Multiple-image Super Resolution Using Both Reconstruction Optimization and Deep Neural network

Jie Wu, Tao Yue, Qiu Shen, Xun Cao, Zhan Ma
School of Electronic Science and Engineering, Nanjing University

Jie Wu
Nov. 15, 2017

Which one do you prefer?

Image Resolution

High – Resolution (HR):
Pixel density within an image is large, therefore offering more details.

Low – Resolution (LR):
Pixel density within an image is small, therefore offering less details.

How can we achieve high Resolution?

Increase the number of sensor elements per unit area

Increase the sensor density by **reducing** the sensor size

Increase the hardware cost

Do not work!

Is there another way to achieve high resolution ?

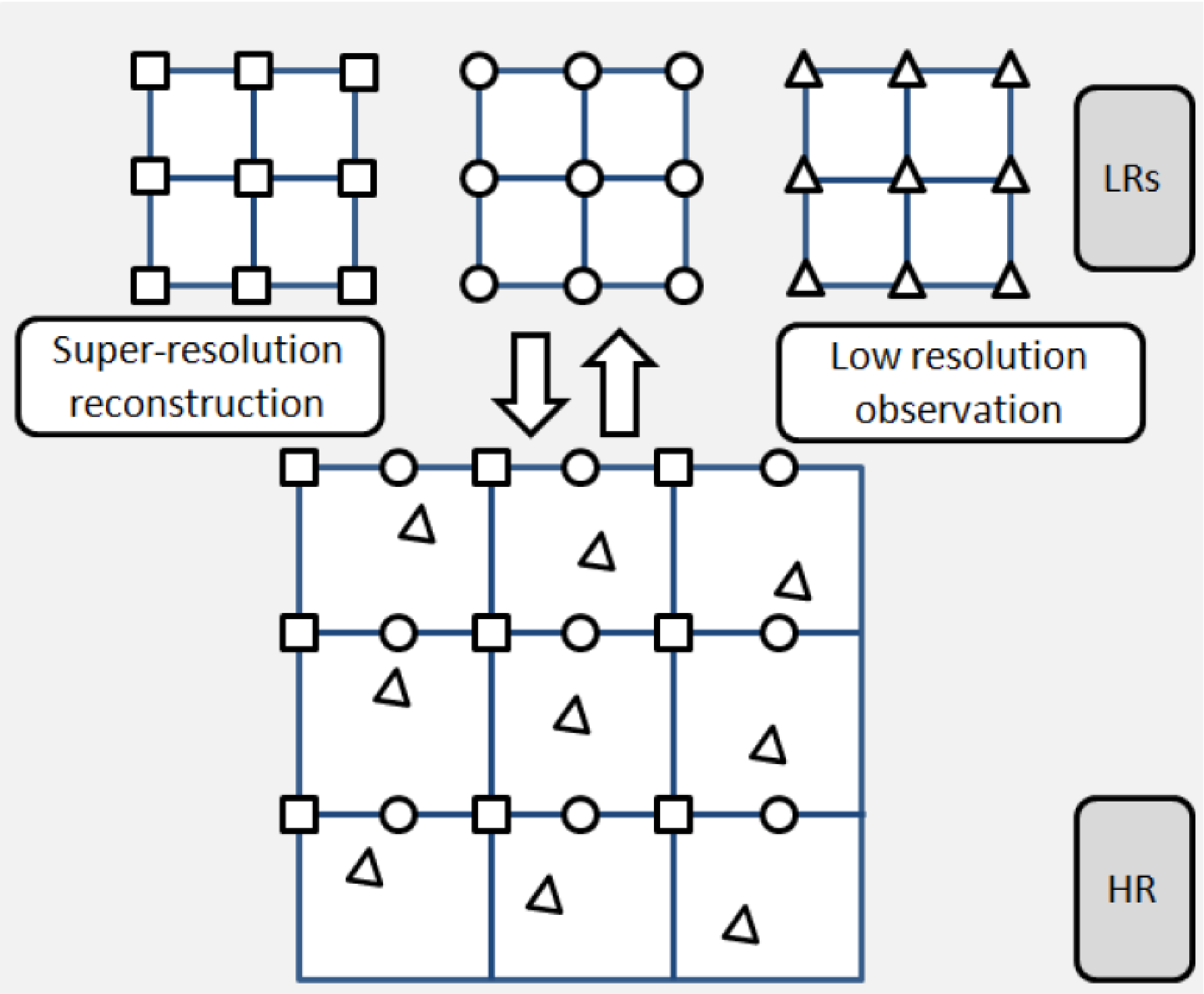
Another way to address this problem is to **accept the image degradations** and use **signal processing technique** to process the captured images, to trade off **computational cost** with the **hardware cost**.

