KAIST

Deep Learning-Based Stereo Super-Resolution

전석준, 백승환, 최인창, 김민혁

2018.08.09

Enhancing the Spatial Resolution of Stereo Images using a Parallax Prior
Daniel S. Jeon, Seung-Hwan Baek, Inchang Choi, Min H. Kim
*Proc. IEEE Computer Vision and Pattern Recognition (**CVPR** 2018)*
Salt Lake City, USA, June 18, 2018

VISUAL COMPUTING Lab

Goal

Low-resolution

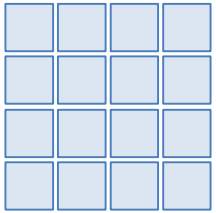
Super-resolution

High-resolution

Problem of Common Imaging System

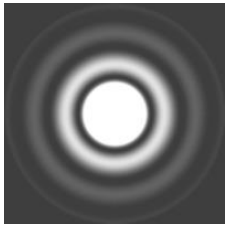
loss of spatial resolution

- Pixel discretization
- Optical distortions
  - Out of focus
  - Diffraction limit
  - Optical aberration
- Motion blur
- Noise
- Aliasing

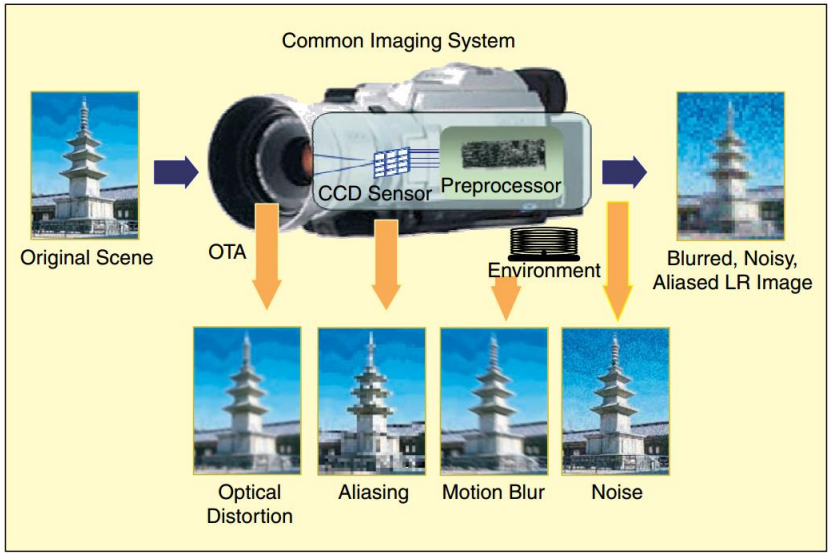


Sensor

Out of focus



Diffraction



BACKGROUND

Multi-Frame Super-Resolution

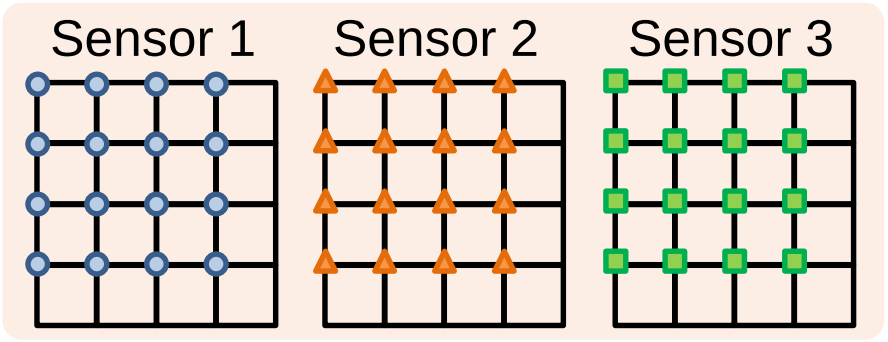
Low-resolution images

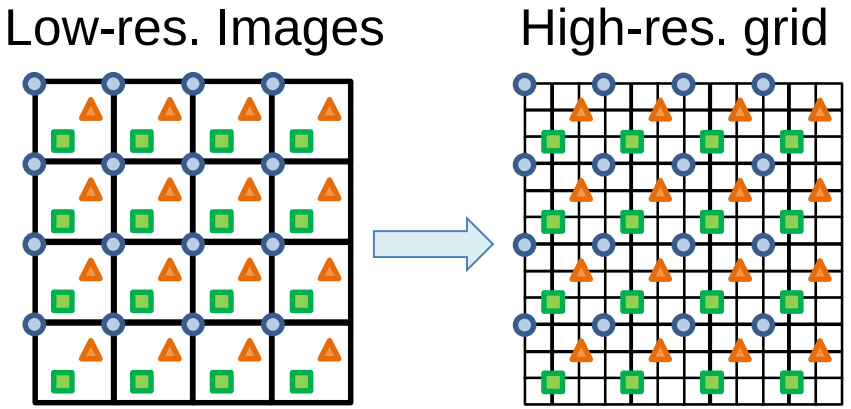
Super-resolution

High-resolution image

Multi-Frame Super-Resolution

- Approach
  - Sub-pixel Image Registration
  - Projection to high-res. grid

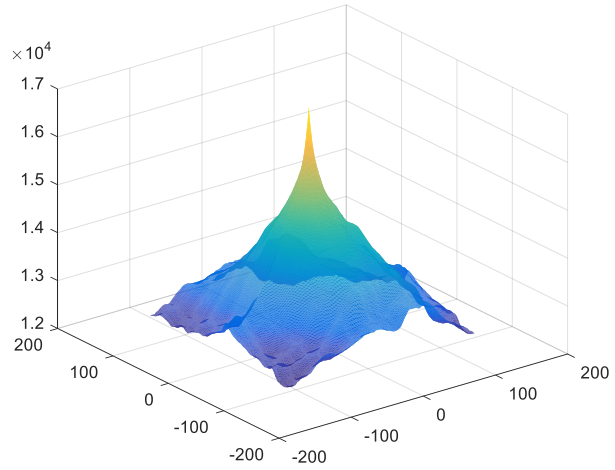




