Supplemental Material:
Learning A Single Network for Scale-Arbitrary Super-Resolution

Longguang Wang, Yingqian Wang, Zaiping Lin, Jungang Yang, Wei An, Yulan Guo*
National University of Defense Technology
{wanglongguang15,yulan.guo}@nudt.edu.cn

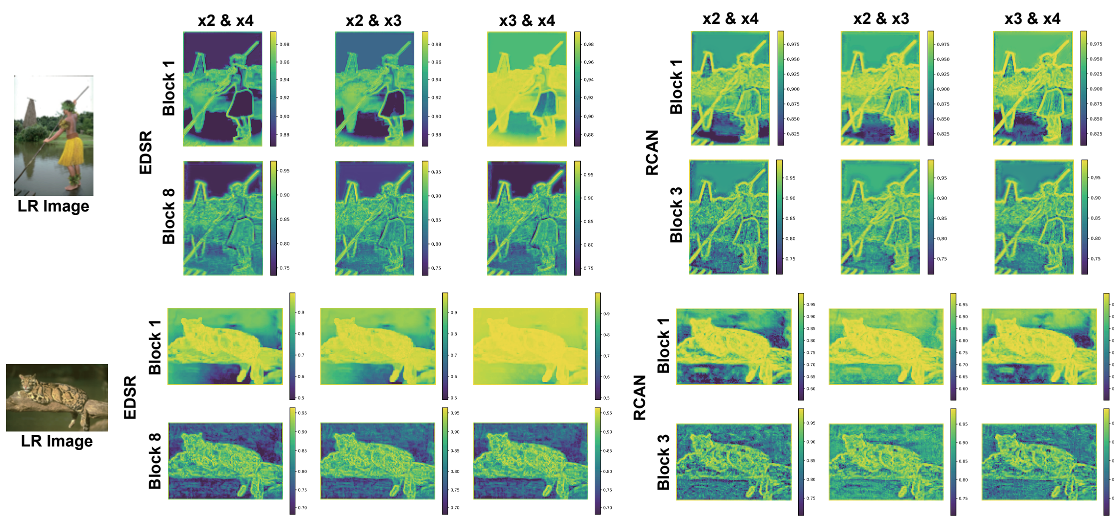
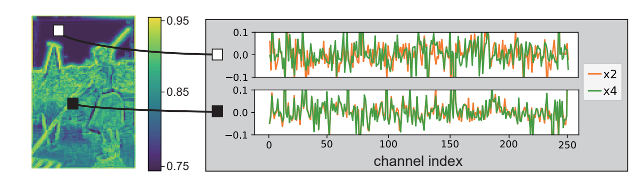


Figure I. Visualization of feature similarity maps between $\times 2/3/4$ SR networks.

Section I includes more examples of feature similarity with additional discussions. Section II describes the pixel shuffling layer and the proposed scale-aware upsampling layer. Then, additional analyses with respect to our networks are presented in Section III. Finally, Section IV provides additional quantitative and qualitative results.

I. Feature Similarity

It has been demonstrated in several SR methods developed for multiple degradations [1, 2, 3, 4] that features in the network vary for different degradations. Intuitively, features learned for various scale factors are also different since their bicubic degradations are different [5, 6]. To demonstrate this, we show feature similarity maps achieved by $\times 2/3/4$ EDSR and RCAN on different images in Fig. I. It can be observed that features learned for $\times 2/3/4$ SR are different, with mean similarity being 0.77. Moreover, feature similarity varies for different blocks and regions. Specifically, regions with strong textures (*e.g.*, the cheetah) usually have higher similarities than those flat regions (*e.g.*, the lake). We further compare features within regions of different similarities in Fig. II. As we can see, features within flat regions



(a) feature similarity between x2/x4 models (b) normalized features

Figure II. Comparison of features within different regions.

changes more significantly for different scale factors than those in edge regions. This further demonstrates the variations of features learned for different scale factors.

