A Comparative Study on Wavelets and Residuals in Deep Super-Resolution

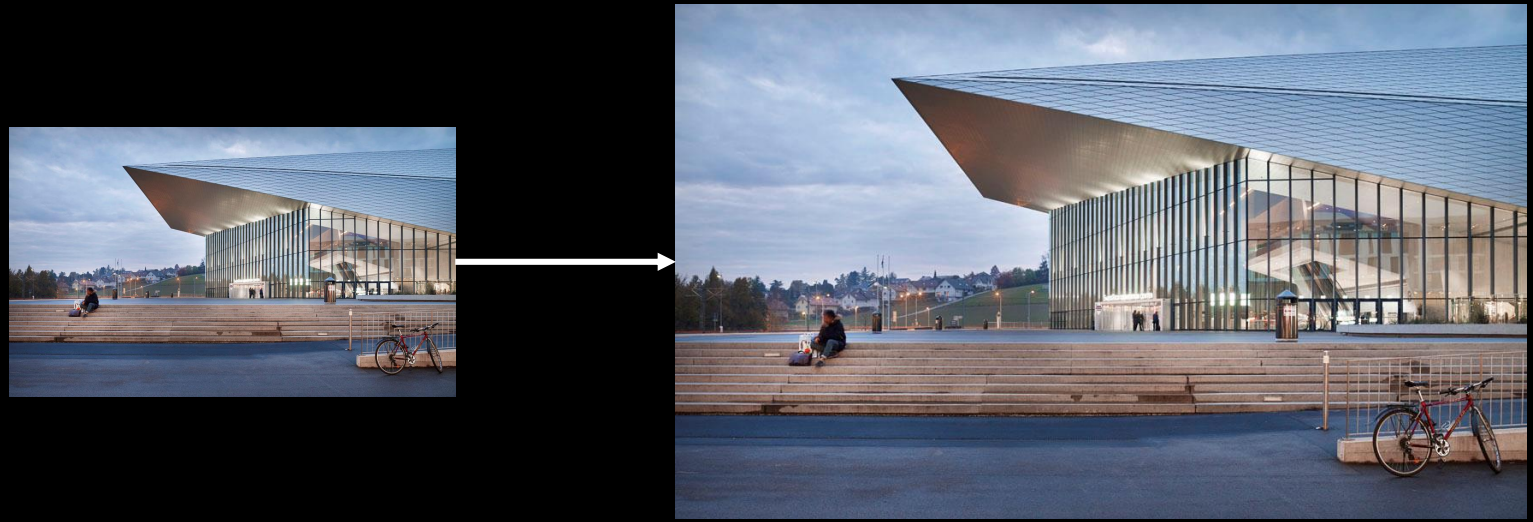
Ruofan Zhou, **Fayez Lahoud**, Majed EI Helou, and Sabine Süsstrunk

Image and Visual Representation Lab

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Super-Resolution

- Obtaining a high-resolution image from a low-resolution image



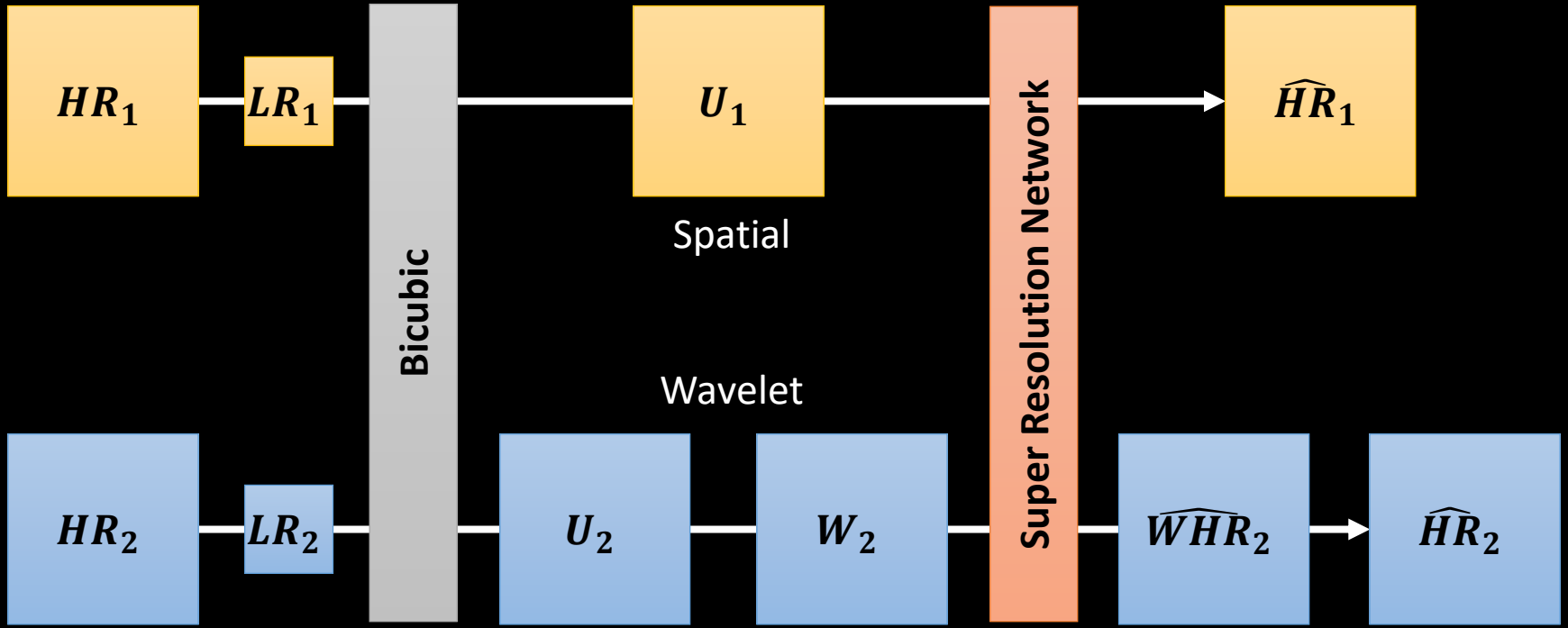
- Deep learning[1] comes in 2014

[1] Chao Dong et al. Image super-resolution using deep convolutional networks. ECCV 2014

