Machine Learning Approaches for Solving Inverse Imaging Problems

Aggelos K. Katsaggelos

Professor
Joseph Cummings Chair
Northwestern University
Department of EECS
Department of Linguistics
NorthSide University Hospital System
Argonne National Laboratory
Evanston, IL 60208
www.ece.northwestern.edu/~aggk

EUSIPCO, Kos Island, August 29, 2017

Talk Outline

- Inverse Problems
- Analytic Approaches
- Data-Driven Approaches
  - Dictionary Techniques
  - Deep Learning Techniques
- Analytic *AND* Data-Driven Approaches
- Closing Thoughts

Talks Objectives

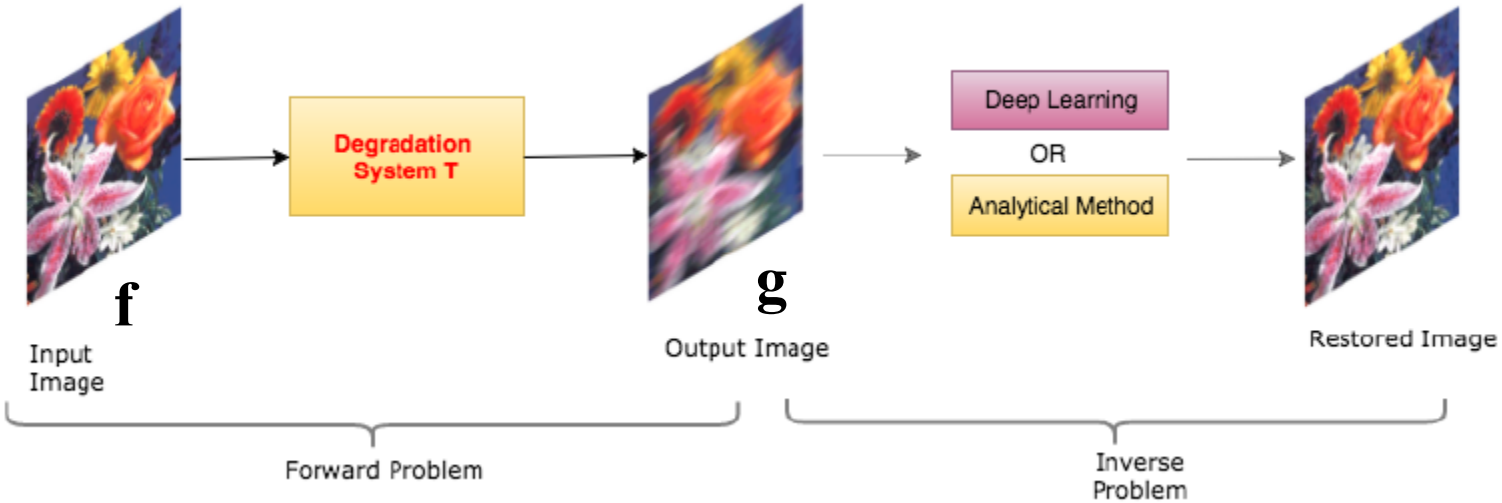
- Compare and contrast analytical and learning approaches in solving inverse problems in imaging
- Review some of the recent advances in solving inverse imaging problems using (deep) learning approaches
- Address their point of intersection

Questions to be Addressed

- Relative advantages of analytical and learning approaches for solving inverse problems, or
- for a specific problem at hand which of the approaches in our available toolbox should one use?
- Under what circumstances would one expect learning approaches to provide more accurate solutions (computational considerations aside) than analytical approaches?
- or would one expect to see the same gains in solving inverse problems with neural networks as we have seen in solving certain classification problems (equaling or even surpassing human performance)?

A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using Deep Neural Networks for Inverse Problems in Imaging," *IEEE Signal Processing Magazine*, submitted for publication.

Problem Considerations



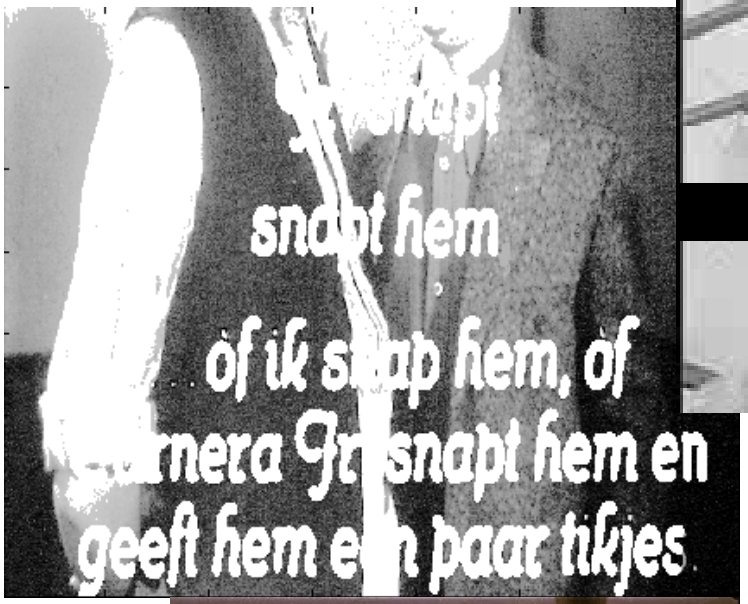
Known	Problem Type
f, g	system identification
f, T	system implementation
T, g	recovery -- an inverse problem
g	blind recovery
g, T partially	semi-blind recovery

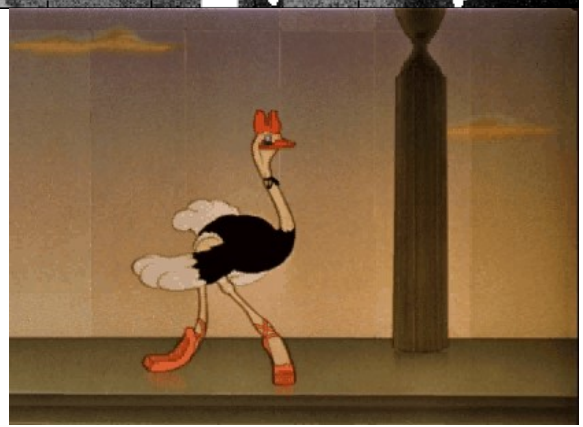
Parameter Estimation

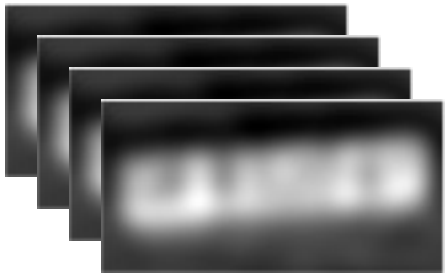
Forms of the Recovery Problem

- Single vs. multiple observations
- Noise Smoothing
- Restoration/Deconvolution (1D, 2D, 3D)
  - Multi-spectral, multi-channel
- Removal of Compression Artifacts
- Super-Resolution (Pansharpening, Demosaicking)
- Inpainting, Concealment
- Reconstruction
- Dual Exposure Imaging
- Compressive Sensing
- Motion Estimation
- Light-field Reconstruction

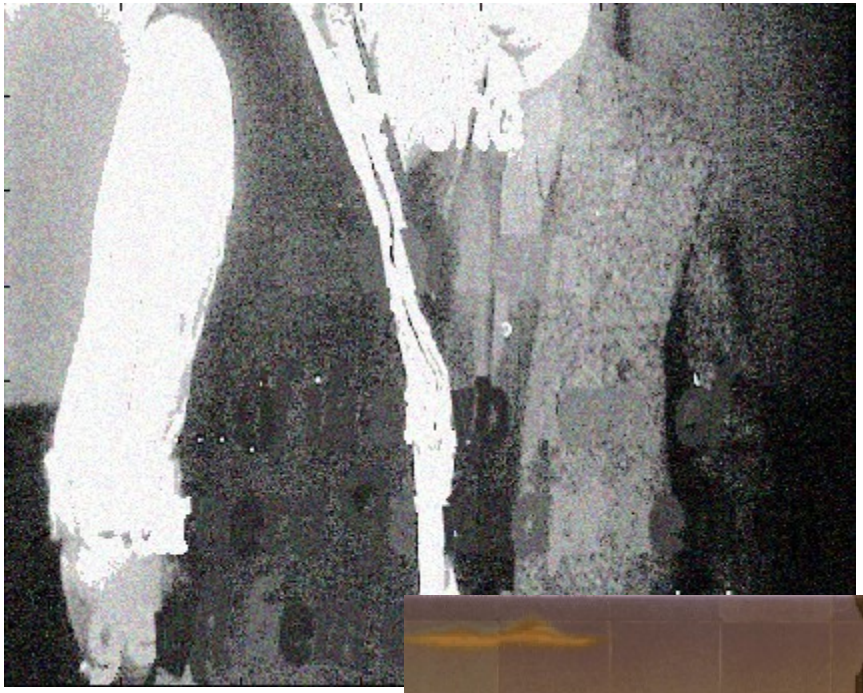
Degraded Images







Restored Images



4 images

Degraded Images

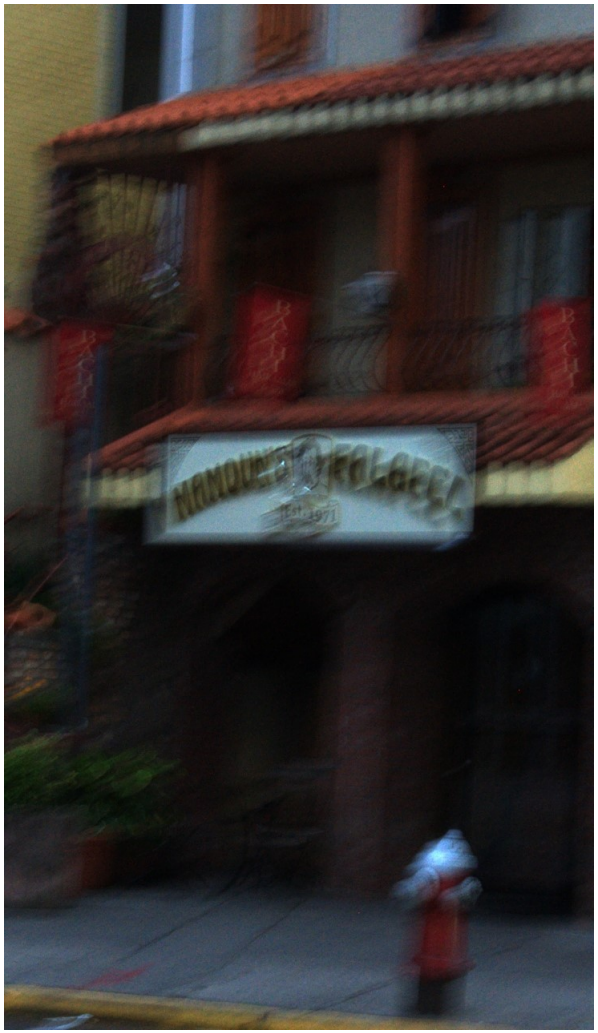
Restored Images

Deconvolved images using log prior

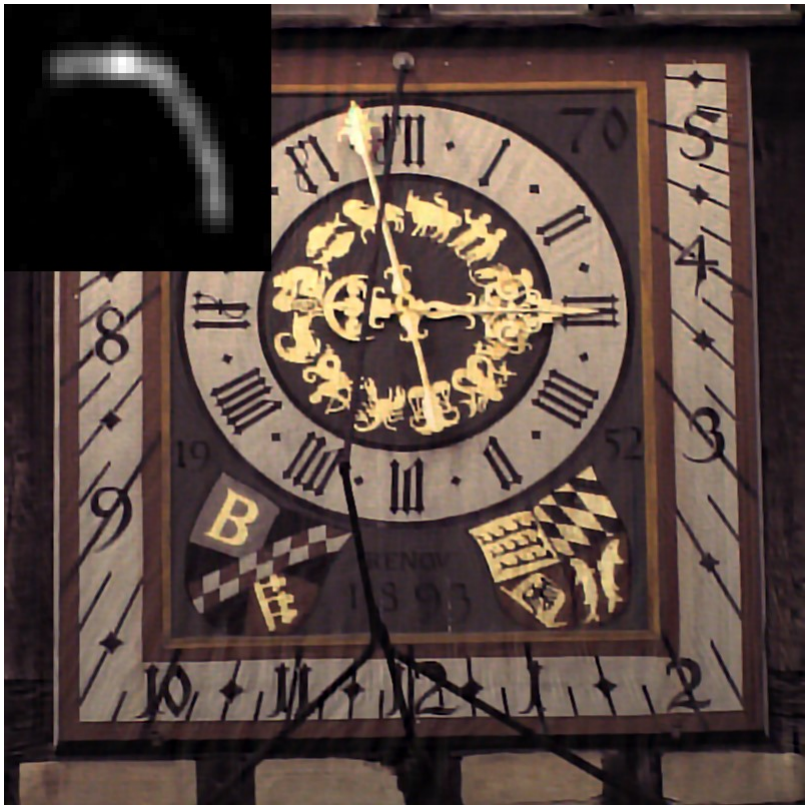
Software available at http://www.dbabacan.info/BDGSP.php

S.D. Babacan, R. Molina, M.N. Do, and A.K. Katsaggelos, "Bayesian Blind deconvolution with general sparse image priors" in *European Conference on Computer Vision (ECCV)*, 341-355, Florence (Italy), September 2012

Degraded Images



Restored Images



X. Zhou, M. Vega, F. Zhou, R. Molina, and A. K. Katsaggelos, “Fast Bayesian Blind Deconvolution with Huber Super Gaussian Priors”, Digital Signal Processing, vol. 60, 122-133, 2017.

Software available at https://sites.google.com/site/fbdhsgp/

Analytical vs Learning Approaches

- **Pros: Analytical**
- Incorporation of domain knowledge

- **Cons: Learning**
- Hard to incorporate domain knowledge

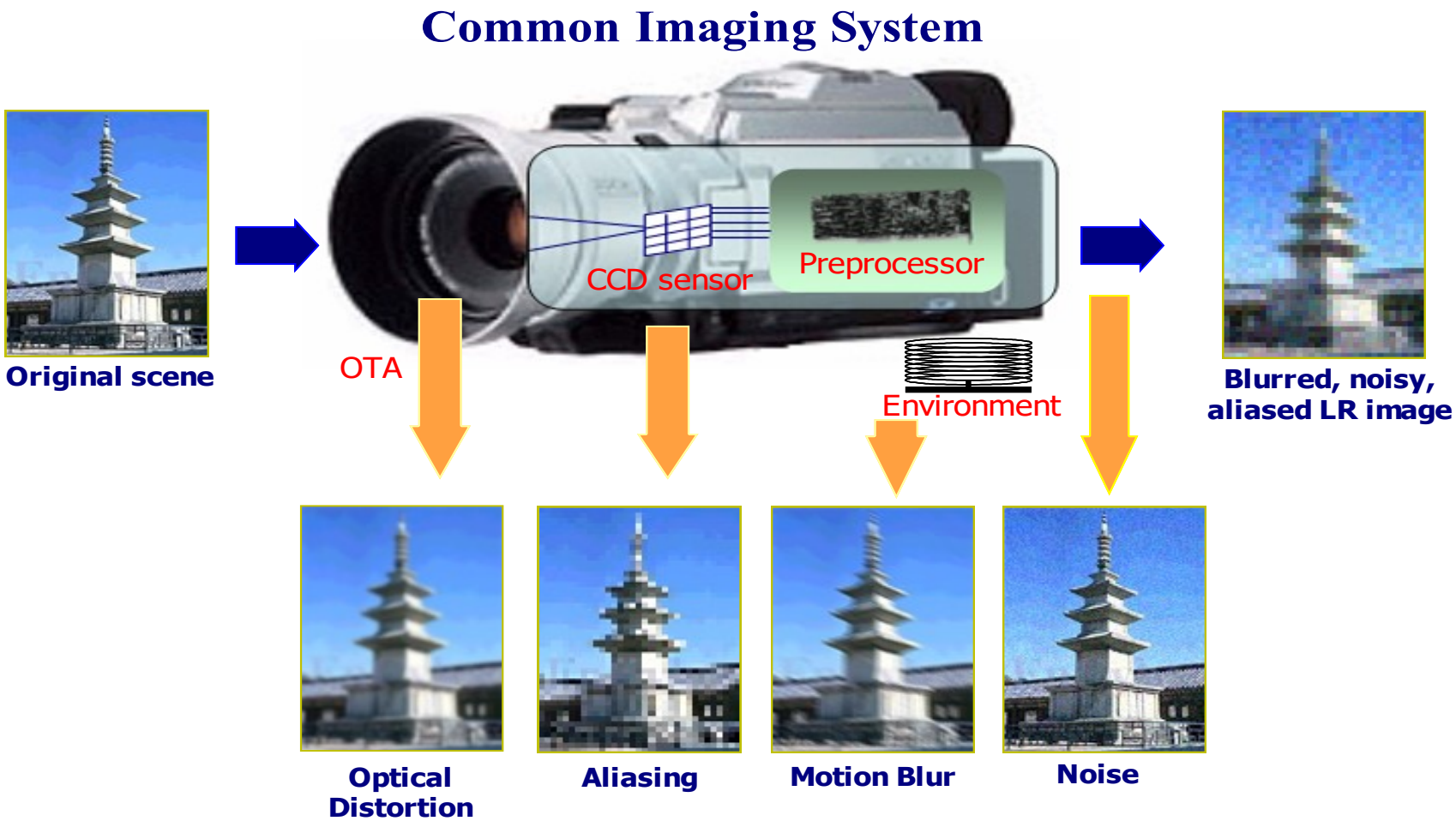
- **Pros: Learning**
- Fast Inference

- **Cons: Analytical**
- Typically Slow Inference

- **Pros: Analytical**
- Typically require one observation

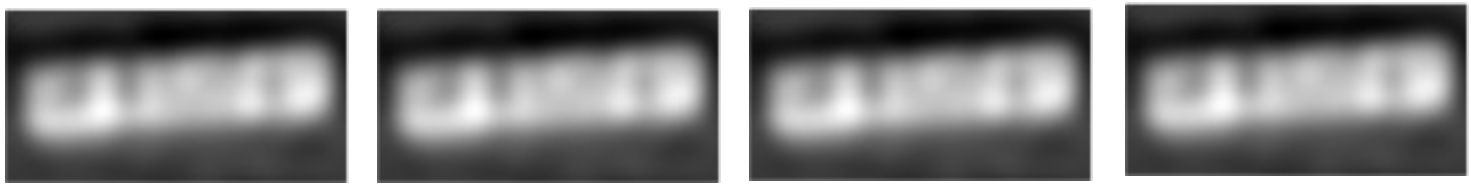
- **Cons: Learning**
- Require large amounts of data

Image Super Resolution



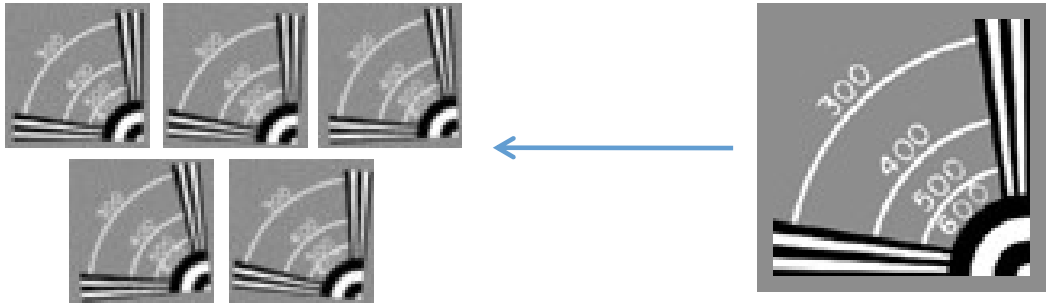
A. K. Katsaggelos, R. Molina, J. Mateos, *Super Resolution of Images and Video*, Morgan and Claypool, 2007.

Image Super Resolution



4 images

Image Super Resolution



$$\mathbf{y}_k = \mathbf{A}\ \mathbf{H}_k\ \mathbf{C}(\mathbf{s}_k)\ \mathbf{x} + \mathbf{n}_k = \mathbf{B}_k(\mathbf{s}_k)\mathbf{x} + \mathbf{n}_k$$

LR observation

Downsampling

PSF

Motion Warping

HR image

Noise

SD Babacan, R Molina, AK Katsaggelos, " Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Processing*, 17 (3), 326-339, 2008.

Bayesian Model

- Modeling the uncertainties in the registration parameters

$$\mathrm{p}(\mathbf{s}_k) = \mathcal{N}(\mathbf{s}_k | \bar{\mathbf{s}}_k^p, \mathbf{\Lambda}_k^p)$$

Motion Parameters

Initial Estimates

A priori covariance matrix

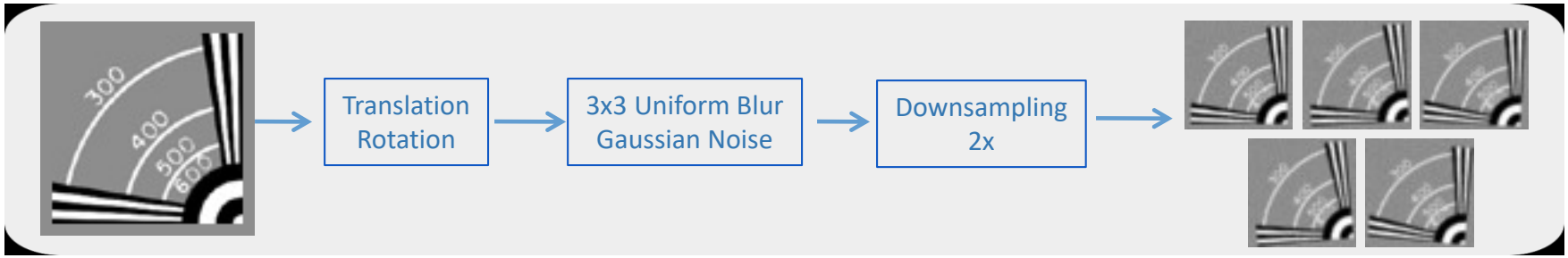
- Joint Distribution

Motion Prior

$$\mathrm{p}(\mathbf{y}, \mathbf{x}, \{\mathbf{s}_k\}, \alpha_{\mathrm{im}}, \{\beta_k\}) = \underbrace{\mathrm{p}(\mathbf{y}|\mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\})}_{\text{Gaussian Noise}} \underbrace{\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}})}_{\text{TV image prior}} \mathrm{p}(\mathbf{s}_k) \underbrace{\mathrm{p}(\alpha_{\mathrm{im}}) \prod_{k=1}^{L} \mathrm{p}(\beta_k)}_{\text{Hyperpriors}}$$

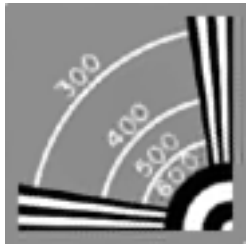
Chen, Z., D. S. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian Methods For Multimedia Problems", *IEEE Trans. on Multimedia*, vol. 16, issue 4, pp. 1000-1017, 06/2014.

