Joint Image Super-Resolution via Recurrent Convolutional Neural Networks with Coupled Sparse Priors

Iman Marivani, Evaggelia Tsiligianni, Bruno Cornelis and Nikos Deligiannis

Vrije Universiteit Brussel, Brussels, Belgium - imec, Leuven, Belgium

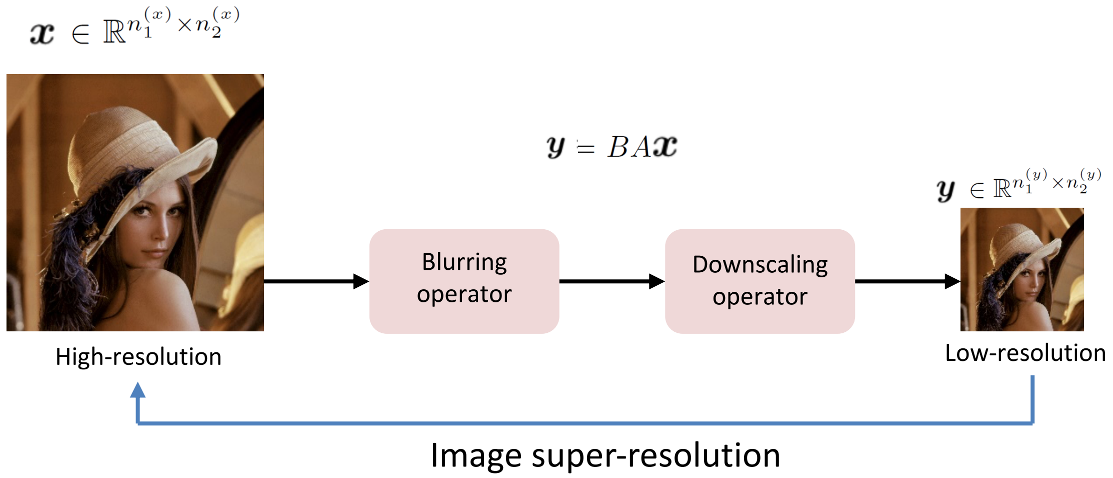
Presenter: Iman Marivani

Session: TEC-05 -- Machine Learning for Image and Video Processing III

Image Super-resolution

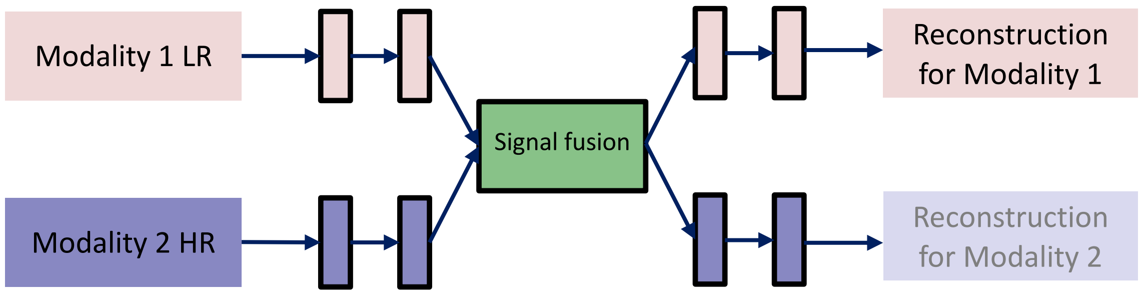
Single image super-resolution:

Reconstruction of a HR image X given its LR version Y



Multimodal image super-resolution:

Reconstruction of a HR image X given its LR image Y guided by a HR image Z from another modality



Different signal modalities:

- RGB
- Depth
- NIR
- Thermal
- Multi- / Hyper-spectral Imaging
- Medical Imaging

VRIJE UNIVERSITEIT BRUSSEL imec

Single image super-resolution with convolutional sparse priors

Assumption: LR image $\boldsymbol{y}$ and HR image $\boldsymbol{x}$ share the same sparse representation

LR image: $\boldsymbol{y} = \sum_{i=1}^{k} \boldsymbol{d}_i^y * \boldsymbol{\alpha}_i^y$

