上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Institute of Media,
Information, and Network

A Hybrid Wavelet Convolution Network with Sparse-Coding for Image Super-Resolution

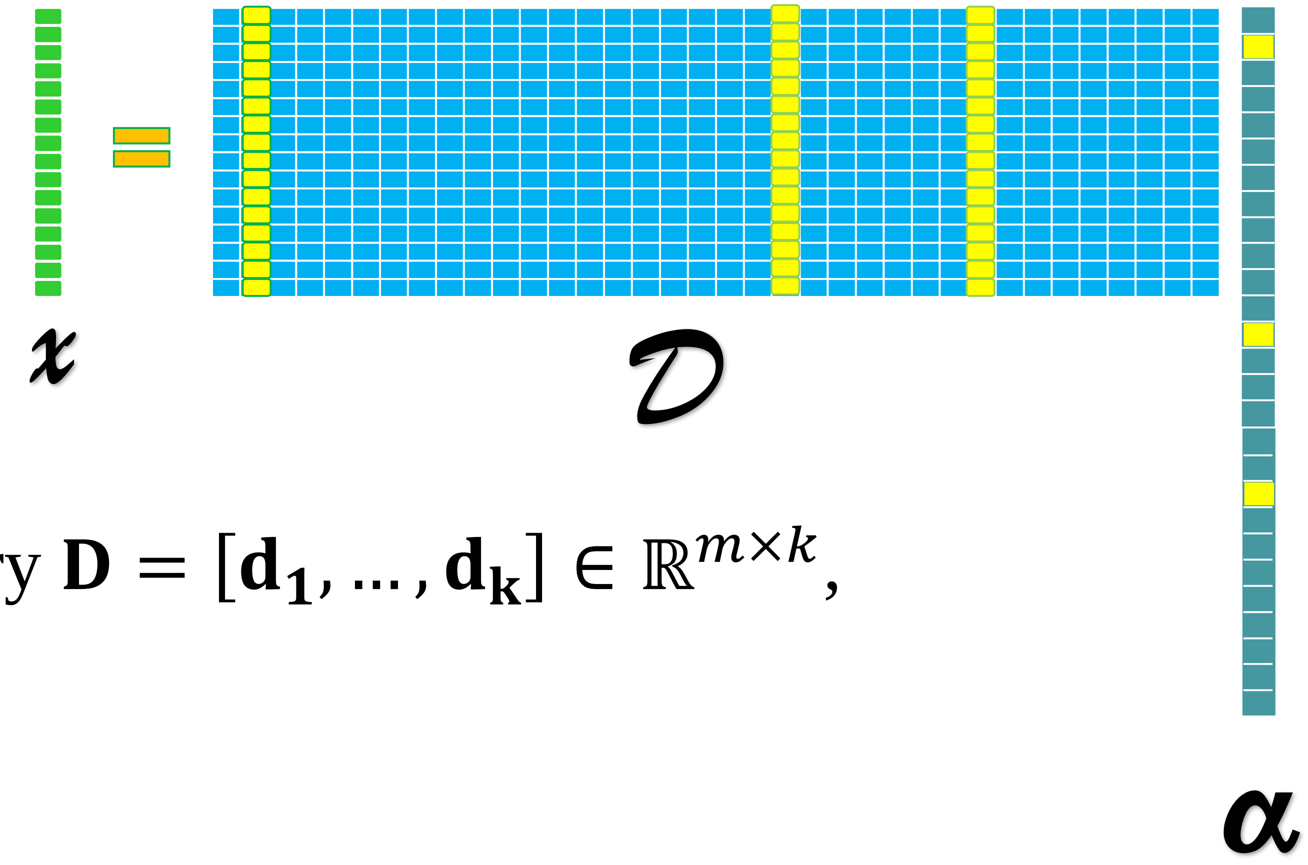
Xing Gao, Hongkai Xiong

Department of Electronic Engineering
Shanghai Jiao Tong University
Sep. 28th, 2016

Background——Sparse Representation

◆ *Sparse representation:*

Describe a signal as a linear combination of a few atoms from a pre-specified dictionary.



$$\mathbf{x} \approx \mathbf{D}\alpha,$$
$$\text{s.t. } \|\alpha\|_0 \ll m$$

where signal $\mathbf{x} \in \mathbb{R}^m$, dictionary $\mathbf{D} = [\mathbf{d_1}, \ldots, \mathbf{d_k}] \in \mathbb{R}^{m \times k}$,
$\alpha \in \mathbb{R}^k$ is sparse.

Background——Sparse Representation

◆ *The quest for a dictionary:*

A good choice of the dictionary is crucial for the success of the sparse representation model.

dictionary

Analytic dictionaries: DCT, Wavelet, etc.

Mathematical simplicity & Multi-scale & fast implementation;

Limited expressiveness, simplistic for complex natural phenomena.