Super-resolution architectures

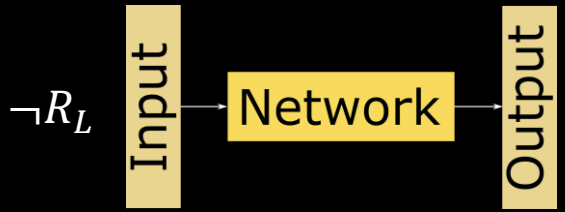
- Multiple models
- Multiple inputs
- Unclear effects

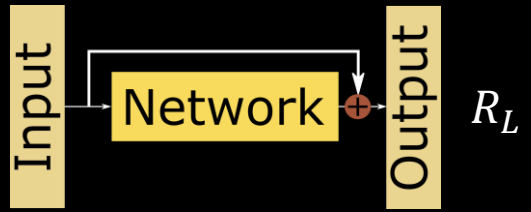
Super-Resolution Networks



Techniques used in Super-Resolution Networks

- Residual learning





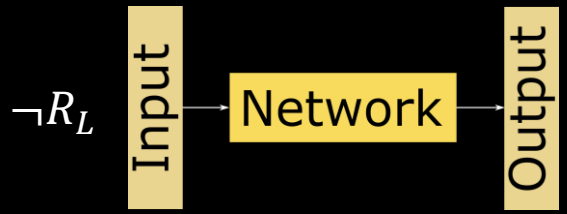
- Reduced and stable training
- Higher accuracy

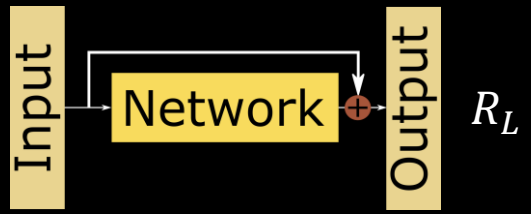
- Easier than predicting a natural image

Kai Zhang et al. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing 2017

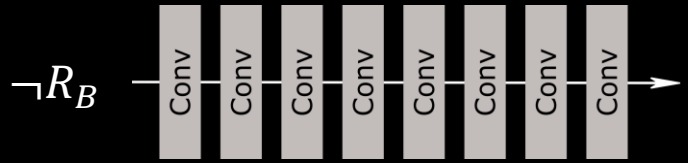
Techniques used in Super-Resolution Networks

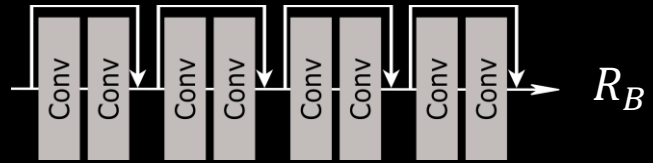
- Residual learning





- Residual blocks





Bee Lim. Enhanced Deep Residual Networks for Single Image Super-Resolution . CVPRW 2017

Techniques used in Super-Resolution Networks

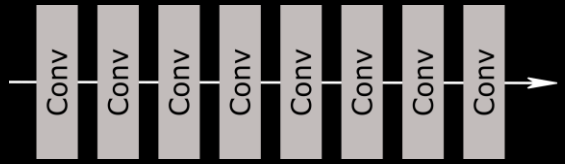
- Residual learning

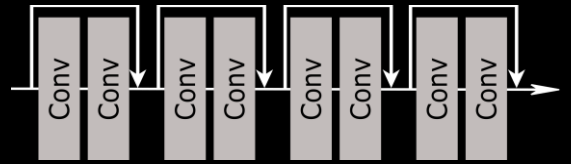
$\neg R_L$

R_L

- Residual blocks

$\neg R_B$

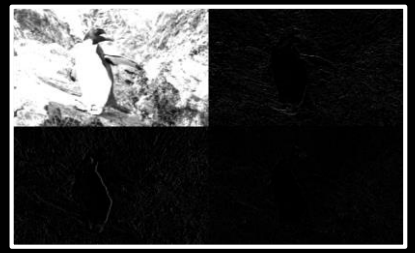




R_B

- Wavelet Decomposition

$\mathcal{S}$



$\mathcal{W}$

Tiantong Guo et al. Deep wavelet prediction for image super-resolution. CVPRW 2017

Techniques used in Super-Resolution Networks

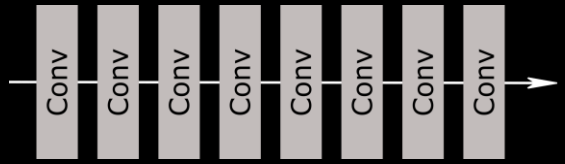
- Residual learning

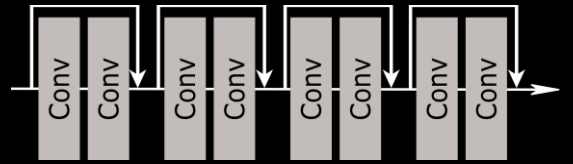
$\neg R_L$

R_L

- Residual blocks

$\neg R_B$

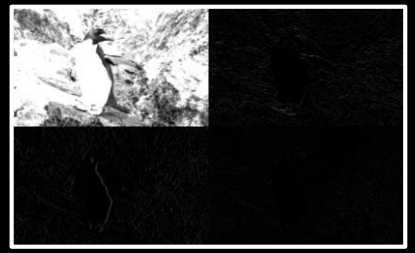




R_B

- Wavelet Decomposition

$\mathcal{S}$



$\mathcal{W}$

Which one helps?

