Learning to Mean-Shift in O(1) for Bayesian Image Restoration

Siavash Bigdeli
Images and Visual Representation Lab (IVRL)

Meiguang Jin, Paolo Favaro
Matthias Zwicker

Image restoration

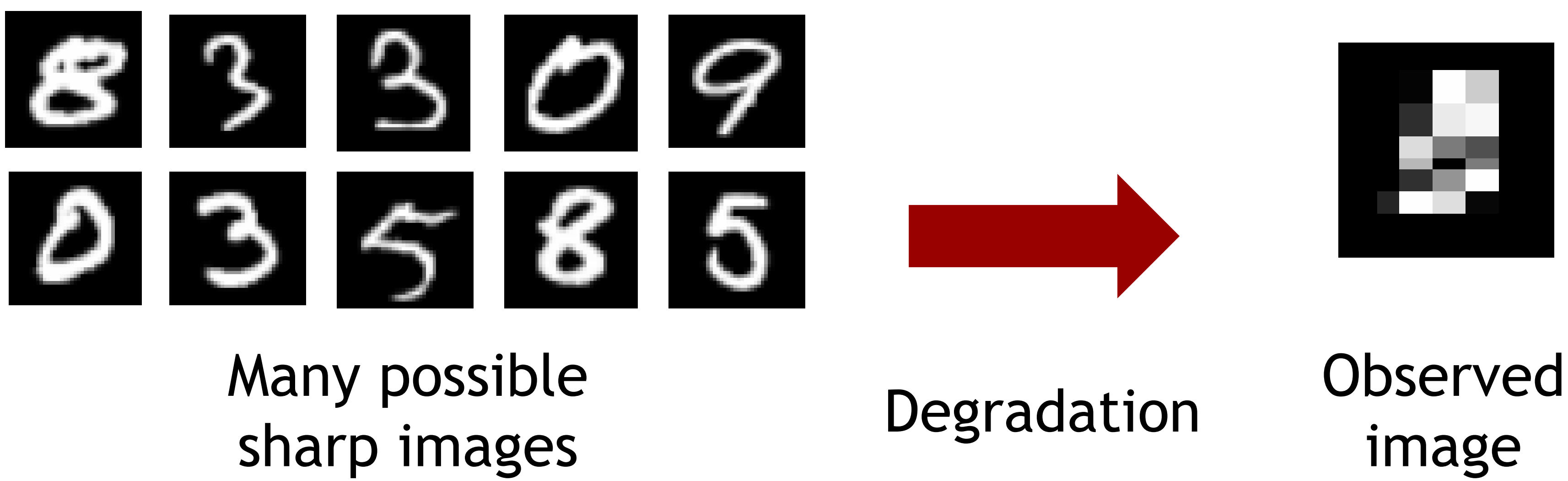
- Restore degraded images by removing blur, noise, holes, etc.

Image restoration

- Restore degraded images by removing blur, noise, holes, etc.

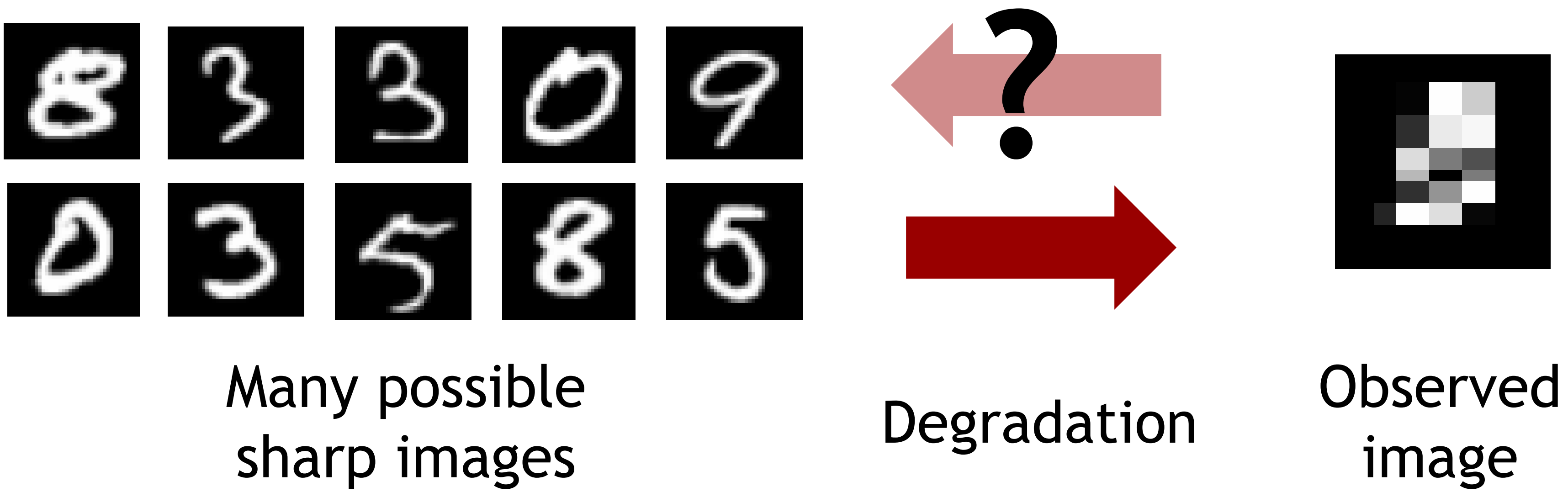
Challenges

- Degradation leads to loss of information



Challenges

- Degradation leads to loss of information
- Restoration is ill-posed



Maximum a-Posteriori

- Given y, maximize posterior probability

$$\underset{x}{\operatorname{argmax}}\, p(x|y)$$

Maximum a-Posteriori

- Given y, maximize posterior probability

$$\underset{x}{\operatorname{argmax}}\, p(x|y) = \underset{x}{\operatorname{argmax}}\, p(y|x)p(x)$$

Maximum a-Posteriori

- Given y, maximize posterior probability

$$
\begin{aligned}
\operatorname*{argmax}_x p(x|y) &= \operatorname*{argmax}_x p(y|x)p(x) \\
&= \operatorname*{argmax}_x \log p(y|x) + \log p(x)
\end{aligned}
$$

Maximum a-Posteriori

- Given y, maximize posterior probability

$$\begin{aligned}\operatorname*{argmax}_x p(x|y) &= \operatorname*{argmax}_x p(y|x)p(x) \\ &= \operatorname*{argmax}_x \boxed{\log p(y|x)} + \log p(x)\end{aligned}$$

- $\boxed{\log p(y|x)}$: data term (likelihood)
  - How well does the solution explain the observed data

Maximum a-Posteriori

- Given y, maximize posterior probability

$$\underset{x}{\operatorname{argmax}}\, p(x|y) = \underset{x}{\operatorname{argmax}}\, p(y|x)p(x)$$
$$= \underset{x}{\operatorname{argmax}}\, \log p(y|x) + \log p(x)$$

- $\log p(y|x)$: data term (likelihood)
  - How well does the solution explain the observed data
- $\log p(x)$: image prior
  - Probability of the solution

Classical and novel priors

- Characteristics

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- External dataset vs. internal image

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- Large images regions vs. small patches

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- True density vs. sparse representations

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- Neural vs. GMM approximations

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- Neural vs. GMM approximations

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- General vs. limited applications