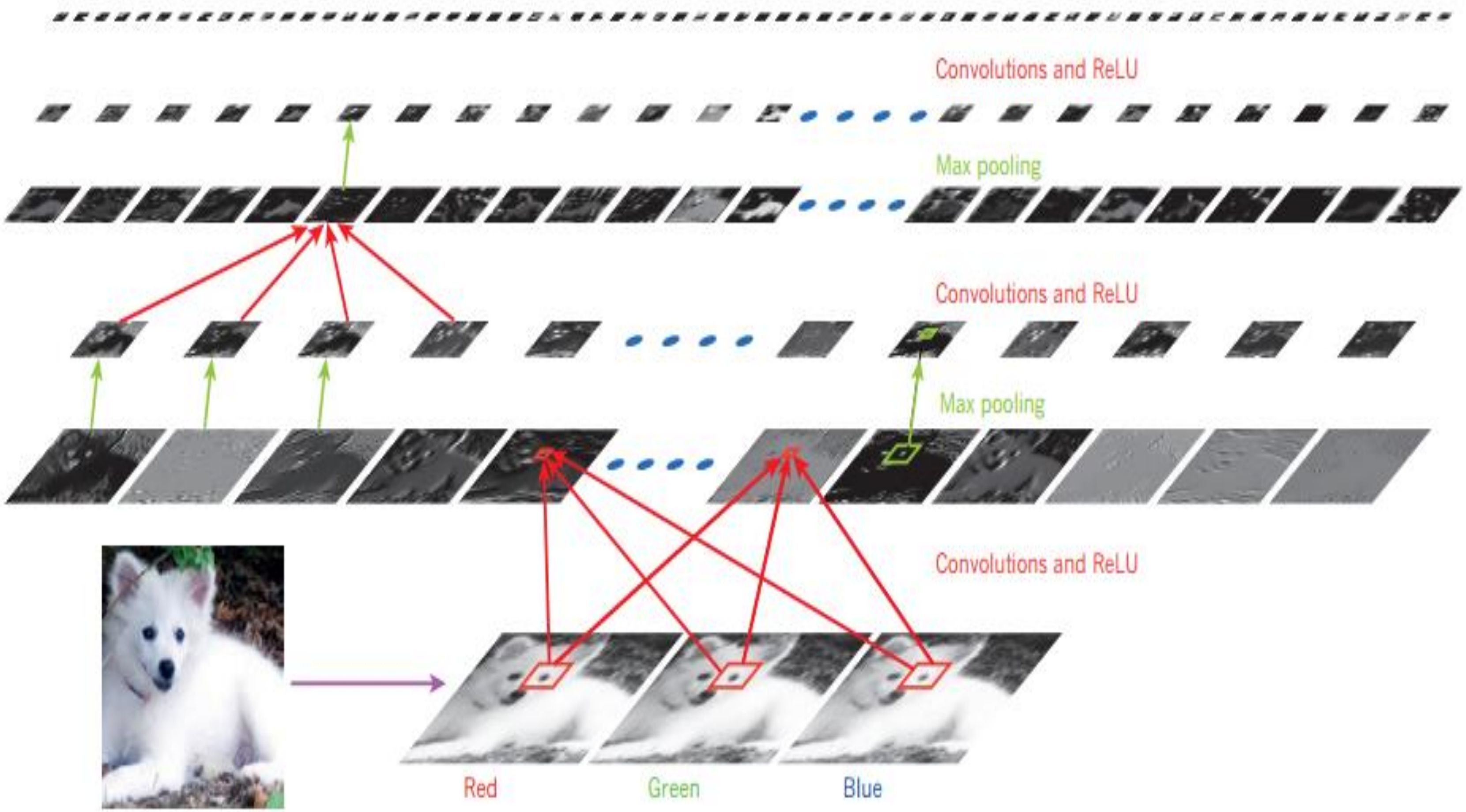
Learned dictionaries: MOD, K-SVD, etc.

Adaptive to specific data, better performance;

Complex, costly to deploy unknown inner structure.

How to incorporate two types of dictionaries efficiently?

Background——Deep Learning



(LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015)

- Stacked convolution, pooling, and fully-connected layers

Superiority

- Hierarchical structure → features with various abstract levels
- **State-of-the-art** in most of computer vision fields: classification, recognition, etc.

Inferiority

- Over 100M parameters → **train hard**
- Highly nonlinearity: the property and optimal structure **not well understood**

Challenges

◆ ***Sparse representation:***

Dictionary: Fast implementation & Adaptivity

◆ ***Deep learning:***

Optimize structure (reduce parameters) & Improve performance

⬇

A Hybrid Wavelet Convolution Network (HWCN)

Contributions

1. Sparse representation:

Provide an ***end-to-end*** and ***adaptive*** framework to implement sparse coding and reconstruction.

2. Deep learning:

Design ***a novel structure*** of deep networks, composed of predefined and learned convolution kernels, and it is ***easier*** to train. **fewer parameters**

3. Image super-resolution:

Provide another approach to achieve ***state-of-the-art*** performance and also enhance current performance.