Sub-Pixel Image Registration

Image 1

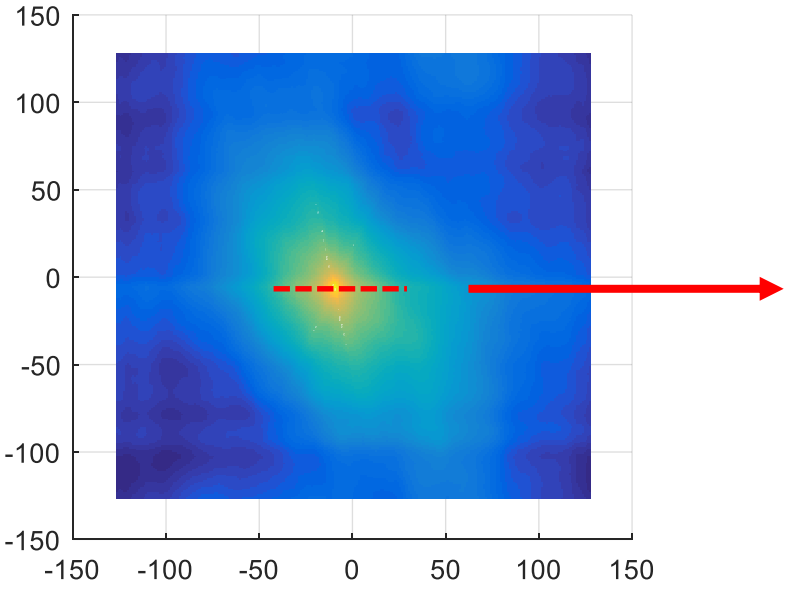
Image 2

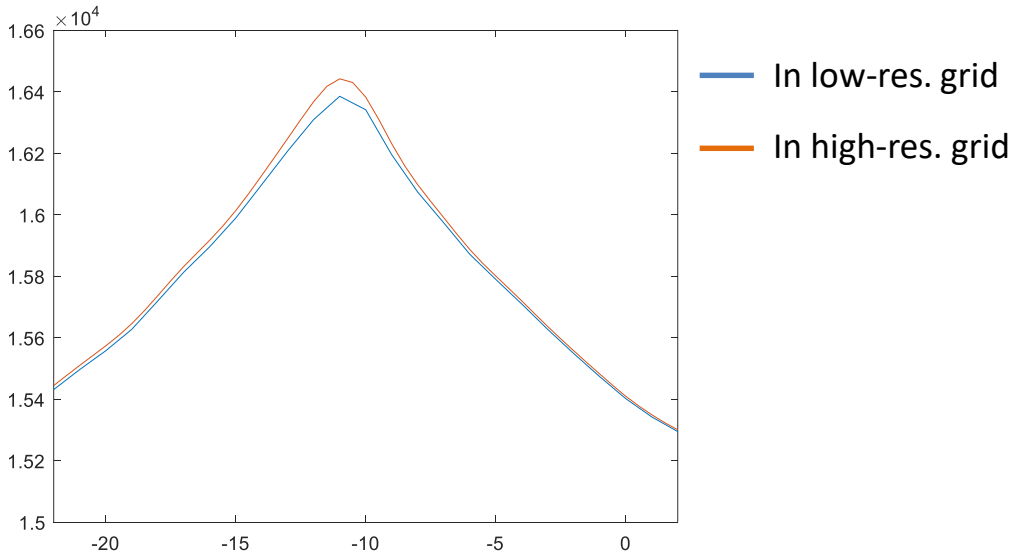


Cross-correlation

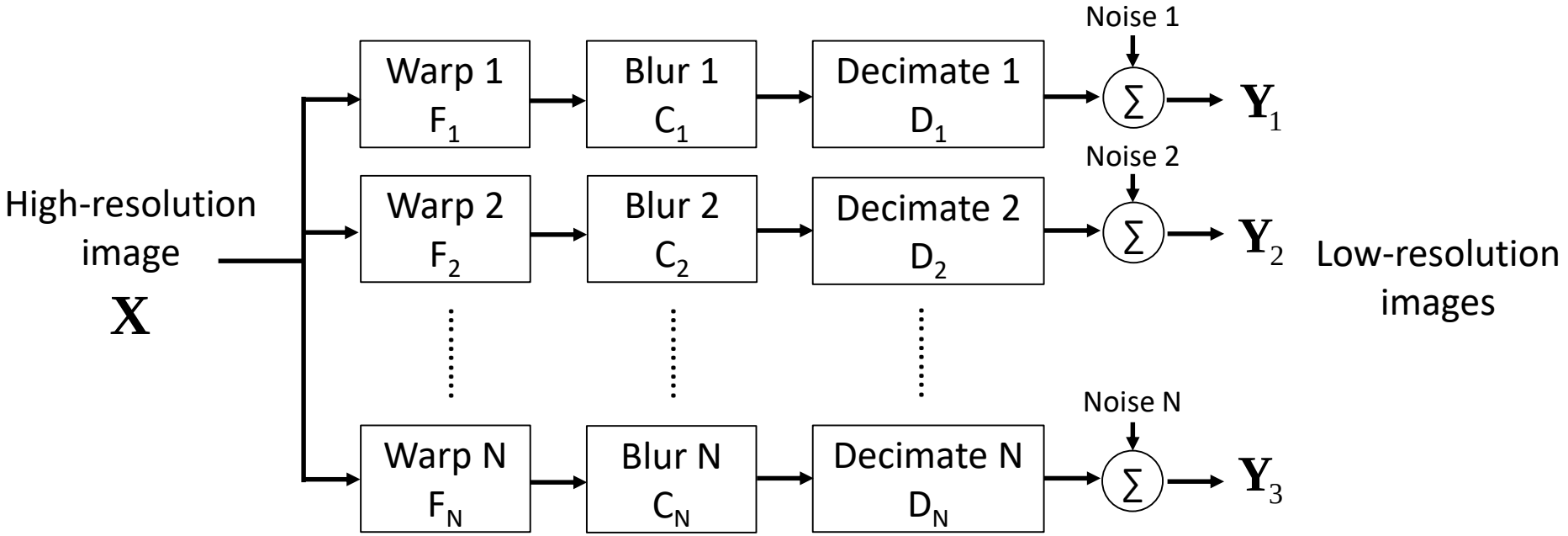
Sub-Pixel Image Registration

- Estimate sub-pixel shift in high-res. grid





Degradation Model



Decimation: systematic degradation of spatial resolution

Solve Image Super-Resolution

Image formation

$$\begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix} = \begin{bmatrix} D_1 C_1 F_1 \\ \vdots \\ D_N C_N F_N \end{bmatrix} \mathbf{X} + \begin{bmatrix} E_1 \\ \vdots \\ E_N \end{bmatrix} = \begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix} \mathbf{X} + \mathbf{E}$$

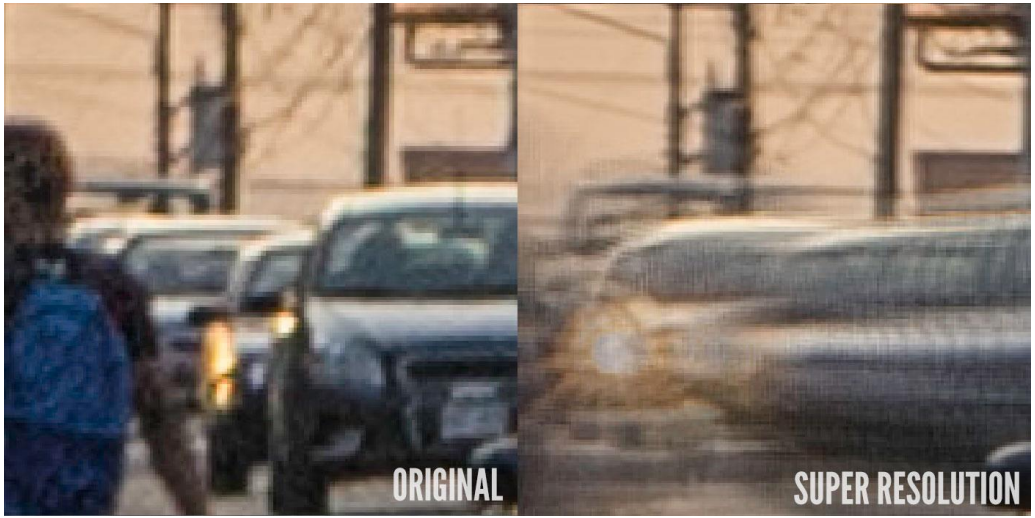
Solve

$$X_{MAP} = \arg\min_X \left\| \mathbf{Y} - \mathbf{H}\mathbf{X} \right\|_2^2 + \lambda TV(\mathbf{X})$$

H: system matrix

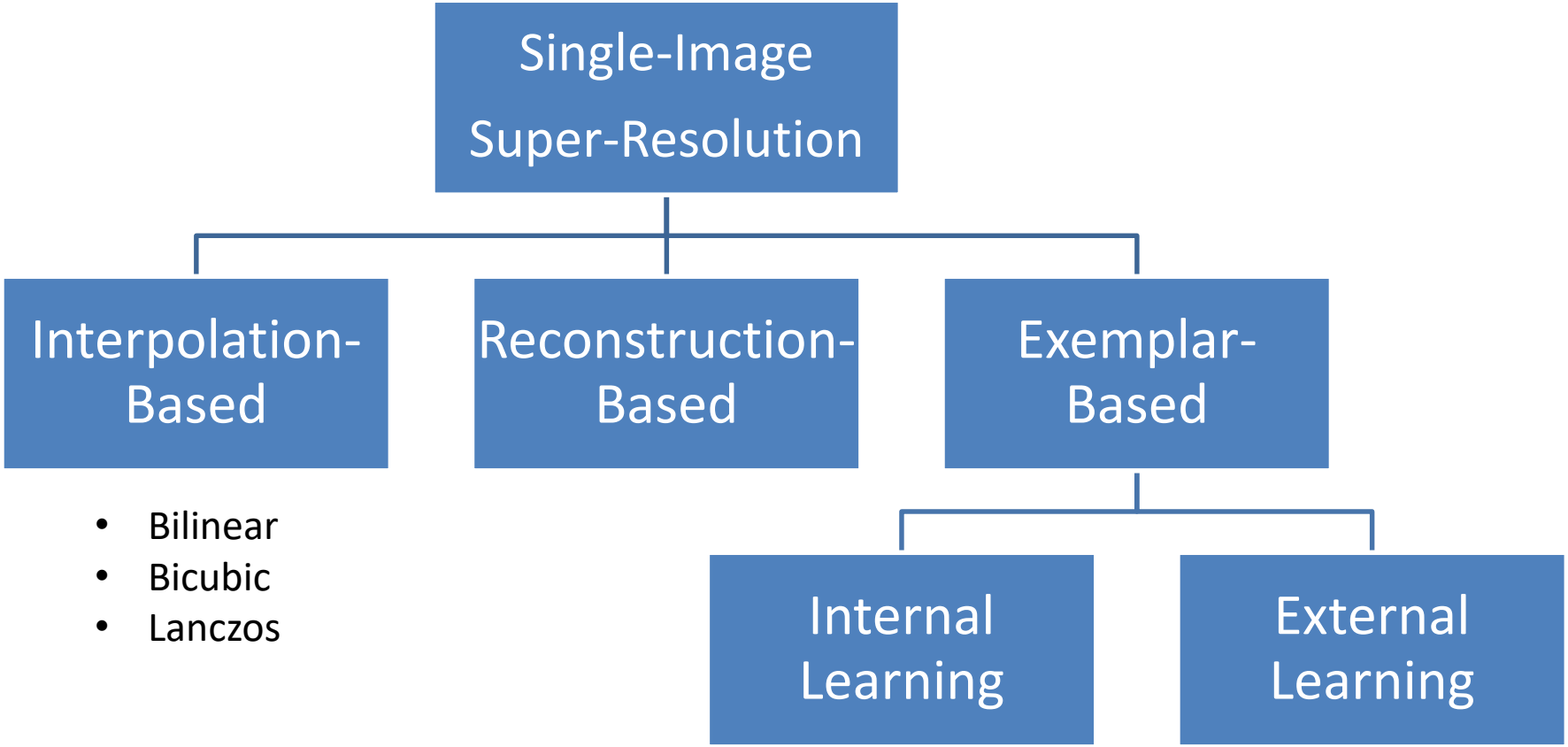
Limitation of Traditional SR

- Needs precise (subpixel accuracy) motion estimates.
- Long capture time due to multiple frame acquisition.
- Subject should not move.



