]
  - e.g. derivative statistics of the model
  - neglected ever since

Difficult!

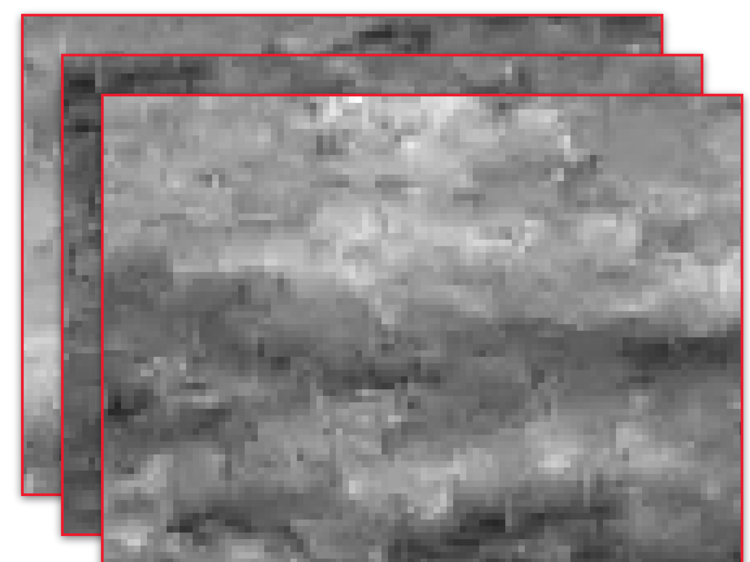
MRF prior

Draw samples (MCMC)

Data

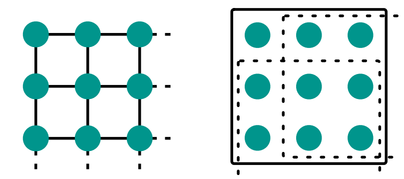
Compare

statistical properties



MRF samples

Agenda

1. Evaluate generative properties of common image priors
   - Pairwise & high-order MRFs 
   - Based on a flexible MRF framework with an efficient sampler

2. Learn improved generative models

3. Find that in the context of MAP estimation our models do not perform as well as expected for image denoising

4. Address this problem (and others) by changing the estimator

Flexible MRF Model

- Fields-of-Experts (FoE) framework [Roth & Black ’05, ’09]
  - Subsumes popular pairwise & high-order MRFs

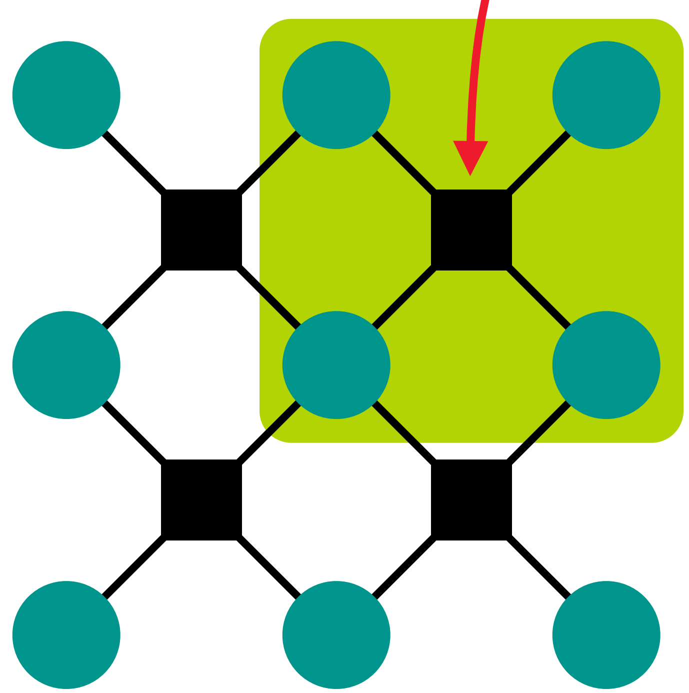
Expert function

$$p(\mathbf{x};\boldsymbol{\Theta}) = \frac{1}{Z(\boldsymbol{\Theta})} e^{-\epsilon\|\mathbf{x}\|^2/2} \prod_{c\in\mathcal{C}} \prod_{i=1}^{N} \phi(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(c)}; \boldsymbol{\alpha}_i)$$

Image

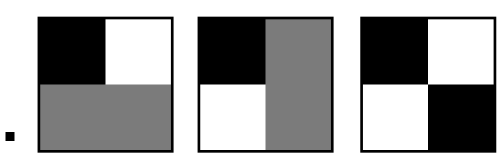
Parameters

$$\boldsymbol{\Theta} = \{\mathbf{J}_i, \boldsymbol{\alpha}_i\}_{i=1,\ldots,N}$$



Linear filter

e.g.



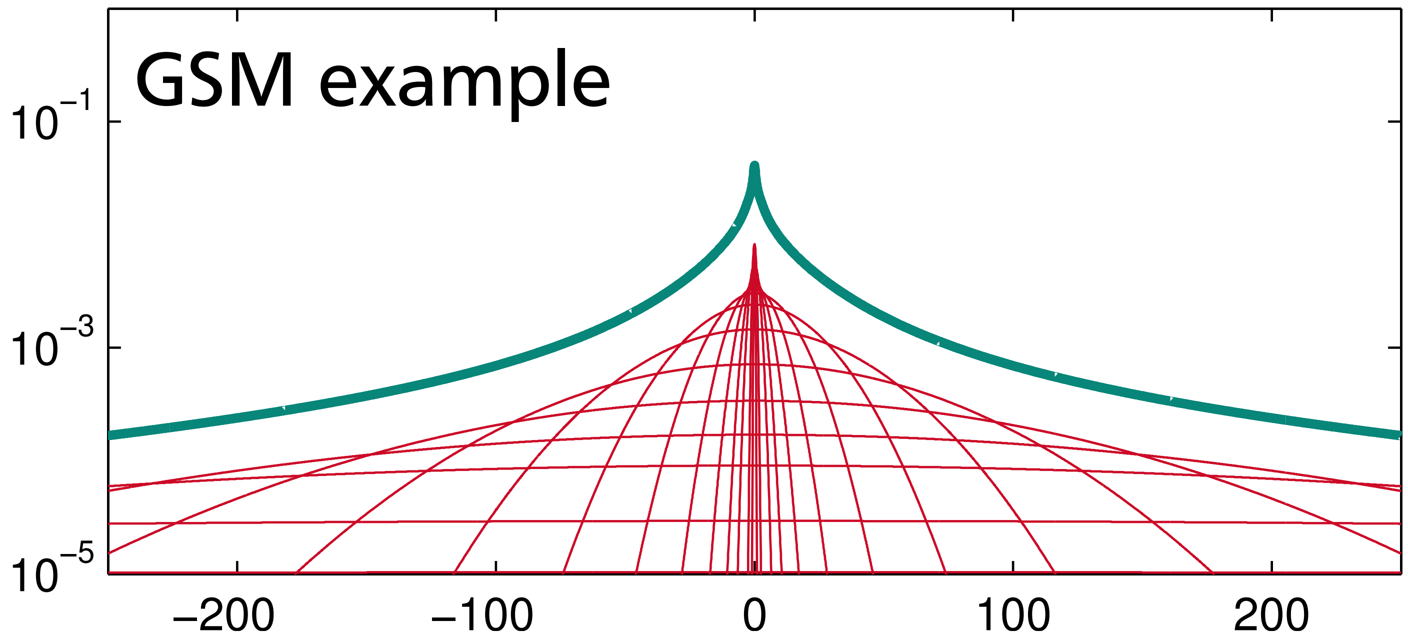
Vector of nodes in clique c

Flexible MRF Model

- Fields-of-Experts (FoE) framework [Roth & Black '05, '09]
  - Subsumes popular pairwise & high-order MRFs

Mixture Weights

$$p(\mathbf{x};\boldsymbol{\Theta}) = \frac{1}{Z(\boldsymbol{\Theta})} e^{-\epsilon\|\mathbf{x}\|^2/2} \prod_{c \in \mathcal{C}} \prod_{i=1}^{N} \phi(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(c)}; \boldsymbol{\alpha}_i)$$



Gaussian Scale Mixture (GSM) Distribution

[Wainwright & Simoncelli '99, Weiss & Freeman '07]

$$\phi(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(c)}; \boldsymbol{\alpha}_i) = \sum_{j=1}^{J} \alpha_{ij} \cdot \mathcal{N}(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(c)}; 0, \sigma_i^2 / s_j)$$