The Architecture of the HWCN

- *A hybrid wavelet convolution network =*

Scattering convolution network + Convolutional neural network

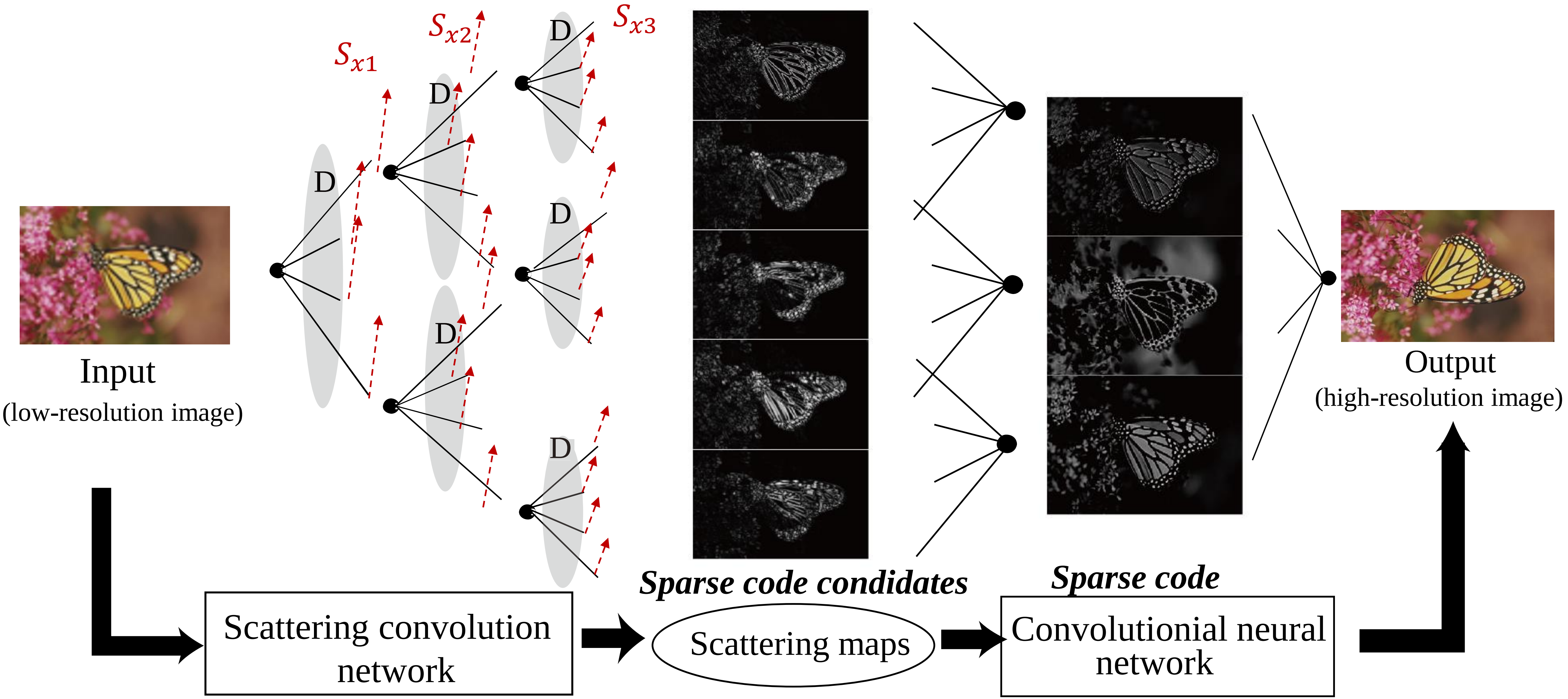
Directional complex wavelets

Select and weight

Reconstruct

Sparse coding candidates → Sparse code → HR image

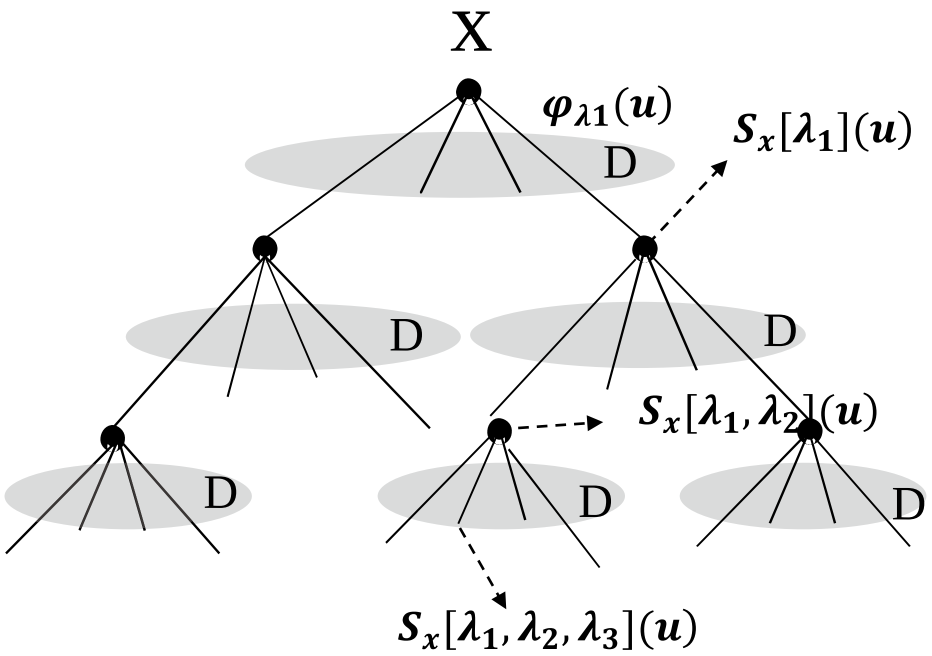
The Architecture of the HWCN



The Scattering Convolution Component

Scattering convolution network:

A revised version of ***invariant scattering convolution networks*** introduced by S.Mallat *et al.* in 2012.



$$S_x = \{S_x[p](u)\}_p$$

- Convolutional filter banks are defined by two-dimensional complex directional wavelets :

with $$D = \{\psi_\lambda(u) : 2^{-2j}\psi(2^{-j}r^{-1}u)\}_{\lambda\in\Lambda},$$

$$\Lambda = \{\lambda = 2^{-j}r | j = 0, 1, \cdots, J-1; \quad r = 0, 2\pi/L, \cdots, 2(L-1)\pi/L\}$$