RELATED WORK

Single-Image Super-Resolution

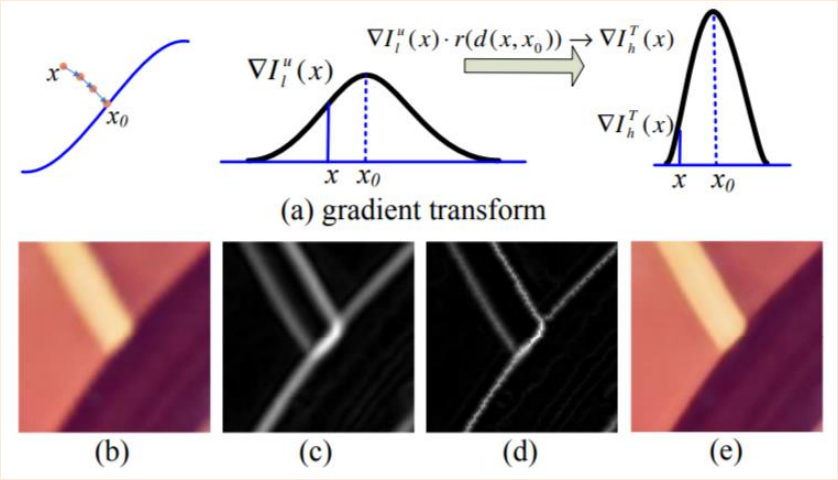


- Bilinear
- Bicubic
- Lanczos

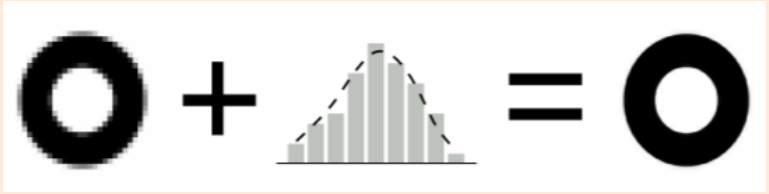
Reconstruction-Based Methods

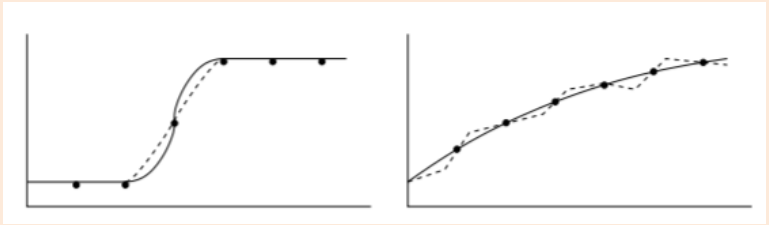
- Main idea: Preserve edge information

Sun et al., CVPR 2008



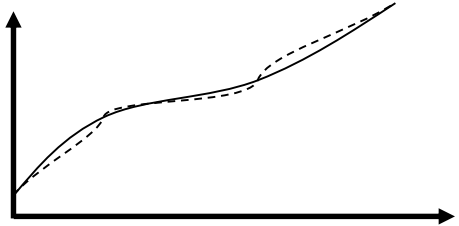
Fattal, SIGGRAPH 2007

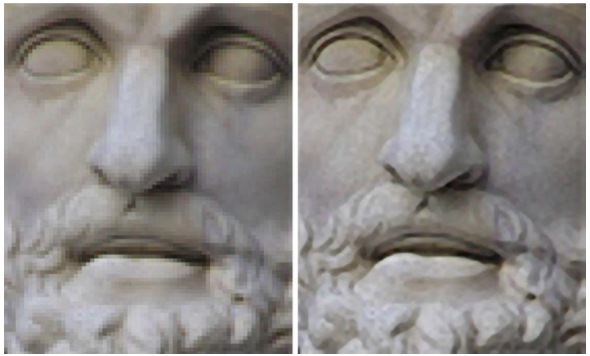


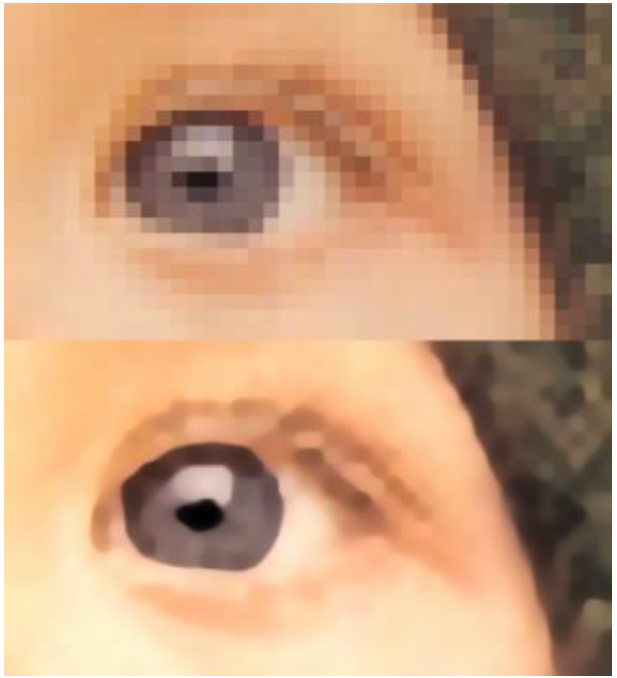


Limitations of Reconstruction-Based Methods

- Texture information is smudged
- Ringing artifacts Imposed prior is not true for arbitrary images

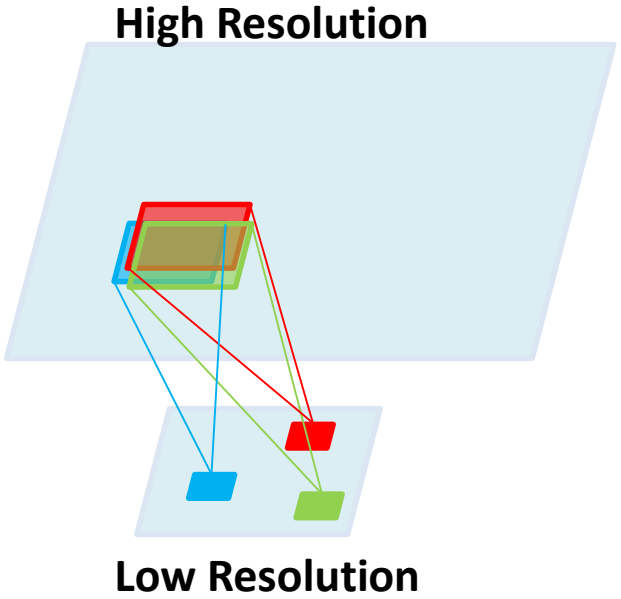




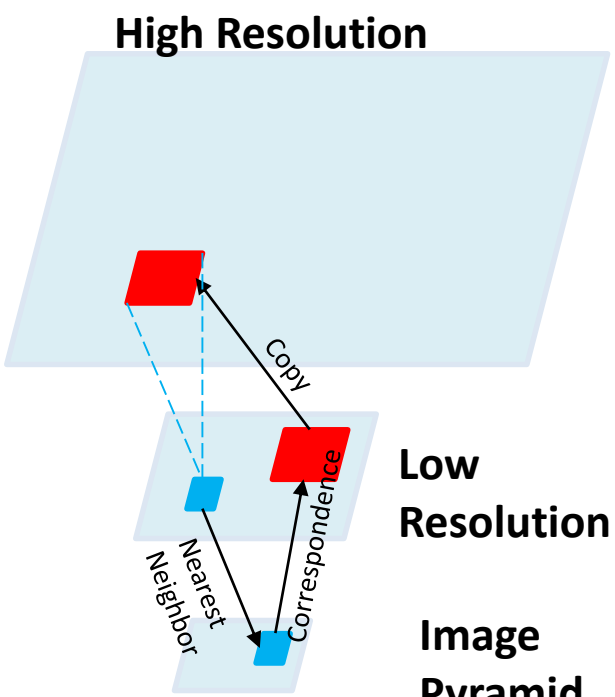


Self Example Super-resolution

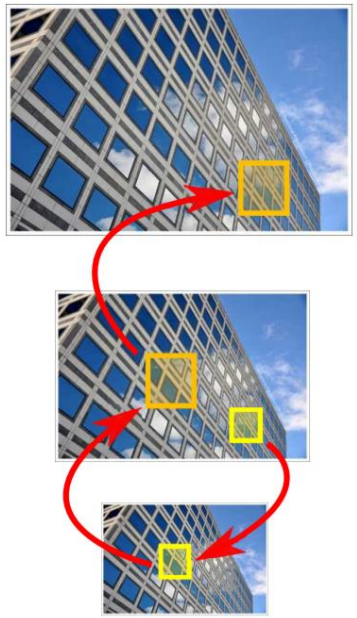
In-scale super-resolution



Cross-scale super-resolution

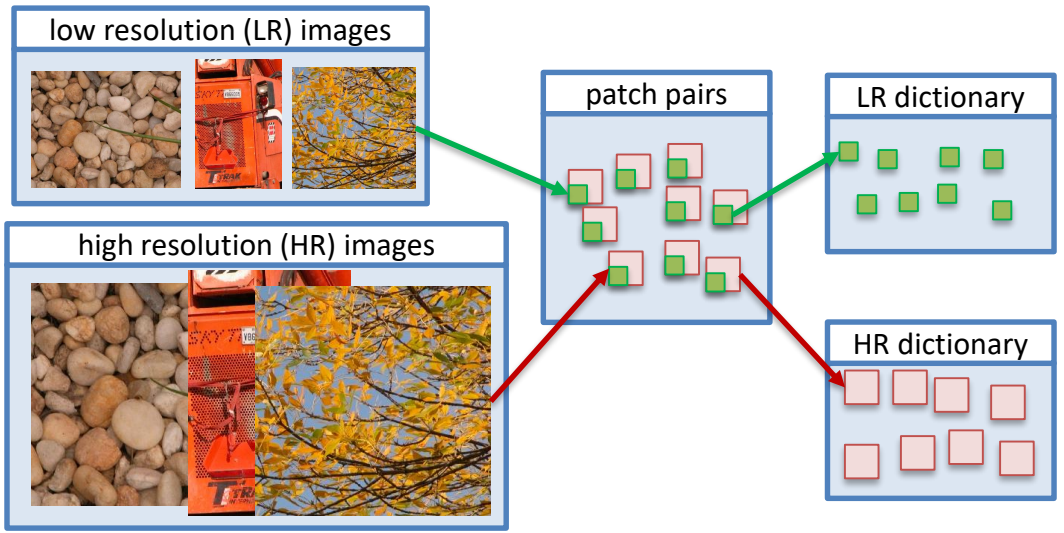


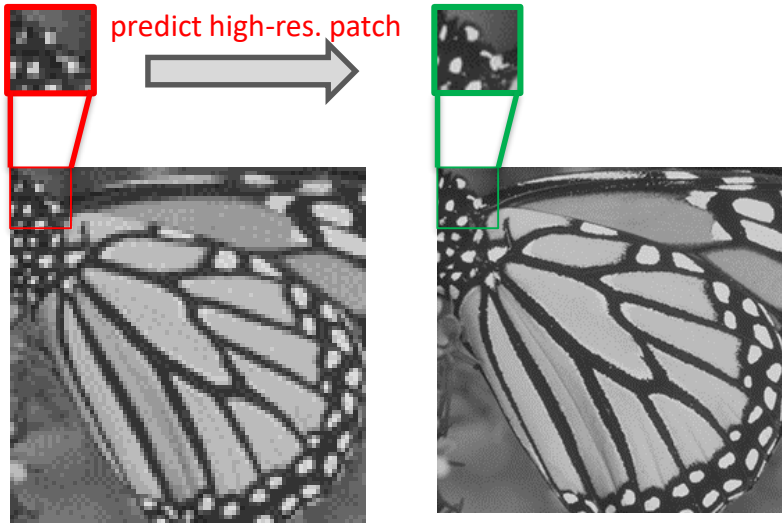
Huang, CVPR 2015



Learning-Based Super-Resolution

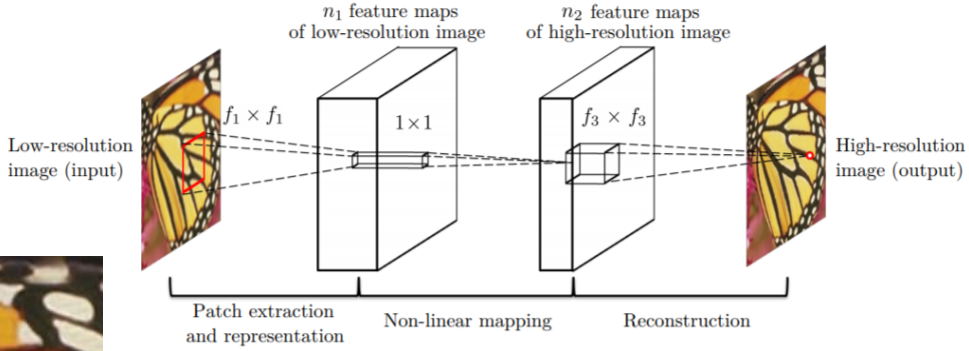
- Predict a high-res. pixel from a low-res. patch with a contextual information
- Learning the correspondences of patch pairs from a large dataset of image pairs of low-res. and high-res. images.