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- Proposed

Method	Domain	Feature	Prior	Parametrization	Application
TV	internal	gradient	Sparsity	L_2-norm	generic
NLM	internal	patch	MRFs	Gaussian	generic
BM3D	internal	patch blocks	Sparsity	DCTs	generic
FoE	external	patch	MRFs	t-distribution	generic
KSVD	external	patch	Sparsity	Dictionaries	generic
EPLL	external	patch	True density	GMMs	generic
VDSR	external	image*	True density	CNNs	upsampling
DnCNN	external	image*	True density	CNNs	denoising
DJDD	external	image*	True density	CNNs	demosaicing
GradNet	external	image*	Kernel density	Gumbel, CNNs	deblurring
IRCNN	external	image*	True density	Cascade of CNNs	generic
Our method	external	image*	Kernel density	CNNs	generic

Classical and novel priors

- Visualizing image priors, Shaham and Michaeli, ECCV 2016

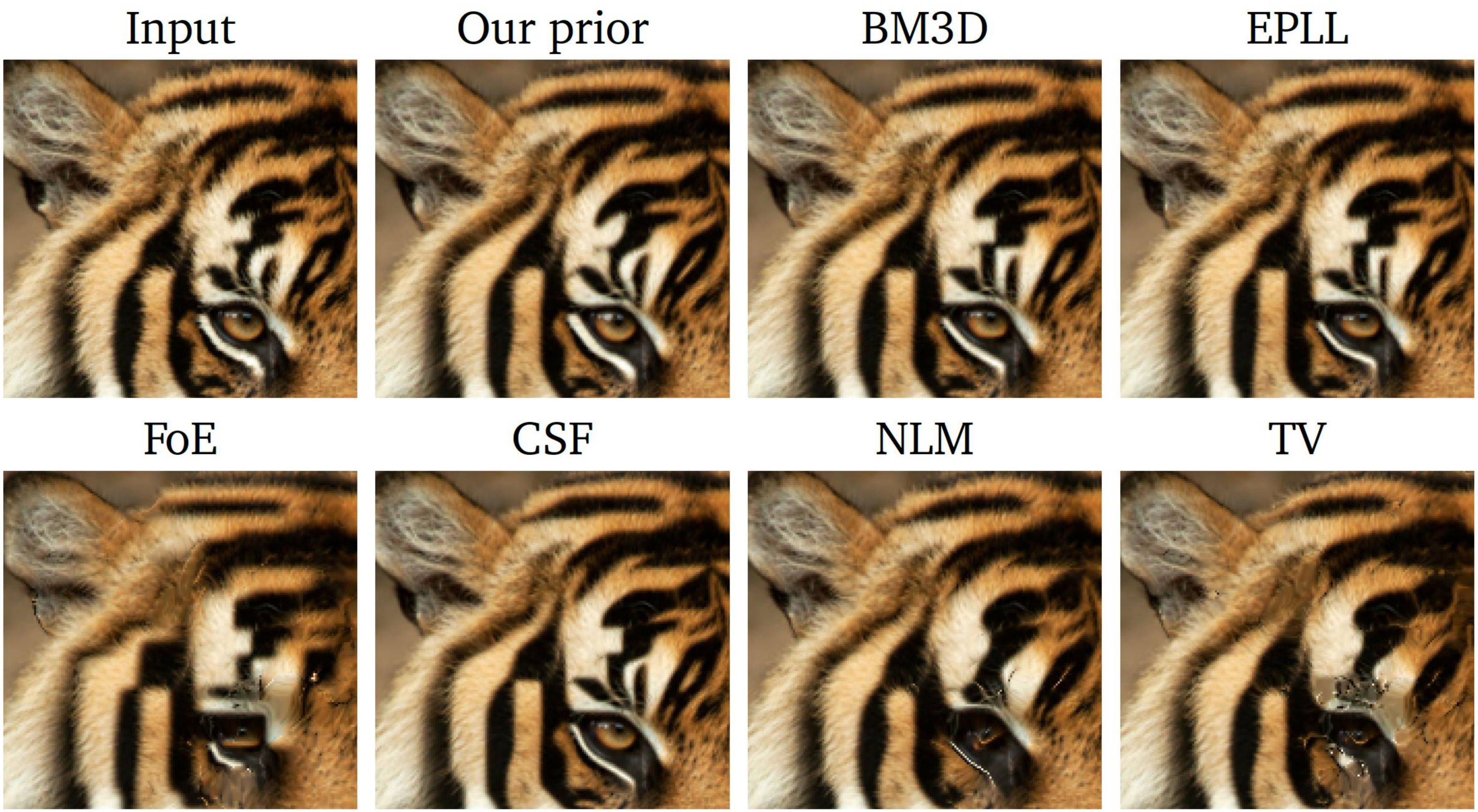


Image Restoration using Autoencoding Priors

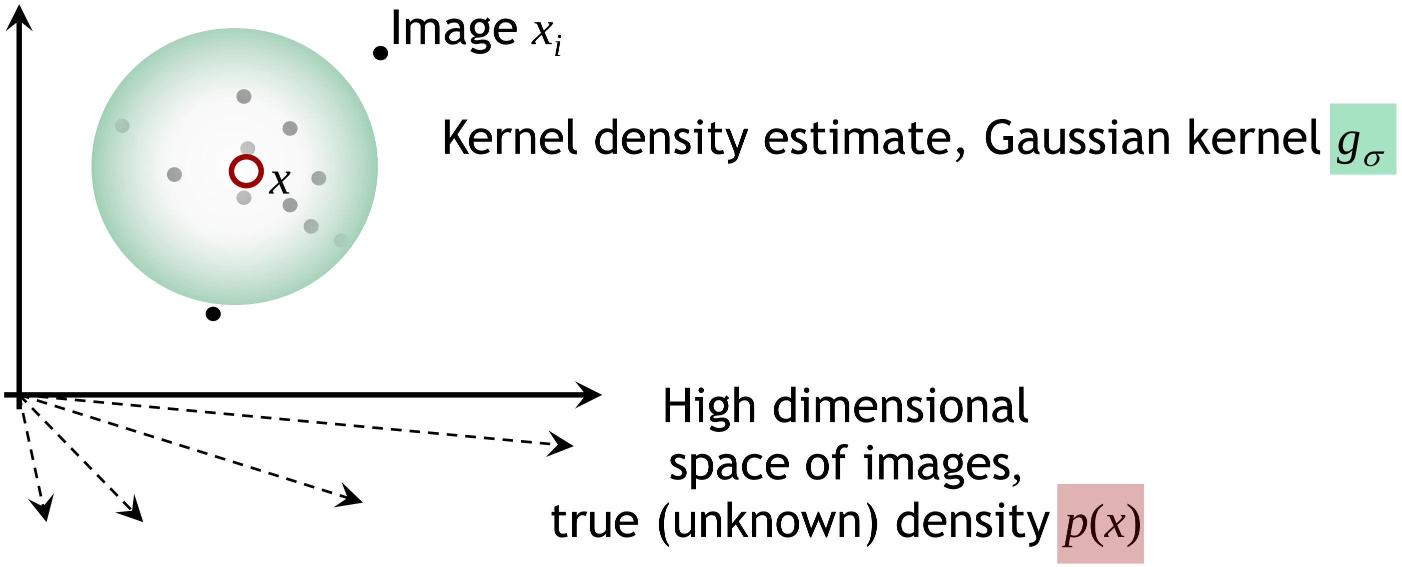
VISAPP 2018

Siavash Arjomand Bigdeli

Matthias Zwicker

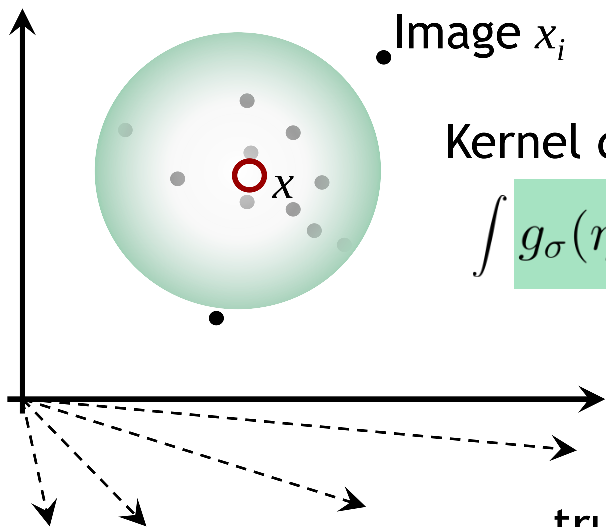
Density estimation prior

- Given: large database of images (or image patches)
- Our prior: approx. $p(x)$ using density estimate