These techniques are specifically referred to as the **Super Resolution (SR) reconstruction.**

What is Super Resolution?

" Generate a high-resolution (HR) image from a single or a set of low-resolution (LR) images. "

Sketch map



$$Y_i \in \mathrm{R}^{M \times N} (i = 1 \sim 16)$$

$$X \in \mathrm{R}^{Ms \times Ns}$$

★ schematic diagram ★

Development history

1990 – 2000

[1] **H. Stark, P. Oskoui**, POCS

[2] **M. Irani, S. Peleg**, IBP

[3] **B. C. Tom, A. K. Katsaggelos,** ML

[4] **R. R. Schulz, R. L. Stevenson**, MAP

[5] **M. Elad, A. Feuer** , Adaptive Filtering

Xiaoou Tang. Deep learning for SR

[1] Learning a Deep Convolutional Network for Image Super-Resolution, ECCV 2014

[2] Image Super-Resolution Using Deep Convolutional Networks, PAMI 2015

[3] Compression Artifacts Reduction by a Deep Convolutional Network, ICCV 2015

Li Xu. Change filter size.

[1] "Handling Motion Blur in Multi-Frame Super-Resolution" IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[2] "Deep Convolutional Neural Network for Image Deconvolution" Advances in Neural Information Processing Systems (NIPS), 2014.

1964 – 1990

J.L.H arris and **J.W.G oodman** proposed a single frame image spectrum extrapolation is the earliest super-resolution image processing method. Subsequently, Tsai and Huang presented multiple frames super-resolution reconstruction method, and provided the reconstruction method based on frequency domain.

2008 – 2010

Jianchao Yang. Sparse coding.

[1] Image Super-Resolution via Sparse Representation, TIP 2010

[2] Image Super-Resolution as Sparse Representation of Raw Image Patches, CVPR 2008

2014 – now

Jianchao Yang. sparse-coding.

[1] Deep Networks for Image Super-Resolution With Sparse Prior. (ICCV), 2015

Li Fei-Fei. Real Time.

[1] Perceptual Losses for Real-Time Style Transfer and Super-Resolution. ECCV 2016

[2] Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. (CVPR), 2017

Classification

Reconstruction-based

- Adding prior information.
- Register and reconstruct.

- Larger magnification factor or insufficient input images will bring a dramatically degradation of reconstruction performance.

Example-based

- Sparse Coding
- Deep Neural Network.

- Need enormous databases of millions of high-resolution and low-resolution patch pairs, and are therefore computationally intensive.

Multiple-image Super Resolution

Reconstruction-based Multi-images SR method

16 low-res image array (input)

