Imperial College London

Sparsity and Deep Neural Networks: a Match Made in Heaven

Pier Luigi Dragotti, Imperial College London

7 February 2022

Imperial College London

Motivation: A Theory for DL

- Deep Neural Networks achieves state-of-the-art performance in many imaging tasks

- Fundamental questions:
  - is there a systematic way to interpret Deep Neural Networks?
  - Is there a systematic way to design the architecture of a Deep Neural Networks?

- Personal view: sparse signal representation theory is much more developed and can be used to help addressing both questions

Imperial College London

Outline

- Invertible Neural Networks and Wavelets
  - What are invertible Neural Networks (INN)?
  - Lifting Scheme and INN
  - Wavelet-like INN for denoising and deblurring
- Multimodal Image Processing and Unfolding
  - Multimodal Image Super-Resolution
  - Unfolding strategy for image separation in Art Investigation
- Conclusions

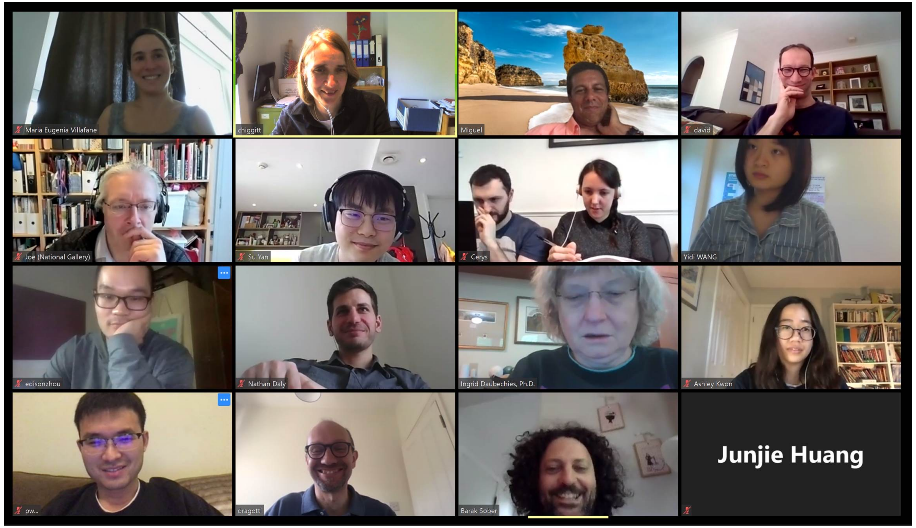
Common Theme: Interplay between sparsity and learning.

Imperial College London

Joint work with

Junjie Huang (ICL, now Associate Prof. at NUDT))

Xin Deng (ICL, now Associate Prof. at Beihang University)



Consortium involving: UCL, ICL, Duke and National Gallery led by M. Rodrigues

Imperial College London

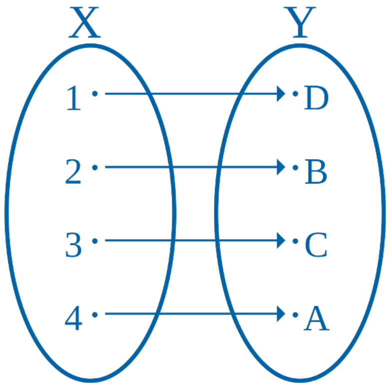
What are Invertible Neural Networks?

- Bijective (invertible) function approximators that have a forward mapping

$$F_\theta : \mathbb{R}^d \to \mathbb{R}^l$$
$$x \mapsto z$$

- and inverse mapping

$$F_\theta^{-1} : \mathbb{R}^l \to \mathbb{R}^d$$
$$z \mapsto x$$

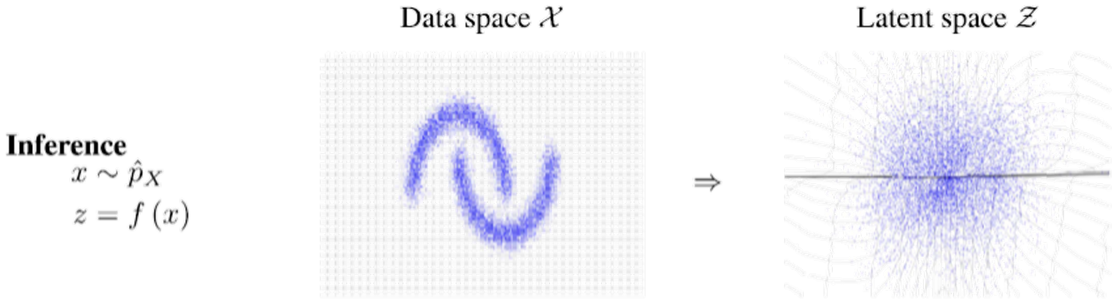


A bijective function (or invertible function)

Imperial College London

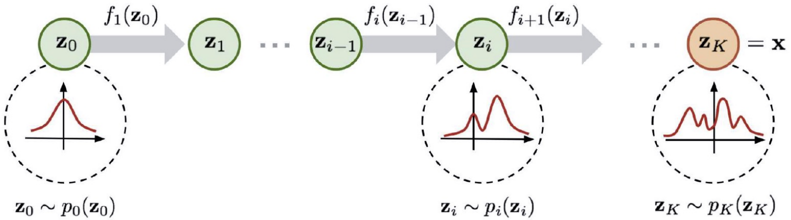
What are Invertible Neural Networks?

- INNs are bijective function approximators



Why Invertible Neural Networks?

- Generative modeling
  - Tractable Jacobian, allows explicit computation of posterior probability
  - Normalizing flows

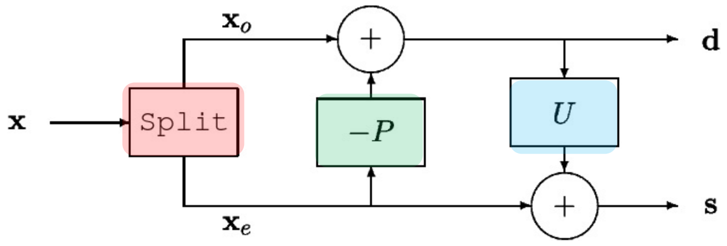


Kingma, Diederik P., and Prafulla Dhariwal. "Glow: Generative Flow with Invertible 1×1 Convolutions." arXiv preprint arXiv:1807.03039 (2018).

Imperial College London

How to Achieve Invertibility?

- Invertible via lifting scheme like architectures
  - Signal splitting
  - Alternate prediction and update



Split → $\begin{cases} d = x_o - P(x_e) \\ s = x_e + U(d) \end{cases}$

$\begin{cases} x_o = d + P(x_e) \\ x_e = s - U(d) \end{cases}$ → Merge

Forward pass

Backward pass

Imperial College London

Wavelets and Invertible Neural Networks

- In the beginning there were Wavelets 😄
- Wavelets provide sparse representations of piecewise smooth images.
- This is why they have been successfully used in many application including denoising

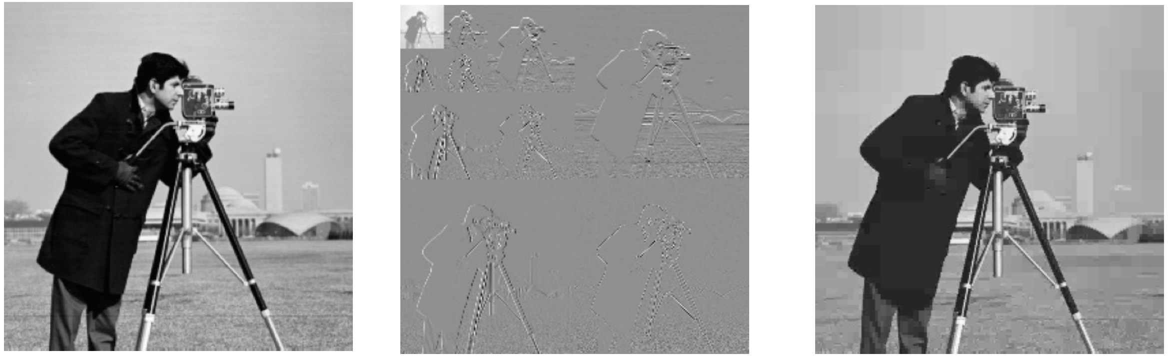
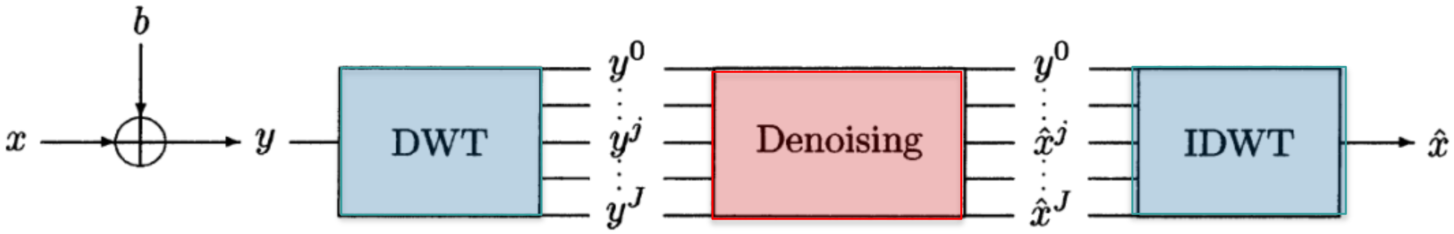


Figure: Cameraman is reconstructed using only 8% of the wavelet coefficients

Wavelet-based Denoising

Imperial College London

- Principles of wavelet denoising:



Wavelet transform

- Multi-resolution analysis
- Perfect reconstruction
- Noise is uniformly spread through the coefficients
- Image information is concentrated on small number of large coefficients