Experiments

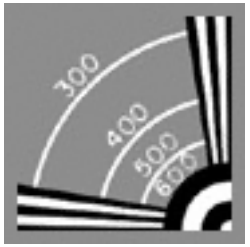


Bicubic

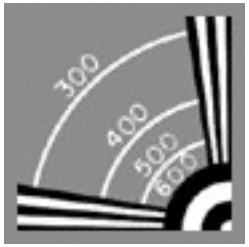
Zomet



RSR

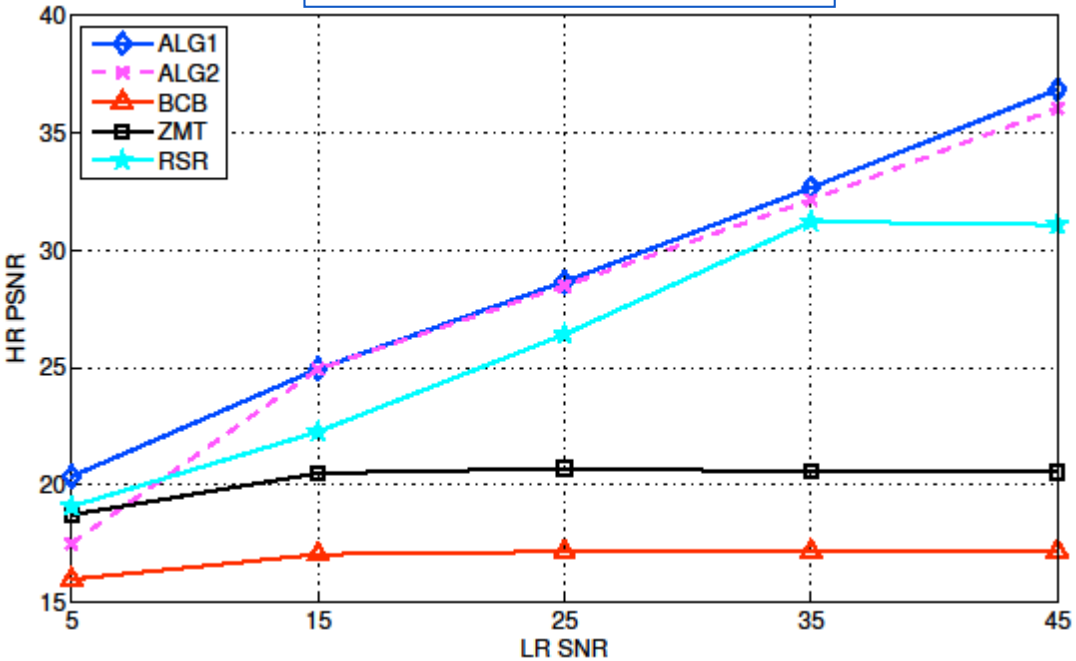


ALG1



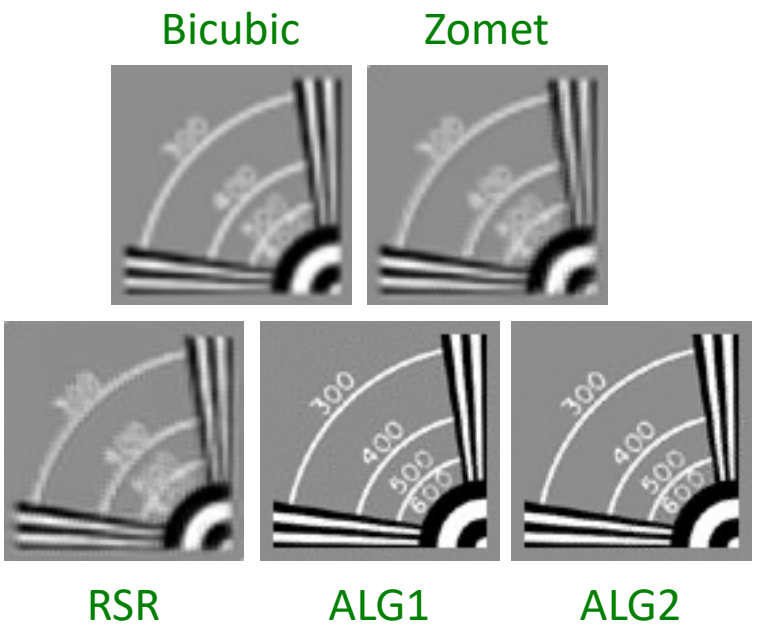
ALG2

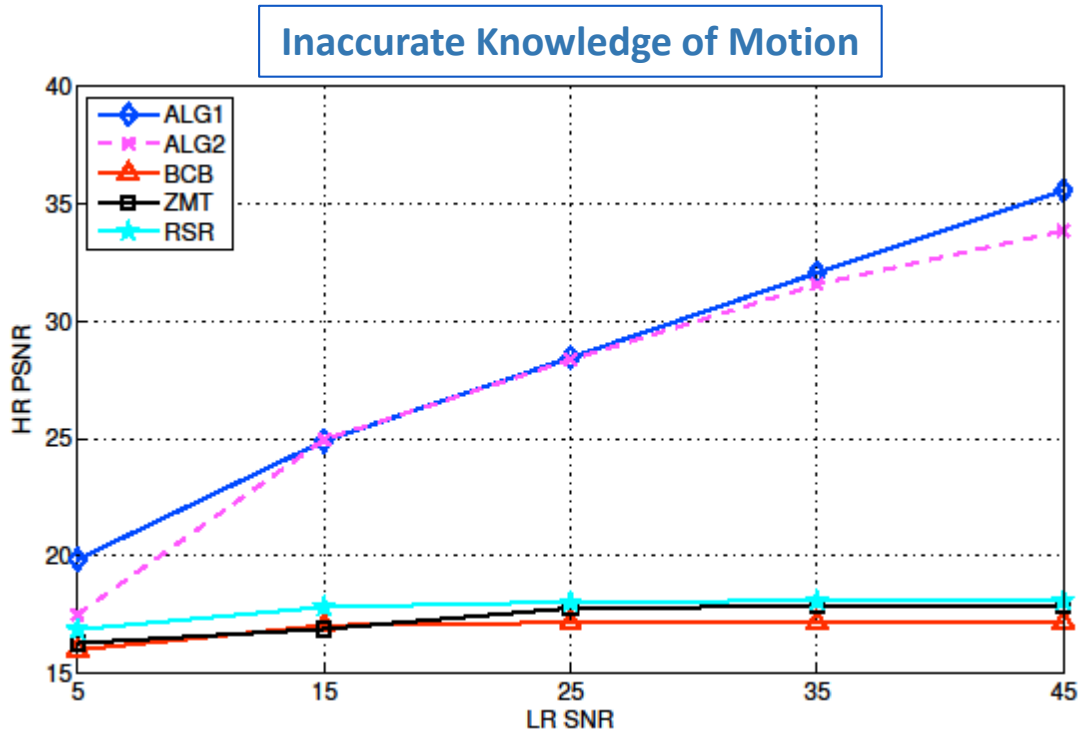
Exact Knowledge of Motion



Experiments

- Motion parameters corrupted by noise





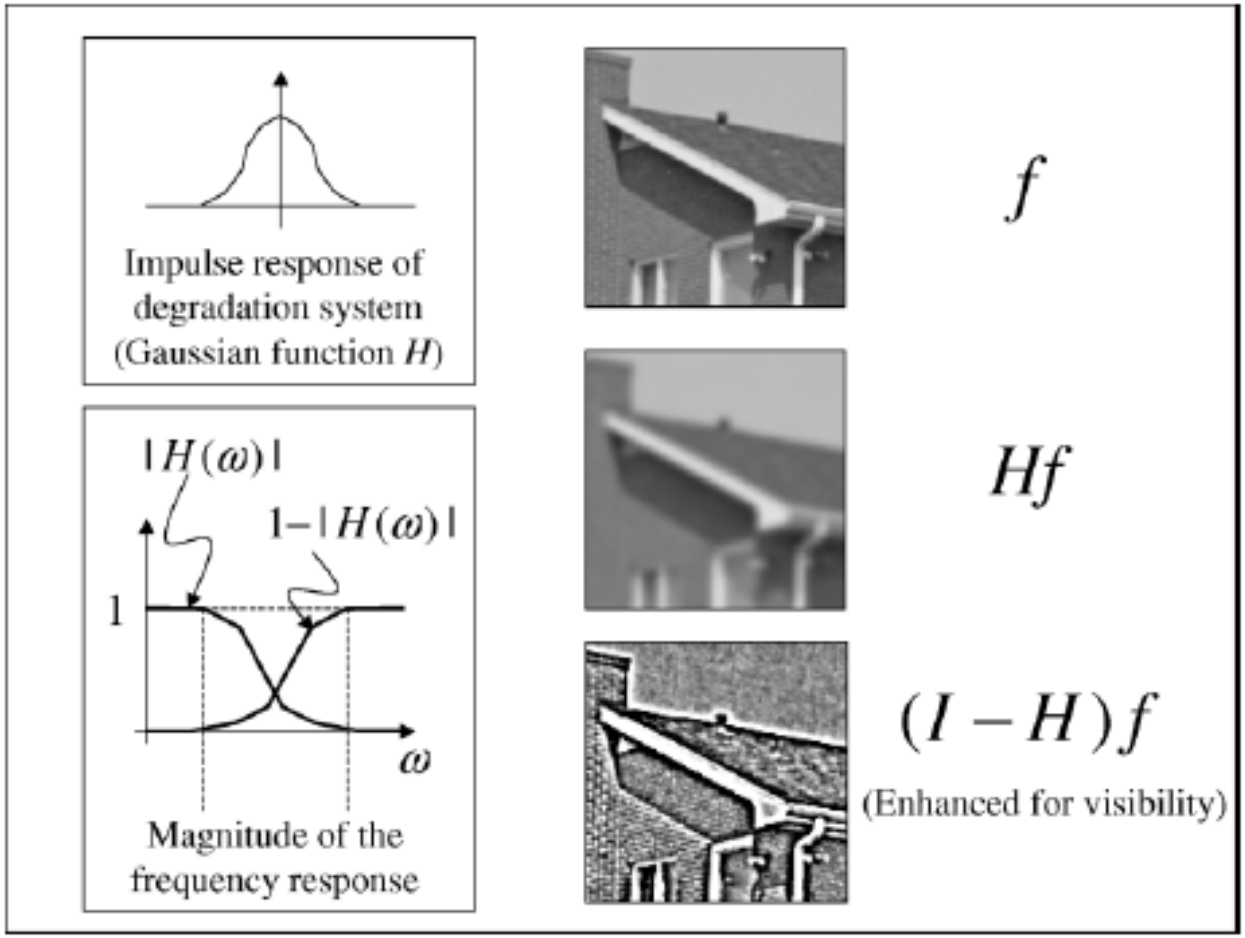
Lessons Learned

- Prior knowledge critical
  - Means to incorporate it into the problem (via modeling)
- Spatio-temporal adaptivity critical
- Automatic estimation of unknown parameters based on the data very important

Dictionary Techniques

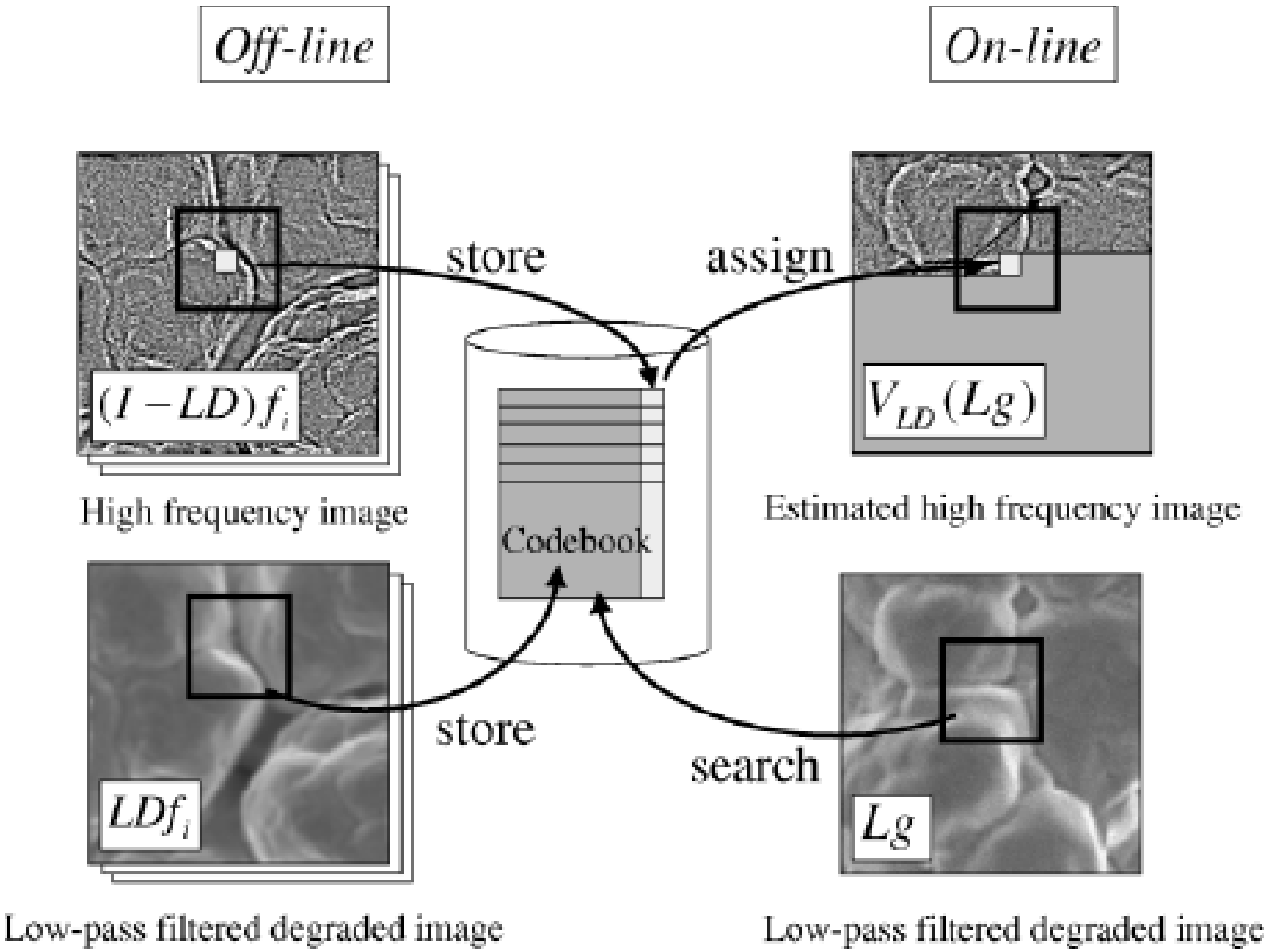
- Fixed vs Designed Dictionaries
- Complete or Overcomplete
- All prior knowledge incorporated by the dictionary
- Advances in dictionaries are coupled with advances in solving sparse problems
- Sparse representations are “extensions” of VQ
- “Related” to NN-based Techniques

A Learning Restoration Approach



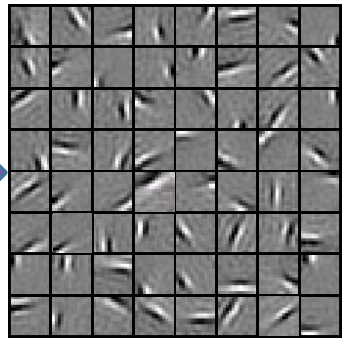
R. Nakagaki and A. K. Katsaggelos, "A VQ-Based Blind Image Restoration Algorithm," *IEEE Trans. Image Processing*, vol.12, no.9, pp. 1044-1053, Sept. 2003.

A Learning Restoration Approach

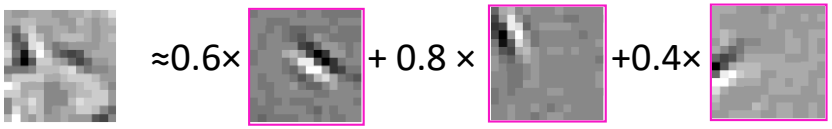


Dictionary Learning

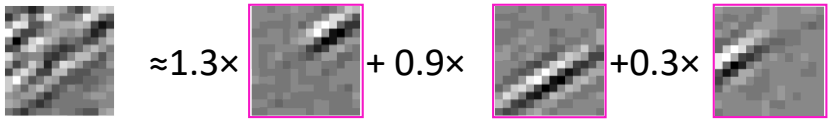
Training Images



Trained Dictionary



Represent as: [0, 0, …, 0, 0.6, 0, …, 0, 0.8, 0, …, 0, 0.4, …]



Represent as: [0, 0, …, 0, 1.3, 0, …, 0, 0.9, 0, …, 0, 0.3, …]

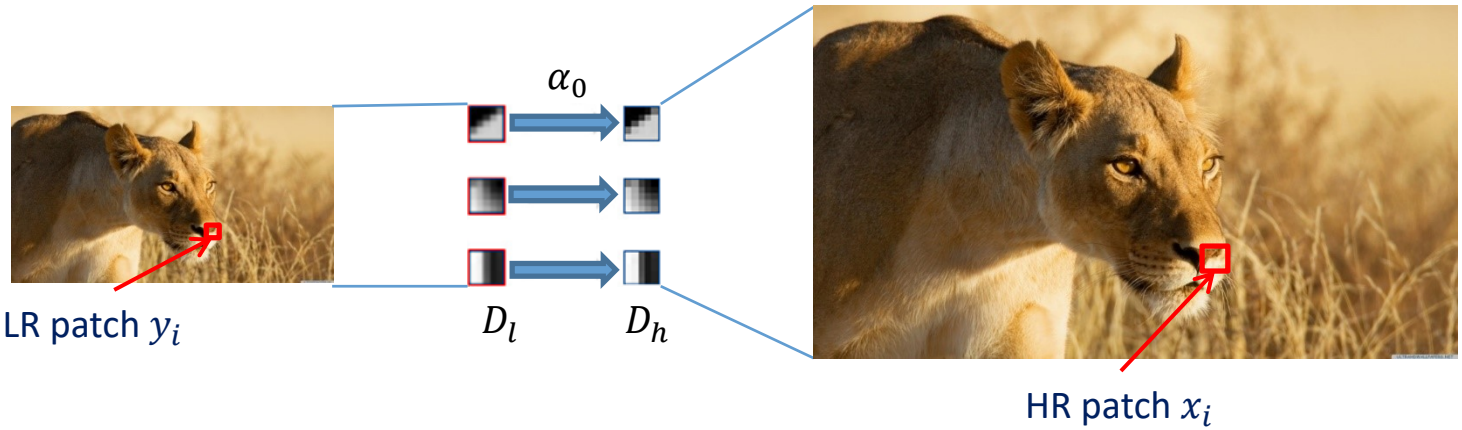
Training:

$$\min_{D,\{\alpha_i\}} \sum_{i=1}^{N} \|x_i - D\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1$$

Testing:

$$\min_{\alpha_i} \|x'_i - D\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1$$

Single Frame Super-Resolution



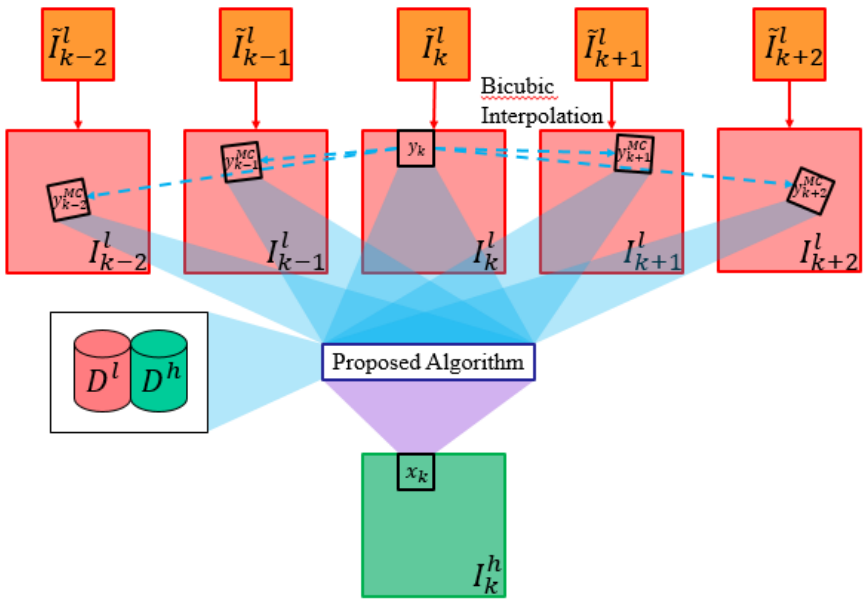
Training: $$\min_{D^h, D^l, a_i^{tr}} \sum_{i=1}^{N} \left\| x_i^{tr} - D^h a_i^{tr} \right\|_2^2 + \left\| y_i^{tr} - D^l a_i^{tr} \right\|_2^2 + \lambda \left\| a_i^{tr} \right\|_1$$

Reconstruction: $$\min_{z_i} \left\| y_i - D^l z_i \right\|_2^2 + \lambda \| z_i \|_1$$

$$x_i = D^h z_i$$

J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*19, no. 11 (2010): 2861-2873.

Batch Multiple-frame Super-Resolution



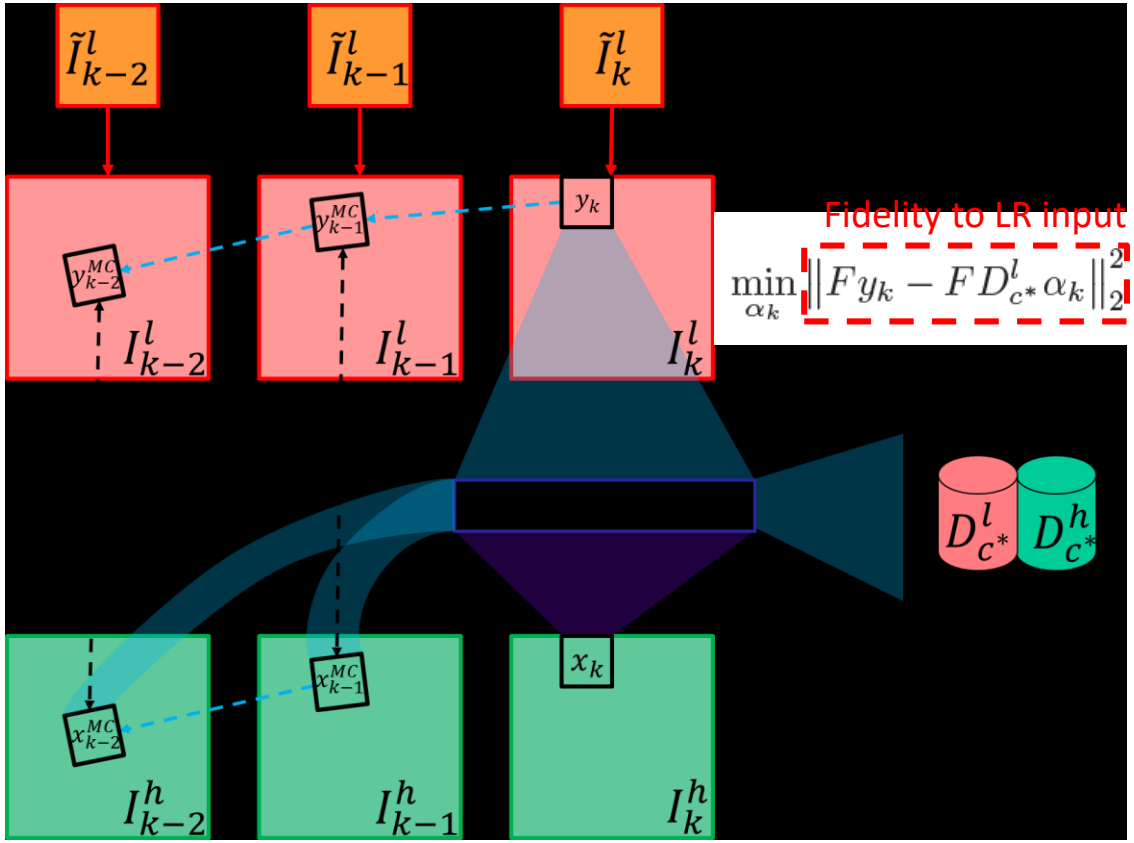
$$\min_{\substack{\alpha_k,\alpha^{MC}_{k+i}\\ i=-M,\dots,N,i\neq 0}} \left\|y_k - D^l\alpha_k\right\|_2^2 + \sum_{i=-M,i\neq 0}^{N} \left\|y^{MC}_{k+i} - D^l\alpha^{MC}_{k+i}\right\|_2^2 + \lambda\left(\|\alpha_k\|_1 + \sum_{i=-M,i\neq 0}^{N} \left\|\alpha^{MC}_{k+i}\right\|_1\right)$$

$$+ \sum_{i=-M,i\neq 0}^{N} \gamma_i \left\|D^h\alpha_k - D^h\alpha^{MC}_{k+i}\right\|_2^2$$

$$x_k = D^h\alpha_k$$

Q. Dai, S. Yoo, A. Kappeler, A. K. Katsaggelos, Sparse Representation Based Multiple Frame Video Super-Resolution, *IEEE Transactions on Image Processing*, Nov. 2016.

Recursive Multiple-frame Super-Resolution



Fidelity to LR input

Enforces Sparseness

$$\min_{\alpha_k} \left\| F y_k - F D^l_{c^*} \alpha_k \right\|_2^2 + \lambda \left\| \alpha_k \right\|_1 + \sum_{j=1}^{N} \gamma_j \left\| D^h_{c^*} \alpha_k - x^{MC}_{k-j} \right\|_2^2$$

Temporal Smoothness

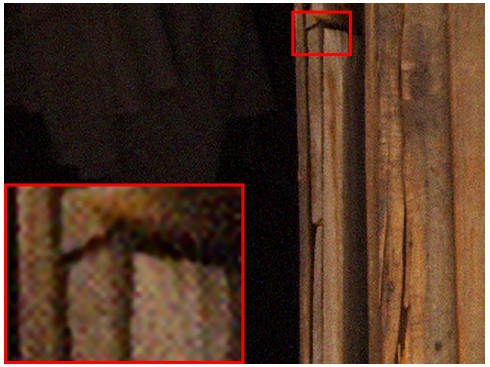
$$x_k = D^h_{c^*} \alpha_k$$

Experimental Results

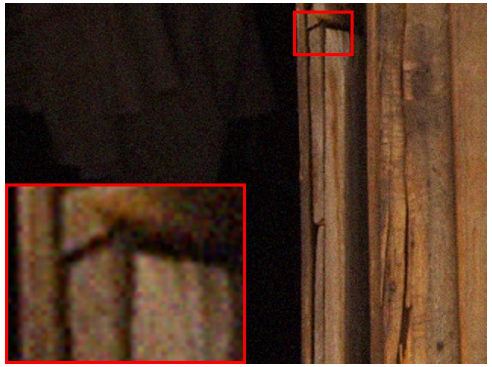
Video	Bicubic	Bilevel [14], [15]	NE+NNLS [26]	NE+LLE [27]	ANR [28]
Scene 2	45.27	46.12	46.85	46.53	46.91
Scene 8	38.18	39.94	40.04	41.08	40.27
Scene 18	41.43	43.04	43.18	43.28	43.45
Scene 25	44.40	46.69	46.69	47.38	47.85
Scene 33	40.22	42.95	42.70	43.37	43.59
Scene 45	42.43	43.72	43.62	43.89	44.11
Scene 48	33.90	36.10	35.78	36.24	36.39
Average	40.83	42.65	42.69	43.11	43.22

Video	SR-CNN [29]	Enhancer [30]	Bayesian [5]	MDMF-B-VT	MDMF-R-VT
Scene 2	47.41	46.01	46.49	**48.26**	48.11
Scene 8	41.08	40.26	39.94	41.62	**41.72**
Scene 18	43.94	43.66	43.19	44.73	**44.94**
Scene 25	48.29	46.21	46.08	49.57	**49.68**
Scene 33	43.83	43.00	42.51	45.33	**45.44**
Scene 45	44.32	43.46	43.70	44.76	**44.91**
Scene 48	37.48	35.51	34.10	**37.56**	37.31
Average	43.76	42.59	42.29	44.55	**44.59**

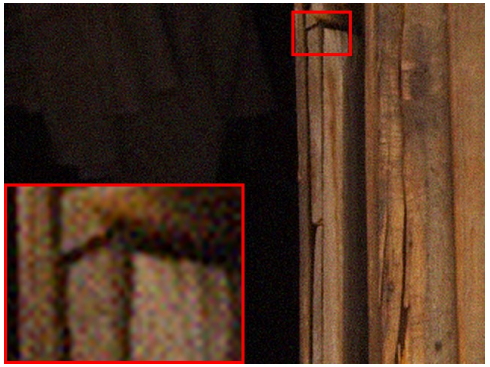
Experimental Results



SRCNN

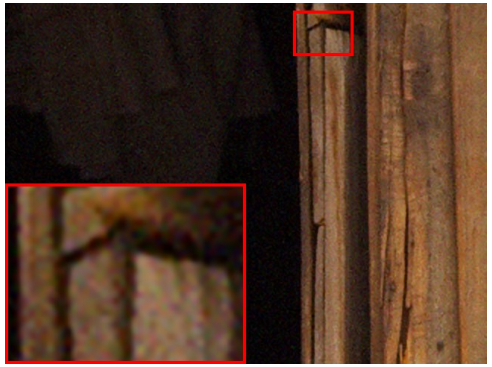


Enhancer



DraftCNN

MDMF-B-VT



MDMF-R-VT

Ground Truth

Computational X-Ray Fluorescence (XRF) Imaging using Dictionary Learning

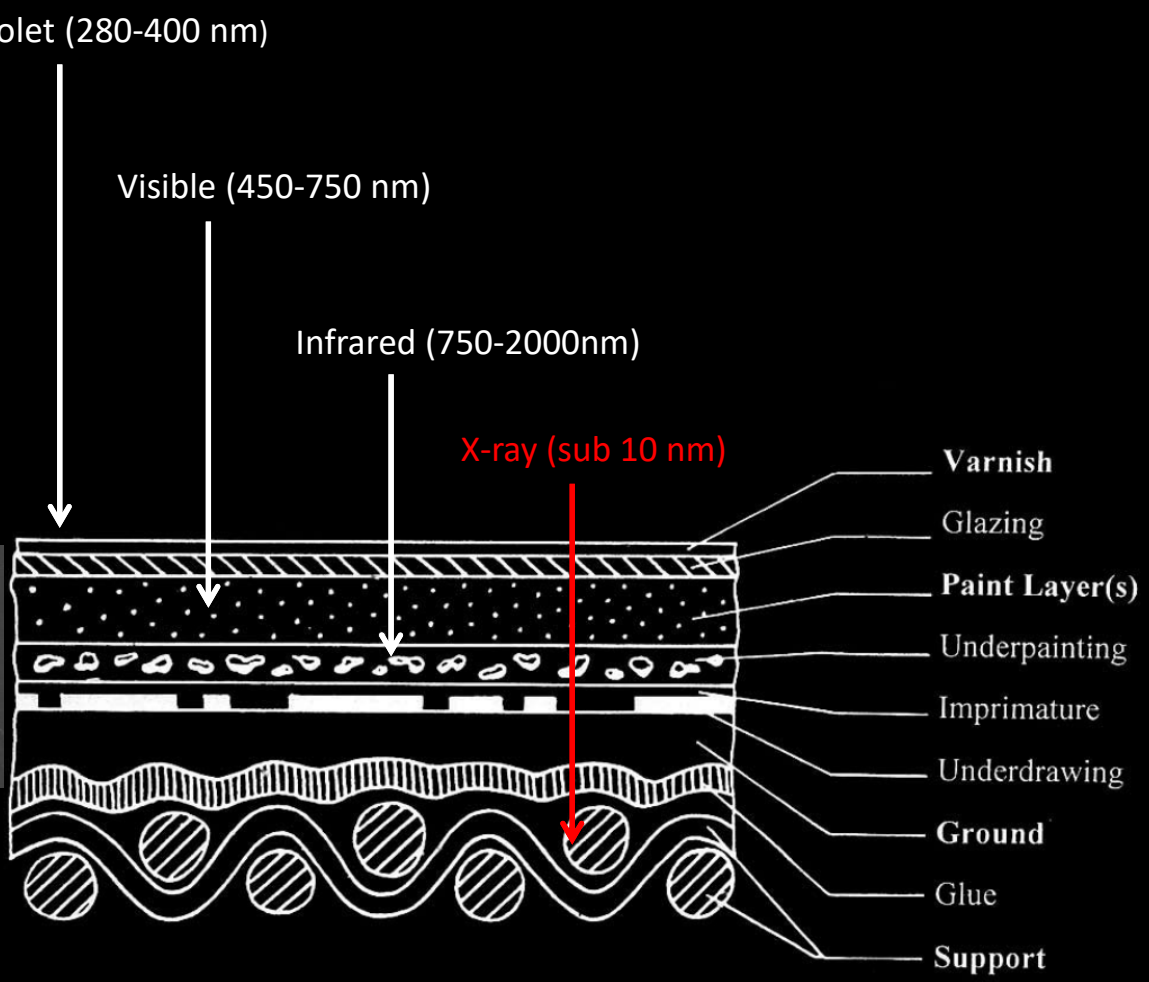
Q. Dai et al, “Computational X-Ray Fluorescence Imaging using Dictionary Learning,” *IEEE Trans. on Computational Imaging*, Sept. 2017.