Density estimation prior

- Given: large database of images (or image patches)
- Our prior: approx. $p(x)$ using density estimate



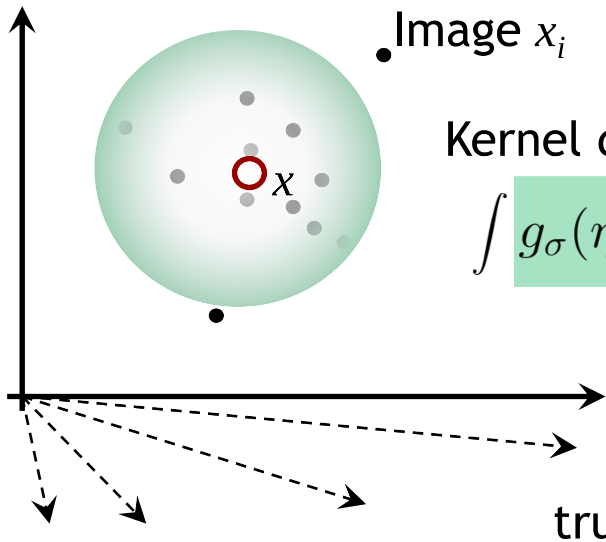
Kernel density estimate, Gaussian kernel g_σ

$$\int g_\sigma(\eta) p(x+\eta) d\eta \approx \frac{1}{n} \sum g_\sigma(x_i - x)$$

High dimensional
space of images,
true (unknown) density $p(x)$

Density estimation prior

- Given: large database of images (or image patches)
- Our prior: approx. $p(x)$ using density estimate



Kernel density estimate, Gaussian kernel g_σ

$$\int g_\sigma(\eta) p(x+\eta) d\eta \approx \frac{1}{n} \sum g_\sigma(x_i - x)$$

High dimensional
space of images,
true (unknown) density $p(x)$

- Contribution
  - Learn (gradient of) density estimate using denoising autoencoders
  - Prior is mean-shift magnitude

Denoising autoencoder (DAE)

- DAE r_σ minimizes

$$\mathcal{L}_{\mathrm{DAE}} = \mathbb{E}_{\eta,x}\left[|x - r_\sigma(x+\eta)|^2\right]$$

  - Images x
  - Gaussian noise η, variance σ^2

Denoising autoencoder (DAE)

- DAE r_σ minimizes

$$\mathcal{L}_{\mathrm{DAE}} = \mathbb{E}_{\eta,x}\left[|x - r_\sigma(x+\eta)|^2\right]$$

  - Images x
  - Gaussian noise η, variance σ^2
- DAE $r_\sigma(x)$ is local mean around image x

[Alain and Bengio, JMLR 2014] [Levin and Nadler, CVPR 2011]

Denoising autoencoder (DAE)

- DAE r_σ minimizes

$$\mathcal{L}_{\text{DAE}} = \mathbb{E}_{\eta,x}\left[|x - r_\sigma(x+\eta)|^2\right]$$

  - Images x
  - Gaussian noise η, variance σ^2
- DAE $r_\sigma(x)$ is local mean around image x
- Our prior is the mean shift vector

[Miyasawa, BIIS 1961] [Raphan and Simoncelli NC 2010]
[Bigdeli and Zwicker, arXiv 2017]

$$\nabla \log \int g_\sigma(\eta) p(x+\eta) d\eta = \frac{1}{\sigma^2}\left(r_\sigma(x) - x\right)$$

Prior

Mean shift vector
Autoencoding error

Denoising autoencoder (DAE)

- DAE r_σ minimizes

$$\mathcal{L}_{\mathrm{DAE}} = \mathbb{E}_{\eta,x}\left[|x - r_\sigma(x+\eta)|^2\right]$$

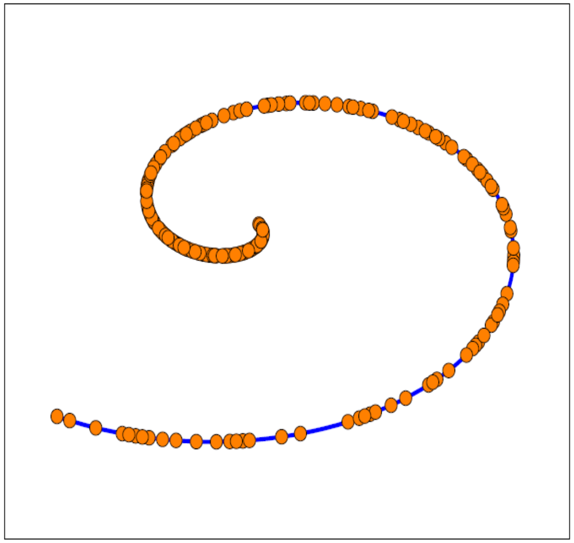
  - Images x
  - Gaussian noise η, variance σ^2
- DAE $r_\sigma(x)$ is local mean around image x
- Our prior is the mean shift vector

$$\underbrace{\nabla \log \overbrace{\boxed{\int g_\sigma(\eta)p(x+\eta)d\eta}}^{\text{Kernel density estimate}}}_{\text{Prior}} = \frac{1}{\sigma^2}\underbrace{\left(r_\sigma(x) - x\right)}_{\substack{\text{Mean shift vector} \\ \text{Autoencoding error}}}$$

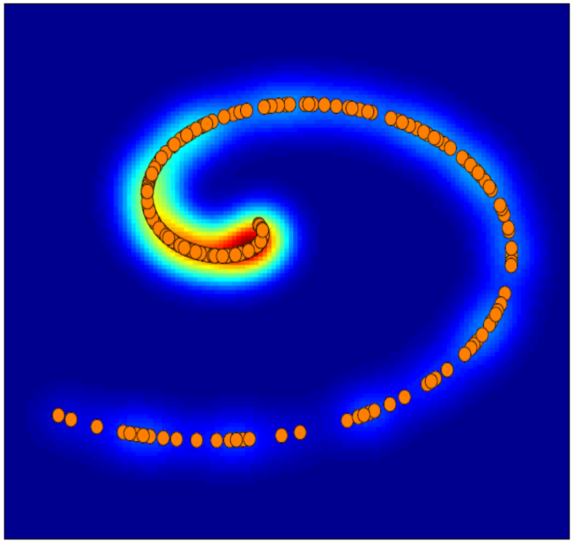
Denoising autoencoder (DAE)