Experiments

- Three parameters
  - Without | With residual learning $(\neg R_L | R_L)$
  - Spatial input | Wavelet input $(\mathcal{S} | \mathcal{W})$
  - Without | With residual blocks $(\neg R_B | R_B)$

- Training dataset (at least 2K)
  - DIV2K[2]
  - Training 800 high-resolution images
  - Validation 100 high-resolution images

Eirikur Agustsson et al. NTIRE 2017 challenge on single image super-resolution: Dataset and study. CVPRW 2017

DIV2K






Experiments

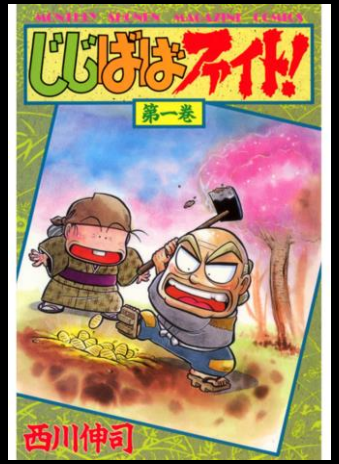
- All networks
  - 12 convolutional layers
  - 64 kernels of 3 x 3
  - Patches 64 x 64
  - 100 epochs
  - Adam optimizer with $lr = 0.001$
    - decayed by factor of 10 every 30 epochs
  - Same initialization (Xavier)

- Scales x2, x3, and x4 using MATLAB's *imresize* (bicubic)

Experiments

Set5

Set14



Manga109

BSDS100

Urban100

Experiments

- Network configuration ($\mathcal{S}$, R_L, $\neg R_B$)
- Performance evaluation
  - PSNR
  - SSIM
- Statistical significance
  - T-test

Set	Set5			Set14			BSDS100			Urban100			Manga109		
Scale	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4
Bicubic	31.79	26.95	26.69	28.00	24.44	23.81	26.11	24.66	22.38	25.43	**21.30**	21.70	26.79	24.61	22.05
$(\mathcal{S}, \neg R_L, \neg R_B)$	34.52	27.77	28.43	29.36	24.58	24.47	25.93	24.72	21.91	28.25	21.13	22.93	27.22	25.99	22.28
$(\mathcal{S}, \neg R_L, R_B)$	34.94	27.99	28.81	29.58	24.66	24.64	25.99	24.73	21.86	28.65	21.13	23.24	27.47	26.21	22.37
$(\mathcal{S}, R_L, \neg R_B)$	34.99	28.02	28.89	29.62	24.66	24.66	25.89	24.73	23.17	28.67	21.14	23.22	27.38	26.31	22.45
$(\mathcal{S}, R_L, R_B)$	34.80	27.99	28.88	29.51	24.64	24.67	25.91	24.70	21.82	28.51	21.10	23.24	27.35	26.23	22.35
$(\mathcal{W}, \neg R_L, \neg R_B)$	34.42	27.80	28.75	29.23	24.58	24.57	26.25	24.71	21.89	27.96	21.09	23.13	27.50	26.04	22.36
$(\mathcal{W}, \neg R_L, R_B)$	34.89	27.95	28.85	29.57	24.61	24.70	26.46	24.70	21.98	28.51	21.06	23.28	27.87	26.18	22.62
$(\mathcal{W}, R_L, \neg R_B)$	34.84	27.96	28.94	29.51	24.62	24.74	26.35	24.70	21.93	28.42	21.06	23.28	27.87	26.19	22.46
$(\mathcal{W}, R_L, R_B)$	34.80	28.00	28.93	29.54	24.64	24.69	26.33	24.70	21.93	28.43	21.07	23.30	27.90	26.20	22.47

Set	Set5			Set14			BSDS100			Urban100			Manga109		
Scale	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4
Bicubic	31.79	26.95	26.69	28.00	24.44	23.81	26.11	24.66	22.38	25.43	21.30	21.70	26.79	24.61	22.05
$(\mathcal{S}, \neg R_L, \neg R_B)$	34.52	27.77	28.43	29.36	24.58	24.47	25.93	24.72	21.91	28.25	21.13	22.93	27.22	25.99	22.28
$(\mathcal{S}, \neg R_L, R_B)$	34.94	27.99	28.81	29.58	24.66	24.64	25.99	24.73	21.86	28.65	21.13	23.24	27.47	26.21	22.37
$(\mathcal{S}, R_L, \neg R_B)$	34.99	28.02	28.89	29.62	24.66	24.66	25.89	24.73	23.17	28.67	21.14	23.22	27.38	26.31	22.45
$(\mathcal{S}, R_L, R_B)$	34.80	27.99	28.88	29.51	24.64	24.67	25.91	24.70	21.82	28.51	21.10	23.24	27.35	26.23	22.35
$(\mathcal{W}, \neg R_L, \neg R_B)$	34.42	27.80	28.75	29.23	24.58	24.57	26.25	24.71	21.89	27.96	21.09	23.13	27.50	26.04	22.36
$(\mathcal{W}, \neg R_L, R_B)$	34.89	27.95	28.85	29.57	24.61	24.70	26.46	24.70	21.98	28.51	21.06	23.28	27.87	26.18	22.62
$(\mathcal{W}, R_L, \neg R_B)$	34.84	27.96	28.94	29.51	24.62	24.74	26.35	24.70	21.93	28.42	21.06	23.28	27.87	26.19	22.46
$(\mathcal{W}, R_L, R_B)$	34.80	28.00	28.93	29.54	24.64	24.69	26.33	24.70	21.93	28.43	21.07	23.30	27.90	26.20	22.47

Closest net to bicubic $(\mathcal{S}, \neg R_L, \neg R_B)$

$t_{psnr} = 3.92, p_{psnr} = 10^{-5} \mid t_{ssim} = 4.98, p_{ssim} = 7 \times 10^{-7}$

Set	Set5			Set14			BSDS100			Urban100			Manga109		
Scale	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4
Bicubic	31.79	26.95	26.69	28.00	24.44	23.81	26.11	24.66	22.38	25.43	**21.30**	21.70	26.79	24.61	22.05
$(\mathcal{S}, \neg R_L, \neg R_B)$	34.52	27.77	28.43	29.36	24.58	24.47	25.93	24.72	21.91	28.25	21.13	22.93	27.22	25.99	22.28
$(\mathcal{S}, \neg R_L, R_B)$	34.94	27.99	28.81	29.58	24.66	24.64	25.99	24.73	21.86	28.65	21.13	23.24	27.47	26.21	22.37
$(\mathcal{S}, R_L, \neg R_B)$	34.99	28.02	28.89	29.62	24.66	24.66	25.89	24.73	23.17	28.67	21.14	23.22	27.38	26.31	22.45
$(\mathcal{S}, R_L, R_B)$	34.80	27.99	28.88	29.51	24.64	24.67	25.91	24.70	21.82	28.51	21.10	23.24	27.35	26.23	22.35
$(\mathcal{W}, \neg R_L, \neg R_B)$	34.42	27.80	28.75	29.23	24.58	24.57	26.25	24.71	21.89	27.96	21.09	23.13	27.50	26.04	22.36
$(\mathcal{W}, \neg R_L, R_B)$	34.89	27.95	28.85	29.57	24.61	24.70	26.46	24.70	21.98	28.51	21.06	23.28	27.87	26.18	22.62
$(\mathcal{W}, R_L, \neg R_B)$	34.84	27.96	28.94	29.51	24.62	24.74	26.35	24.70	21.93	28.42	21.06	23.28	27.87	26.19	22.46
$(\mathcal{W}, R_L, R_B)$	34.80	28.00	28.93	29.54	24.64	24.69	26.33	24.70	21.93	28.43	21.07	23.30	27.90	26.20	22.47