How a Painting is Structured

Final details include the depiction of jewels in the haloes, the white rose, and the mordanted star on the Madonna's shoulder.

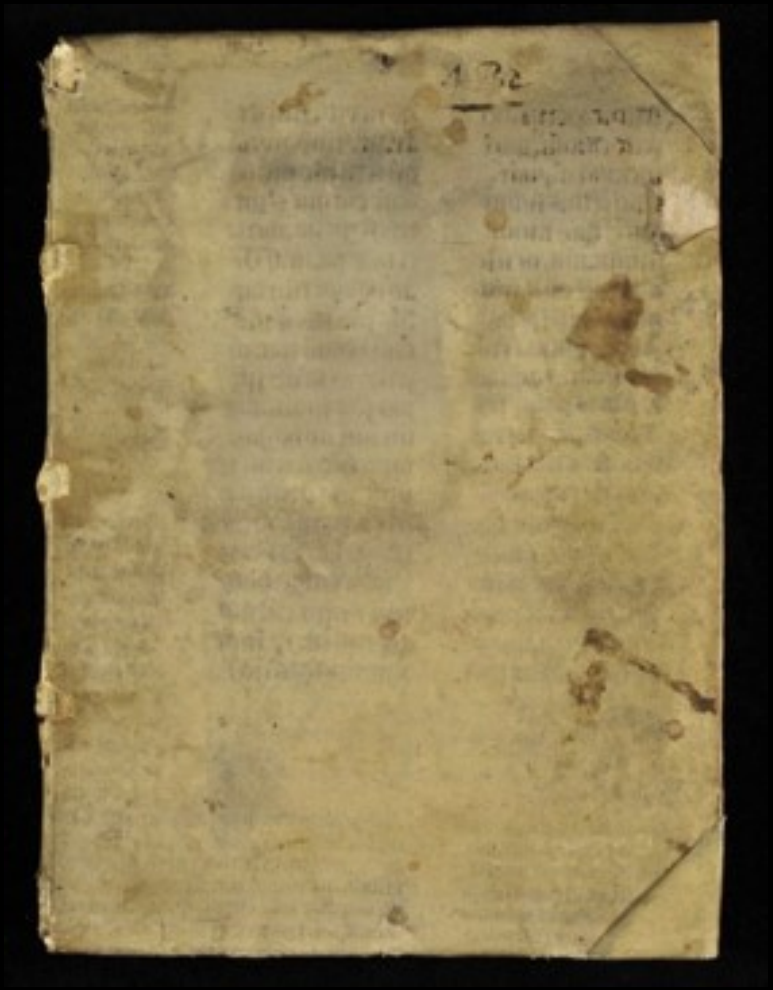
http://www.artcons.udel.edu/about/kress/historic-painting-reconstructions/giotto-di-bondone

Why Imaging of Art?
Different modalities to interrogate structure

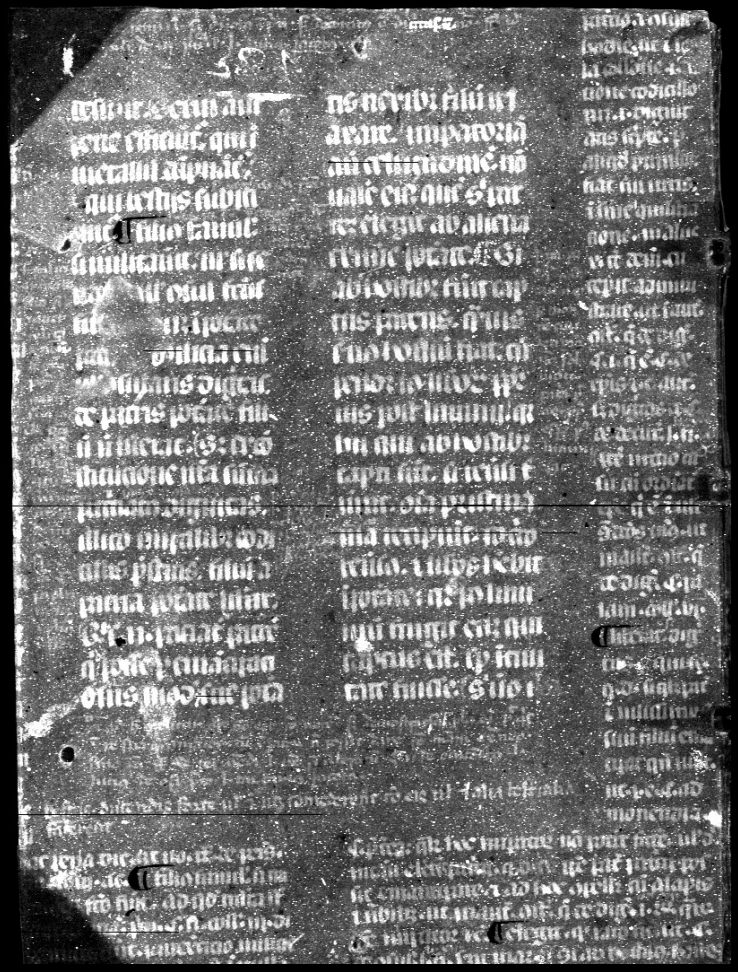


XRF maps let you see the invisible!

Hesiod Manuscript (RGB Image)



XRF Map (Iron)



Beam size –Step size 200x200µm²
Dwell time 10ms/point - Att (6)

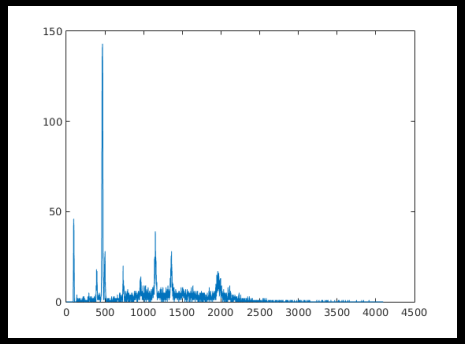
XRF Re-colorization of Hidden Layers

Rembrandt, *An Old Man in Military Costume* (1630), J. Paul Getty Museum, 78.PB.246

Trentelman, Karen, et al. "*Rembrandt's An Old Man in Military Costume*: the underlying image re-examined." *Applied Physics A* 121.3 (2015): 801-811.

XRF Scanner Advantages & Disadvantages

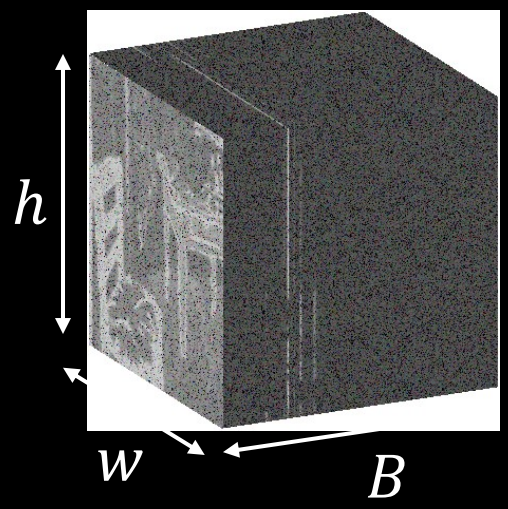
- Advantages
  - Deep X-Ray penetration into hidden layers
  - Elemental composition for chemical characterization and provenance analysis
  - High spectral resolution (4096 channels)
- Disadvantages
  - Low Signal-to-Noise-Ratio (SNR)
  - Slow raster scan process -> Low spatial resolution



A typical XRF scan takes 2sec/pixel -> 20,000 s (6hr) for a 100x100 XRF map)

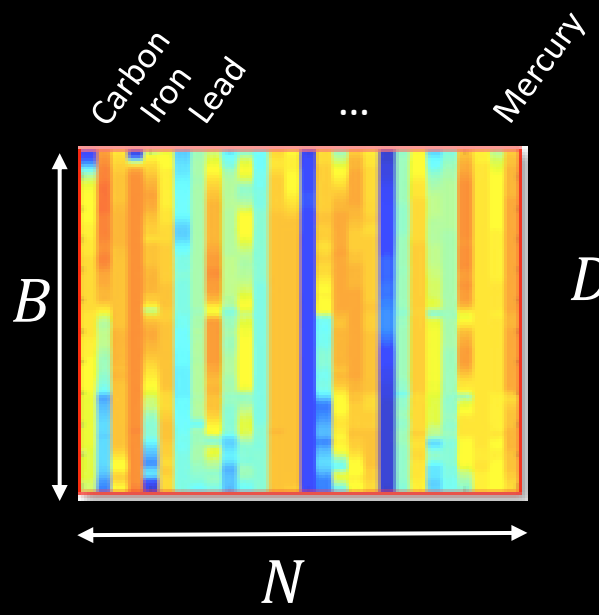
How can we make scanning Faster?

Captured XRF Datacube



$X \in \mathbb{R}^{h \cdot w \times B}$

Elemental XRF Dictionary

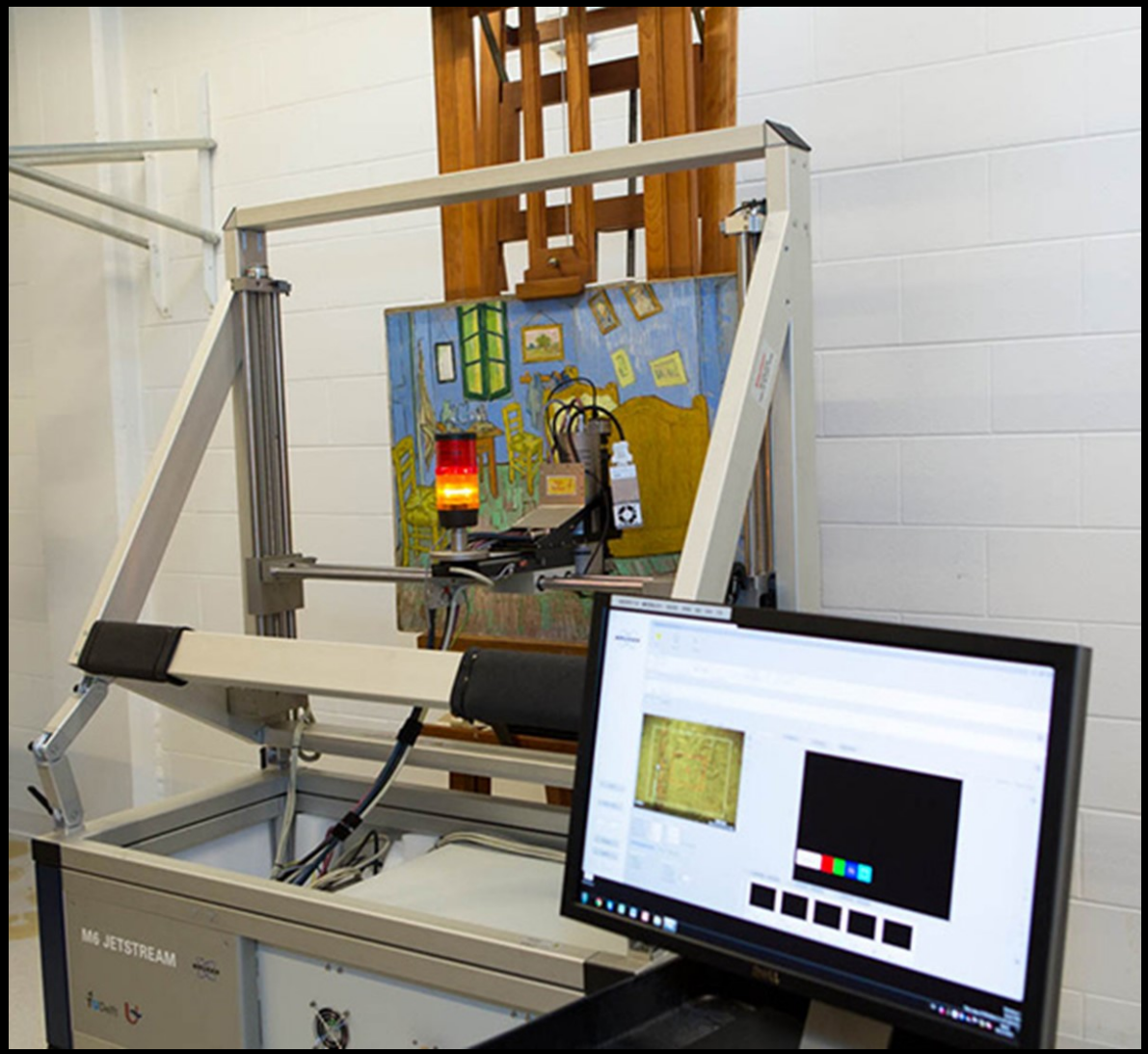


$D^{xrf} \in \mathbb{R}^{B \times N}$

<u>Observation:</u> The XRF Spectra span a small subspace in $\mathbb{R}^B$
<u>Approach</u>: Use dictionary methods to model XRF spectra using a linear basis decomposition

Vincent Van Gogh, *The Bedroom* (1889)
Art Institute of Chicago 1926.417

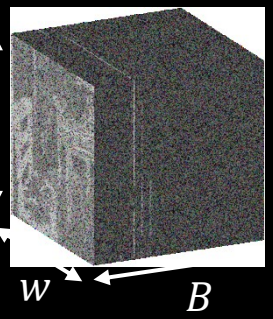
XRF Scanning of Van Gogh's Bedroom



XRF Super-Resolution + RGB Image Fusion

XRF Scanner

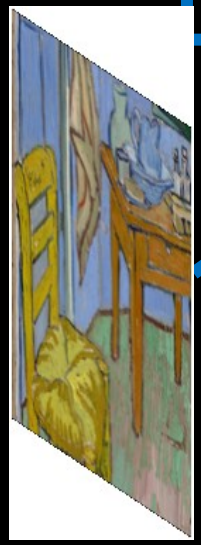
Captured XRF Cube



Super-resolution + Fusion algorithm

Super-resolution HR XRF image

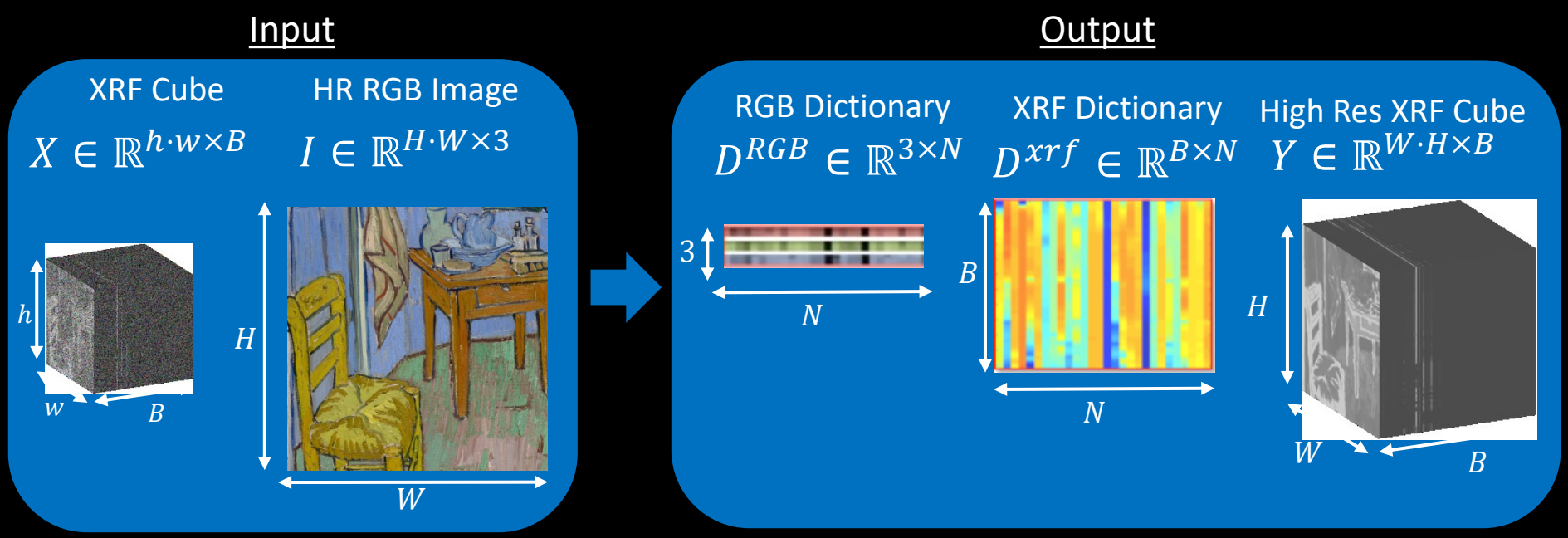
Conventional Camera



Conventional RGB image

Vincent Van Gogh, *The Bedroom* (1889)
Art Institute of Chicago 1926.417

RGB Fusion using XRF Dictionaries



Superresolution + Fusion:

$$D^{rgb^*}, D^{xrf^*}, A^* = \min_{D^{rgb}, D^{xrf}, A} \left\| I - D^{rgb} A \right\|_F^2 + \left\| X - D^{xrf} A S \right\|_F^2$$