SRCNN

- First deep learning-based super-resolution
- Shallow network: 3-layer CNN
- Simple architecture
- Good performance

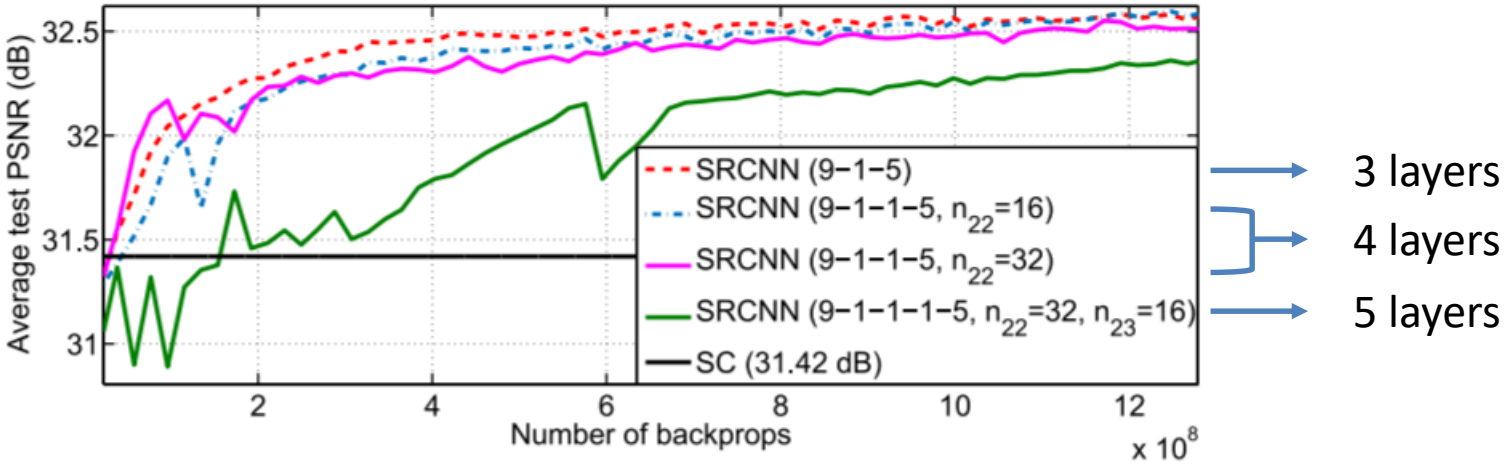


Bicubic

SRCNN

Deep CNN

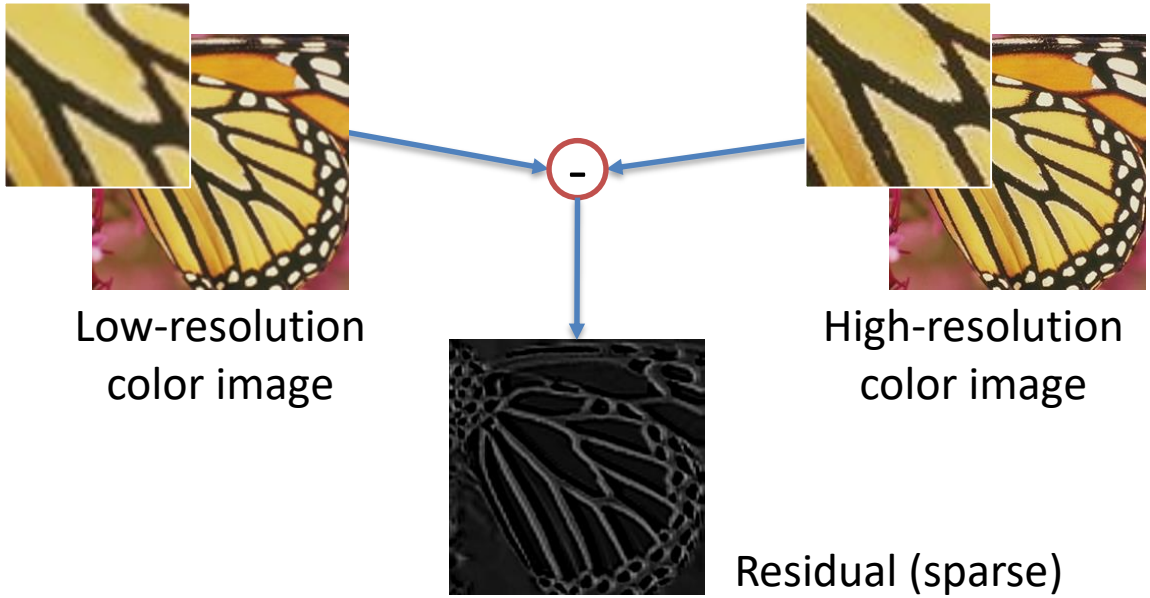
- Vanishing gradient problem
  - Deeper network decreases performance
  - End-to-end relation requires very long-term memory



Residual Learning

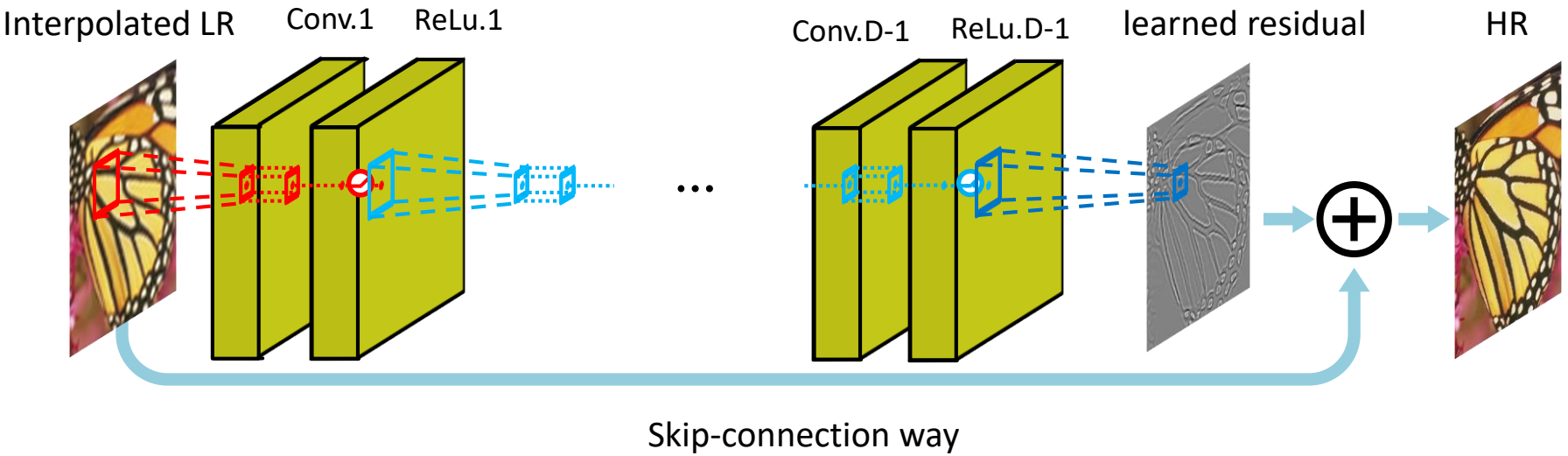
Kaiming He et al., 2015

- Learning difference of low-resolution and high-resolution
  - Prevent vanishing gradient problem



Residual Learning Model

Kim et al., CVPR 2016



- Skip connection to learn residual only
- 64 channels, 3×3 filters in each convolutional layer
- 20 convolutional layers (41×41 receptive field)

OUR METHOD

Image Registration Problem in Stereo System

KAIST

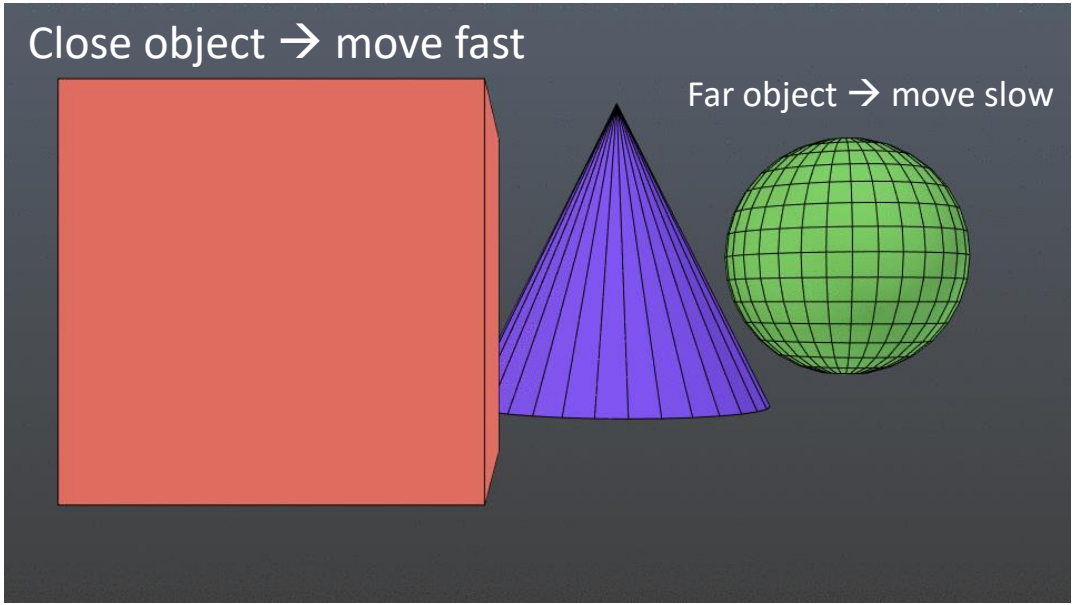
Left camera

Right camera

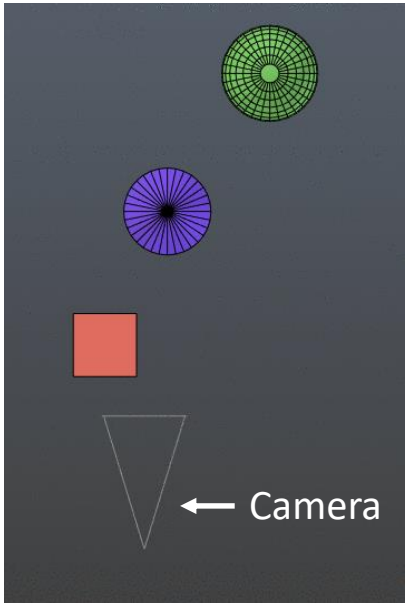
Overlap of left and right images

Two images do not overlap correctly due to disparity of two cameras.