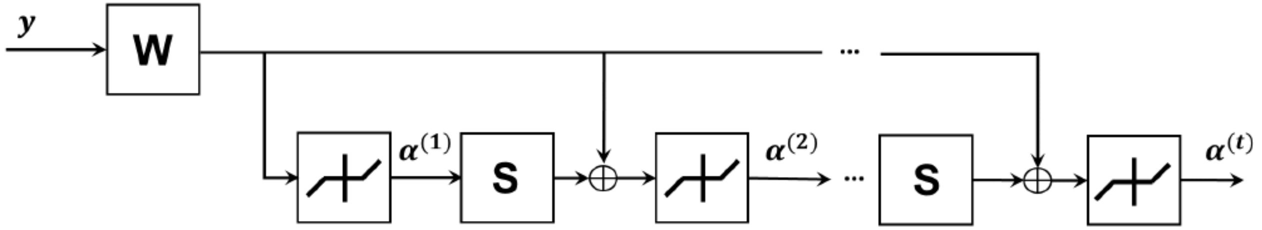
HR image: $\boldsymbol{x} = \sum_{i=1}^{k} \boldsymbol{d}_i^x * \boldsymbol{\alpha}_i^x$

Shared sparse features from: $\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{y} - \sum_{i=1}^{k} \boldsymbol{d}_i * \boldsymbol{\alpha}_i\|_2^2 + \lambda \sum_{i=1}^{k} \|\boldsymbol{\alpha}_i\|_1,$

Approximate convolutional sparse coding (ACSC):

$\boldsymbol{\alpha}^{t+1} = \phi_\gamma(\boldsymbol{S} * \boldsymbol{\alpha}^t + \boldsymbol{W} * \boldsymbol{y})$

Shrinkage function: $\phi_\gamma(u_i) = \operatorname{sign}(u_i)\max\{|u_i| - \gamma, 0\}$



Multimodal image SR via convolutional sparse coding with side information

LR image (modality 1): $\boldsymbol{y} = \sum_{i=1}^{k} \boldsymbol{d}_i^y * \boldsymbol{\alpha}_i^y$

HR image (modality 1): $\boldsymbol{x} = \sum_{i=1}^{k} \boldsymbol{d}_i^x * \boldsymbol{\alpha}_i^x$

HR side info (modality 2): $\boldsymbol{z} = \sum_{i=1}^{k} \boldsymbol{d}_i^z * \boldsymbol{\alpha}_i^z$

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{y} - \sum_{i=1}^{k} \boldsymbol{d}_i * \boldsymbol{\alpha}_i\|_2^2 + \lambda\left(\sum_{i=1}^{k} \|\boldsymbol{\alpha}_i\|_1 + \sum_{i=1}^{k} \|\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^z\|_1\right)$$

Learned multimodal convolutional sparse coding (LMCSC) for solving the problem above:

$$\boldsymbol{\alpha}^{t+1} = \xi_\mu(\boldsymbol{\alpha}^t - \boldsymbol{Q} * \boldsymbol{R} * \boldsymbol{\alpha}^t + \boldsymbol{P} * \boldsymbol{y}; \boldsymbol{\alpha}^z)$$

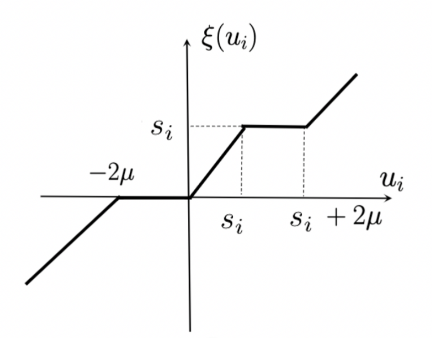
Shrinkage function (LeSITA operator):

for $s_i \geq 0, i = 1, \dots, m$:

$$\xi_\mu(u_i; s_i) = \begin{cases} u_i + 2\mu, & u_i < -2\mu \\ 0, & -2\mu \leq u_i \leq 0 \\ u_i, & 0 < u_i < s_i \\ s_i, & s_i \leq u_i \leq s_i + 2\mu \\ u_i - 2\mu, & u_i \geq s_i + 2\mu \end{cases}$$

for $s_i < 0, i = 1, \dots, m$:

$$\xi_\mu(u_i; s_i) = \begin{cases} u_i + 2\mu, & u_i < s_i - 2\mu \\ s_i, & s_i - 2\mu \leq u_i \leq s_i \\ u_i, & s_i < u_i < 0 \\ 0, & 0 \leq u_i \leq 2\mu \\ u_i - 2\mu, & u_i \geq 2\mu \end{cases}$$