$$Y_{SR} = D^{xrf^*} A^*$$

Coupled Sparse Coefficients

Downsamping Matrix

Fusion Results

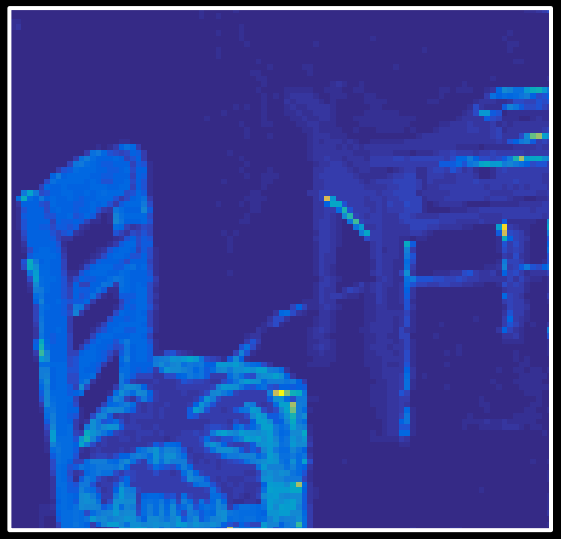
38cm

38cm

Area scanned using MA-XRF

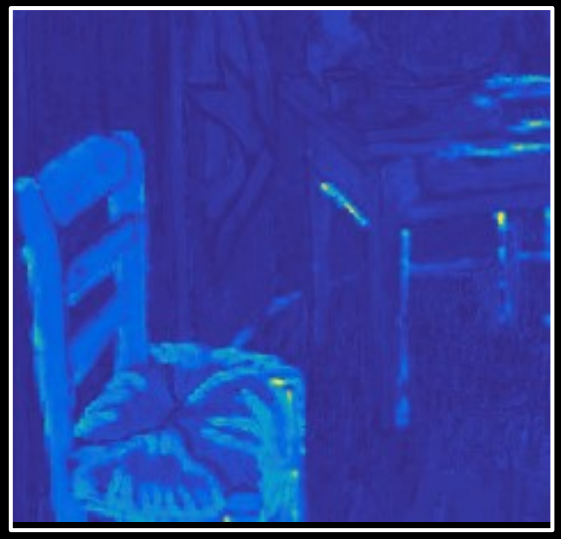
38x38cm²
Step size: 0.400 cm

Captured XRF Cube X



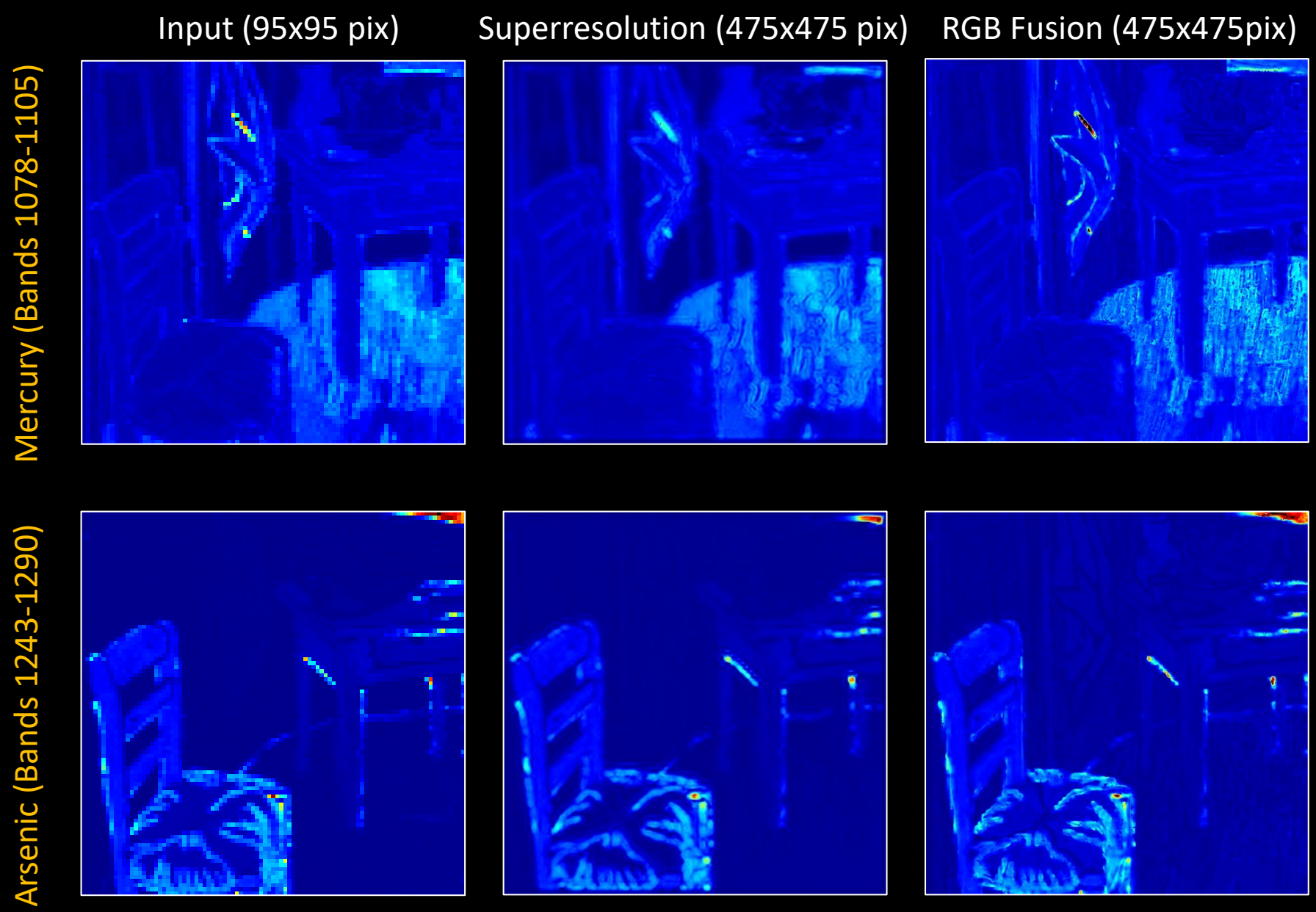
95x95 Pixels

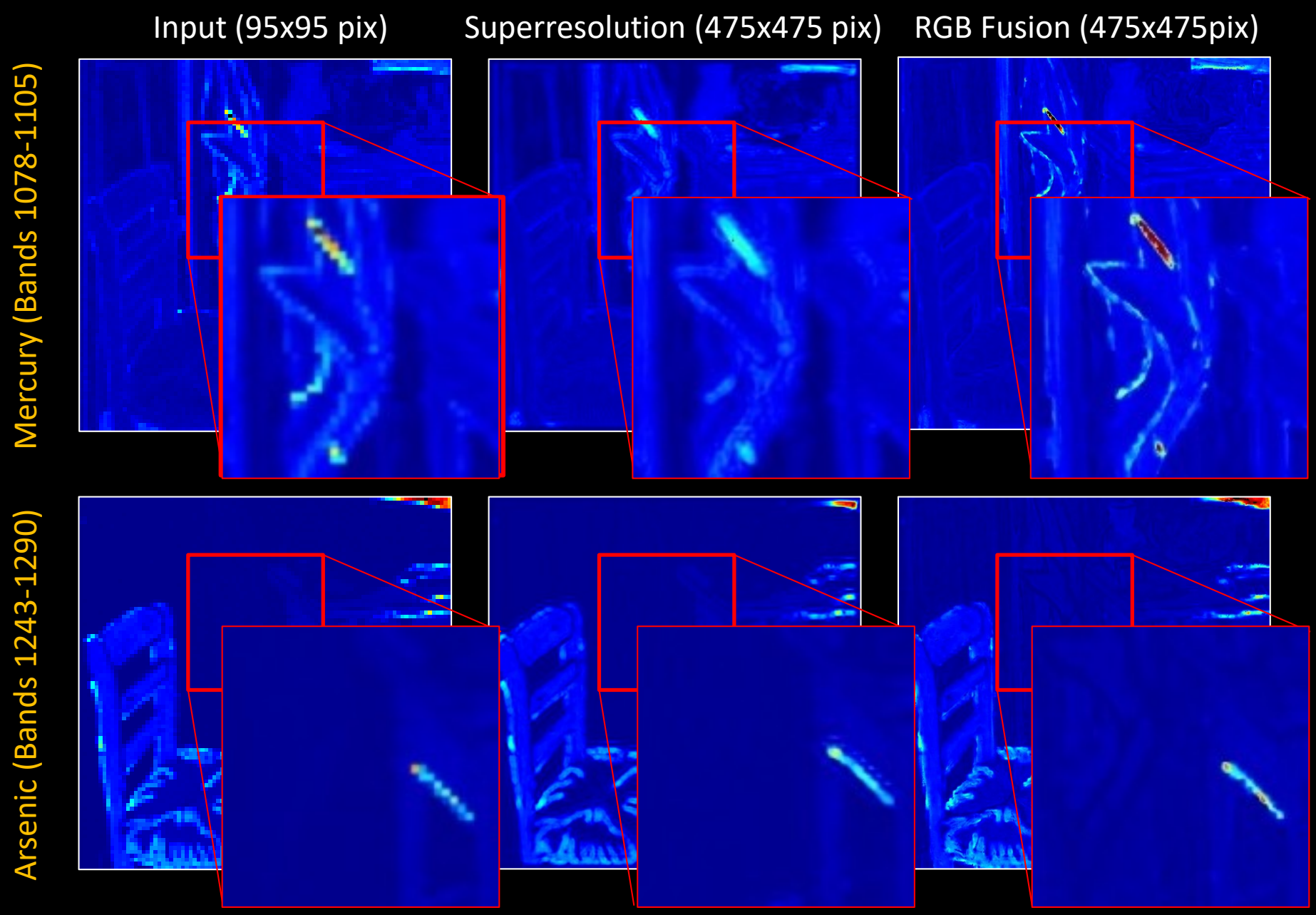
Fused XRF Cube Y

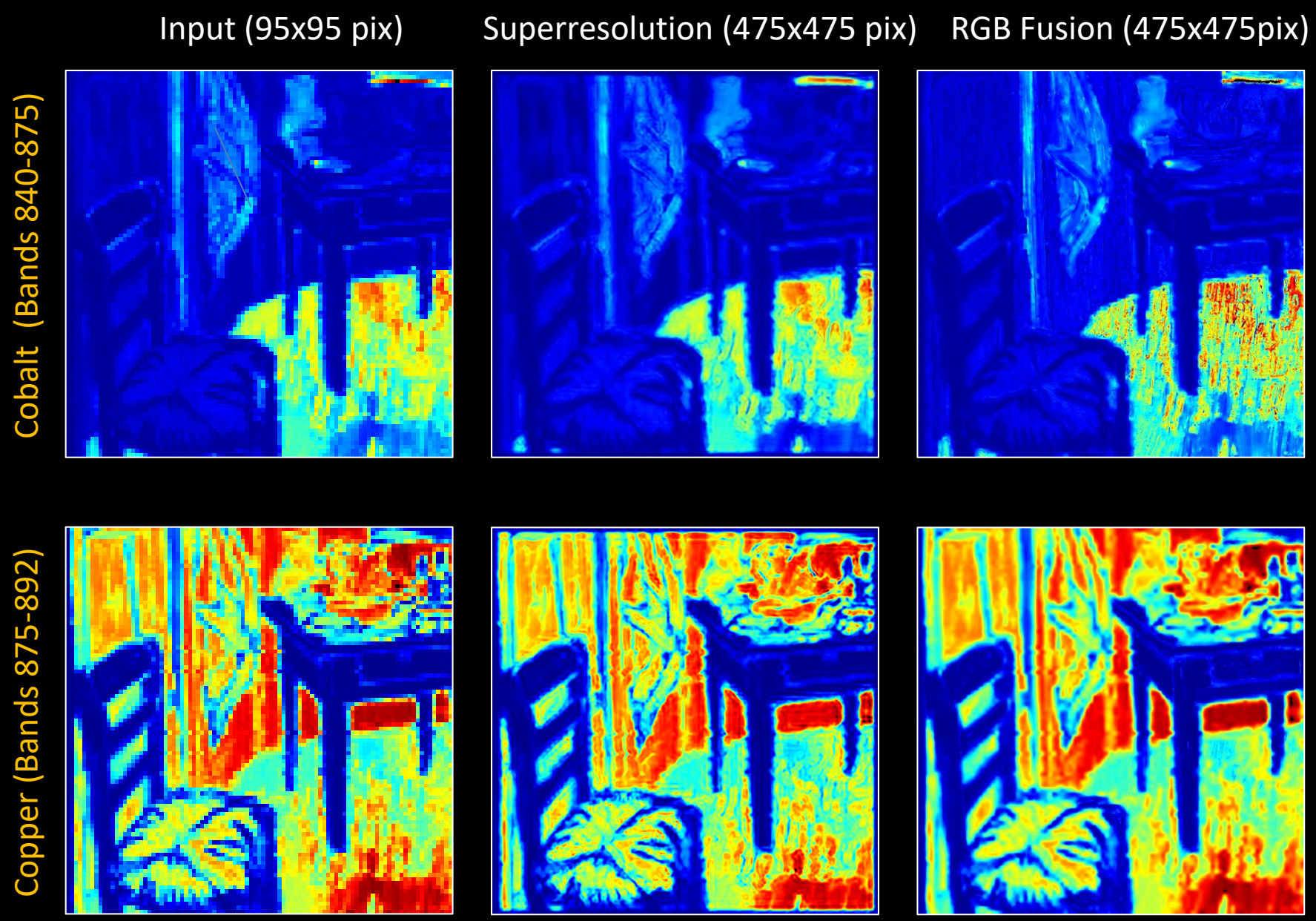


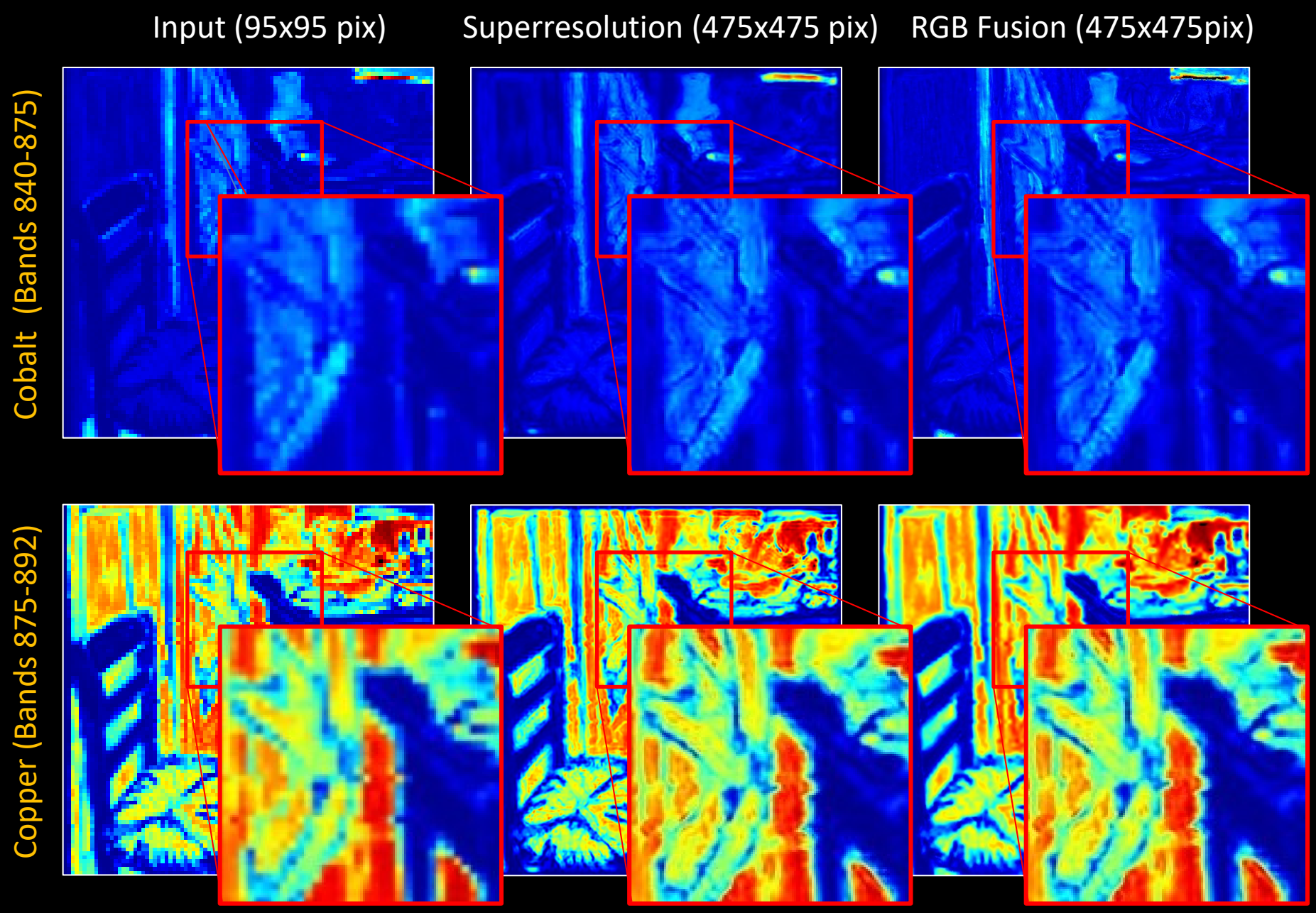
475x475 Pixels

CrK Peak, Bands 611-657

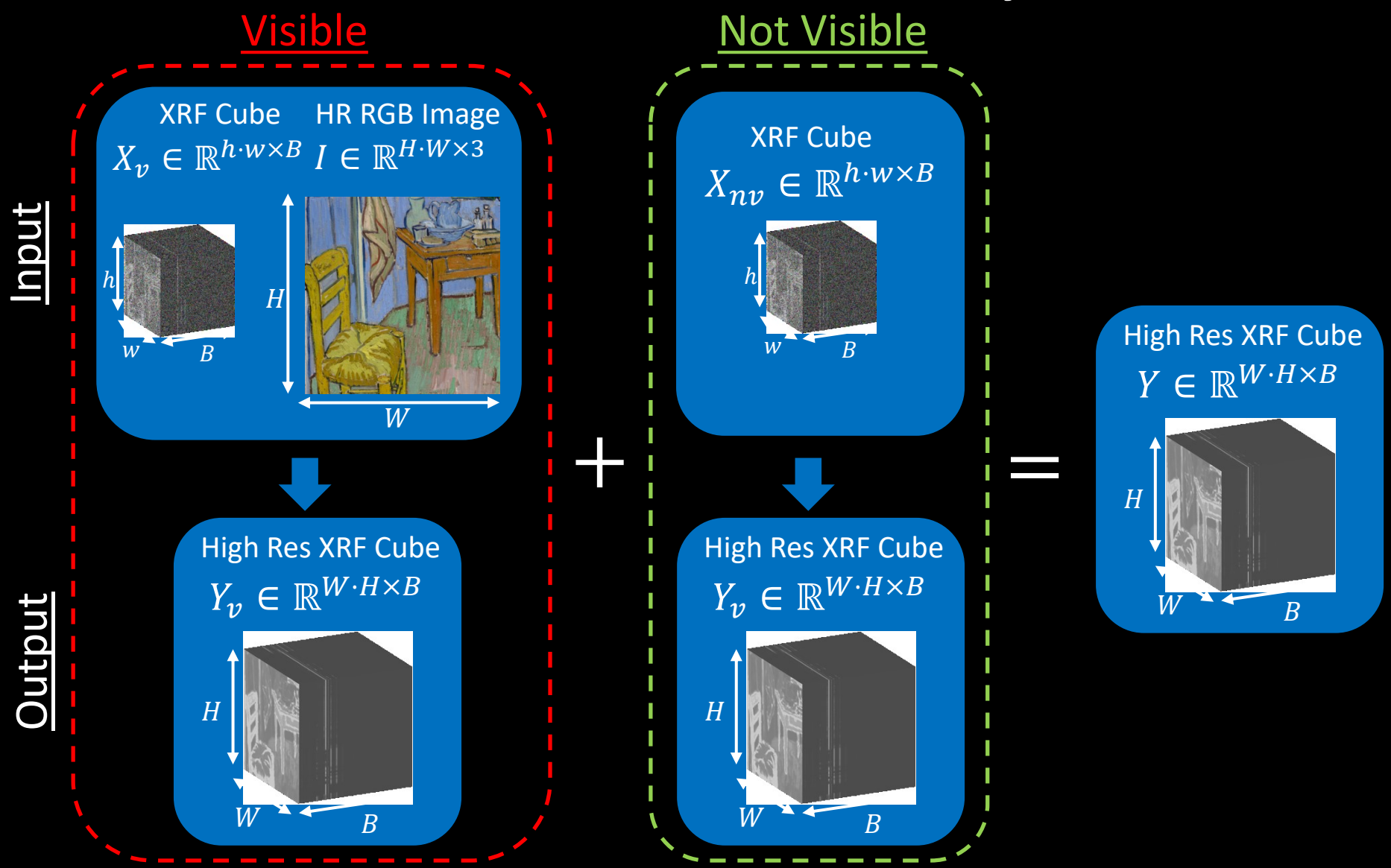

Input (95x95 pix)
Superresolution (475x475 pix)
RGB Fusion (475x475pix)
Mercury (Bands 1078-1105)
Arsenic (Bands 1243-1290)


Input (95x95 pix)
Superresolution (475x475 pix)
RGB Fusion (475x475pix)
Mercury (Bands 1078-1105)
Arsenic (Bands 1243-1290)


Input (95x95 pix)
Superresolution (475x475 pix)
RGB Fusion (475x475pix)
Cobalt (Bands 840-875)
Copper (Bands 875-892)


Input (95x95 pix)
Superresolution (475x475 pix)
RGB Fusion (475x475pix)
Cobalt (Bands 840-875)
Copper (Bands 875-892)

What about hidden layers?



RGB Fusion Model for Hidden Layers

Inputs

X: Low resolution input XRF image
X_v: Visible part of X
X_{nv}: Non-visible part of X
I: Input high resolution RGB image
S: Spatial down-sample matrix

Outputs

D_v^{xrf}: XRF dictionary (visible)
D_{nv}^{xrf}: XRF dictionary (non-visible)
D^{rgb}: RGB dictionary
A_v: Coefficient of visible component
A_{nv}: Coefficient of non-visible component

Superresolution + Fusion + V/NV:

$$D^{rgb^*}, D_{nv}^{xrf^*}, D_v^{xrf^*}, A_v^*, A_{nv}^* = \min_{D^{rgb}, D_v^{xrf}, D_{nv}^{xrf}, A_v, A_{nv}}$$

Non- Visible

$$\left\|X_{nv} - D_{nv}^{xrf} A_{nv} S\right\|_F^2 + \left\|\nabla(D_{nv}^{xrf} A_{nv})\right\|_F^2$$

Non-Visible
Total Variation
Superresolution

Visible

$$\left\|I - D^{rgb} A_v\right\|_F^2 + \left\|X_v - D_v^{xrf} A_v S\right\|_F^2$$

Reconstruction:

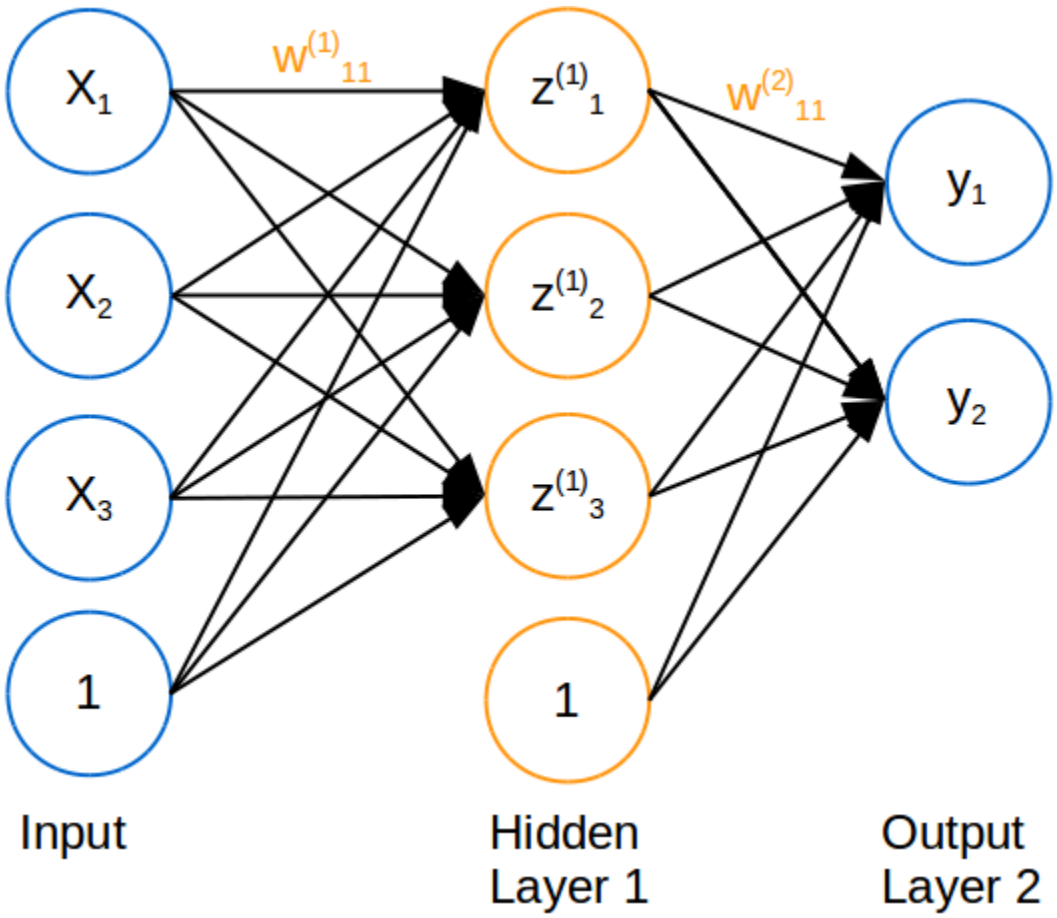
$$Y = D_{nv}^{xrf^*} A_{nv}^* + D_v^{xrf^*} A_v^*$$

Visible
Coupled Coefficients

Extensions

- Semi-coupled dictionaries (sparse representations are linearly related)
- Sparse representations share the same support
- Manifold similarity
- Self-dictionaries
- Multi-Scale dictionaries

Neural Network



Function Approximation

$$f_m(\mathbf{x}) = \tanh\left(c_m + \mathbf{x}^T \mathbf{v}_m\right) \quad \text{for all } m \geq 1.$$

$$f_m(\mathbf{x}) = \tanh\left(c_m^{(1)} + \sum_{m_2=1}^{M_2} \tanh\left(c_{m_2}^{(2)} + \mathbf{x}^T \mathbf{v}_{m_2}^{(2)}\right) v_{m_2,m}^{(1)}\right)$$

$$f_m(\mathbf{x}) = \max\left(0,\ c_m^{(1)} + \sum_{m_2=1}^{M_2} \max\left(0,\ c_{m_2}^{(2)} + \sum_{m_3=1}^{M_3} \max\left(0,\ c_{m_3}^{(3)} + \mathbf{x}^T \mathbf{v}_{m_3}^{(3)}\right) v_{m_3,m_2}^{(2)}\right) v_{m_2,m}^{(1)}\right)$$

Function Approximation

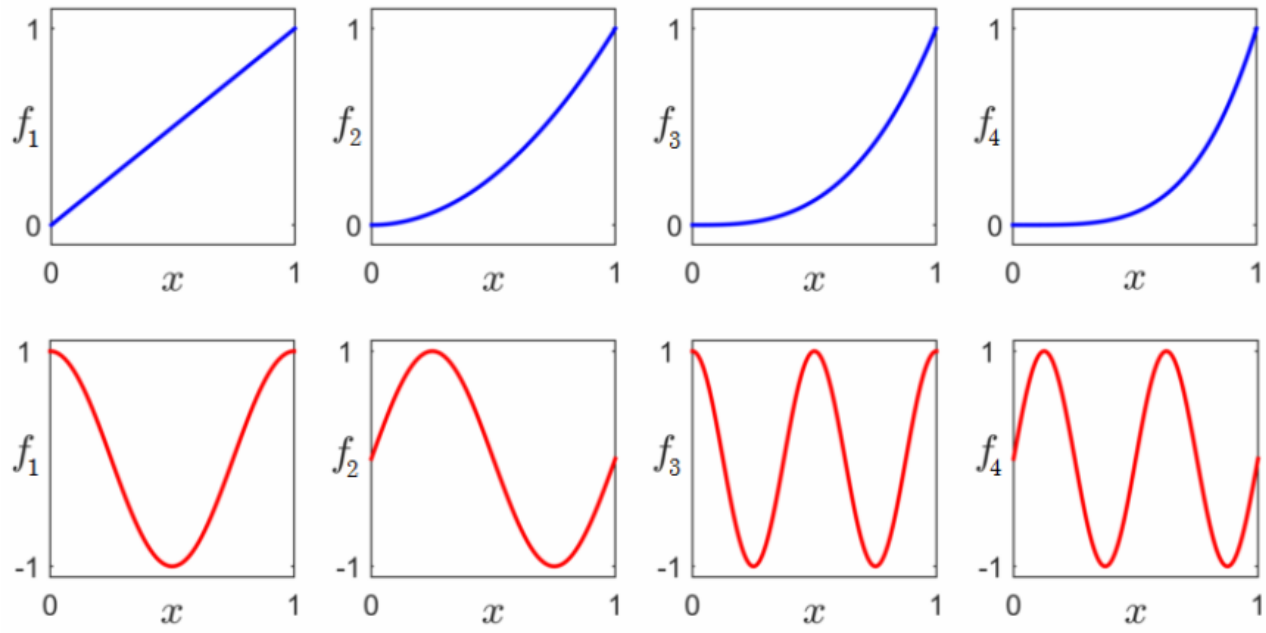


Figure 5.4. *(top row, from left to right) The first four non constant elements of the polynomial basis. (bottom row, from left to right) The first four non constant elements of the Fourier basis.*

J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine Learning Refined*, Cambridge University Press, 2016.