Stereo Parallax

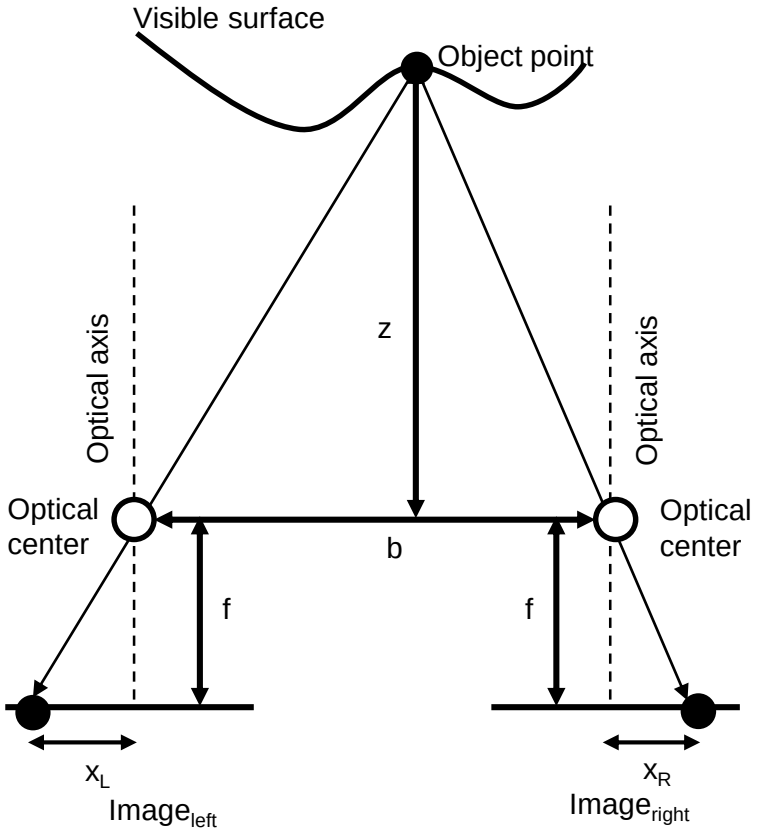


Camera view



Top view

Stereo Matching



Definitions

- Disparity (***d***): displacement of corresponding pixels

$$d = \left| X_L + X_R \right|$$

- Baseline (***b***): distance between the center of projections of two cameras

With simple trigonometry, we can derive the following equation:

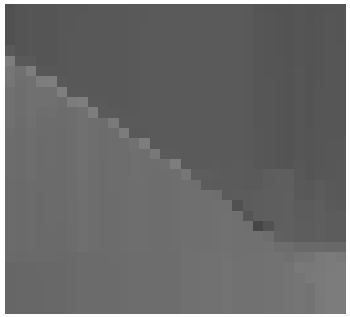
$$z = \frac{fb}{d}$$

Therefore, if we estimate disparity value for each pixel, depth information is obtainable through the known values of focal length and the baseline.

Problem of super-resolution using disparity

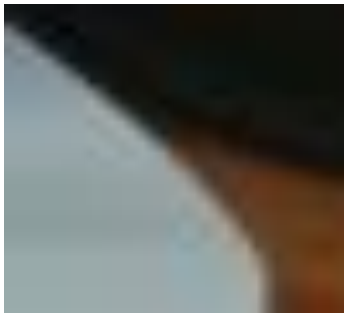
KAIST

- Disparity is not perfect
  - Especially incorrect on edges

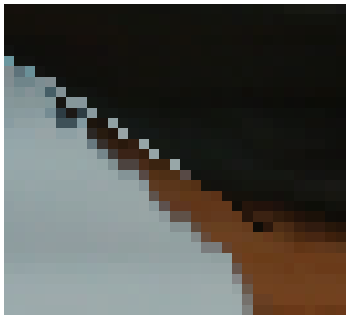


Disparity map

Left image



Right image



Warped right image

Parallax Prior

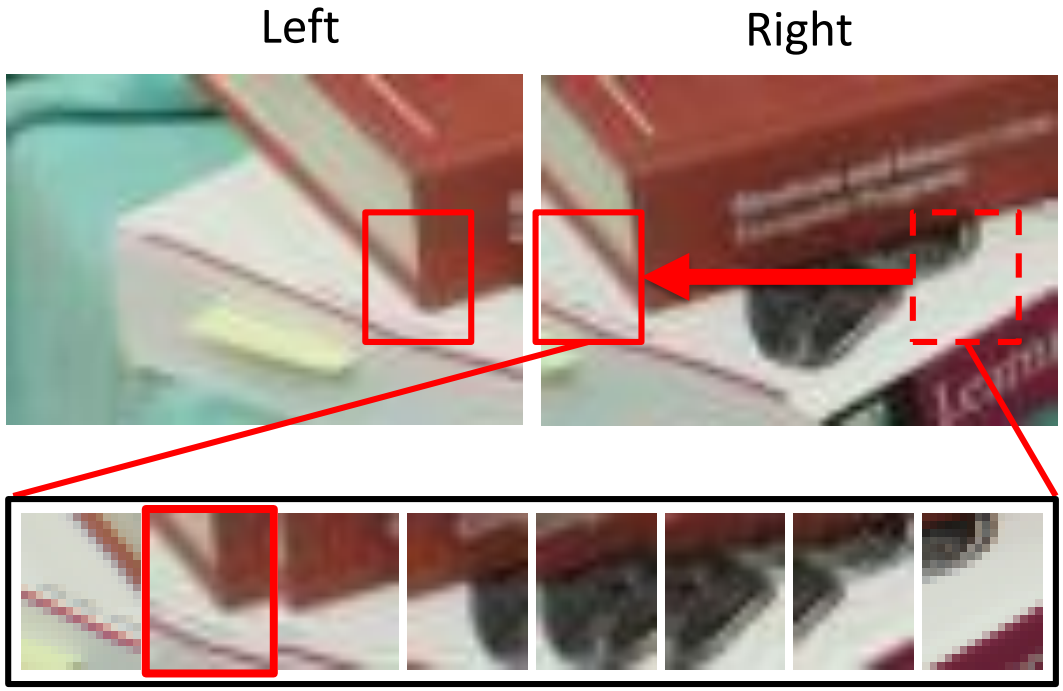


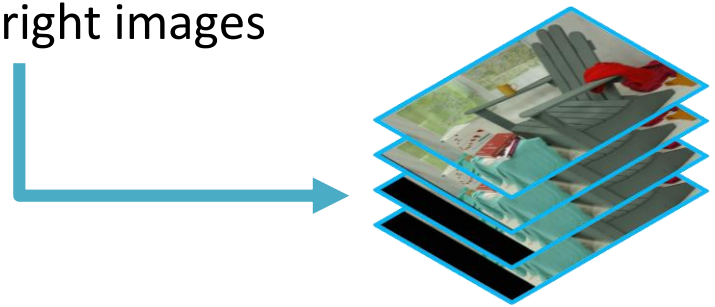
Image stack with shift intervals

Input With Shifted Images

- 1 Left Image
- 64 shifted right images (1 ~ 64 pixels shift)

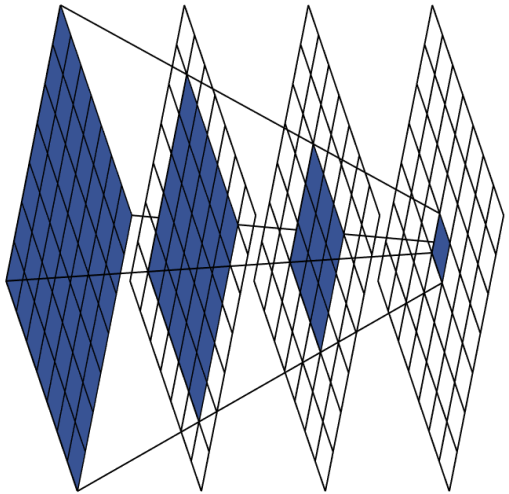
Left image

64 shifted right images



Receptive Field

- Receptive Field
  - Part of the input that is visible to a neuron
  - It increases as we stack more convolutional layers



Example) 3×3 kernel with 3 convolutions
➔ Total 7×7 receptive field size

Max Disparity

- Max disparity
  - 64 shifted images with one-pixel intervals
  - 16 layers ➔ 33×33 receptive field
  - Total 80 pixels max disparity
- More than about **98%** of disparities in the Middlebury stereo dataset are with this range

Dataset

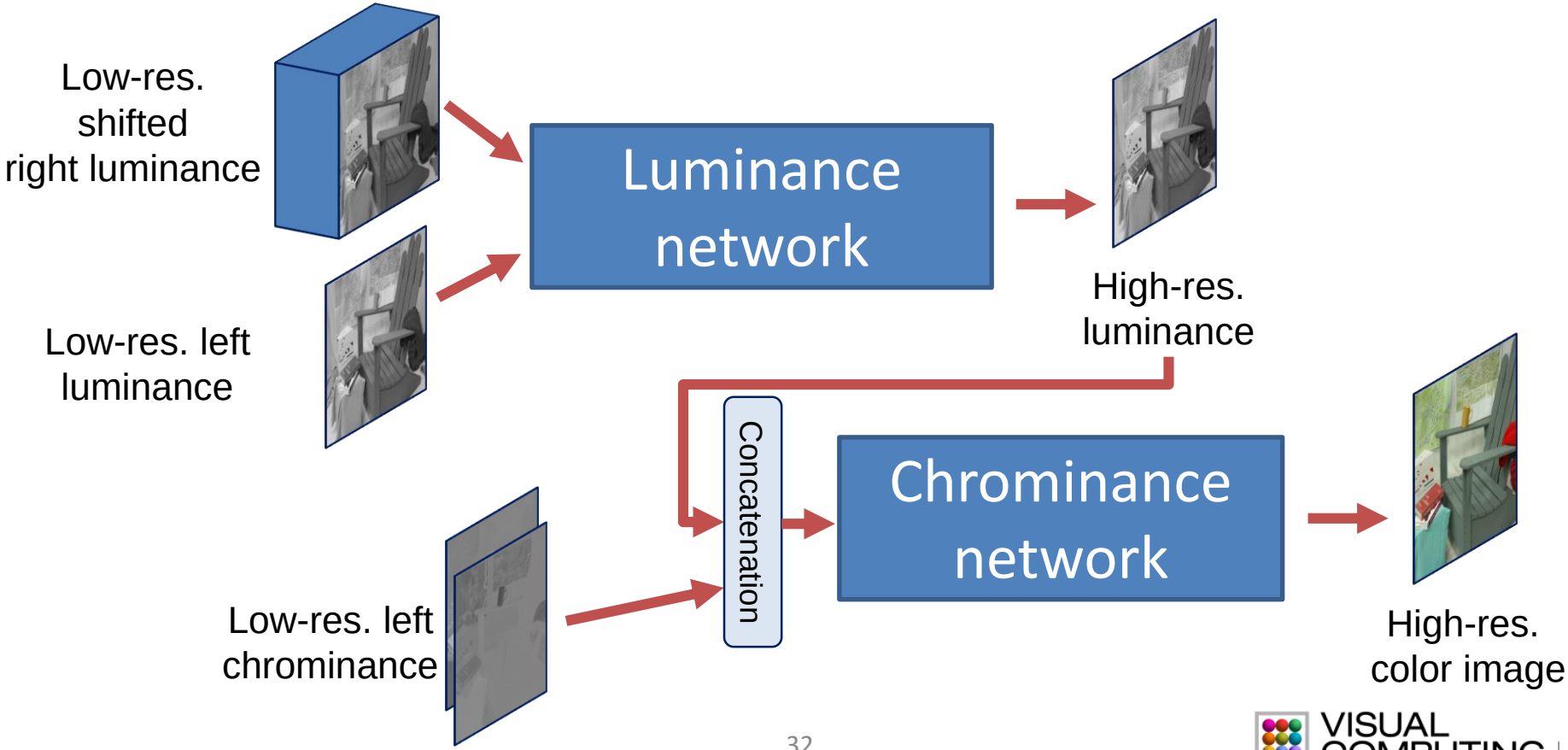
- 60 Middlebury images
- Input
  - 33×33 patch pairs
  - 24 stride
  - Data augmentation