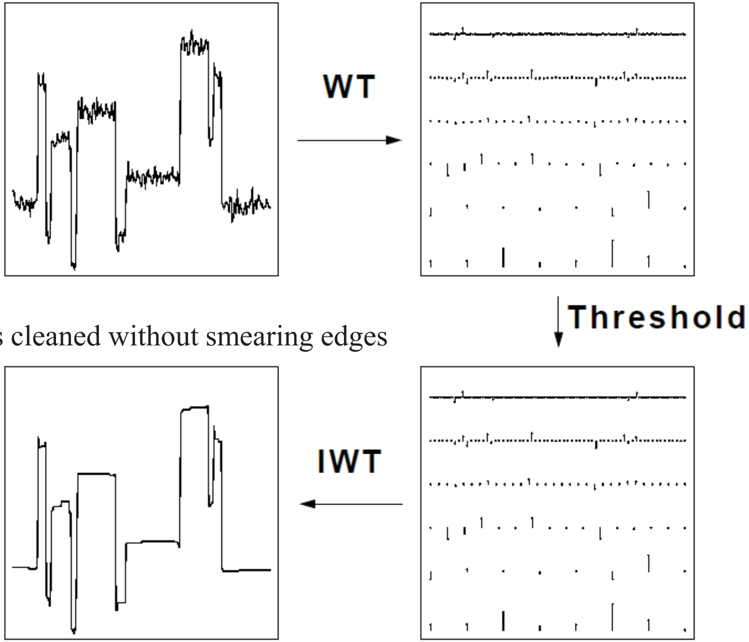
Denoising

- Element-wise thresholding, e.g. soft-thresholding

Imperial College London

Wavelet-based Denoising

1-D Example



Smooth parts cleaned without smearing edges

Imperial College London

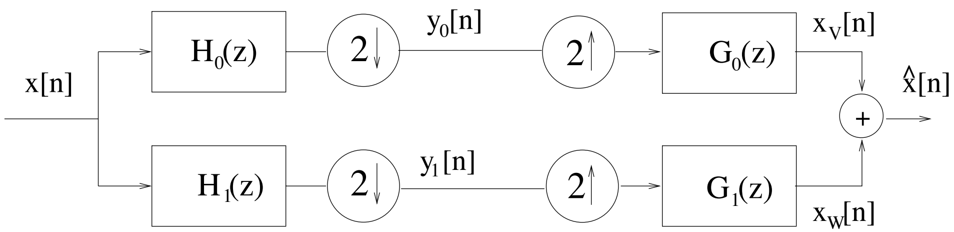
Wavelets for Deconvolution

- Sparsity constraints in the wavelet domain (or in another domain) can also be used as a regularizers for different applications, e.g., deconvolution

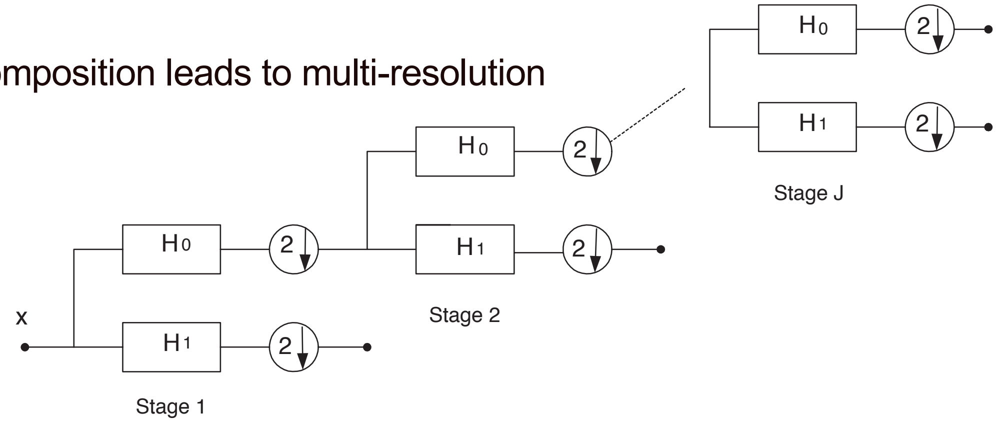
- Iterative shrinkage:
  - $\min_{\alpha}(\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{W}^{-1}\boldsymbol{\alpha}\|^2+\lambda\|\boldsymbol{\alpha}\|_1)$ where $\boldsymbol{y}$ is the blurred image and $\boldsymbol{x}=\boldsymbol{W}^{-1}\boldsymbol{\alpha}$ is the target image
  - $\boldsymbol{\alpha}_k=S_\lambda(\boldsymbol{\alpha}_{k-1}+\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{W}^{-1}\boldsymbol{\alpha}_{k-1}))$

Implementation of the Wavelet Transform

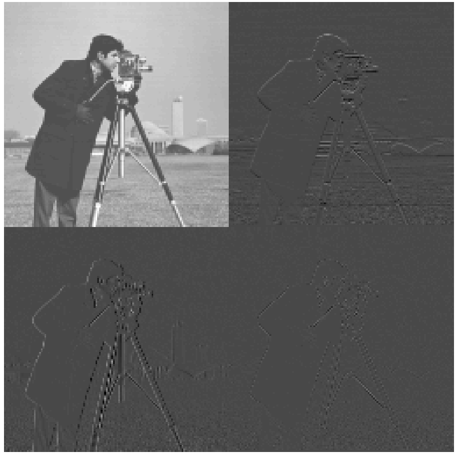
- Two-channel filter-bank



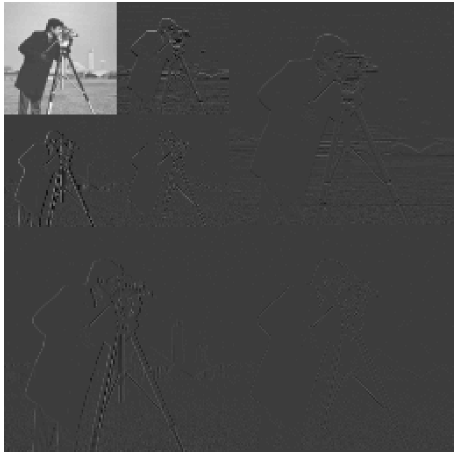
- Iterated decomposition leads to multi-resolution



Implementation of the 2-D Wavelet Transform



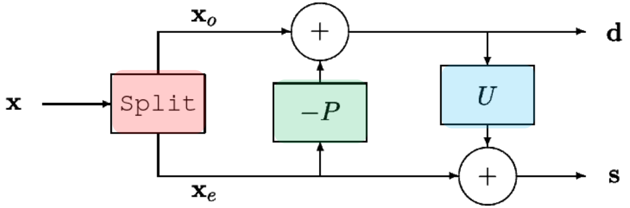
Implementation of the 2-D Wavelet Transform



Imperial College London

Wavelets and INN

- The wavelet transform can be implemented using the lifting scheme



Split $\longrightarrow$ $\begin{cases} d = x_o - P(x_e) \\ s = x_e + U(d) \end{cases}$ $\qquad$ $\begin{cases} x_o = d + P(x_e) \\ x_e = s - U(d) \end{cases}$ $\longrightarrow$ Merge

Forward pass $\qquad$ Backward pass

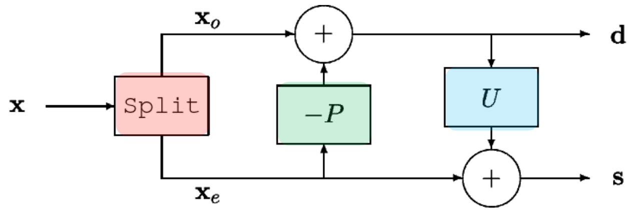
- The predictor (P) predicts the odd samples using the even, the update (U) uses the prediction error to smooth the even samples
- Predictor/update are fixed
- The scheme is perfectly invertible

I. Daubechies and W. Sweldens, “Factoring Wavelet Transforms into lifting Steps” 1997

Imperial College London

Wavelets and INN

- Can we learn a wavelet-like non-linear sparsifying transform?



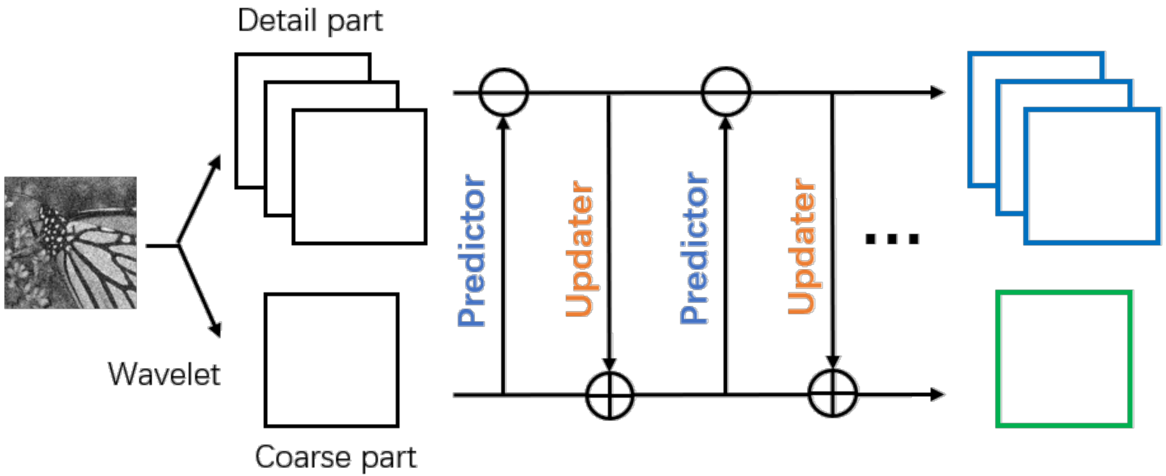
Split $\longrightarrow \begin{cases} d = x_o - P(x_e) \\ s = x_e + U(d) \end{cases}$ $\qquad$ $\begin{cases} x_o = d + P(x_e) \\ x_e = s - U(d) \end{cases} \longrightarrow$ Merge

Forward pass $\qquad$ Backward pass

- Approach:
  - convert the P/U operators into two deep networks and learn them
  - Use denoising as the bottleneck to impose sparsity

Wavelets and INN

- Can we learn a wavelet-like non-linear sparsifying transform?