Function Approximation

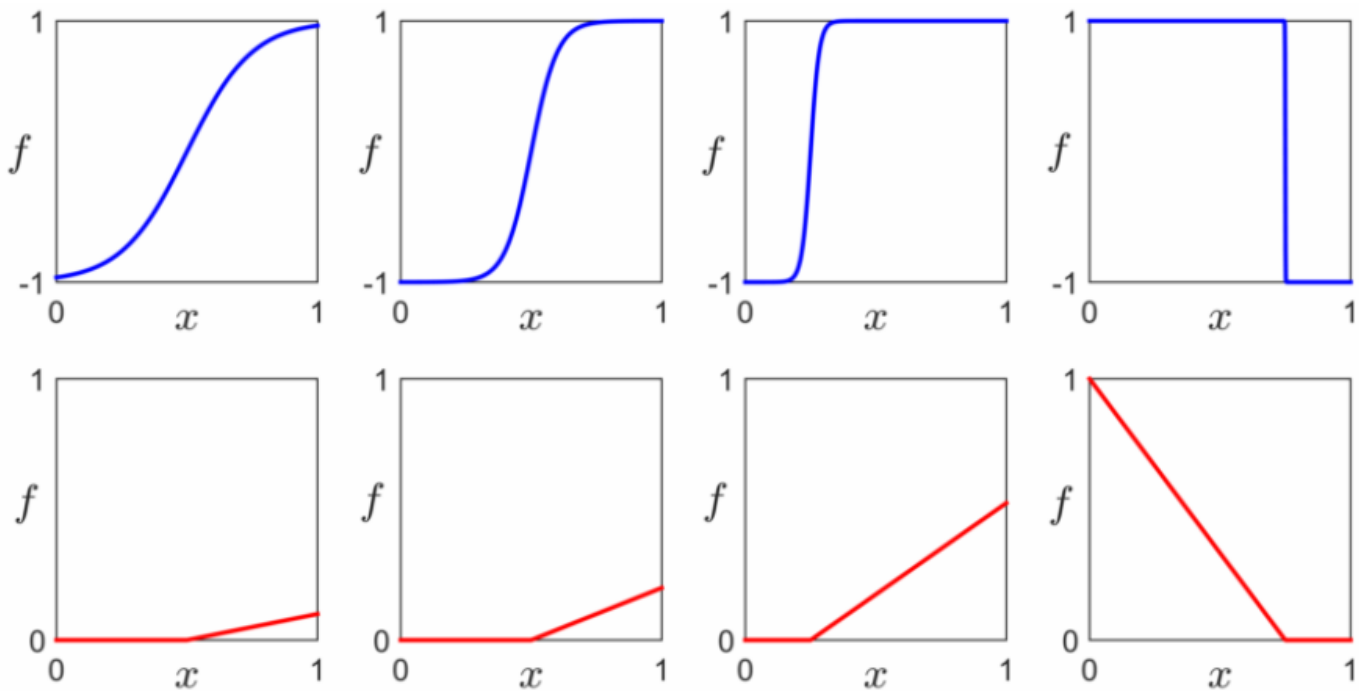


Figure 5.5. *Unlike a fixed basis, elements of an adjustable basis are free to change form by adjusting internal parameters. Here we show four instances of a single (top row) hyperbolic tangent and (bottom row) hinge basis function, with each instance corresponding to a different setting of its internal parameters.*

Function Approximation

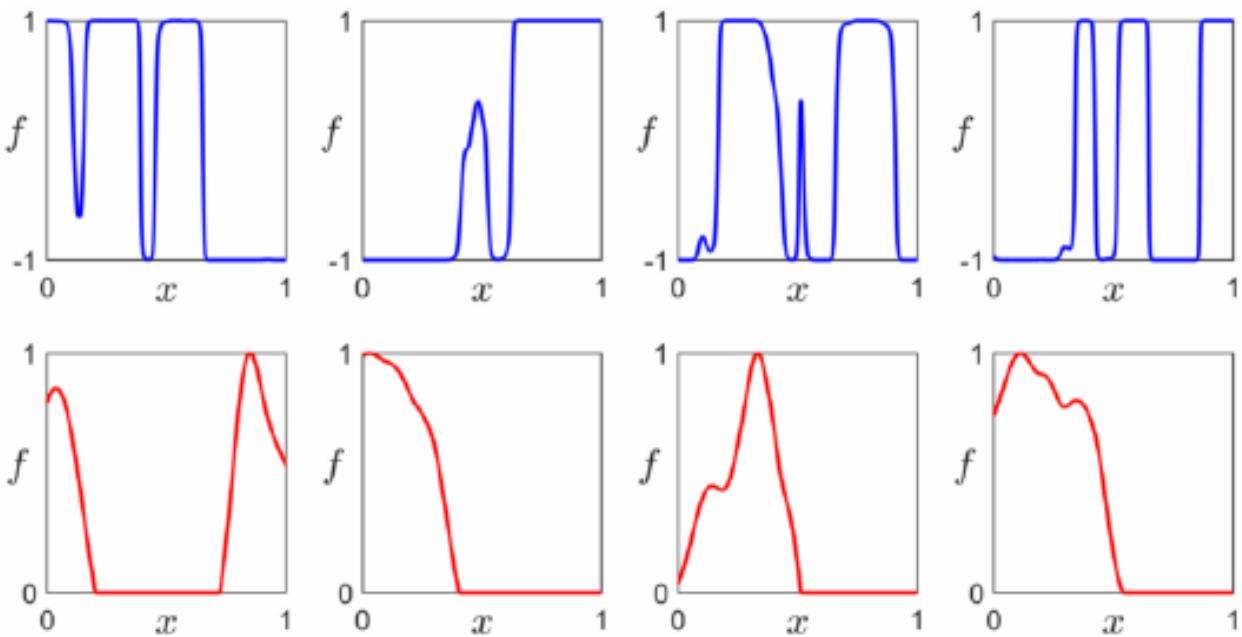


Figure 5.7. *Four instances of a two layer neural network basis function made by composing (top row) hyperbolic tangent and (bottom row) hinge functions. In each instance the internal parameters are set randomly. Note how the two layer basis elements are far more diverse in shape than those of a single layer basis shown in Figure 5.5.*

Function Approximation

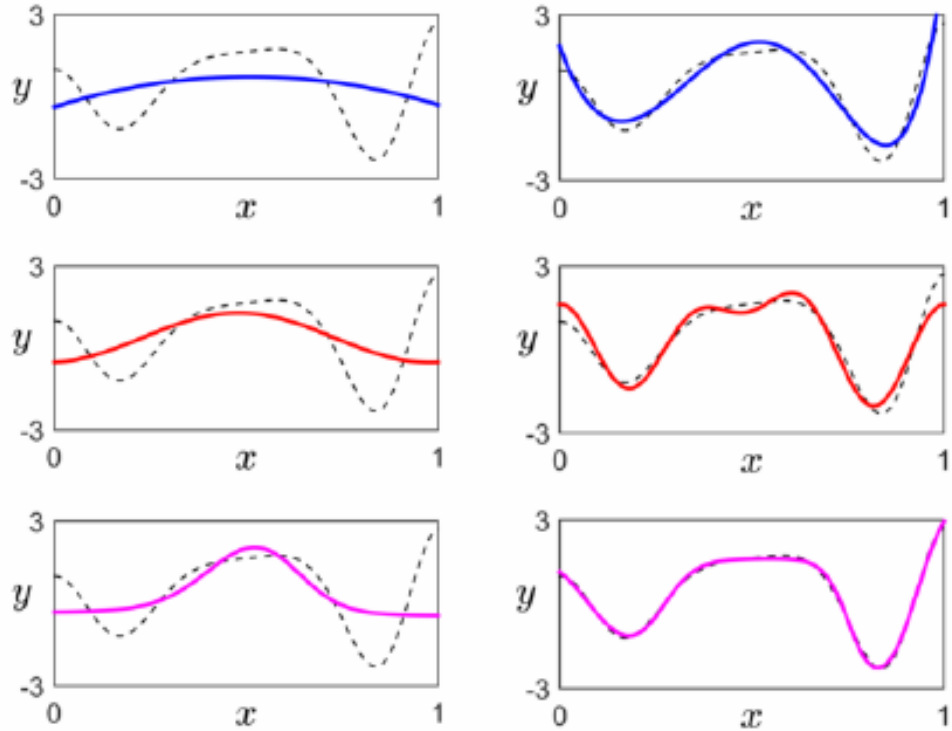
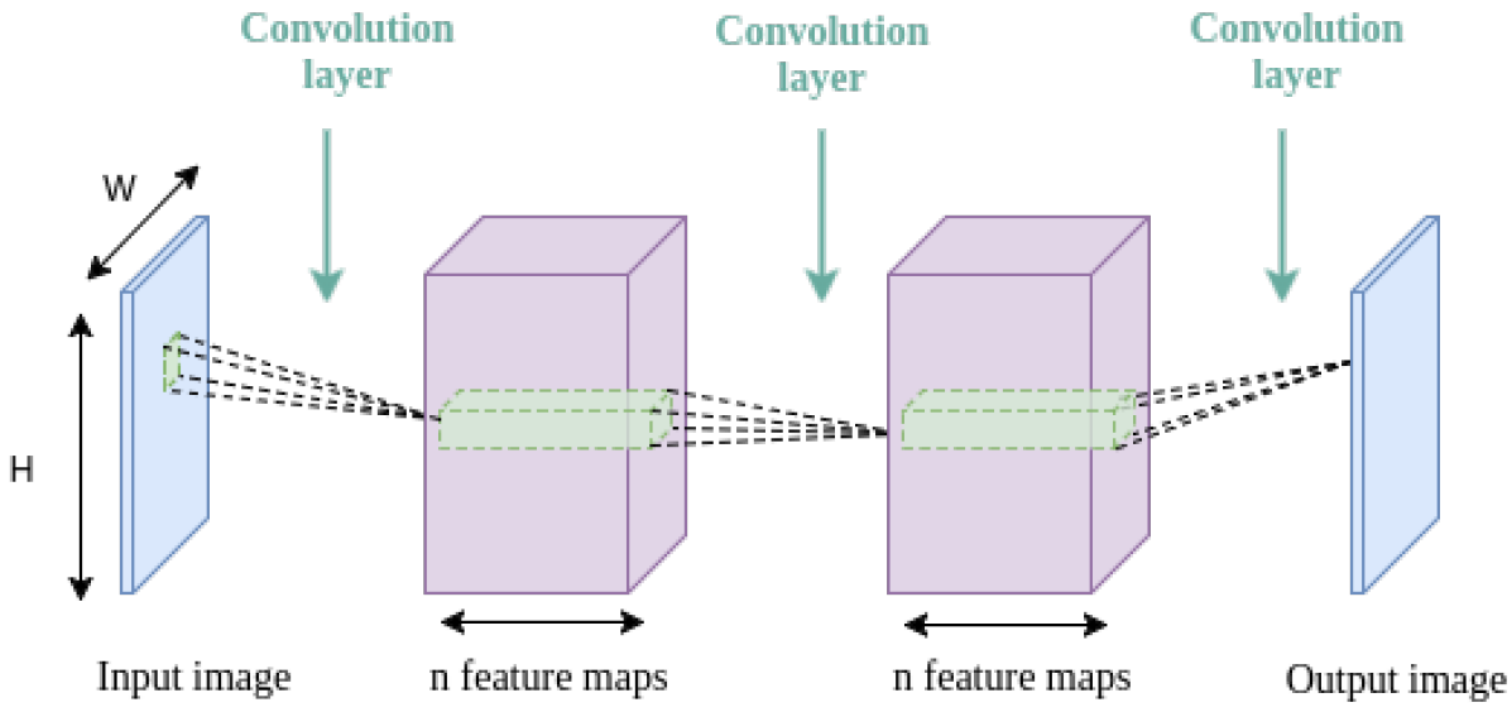


Figure 5.6. *From left to right, approximation of a continuous function (shown by the dashed black curve) over* $[0, 1]$*, using* $M = 2$ *and* $M = 6$ *elements of (top row) polynomial, (middle row) Fourier, and (bottom row) single hidden layer neural network bases, respectively. While all three bases could approximate this function as finely as desired by increasing* M*, the neural network basis (with its adjustable internal parameters) approximates the underlying function more closely using the same number of basis elements compared to both fixed bases.*

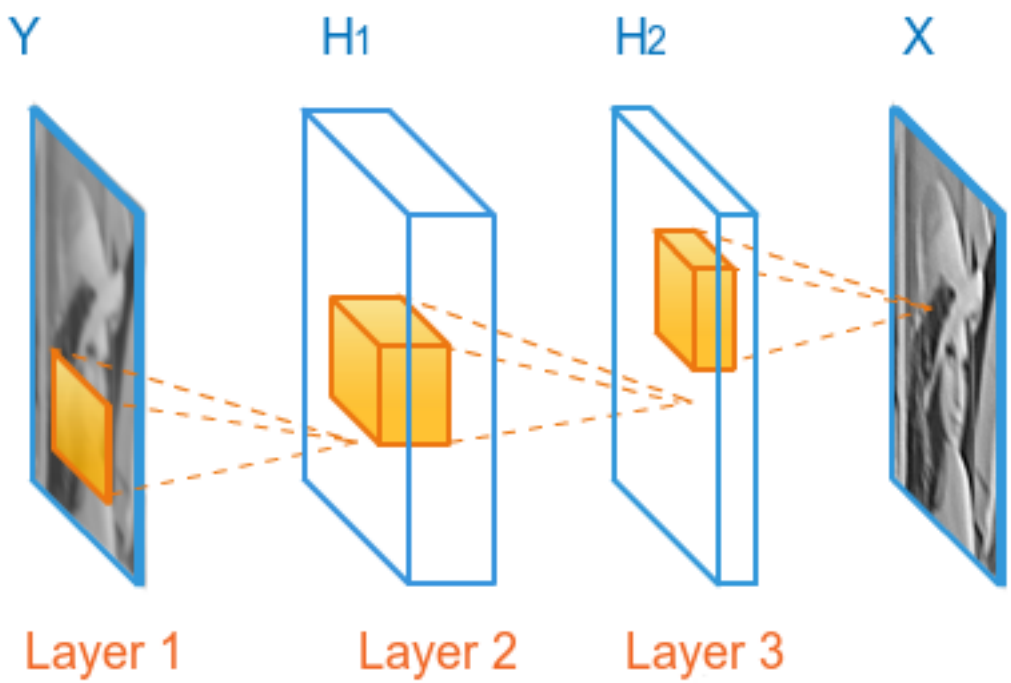
Convolutional Neural Networks

- CNNs offer multiple advantages over fully-connected NNs:
  - Smaller number of parameters
  - Translation invariance and locality



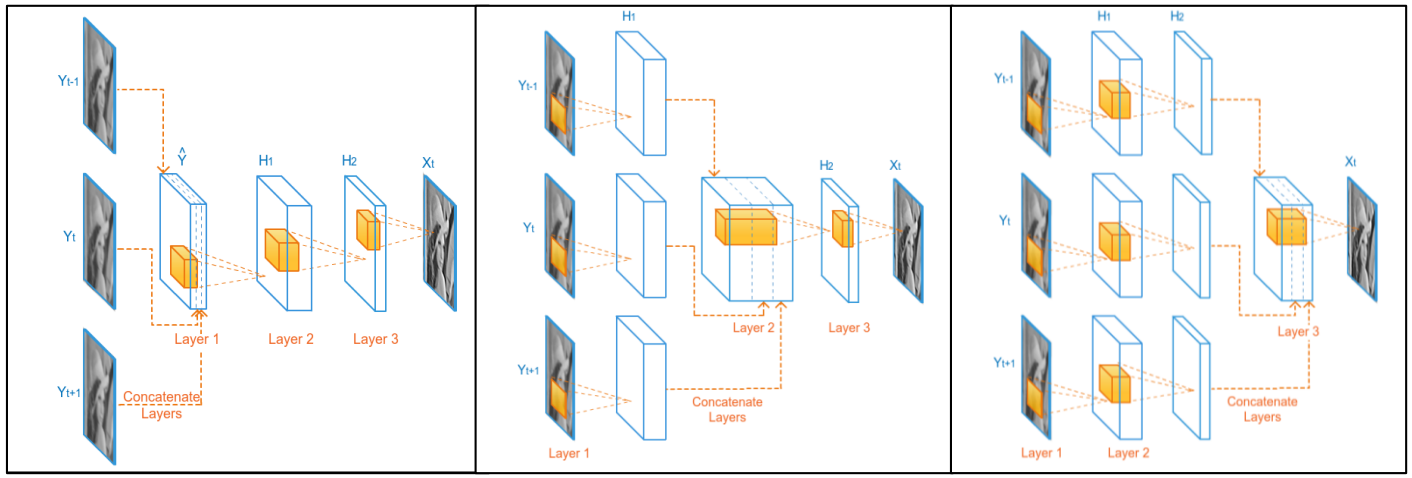
“Vanilla” CNN

- The “Vanilla” CNN shown in the previous slide has been used for many inverse imaging applications including denoising, super-resolution, and compressive sensing.



C. Dong, CC Loy, K He, and X. Tang, “Learning a Deep Convolutional Network for Image Super-Resolution, ECCV, 2014.

Multiple Frame Super Resolution



(a) (b) (c)

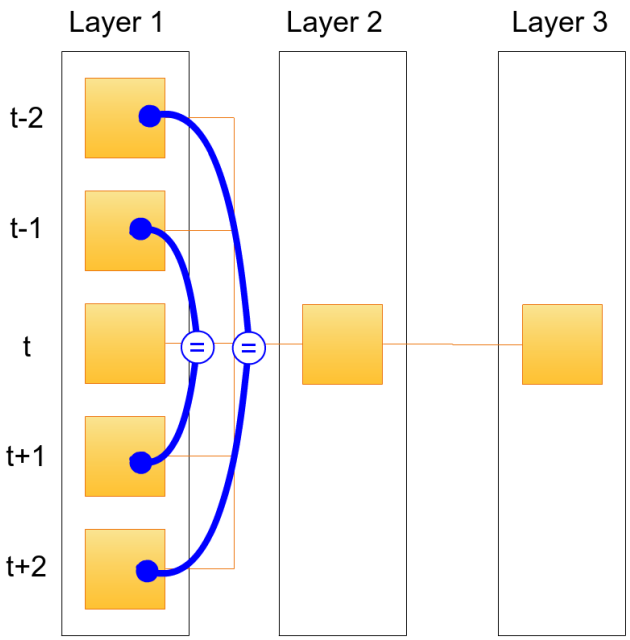
PSNR for 4x Upscaling:

Architecture	(a)	(b)	(c)	AVG
1 frame	-	-	-	31.26
3 frames	31.77	**31.81**	31.80	31.31
5 frames	31.80	**31.85**	31.82	31.33

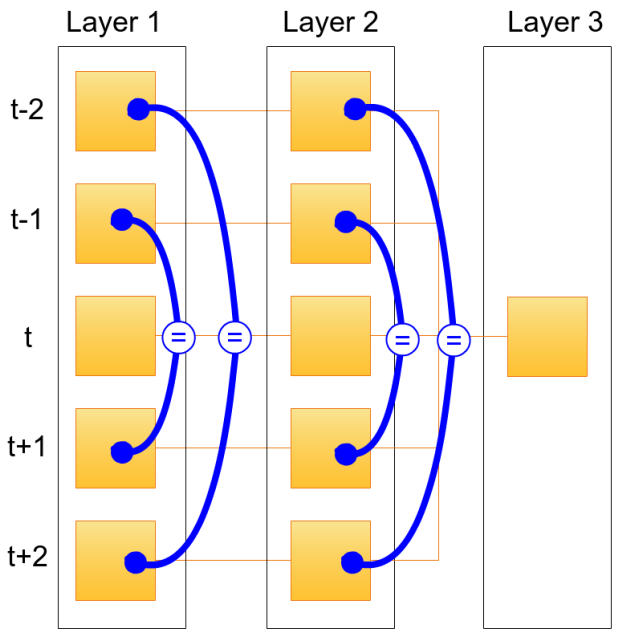
A. Kappeler, Q. Dai, S. Yoo, and A. K. Katsaggelos, "Video Super-Resolution Using Convolutional Networks," *IEEE Trans. On Computational Imaging*, pp. 109-122, June 2016.

Filter Symmetry Enforcement

For Architecture (b):



For Architecture (c):



Adaptive MC

Walk Sequence, 3x Upscaling

original | Bayesian-MB: 28.37 dB | VSRnet MC: 29.94 dB | VSRnet AMC: 30.21 dB

Foreman Sequence, 3x Upscaling

original | Bayesian-MB: 33.83 dB | VSRnet MC: 34.79 dB | VSRnet AMC: 35.74 dB

Experimental Results

MYANMAR	Scale	Image SR Algorithms				Video SR Algorithms		Own
		Bicubic	SrSC [5]	A+ [7]	SRCNN [15]	Bayesian [4]	Enhancer [37]	VSRnet
PSNR	2	34.59	-	37.19	37.79	36.24	35.94	**38.48**
SSIM	2	0.9458	-	0.9638	0.9640	0.9634	0.9588	**0.9679**
PSNR	3	31.59	32.71	33.48	33.88	32.07	32.50	**34.42**
SSIM	3	0.8957	0.9127	0.9191	0.9198	0.9131	0.9099	**0.9247**
PSNR	4	29.53	-	30.88	31.26	30.22	30.23	**31.85**
SSIM	4	0.8526	-	0.8777	0.8777	0.8742	0.8681	**0.8834**

Bicubic

VSRnet

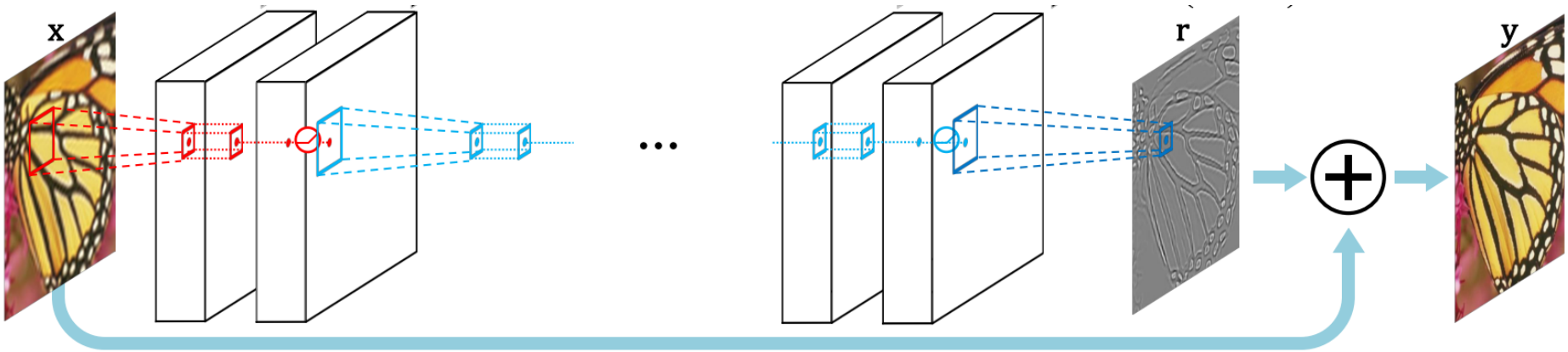
CNNs vs. Dictionaries

Myanmar Video

	Bicubic	MDMF-B-VT	MDMF-R-VT	VSRnet
Scene 2	45.27	**48.26**	48.11	47.92
Scene 8	38.18	41.62	41.72	**42.53**
Scene 18	41.43	44.73	44.94	**45.00**
Scene 25	44.40	49.57	**49.68**	48.45
Scene 33	40.22	45.33	**45.44**	44.50

Improvements over vanilla CNN: skip connections

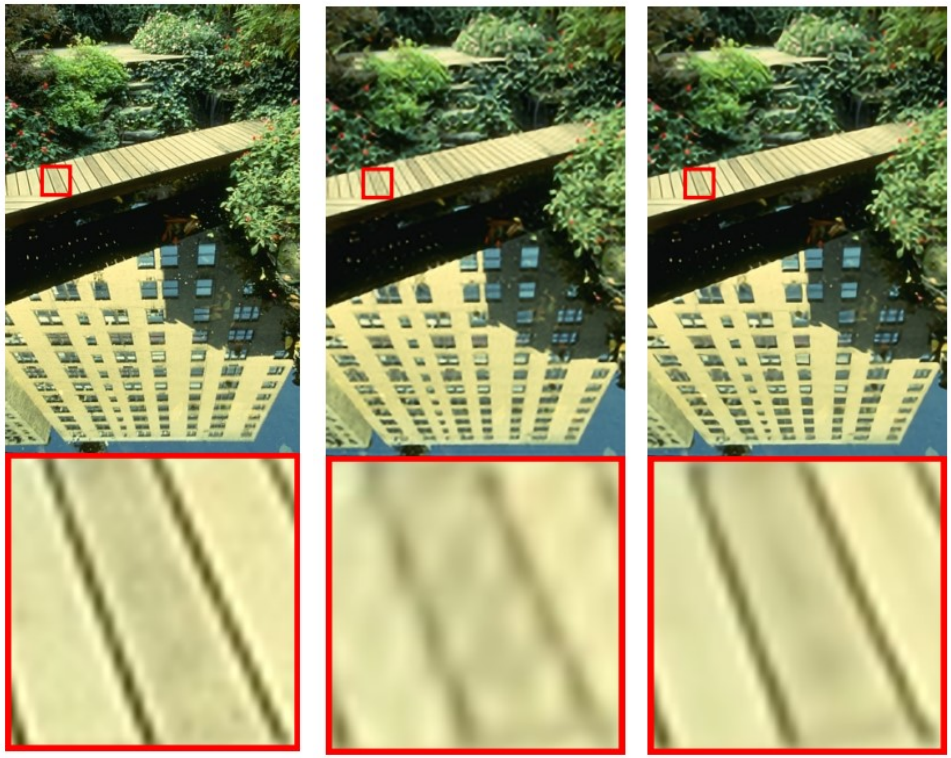
- Skip connections allow for residual learning between the input and output image.



Kim, J., Lee, J. K., & Lee, K. M. (2015). Accurate Image Super-Resolution Using Very Deep Convolutional Networks, arXiv 2016.

Results

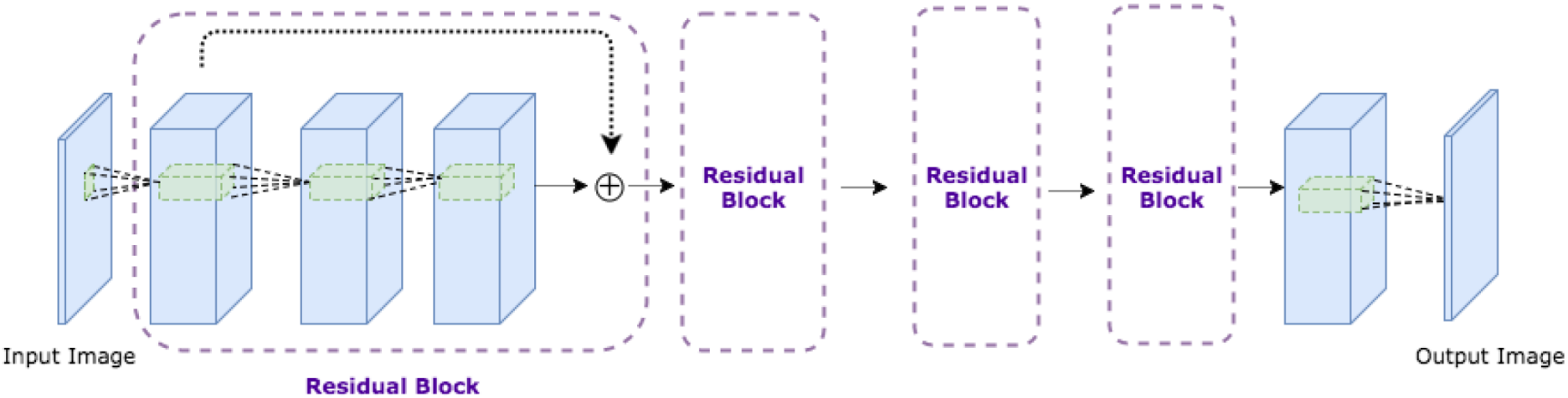
- Residual learning can lead to a significant improvement over “vanilla” end-to-end mapping.



