Dense ByNet: Residual Dense Network for Image Super Resolution

Oct 10, 2018

Jiu Xu, Yeongnam Chae, Björn Stenger, Ankur Datta

Rakuten Institute of Technology
Rakuten, Inc.

Image Super-resolution

- **Definition**
  - Given a single low-resolution image as input, the image super-resolution is aiming for transforming it into a high-resolution image.

Big, but Blurry

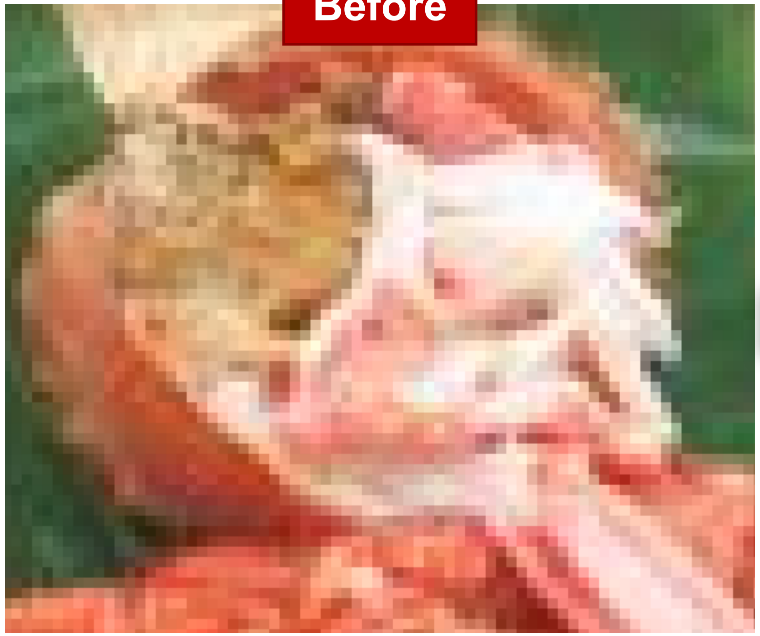
Big & Sharp

Image Super-resolution

● Conventional Approach

- Traditional up-sampling techniques “**blow up**” images by filling in new pixel areas with data from existing pixel values nearby. These methods are fast, but they aren’t the best at bringing out details in the enlarged images.

Before



After

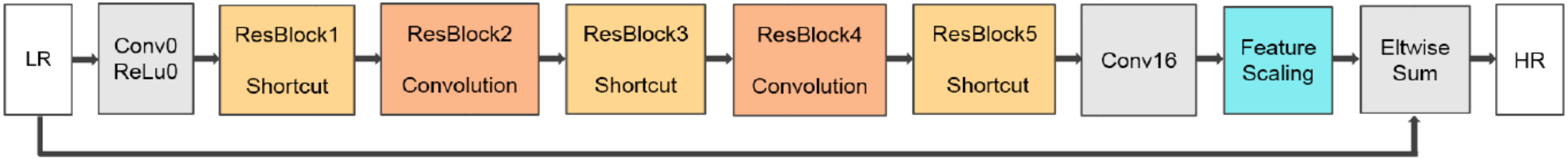
No Sharp Details

Conventional Method

3

ByNet: Image Super Resolution With A Bypass Connection Network

• *ByNet-SR* in ICIP2017 [9]