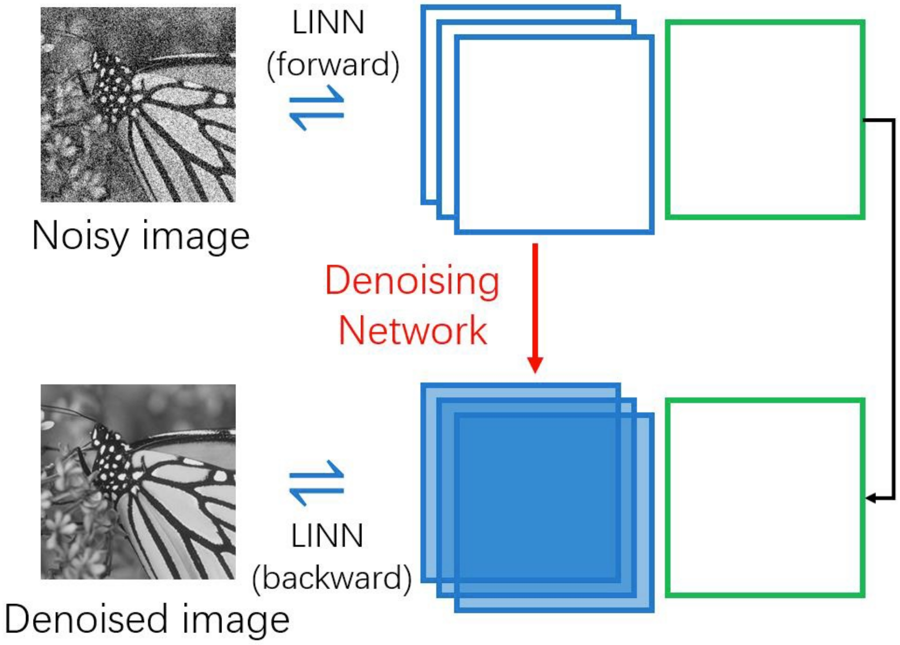


- Approach:
  - convert the P/U operators into two deep networks and learn them
  - Use denoising as the bottleneck to impose sparsity

Imperial College London

Wavelets and INN

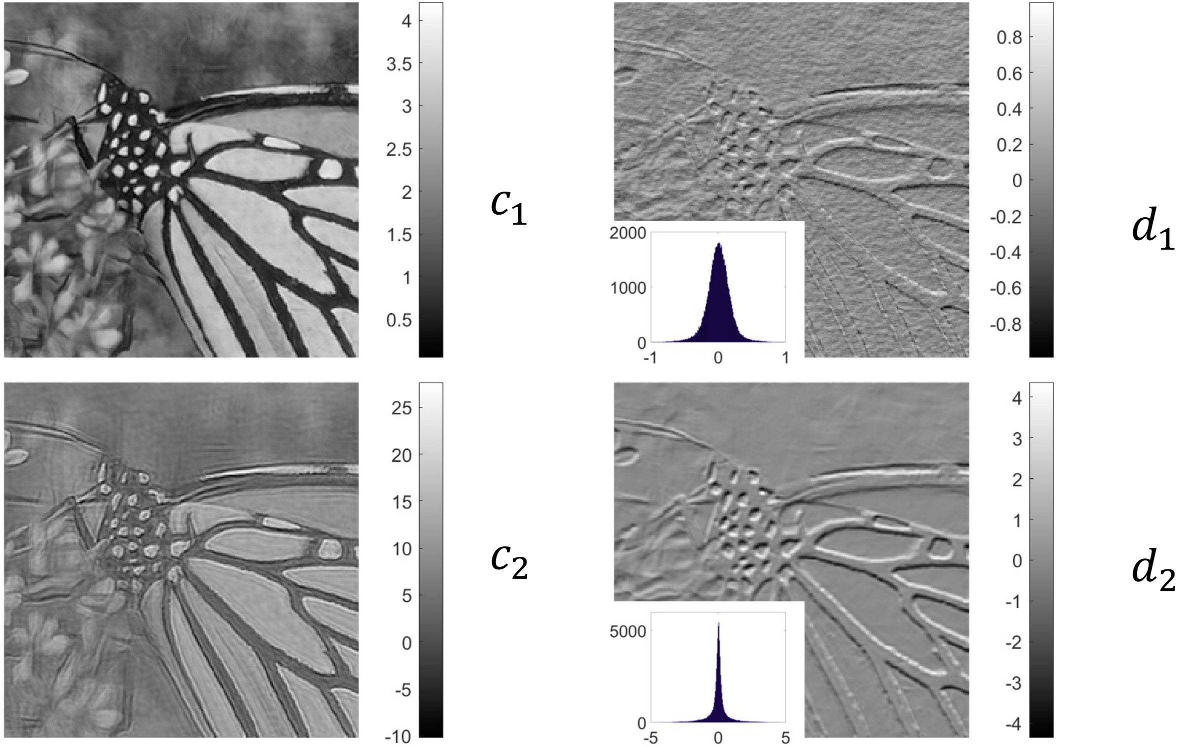
- To make sure P acts as a sparsifying predictor:
  - Train the network with noisy/noiseless image pairs
  - Add a denoising network on the details



Imperial College London

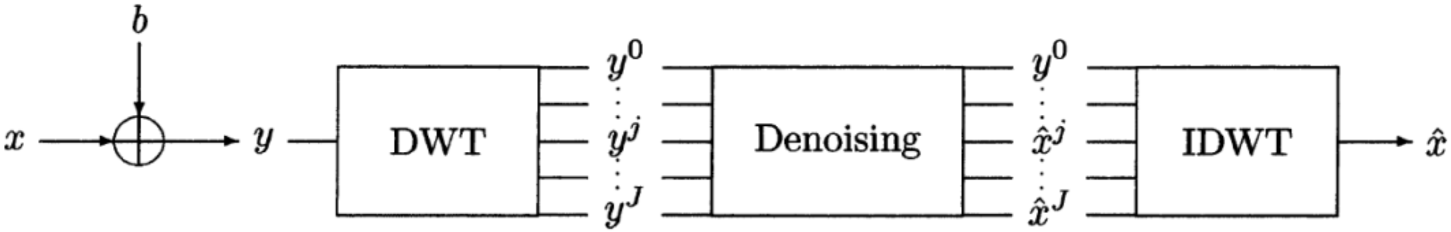
Signal Decomposition

- Training with noiseless/noisy pairs leads to a sparsifying transform
- Each piece of the network is interpretable
- As for wavelets, we can now use the INN for e.g., denoising or deconvolution