Existing methods for multimodal image SR

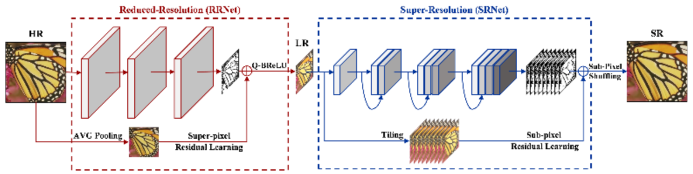
Analytical methods:

$$x = D_x\alpha, \quad \text{s.t.} \quad \alpha = \arg\min_{v\in\mathbb{R}^{n_\alpha}} \|y - D_y v\|_2^2 + \lambda\|v\|_1,$$

[J. Yang, J. Wright, T. Huang, Y. Ma 2010]

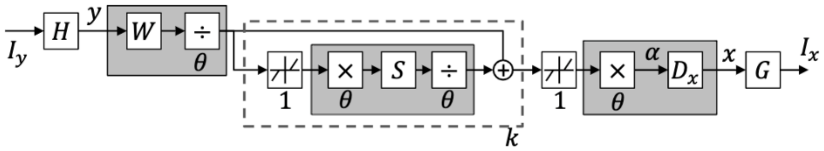
+ Taking the structure of the signal into account
+ These methods are explainable
− **Computationally expensive at training and inference**
− **Not practical in case of very large datasets**

Deep learning methods:



+ State-of-the-art performance
+ Fast inference
− **Structure of the signal is not considered**
− **The intermediate steps are not interpretable**
− **The fusion is performed blindly by a concatenation or linear combination**

Deep Unfolding methods:



[Z. Wang, D. Liu, J. Yang, W. Han, T. Huang 2015]

+ Interpretable structure
+ Neural network architecture, fast inference
− **Does not consider advances in deep learning**

Our goal is designing a deep network that:

Considers the structure of the signal and the prior knowledge

Follows the advances in deep learning

Fuses the signal representations using a principled method

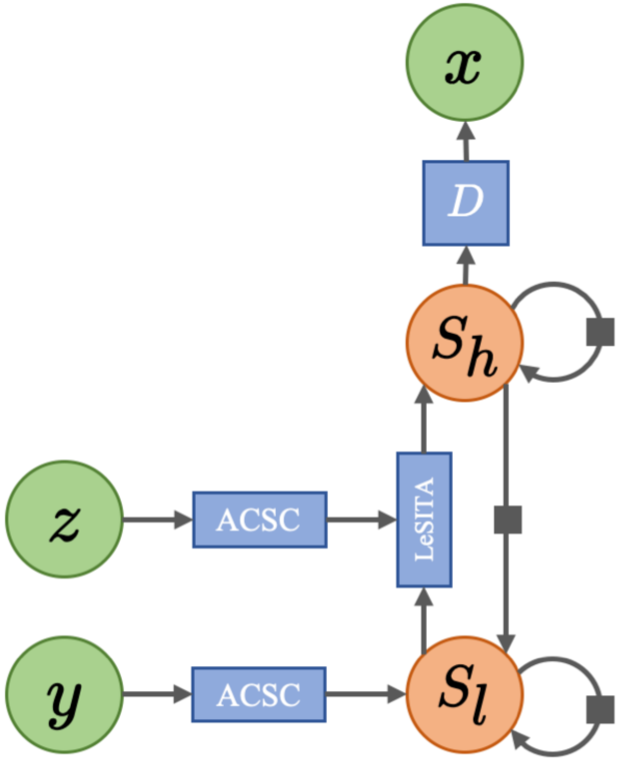
Main components for the network design

ACSC for the convolutional sparse feature map extractor

LeSITA for the fusion of the modalities

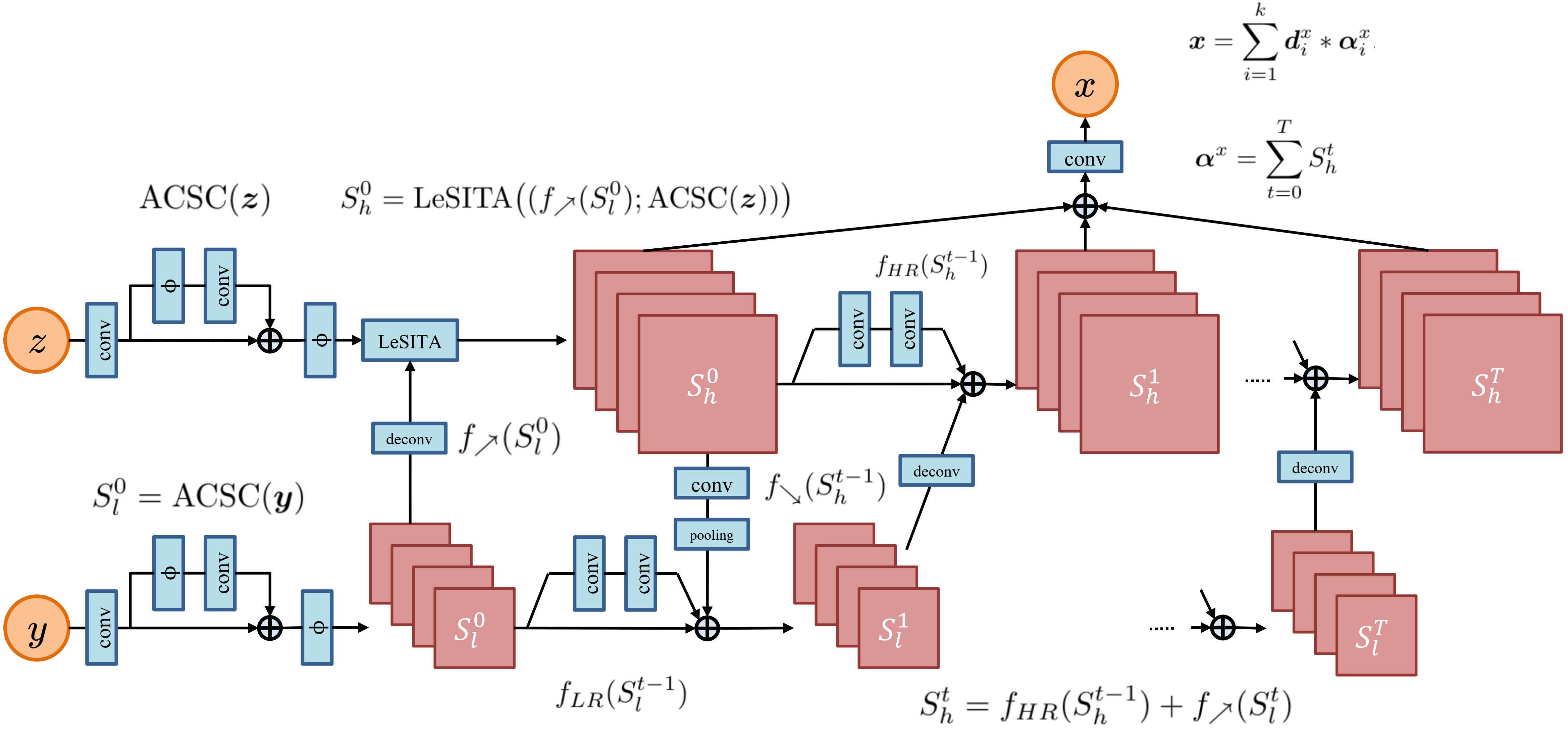
A dual state RNN to obtain the HR representations

A convolutional dictionary to reconstruct the HR image



VRIJE UNIVERSITEIT BRUSSEL imec

Network design



VRIJE UNIVERSITEIT BRUSSEL

imec

Experimental results

Super-resolution of Multi-spectral images with the help of an RGB image.