Sampling from the MRF

- Obtain joint distribution:
  - Product of GSMs = GSM
  - Augment MRF with auxiliary variables **z** for the mixture components and do not marginalize them out

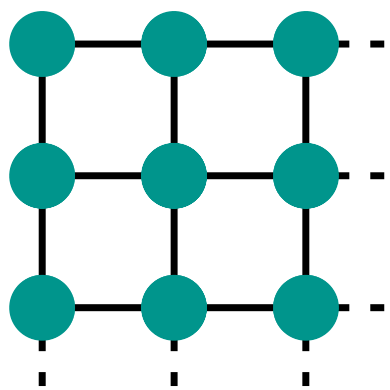
Mixture Weights

Gaussian

$$\sum_{\mathbf{z}} \underbrace{p(\mathbf{z})\, p(\mathbf{x}|\mathbf{z})}_{p(\mathbf{x}, \mathbf{z})}$$

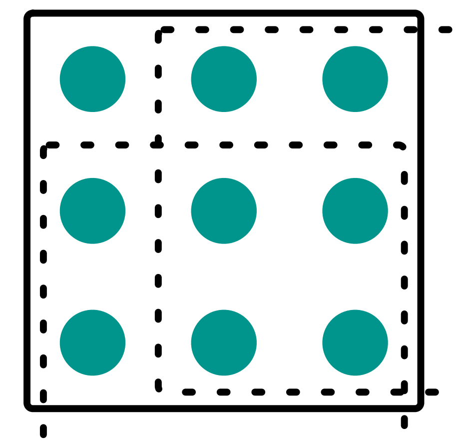
- Gibbs sampling from the joint distribution $p(\mathbf{x}, \mathbf{z}; \boldsymbol{\Theta})$ [Geman & Yang ’95; Welling et al. ’02]
  - Alternate block sampling from $p(\mathbf{x}|\mathbf{z}; \boldsymbol{\Theta})$ and $p(\mathbf{z}|\mathbf{x}; \boldsymbol{\Theta})$
  - The $\mathbf{z}$ can be discarded in the end
  - Least-squares method for sampling $p(\mathbf{x}|\mathbf{z}; \boldsymbol{\Theta})$ [Weiss ’05, Levi ’09]