Motivated by these observations, we propose to learn a guidance map to perform pixel-wise feature adaption accordingly. Although early attempts [7, 8, 9] show it is feasible to use shared features to handle multiple scale factors, these methods suffer inferior performance since the difference among features learned for various scale factors is not considered [10]. It is illustrated in Section 4.3 (Model 4 vs. Model 1 in Table 1) that our network benefits from scale-aware feature adaption to produce better results. This fur-

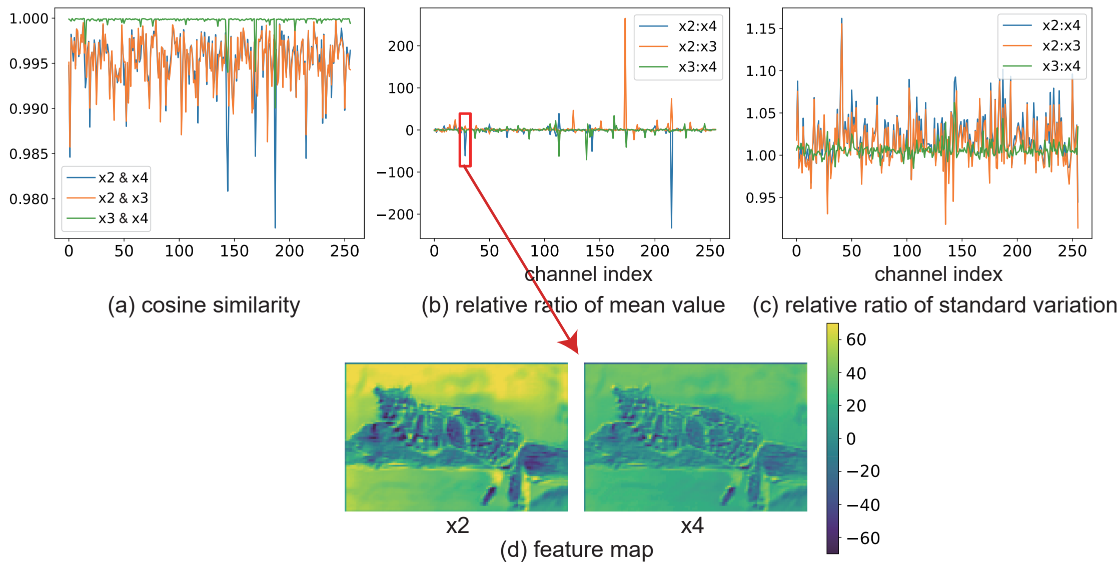


Figure III. Statistics of filters learned for different scale factors.

ther validates the difference of features learned for various scale factors.

We also conduct experiments to study the difference of feature similarities among various regions. It is observed in [3] that filters of restoration models trained with different restoration levels are similar at visual patterns but have different statistics (*e.g.*, mean value). Inspired by this observation, we further investigate the statistics of the filters in $\times 2/3/4$ SR models. It can be observed in Fig. III that filters in $\times 2/3/4$ SR models share a similar observation with [3]. Specifically, filters learned for different scale factors have high cosine similarities (Fig. III(a)) with quite different mean values (Fig. III(b)). That is, there is a mean shift of convolutional filters learned for different scale factors. For flat regions, neighboring pixels have the same sign such that this mean shift is accumulated (Fig. III(d)). In contrast, since neighboring pixels in edge regions usually have opposite signs, the effect of mean shift is weakened. Consequently, features within flat regions changes more significantly than edge regions. We suppose that since LR images used to train $\times 2$ SR networks contain much more low-frequency components (Fig. IV), $\times 2$ models learn higher response to highlight these components in the network.