No residuals, lowest performance

Set	Set5			Set14			BSDS100			Urban100			Manga109		
Scale	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4
Bicubic	31.79	26.95	26.69	28.00	24.44	23.81	26.11	24.66	22.38	25.43	**21.30**	21.70	26.79	24.61	22.05
$(\mathcal{S}, \neg R_L, \neg R_B)$	34.52	27.77	28.43	29.36	24.58	24.47	25.93	24.72	21.91	28.25	21.13	22.93	27.22	25.99	22.28
$(\mathcal{S}, \neg R_L, R_B)$	34.94	27.99	28.81	29.58	24.66	24.64	25.99	24.73	21.86	28.65	21.13	23.24	27.47	26.21	22.37
$(\mathcal{S}, R_L, \neg R_B)$	34.99	28.02	28.89	29.62	24.66	24.66	25.89	24.73	23.17	28.67	21.14	23.22	27.38	26.31	22.45
$(\mathcal{S}, R_L, R_B)$	34.80	27.99	28.88	29.51	24.64	24.67	25.91	24.70	21.82	28.51	21.10	23.24	27.35	26.23	22.35
$(\mathcal{W}, \neg R_L, \neg R_B)$	34.42	27.80	28.75	29.23	24.58	24.57	26.25	24.71	21.89	27.96	21.09	23.13	27.50	26.04	22.36
$(\mathcal{W}, \neg R_L, R_B)$	34.89	27.95	28.85	29.57	24.61	24.70	26.46	24.70	21.98	28.51	21.06	23.28	27.87	26.18	22.62
$(\mathcal{W}, R_L, \neg R_B)$	34.84	27.96	28.94	29.51	24.62	24.74	26.35	24.70	21.93	28.42	21.06	23.28	27.87	26.19	22.46
$(\mathcal{W}, R_L, R_B)$	34.80	28.00	28.93	29.54	24.64	24.69	26.33	24.70	21.93	28.43	21.07	23.30	27.90	26.20	22.47

$$t_{psnr} = 4.45, p_{psnr} = 5 \times 10^{-4} \mid t_{ssim} = 7.11, p_{ssim} = 5 \times 10^{-6}$$

Set	Set5			Set14			BSDS100			Urban100			Manga109		
Scale	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4
Bicubic	31.79	26.95	26.69	28.00	24.44	23.81	26.11	24.66	22.38	25.43	**21.30**	21.70	26.79	24.61	22.05
$(\mathcal{S}, \neg R_L, \neg R_B)$	34.52	27.77	28.43	29.36	24.58	24.47	25.93	24.72	21.91	28.25	21.13	22.93	27.22	25.99	22.28
$(\mathcal{S}, \neg R_L, R_B)$	34.94	27.99	28.81	29.58	24.66	24.64	25.99	24.73	21.86	28.65	21.13	23.24	27.47	26.21	22.37
$(\mathcal{S}, R_L, \neg R_B)$	34.99	28.02	28.89	29.62	24.66	24.66	25.89	24.73	23.17	28.67	21.14	23.22	27.38	26.31	22.45
$(\mathcal{S}, R_L, R_B)$	34.80	27.99	28.88	29.51	24.64	24.67	25.91	24.70	21.82	28.51	21.10	23.24	27.35	26.23	22.35
$(\mathcal{W}, \neg R_L, \neg R_B)$	34.42	27.80	28.75	29.23	24.58	24.57	26.25	24.71	21.89	27.96	21.09	23.13	27.50	26.04	22.36
$(\mathcal{W}, \neg R_L, R_B)$	34.89	27.95	28.85	29.57	24.61	24.70	26.46	24.70	21.98	28.51	21.06	23.28	27.87	26.18	22.62
$(\mathcal{W}, R_L, \neg R_B)$	34.84	27.96	28.94	29.51	24.62	24.74	26.35	24.70	21.93	28.42	21.06	23.28	27.87	26.19	22.46
$(\mathcal{W}, R_L, R_B)$	34.80	28.00	28.93	29.54	24.64	24.69	26.33	24.70	21.93	28.43	21.07	23.30	27.90	26.20	22.47

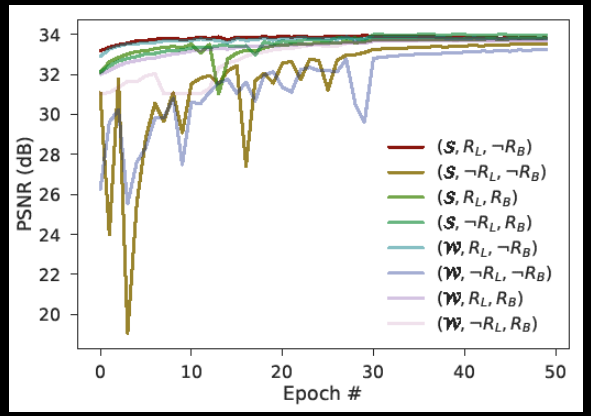
$t_{psnr} = 1.91, p_{psnr} = 0.07$

$t_{ssim} = 1.02, p_{ssim} = 0.31$

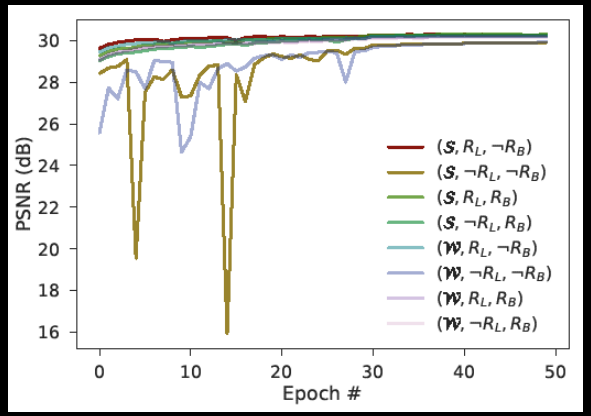
$\overline{PSNR_{\mathcal{W}}} = 26.15 \mid \overline{PSNR_{\mathcal{S}}} = 26.09$