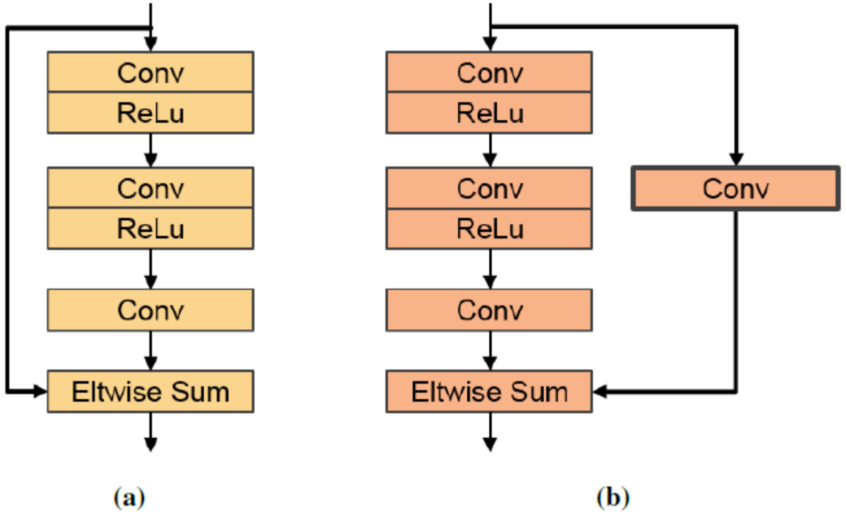


Fig. 2: New residual blocks used in ByNet: (a) feature bypass with shortcut connection **(b)** feature bypass with convolutional connection.

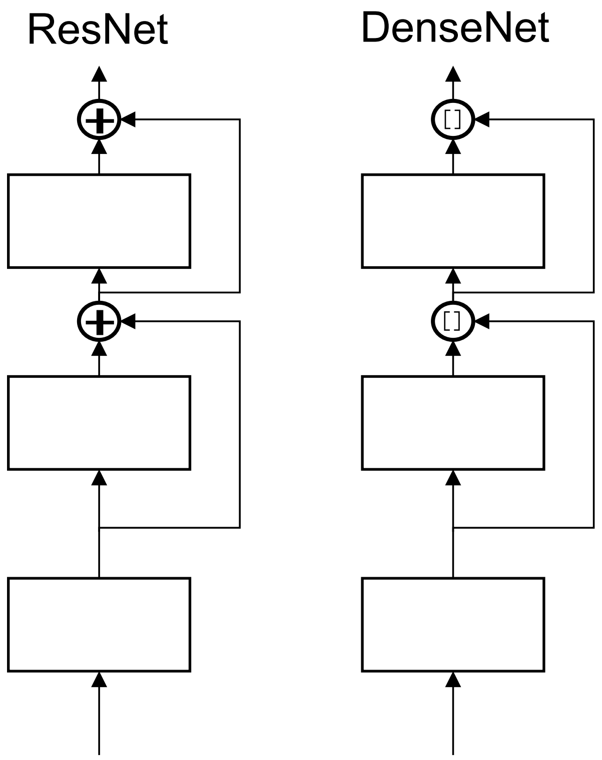
Feature bypass with shortcut connection
improves convergence properties and achieve higher accuracy within the same training epoch

Feature bypass with convolutional connection
improves the model robustness in terms of multi-scale capability

Feature scaling
improves convergence by scaling the network output to efficiently fit the distribution of image residuals.

[9] J. Xu, Y. Chae, and B. Stenger, "ByNet-SR: Image super resolution with a bypass connection network," in *ICIP*, 2017.

● From *ResNet* to *DenseNet*



From ResNet to DenseNet[1]

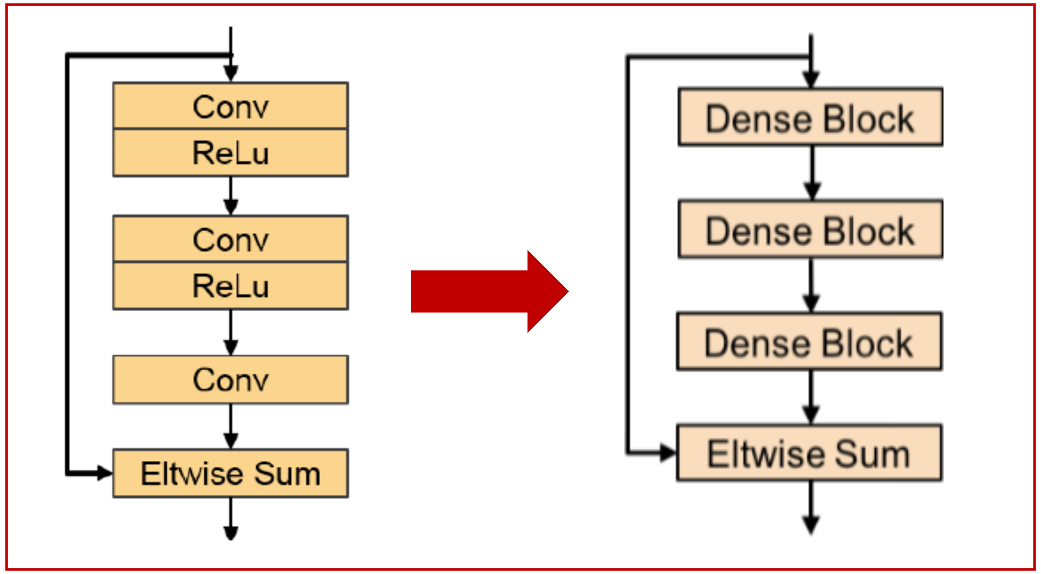
- *DenseNet*[1] introduced densely connected layers, where feature maps are channel-wise concatenated instead of element-wise summed.