Ground truth | SRCNN (Dong et al., 2014) | Kim et al. (2015)

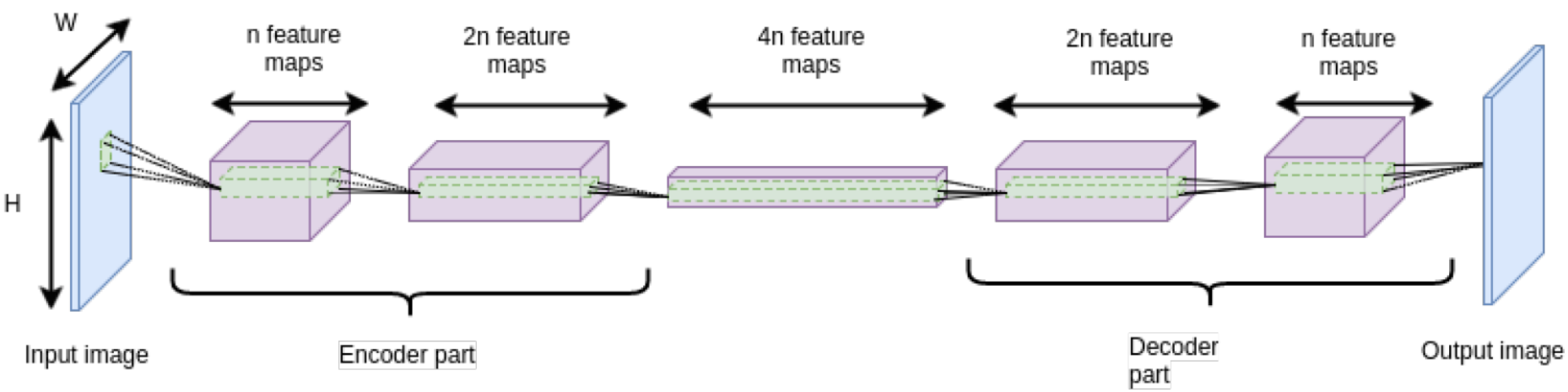
Improvements over vanilla CNN: Residual blocks

- Residual blocks help train networks with greater depth – increasing the capacity of the neural network.



Improvements over vanilla CNN: Encoder-decoder CNNs

- The first part of the network, the ”compressive” part, learns an abstract representation of the input image, which is then used by the ”expansive” part of the network to produce an output image.
- This modelling has a very intuitive justification in the probabilistic formulation of the inverse problem, in which we find a set of latent variables which after decoding are able to explain our observations



Improvements over vanilla CNN: Encoder-decoder CNNs

- Encoder-decoder CNNs have been used in image inpainting, denoising, and optical flow estimation.

Input Reconstruction

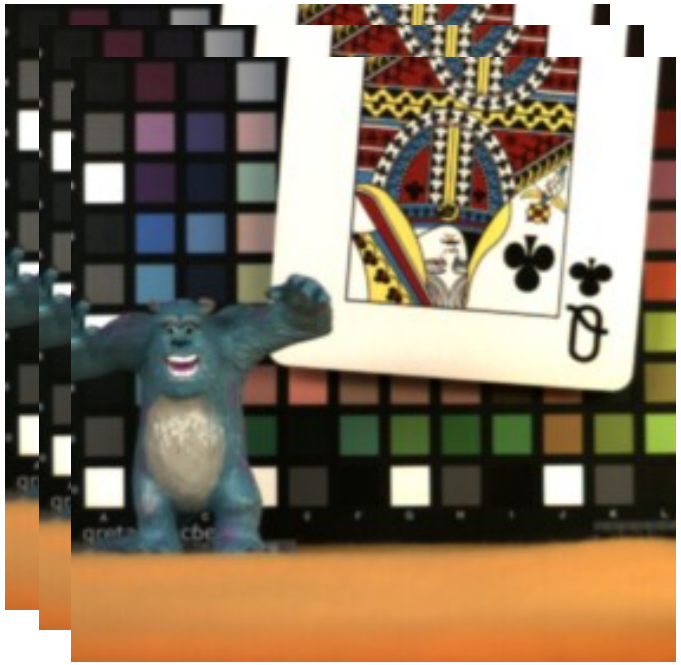
Pathak, D., Donahue, J., & Efros, A. A. (2016). Context Encoders : Feature Learning by Inpainting. *Cvpr 2016*, 2536–2544. http://doi.org/10.1109/CVPR.2016.278

High Speed Compressive Video

1. Capture high-speed video using a low frame-rate camera.
2. Optimize utilized sampling patterns.

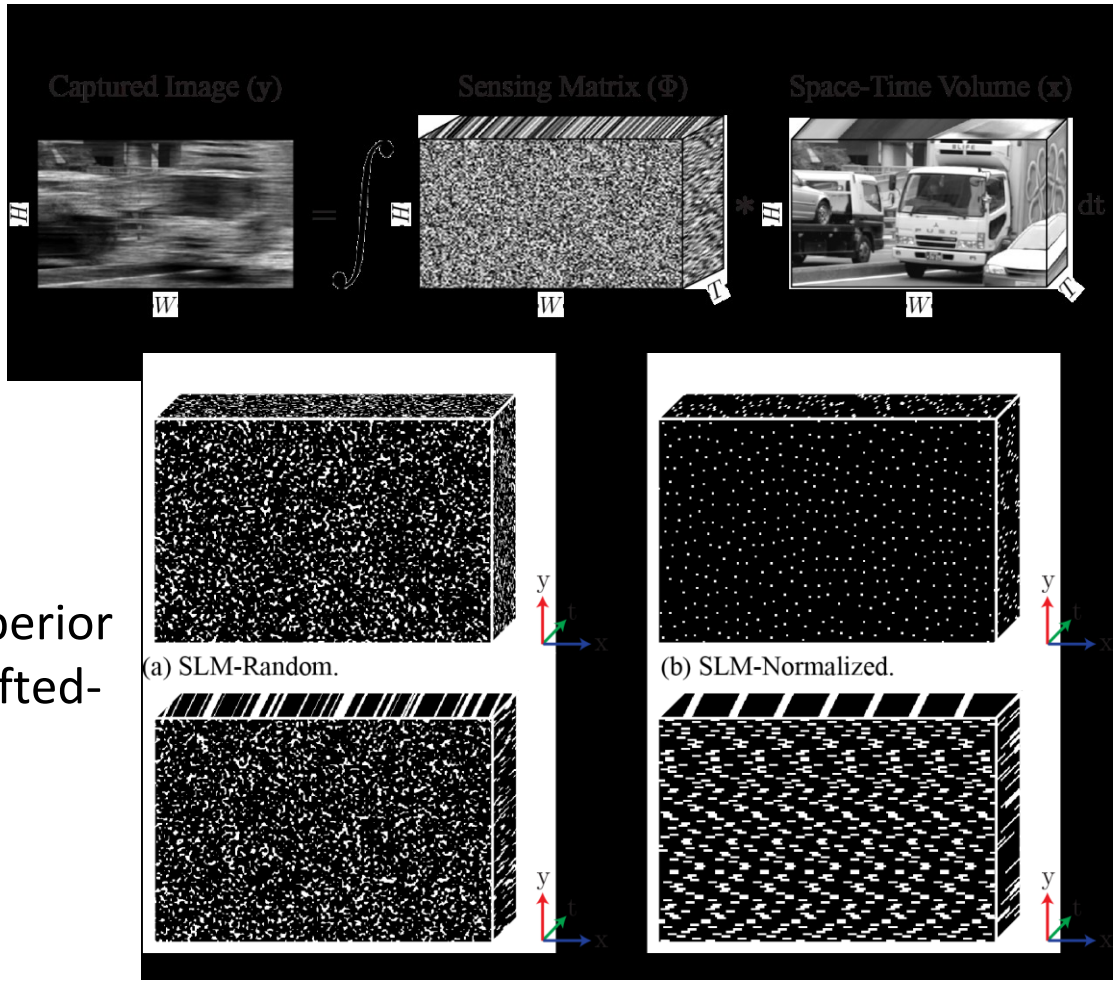
Low frame-rate blurry frame

High frame-rate blur-free frames



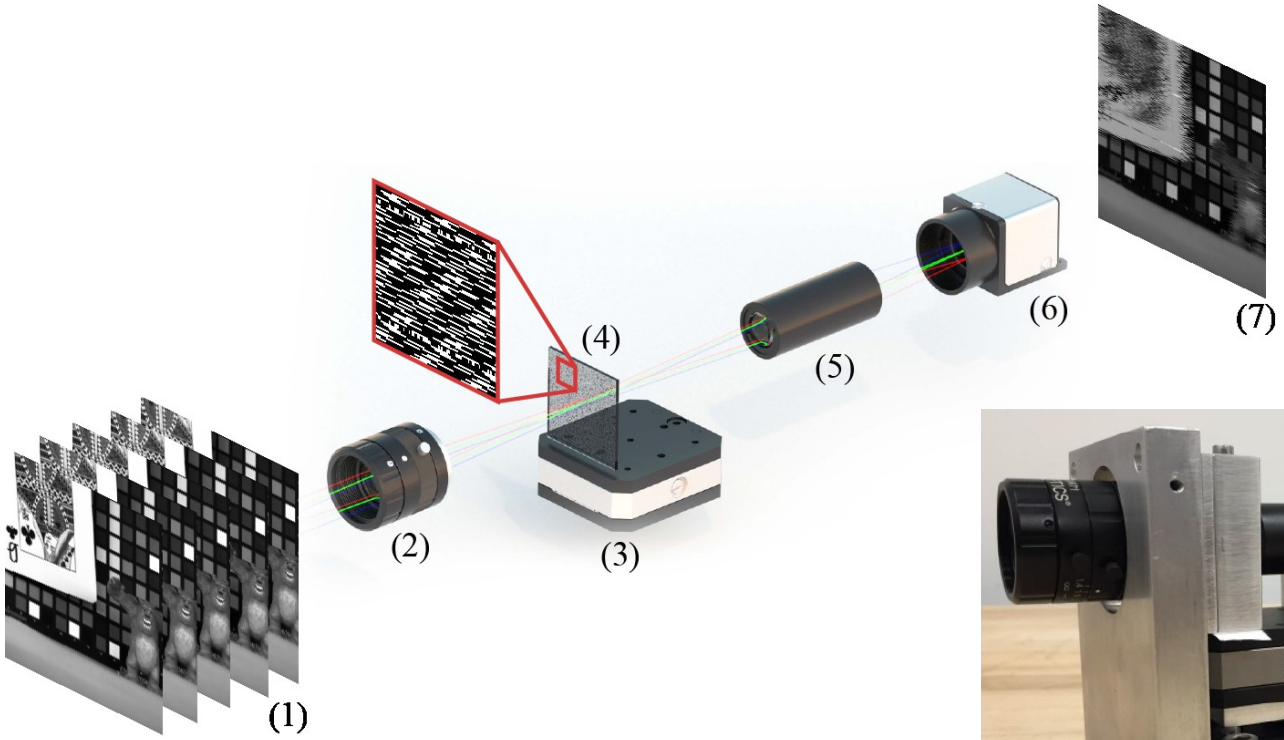
CS Video System

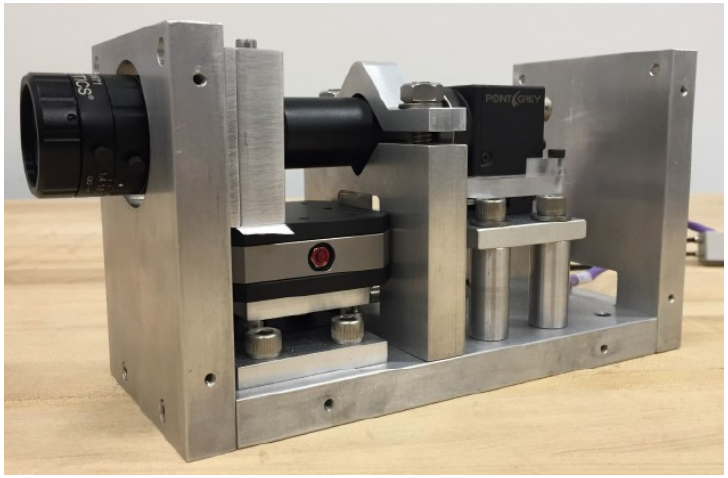
- Measurement model:



- Tested masks:
  - Shifted-Normalized masks provided superior performance to Shifted-Random.

NU CS Video System

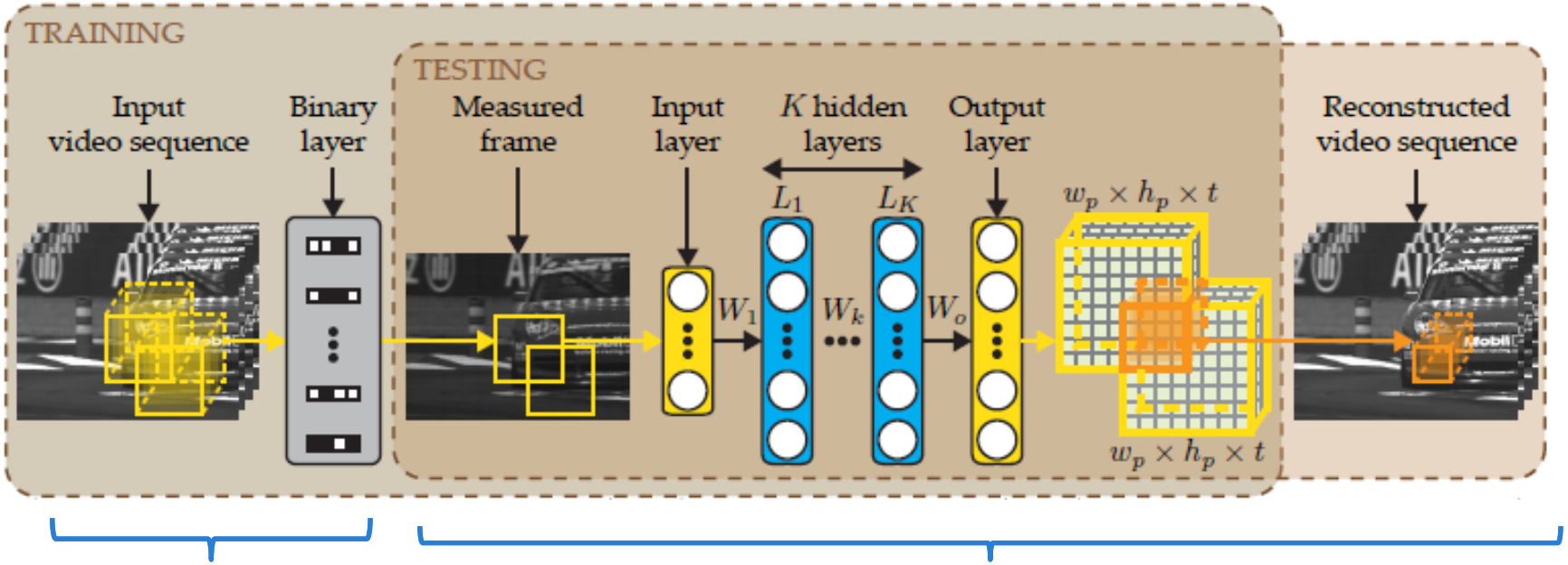




Koller. et al. “High spatio-temporal resolution video with compressed sensing”, Optics Express, 2015.

Proposed Encoder-Decoder

Encoder: we learn the **binary** weights of the sensing matrix
Decoder: we learn to reconstruct the video sequence

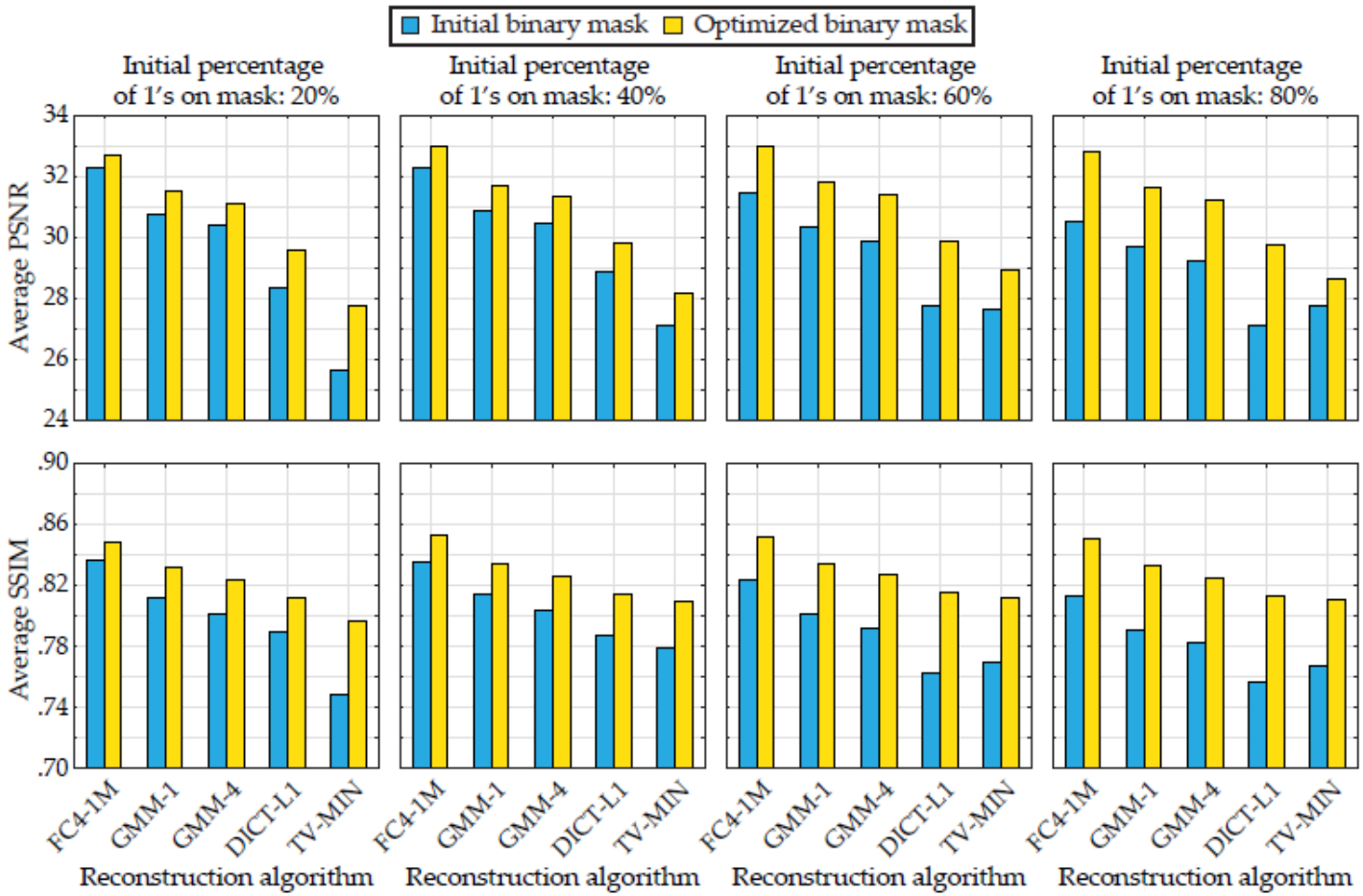


Encoder

Decoder

Iliadis, M., Spinoulas, L., & Katsaggelos, A. K. (2016). DeepBinaryMask: Learning a Binary Mask for Video Compressive Sensing. *arXiv Preprint*, (1), 1–13. Retrieved from http://arxiv.org/abs/1603.04930

Quantitative Results



Video Reconstruction Results

DICT-L1, DeepMask-40, PSNR: 30.05

DICT-L1, RandomMask-40, PSNR: 28.42

TV-MIN, DeepMask-40, PSNR: 29.86

TV-MIN, RandomMask-40, PSNR: 28.07

DICT-L1, DM-40, PSNR: 27.88

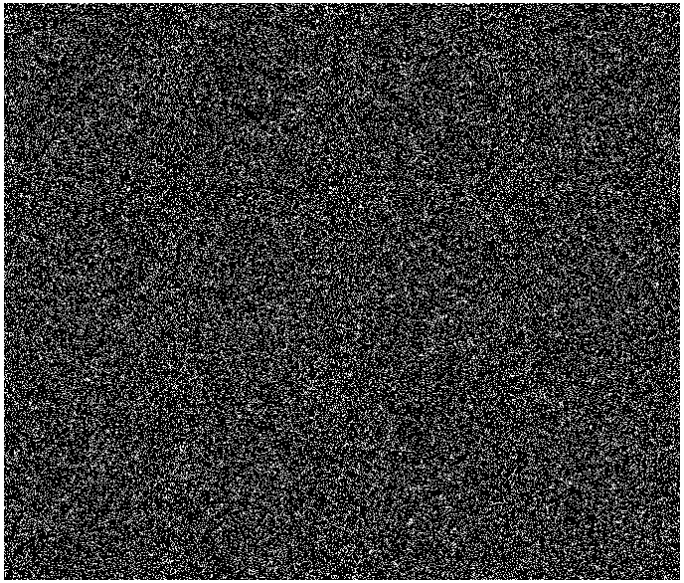
GMM-4, DM-40, PSNR: 28.79

TV-MIN, DM-40, PSNR: 28.14

FC4-1M, DM-40, PSNR: **30.09**

An Inpainting Approach

Full Scan: 580x680=394400 pixels

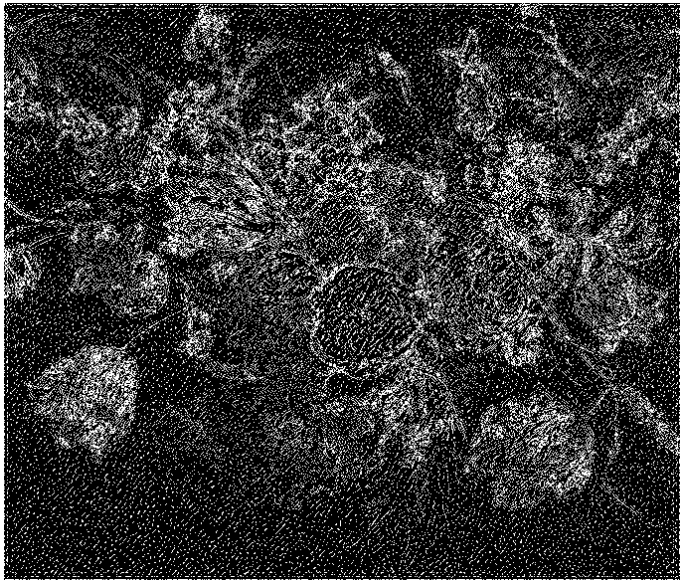


20% Scan: 580x680x20%=78880 pixels
5X speedup

An Inpainting Approach

Full Scan: 580x680=394400 pixels

Adaptive Sampling Mask



20% Scan: 580x680x20%=78880 pixels
5X speedup

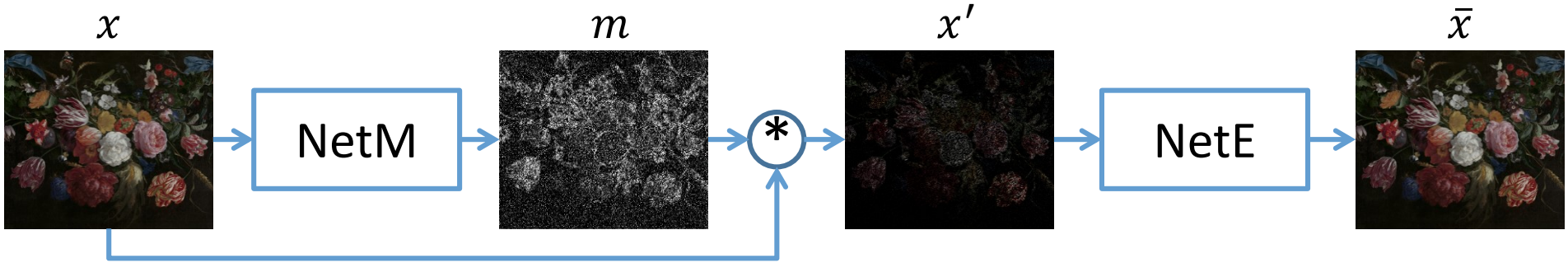
Experimental Results: RGB Image Inpainting