$\overline{SSIM_{\mathcal{W}}} = 0.798 \mid \overline{SSIM_{\mathcal{S}}} = 0.797$

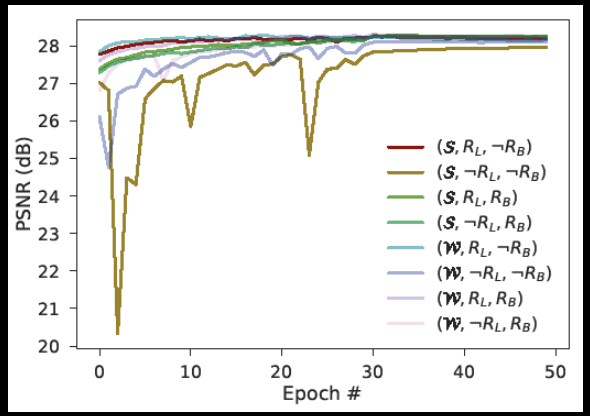
Stability in training



Scale=2

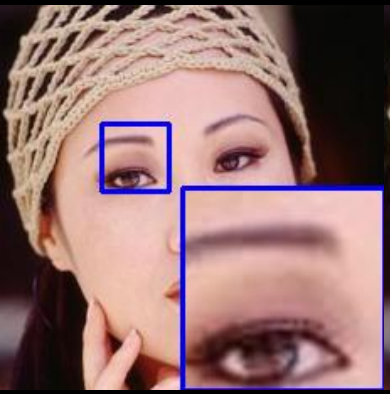


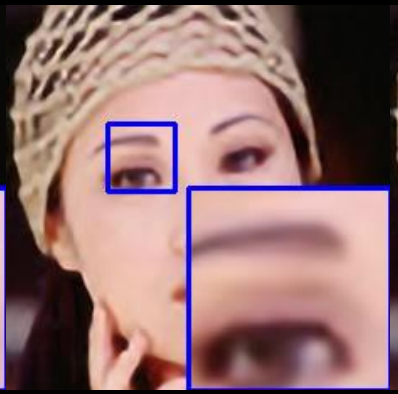
Scale=3

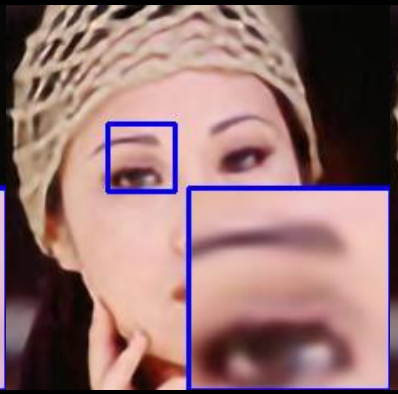


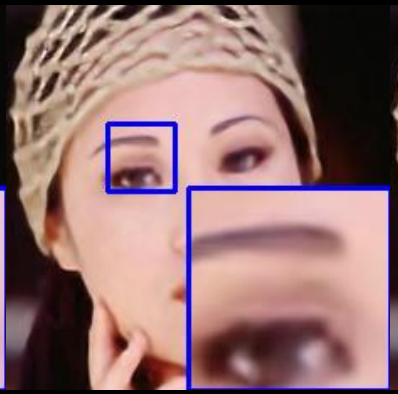
Scale=4

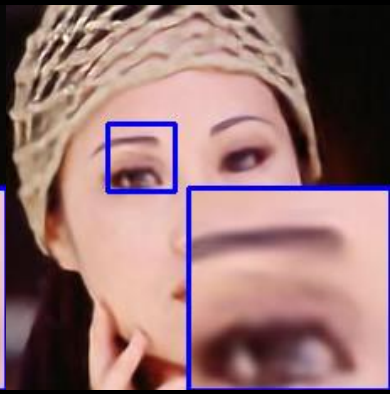
Qualitative Results

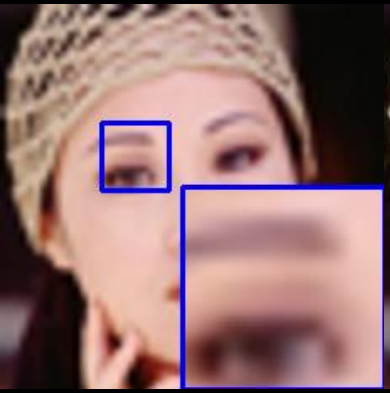

Reference | PSNR

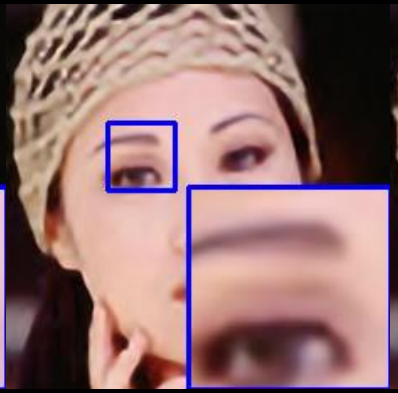

$(\mathcal{S}, \neg R_L, \neg R_B)$ | 27.34

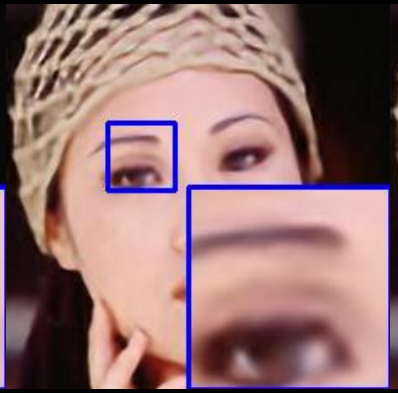

$(\mathcal{W}, \neg R_L, \neg R_B)$ | 27.44

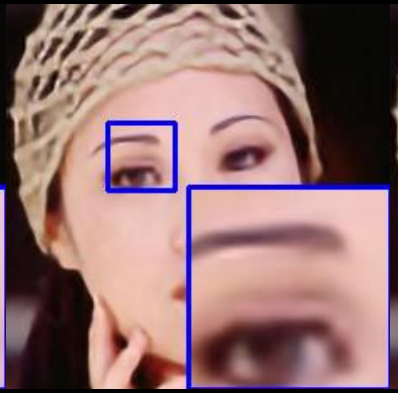