- Features from all preceding layers are input into subsequent layers.

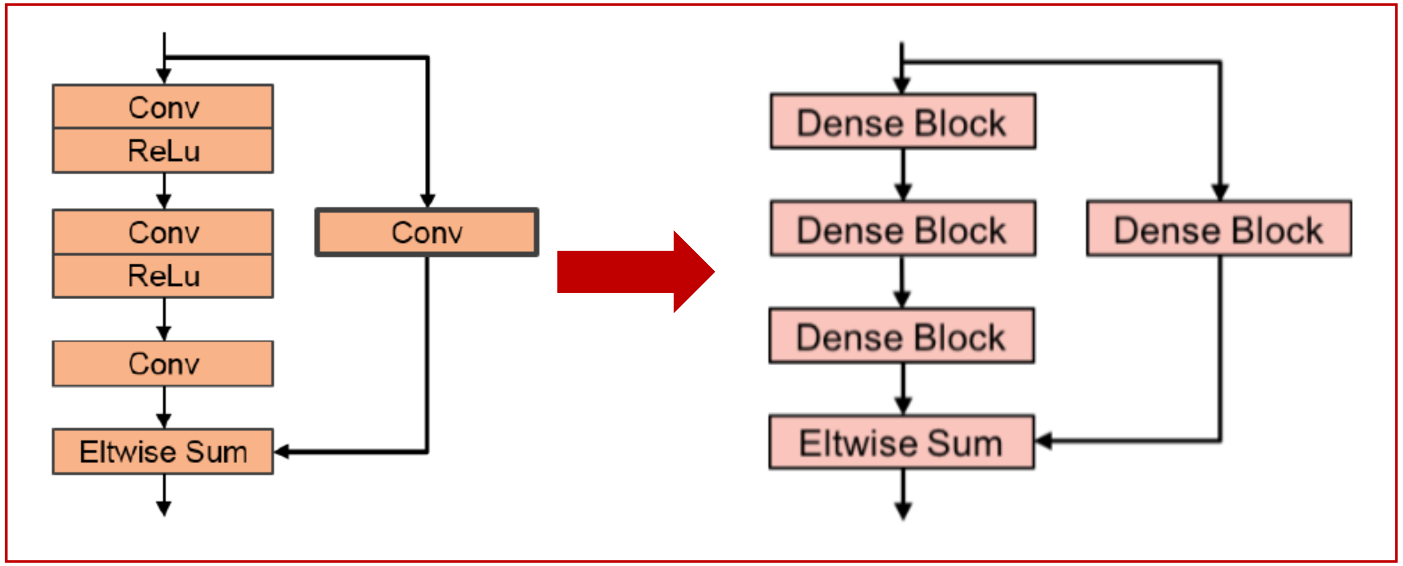
- *DenseNet* significantly reduces the total number of parameters by encouraging feature reuse throughout the network. It showed good performance on the ImageNet classification task.

[1] G. Huang, Z. Liu, L. Van der Maaten, and K. Weinberger, "Densely connected convolutional networks," in *CVPR*, 2017.