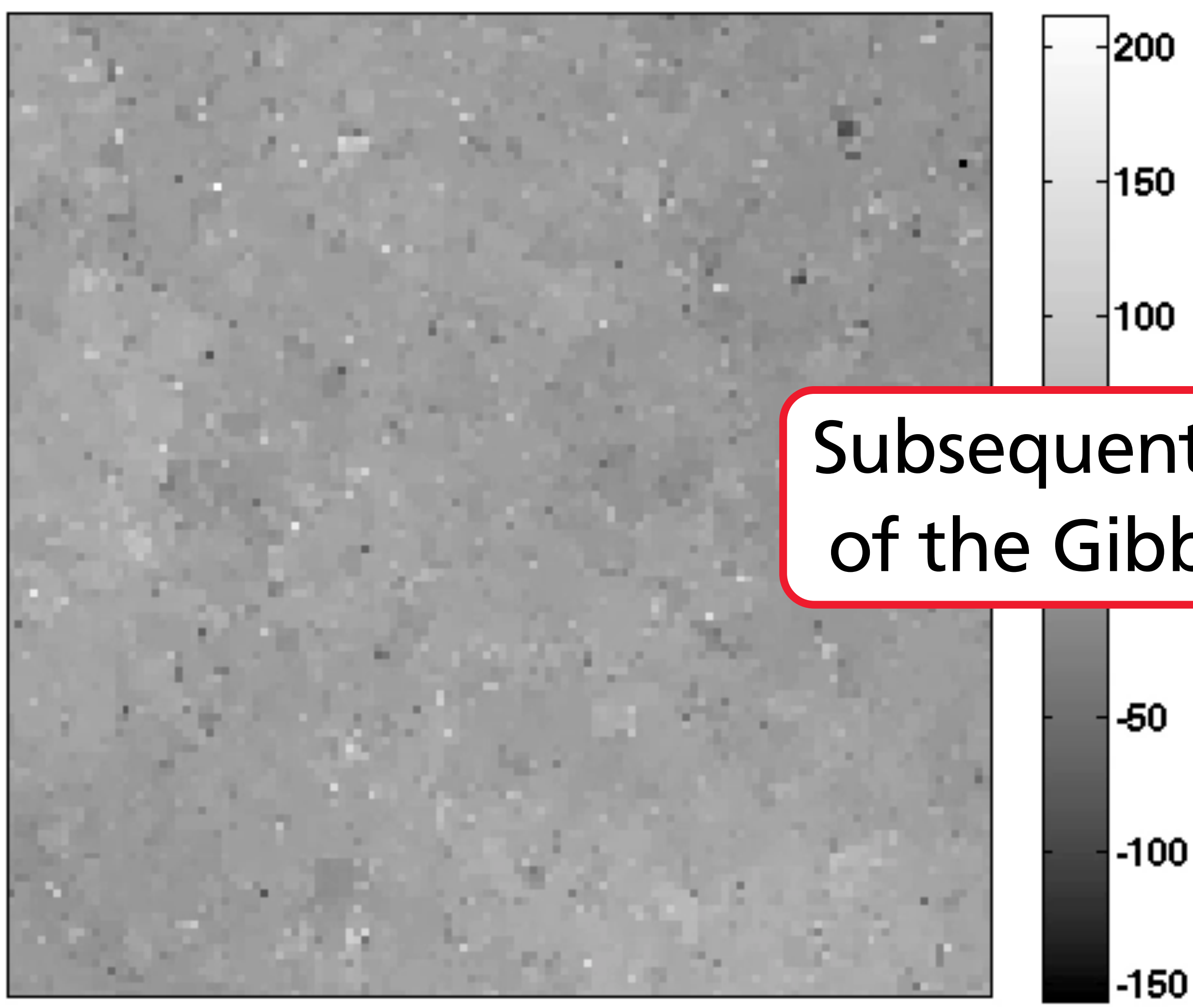
MRF Sampling – Example



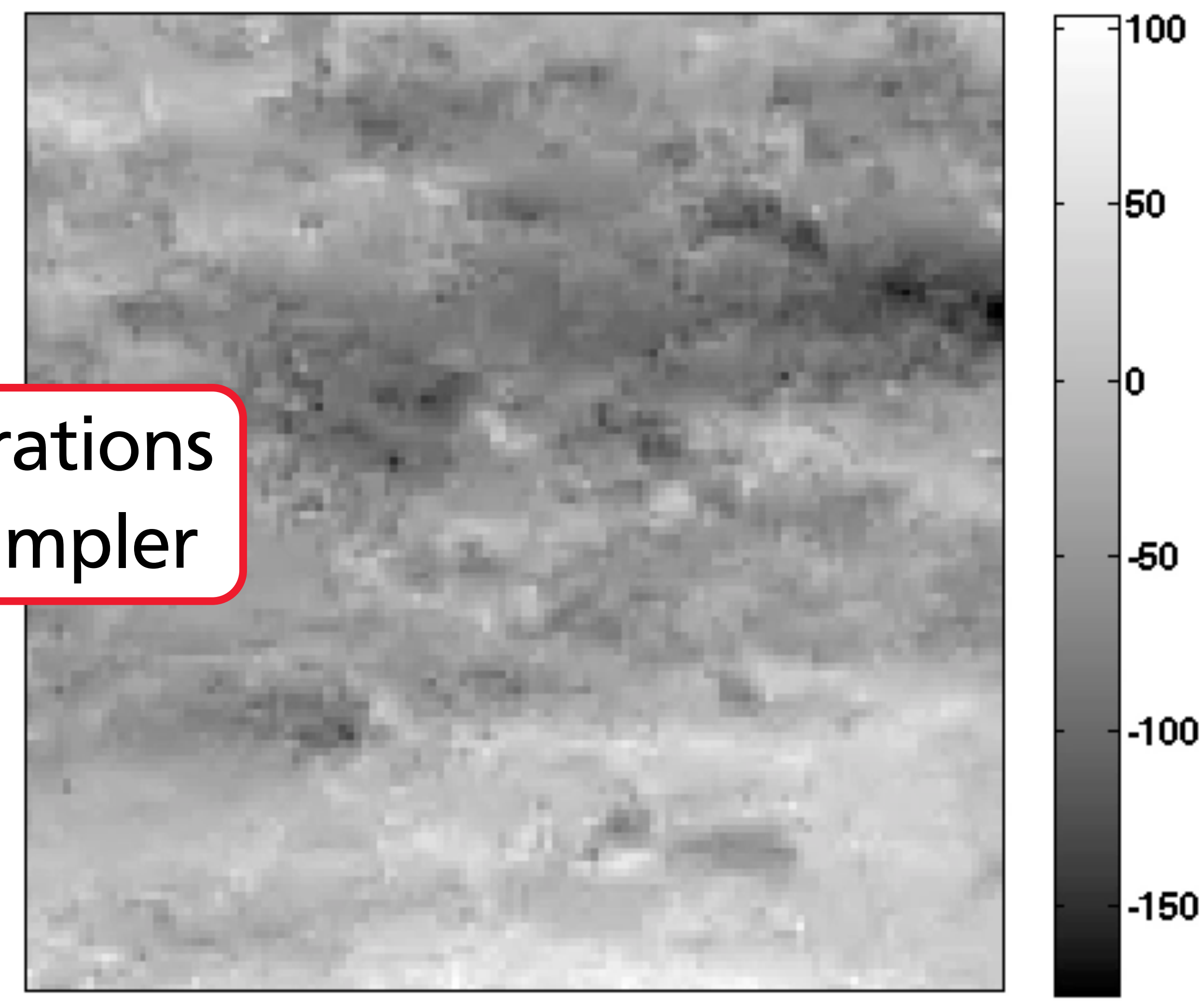
Pairwise MRF

High-order MRF with 3×3 cliques



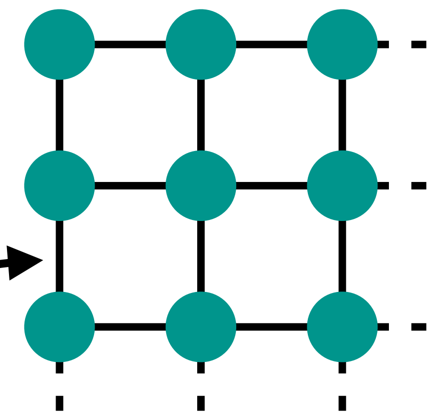


Subsequent iterations of the Gibbs sampler

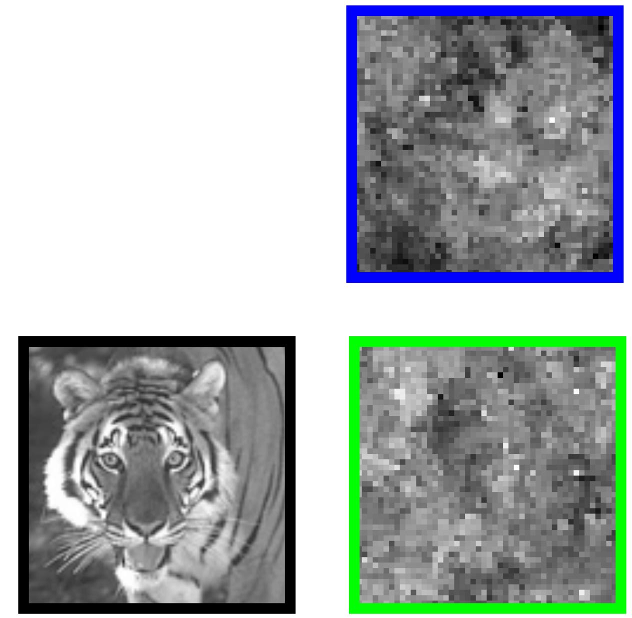


Generative Properties of Pairwise MRFs

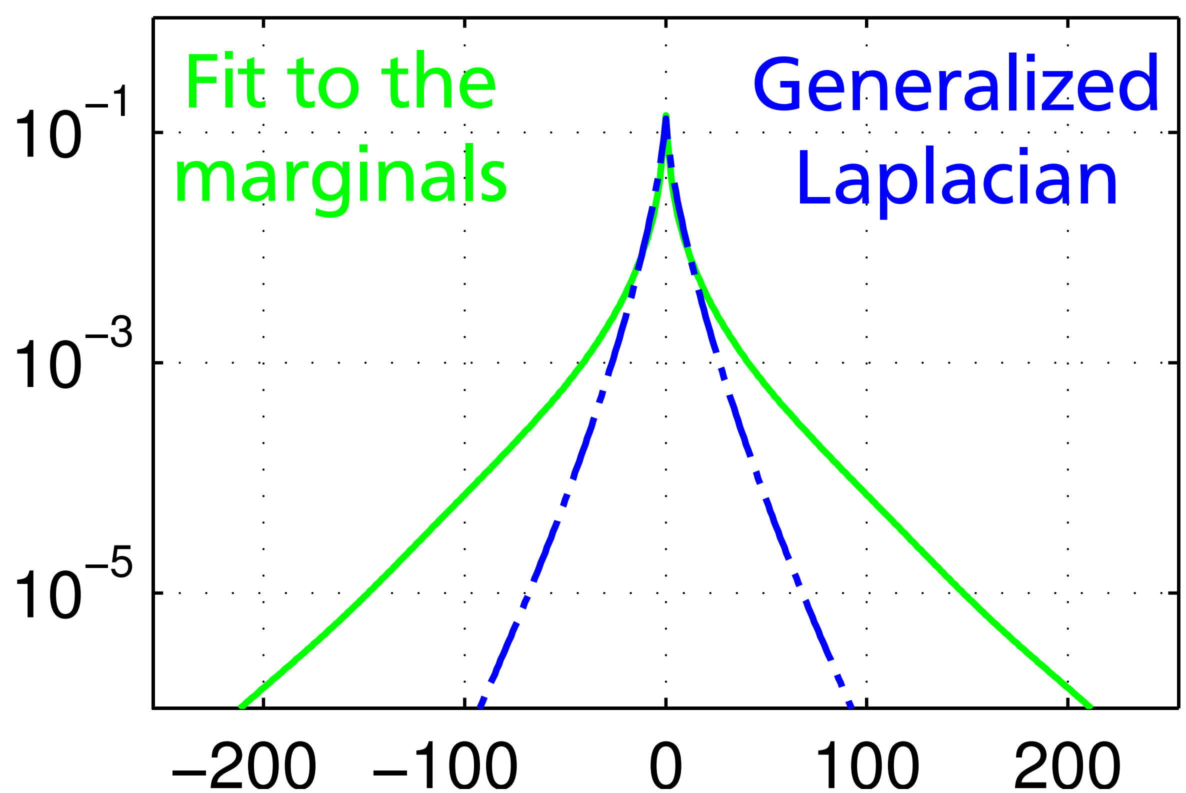
- Consider simplest pairwise MRFs



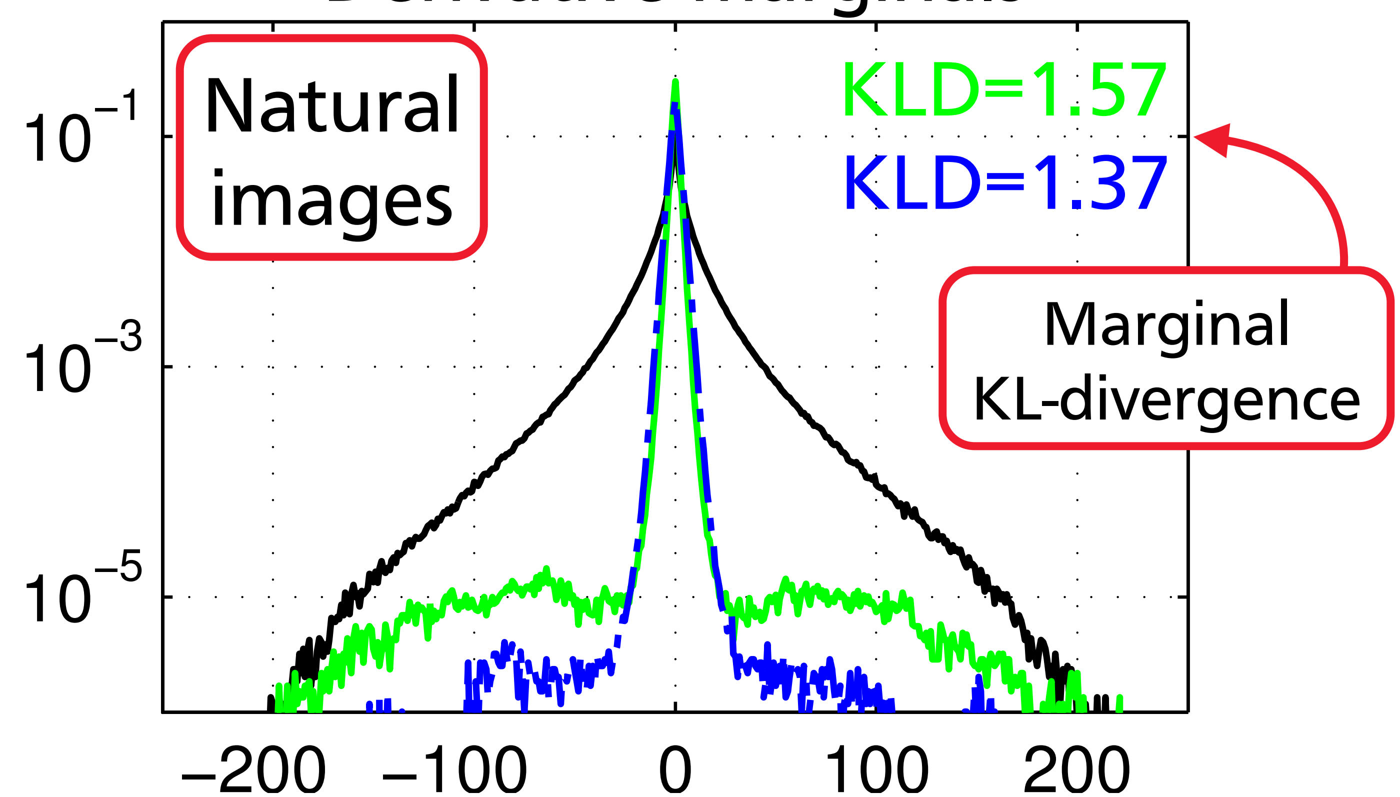
MRF



Potential function



Derivative marginals



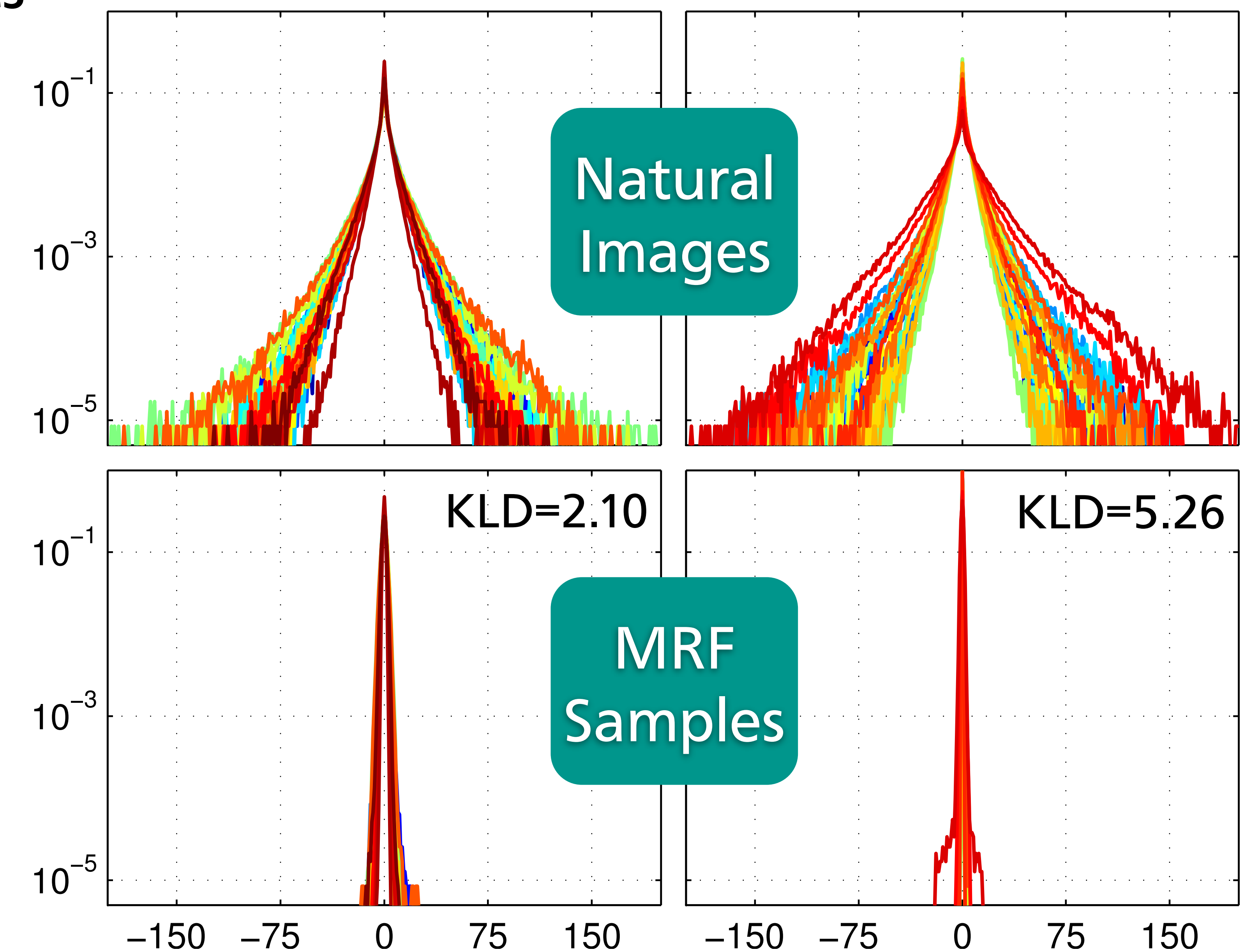
Generative Properties of High-order MRFs

- Common FoE models
  - Evaluate filter statistics of model filters $\mathbf{J}_i$
- Apparent contradiction:
  - ✘ Poor generative properties
  - ✔ Good application performance

Why?

[Roth & Black '09]
24 5×5 filters
Student-t experts

[Weiss & Freeman '07]
25 15×15 filters
fixed GSM experts

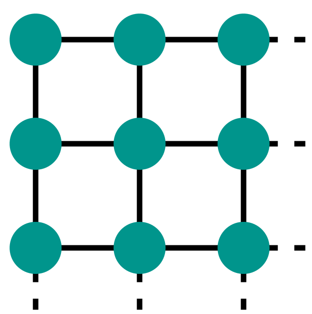


Learning Better Generative MRFs

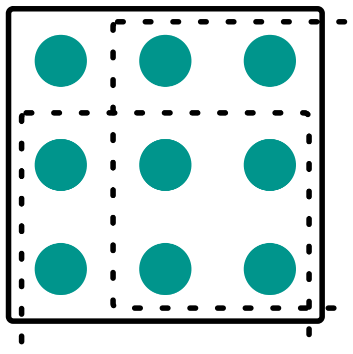
- Learn shapes of flexible GSM experts and linear filters $\mathbf{J}_i$ (for high-order model)
  - Use efficient sampler
  - Otherwise training similar to [Roth & Black '09]

- Learned models:

1. Pairwise MRF with single GSM potential (fixed first-derivative filters)



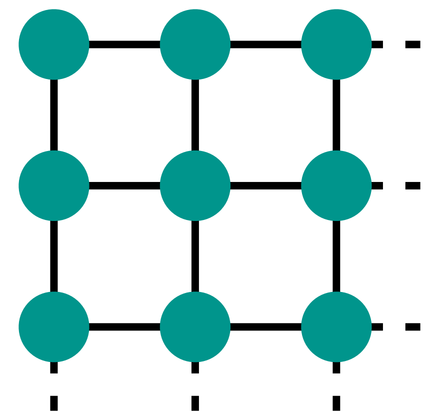
2. FoE with 3×3 cliques and 8 GSM experts (including filters)