$(\mathcal{S}, R_L, \neg R_B)$ | 27.60

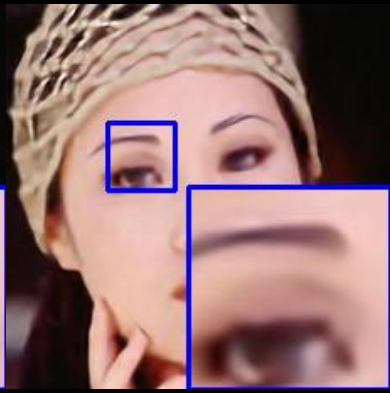

$(\mathcal{W}, R_L, \neg R_B)$ | 27.50


Bicubic | 25.11

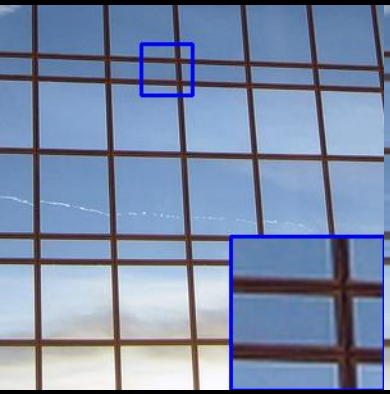

$(\mathcal{S}, \neg R_L, R_B)$ | 27.72


$(\mathcal{W}, \neg R_L, R_B)$ | 27.81

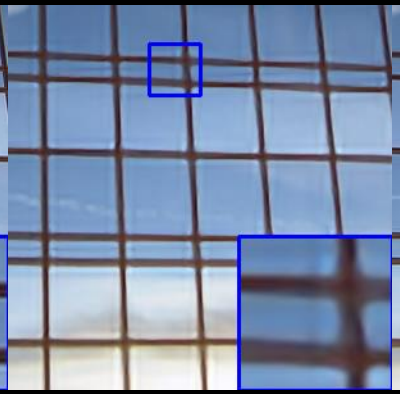

$(\mathcal{S}, R_L, R_B)$ | 27.67


$(\mathcal{W}, R_L, R_B)$ | 27.71

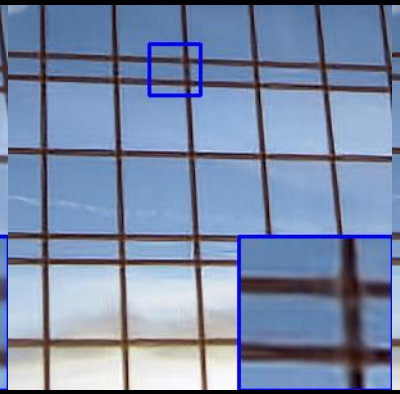
Qualitative Results



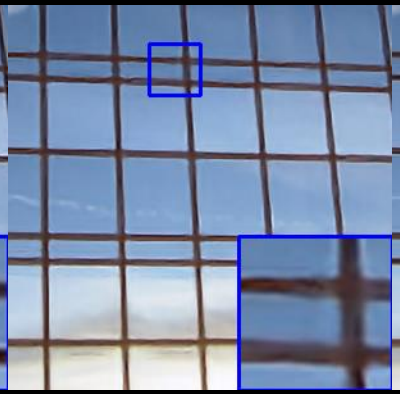
Reference | PSNR



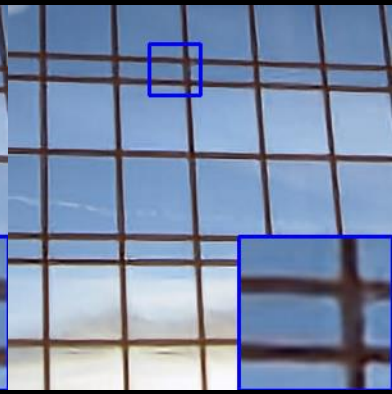
$(\mathcal{S}, \neg R_L, \neg R_B)$ | 26.09



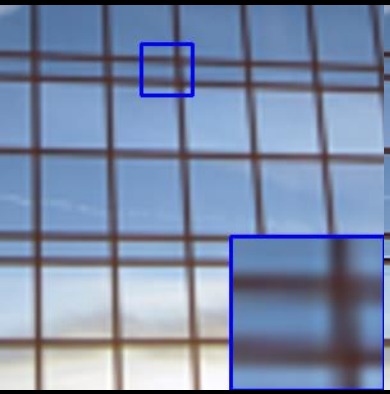
$(\mathcal{W}, \neg R_L, \neg R_B)$ | 26.60