- Simple 2D Spiral distribution

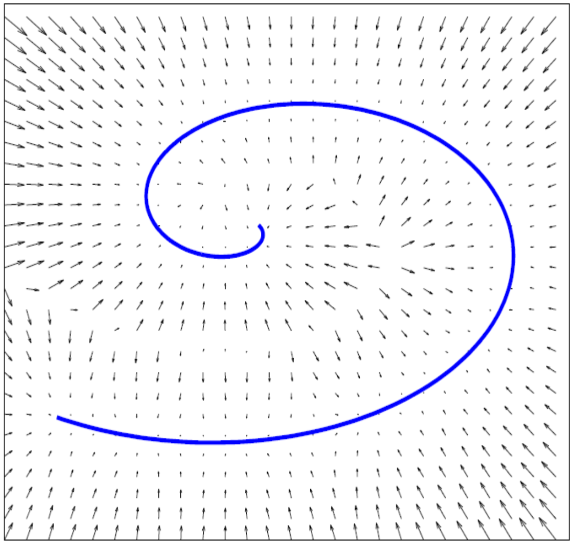
Spiral manifold and observed samples



Smoothed density from observed samples



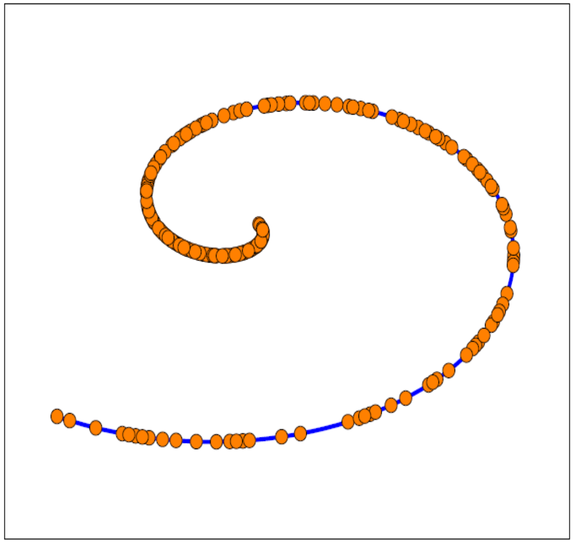
Mean-shift vectors Learned by DAE



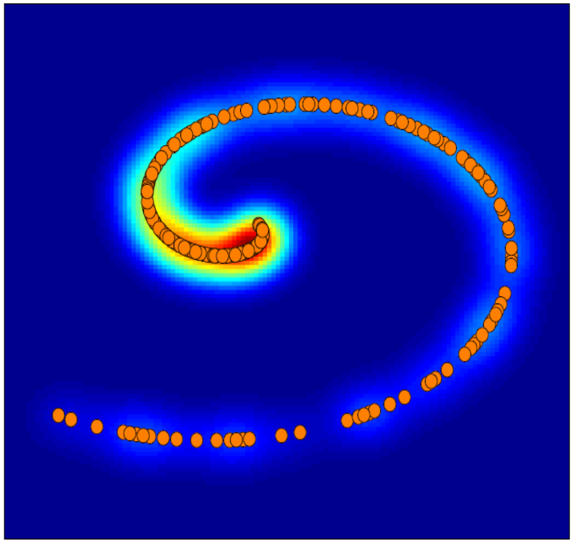
Denoising autoencoder (DAE)

- Simple 2D Spiral distribution

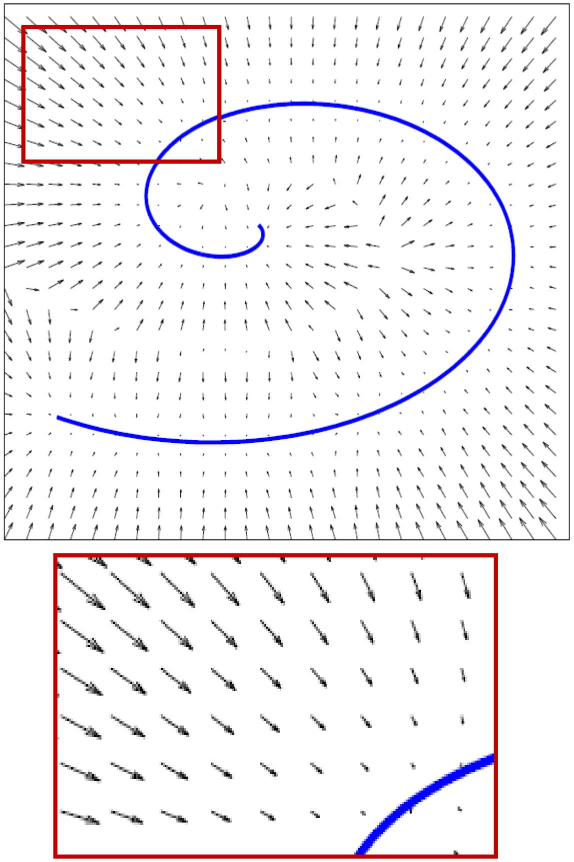
Spiral manifold and observed samples



Smoothed density from observed samples



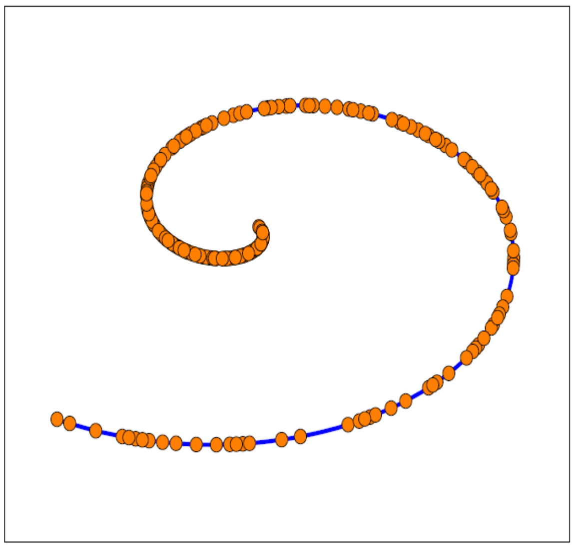
Mean-shift vectors Learned by DAE



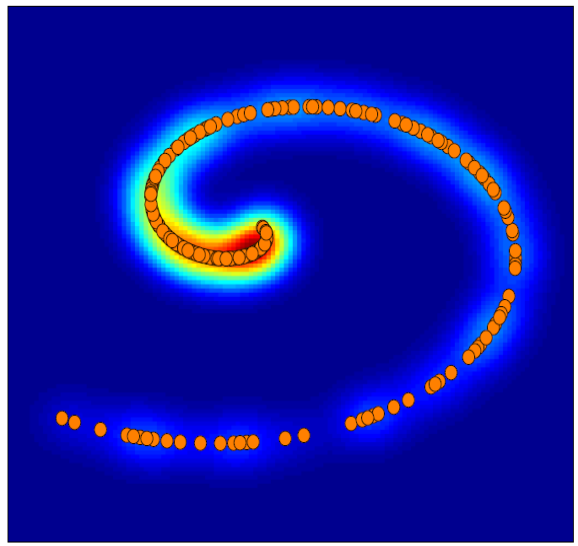
Denoising autoencoder (DAE)