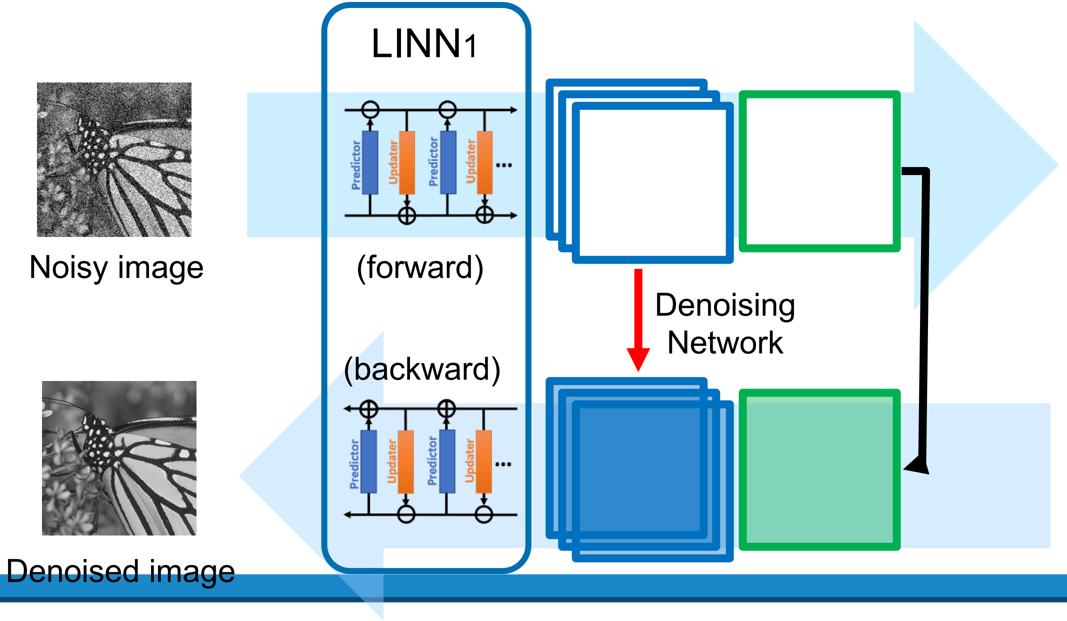


Imperial College London

Denoising - Overall Method

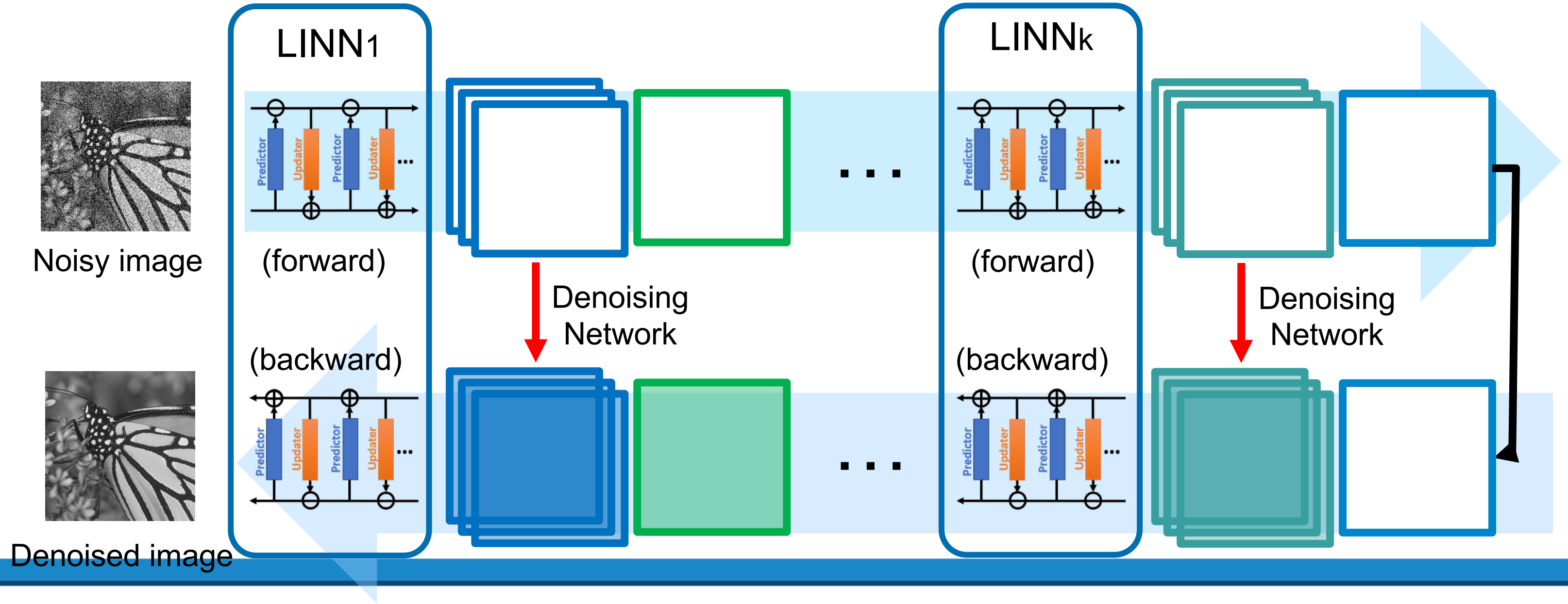


Denoising - Overall Method



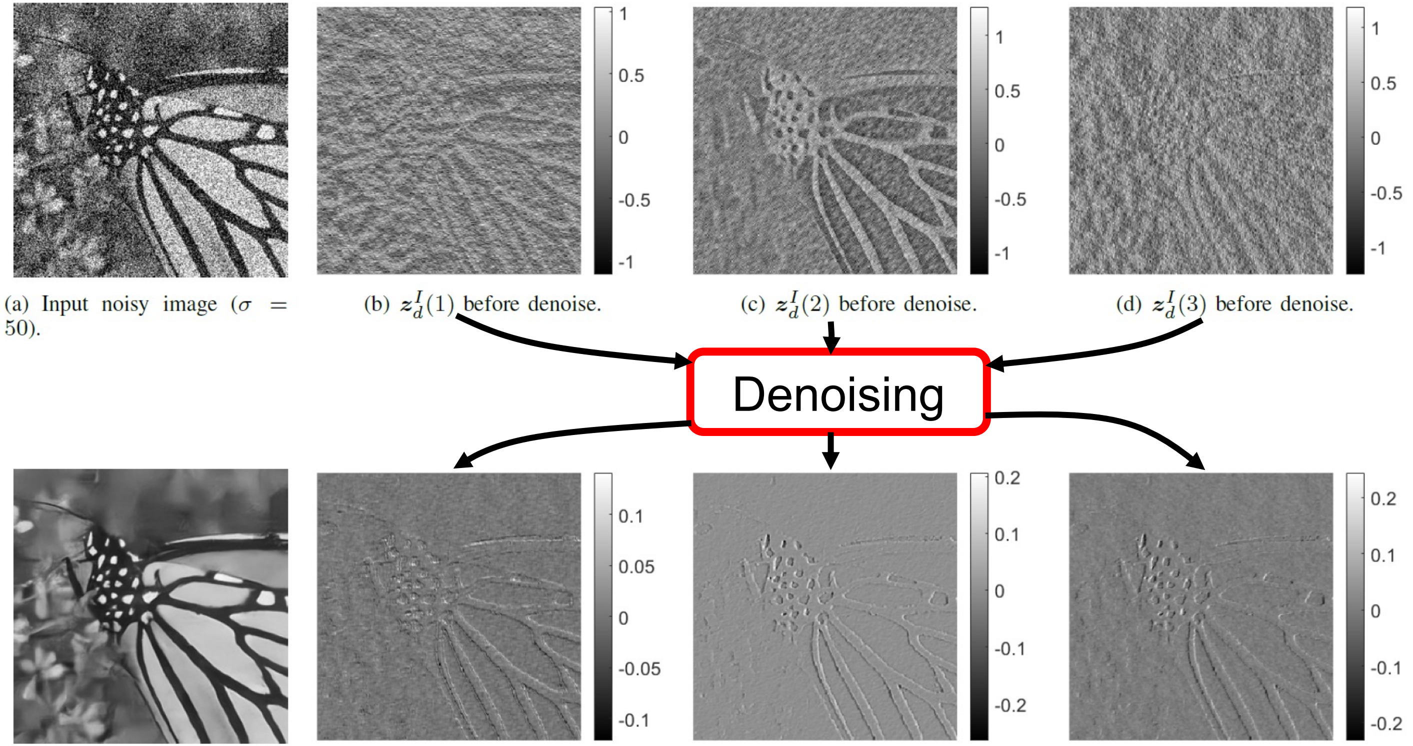
Imperial College London

Denoising - Overall Method



Imperial College London

Denoising



(a) Input noisy image ($\sigma = 50$).

(b) $z_d^I(1)$ before denoise.

(c) $z_d^I(2)$ before denoise.

(d) $z_d^I(3)$ before denoise.

(e) The denoised image (PSNR=26.75 dB).

(f) $z_d^I(1)$ after denoise.

(g) $z_d^I(2)$ after denoise.

(h) $z_d^I(3)$ after denoise.

Imperial College London

Results

Denoising:

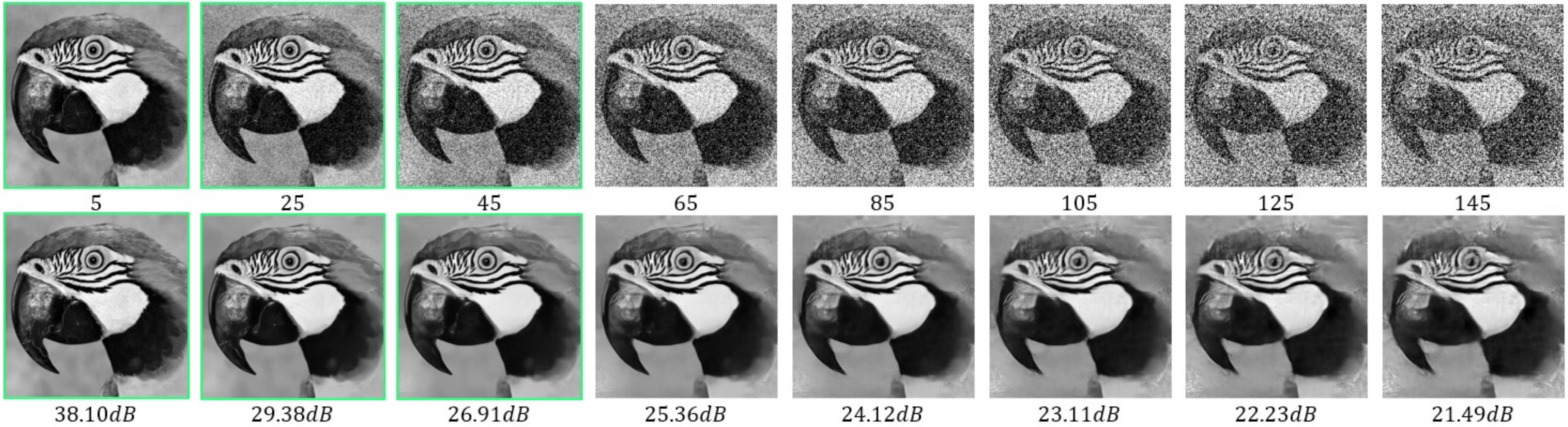


Image Deblurring

Imperial College London

Simulation Results – Image Deblurring

Algorithm 1: Plug-and-Play image deblurring with blind WINNet.

Input: Input image $\boldsymbol{y}$, kernel $\boldsymbol{k}$, parameter λ;

Initialize: $\boldsymbol{z}_0 = \boldsymbol{y}$, $\beta_0 = \text{NENet}(\boldsymbol{z}_0)$, $\beta_1 = 10 \times \beta_0$, $k = 1$;

while $\beta_k > \beta_0$ **do**

$\quad \boldsymbol{x}_k = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{k} \otimes \boldsymbol{x}\|_2^2 + \frac{\lambda \beta_0^2}{\beta_k^2} \|\boldsymbol{x} - \boldsymbol{z}_{k-1}\|_2^2$;

$\quad \beta_{k+1} = \text{NENet}(\boldsymbol{x}_k)$;

$\quad \boldsymbol{z}_k = \text{WINNet}(\boldsymbol{x}_k, 2\beta_{k+1})$;

$\quad k = k + 1$;

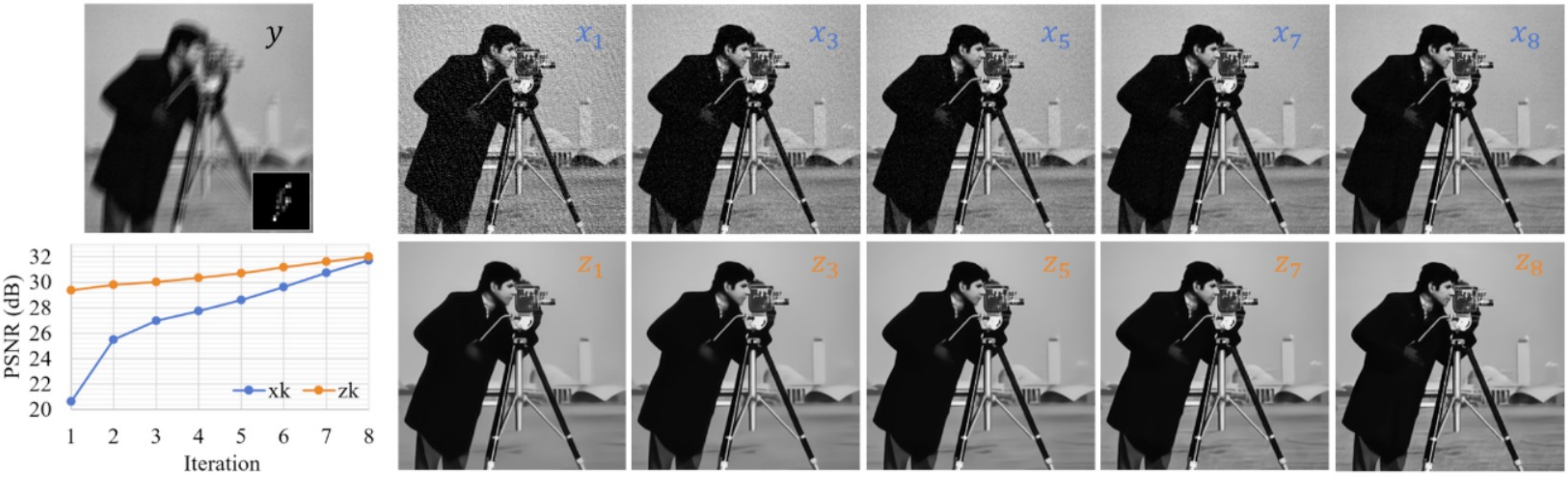
end

Output: Deblurred image $\boldsymbol{x} = \boldsymbol{z}_{k-1}$.

Imperial College
London

Results

Deconvolution:



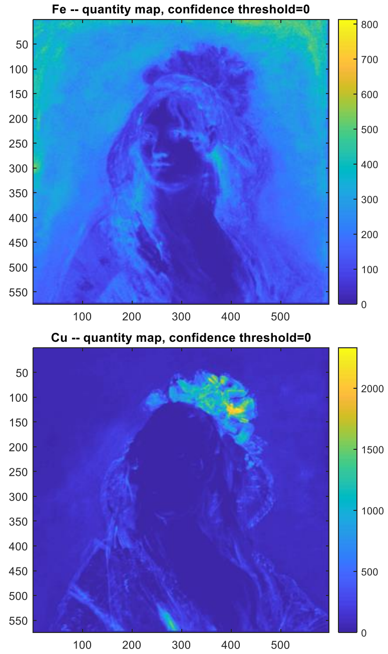
Imperial College London

First Set of Conclusions

- Invertible Neural Networks is an interesting new concept
- Designing INN using wavelets/lifting leads to a more interpretable network
- Good generalization ability

Why Multi-modal Image Processing?