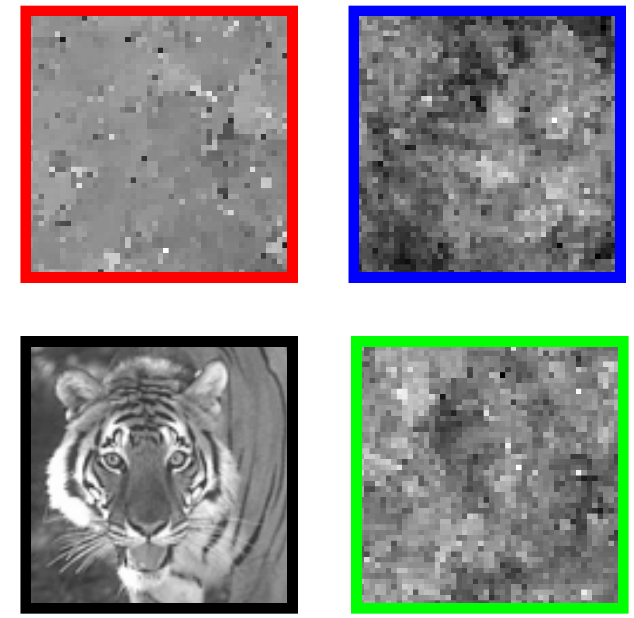


Generative Properties of Our Pairwise MRF

TECHNISCHE UNIVERSITÄT DARMSTADT

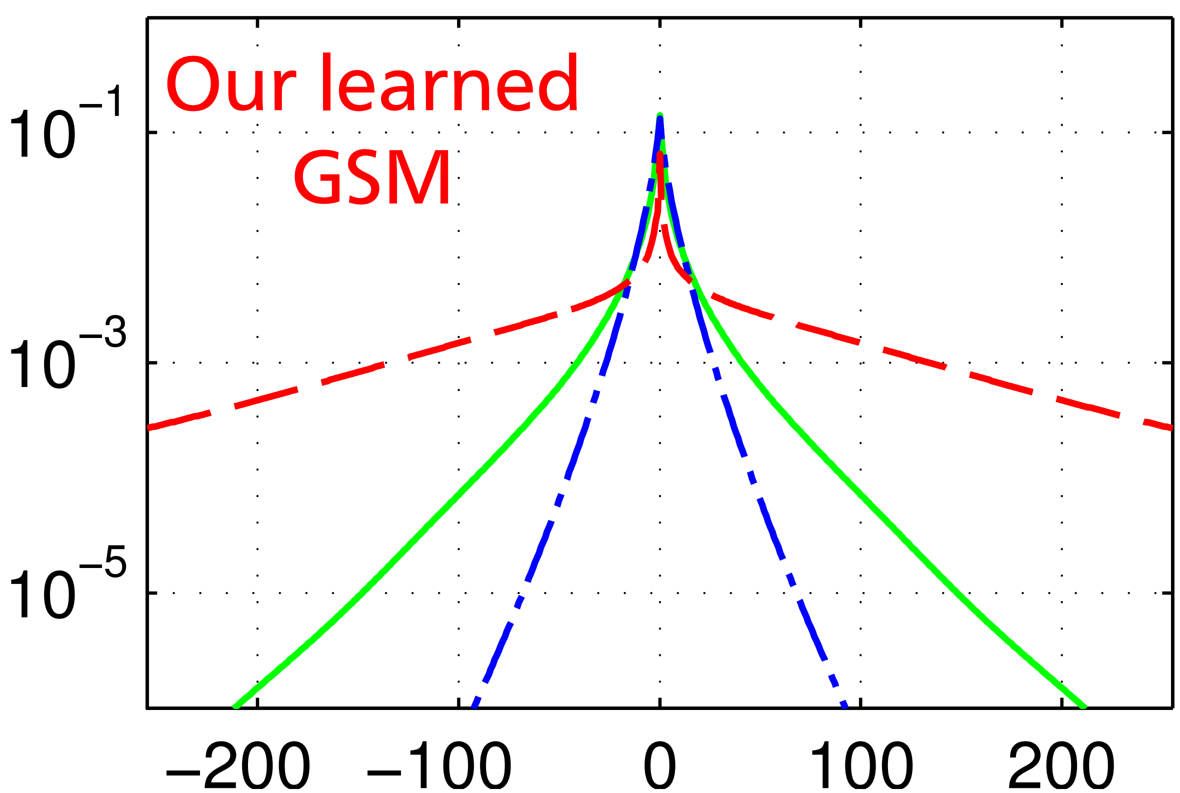
- Our pairwise MRF compared to previously shown