- Each wavelet in D defines a filter of *each node*, and each node preforms ***convolution*** and ***modulus*** operations.
- From the root node to each node forms a path:

$$p = [\lambda_1, \lambda_2, \cdots, \lambda_m]$$

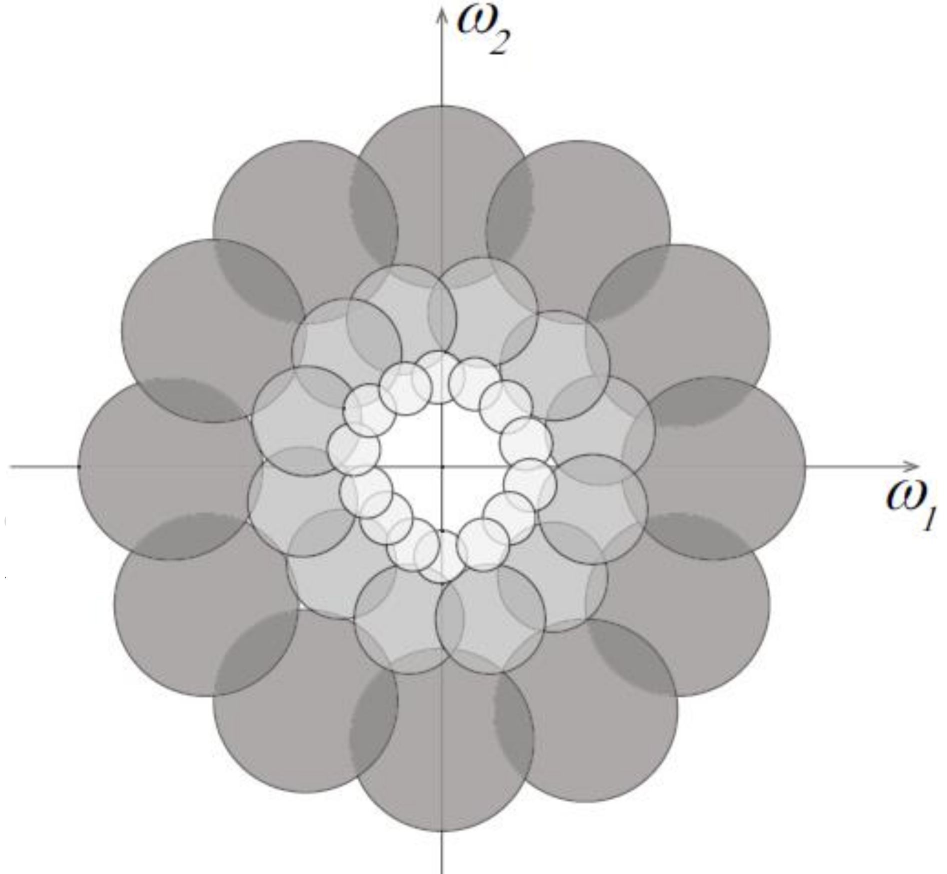
which produces a scattering map

$$S_x[p](u) = |||x * \psi_{\lambda 1}| * \psi_{\lambda 2}| \cdots * \psi_{\lambda m}|.$$

The Scattering Convolution Component

Scattering convolution network:

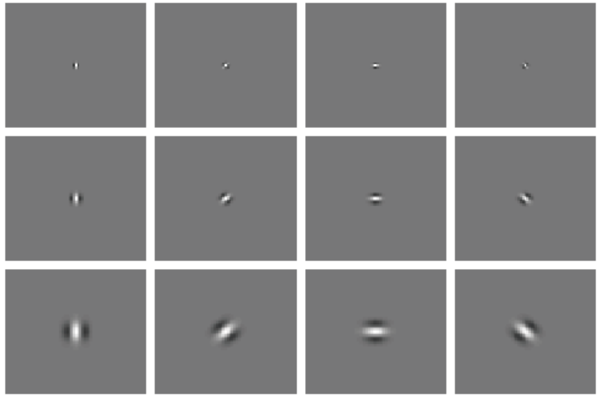
- ***Frequency domain analysis***



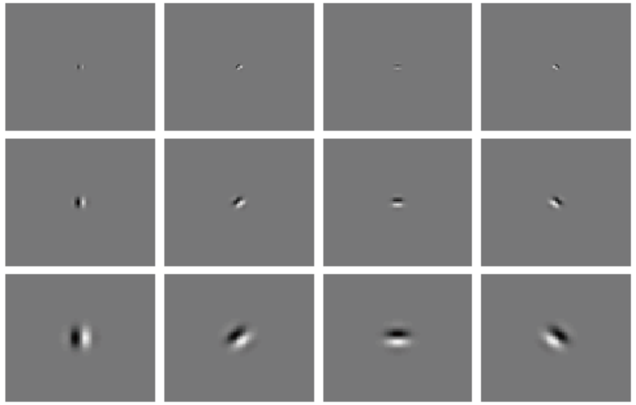
- Convolutional filter banks:

$$D = \{\psi_\lambda(u) : 2^{-2j}\psi(2^{-j}r^{-1}u)\}_{\lambda\in\Lambda},$$

Fourier transform



Real parts of filter banks

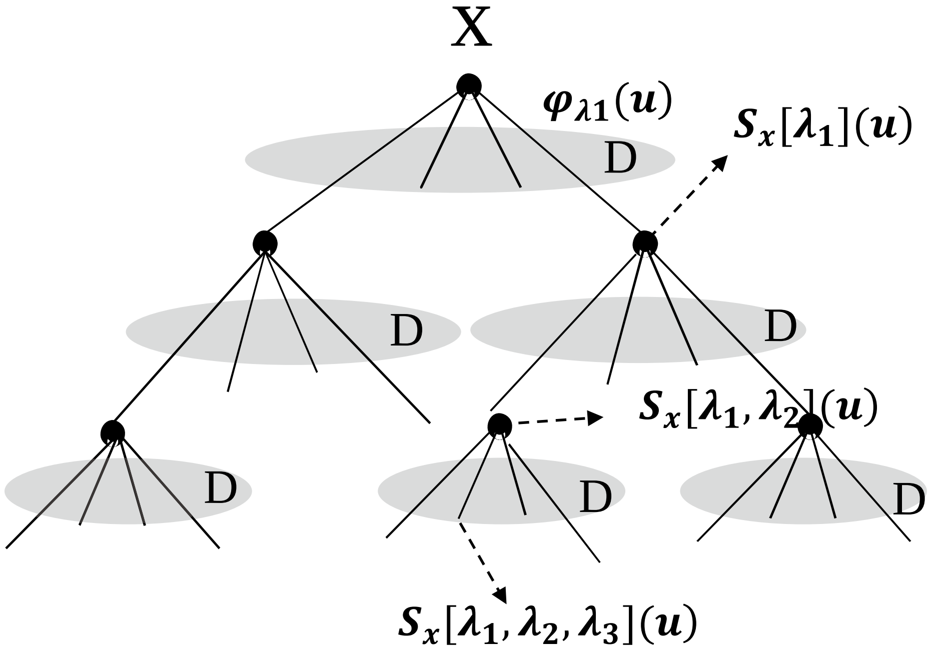


Imaginary parts of filter banks

The Scattering Convolution Component

◆ *Scattering convolution network:*

● ***Optimize structure***

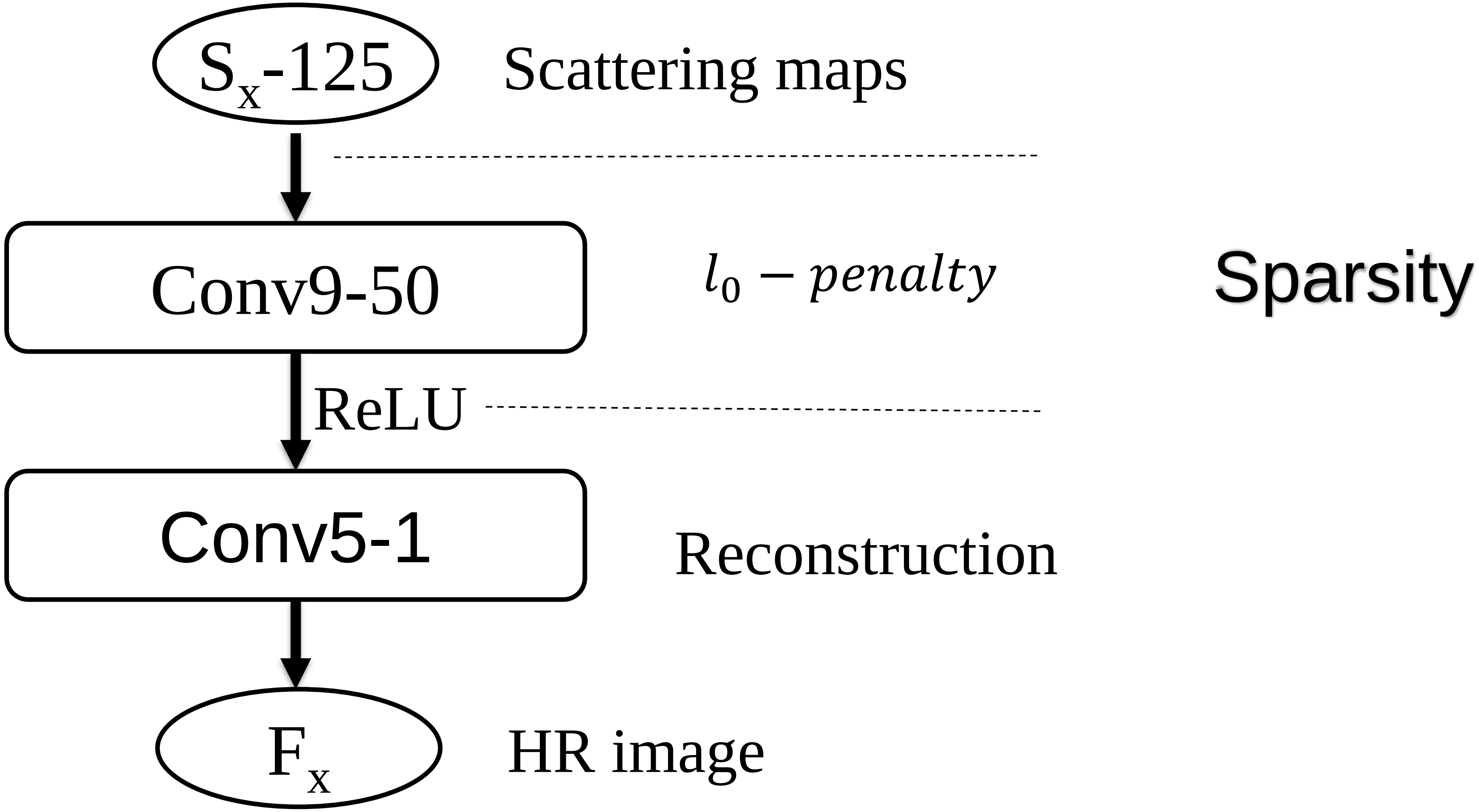


- Frequency decreasing path:

$$|\lambda_{i+1}| < |\lambda_i|$$

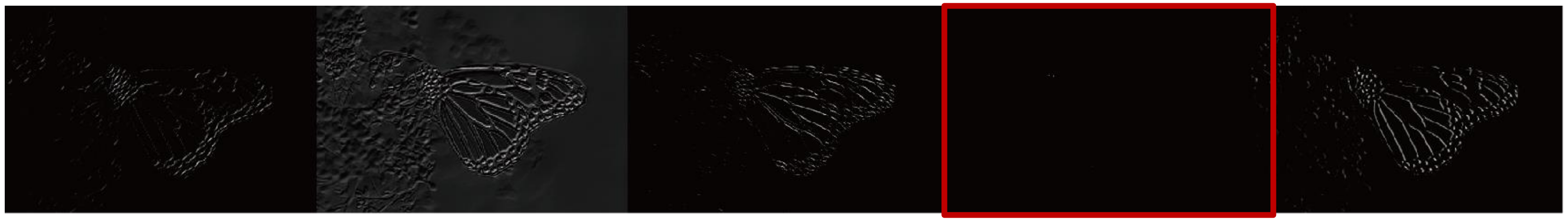
- Modulus operation removes oscillations $\longrightarrow$ shift signal to lower frequency $\longrightarrow$ major energy of signal along frequency decreasing paths $\longrightarrow$ **pruning**

The Convolution Neural Component

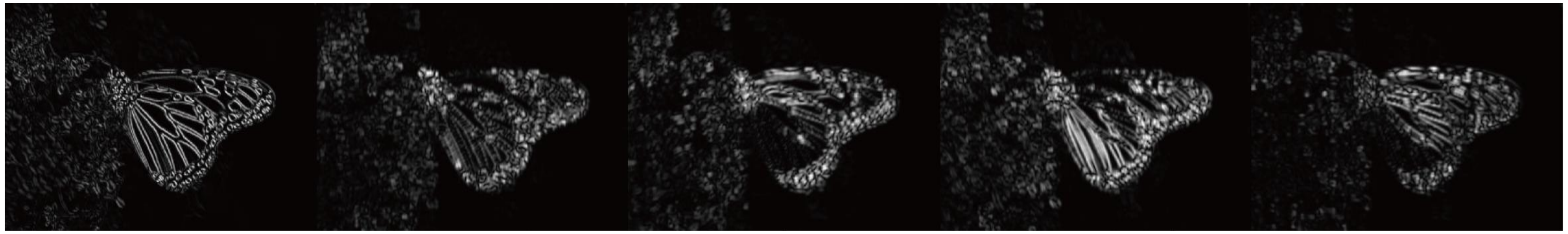


Analysis

◆ Relationship to CNNs



Feature maps extracted by the first layer of CNN [1].



Feature maps extracted by the scattering part.

☐ CNNs:

- Learned
- Some ***'dead'*** features

☐ HWCN:

- Predefined and learned
- Multiscale and multidirectional features
- Features can be selected based on scale, direction, and frequency property

[1] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[M]//Computer Vision–ECCV 2014. Springer International Publishing, 2014: 184-199

Analysis

◆ Relationship to typical dictionary based methods:

$$x = D_l \alpha, y = D_h \alpha$$

- Dictionary based sparse coding methods:
  - Training dictionary:

$$\min_{D_l, D_h} \frac{1}{2n} \sum_{i=1}^{n} \left\| D_h \alpha_i - y_i \right\|^2$$

$$s.t. \qquad \alpha_i = \arg\min_z \left\| x_i - D_l \alpha_i \right\|^2 + \lambda \left\| z \right\|_1$$

  - Sparse coding:

$$\alpha_i = \arg\min_z \left\| x_i - D_l \alpha_i \right\|^2 + \lambda \left\| z \right\|_1$$

- HWCN:
  - Scattering convolution network:
    - Predefined wavelet filter banks
    - Produce sparse code candidates

    Analytic dictionary $\longrightarrow$ ***Fast implementation***
  - Convolution neural network:
    - Learned from data
    - Select and weight code candidates
    - Reconstruct from sparse code

    Learned dictionary $\longrightarrow$ ***Adaptivity***
  - Forward and end-to-end

D_l

D_h

Application to Image Super-Resolution

- ***Training set:***

91 natural images, the same as prior methods based on sparse representation [1~3].

- ***Test sets:***

Set5 & Set14, the same as [3].

- ***Loss function-MSE:***

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\|F_{xi}(\theta) - y_i\right\|^2 ,$$

where F_{xi} , y_i represent reconstructed and ground truth HR image, respectively.

[1]Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. Image Processing, IEEE Transactions on, 2010, 19(11): 2861-2873.
[2] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[M]//Curves and Surfaces. Springer Berlin Heidelberg, 2012: 711-730.
[3] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[M]//Computer Vision–ECCV 2014. Springer International Publishing, 2014: 184-199.

Application to Image Super-Resolution

Table 1. The result of PSNR (dB) on the Set5 dataset.