Random Sampling Mask
PSNR: 26.97 dB

Experimental Results: RGB Image Inpainting

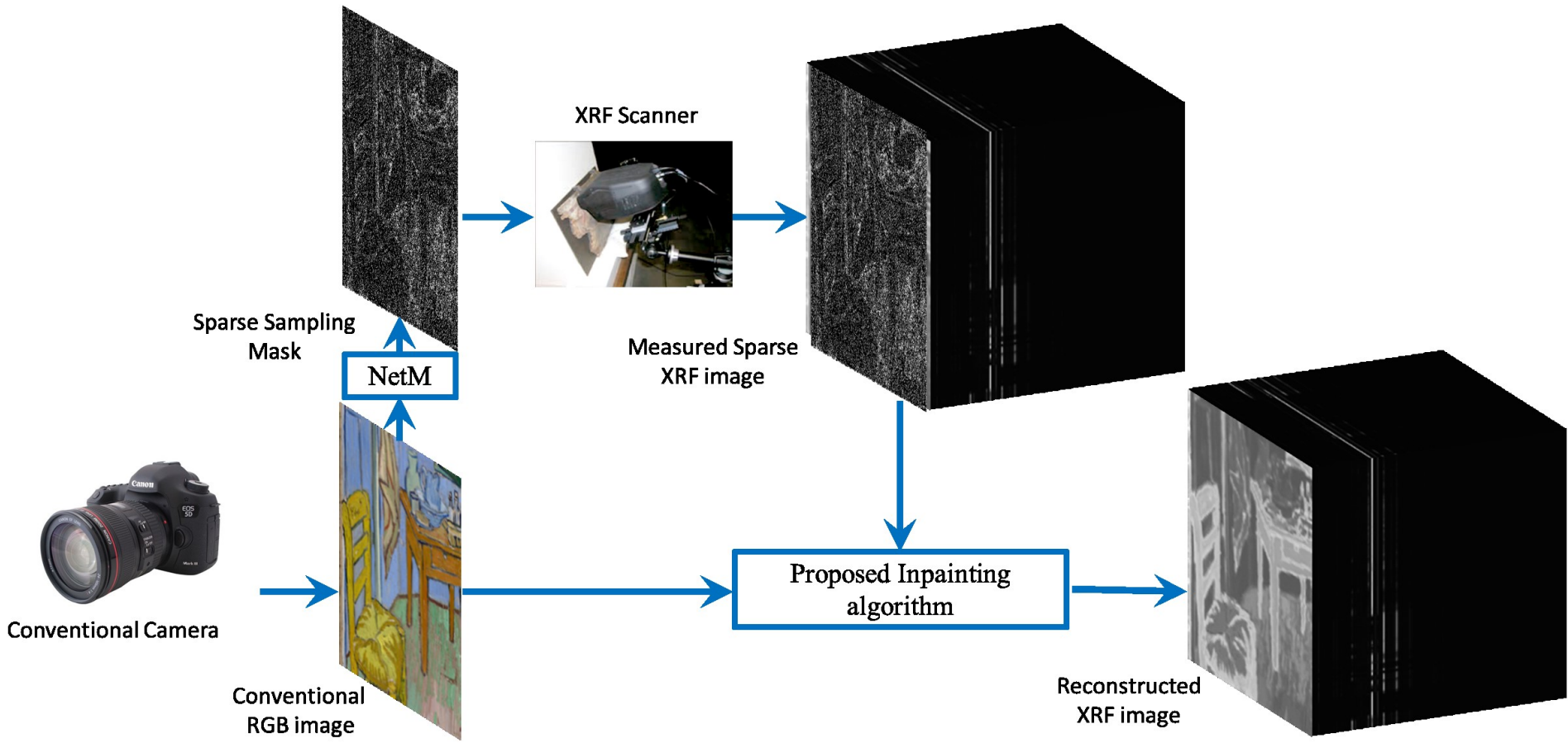
Adaptive Sampling Mask
PSNR: 29.04 dB

Proposed Method

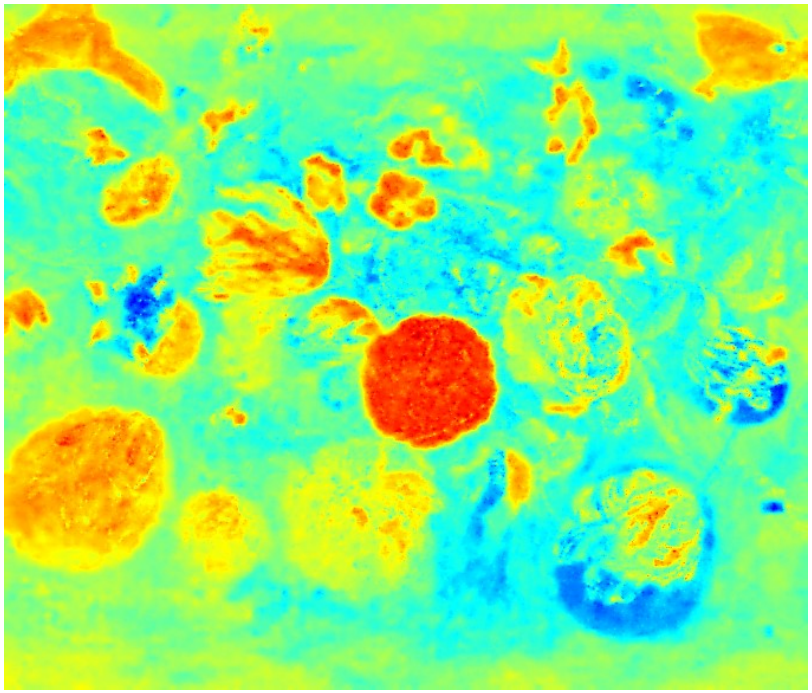


$$\min_{NetM,NetE} \sum_{i=1}^{N} \|\bar{x}_i - x_i\|_2^2$$

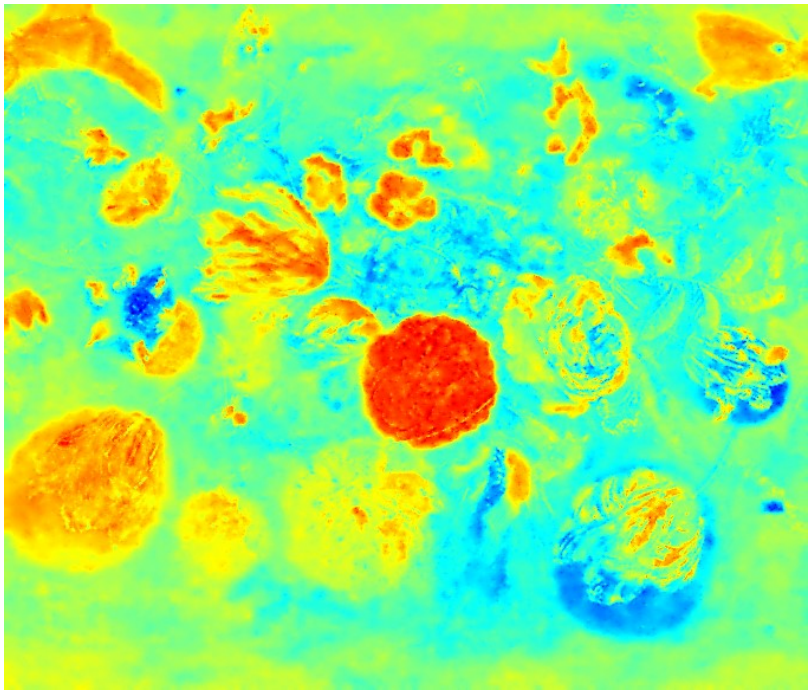
XRF/RGB Fusion Inpainting



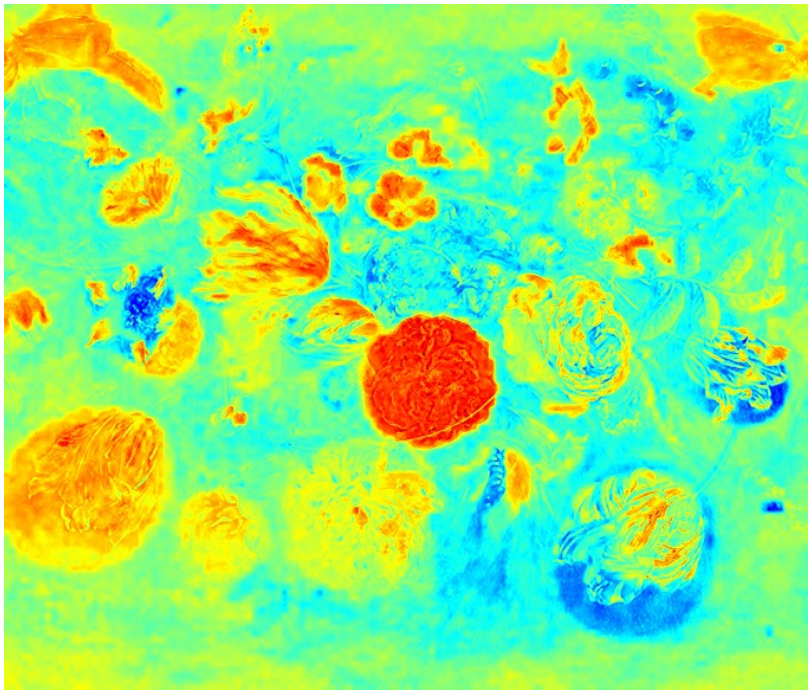
MS reconstruction, Channel #13, Random sampling mask



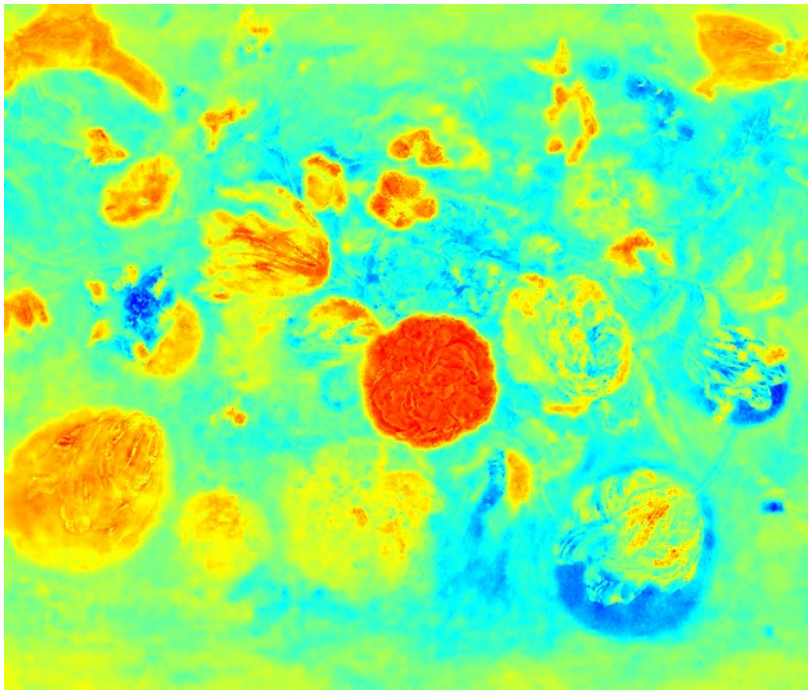
MS reconstruction, Channel #13, Adaptive sampling mask



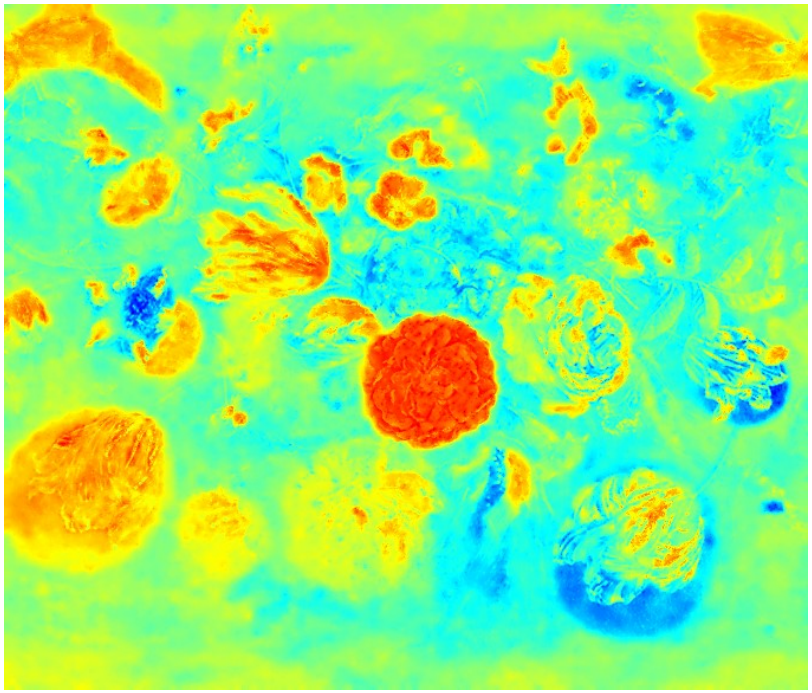
Ground Truth Channel #13



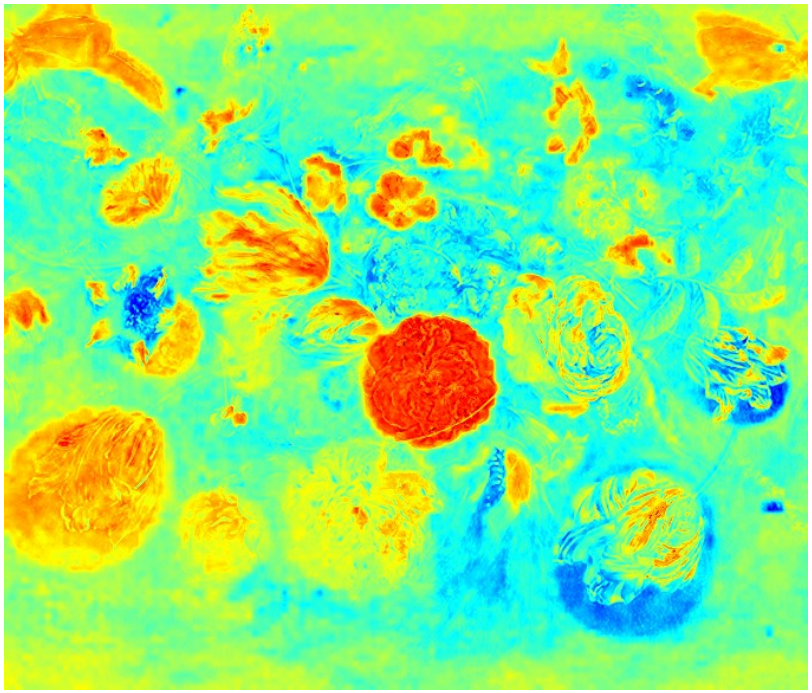
XRF/RGB Fusion Inpainting
Fusion Result of Random Sampling Channel #13



XRF/RGB Fusion Inpainting
Fusion Result of Adaptive Sampling Channel #13

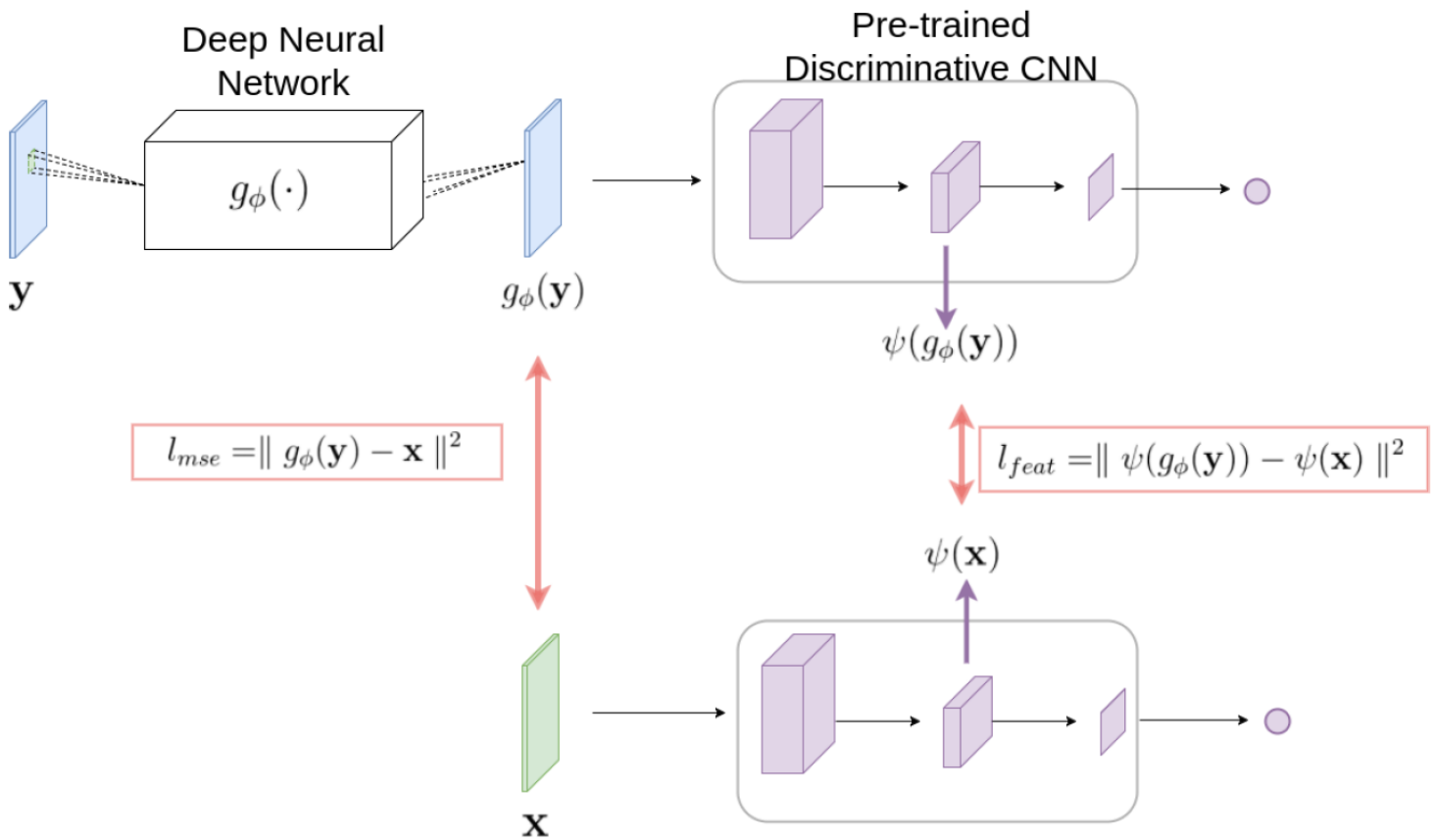


Ground Truth Channel #13

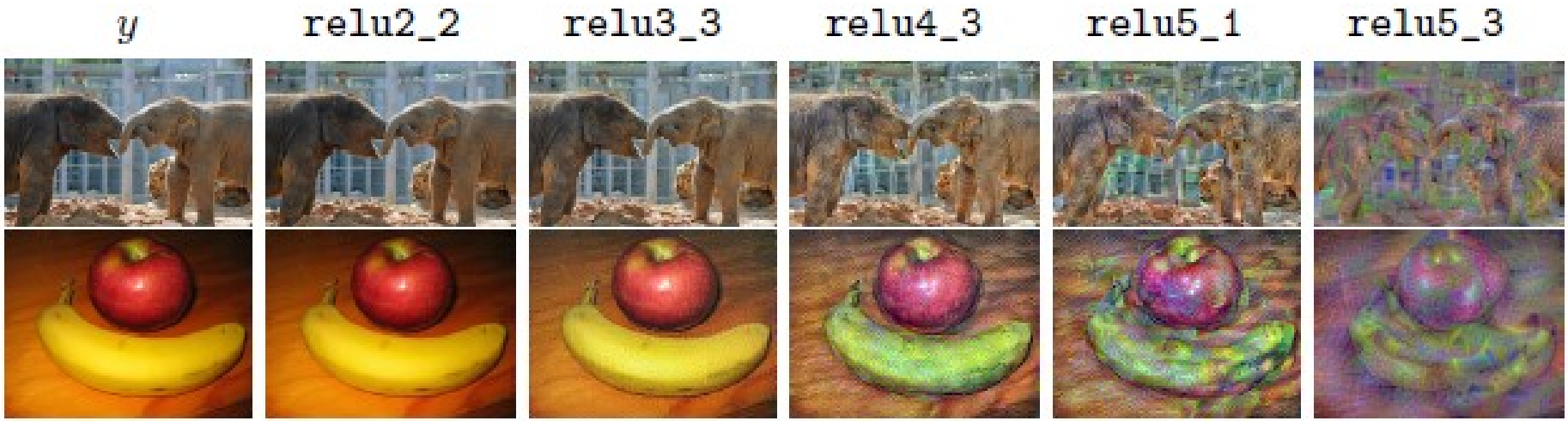


From MSE to feature-space loss

- Because CNNs are able to learn all the relevant statistics of natural images, feature space losses can be used to solve inverse imaging problems.



From MSE to feature-space loss



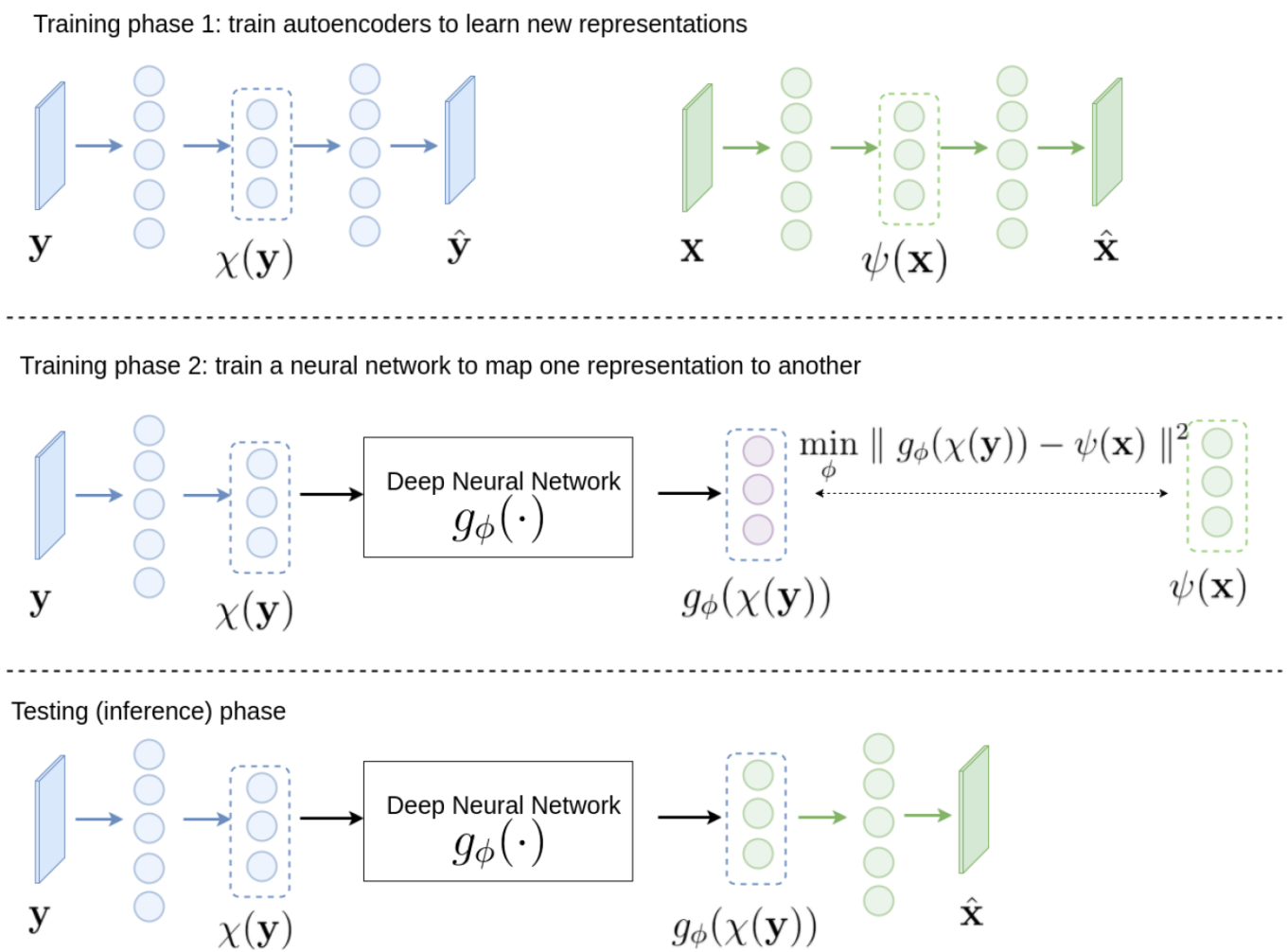
Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses. *Arxiv*, 1–5. http://doi.org/10.1007/978-3-319-46475-6_43

Feature-space loss for super-resolution

Ground Truth	Bicubic	Ours (ℓ_{pixel})	SRCNN [11]	Ours (ℓ_{feat})
This image	31.78 / 0.8577	31.47 / 0.8573	32.99 / 0.8784	29.24 / 0.7841

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses. *Arxiv*, 1–5. http://doi.org/10.1007/978-3-319-46475-6_43

End-to-end mapping in learned feature spaces

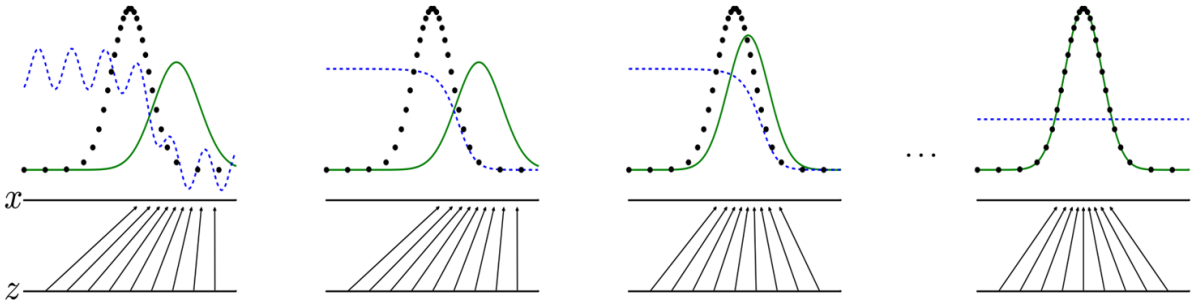


Zeng, K., Yu, J., Wang, R., Li, C., & Tao, D. (2017). Coupled Deep Autoencoder for Single Image Super-Resolution. *IEEE Transactions on Cybernetics*, *47*(1), 27–37. http://doi.org/10.1109/TCYB.2015.2501373

GANs: the minimax two-player game

Generative Adversarial Networks (GANs) are composed of two models:

- The **generative model** is the model generating a sample $G(\mathbf{z})$
- Given $G(\mathbf{z})$, the **discriminative model** learns to discriminate samples that come from the generator with samples from the training data. It does so by outputting a probability $D(G(\mathbf{z}))$.