$$J = \| D \cdot X - Y \|^2$$

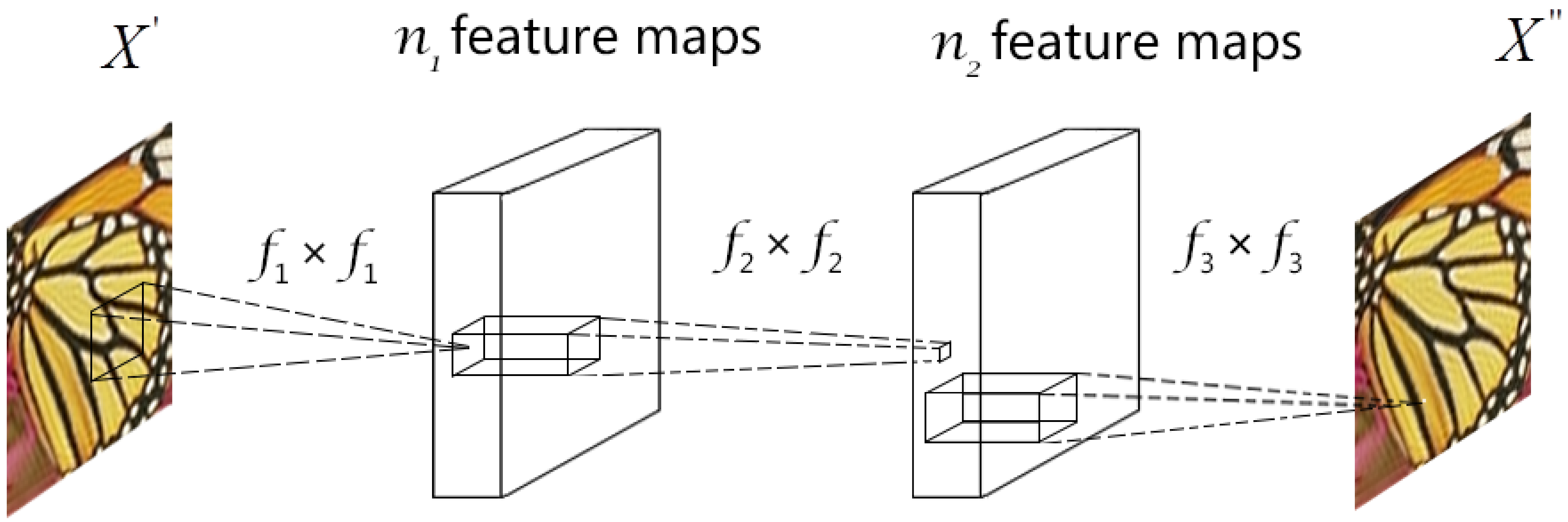
Objection function

$$J = \frac{1}{2m} \sum_{i=1}^{m} \| D_i \cdot X - y_i \|^2 + \theta \| \nabla X - Q \|_2^{\,2} + \beta \| Q \|_1$$

Adding l1-norm regularization term

$\longrightarrow X'$

neural network input



high-resolution image (output)

Convolutional neural network for artifacts removing

Notation

- High-resolution image. (ground truth image) $X \in \mathrm{R}^{Ms \times Ns}$
- Low-resolution images. (input images) $Y_i \in \mathrm{R}^{M \times N} (i = 1 \sim 16)$
- Down - sampling matrix $D_i \in \mathrm{R}^{MN \times MNs^2} (i = 1 \sim 16)$
- Reconstructed high-resolution image. (output image of the reconstruction-based method) $X' \in \mathrm{R}^{Ms \times Ns}$
- Reconstructed high-resolution image. (output image of the CNN) $X'' \in \mathrm{R}^{Ms \times Ns}$

Cost function

$$Y = D \cdot X$$

$$J = \| D \cdot X - Y \|^2$$

$$J = \frac{1}{2m} \sum_{i=1}^{m} \| D_i \cdot X - y_i \|^2 + \theta \| \nabla X - Q \|_2^{\,2} + \beta \| Q \|_1$$

Reconstruction-based method for super resolution

16 low-res images array (Input)

$$\longrightarrow$$

$$J = \| D \cdot X - Y \|^2$$ Objection function

$$J = \frac{1}{2m} \sum_{i=1}^{m} \| D_i \cdot X - y_i \|^2 + \theta \| \nabla X - Q \|_2^{\,2} + \beta \| Q \|_1$$