Original patch

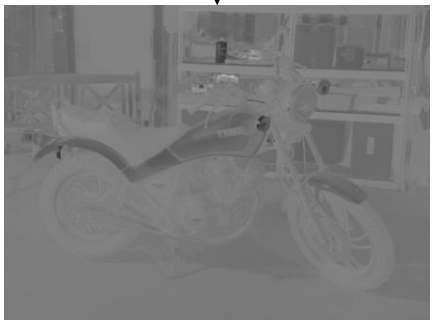
Rotating

Flipping

Two Network Design

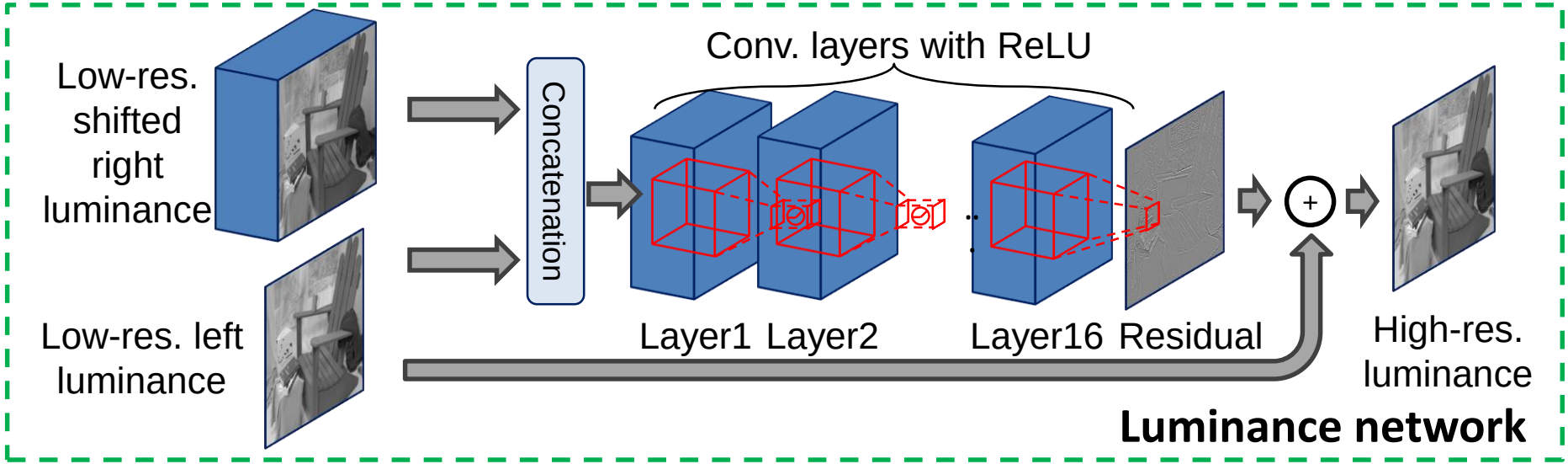


YCbCr Color Space



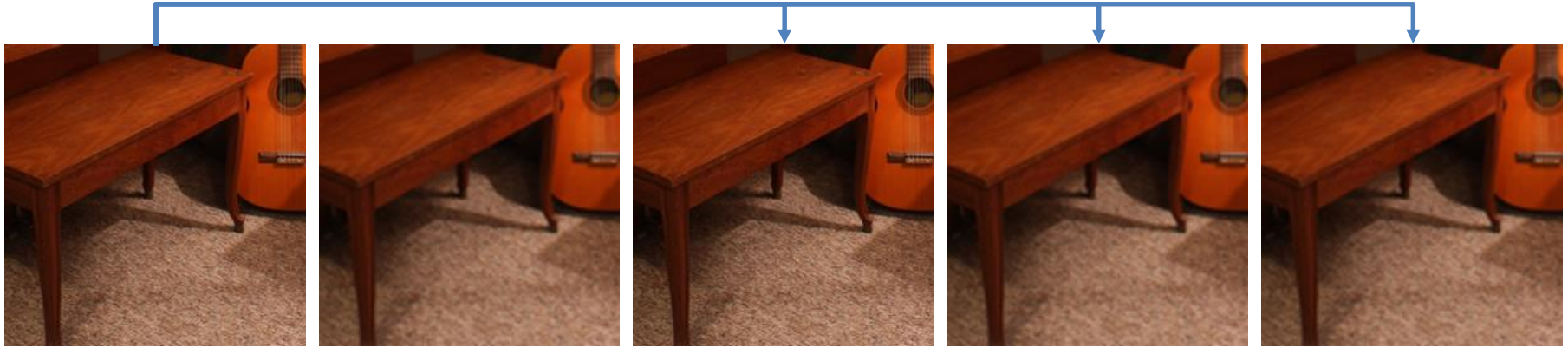
Y (luminance) Cb (color components) Cr

Luminance Network



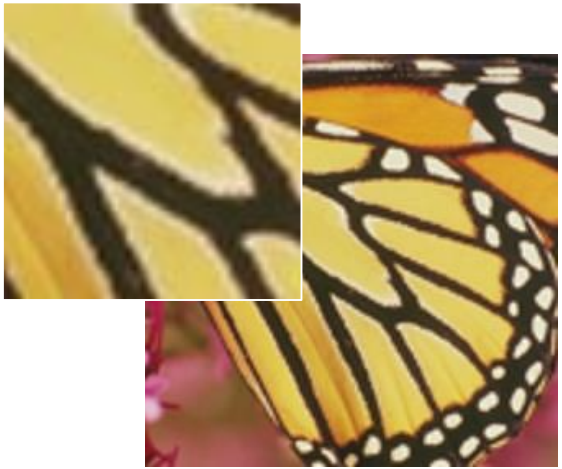
Effect of Luminance

- Luminance is more important for resolution than chrominance.



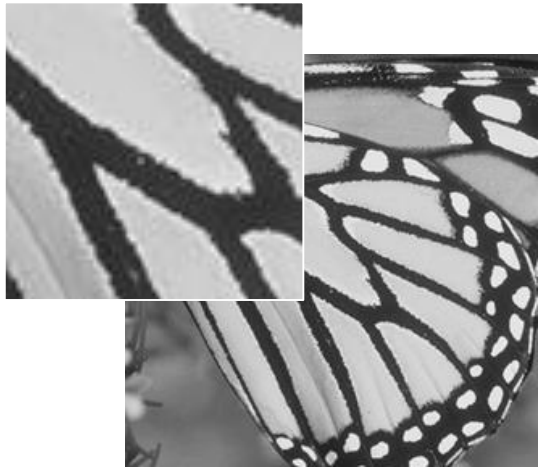
Ground truth | Low-resolution | High res. Y, Low res. C_b, C_r | High res. C_b, Low res. Y, C_r | High res. C_r, Low res. Y, C_b