- Simple 2D Spiral distribution

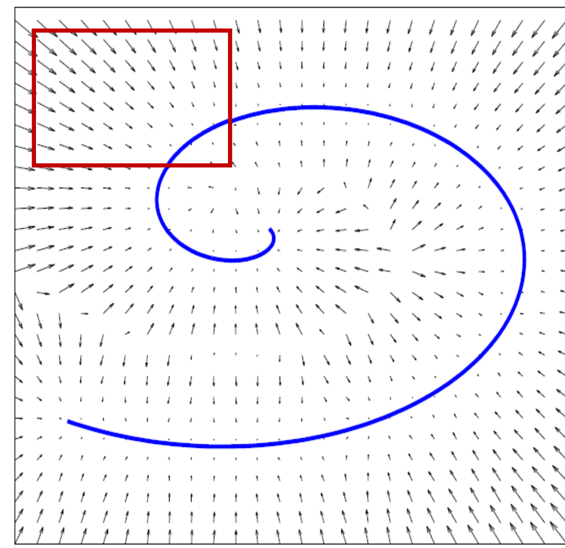
Spiral manifold and observed samples

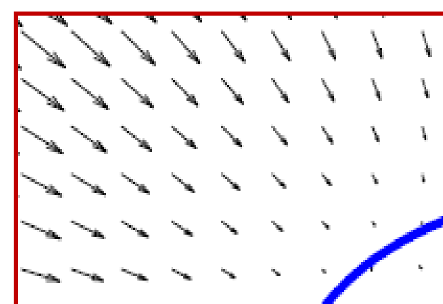


Smoothed density from observed samples



Mean-shift vectors Learned by DAE





- Our Prior: $$|r_\sigma(x) - x|^2$$

length of the Mean Shift vector

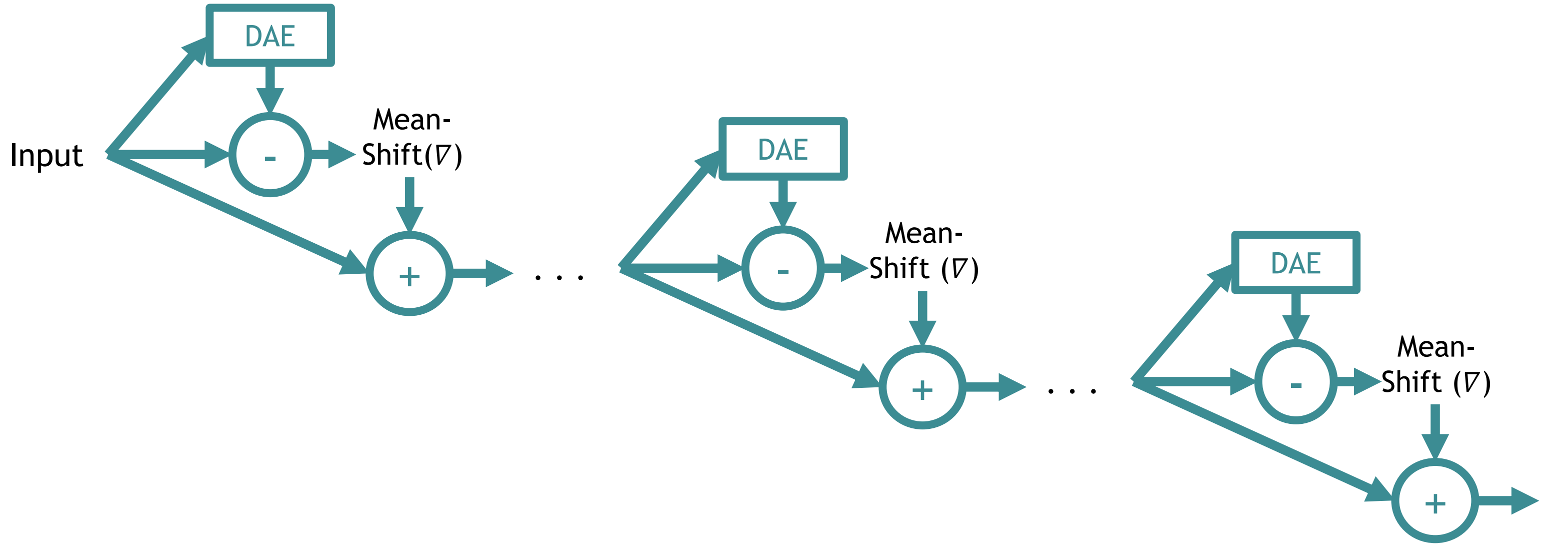
Denoising autoencoder (DAE)

- Parametrized using Neural Networks
  - Learn to mean-shift in O(1)!!

Denoising autoencoder (DAE)

- Parametrized using Neural Networks
  - Learn to mean-shift in O(1)!!

- Gradient steps to minimize our prior error



Denoising autoencoder (DAE)

- Parametrized using Neural Networks
  - Learn to mean-shift in O(1)!!
- Gradient steps to minimize our prior error

Input

Intermediate image

Local minimum

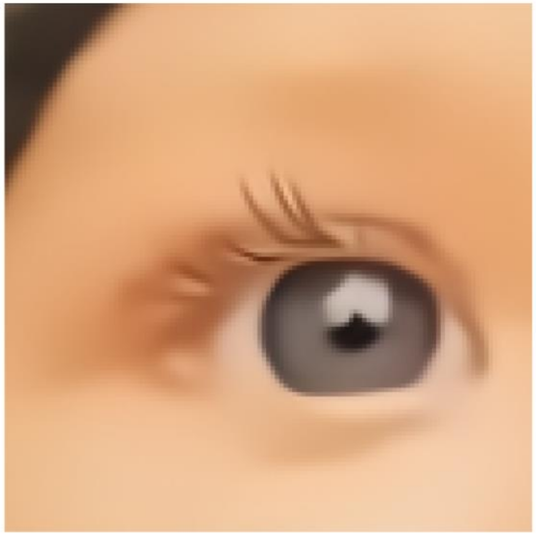


Image restoration algorithm

- Gradient descent to find MAP
  - DAE to compute gradient of prior
  - Gradient of data term depending on degradation model, can be computed directly

Image restoration algorithm

- Gradient descent to find MAP

Blurry | Iterations 10, 30, 100 and 250

$23.12dB$ | $24.17dB$ | $26.43dB$ | $29.1dB$ | $29.9dB$

Results: non-blind deblurring

- Blurred

Results: non-blind deblurring

- Our restoration

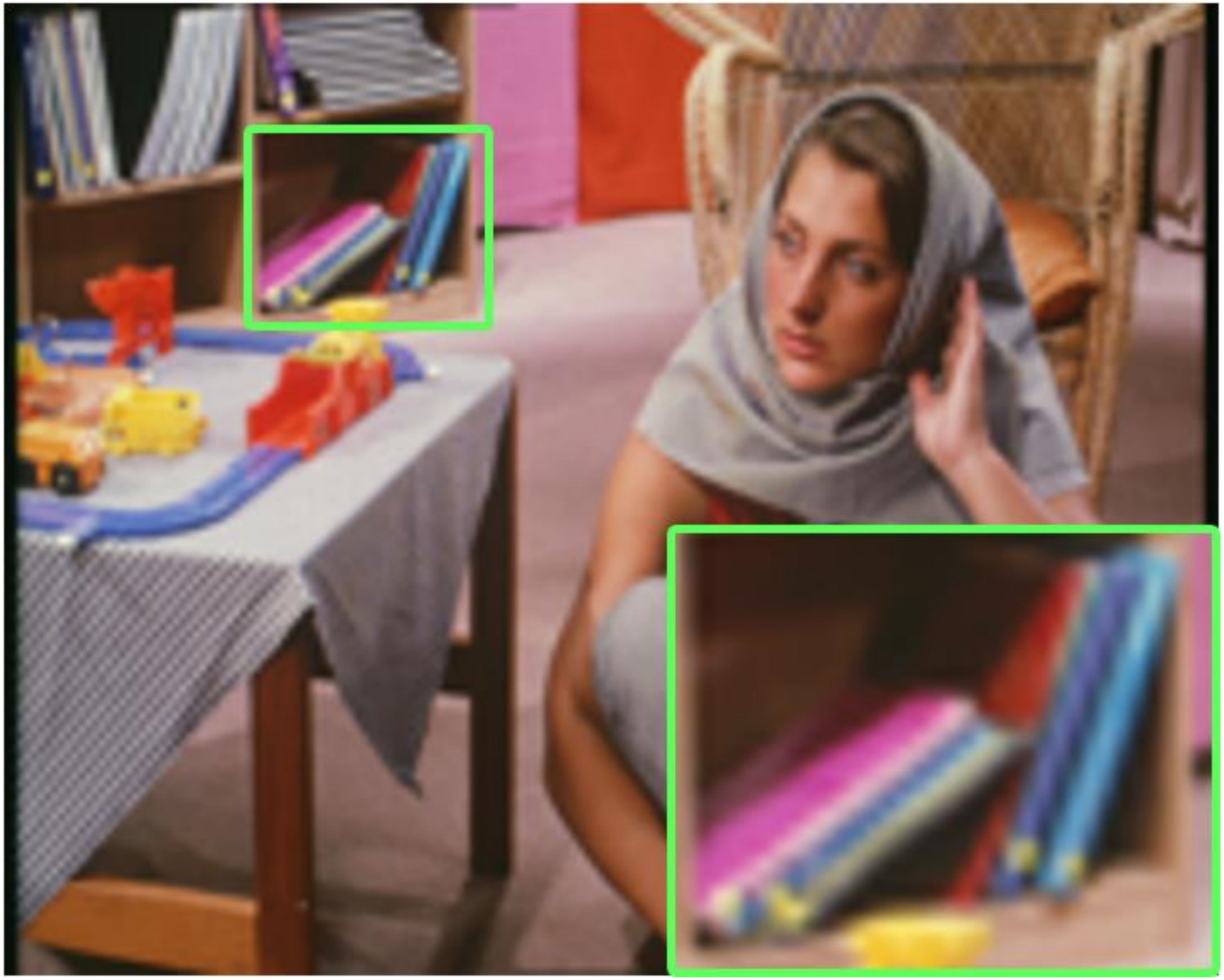
Results: deblurring (PSNR)