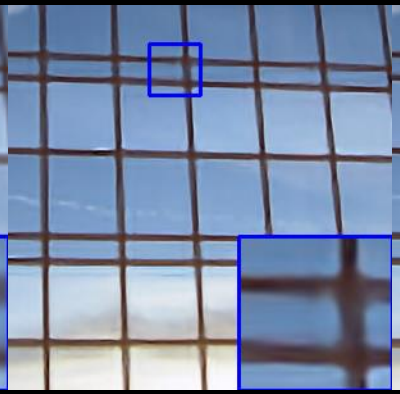
$(\mathcal{S}, R_L, \neg R_B)$ | 26.93



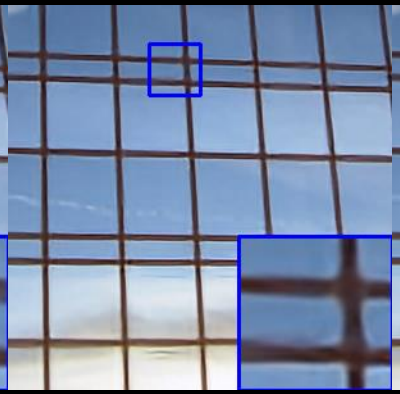
$(\mathcal{W}, R_L, \neg R_B)$ | 27.60



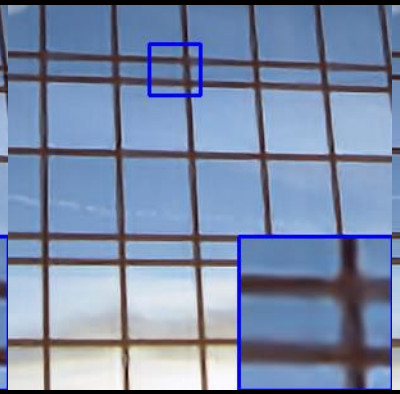
Bicubic | 23.25



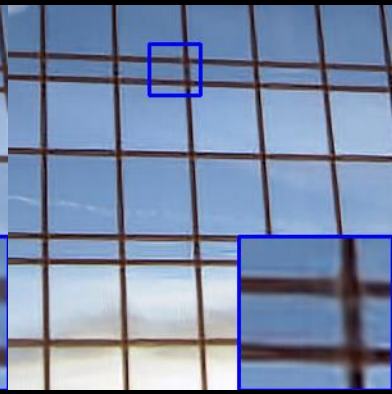
$(\mathcal{S}, \neg R_L, R_B)$ | 26.68



$(\mathcal{W}, \neg R_L, R_B)$ | 27.44



$(\mathcal{S}, R_L, R_B)$ | 27.53



$(\mathcal{W}, R_L, R_B)$ | 26.75

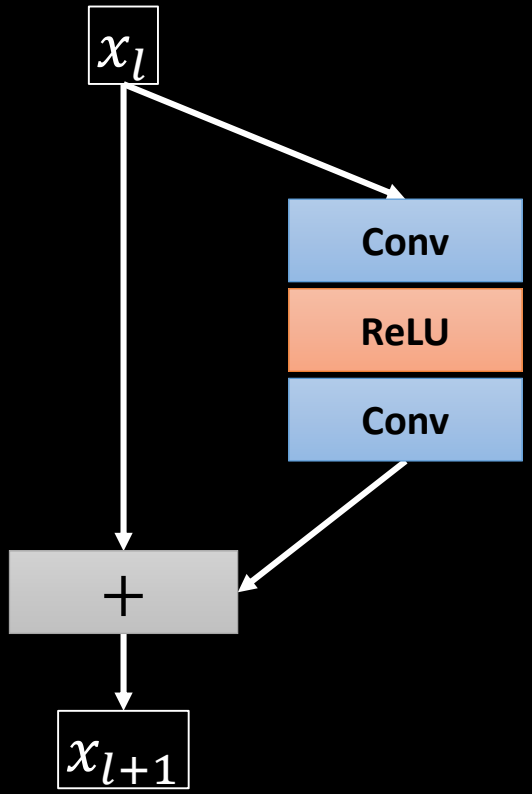
Runtime performance (1024 x 1024)

	Memory
$(\mathcal{S}, \neg R_L, \neg R_B)$	5412MB
$(\mathcal{S}, \neg R_L, R_B)$	5412MB
$(\mathcal{S}, R_L, \neg R_B)$	5432MB
$(\mathcal{S}, R_L, R_B)$	5432MB
$(\mathcal{W}, \neg R_L, \neg R_B)$	**1380MB**
$(\mathcal{W}, \neg R_L, R_B)$	**1380MB**
$(\mathcal{W}, R_L, \neg R_B)$	**1460MB**
$(\mathcal{W}, R_L, R_B)$	**1460MB**