Technical Study of Old Master Paintings[1]:
Data acquired using different imaging techniques

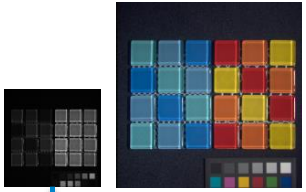


[1]joint project with UCL (M. Rodrigues), Duke and the National Gallery

Imperial College London

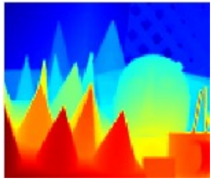
Why Multimodal Image Super-Resolution?

- Near-infrared (NIR)/RGB
- Multi-spectral/RGB
- Depth/RGB



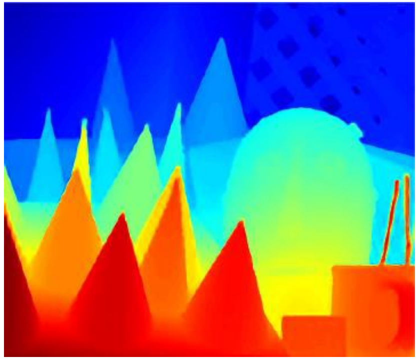
Low resolution !

Multimodal Image Super-Resolution (MISR)



LR Depth Image

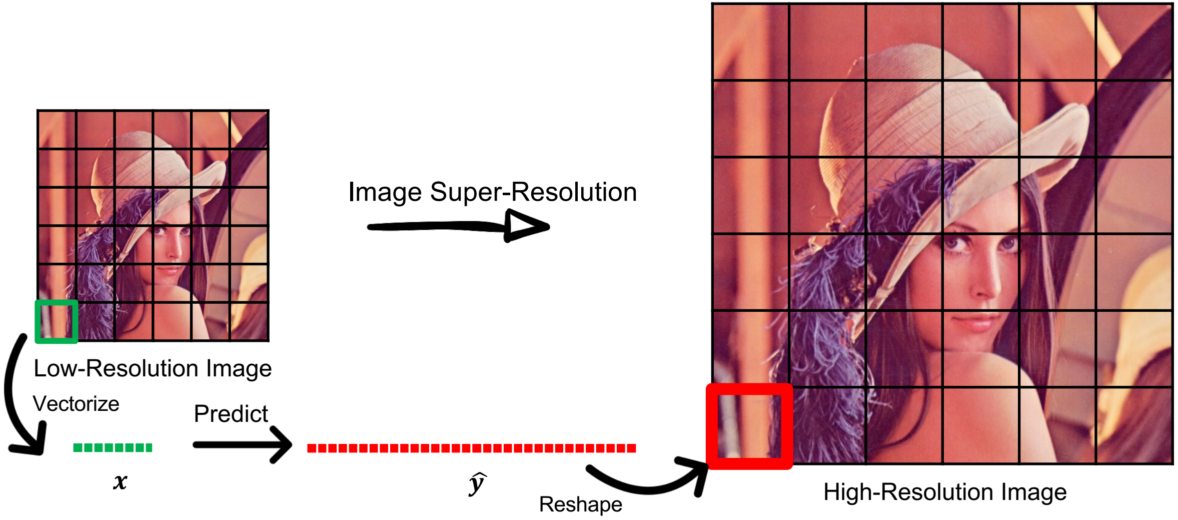
HR Color Image (guidance image)



Estimated HR Depth Image

Single Image Super-Resolution

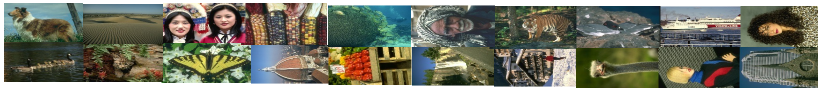
- Patch-based Prediction:



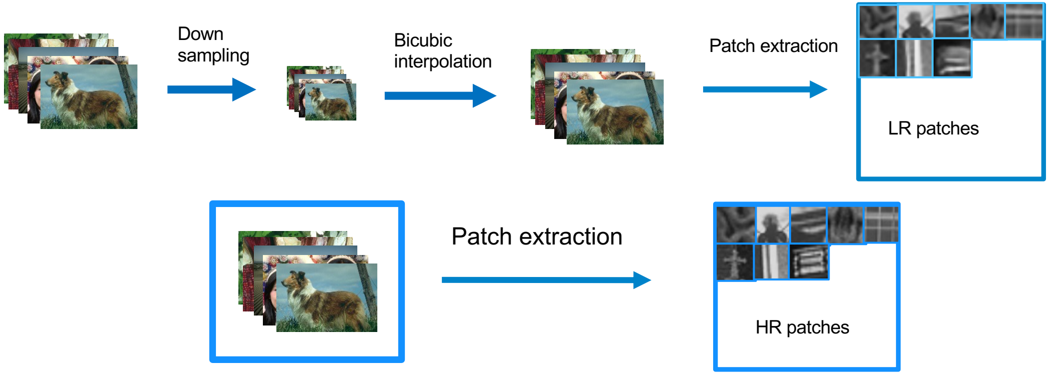
Imperial College London

Single Image Super-Resolution

Start with an external dataset of images (e.g., BSD 300 dataset)

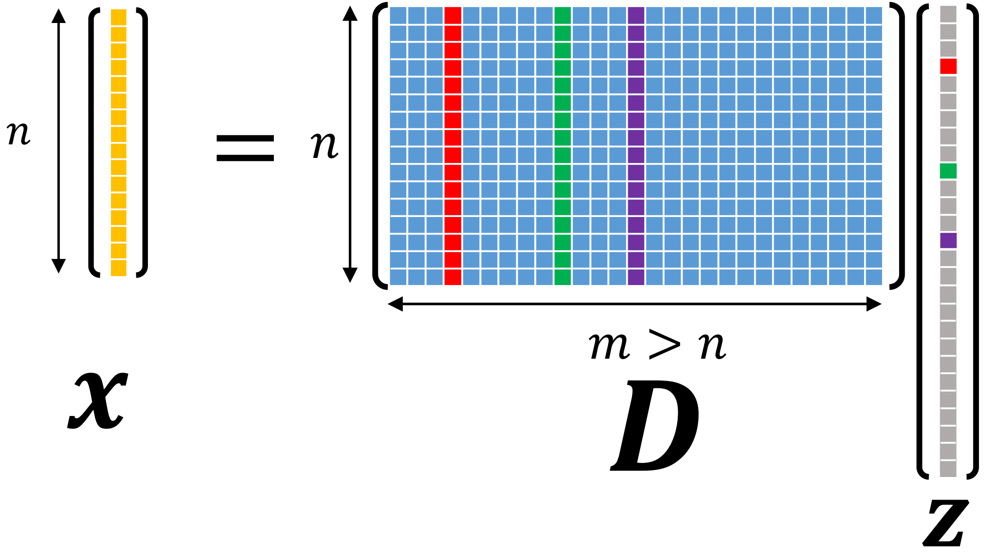


Extract pairs of LR and HR patches



Dictionary Learning for Super-Resolution

- The **key insight** is that images or patches of images have a sparse representation in a redundant dictionary
- The dictionary is usually learned



Super-Resolution Model

- One assumes that HR patches and LR patches admit a common sparse representation z_i :

$$\mathbf{x}_i^{LR} = \mathbf{D}^{LR}\mathbf{z}_i$$

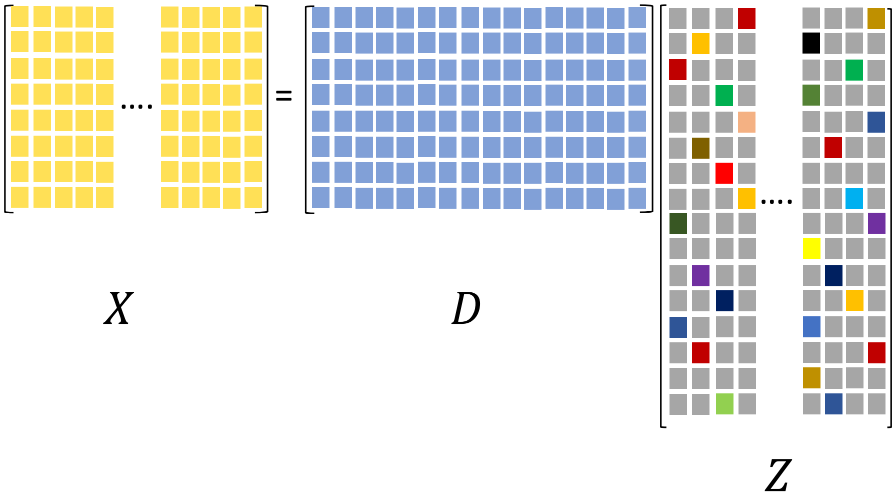
$$\mathbf{x}_i^{HR} = \mathbf{D}^{HR}\mathbf{z}_i$$

Imperial College London

Super-Resolution: Training

- Training:

1. Given $\boldsymbol{x}_i^{LR}$, learn $\boldsymbol{D}^{LR}$ and $\boldsymbol{z}_i$ using for example K-SVD or MOD



2. Given $\boldsymbol{x}_i^{HR}$ and $\boldsymbol{z}_i$ compute $\boldsymbol{D}^{HR}$ directly

Imperial College London

Multi-Modal Image Super-Resolution Model

- Approach:
  - Use sparse representation and dictionary learning to model dependency across modality and to drive the design of the neural network architecture through *unfolding*.

Imperial College London

Multi-Modal Image Super-Resolution Model

- Approach:

  - Use sparse representation and dictionary learning to model dependency across modality and to drive the design of the neural network architecture through *unfolding*.

  - **This is a trend now**: e.g Deep K-SVD [Elad et al.19], Neumann Networks [Willett-19], Deep Ultrasound [Eldar-19], Algorithm Unrolling [SP Mag, Eldar-21]

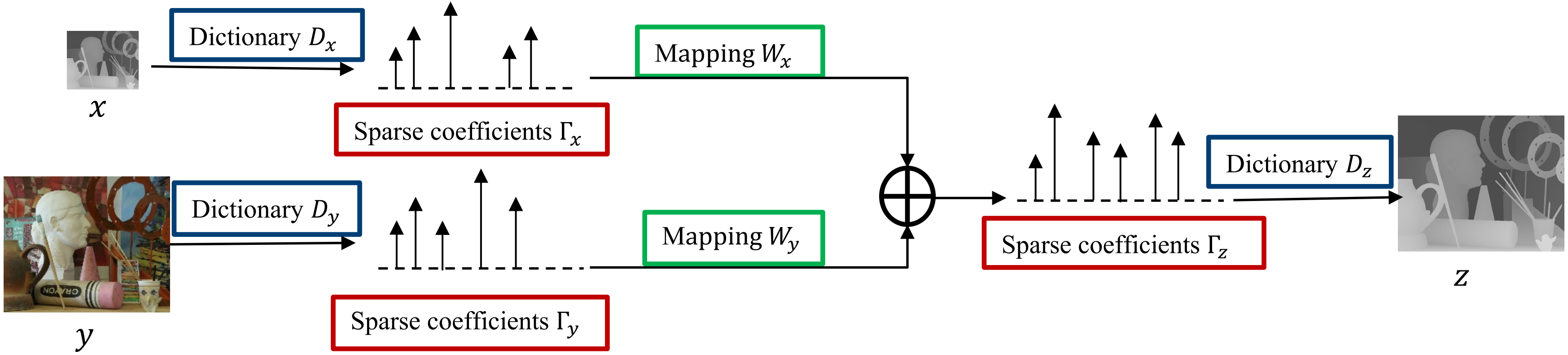
Multi-Modal Image Super-Resolution Model

- Approach:
  - Use sparse representation and dictionary learning to model dependency across modality and to drive the design of the neural network architecture through *unfolding*.