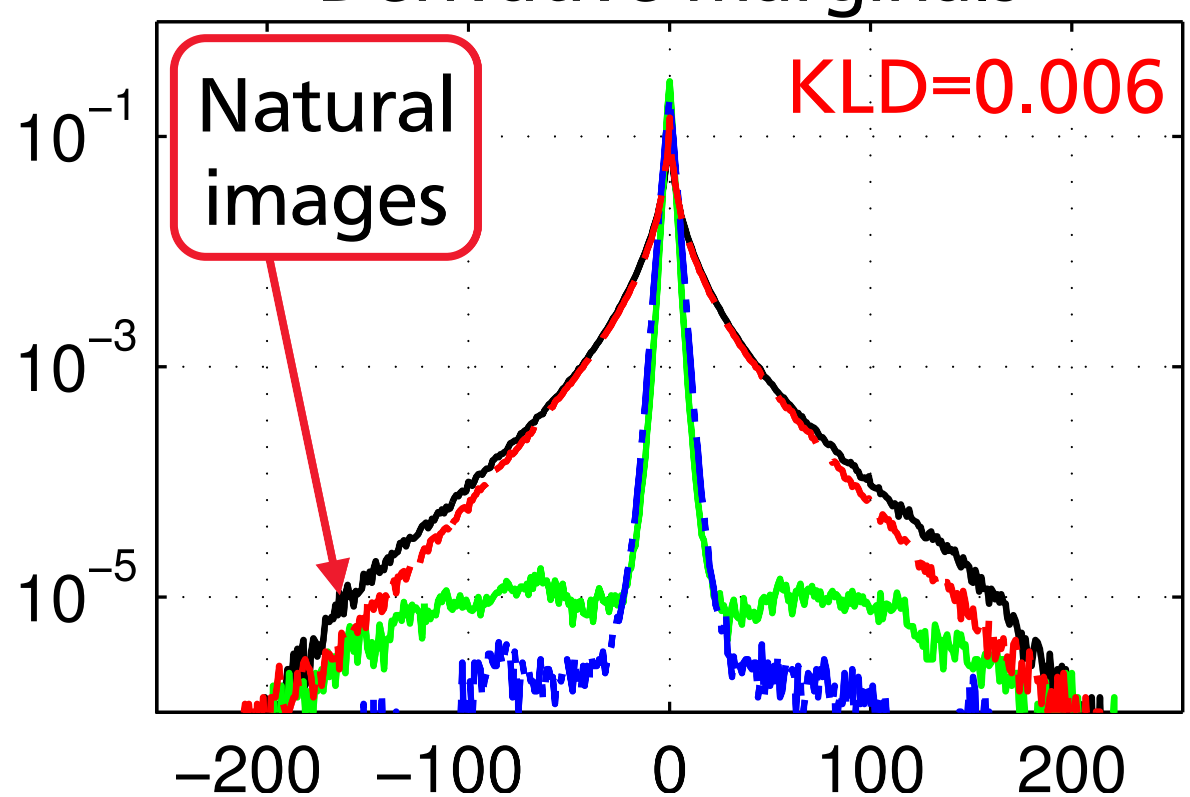


MRF

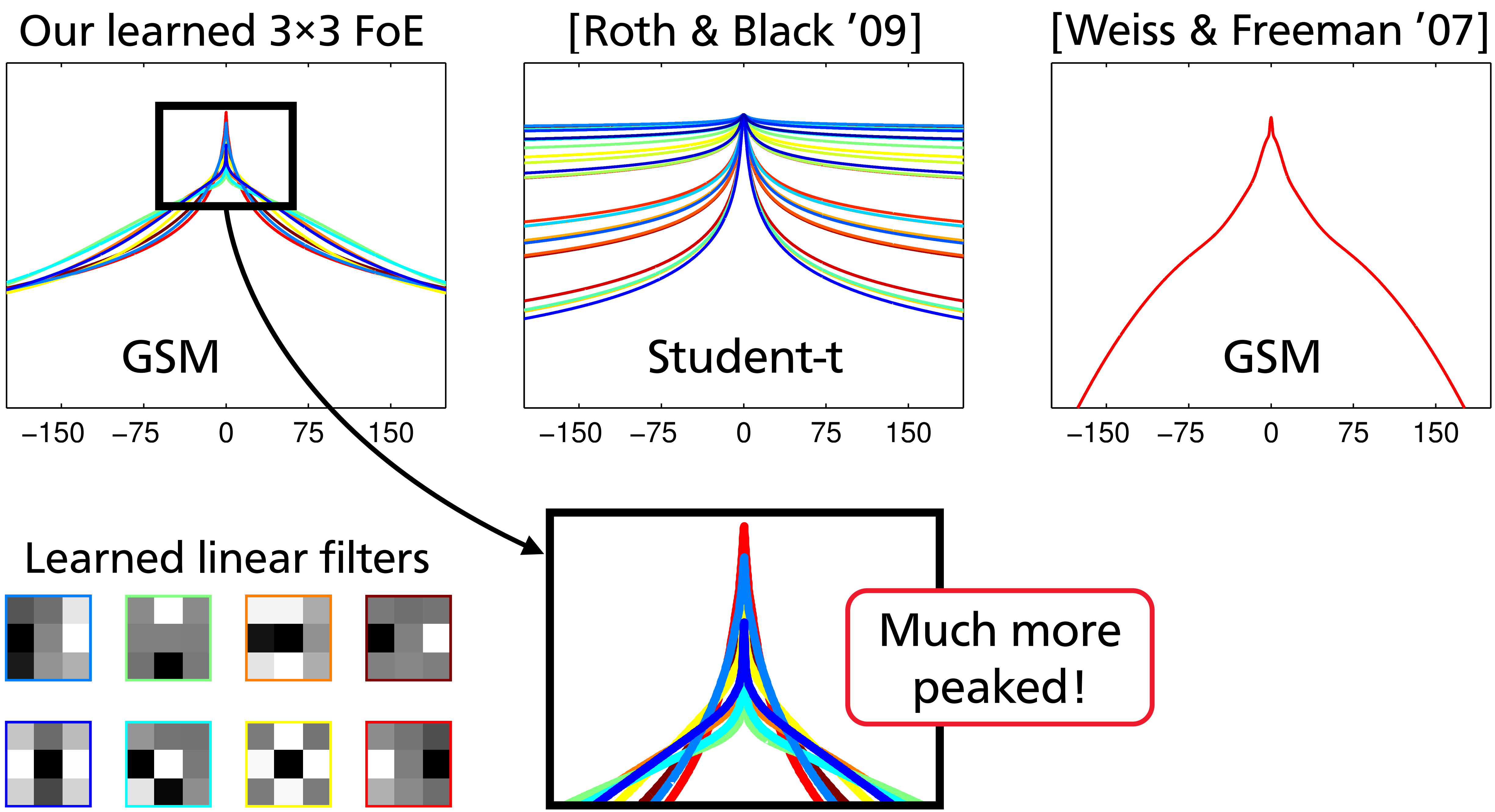
Potential function



Derivative marginals



Our Learned FoE in Comparison



Generative Properties of our FoE

- Filter statistics of our learned 3×3 FoE
  - Much better than previous models
  - Room for improvement

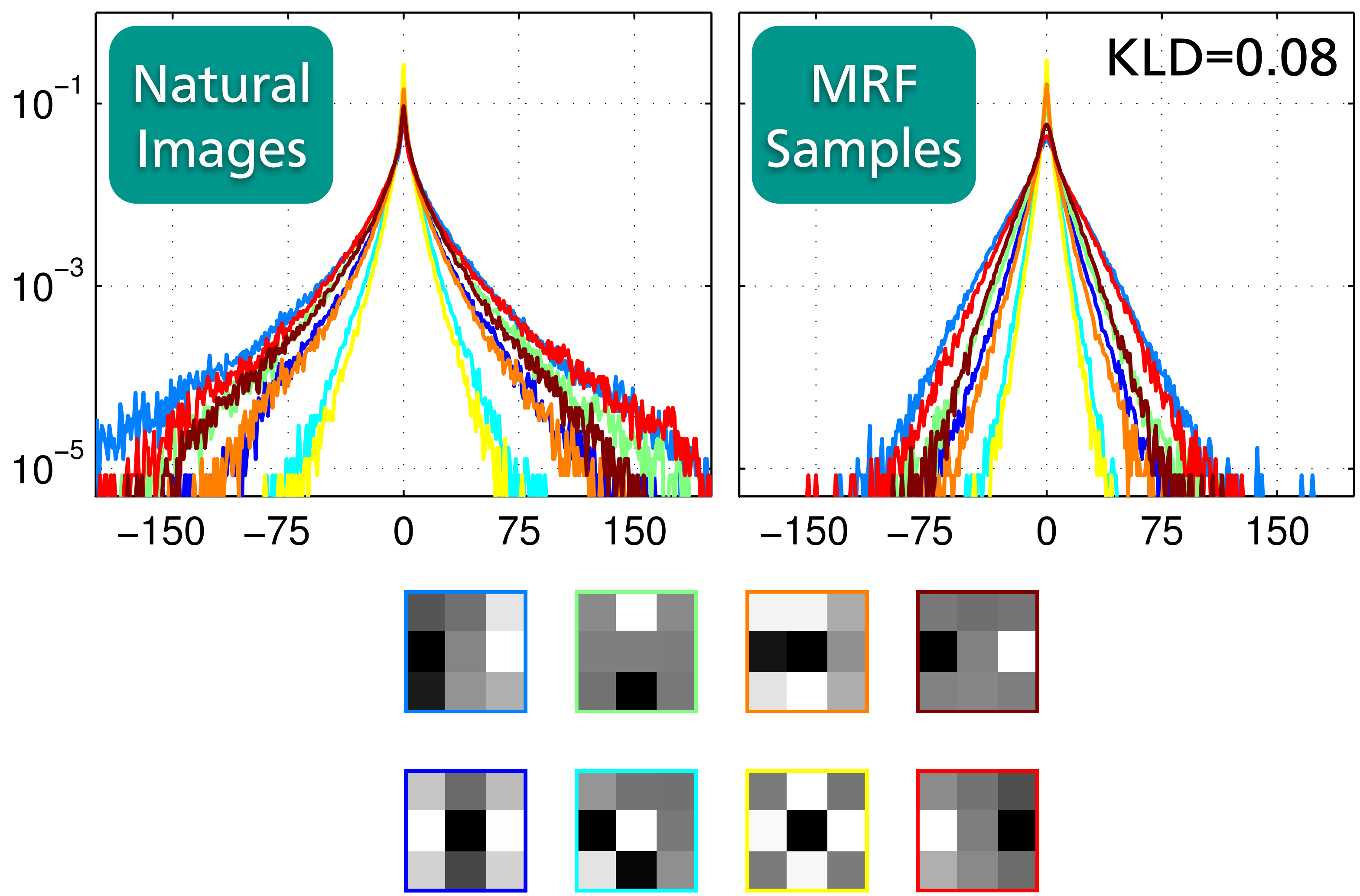
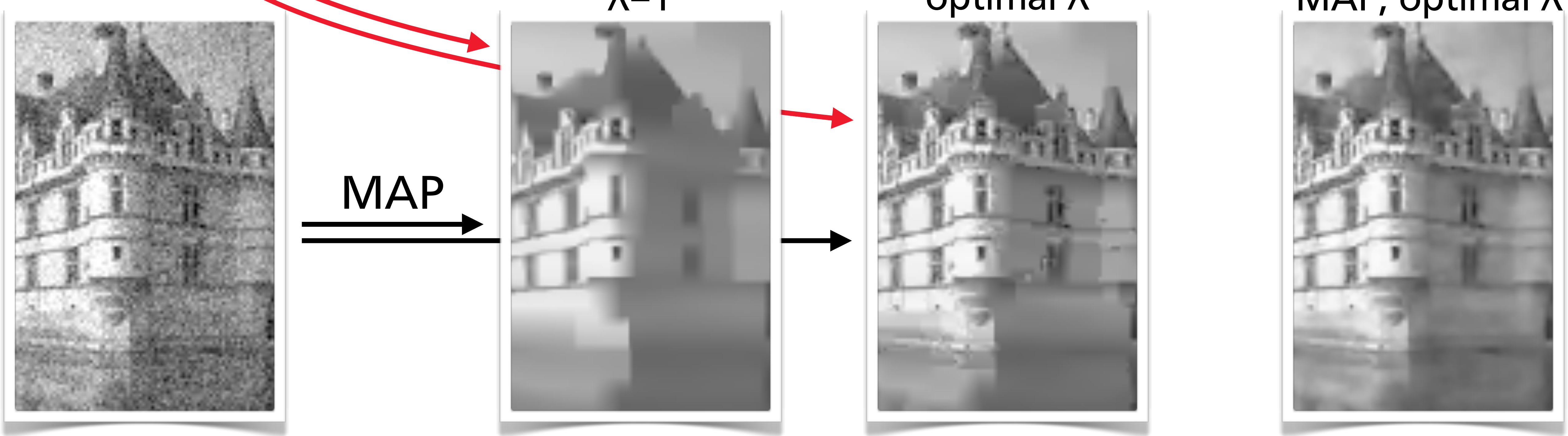