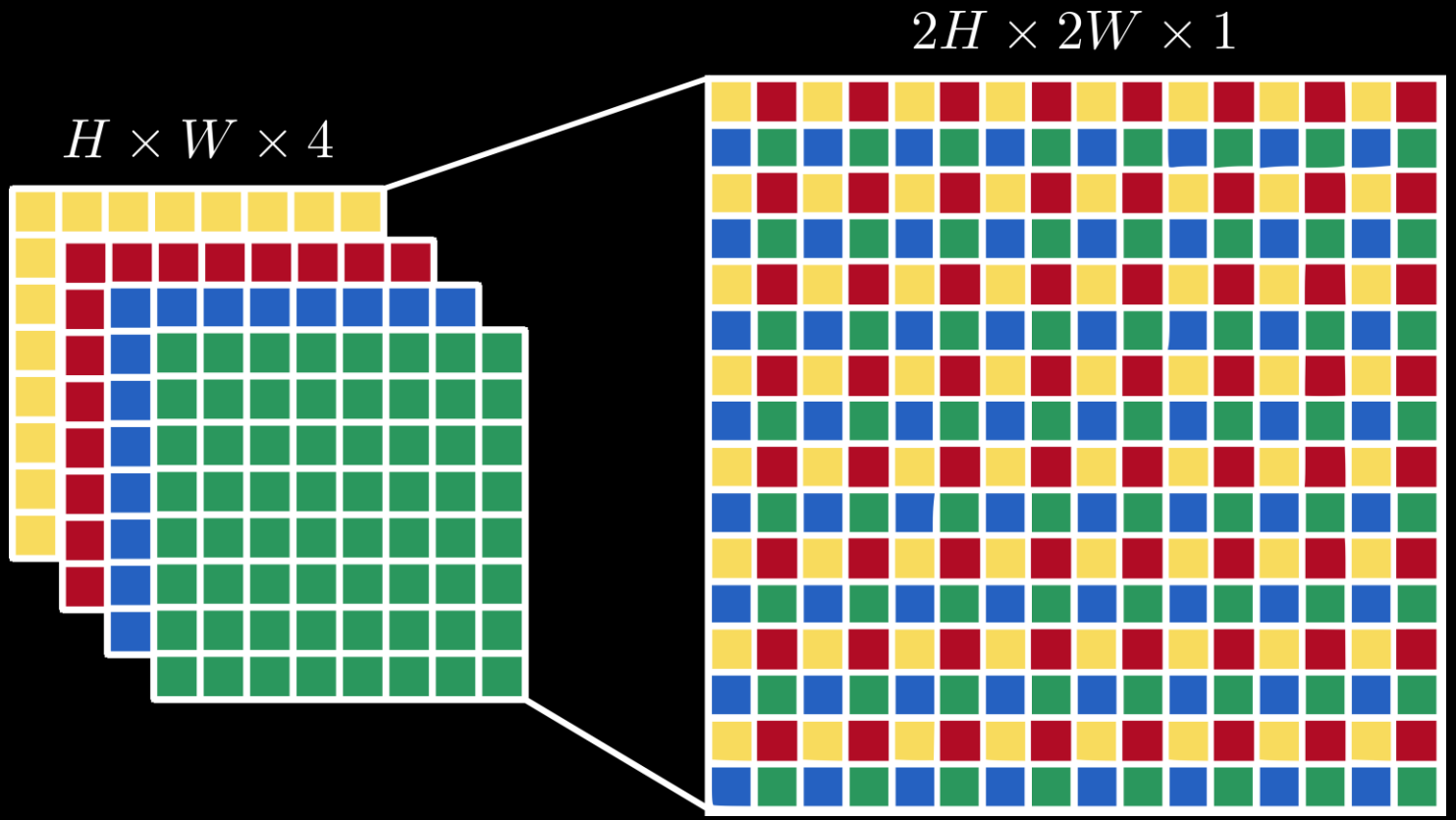
H x W

H/2 x W/2 x 4

Residual Connection Structure

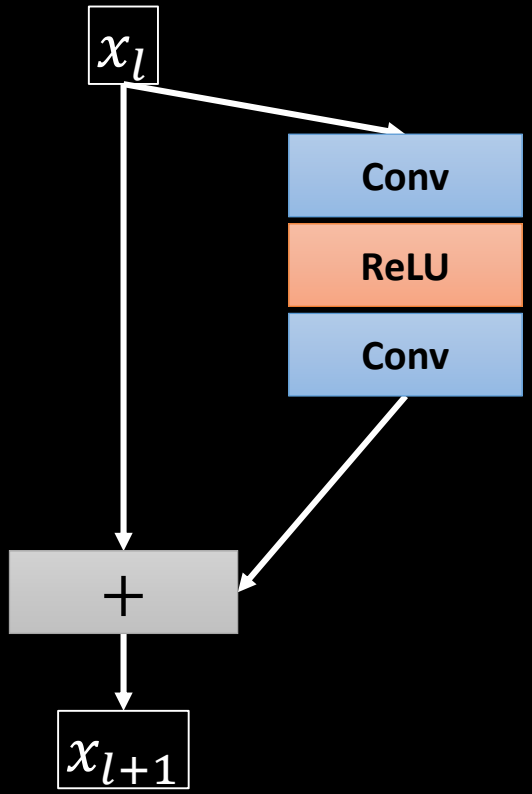


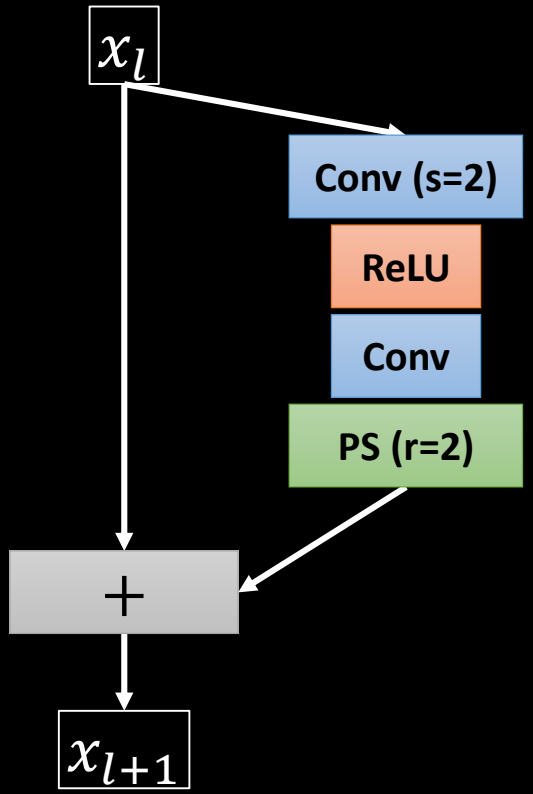
PixelShuffle (PS)



Jiwon Kim et al. Accurate image super-resolution using very deep convolutional networks. CVPR 2016

Proposed Residual Connection





ShuffleNet vs $(R_L, \neg R_B)$

- 5450MB -> 3500MB for spatial input
- 1500MB -> 900 MB for wavelet input

1024 x 1024 Image

ShuffleNet vs $(R_L, \neg R_B)$

- 5450MB -> 3500MB for spatial input
- 1500MB -> 900 MB for wavelet input

1024 x 1024 Image

Set	Set5			Set14			BSDS100			Urban100			Manga109		
Scale	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4	x2	x3	x4
$(\mathcal{S}, R_L, \neg R_B)$	34.99	28.02	28.89	29.62	24.66	24.66	25.89	24.73	23.17	28.67	21.14	23.22	27.38	26.31	22.45
($\mathcal{S}$, ShuffleNet)	34.92	27.94	28.74	29.79	24.61	24.54	25.80	24.77	22.85	28.55	21.03	23.07	27.34	26.18	22.40
$(\mathcal{W}, R_L, \neg R_B)$	34.84	27.96	28.94	29.51	24.62	24.74	26.35	24.70	21.93	28.42	21.06	23.28	27.87	26.19	22.46
($\mathcal{W}$, ShuffleNet)	34.92	27.89	28.69	29.51	24.56	24.57	26.18	24.67	22.04	27.82	20.99	23.09	27.71	26.18	22.47

$\overline{PSNR_{\mathcal{S}}}\ (R_L, \neg R_B) = 26.25 \mid \overline{PSNR_{\mathcal{S}}}(\text{ShuffleNet}) = 26.18$

$\overline{PSNR_{\mathcal{W}}}(R_L, \neg R_B) = 26.19 \mid \overline{PSNR_{\mathcal{W}}}(\text{ShuffleNet}) = 26.09$

Qualitative Comparison

($\mathcal{S}$, ShuffleNet)

($\mathcal{S}$, R_L, $\neg R_B$)

($\mathcal{W}$, ShuffleNet)

($\mathcal{W}$, R_L, $\neg R_B$)

Conclusion

- Residuals improve training speed and network performance
- No performance impact between spatial and wavelets inputs
- Wavelets reduce memory requirements