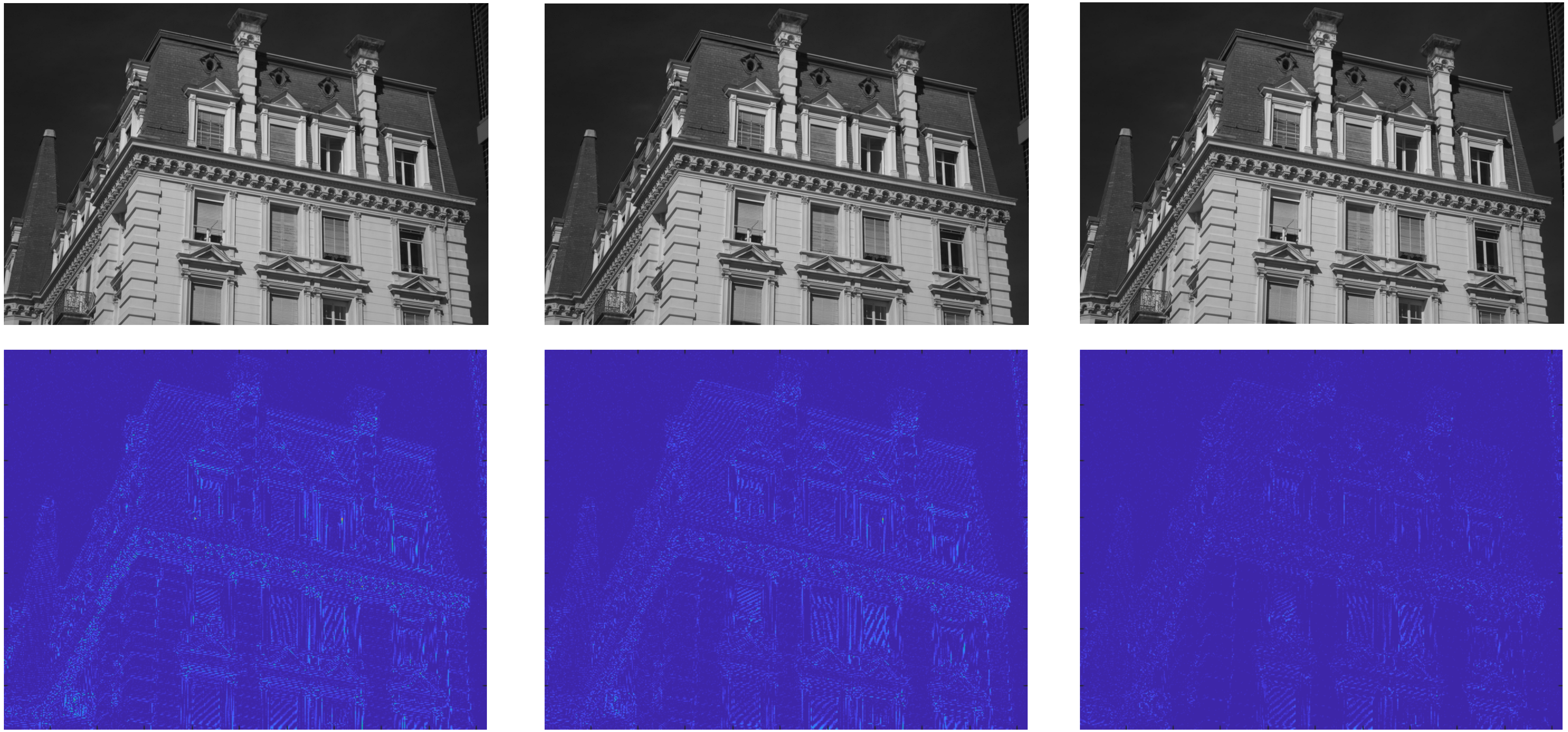
Image	SRFBN [20]				DJF [14]				CoISTA [15]				LMCSC [19]				Proposed			
	×4		×8		×4		×8		×4		×8		×4		×8		×4		×8	
Chart toy	33.43	0.9838	27.90	0.8736	37.86	0.9935	32.89	0.9733	36.58	0.9914	33.18	0.9768	40.31	0.9965	34.35	0.9805	**40.96**	**0.9978**	**34.89**	**0.9833**
Egyptian	40.04	0.9822	35.50	0.9755	45.69	0.9922	41.58	0.9850	45.91	0.9961	43.46	0.9906	48.79	0.9981	43.90	0.9966	**49.55**	**0.9991**	**44.60**	**0.9967**
Feathers	35.53	0.9873	30.14	0.9718	40.13	0.9939	31.50	0.9396	39.62	0.9937	32.04	0.9432	41.48	0.9962	36.81	0.9875	**42.05**	**0.9975**	**37.65**	**0.9904**
Glass tiles	29.53	0.9676	23.72	0.9002	34.97	0.9915	29.53	0.9685	33.99	0.9907	27.96	0.9390	34.65	0.9939	30.20	0.9724	**36.11**	**0.9959**	**30.85**	**0.9818**
Jelly beans	32.97	0.9845	25.80	0.9243	39.16	0.9885	30.14	0.9503	38.92	0.9956	30.69	0.9585	39.75	0.9966	34.70	0.9888	**41.10**	**0.9983**	**34.78**	**0.9898**
Oil Paintings	32.68	0.9182	31.07	0.9258	37.76	0.9805	35.12	0.9492	37.26	0.9690	35.99	0.9482	**39.14**	0.9910	**36.27**	**0.9759**	39.01	**0.9917**	35.73	0.9724
Paints	36.06	0.9907	28.03	0.9653	39.36	0.9944	31.86	0.9553	38.40	0.9949	33.05	0.9679	38.98	0.9966	35.06	0.9910	**40.61**	**0.9981**	**35.49**	**0.9937**
Average	34.32	0.9735	28.88	0.9338	39.28	0.9906	33.23	0.9602	38.67	0.9902	33.77	0.9615	40.44	0.9955	35.90	0.9847	**41.34**	**0.9969**	**36.28**	**0.9869**