- Model:
  - In the multimodal case the two modalities share some but not all latent features

Imperial College London

Joint multimodal dictionary learning (JMDL)



- Assume patches x, y, z are sparse in learned dictionaries D_x, D_y and D_z, we can have the followings:

$$\begin{aligned} x &\simeq D_x a, \\ y &\simeq D_y b, \\ z &\simeq D_z c, \end{aligned}$$

where a, b, c are the sparse representations for x, y, z, respectively.

Joint multimodal dictionary learning (JMDL)

- Since x, y, z are from the same image scene, we assume the sparse representation c of z can be inferred from the others:

$$\boldsymbol{c} = S_{\boldsymbol{\lambda}_c}(\boldsymbol{W}_x \boldsymbol{a} + \boldsymbol{W}_y \boldsymbol{b})$$

where W_x, W_y are the transform matrices to be learned.

- JMDL optimization problem:

$$\begin{aligned}
\min_{\left\{\substack{\boldsymbol{D}_x, \boldsymbol{D}_y, \boldsymbol{D}_z, \\ \boldsymbol{\Gamma}_x, \boldsymbol{\Gamma}_y, \boldsymbol{\Gamma}_z, \\ \boldsymbol{W}_x, \boldsymbol{W}_y}\right\}} \; & \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{D}_x \boldsymbol{\Gamma}_x\|_F^2 + \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{D}_y \boldsymbol{\Gamma}_y\|_F^2 \\
& + \frac{1}{2}\|\boldsymbol{Z} - \boldsymbol{D}_z \boldsymbol{\Gamma}_z\|_F^2 + \nu_x \|\boldsymbol{\Gamma}_x\|_1 + \nu_y \|\boldsymbol{\Gamma}_y\|_1 \\
& + \nu_z \|\boldsymbol{\Gamma}_z\|_1 + \mu_x \|\boldsymbol{W}_x\|_F^2 + \mu_y \|\boldsymbol{W}_y\|_F^2 \\
& + \alpha \|\boldsymbol{\Gamma}_z - \boldsymbol{W}_x \boldsymbol{\Gamma}_x - \boldsymbol{W}_y \boldsymbol{\Gamma}_y\|_F^2, \\
& s.t., \|\boldsymbol{d}_{x,i}\|_2^2 \le 1, \|\boldsymbol{d}_{y,i}\|_2^2 \le 1, \|\boldsymbol{d}_{z,i}\|_2^2 \le 1, \forall i,
\end{aligned}$$

- Solving strategy:

  I. Fix D_x, D_y, D_z, W_x, W_y to update $\Gamma_x, \Gamma_y, \Gamma_z$.

  II. Fix $\Gamma_x, \Gamma_y, \Gamma_z, W_x, W_y$ to update D_x, D_y, D_z.

  III. Fix $\Gamma_x, \Gamma_y, \Gamma_z, D_x, D_y, D_z$ to update W_x, W_y.

Imperial College London

Deep coupled ISTA network

- In the reconstruction phase, given x and y, we first need to calculate their sparse representations based on the learned dictionaries D_x and D_y:

$$\min_{\{\boldsymbol{a},\boldsymbol{b}\}} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{D}_x\boldsymbol{a}\|_2^2+\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{D}_y\boldsymbol{b}\|_2^2+\lambda_x\|\boldsymbol{a}\|_1+\lambda_y\|\boldsymbol{b}\|_1 .$$

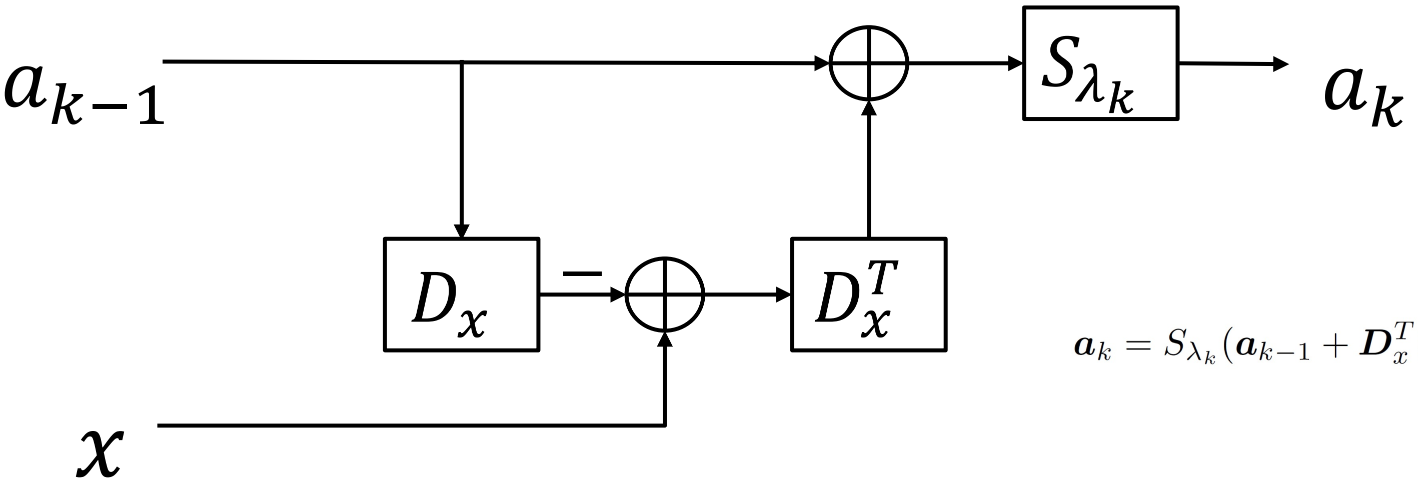
- Solved by ISTA algorithm:

$$\boldsymbol{a}_k = S_{\lambda_k}(\boldsymbol{a}_{k-1}+\boldsymbol{D}_x^T(\boldsymbol{x}-\boldsymbol{D}_x\boldsymbol{a}_{k-1}))$$

- Inspired by LISTA[3], we “unfold” this iteration to obtain two deep networks (one per modality)

[3] Gregor Karol and LeCunYann, “Learning fast approximations of sparse coding ”, Proceedings of the 27th International Conference on International Conference on Machine Learning, 2010

Deep coupled ISTA network



$$\boldsymbol{a}_k = S_{\lambda_k}(\boldsymbol{a}_{k-1} + \boldsymbol{D}_x^T(\boldsymbol{x} - \boldsymbol{D}_x\boldsymbol{a}_{k-1}))$$

Inspired by LISTA[3], we "unfold" this iteration to obtain two deep networks (one per modality)

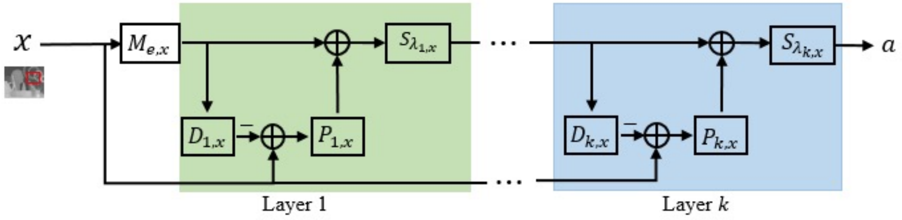
[3] Gregor Karol and LeCunYann, "Learning fast approximations of sparse coding ", Proceedings of the 27th International Conference on International Conference on Machine Learning, 2010

Imperial College London

Deep coupled ISTA network

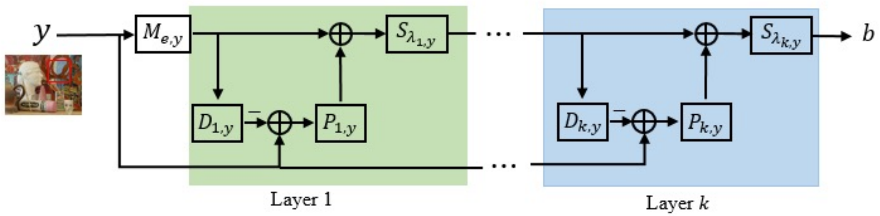
- Solving by ISTA algorithm through unfolding:

$$\boldsymbol{a}_k = S_{\lambda_k}(\boldsymbol{a}_{k-1} + \boldsymbol{D}_x^T(\boldsymbol{x} - \boldsymbol{D}_x\boldsymbol{a}_{k-1}))$$