II. Scale-Aware Upsampling Layer

We illustrate the pixel shuffling layer and our scale-aware upsampling layer in Fig. V. It can be observed from Fig. V(a) and (b) that the pixel shuffling layer can be considered as a two-step pipeline, which consists of a sampling step and a spatially-varying filtering step. Therefore, we generalize the pixel shuffling layer to a scale-aware upsampling layer, as shown in Figs. V(c) and (d). Specifically, for

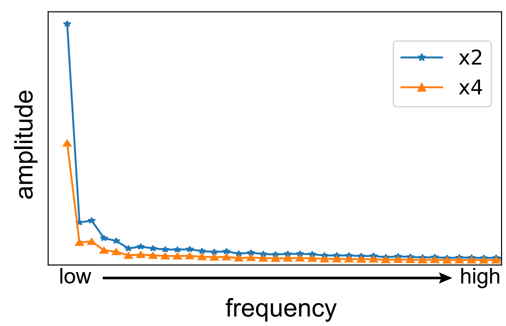


Figure IV. Comparison of components with different frequencies in the training dataset for $\times 2/4$ SR. Results are averaged over 50 LR images.

location (x, y) in HR space, its coordinates ($L(x)$ and $L(y)$) and relative distances ($R(x)$ and $R(y)$) in LR space are first calculated, as shown in Fig. V(c). Then, horizontal and vertical scale factors (r_h and r_v), $R(x)$ and $R(y)$ are used to produce offsets (δ_x and δ_y) and a pair of $k \times k$ convolutional kernels. Finally, a $k \times k$ neighborhood centered at $(L(x)+\delta_x, L(y)+\delta_y)$ is sampled and the predicted kernels are used to generate the output feature at location (x, y), as shown in Fig. V(d).

III. Additional Analyses

III.I. Different Settings for Non-Integer SR

To perform SR with non-integer scale factors (*e.g.*, $\times 2.5$) on an LR image (*e.g.*, 60×60) using baseline networks like RCAN, we have four different settings:

- **Setting 1**: Bicubicly downscale the LR image to the

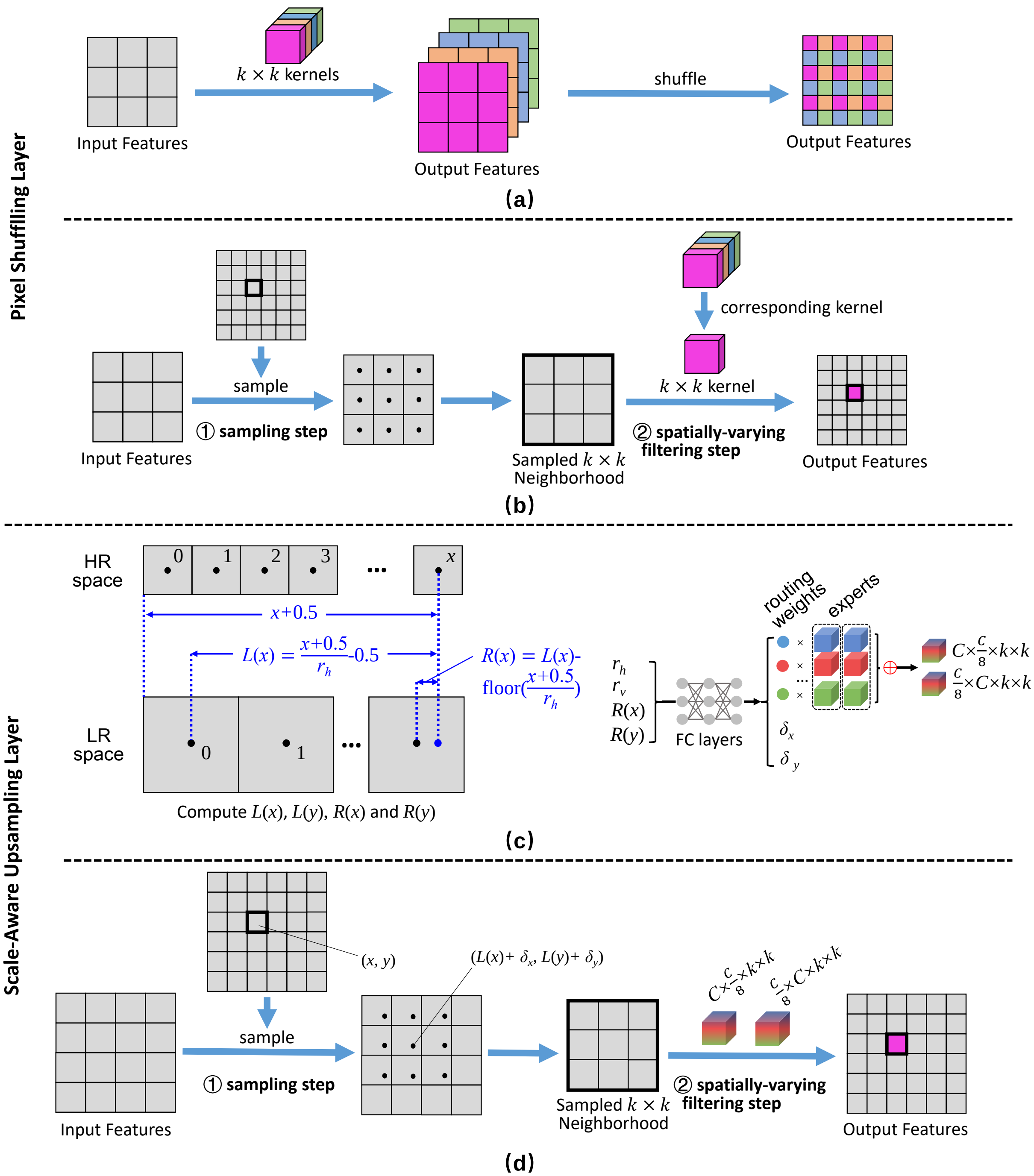


Figure V. An illustration of the pixel shuffling layer and our scale-aware upsampling layer.

size of 50×50 and then super-resolve the result for $\times 3$ SR.

- **Setting 2**: Bicubicly upscale the LR image to the size of 75×75 and then super-resolve the result for $\times 2$ SR. Note that, this setting degrades to bicubic interpolation for scale factors lower than 2 (*e.g.*, $\times 1.55$).

- **Setting 3**: Super-resolve the LR image for $\times 2$ SR and then bicubicly upscale the result (120×120) to the size of 150×150. Note that, this setting also degrades to bicubic interpolation for scale factors lower than 2 (*e.g.*, $\times 1.55$).

- **Setting 4**: Super-resolve the LR image for $\times 3$ SR and

Table I. PSNR results achieved on Set5 and Set14 for non-integer scale factors.

	Set5									Set14								
	×1.55	×1.8	×2	×2.35	×2.5	×2.95	×3.1	×3.8	Avg	×1.55	×1.8	×2	×2.35	×2.5	×2.95	×3.1	×3.8	Avg
RCAN [11] (Setting 1)	36.24	36.40	**38.27**	32.80	33.44	31.66	29.92	30.76	33.69	32.32	32.45	**34.12**	29.17	29.60	28.12	27.04	26.85	29.96
RCAN [11] (Setting 2)	36.24	34.70	**38.27**	32.59	32.19	30.36	30.62	29.08	33.00	32.53	31.12	**34.12**	29.42	28.97	27.80	27.84	26.27	29.76
RCAN [11] (Setting 3)	36.24	34.70	**38.27**	36.26	35.83	34.21	34.38	32.88	35.35	32.53	31.12	**34.12**	32.15	31.65	30.17	30.26	28.85	31.36
RCAN [11] (Setting 4)	40.77	39.08	**38.27**	36.45	36.16	34.74	34.37	33.04	36.61	36.96	34.99	**34.12**	32.36	31.94	30.53	30.26	28.97	32.52
ArbRCAN (Ours)	**40.97**	**39.20**	38.26	**36.55**	**36.22**	**34.77**	**34.50**	**33.04**	**36.69**	**37.05**	**35.15**	34.09	**32.48**	**32.01**	**30.59**	**30.37**	**29.01**	**32.60**

Table II. PSNR results achieved on Set5 and Set14 for asymmetric scale factors.