Image Denoising

- Image denoising assuming i.i.d. Gaussian noise with known standard deviation σ

Gaussian Likelihood | Our 3×3 FoE | Regularization weight

$$p(\mathbf{x}|\mathbf{y};\boldsymbol{\Theta}) \propto \underbrace{\mathcal{N}\left(\mathbf{y};\mathbf{x},\sigma^2\mathbf{I}\right)}_{\text{Gaussian Likelihood}} \cdot \underbrace{p(\mathbf{x};\boldsymbol{\Theta})}_{\text{Our 3×3 FoE}}{}^{\lambda}$$



λ=1 | optimal λ | [Roth & Black ’09] MAP, optimal λ

PSNR=22.18dB | PSNR=26.64dB | PSNR=29.18dB | PSNR=30.06dB

Image Denoising – MAP

- Recent works point to deficiencies of MAP [Nikolova '07, Woodford et al. '09]

- We find only modest correlation between:
  - Image quality of the MAP estimate
  - Generative quality of the MRF

Better generative properties $\nleftrightarrow$ Better application performance

Image Denoising – MMSE

- We propose to use Bayesian minimum mean squared error estimation (MMSE)

$$\hat{\mathbf{x}} = \arg\min_{\tilde{\mathbf{x}}} \int ||\tilde{\mathbf{x}} - \mathbf{x}||^2 p(\mathbf{x}|\mathbf{y};\boldsymbol{\Theta})\, \mathrm{d}\mathbf{x} = E[\mathbf{x}|\mathbf{y}]$$

- [Levi '09] extended sampler to the posterior
  - Only used a single sample in applications
- We approximate the MMSE estimate
  - Average samples from the posterior
- We find high correlation between:
  - Image quality of the MMSE estimate
  - Generative quality of the MRF

Samples

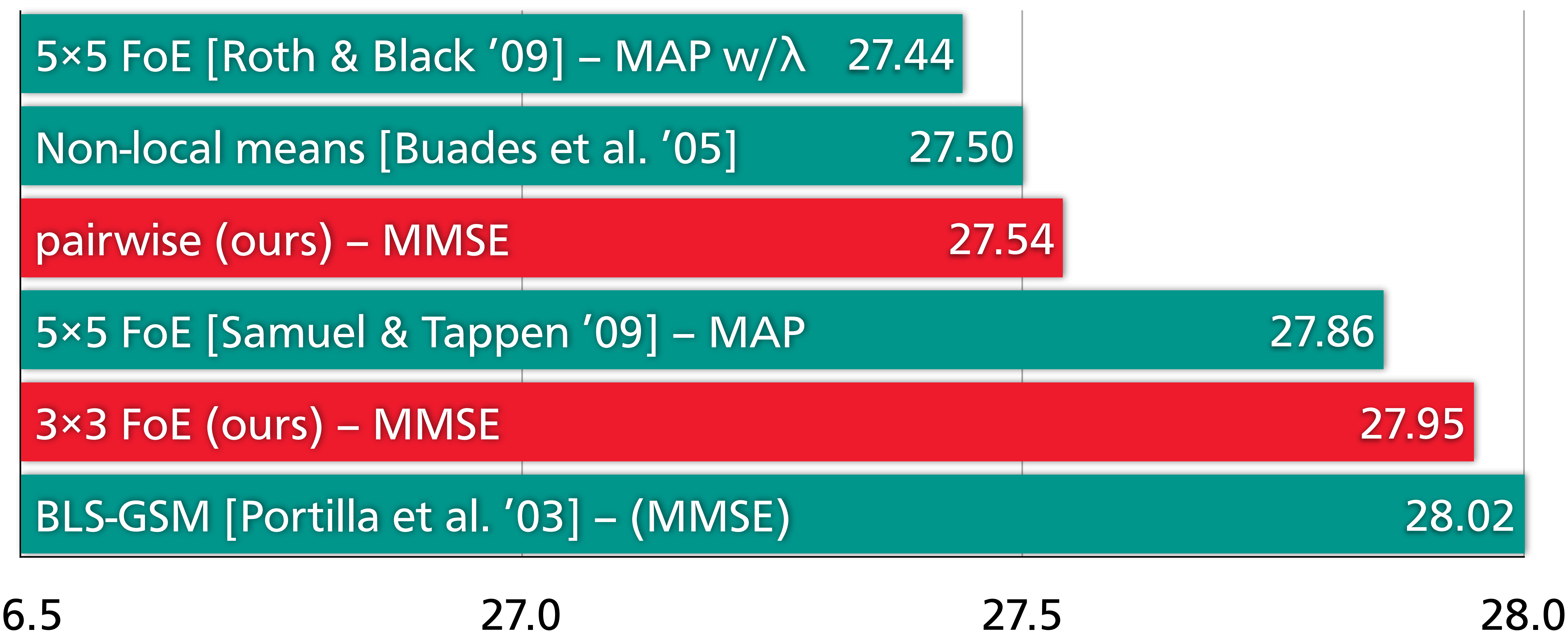
average

MMSE

Image Denoising – Results

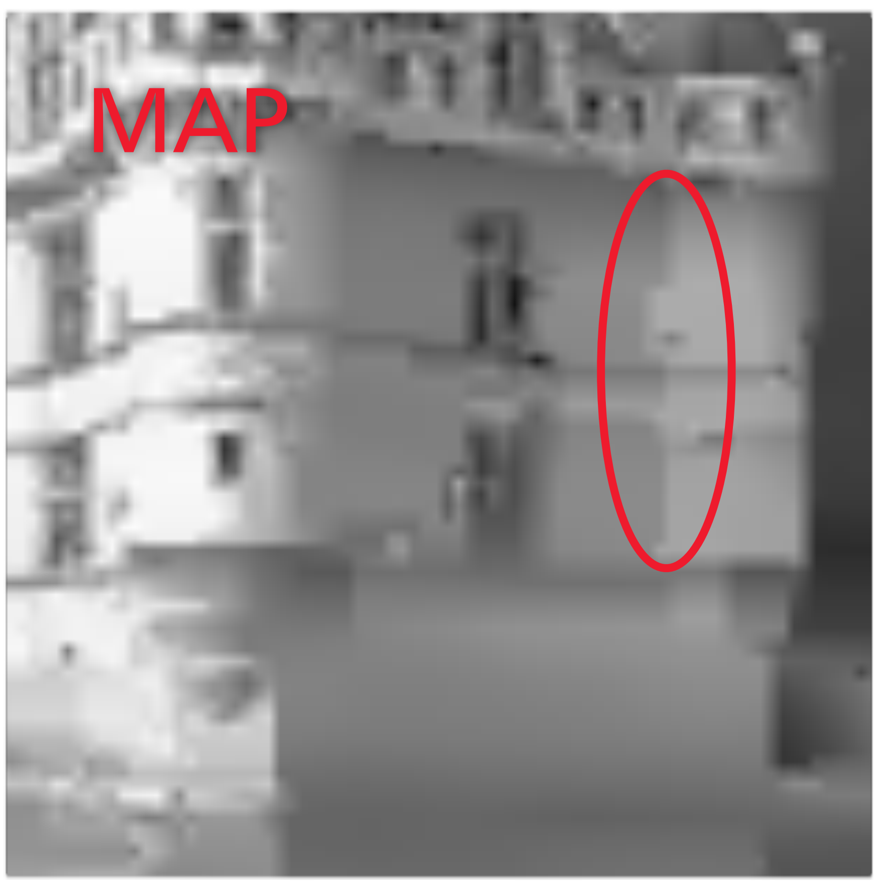
- Compared the MMSE estimate for our learned models with other popular methods

Average PSNR (dB) for 68 test images (σ = 25)