Method	2.55	7.65	12.75	time(s)
Levin	31.09	27.40	25.36	3.09
EPLL	32.51	28.42	26.13	16.49
RTF-6	32.51	21.44	16.03	9.82
IRCNN	30.78	28.77	**27.41**	2.47
DAEP (Ours)	**32.69**	**28.95**	26.87	11.19

Learning Deep CNN Denoiser Prior for Image Restoration
Zhang et al., CVPR 2017

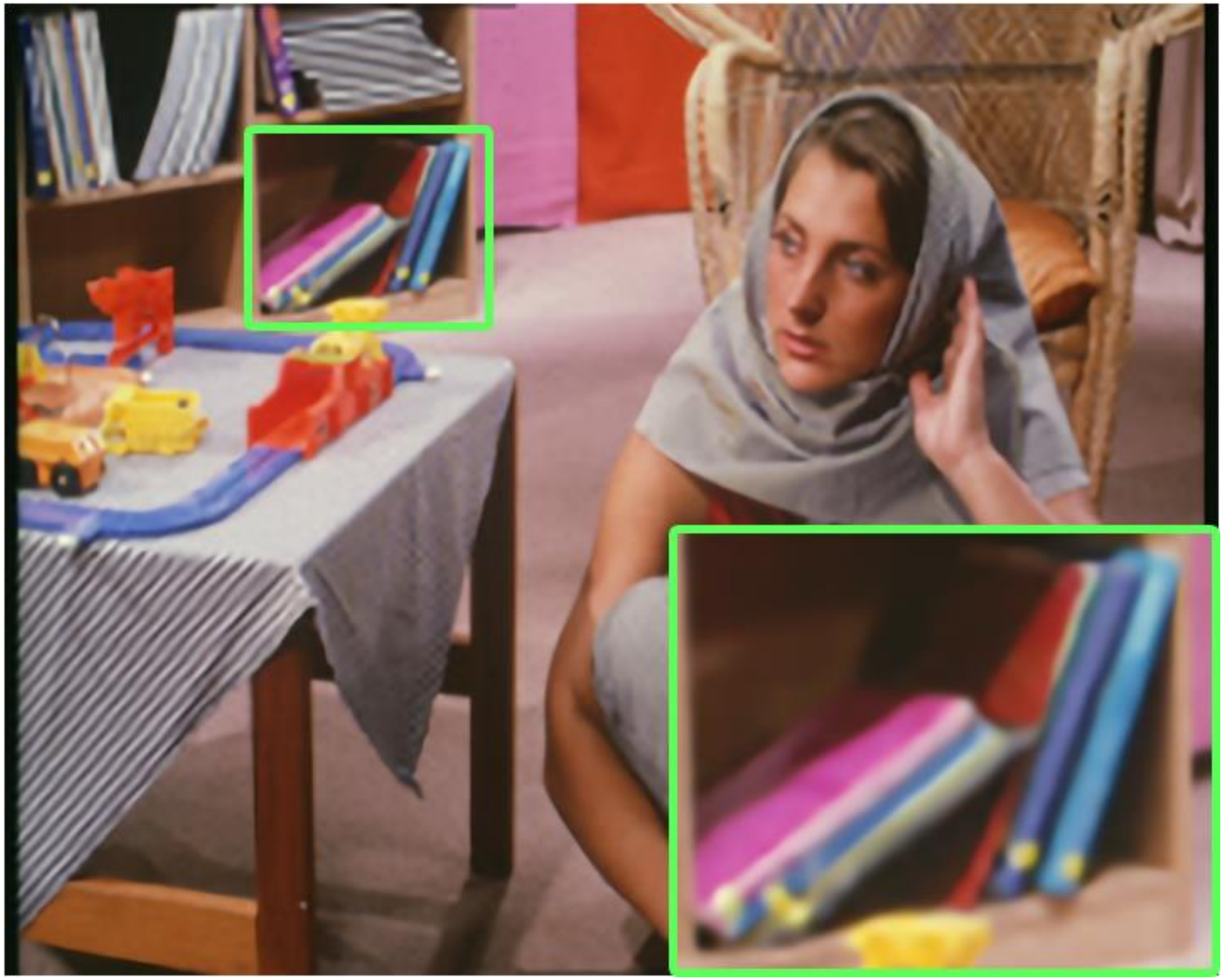
Results: super-resolution

- 4x down-sampled



Results: super-resolution

- Our restoration



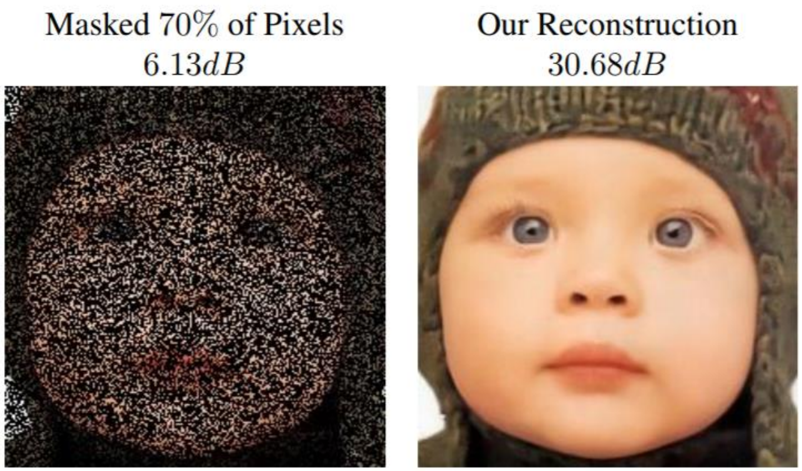
Results: super-resolution (PSNR)