	Set5									Set14								
	$\frac{\times 1.3}{\times 1.5}$	$\frac{\times 1.6}{\times 2.95}$	$\frac{\times 1.6}{\times 3.45}$	$\frac{\times 1.55}{\times 3.5}$	$\frac{\times 1.75}{\times 3.5}$	$\frac{\times 4}{\times 1.5}$	$\frac{\times 3.5}{\times 2}$	$\frac{\times 3}{\times 1.95}$	Avg	$\frac{\times 1.3}{\times 1.5}$	$\frac{\times 1.6}{\times 2.95}$	$\frac{\times 1.6}{\times 3.45}$	$\frac{\times 1.55}{\times 3.5}$	$\frac{\times 1.75}{\times 3.5}$	$\frac{\times 4}{\times 1.5}$	$\frac{\times 3.5}{\times 2}$	$\frac{\times 3}{\times 1.95}$	Avg
RCAN [11] (Setting 1)	41.75	35.62	34.55	34.52	34.38	34.28	34.94	35.73	35.72	38.17	32.13	31.13	31.17	30.95	30.40	30.72	31.69	32.04
RCAN [11] (Setting 2)	41.75	36.02	35.03	35.02	34.88	34.97	35.33	36.11	36.14	38.17	32.32	31.34	31.41	31.15	30.70	31.02	31.96	32.26
Meta-RCAN [12] (Setting 1)	38.02	32.03	31.15	31.11	30.91	30.97	31.32	32.11	32.20	34.26	29.45	28.68	28.62	28.47	27.96	28.16	28.88	29.31
Meta-RCAN [12] (Setting 2)	42.04	36.03	35.15	35.22	35.06	35.01	35.47	36.09	36.26	38.29	32.43	31.49	31.54	31.30	30.91	31.10	31.94	32.38
ArbRCAN (Ours)	**42.16**	**36.21**	**35.29**	**35.35**	**35.16**	**35.25**	**35.60**	**36.23**	**32.41**	**38.45**	**32.51**	**31.57**	**31.63**	**31.37**	**31.04**	**31.27**	**32.11**	**32.49**

then bicubicly downscale the result (180×180) to the size of 150×150.

The comparison of these four settings is presented in Table I. Compared to settings 1 and 2, using bicubic interpolation as a post-processing (settings 3 and 4) produces much higher PSNR values. Moreover, setting 4 achieves the best performance. Therefore, setting 4 is referred to as the default setting to achieve non-integer SR using baseline networks in the main text.

III.II. Different Settings for Asymmetric SR

To super-resolve an LR image (*e.g.*, 60×60) with asymmetric scale factors (*e.g.*, $\frac{\times 2.5}{\times 1.5}$) using baseline networks like RCAN, we have two different settings:

- **Setting 1**: Super-resolve the LR image for $\times 2$ SR ($\text{ceil}(1.5)$) and then bicubicly rescale the result (120×120) to the size of 150×90.
- **Setting 2**: Super-resolve the LR image for $\times 3$ SR ($\text{ceil}(2.5)$) and then bicubicly downscale the result (180×180) to the size of 150×90.

For Meta-SR networks like Meta-RCAN, we also have two different settings to perform SR with asymmetric scale factors (*e.g.*, $\frac{\times 2.5}{\times 1.5}$) on an LR image (*e.g.*, 60×60):

- **Setting 1**: Super-resolve the LR image for $\times 1.5$ SR and then bicubicly upscale the result (90×90) to the size of 150×90.
- **Setting 2**: Super-resolve the LR image for $\times 2.5$ SR and then bicubicly downscale the result (150×150) to the size of 150×90.

Note that, we only focus on settings using bicubic interpolation as a post-processing due to their superior performance (as demonstrated in Sec. III.I). The performance of RCAN and Meta-RCAN with different settings are compared in Table II. As we can see, setting 2 outperforms setting 1 for both RCAN and Meta-RCAN. Thus, we use setting 2 as the default setting to perform SR with asymmetric scale factors in the main text.

IV. Additional Results

IV.I. SR with Non-Integer Scale Factors

Quantitative Results. We compare our ArbRCAN to Meta-RCAN and RCAN on SR with non-integer scale factors. Comparative results achieved on the Set5 and Set14 datasets are shown in Table III. It can be observed that our ArbRCAN produces comparable or better performance as compared to other methods on most scale factors. Specifically, our ArbRCAN outperforms Meta-RCAN on Set5 for $\times 1.55/3.1$ SR, with higher PSNR values being achieved (40.97/34.50 vs. 40.93/34.44).