Adding L1-norm regularizer

$$\longrightarrow X'$$

Output image

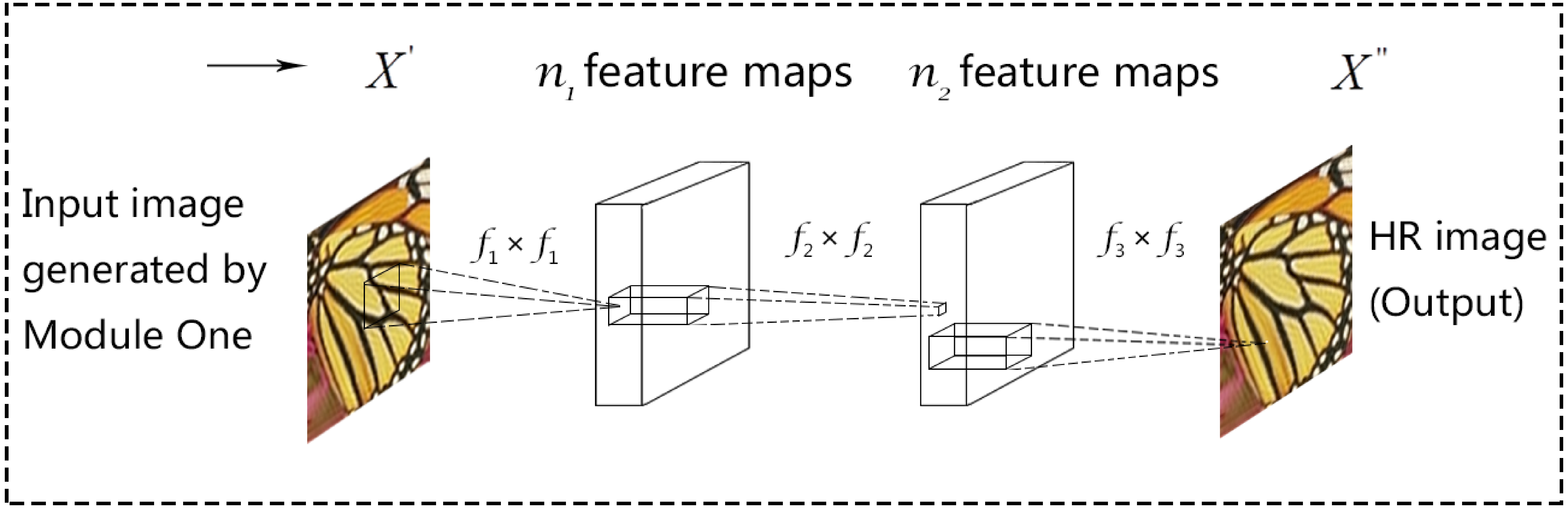
★ Groundtruth image

★ Bicubic image

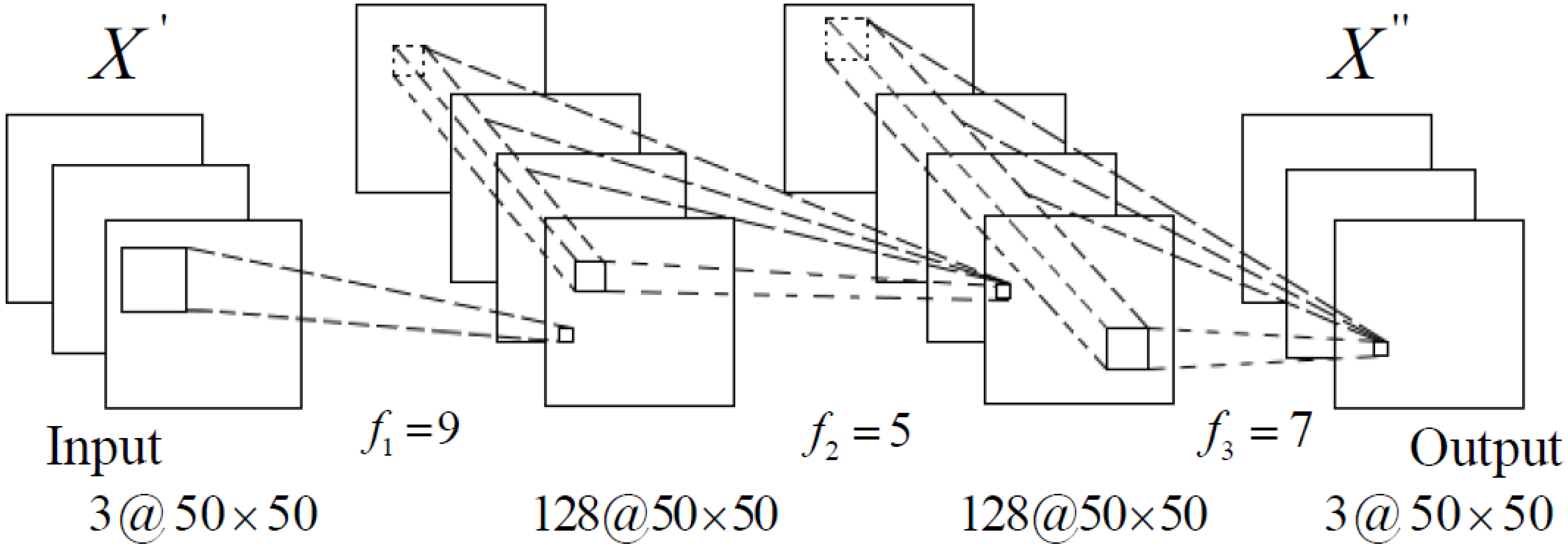
★ Reconstruction-based image

Remove artifacts

A three-layer convolutional network for artifacts removal



Artifacts removal

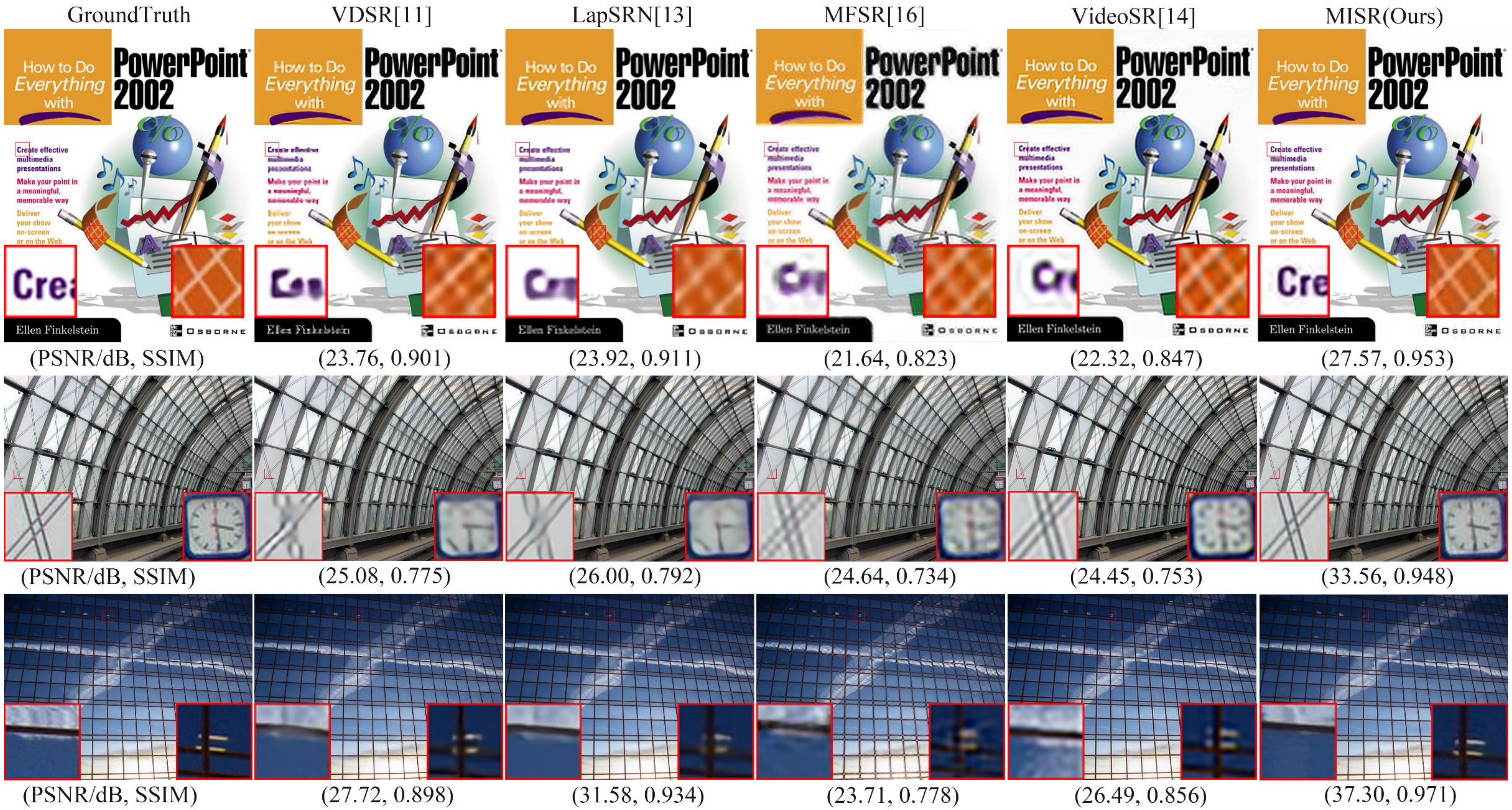


★ A three-layer Artifacts removal neural network ★

Output image

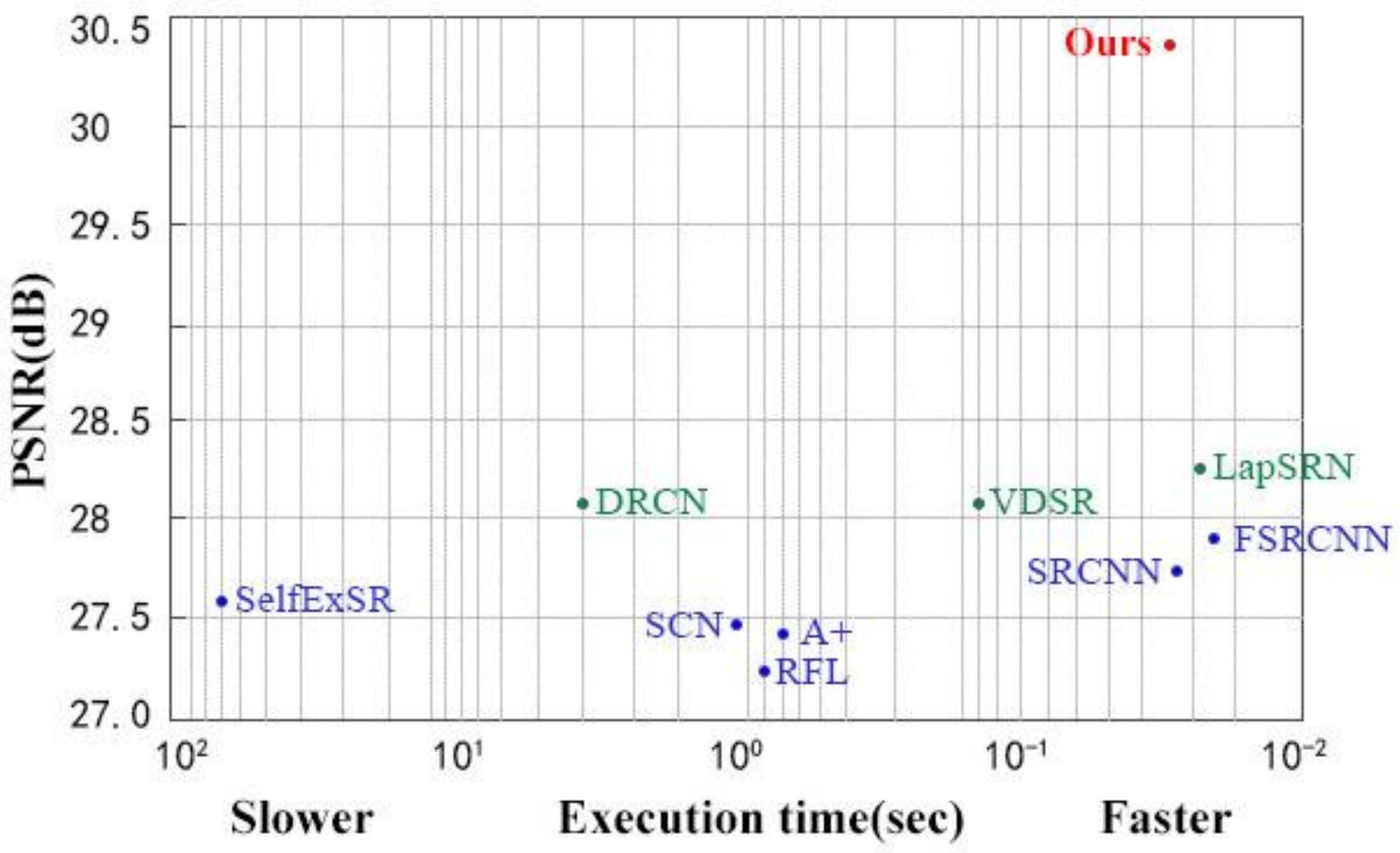
★Groundtruth image ★Reconstruction-based image ★Re + SR image

Performance: Subjective