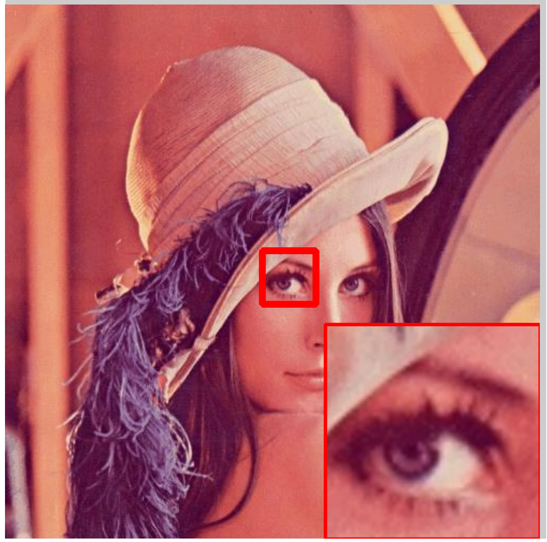
Set5	Scale	Bicubic	SC	K-SVD	NE+LLE	ANR	CNN	Proposed
baby	3	33.870	34.258	35.038	35.018	**35.092**	34.968	34.903
bird	3	32.648	34.266	34.689	34.678	34.724	35.052	**35.504**
butterfly	3	24.064	25.675	26.000	25.809	25.964	**27.677**	27.648
head	3	32.824	33.150	33.508	33.556	33.590	33.499	**33.703**
woman	3	28.654	30.073	30.451	30.335	30.444	31.007	**31.260**
average	3	30.412	31.484	31.937	31.879	31.963	32.441	**32.604**

Application to Image Super-Resolution

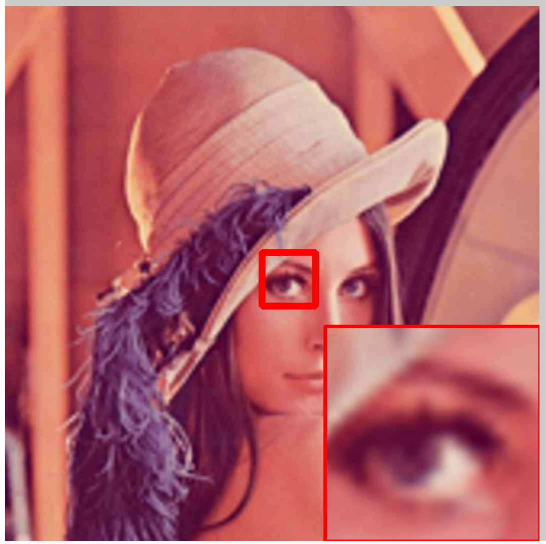
Table 2. Average performance (PSNR dB) and time (s) on the Set14 dataset.

Set14	Scale	Bicubic	SC	K-SVD	NE+LLE	ANR	CNN	Proposed
Avg PSNR	3	27.552	28.381	28.674	28.612	28.669	29.041	**29.171**
Avg time	3	-	80.8284	3.8231	4.9020	**0.7295**	2.1742	3.6492