Qualitative Results. Figure VI illustrates the visual results achieved on three images of the Urban100 dataset. It can be observed from the zoom-in regions that our ArbRCAN produces results with better perceptual quality than other methods for different non-integer scale factors. For example, Meta-RCAN produces obvious blurring artifacts on "img_049". In contrast, our network recovers clearer stripes.

IV.II. SR with Asymmetric Scale Factors

Quantitative Results. We compare our ArbRCAN to Meta-RCAN and RCAN on SR with asymmetric scale factors. Comparative results are presented in Table IV. As we can see, our ArbRCAN achieves the best performance on all scale factors. Further, the performance improvements on highly asymmetric scale factors are more signif-

Table III. PSNR results achieved on Set5 and Set14 for non-integer scale factors.

	Set5									Set14								
	×1.55	×1.8	×2	×2.35	×2.5	×2.95	×3.1	×3.8	Avg	×1.55	×1.8	×2	×2.35	×2.5	×2.95	×3.1	×3.8	Avg
Bicubic	36.24	34.70	33.65	32.18	31.77	30.04	29.89	28.75	35.15	32.53	31.12	30.09	28.95	28.61	27.46	27.22	26.14	29.02
RCAN [11](+Bicubic)	40.77	39.08	**38.27**	36.45	36.16	34.74	34.37	33.04	36.61	36.96	34.99	**34.12**	32.36	31.94	30.53	30.26	28.97	32.52
Meta-RCAN [12]+ft	40.93	39.20	38.21	36.53	36.20	34.76	34.44	33.03	36.66	37.01	35.15	34.03	32.48	31.98	30.56	30.37	**29.04**	32.57
ArbRCAN (Ours)	**40.97**	**39.20**	38.26	**36.55**	**36.22**	**34.77**	**34.50**	**33.04**	**36.69**	**37.05**	**35.15**	34.09	**32.48**	**32.01**	**30.59**	**30.37**	29.01	**32.60**

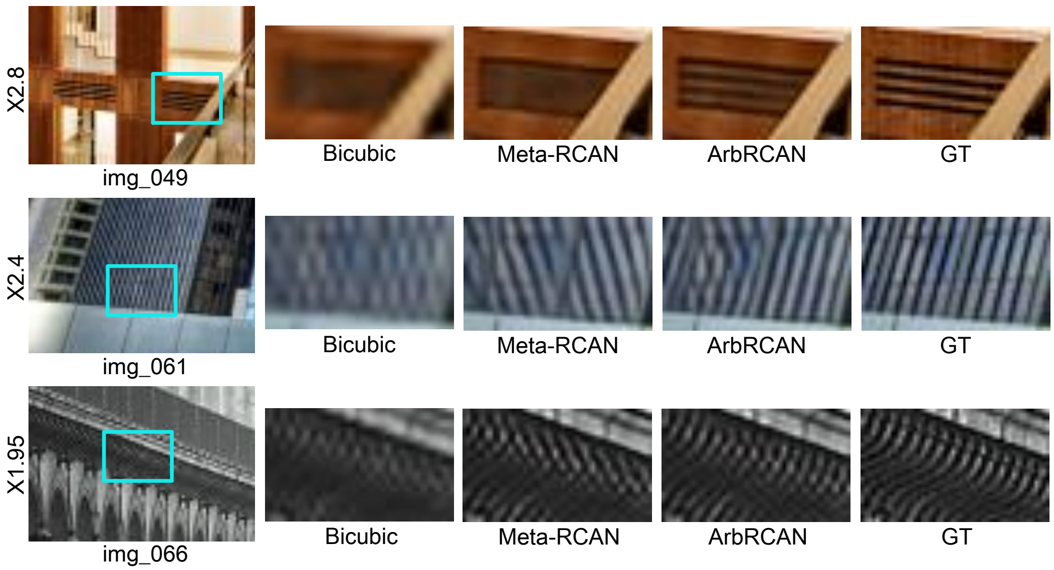


Figure VI. Visual comparison for non-integer SR.

icant (*e.g.,* 35.25(↑0.24)/35.60(↑0.13) vs. 35.01/35.47 on Set5 for $\frac{\times 4}{\times 1.5}$/$\frac{\times 3.5}{\times 2}$ SR).