Method	×2	×3	×4	×5
Bicubic	28.53	25.92	24.44	23.46
SRCNN	30.52	27.48	25.76	24.05
TNRD	30.53	27.60	25.92	24.61
VDSR	30.72	27.81	26.16	24.01
DnCNN-3	30.99	27.93	**26.25**	24.26
IRCNN	30.79	27.68	25.96	24.73
DAEP (Ours)	**31.07**	**27.93**	26.13	**24.88**

Beyond a gaussian denoiser:
Residual learning of deep CNN for image denoising
Zhang et al., IEEE TIP 2017

Summary

- Natural image prior based on density estimate
- Generic
  - Deblurring
  - Super-resolution
  - Hole-filling



- State of the art results
- Code: https://github.com/siavashbigdeli/DAEP

Discussion

- Many recent approaches use denoising to solve generic image restoration problems
- Usually motivated by specific underlying algorithm (ADMM)

Our approach

- First one to make connection between denoising and mean shift
- Insight: optimal Gaussian denoising under L_2 loss equivalent to computing gradient on smoothed data distribution

What's next

- Generic image prior
  - Degradation parameters unknown during training
  - Deblurring, demosaicing, super-resolution, etc.
- Generic restoration
  - Degradation parameters unknown during restoration!
  - Noise-blind, Kernel-blind

Deep Mean-Shift Priors for Image Restoration

NIPS 2017

Siavash Arjoman Bigdeli

Meiguang Jin

Paolo Favaro

Matthias Zwicker

Bayes formulation

- Expected utility of solution *x*

$$E_{\tilde{x}}[G(\tilde{x},x)] = \int \underbrace{G(\tilde{x},x)}_{\text{Utility}}\underbrace{p(y|\tilde{x})}_{\text{likelihood}}\underbrace{p(\tilde{x})}_{\text{Prior}}d\tilde{x}$$

Suitable utility, Jensen's inequality

$$\geq \underbrace{\int g_\sigma(\epsilon)\log p(y|x+\epsilon)d\epsilon}_{\text{Data term } \ \mathrm{data}(x)} + \underbrace{\log\int g_\sigma(\eta)p(x+\eta)d\eta}_{\text{Image likelihood } \ \mathrm{prior}(x)}$$

Image restoration algorithm

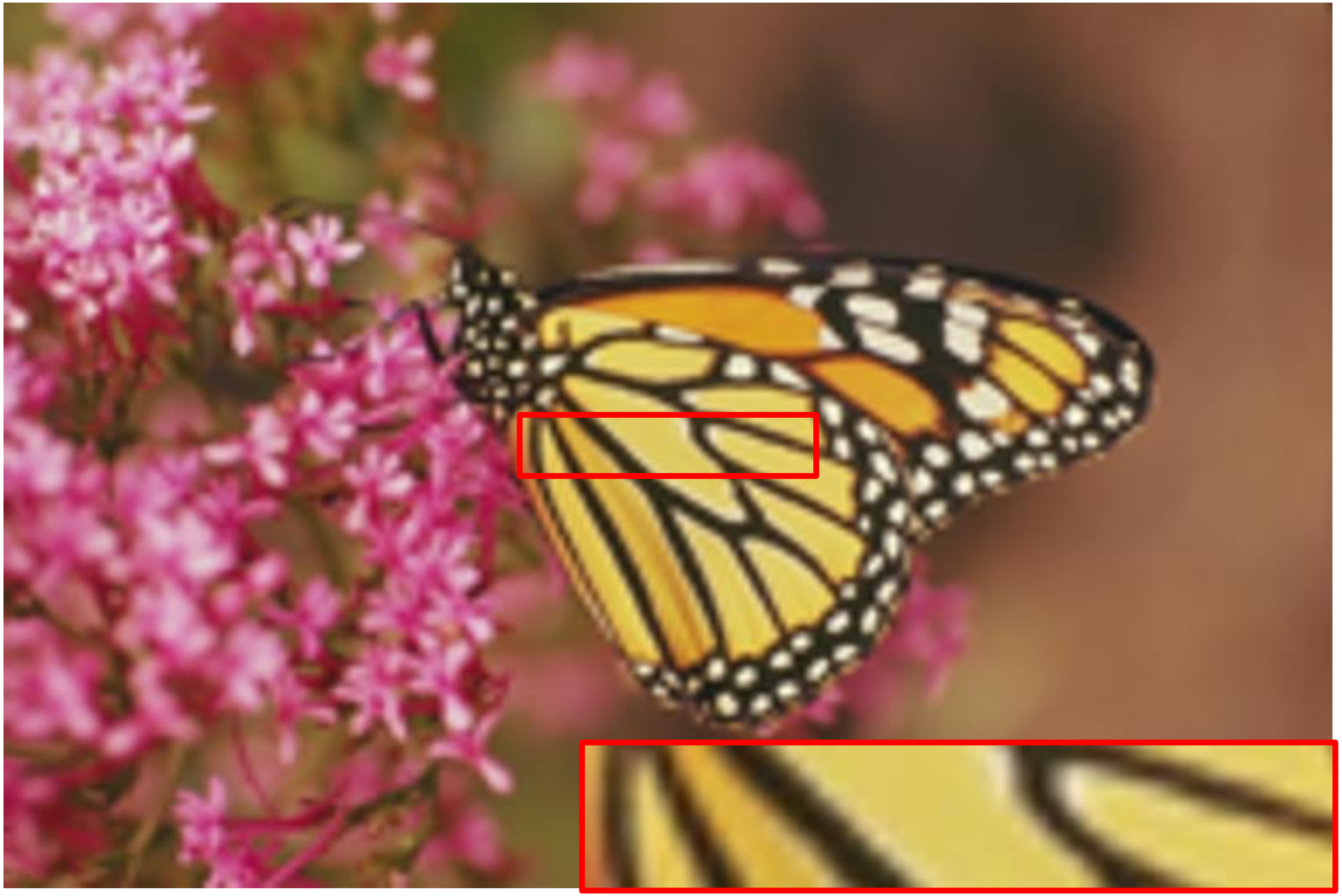
- Gradient descent to minimize (negative) utility
- Denoising autoencoder (DAE) gives gradient of prior

$$\nabla_x \operatorname{prior}(x) = \nabla \log \int g_\sigma(\eta) p(x+\eta) d\eta = \Big(\operatorname{DAE}_\sigma(x) - x \Big) / \sigma^2$$

- Gradient of data term can be computed
  - known parameters $\nabla_x \operatorname{data}(x)$
  - noise-blind $\nabla_x \operatorname{data}_{\sigma_n^*}(x)$
  - noise- and kernel-blind $\nabla_{x,k} \operatorname{data}_{\sigma_n^*}(x)$