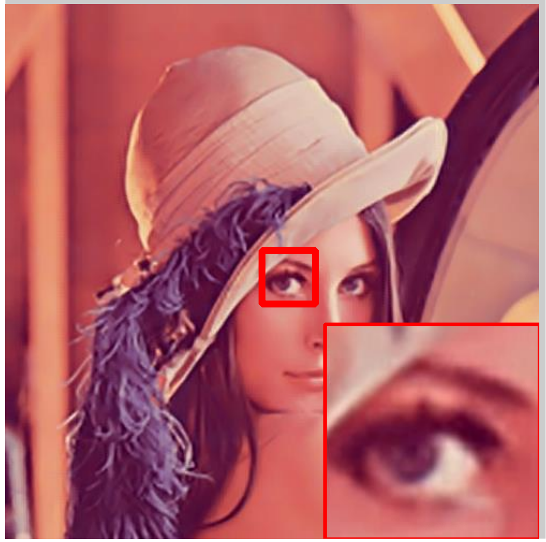
Application to Image Super-Resolution



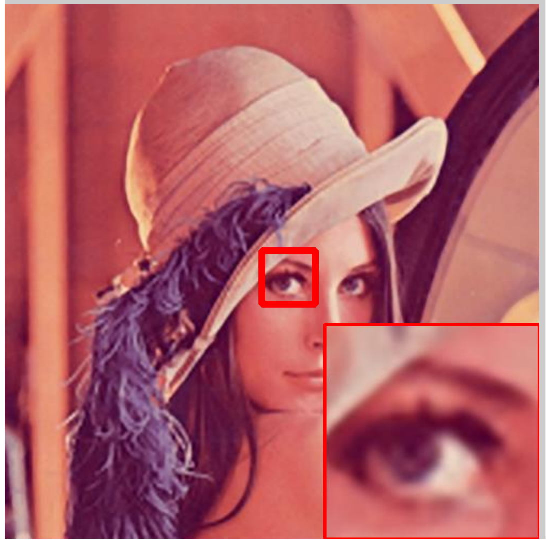
Original/PSNR



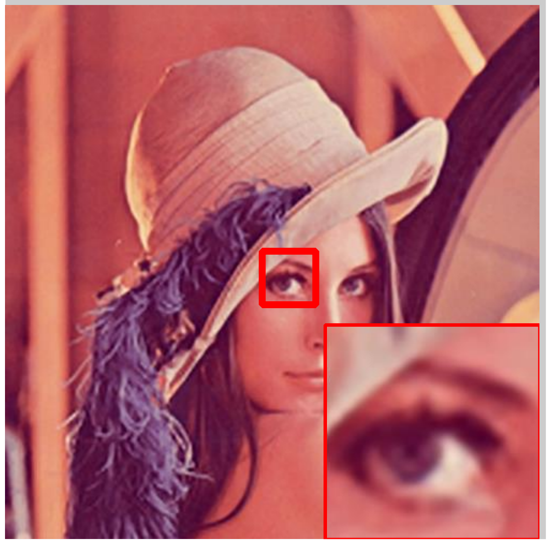
Bicubic/31.618dB



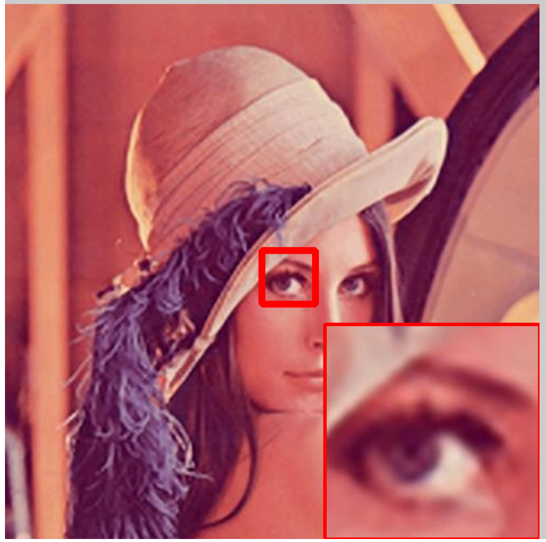
SC/32.601dB



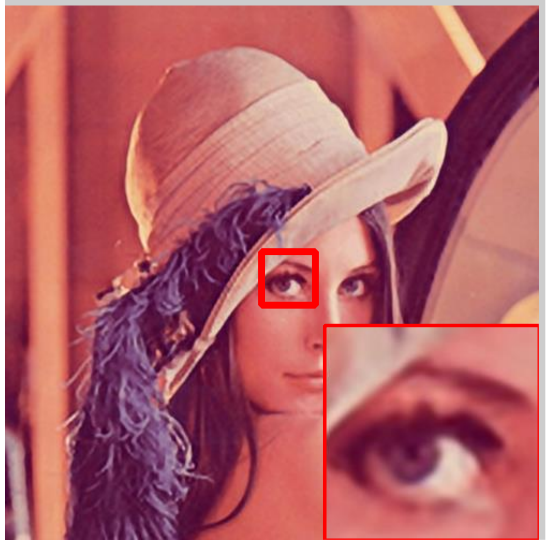
K-SVD/32.945dB



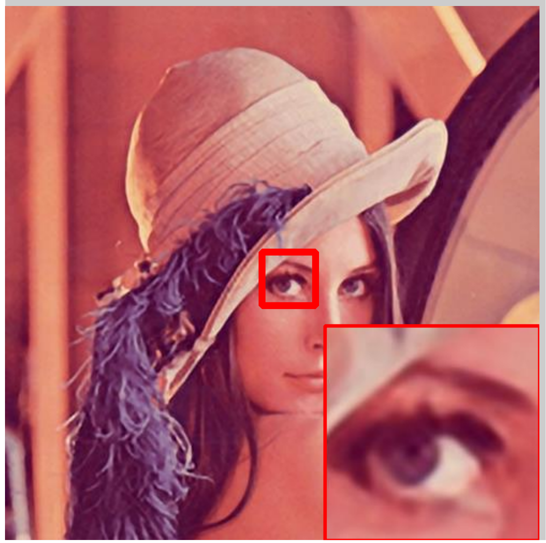
NE+LLE/32.958dB



ANR/33.027dB



CNN/33.344dB



Proposed/**33.472dB**

Conclusion

- The hybrid wavelet convolution network

= scattering convolution network + convolutional neural network

= analytic dictionary + learned dictionary

Sparse & Multiscale *Adaptive*

= predefined convolution kernel + trainable convolutional kernel

Prior knowledge & Generalization *Flexible & Powerful*

- Follow-up
  - Further optimize the structure of HWCN: contourlet?
  - Application in other fields: image classification, recognition.

Q & A

Thanks

Application to Image Super-Resolution

Table 3. The result of PSNR (dB) on the Set14 dataset.

Set14	Scale	Bicubic	SC	K-SVD	NE+LLE	ANR	CNN	Proposed
baboon	3	23.210	23.482	23.523	23.558	23.569	23.605	**23.617**
barbara	3	26.191	26.348	**26.699**	26.680	26.635	26.597	26.513
bridge	3	24.427	24.855	25.045	25.005	25.036	25.104	**25.195**
coastguard	3	26.715	27.054	27.168	27.146	27.174	27.196	**27.208**
comic	3	23.045	23.844	23.882	23.901	23.966	24.307	**24.405**
face	3	32.759	33.086	33.475	33.517	33.566	33.525	**33.710**
flowers	3	27.151	28.172	28.350	28.304	28.413	28.894	**29.020**
foreman	3	31.667	33.336	33.743	33.805	33.850	34.339	**34.759**
lenna	3	31.618	32.601	32.945	32.958	33.027	33.344	**33.472**
man	3	26.978	27.738	27.873	27.839	27.895	28.150	**28.302**
monarch	3	29.348	30.646	31.029	30.872	31.016	32.316	**32.349**
pepper	3	32.435	33.390	34.130	33.879	33.905	34.447	**34.720**
ppt3	3	23.619	24.890	25.136	24.853	24.939	25.933	**26.133**
zebra	3	26.564	27.898	28.435	28.254	28.371	28.818	**28.984**
average	3	27.552	28.381	28.674	28.612	28.669	29.041	**29.171**