Qualitative Results. Figure V compares the visual results achieved on three images of the Urban100 dataset. From the zoom-in regions, we can see that the results produced by our ArbRCAN have better visual quality than other methods for different asymmetric scale factors, such as the number "65" in "img_006".

IV.III. SR in the Wild

We further compare our ArbRCAN to Meta-RCAN on two real-world images in Fig. VI. It can be observed from the zoom-in regions that our ArbRCAN consistently produces clearer and finer details than Meta-RCAN. For example, our ArbRCAN faithfully recovers the text "HE" while Meta-RCAN suffers distorted artifacts.

References

[1] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *CVPR*, 2017. 1

[2] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, 2019. 1

[3] Jingwen He, Chao Dong, and Yu Qiao. Modulating image restoration with continual levels via adaptive feature modification layers. In *CVPR*, pages 11056–11064, 2019. 1, 2

[4] Yu-Syuan Xu, Shou-Yao Roy Tseng, Yu Tseng, Hsien-Kai Kuo, and Yi-Min Tsai. Unified dynamic convolutional network for super-resolution with variational degradations. In *CVPR*, pages 12496–12505, 2020. 1

[5] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In *CVPR*, 2020. 1

[6] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *CVPR*, 2020. 1

[7] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, pages 1646–1654, 2016. 1

[8] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *ECCV*, pages 391–407, 2016. 1

[9] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, pages 252–268, 2018. 1

[10] Ying Fu, Jian Chen, Tao Zhang, and Yonggang Lin. Residual scale attention network for arbitrary scale image super-resolution. *Neurocomputing*, 427:201–211, 2021. 2

[11] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep

Table IV. PSNR results achieved on Set5 and Set14 for asymmetric scale factors.

	Set5									Set14								
	$\frac{\times1.3}{\times1.5}$	$\frac{\times1.6}{\times2.95}$	$\frac{\times1.6}{\times3.45}$	$\frac{\times1.55}{\times3.5}$	$\frac{\times1.75}{\times3.5}$	$\frac{\times4}{\times1.5}$	$\frac{\times3.5}{\times2}$	$\frac{\times3}{\times1.95}$	Avg	$\frac{\times1.3}{\times1.5}$	$\frac{\times1.6}{\times2.95}$	$\frac{\times1.6}{\times3.45}$	$\frac{\times1.55}{\times3.5}$	$\frac{\times1.75}{\times3.5}$	$\frac{\times4}{\times1.5}$	$\frac{\times3.5}{\times2}$	$\frac{\times3}{\times1.95}$	Avg
Bicubic	37.44	37.61	30.83	30.82	30.57	30.72	31.04	31.62	32.58	33.74	29.10	28.33	28.30	28.13	27.64	27.88	27.43	28.82
RCAN [11](+Bicubic)	41.75	36.02	35.03	35.02	34.88	34.97	35.33	36.11	36.14	38.17	32.32	31.34	31.41	31.15	30.70	31.02	31.96	32.26
Meta-RCAN [12]+ft(+Bicubic)	42.04	36.03	35.15	35.22	35.06	35.01	35.47	36.09	36.26	38.29	32.43	31.49	31.54	31.30	30.91	31.10	31.94	32.38
ArbRCAN (Ours)	**42.16**	**36.21**	**35.29**	**35.35**	**35.16**	**35.25**	**35.60**	**36.23**	**32.41**	**38.45**	**32.51**	**31.57**	**31.63**	**31.37**	**31.04**	**31.27**	**32.11**	**32.49**