- **From ByNet to Dense ByNet**

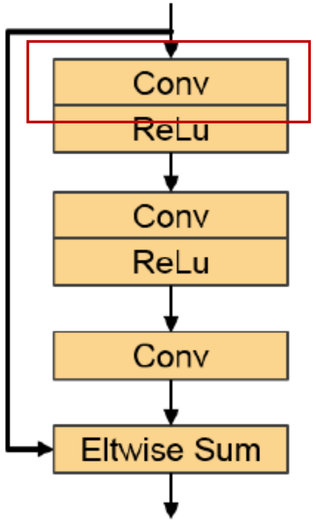


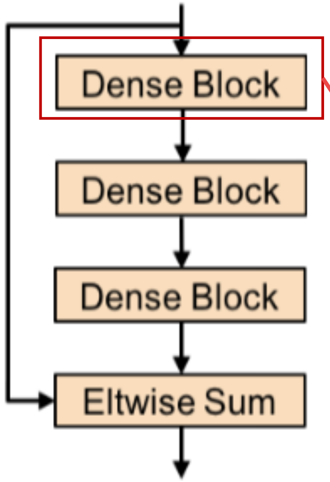
(a) feature bypass with dense skip connection

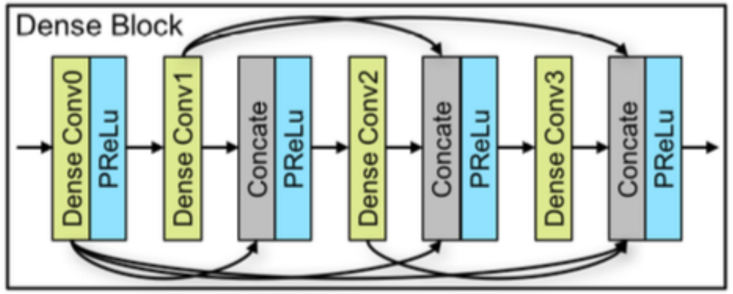


(b) feature bypass with dense convolutional connection

- **From ByNet to Dense ByNet**







The convolution layers in the residual blocks feed input 64-channel feature into a convolutional layer with a 3 × 3 filter with 64 channels.
Thus, each layer has a total of 3×3×64×64 = **36,864** parameters.

Table 1: Number of parameters in a dense connection block.

layer	kernel	plane-in	plane-out	#params
Dense Conv0	3x3	64	16	9,216
Dense Conv1	3x3	16	16	2,304
Dense Conv2	3x3	32	16	4,608
Dense Conv3	3x3	48	16	6,912
Total				23,040

From ByNet to Dense ByNet

Dilated convolutions: In order to better take advantage of spatial context we apply a 2-dilated convolution to every convolutional layer in the dense bypass blocks. This allows the features to be computed over receptive fields of multiple sizes.

Activation function: Replacing the residual blocks with dense connection blocks increases the number of convolution layers and attached activation function layers. The increased number of activation functions exacerbates the vanishing gradient problem, particularly for dense connections. To address this issue, we replace the rectified linear unit (ReLU) activation functions with parametric ReLU (PReLU)

Loss function: Charbonnier loss function is used instead of MSE. Compared to the MSE loss, this function penalizes outliers less, and is therefore better suited for textured image regions.

$$l_C(\widehat{Y}, Y) = \left((\widehat{Y} - Y)^2 + \epsilon^2\right)^{\frac{1}{2}}$$