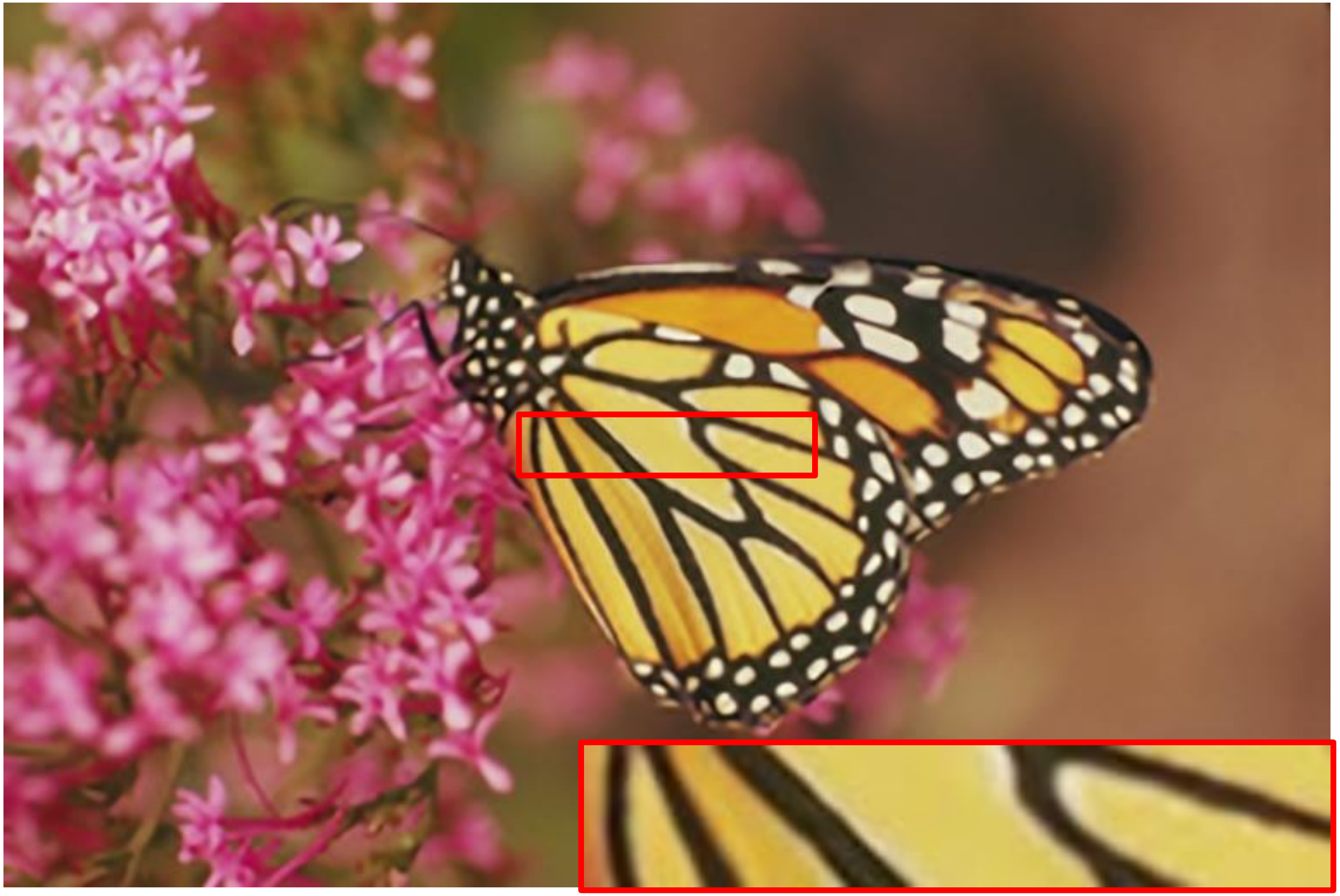
Results: super-resolution

- 4x down-sampled



Results: super-resolution

- Our restoration



Results: non-blind deblurring

- Blurred

Results: non-blind deblurring

- Our restoration

Results: deblurring (PSNR)

- Non-blind

Method σ_n:	2.55	5.10	7.65	10.2
EPLL	25.38	23.53	22.54	21.91
DAEP	25.42	23.67	22.78	22.21
IRCNN	25.60	24.24	23.42	22.91
GradNet 7S	25.57	24.23	23.46	22.94
Ours	<u>25.69</u>	<u>24.45</u>	<u>23.60</u>	**22.99**
Ours + NA	**26.00**	**24.47**	**23.61**	<u>22.97</u>

Results: deblurring (PSNR)

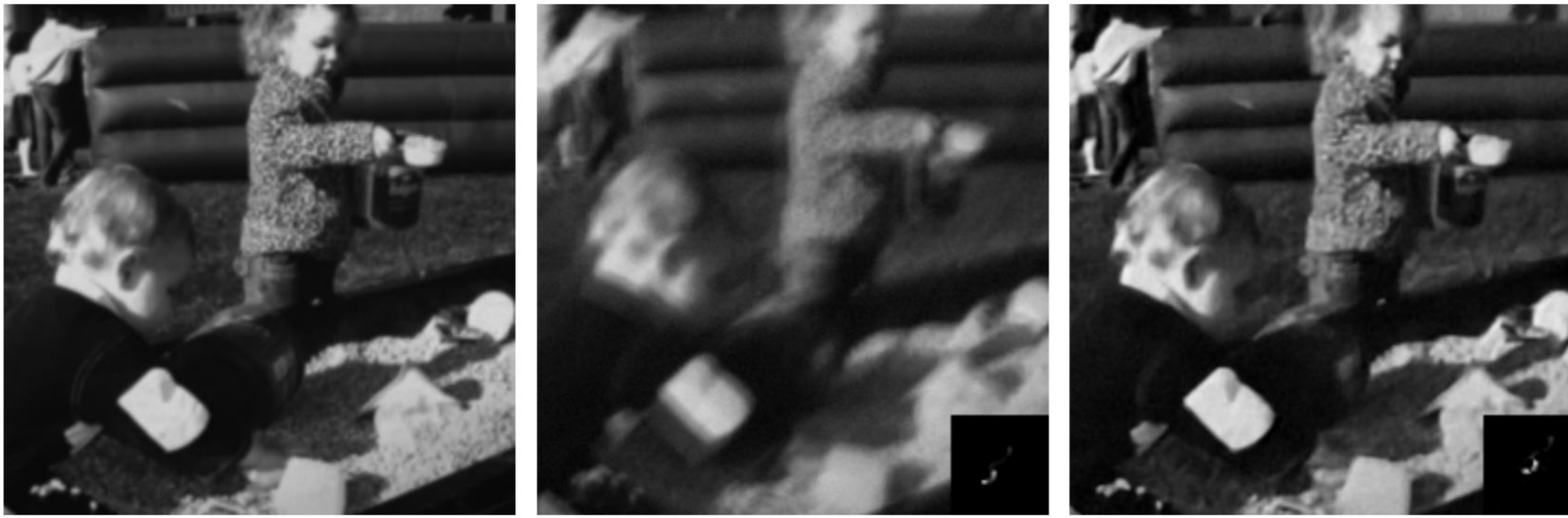
- Noise-blind

Method	σ_n:	2.55	5.10	7.65	10.2
EPLL		25.38	23.53	22.54	21.91
DAEP		25.42	23.67	22.78	22.21
IRCNN		25.60	24.24	23.42	22.91
GradNet 7S		25.57	24.23	23.46	22.94
Ours		25.69	24.45	23.60	**22.99**
Ours + NA		**26.00**	**24.47**	**23.61**	22.97

Noise-blind image deblurring
Jin et al., CVPR 2017

Results: blind deblurring

Ground Truth | Blurred with 1% noise | Ours (blind)



Summary

- Generic image prior based on (gradient of) density estimate of natural image distribution
  - Deblurring, demosaicing, super-resolution, etc.
- Generic framework
  - Noise- and kernel-blind
- Learned using DAE
- State of the art results
- Code: https://github.com/siavashbigdeli/DMSP

Limitations

- Prior seems to capture mostly low-level image properties (sharp edges)
  - Receptive field 41x41 pixels
  - Cannot “invent” semantically meaningful image structure (textures)
- Speed far from real-time applications
- Practical issues with DAE overfitting to noisy images
- Degradation *model* still given

Conclusions

- Generic image prior
  - Deblurring, demosaicing, super-resolution, etc.
- Bayes formulation
  - Noise- and Kernel-blind

Future work

- Generative models to “invent” texture
  [Globally and Locally Consistent Image Completion, Iizuka et al, SIGGRAPH 2017]
- Multiresolution
  [Koltun ICCV 2017, Kerras ICLR 2018]
- DAEs for depth, NIR, 3D models, etc.