Super-resolution of NIR images with the help of an RGB image.

Image	SRFBN [20]				DJF [14]				CoISTA [15]				LMCSC [19]				Proposed			
	×2		×4		×2		×4		×2		×4		×2		×4		×2		×4	
u-0004	35.36	0.9974	29.01	0.9787	34.50	0.9964	31.02	0.9784	35.83	0.9968	31.56	0.9835	37.36	0.9977	33.75	0.9869	**38.13**	**0.9986**	**34.28**	**0.9931**
u-0006	41.08	0.9970	33.73	0.9702	41.52	0.9975	36.04	0.9894	42.40	0.9976	37.21	0.9871	43.60	0.9982	38.74	0.9912	**43.98**	**0.9988**	**39.83**	**0.9943**
u-0017	38.19	0.9950	32.91	0.9725	38.65	0.9961	34.18	0.9815	39.13	0.9953	34.87	0.9777	40.87	0.9967	36.16	0.9828	**41.49**	**0.9978**	**36.92**	**0.9874**
o-0018	36.47	0.9971	28.88	0.9740	34.78	0.9960	30.72	0.9888	37.54	0.9975	32.35	0.9867	39.21	0.9982	34.17	0.9902	**41.82**	**0.9991**	**36.60**	**0.9949**
u-0020	37.50	0.9969	31.44	0.9807	37.35	0.9973	33.60	0.9915	39.00	0.9974	34.75	0.9887	40.98	0.9980	36.95	0.9900	**42.14**	**0.9987**	**37.56**	**0.9943**
u-0026	31.00	0.9782	29.10	0.9702	33.15	0.9939	29.21	0.9397	34.11	0.9944	29.94	0.9708	35.60	0.9963	31.03	0.9784	**36.75**	**0.9978**	**31.91**	**0.9843**
o-0030	35.57	0.9944	29.45	0.9583	35.67	0.9944	31.27	0.9345	36.90	0.9946	32.28	0.9709	38.29	0.9961	33.56	0.9780	**39.30**	**0.9974**	**33.95**	**0.9824**
u-0050	37.06	0.9966	29.89	0.9762	32.60	0.9928	28.58	0.9616	33.53	0.8837	29.42	0.9705	34.11	0.9948	30.04	0.9772	**34.85**	**0.9967**	**30.47**	**0.9796**
Average	36.53	0.9941	30.55	0.9726	36.03	0.9955	31.83	0.9707	37.30	0.9959	32.80	0.9795	38.74	0.9970	34.28	0.9843	**39.81**	**0.9981**	**35.19**	**0.9888**

Experimental results

Super-resolve a NIR image with the help of an RGB image and the error maps.



ColSTA [15] LMCSC [19] proposed

Conclusion

Key properties of the proposed network:

- Convolutional sparse maps as intermediate features
- Does not rely on bicubic interpolation for LR image initialization
- Exploits the advances in deep learning
- Leverages the benefits of deep unfolding designs
- Employs coupled sparse priors for signal fusion
- Reconstructs the entire image at once rather than extracted patches

Thank you for your time!