- **From ByNet to Dense ByNet**

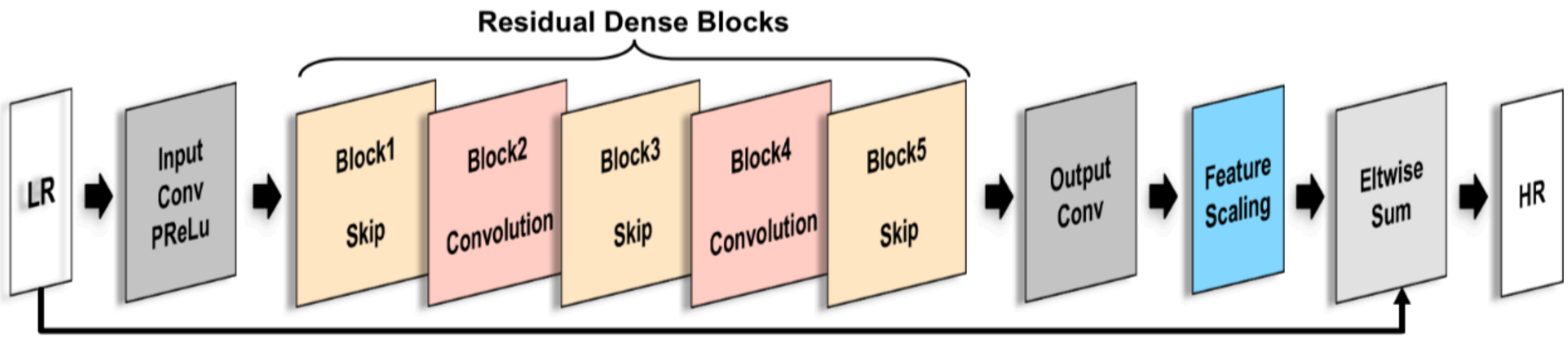
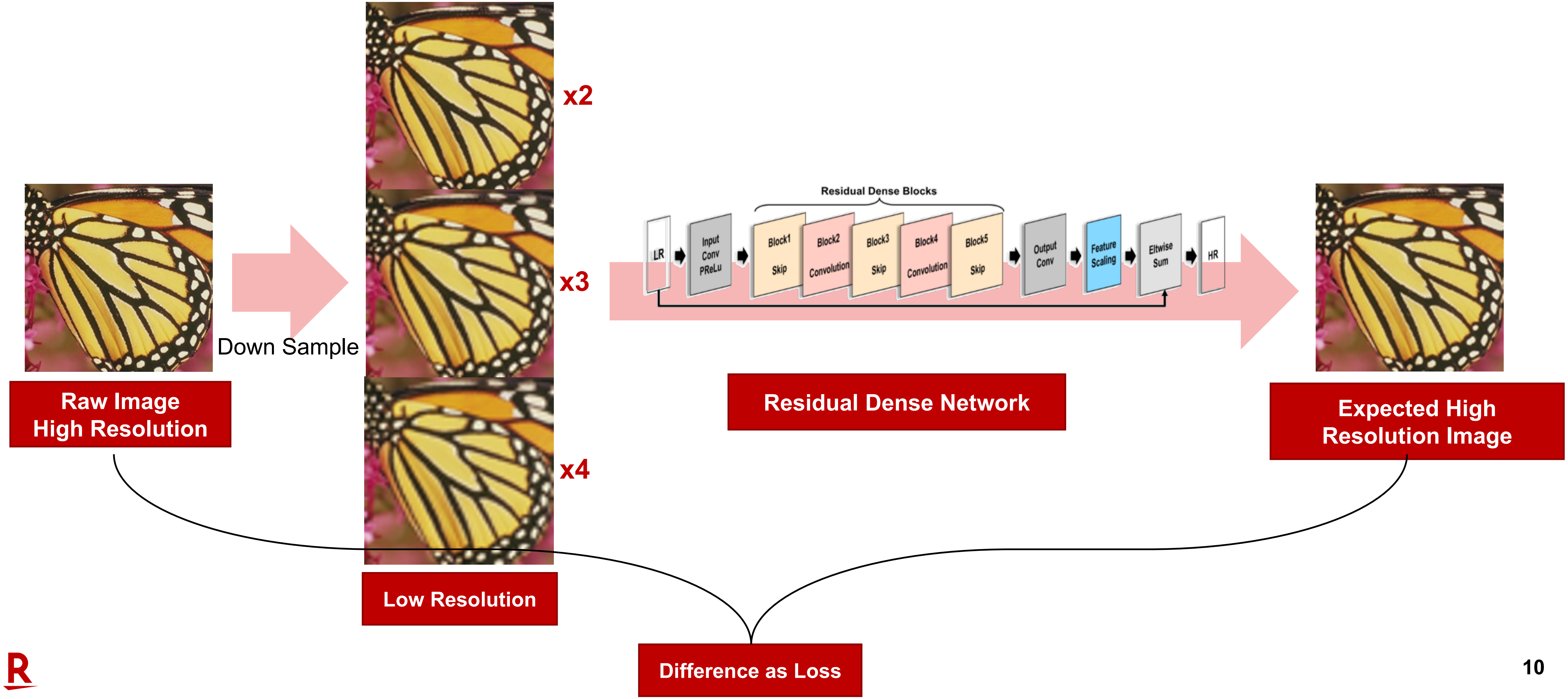