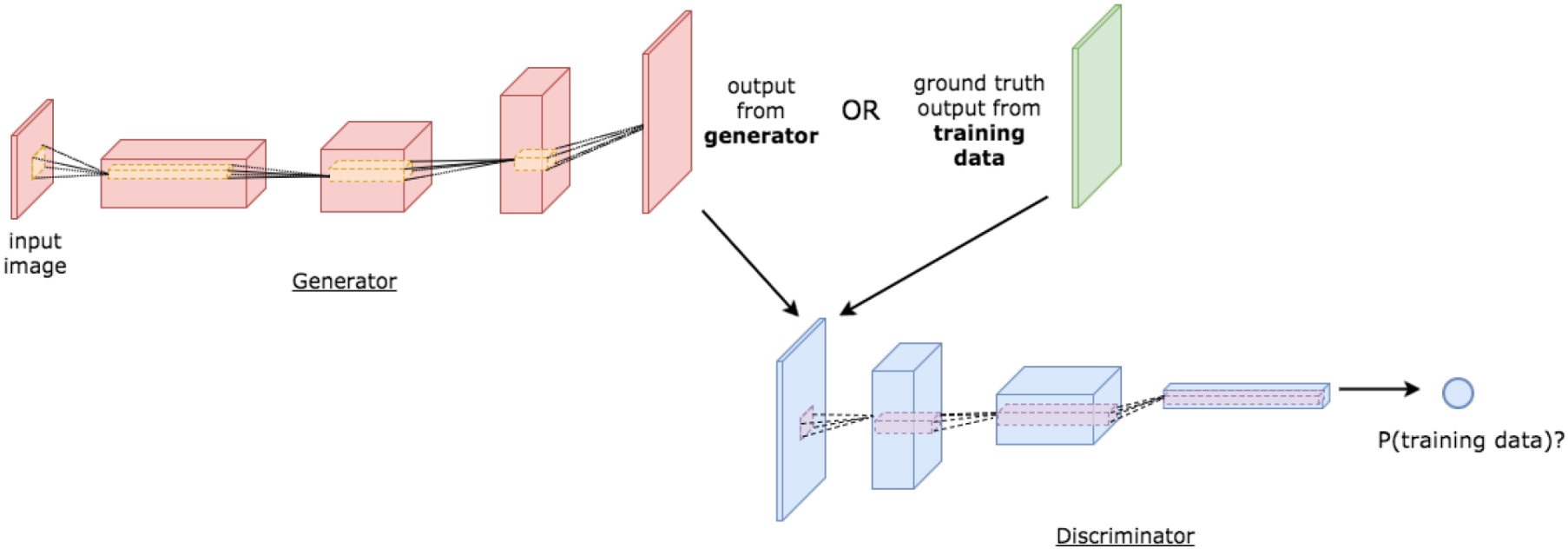


I Goodfellow, J Pouget-Abadie, and Mehdi Mirza. Generative Adversarial Networks. arXiv preprint, pages 1–9, 2014.

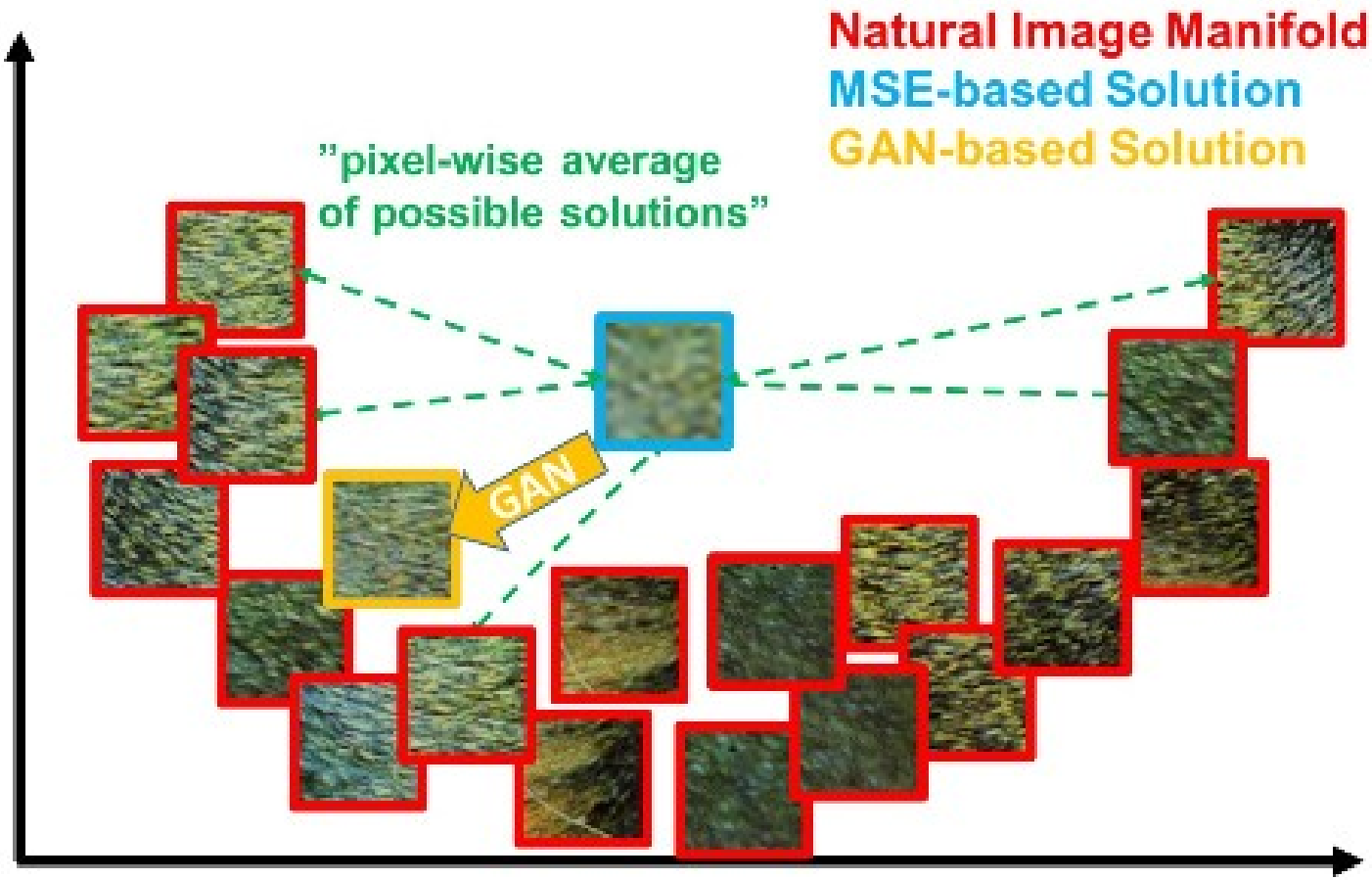
Conditional Generative Adversarial Networks (cGANs)

- Generative Adversarial Networks (GANs) can learn distributions of the training data.

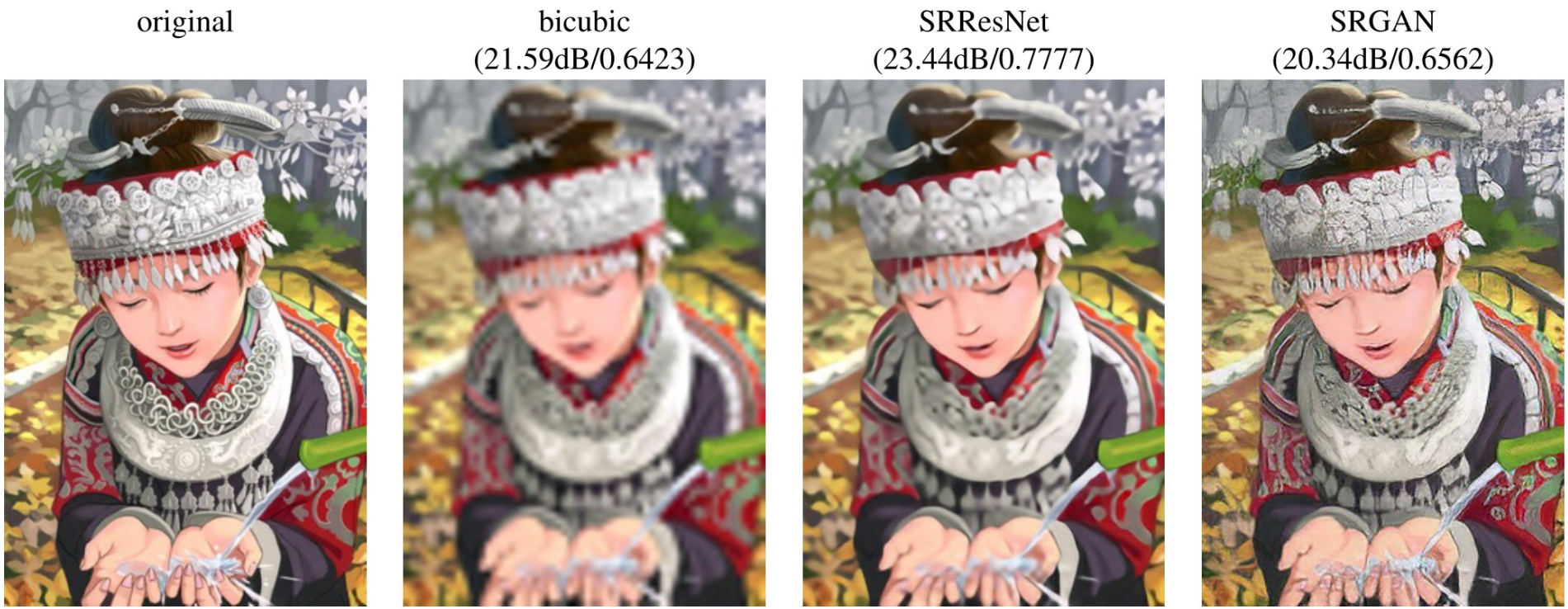
$$l_{GAN} = \mathop{\mathbb{E}}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[log(D(\mathbf{x}))] + \mathop{\mathbb{E}}_{\mathbf{y}\sim p_{data}(\mathbf{y})}[log(1-(D(G(\mathbf{y}))))]$$



Solution Pushed to Natural Image Manifold



cGANs for image super-resolution



Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A., Tejani, A., … Shi, W. (2016). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Retrieved from http://arxiv.org/abs/1609.04802

Using NNs as denoiser in variable splitting-based optimization

ADMM recovery formulation

$$\min_{\mathbf{x},\mathbf{z}} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{z} \parallel^2 + \lambda\phi(\mathbf{x})$$

$$\text{s.t } \mathbf{x} = \mathbf{z}$$

Augmented Lagrangian

$$\mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{u}) = \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{z} \parallel^2 + \lambda\phi(\mathbf{x}) + \frac{\rho}{2} \parallel \mathbf{x} - \mathbf{z} - \mathbf{u} \parallel^2$$

Iterative procedure

Update to be replaced by a DNN

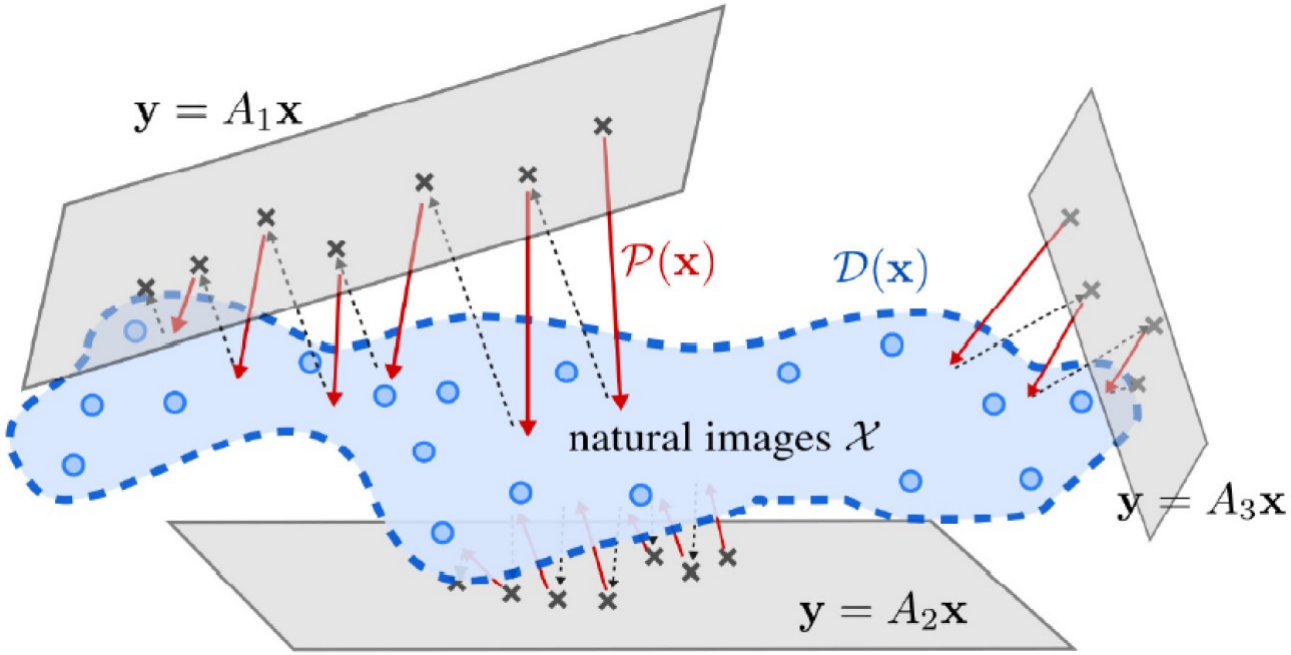
$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \frac{\rho}{2} \parallel \mathbf{x} - \mathbf{z}^k + \mathbf{u}^k \parallel^2 + \lambda\phi(\mathbf{x})$$

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{x}} \frac{1}{2} \parallel \mathbf{y} - \mathbf{A}\mathbf{z}^k \parallel^2 + \frac{\rho}{2} \parallel \mathbf{x}^{k+1} - \mathbf{z} + \mathbf{u}^k \parallel^2$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}$$

Learning a single image prior

Learning a proximal operator with a deep projection model



[1] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” CoRR, vol. abs/1704.03264, 2017.

[21] J. H. R. Chang, C-L Li, B. Poczos, V. K. Vijaya Kumar, and A. C. Sankaranarayanan, “One network to solve them all - solving linear inverse problems using deep projection models,” CoRR, vol. abs/1703.09912, 2017.

Variable splitting and NNs

- In variable splitting method, we separate our problem into a denoising problem and a fidelity problem.
- The solution to the denoising problem can be obtained by a CNN. The regularization term in the denoising sub-problem is learned from the dataset by the CNN.

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \mu\|\mathbf{x} - \mathbf{z}_k\|^2 \longleftarrow \text{Fidelity sub-problem}$$

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} \frac{\mu}{2}\|\mathbf{z} - \mathbf{x}_{k+1}\|^2 + \lambda\Phi(\mathbf{z}) \longleftarrow \text{Denoising sub-problem}$$

Zhang, K., Zuo, W., Gu, S., & Zhang, L. (2017). Learning Deep CNN Denoiser Prior for Image Restoration. Retrieved from http://arxiv.org/abs/1704.03264

Variable splitting and NNs

- The variable splitting with NN approach has the advantage that it can be applied to multiple inverse imaging problems, including image denoising, image deblurring, and image super-resolution.

(a) Ground-truth (b) Zoomed LR image (e) Proposed$_G$ (29.32dB)

Zhang, K., Zuo, W., Gu, S., & Zhang, L. (2017). Learning Deep CNN Denoiser Prior for Image Restoration. Retrieved from http://arxiv.org/abs/1704.03264

Conditional Generative Adversarial Networks (cGANs)

Results in [1] indicate that when the process to obtain the LR image from the HR one is known, this knowledge leads to better HR reconstructions. In [2]

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

The generator has the form (notice that it takes into account the observation process)

$$g_\phi(\mathbf{y}) = (\mathbf{I} - \mathbf{A}^+\mathbf{A}) f_\phi(\mathbf{y}) + \mathbf{A}^+\mathbf{y}$$

The generator is updated by minimizing

$$l_{GAN}(G, D) = \mathop{\mathbb{E}}_{\mathbf{y} \sim p_{data}(\mathbf{y})} \log \frac{1 - D(g_\phi(\mathbf{y}))}{D(g_\phi(\mathbf{y}))}$$

[1] T. Radu, A. Eirikur, L. Van Gool, M-H. Yang, and L. Zhang, “NTIRE 2017 challenge on single image superresolution: Methods and results," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July 2017.
[2] C.K. Snderby, J. Caballero, L. Theis, W. Shi, and F. Huszar, “Amortised MAP Inference for Image Super-resolution," ICLR 2017, pp. 1{19, 2016.

Deep Unfolding

- Model-based methods: problem domain knowledge can be built into the constraints of the model, typically at the expense of difficulties during inference.
- Deterministic deep neural networks: inference is straightforward but their architectures are generic and it is unclear how to incorporate knowledge.
- "Solution": start with a model-based approach and an associated inference algorithm and ***unfold*** the inference iterations as layers in a deep network.
- In addition, instead of optimizing the original model, the model parameters are ***untied*** across layers, so as to create a more powerful network.

J. R. Hersley, J. Le Roux, and F. Weninger, "Deep Unfolding: Model-Based Inspiration of Novel Deep Architectures," arXiv:1409.2574v4, 20 Nov. 2014

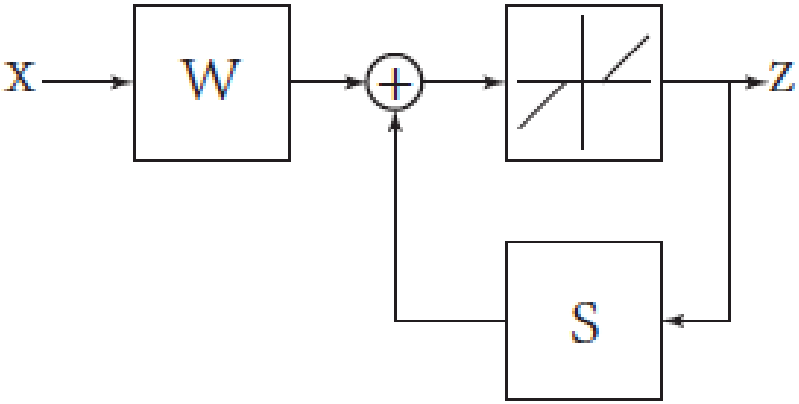
The Lasso Problem and ISTA/FISTA

$$z^* = \underset{z}{\operatorname{argmin}} \|x - Dz\|_2^2 + \lambda \|z\|_1$$

$$z(k+1) = h_\theta(W \cdot x + S \cdot z(k))$$

$$W = \frac{1}{L} D^T$$

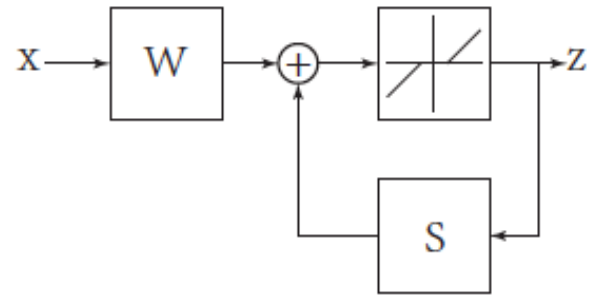
$$S = I - \frac{1}{L} D^T D$$



The shrinkage function h_θ is defined as $[h_\theta(Y)]_i = sign(Y_i)(|Y_i| - \theta_i)^+$

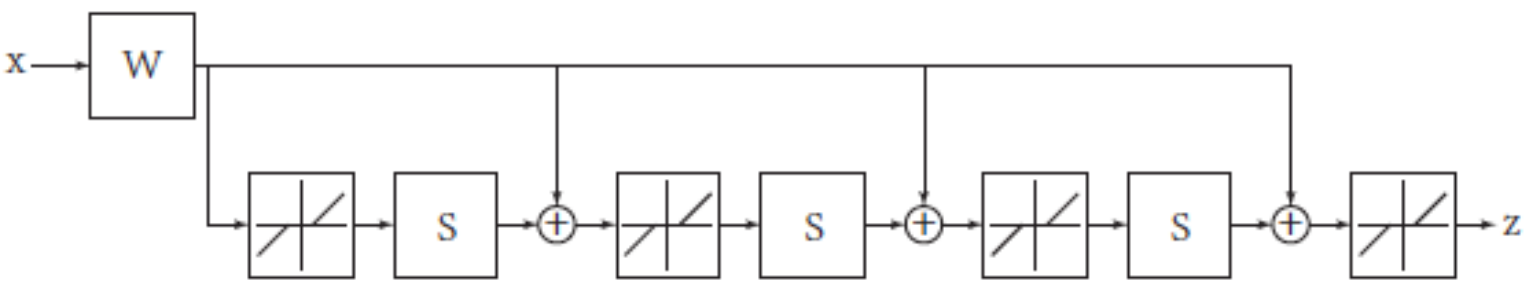
L is the upper bound on the largest eigenvalue of $D^T D$

Learned ISTA (LISTA)



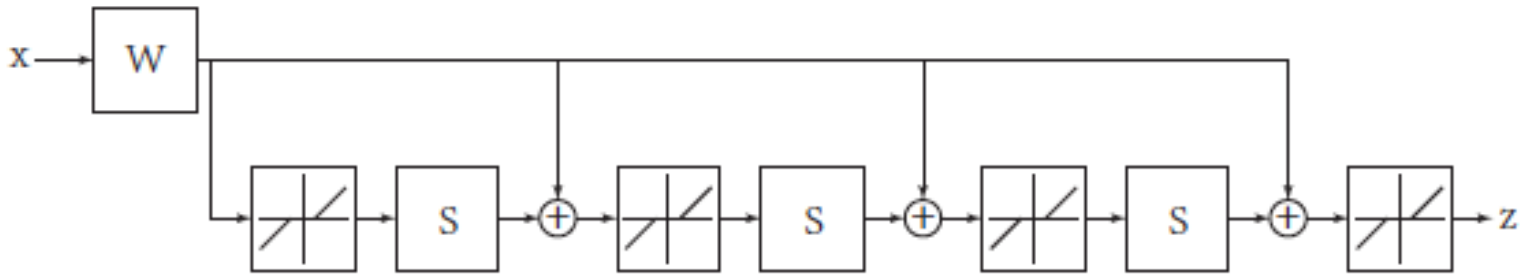
Unfold the ISTA iteration
Only 4 Iterations here

$$z = L(x, W, S, \theta)$$



Gregor Karol, and Yann LeCun. "Learning fast approximations of sparse coding." In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 399-406. 2010.

Learned ISTA



$$z = L(x, W, S, \theta)$$

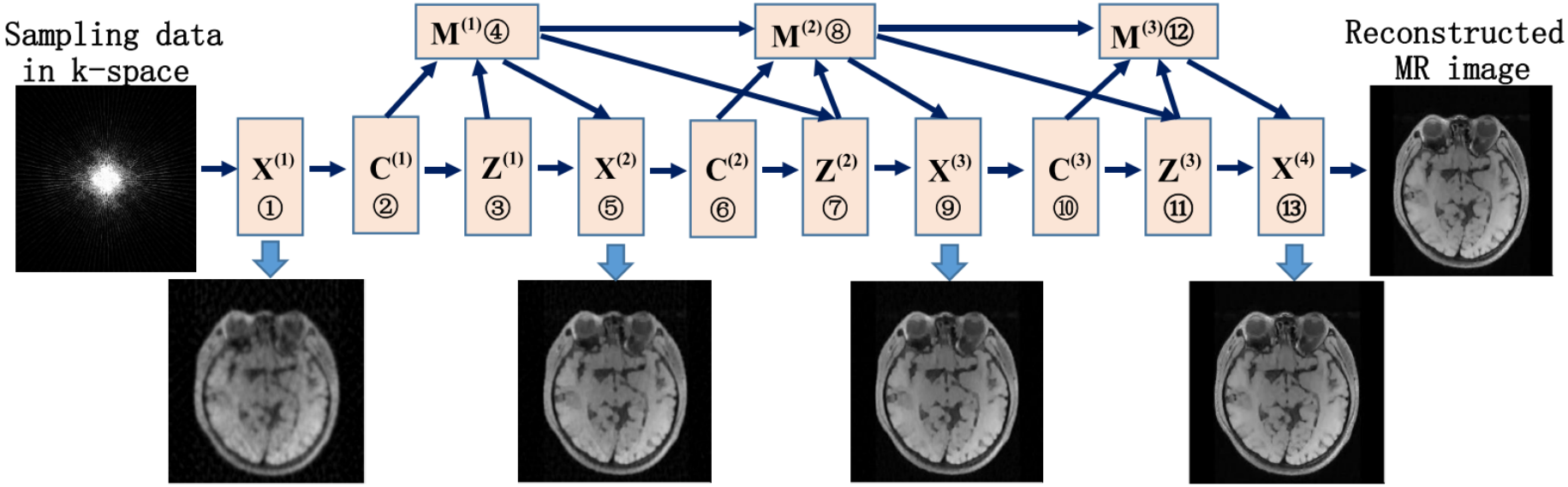
Unlike ISTA, where W, S, θ are determined by D, λ,
Learned ISTA learn the network parameter W, S, θ by minimizing

$$\min_{W,S,\theta} \sum_{j=1}^{N} \left\| L\left(x_j, W, S, \theta\right) - z_j^* \right\|_2^2$$

$$z^* = \underset{z}{\operatorname{argmin}} \|x - Dz\|_2^2 + \lambda \|z\|_1$$

Unfolding for MRI compressive sensing

- Here we base the graph structure on the ADMM algorithm, and learn the parameters as the layers of the neural network.
- The neural network parameters to be learned include a nonlinear transformation of the CS measurements, the shrinkage function, the regularization function, in addition to the various hyper-parameters of the ADMM algorithm.



Yang, Y., Sun, J., Li, H., & Xu, Z. (2016). Deep ADMM-Net for Compressive Sensing MRI. *Nips*, (Nips), 10–18. Retrieved from http://dblp.uni-trier.de/db/conf/nips/nips2016.html#Yang0LX16

Final Thoughts

- Learning in general and via DNNs in particular, although not precisely understood, provides fair competition to modeling techniques, capitalizing on the power of computers (e.g., GPUs) and the abundance of data

- The impressive achievements in recognition tasks (face, speech) do not transfer to recovery problems (yet)

- Advantageous in more “complicated” recovery cases where modeling is at a disadvantage

Final Thoughts

- In spite of the success, by and large DL for image recovery does not usually capture model uncertainty. Most of the current DL approaches can be thought of as ML or MAP methods.

- On the other hand, analytical techniques have greatly benefited from capturing model uncertainty and using approximate inference methods like VB and EP.

- The introduction of GANs (and VAEs) and the description of fundamental DL techniques such as dropout as VB approximation (*) will likely lead to a better modeling of uncertainty with DL techniques for imaging problems.

(*) Y. Gal and Z. Ghahramani, Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48* (ICML'16),

Contributors

- Prof. Rafael Molina, U Granada
- Prof. Ollie Cossairt, Northwestern
- Prof. Marc Walton, Northwestern
- Dr. P. Ruiz, Northwestern
- Dr. Derin Babacan, Google
- Dr. Zhaofu Chen, Google
- Dr. Michael Iliadis, Sony
- Dr. Leonidas Spinoulas, ISI/USC
- Dr. Armin Kappeler, Yahoo
- Tim Qai, EyeVerify
- Andrew Yoo, Northwestern