Chrominance Upsampling using Luminance

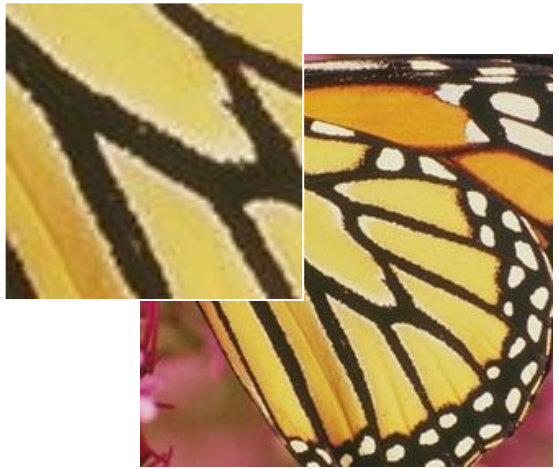


Low-resolution color image

+



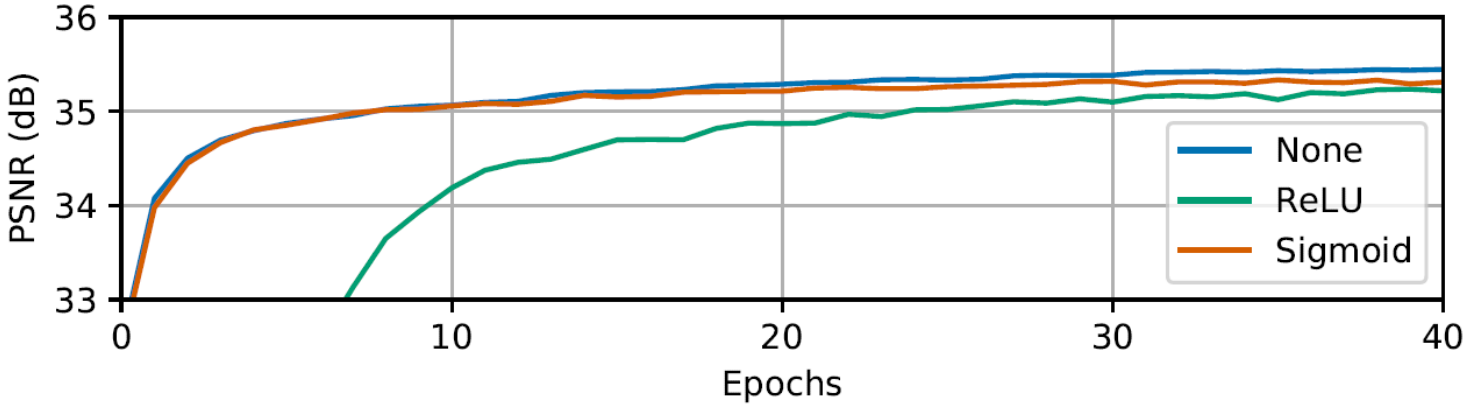
High-resolution luminance image



High-resolution color image

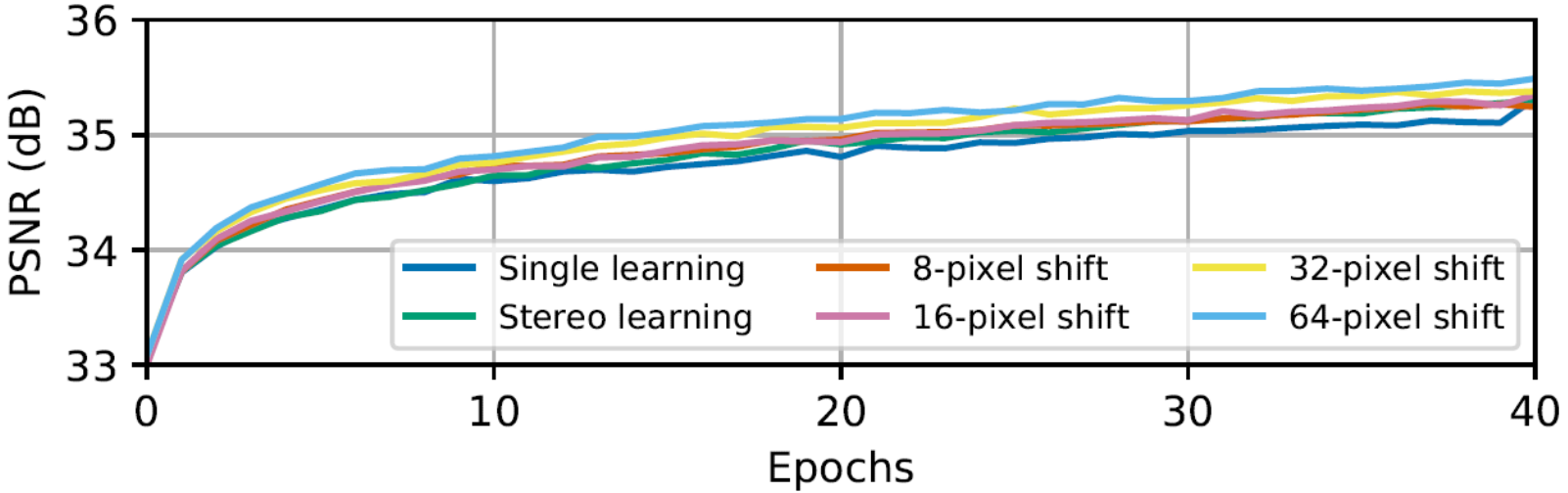
Compare Activation Functions

- The last convolutional layer handles the contrast variance of the residual image
  - Activation functions cause contrast compression
  - No activation function at the last layer shows highest PSNR



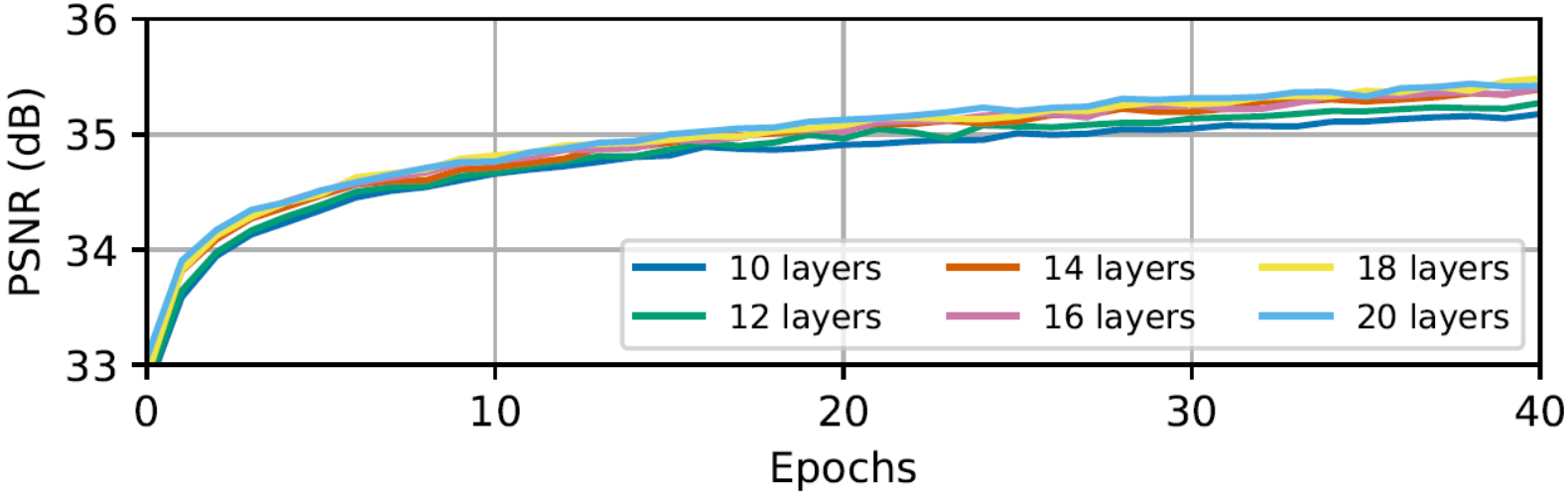
Compare Number of Shifts

- Large number of shifts improves PSNR
  - Covers more disparity range
  - Better than naïve two image super-resolution

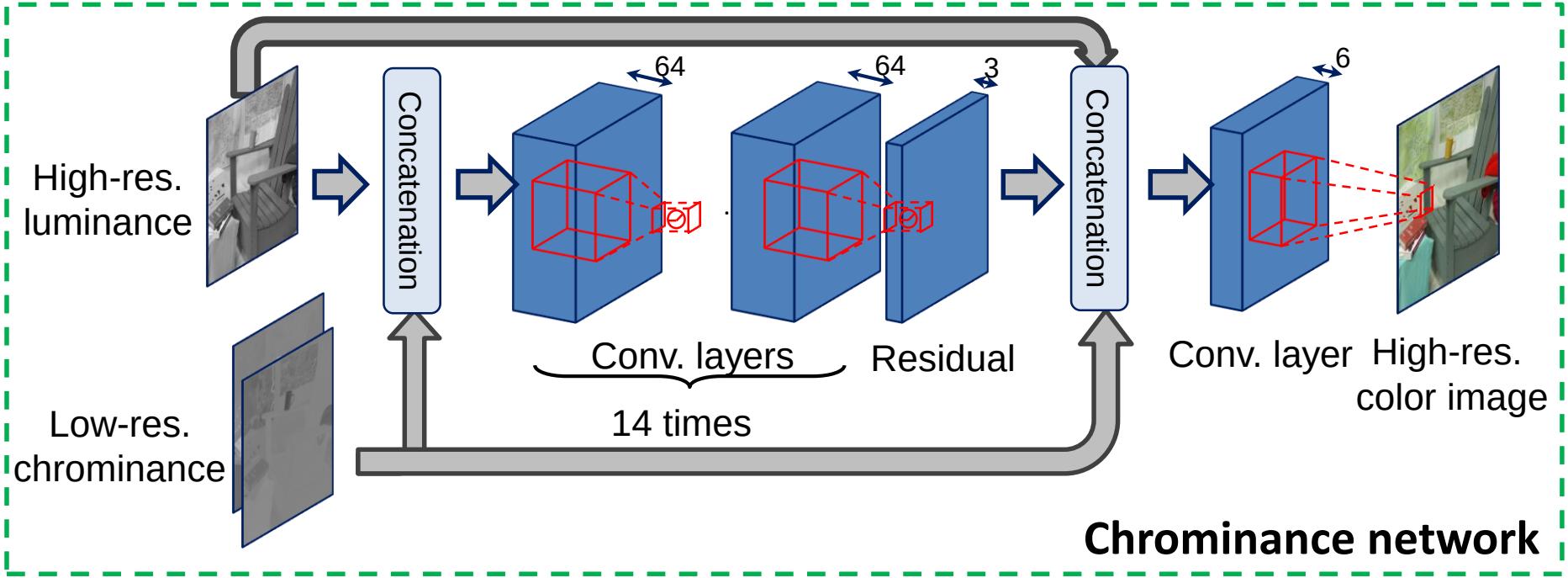


Compare Number of Layers

- Large number of layers improves PSNR
  - Also increases the computational time
  - 16 layers seems optimal