Fig. 3: Network architecture of Dense ByNet5. The proposed CNN learns the non-linear mapping between LR/HR images. The key elements of the architecture are two types of residual dense blocks, which are alternated in sequence.

Dense ByNet: Residual Dense Network for Image Super Resolution

- **Training**



• Comparison with State of the Art

- In terms of efficiency, Dense ByNet5 contains *393K* network parameters, and Dense ByNet7 contains *554K*. Both are smaller than ByNet5 with *628K*, VDSR with *665K*, or LapSRN with *812K* parameters.

Table 2: Performance comparison on benchmark datasets: *PSNR and SSIM are averaged over all images for each scale. The proposed Dense ByNet model consistently achieves the best PSNR and SSIM results.*

Dataset	Scale	Bicubic PSNR	VDSR[13] PSNR	LapSRN[23] PSNR	ByNet5[9] PSNR	Dense ByNet5 PSNR	Dense ByNet7 PSNR	Bicubic SSIM	VDSR[13] SSIM	LapSRN[23] SSIM	ByNet5[9] SSIM	Dense ByNet5 SSIM	Dense ByNet7 SSIM
Set5	×2	33.66	37.53	37.48	37.61	37.70	**37.76**	0.9299	0.9587	0.9591	0.9597	0.9600	**0.9605**
	×3	30.39	33.66	-	33.79	34.03	**34.09**	0.8682	0.9213	-	0.9235	0.9254	**0.9261**
	×4	28.42	31.35	31.65	31.44	31.66	**31.74**	0.8104	0.8838	0.8889	0.8860	0.8891	**0.8902**
Set14	×2	30.24	33.13	33.08	33.20	33.25	**33.34**	0.9056	0.9124	0.9127	0.9143	0.9147	**0.9153**
	×3	27.55	29.92	-	29.95	30.04	**30.07**	0.8188	0.8314	-	0.8348	0.8363	**0.8370**
	×4	26.00	28.20	28.26	28.18	28.27	**28.29**	0.7491	0.7674	0.7735	0.7711	0.7738	**0.7742**
BSD100	×2	29.56	31.92	31.85	31.93	31.99	**32.04**	0.8431	0.8960	0.8948	0.8970	0.8978	**0.8984**
	×3	27.21	28.86	-	28.86	28.92	**28.96**	0.7385	0.7976	-	0.7994	0.8010	**0.8017**
	×4	25.96	27.31	27.36	27.30	27.37	**27.40**	0.6675	0.7251	0.7291	0.7269	0.7293	**0.7301**

[10] J. Kim, J.K. Lee, and K.M. Lee, "Accurate image super resolution using very deep convolutional networks," in CVPR, 2016.
[23] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in CVPR, 2017

● Ablation Study

We evaluate the contribution of the different components in terms of PSNR performance on the Set5 dataset.

- The baseline model is ByNet5.
- We separately and jointly include
  1. Dilated convolutions
  2. Parametric ReLU (*PReLU*)
  3. Charbonnier loss
  4. Dense blocks

Table 3: Ablation Study. We evaluate the contribution of the different components in terms of PSNR performance on the *Set5* dataset. The baseline model is ByNet5 [9]. We separately and jointly include dilated convolutions, parametric ReLU (PReLU), Charbonnier loss, and dense blocks.