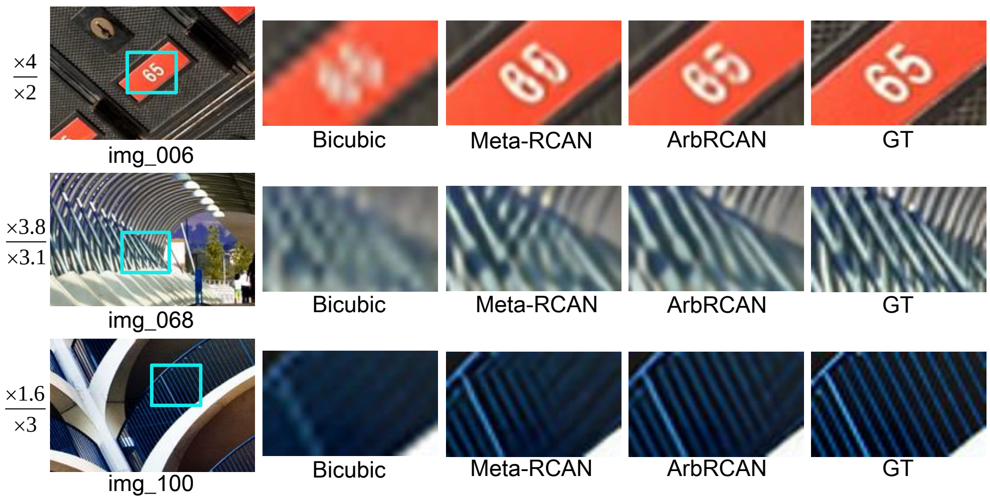


Figure V. Visual comparison for asymmetric SR.

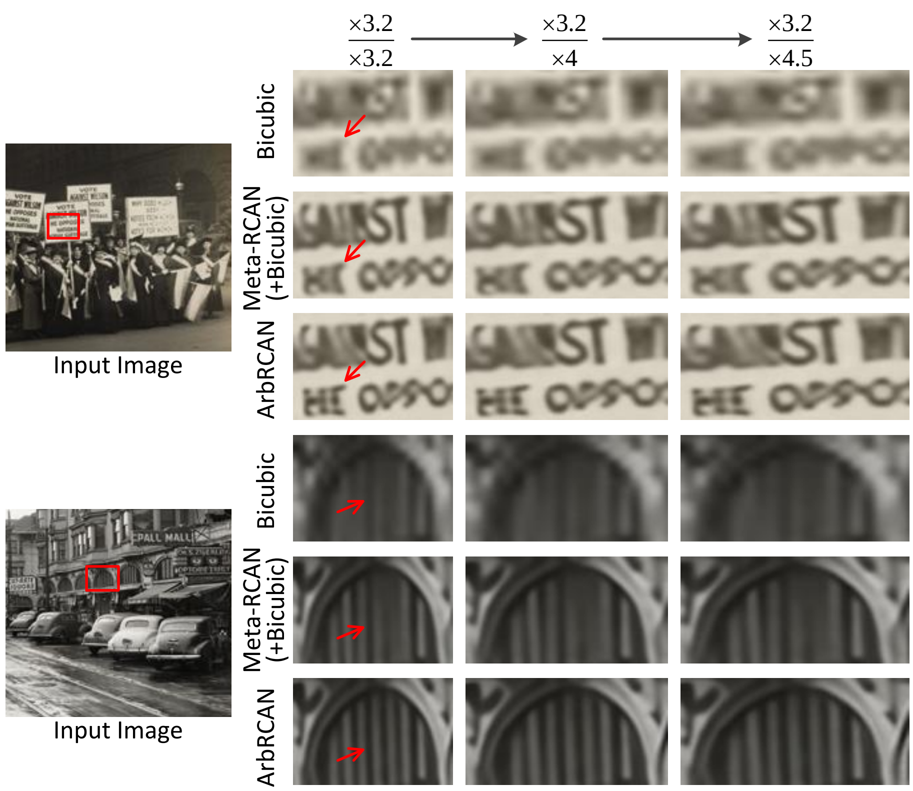


Figure VI. Visual comparison on real-world images.

residual channel attention networks. In *ECCV*, pages 1646–1654, 2018. 4, 5, 6

[12] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Jian Sun, and Tieniu Tan. Meta-SR: A magnification-arbitrary network for super-resolution. In *CVPR*, 2019. 4, 5, 6