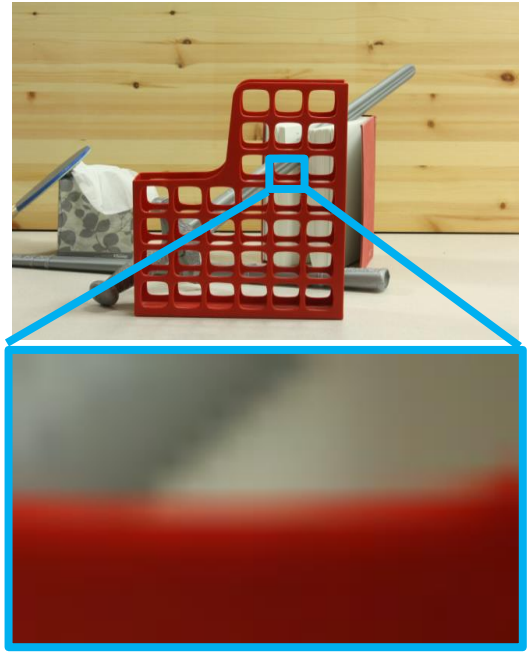
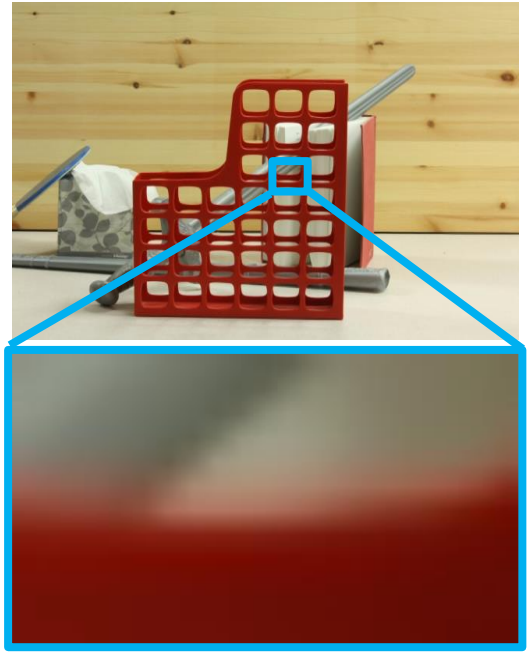
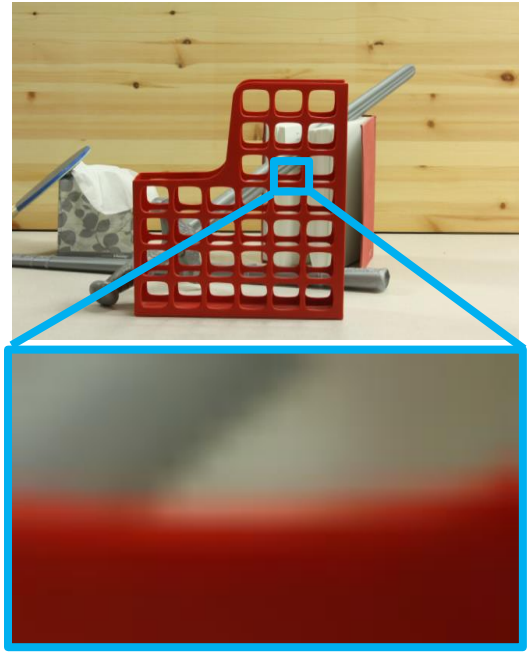


Chrominance Network



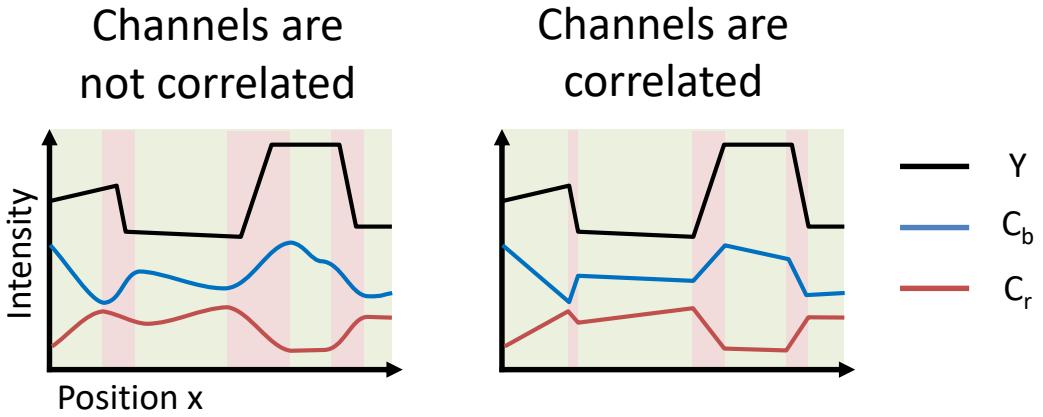
Chrominance Super-Resolution

KAIST

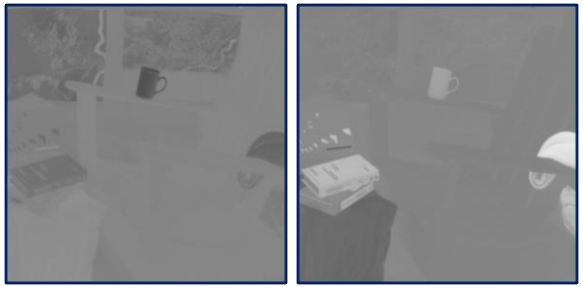
Ground truth	High-res. luminance + Low-res. chrominance	High-res. Luminance + High-res. chrominace
		
PSNR	37.20 dB	39.04 dB

Additional Layer for Color

- Idea: Cross-channel correlation
  - Luminance is in high-resolution



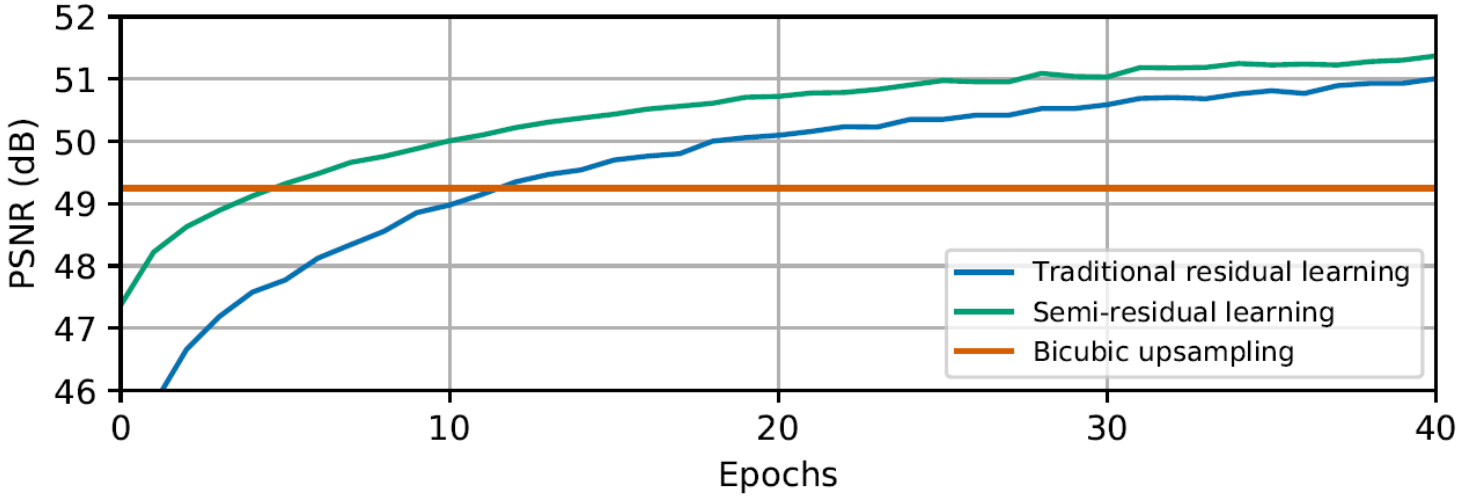
Y



C_b C_r

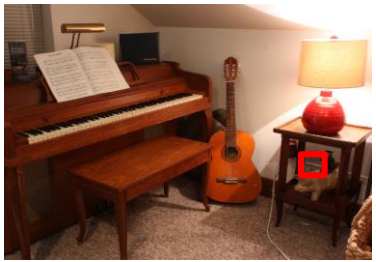
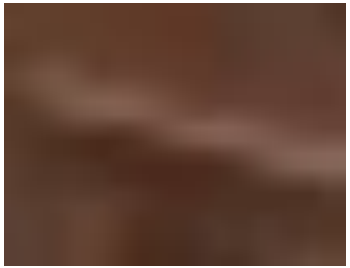
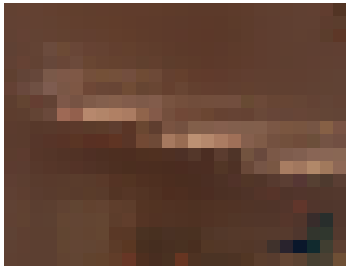
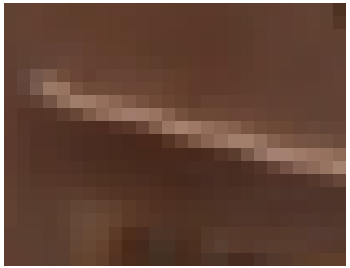
Compare Color Upsampling

- Additional convolutional layer increases performance (Semi-residual learning)



RESULTS

Stereo-based super-resolution vs. ours

Original image	Ground truth	Bicubic	Bhavsar	Our method
				
	PSNR	32.00 dB	26.85 dB	35.90 dB
				
	PSNR	30.01 dB	26.28 dB	32.12 dB

Single-image super-resolution vs. ours

Original image

Ground truth (PSNR)

Bicubic (25.47 dB)

SRCNN (27.03 dB)

VDSR (27.22 dB)

PsyCo (27.39 dB)

Ours (28.10 dB)

Single-image super-resolution vs. ours

Original image

Ground truth (PSNR)

Bicubic (26.82 dB)

SRCNN (28.28 dB)

VDSR (28.75 dB)

PsyCo (28.73 dB)

Ours (29.20 dB)

Benchmark

Scene	Scale	Bicubic		SRCNN		VDSR		PSyCo		Ours	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Middlebury (5 images)	×2	29.64	0.9228	31.48	0.9505	31.62	0.9543	32.03	0.9542	32.90	0.9545
	×3	27.20	0.8737	28.76	0.9136	29.30	0.9240	29.40	0.9231	29.47	0.9117
	×4	25.79	0.8344	27.11	0.8814	27.23	0.8916	27.58	0.8935	27.46	0.8730
Tsukuba (16 images)	×2	36.69	0.9833	40.05	0.9846	40.50	0.9840	41.08	0.9846	42.87	0.9952
	×3	32.98	0.9659	35.89	0.9611	36.70	0.9666	36.74	0.9657	37.35	0.9837
	×4	30.89	0.9453	33.10	0.9343	33.40	0.9463	33.83	0.9418	34.23	0.9678
KITTI 2012 (16 images)	×2	28.08	0.9200	29.27	0.9162	29.48	0.9137	29.82	0.9202	30.11	0.9472
	×3	25.72	0.8701	26.98	0.8533	27.24	0.8552	27.46	0.8637	27.53	0.9041
	×4	24.33	0.8345	25.34	0.8002	25.44	0.8043	25.73	0.8142	25.75	0.8641
KITTI 2015 (100 images)	×2	27.14	0.9176	28.50	0.9193	28.60	0.9156	28.75	0.9188	28.91	0.9460
	×3	24.74	0.8597	26.19	0.8530	26.37	0.8539	26.37	0.8548	26.36	0.8978
	×4	23.34	0.8130	24.44	0.7951	24.50	0.7999	24.64	0.7998	24.64	0.8536

Conclusion

- Reconstruct a high-resolution image without direct disparity calculation
- Proposed stereo super-resolution outperforms current state-of-the-art methods
- It can be used with any other stereo imaging methods additionally