Advantages of the MMSE

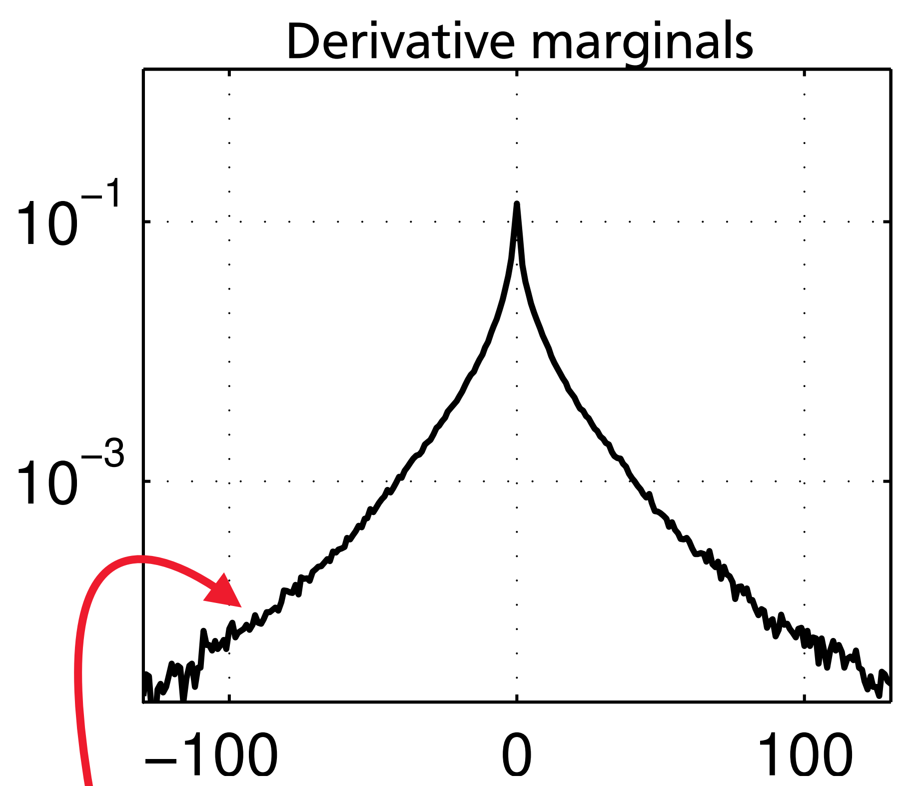
- Denoising performance highly correlated with the generative quality of the model
- No regularization weight λ required to perform well
- Denoised image does not exhibit incorrect statistics



Advantages of the MMSE

TECHNISCHE
UNIVERSITÄT
DARMSTADT

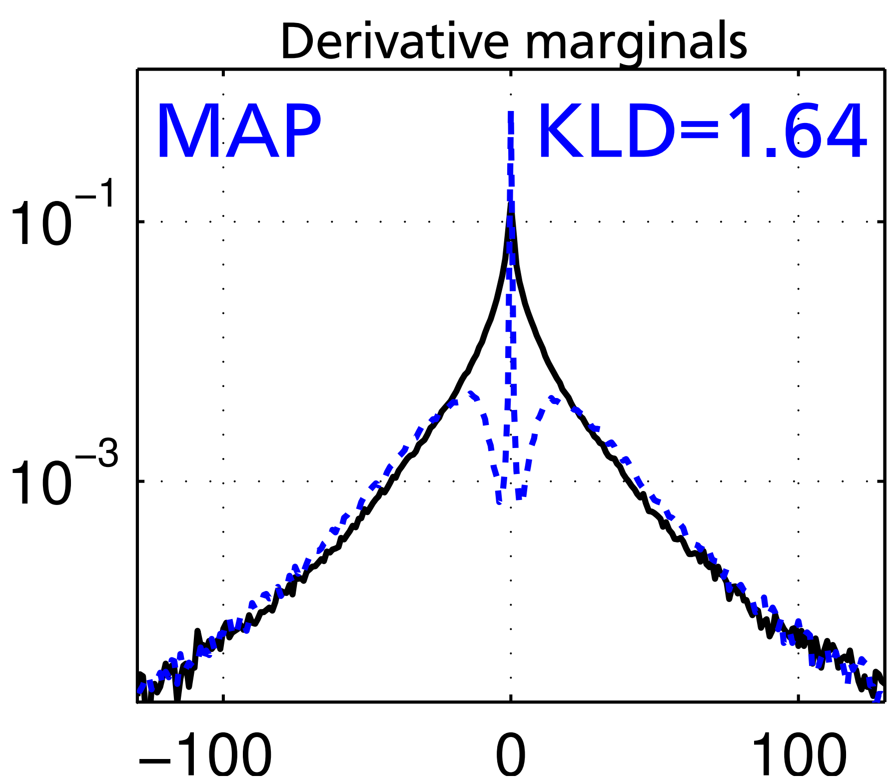
- Denoising performance highly correlated with the generative quality of the model
- No regularization weight λ required to perform well
- Denoised image does not exhibit incorrect statistics
  - No piecewise constant regions



Original images (noise-free)

Advantages of the MMSE

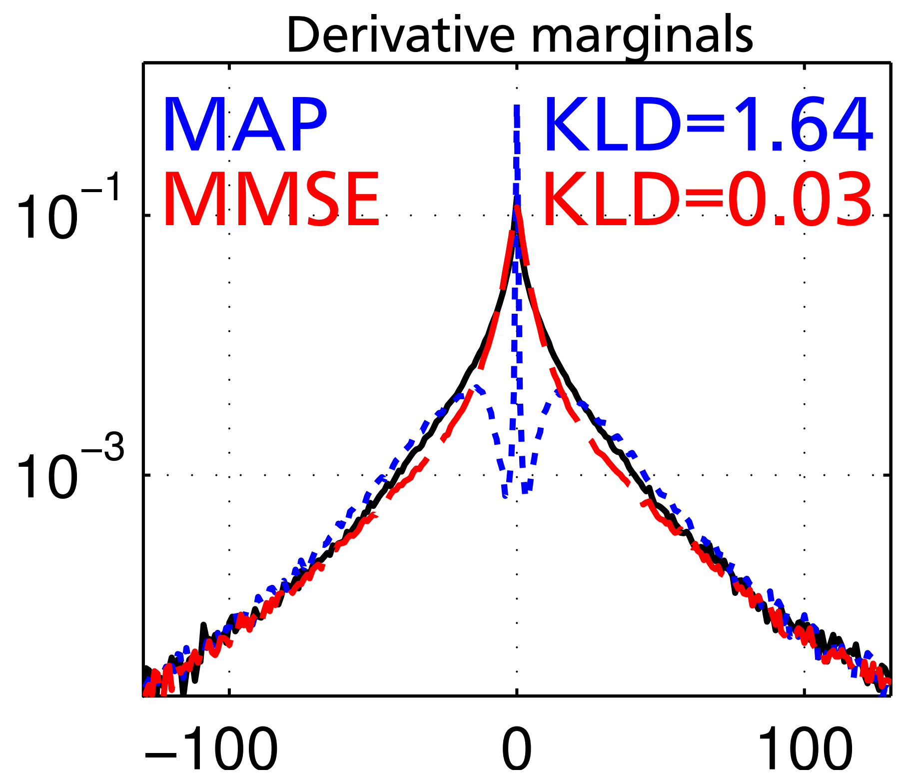
- Denoising performance highly correlated with the generative quality of the model
- No regularization weight λ required to perform well
- Denoised image does not exhibit incorrect statistics
  - No piecewise constant regions



[Woodford et al. ’09]

Advantages of the MMSE

- Denoising performance highly correlated with the generative quality of the model
- No regularization weight λ required to perform well
- Denoised image does not exhibit incorrect statistics
  - No piecewise constant regions
  - Works with standard MRFs



Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

- Evaluated MRFs through their generative properties
  - Based on a flexible framework with an efficient sampler
- Common image priors are surprisingly poor generative models
- Learned better generative MRFs (pairwise & high-order)
  - Potentials more peaked
- Sampling makes MMSE estimation practical
  - Several advantages over MAP
  - Excellent results from generative, application-neutral models

Thanks!

Acknowledgments:
Yair Weiss, Arjan Kuijper, Michael Goesele,
Kegan Samuel, Marshall Tappen

Please come to our poster!

Code and models available soon at http://bit.ly/mmse-mrf

Questions?

More Generative Properties

TECHNISCHE UNIVERSITÄT DARMSTADT