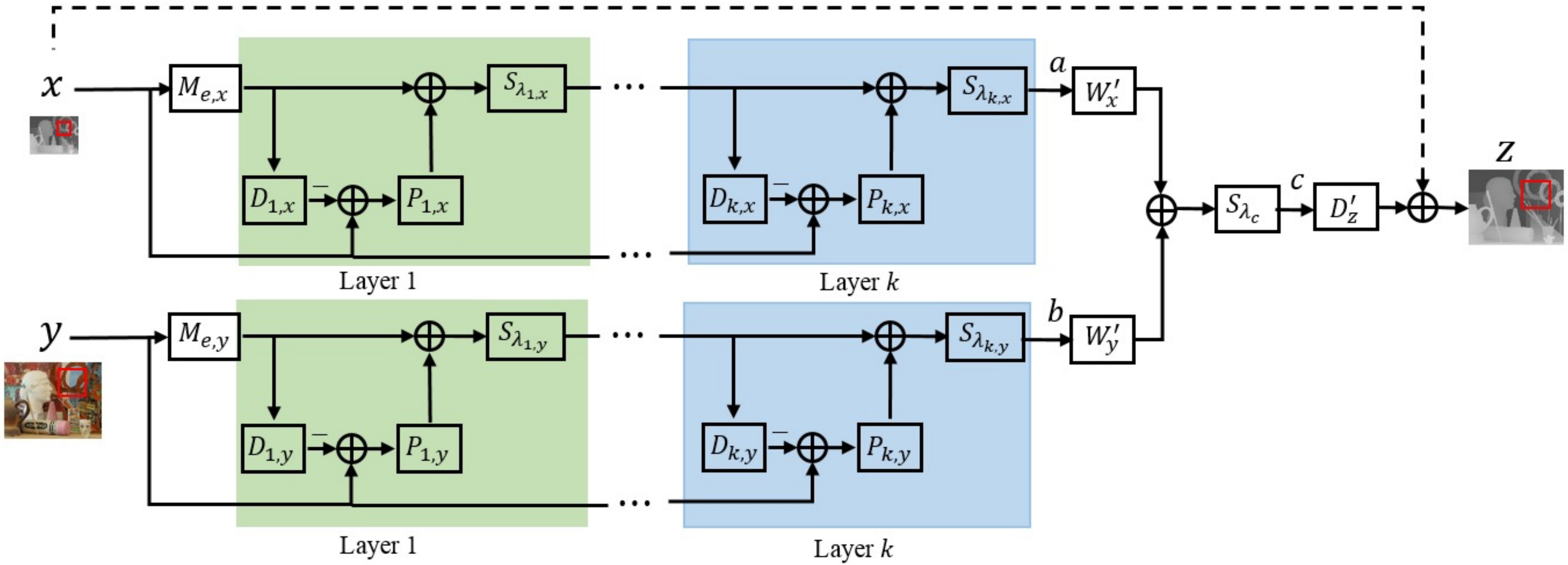
and

$$\boldsymbol{b}_k = S_{\lambda_k}(\boldsymbol{b}_{k-1} + \boldsymbol{D}_y^T(\boldsymbol{y} - \boldsymbol{D}_y\boldsymbol{b}_{k-1}))$$



Imperial College London

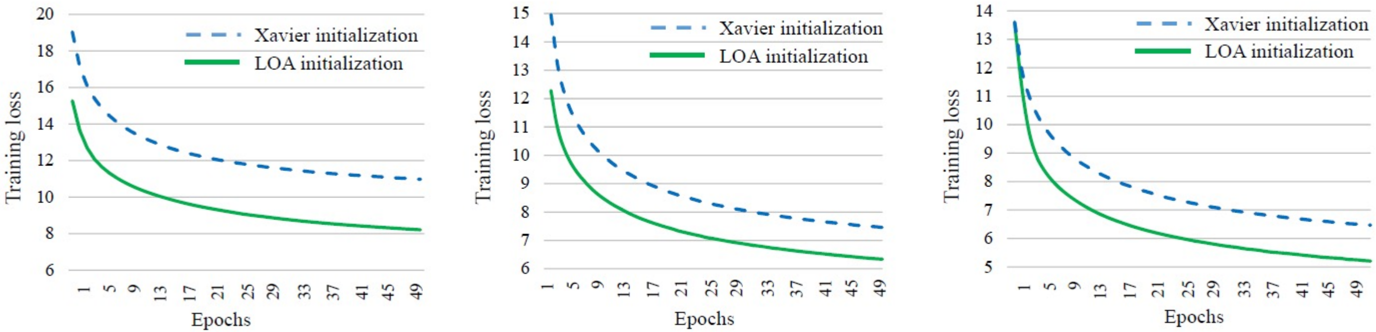
Deep coupled ISTA network



Imperial College London

Numerical results

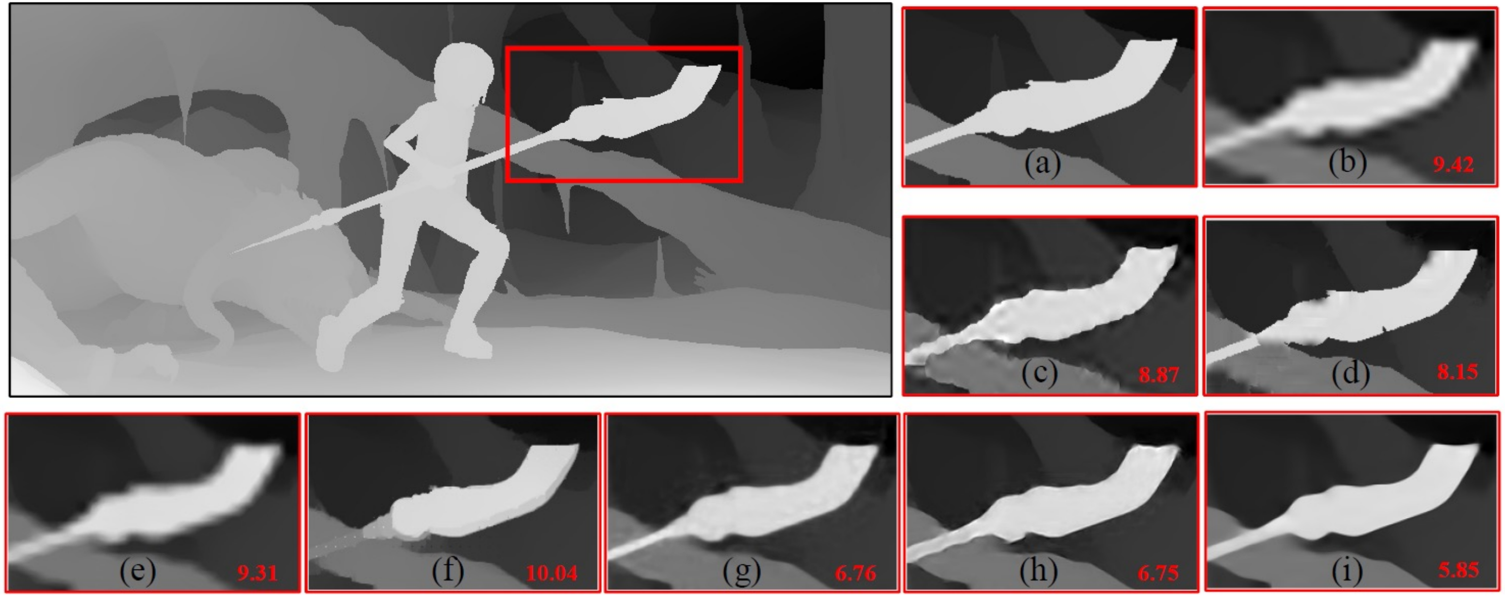
- Effectiveness of LOA



(a) Case I: $m = 256, K = 2$ (b) Case II: $m = 512, K = 2$ (c) Case III: $m = 512, K = 4$

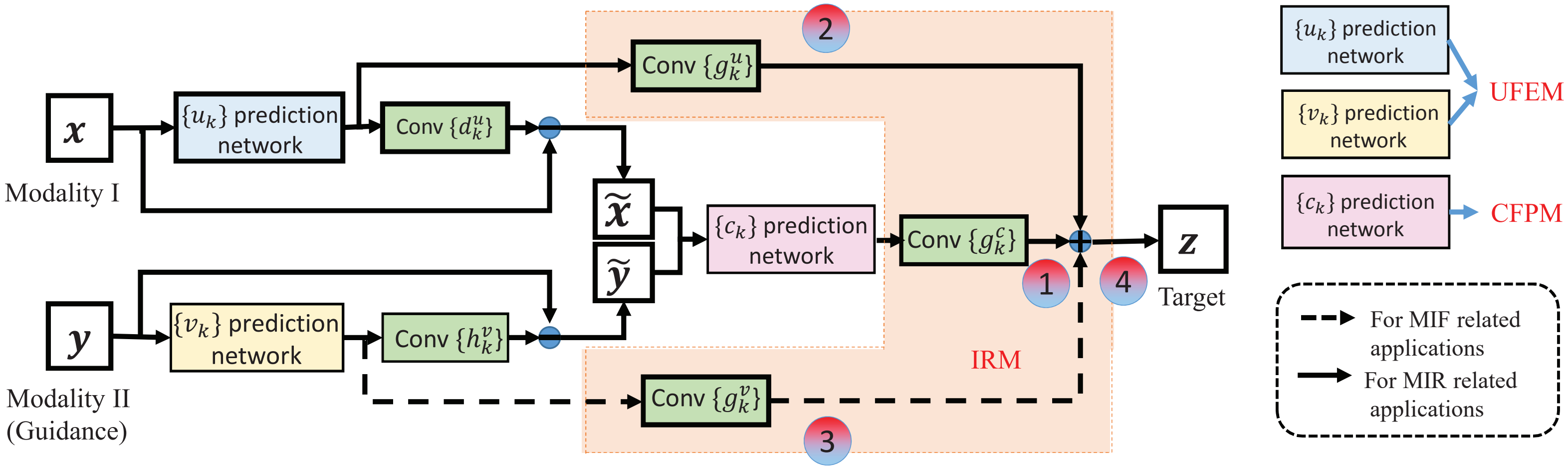
The training loss across 50 epochs with LOA and Xavier initialization methods for different settings of dictionary size m and number of layers K.

Visual comparisons

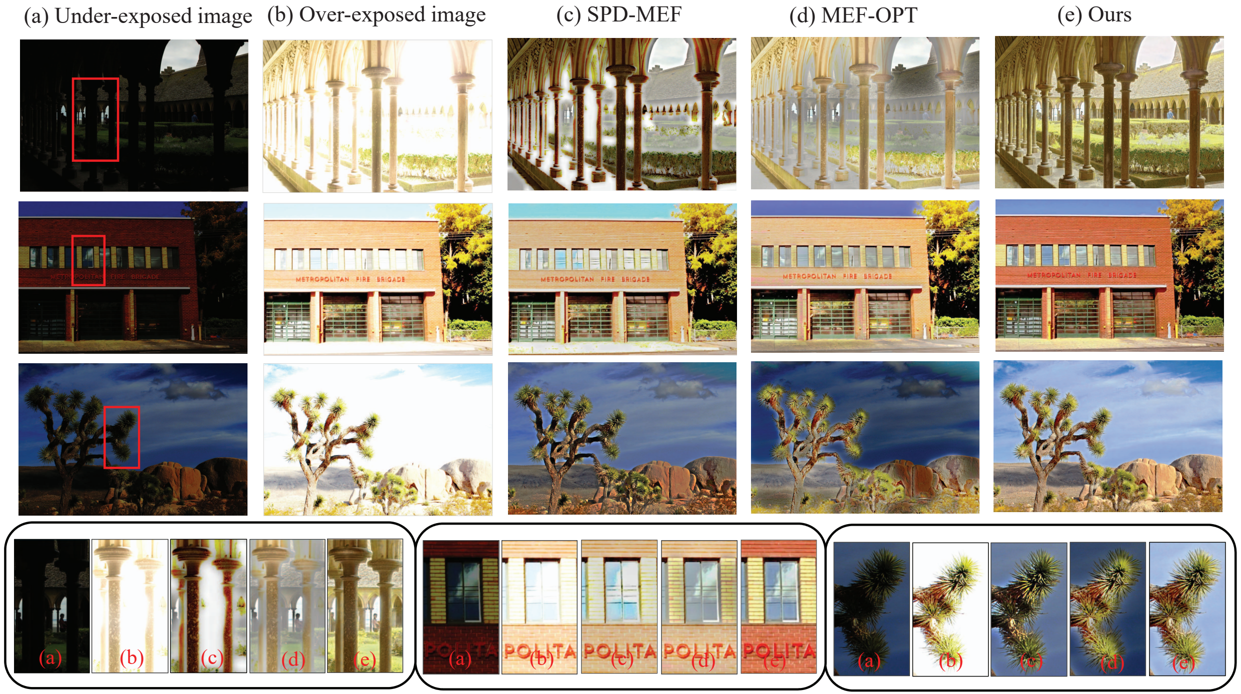


Visual comparisons of *Cave* in Sintel dataset with upscaling factor = 8 using different methods. (a) Ground truth, (b) Bicubic, (c) Park et al. [61], (d) Lu et al. [62], (e) Gu et al. [30], (f) Ferstl et al. [58], (g) SCN [12], (h) VDSR [13], (i) Ours. The numbers in red indicate the RMSE values.