★ Comparison between other Super resolution methods and Ours ★

Performance: Objective



Dataset	Bicubic PSNR/SSIM	VDSR[11] PSNR/SSIM	LapSRN[13] PSNR/SSIM	MFSR[16] PSNR/SSIM	VideoSR[14] PSNR/SSIM	MISR(Ours) PSNR/SSIM
Set5	28.423/0.810	30.289/0.871	31.522/0.885	27.627/0.811	28.450/0.845	**38.200/0.963**
Set14	26.101/0.704	27.166/0.763	27.168/0.744	25.617/0.740	26.121/0.774	**32.808/0.910**
Urban100	23.152/0.659	24.178/0.736	25.201/0.755	23.171/0.704	21.605/0.659	**32.073/0.939**

Table 1. Average PSNR/SSIM for 4× scale factor on datasets Set5 [1], Set14 [22] and Urban100 [9]. The bold indicates the best performance. The proposed method outperforms the state-of-the-arts by a large margin.

Application

Remote sensing | Medical imaging | Computer Vision | Surveillance Video

- Medical imaging (e.g. CT, MRI (magnetic resonance imaging), etc..)
- Satellite imaging
- Video surveillance
- Converting NTSC (National Television Standards Committee) video content to high-definition television

THANKS

Jie Wu

School of Electronic Science and Engineering, Nanjing University

Nov. 15, 2017