Dilated Conv	PReLU	Charb Loss	Dense Block	PSNR 2×	PSNR 3×	PSNR 4×
				37.61	33.79	31.44
✓				37.62	33.87	31.57
	✓			37.60	33.80	31.45
		✓		37.70	33.81	31.45
			✓	37.60	33.76	31.43
✓	✓			37.63	33.92	31.56
✓		✓		37.74	33.97	31.59
✓			✓	37.55	33.89	31.56
	✓	✓		37.71	33.81	31.45
	✓		✓	37.70	33.88	31.52
		✓	✓	37.63	33.72	31.35
✓	✓	✓		37.72	33.93	31.56
✓	✓		✓	37.63	33.90	31.59
✓		✓	✓	37.58	33.89	31.56
	✓	✓	✓	**37.77**	33.89	31.56
✓	✓	✓	✓	37.70	**34.03**	**31.66**

- ## Results

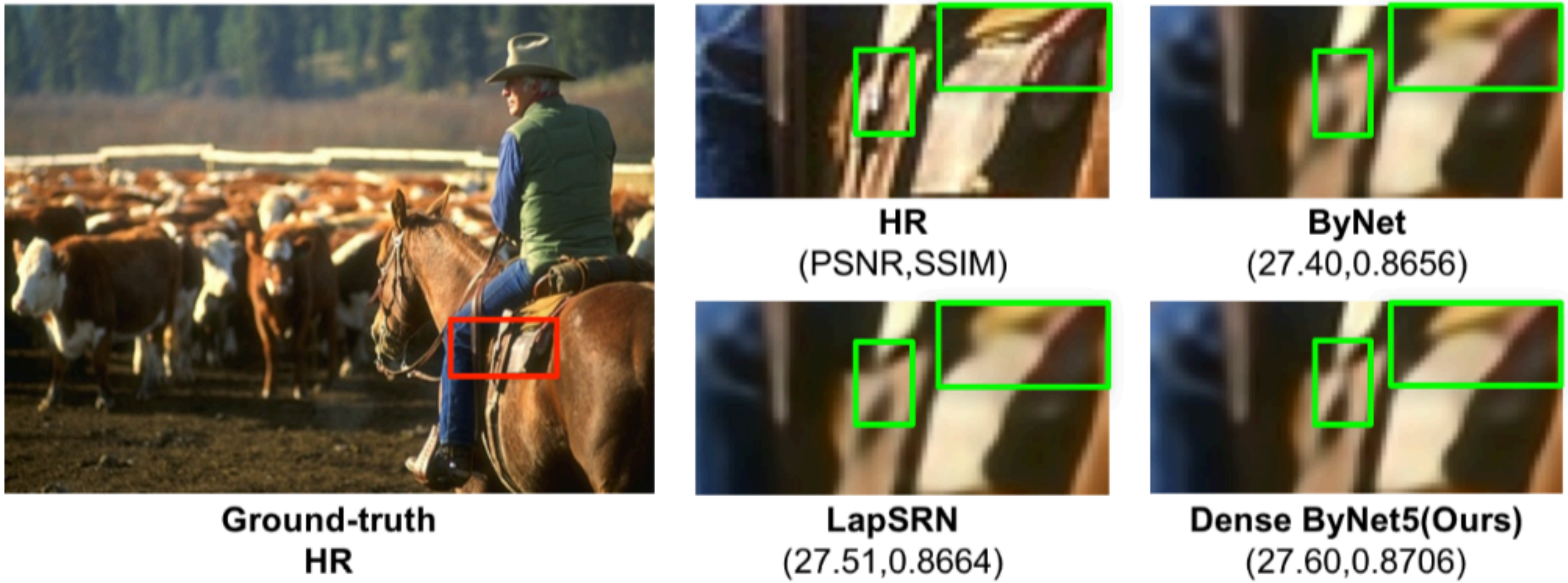


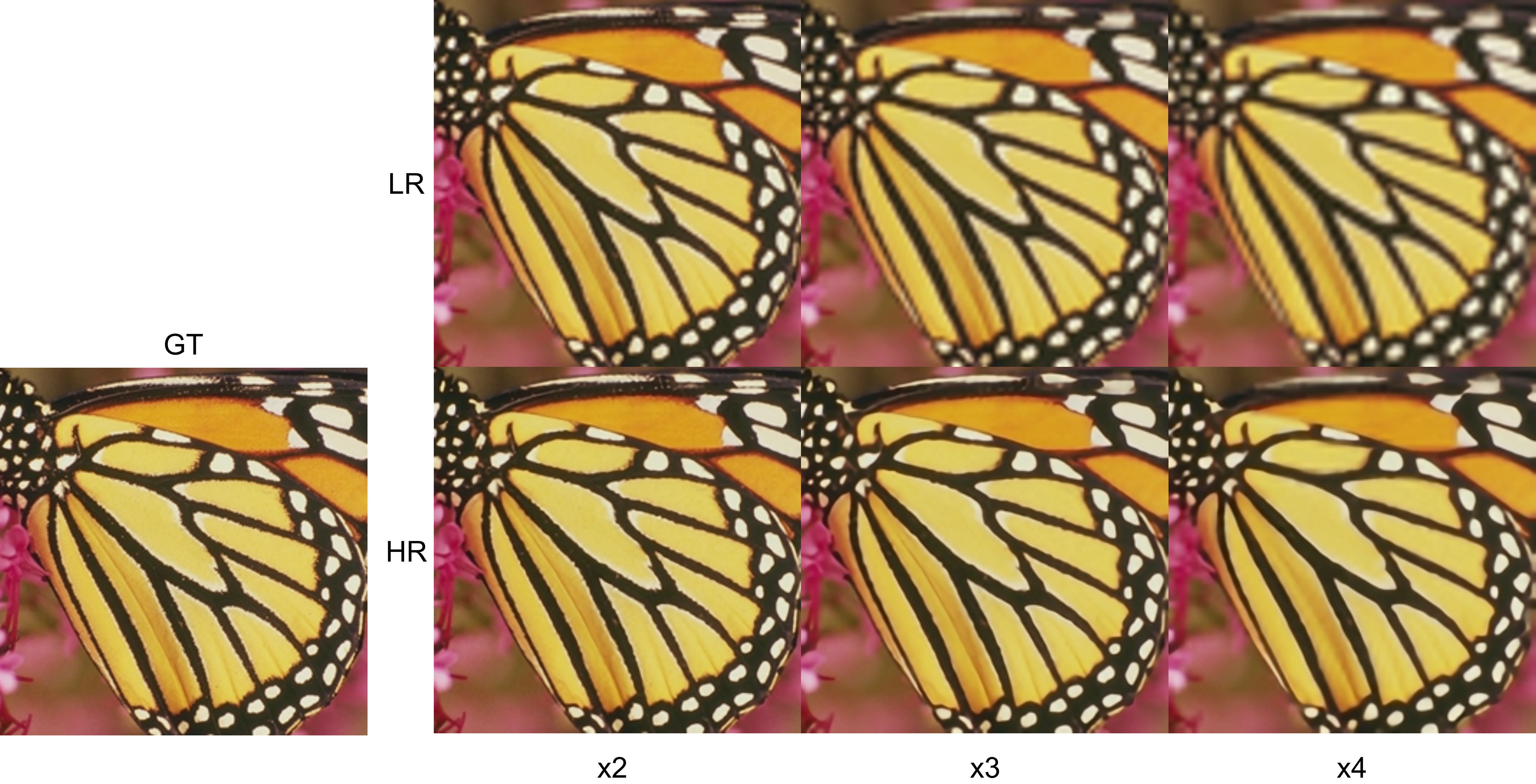
Fig. 1: Example result. *Visual comparison for* $\times 4$ *SR on an example from the* BSD100 *[11] dataset. Regions in green boxes highlight improved recovery of detail.*

GT

LR

HR

x2 x3 x4


GT
LR
HR
x2
x3
x4

Rakuten
Institute of Technology