Unfolding Convolutional Dictionaries



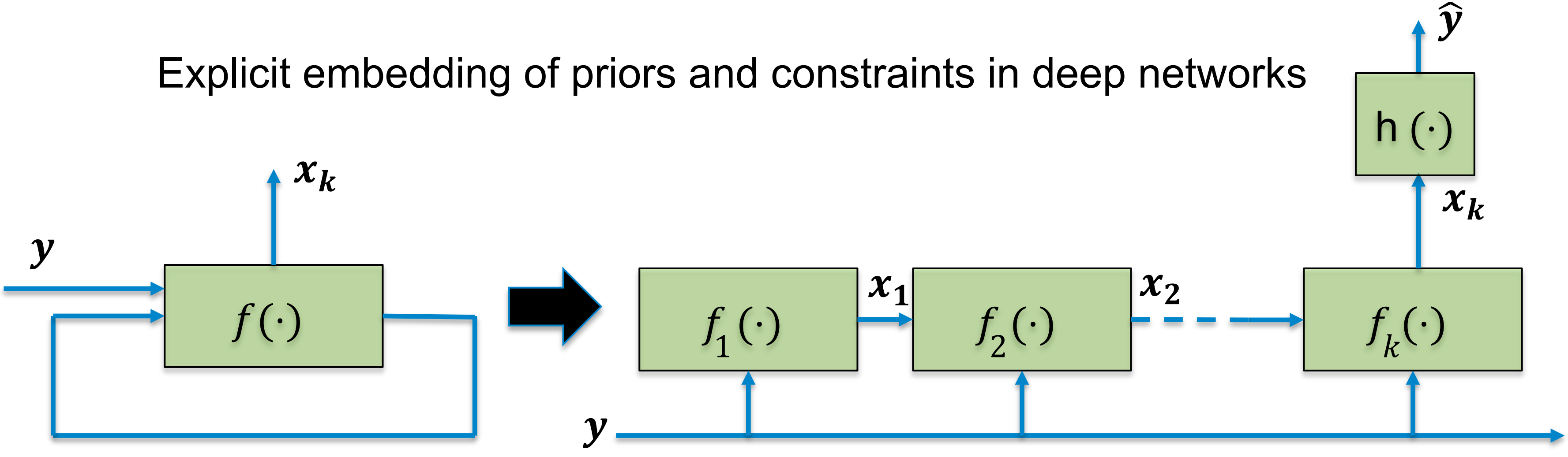
Visual comparisons



Imperial College London

Unfolding Strategy

Explicit embedding of priors and constraints in deep networks



Iterative algorithm with y as input and x as output

Unfolded version of the iterative algorithm with learnable parameters

Need to re-synthesize the input, if self-supervised

Imperial College London

Art-Investigation

- Goal: Use multi-modal imaging techniques
  - for material characterization
  - to discover underdrawings and concealed design

Visible

X-ray

Imperial College London

Art-Investigation

- Goal: we want to separate the two x-ray images

- Approach:
  - Use the visible RGB image as side information (x-ray visible similar to RGB image)
  - Exclusion loss: the “contours” of the two x-ray images should be as different as possible

Visible

X-ray

Imperial College London

X-ray Separation – Proposed Sparsity Model

$$\boldsymbol{x}_1 = \sum_{k=1}^{K} \boxed{\boldsymbol{\Xi}_k} * \boxed{\boldsymbol{z}_{1,k}}, \quad \boldsymbol{x}_2 = \sum_{k=1}^{K} \boxed{\boldsymbol{\Xi}_k} * \boldsymbol{z}_{2,k},$$

$$\boldsymbol{r}_{1,s} = \sum_{k=1}^{K} \boldsymbol{\Omega}_{k,s} * \boxed{\boldsymbol{z}_{1,k}}, \quad \boxed{\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{\Xi}_k * (\boldsymbol{z}_{1,k} + \boldsymbol{z}_{2,k}),}$$

- The visible image and the two separated X-ray images have a sparse representation in proper dictionaries
- RGB image and visible X-ray share the same sparse representation
- The two X-rays x_1, x_2 share the same dictionary
- The measured X-ray is $x = x_1 + x_2$

Visible

X-ray

X-ray Separation – Exclusion Loss

Imperial College London

- Given the reconstructed X-ray images x_1, x_2, we expect that their edges are as different as possible we therefore add an "exclusion term" in the optimization

$$
\begin{aligned}
\min_{\boldsymbol{y}_1,\boldsymbol{y}_2,\boldsymbol{z}_{1,k},\boldsymbol{z}_{2,k}} \quad & \|\boldsymbol{x} - \boldsymbol{\Psi} * \boldsymbol{y}_1 - \boldsymbol{\Psi} * \boldsymbol{y}_2\|_F^2 \\
& + \tau_1 \|\boldsymbol{y}_1 - \sum_{k=1}^{K} \boldsymbol{\Theta}_k * \boldsymbol{z}_{1,k}\|_F^2 \\
& + \tau_2 \|\boldsymbol{y}_2 - \sum_{k=1}^{K} \boldsymbol{\Theta}_k * \boldsymbol{z}_{2,k}\|_F^2 \\
& + \gamma \sum_{s=1}^{3} \|\boldsymbol{r}_{1,s} - \boldsymbol{\Phi}_s * \boldsymbol{y}_1\|_F^2 \\
& + \lambda_1 \sum_{k=1}^{K} \|\boldsymbol{z}_{1,k}\|_1 + \lambda_2 \sum_{k=1}^{K} \|\boldsymbol{z}_{2,k}\|_1 \\
& + \sum_{i=1}^{I} \mu_i \|(\boldsymbol{W}_i * \boldsymbol{y}_1) \odot (\boldsymbol{W}_i * \boldsymbol{y}_2)\|_1,
\end{aligned}
$$

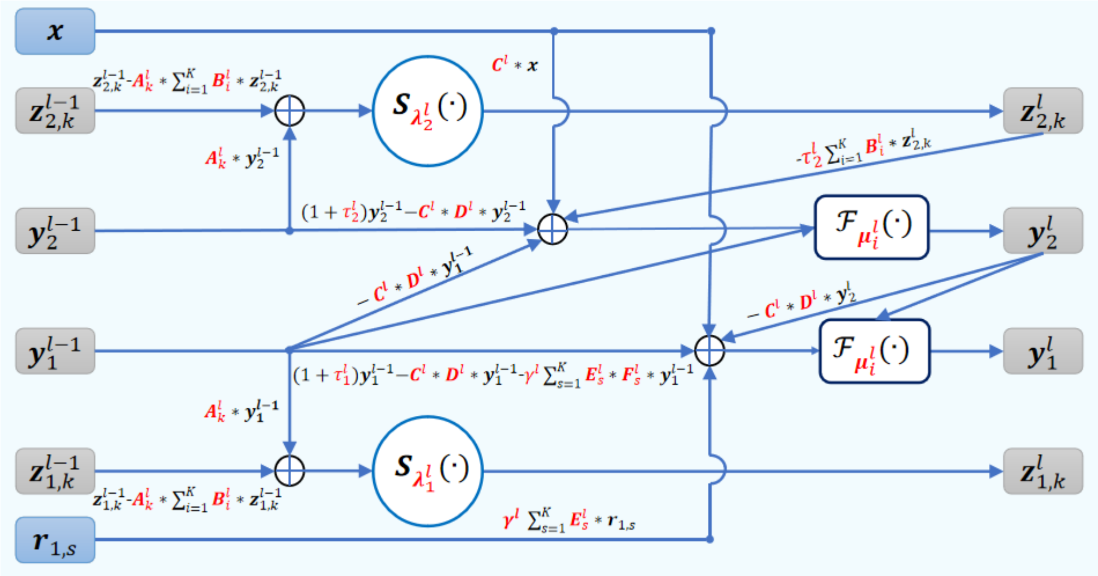
Visible

X-ray

Imperial College London

One Layer of the Network

- The sparsity model and the exclusion constraint leads to an iterative optimization method which leads to a network through unfolding



Imperial College London

Separation Results

Imperial College London

Conclusions

- Cross fertilization between dictionary learning/sparse representation and deep learning is fruitful

- Dictionary Learning/sparsity useful:
  - to impose models and structure to the deep network (through sparse modelling and optimization)
  - To design wavelet-like INN
  - For better interpretability and generalization ability

Related Publications

- **J. Huang** and P.L. Dragotti, “LINN: Lifting Inspired Invertible Neural Network for Image Denoising”, in proc. of 29th European Signal Processing Conference, EUSIPCO 2021
- **J. Huang** and P.L. Dragotti, “WINNet: Wavelet-inspired Invertible Network for Image Denoising”, submitted to IEEE Transactions on Image Processing, September 2021, https://arxiv.org/abs/2109.06381
- **X Deng** and P. L. Dragotti, “Deep Coupled ISTA Network for Multi-modal Image Super-Resolution, IEEE Transactions on Image Processing, pp 1683-1698, vol.29, 2020
- **X Deng** and P. L. Dragotti, “Deep Convolutional Neural Network for Multi-modal Image Restoration and Fusion IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, October 2021
- **W. Pu, J. Huang** et al., “Mixed X-Ray Image Separation for Artworks with Concealed Designs”, submitted to IEEE Transactions on Image Processing, January 2022, https://arxiv.org/abs/2201.09167