Supplemental File to "Restore Globally, Refine Locally: A Mask-Guided Scheme to Accelerate Super-Resolution Networks"

Xiaotao Hu[1,2], Jun Xu[2] *, Shuhang Gu[3], Ming-Ming Cheng[1], and Li Liu[4]

[1] College of Computer Science, Nankai University
[2] School of Statistics and Data Science, Nankai University
[3] The University of Sydney
[4] Inception Institute of Artificial Intelligence

1 Content

Here, we provide more details of our Mask-Guided Acceleration (MGA) scheme on accelerating five representative super-resolution (SR) networks. Specifically, we provide

- the choice of K for feature patch selection in §2;
- the importance of dynamic convolution in our Mask Prediction (MP) module in §3;
- detail comparison of our MGA scheme with ClassSR in §4;
- comparison of our MGA accelerated SR networks and swift-designed SR networks in similar FLOPs in §5;
- the implementation details of our MGA scheme in §6;
- more details of decomposing SR networks in §7;
- more quantitative comparisons on different variants of five existing SR networks in §8;
- more ablation studies in §9;
- more visual comparisons on different variants of five existing SR networks accelerated by our MGA scheme in §10.

2 The Choice of K for Feature Patch Selection

We sort the values of all pixels in "MASK_GT" (described in **Line334** of our main paper) and "MASK", respectively, and calculate the intersection number between the top K pixels with maximum values in MASK_GT and MASK. The results in Table 1 show that our MP module is more accurate when $K = 1000$ than the other choices. So we choose $K = 1000$ to trade-off performance and efficiency. For datasets with large image size (*e.g.*, Urban100, Manga109, Test2K, Test4K), we set $K = 1000$. For datasets with small image size (*e.g.*, Set5, Set14, B100), we set $K = 240$.

* Jun Xu is the corresponding author (email: nankaimathxujun@gmail.com).

Table 1: **Accuracy of our MP module on the mask prediction with different** K ($p = 4$) on Urban100.

K	200	400	600	800	1000	1200	1400
Number	129	302	496	719	960	969	987
Accuracy (%)	64.5	75.5	82.7	89.9	96.0	80.8	70.5

Table 2: **Comparison results of different SR networks** on the Urban100, Test2K and Test4K datasets. "-M(conv)": common convolution. "-M(D-conv)": dynamic convolution.

Scale	Method	# Params	Urban100			Test2K			Test4K		
			PSNR	SSIM	FLOPs(G)	PSNR	SSIM	FLOPs(G)	PSNR	SSIM	FLOPs(G)
×3	FSRCNN-M(conv)	42K	26.20	0.8019	13.44	28.59	0.8215	28.59	30.02	0.8598	103.78
	FSRCNN-M(D-conv)	43K	26.22	0.8025	13.45	28.59	0.8216	28.61	30.03	0.8598	103.85
	PAN-M(conv)	310K	27.92	0.8467	36.31	29.16	0.8395	74.87	30.78	0.8756	266.22
	PAN-M(D-conv)	311K	27.96	0.8475	36.34	29.17	0.8396	74.94	30.78	0.8757	266.49
	RFDN-M(conv)	720K	28.25	0.8540	45.12	29.22	0.8401	93.58	30.84	0.8760	334.03
	RFDN-M(D-conv)	721K	28.28	0.8545	45.16	29.23	0.8403	93.67	30.84	0.8761	334.38
	SRResNet-M(conv)	1.71M	27.95	0.8477	144.65	29.18	0.8391	309.15	30.75	0.8747	1125.26
	SRResNet-M(D-conv)	1.71M	28.01	0.8485	144.70	29.18	0.8391	309.26	30.76	0.8747	1125.70
	RCAN-M(conv)	16.11M	29.07	0.8702	1110.15	29.54	0.8493	2362.16	31.23	0.8841	8574.10
	RCAN-M(D-conv)	16.11M	29.10	0.8706	1110.19	29.55	0.8493	2362.27	31.23	0.8841	8574.54
×4	FSRCNN-M(conv)	42K	24.47	0.7248	14.59	27.08	0.7521	29.18	28.30	0.8003	101.70
	FSRCNN-M(D-conv)	43K	24.50	0.7261	14.60	27.09	0.7523	29.19	28.31	0.8004	101.74
	PAN-M(conv)	327K	25.96	0.7806	29.72	27.55	0.7725	58.26	28.95	0.8194	200.22
	PAN-M(D-conv)	328K	26.02	0.7825	29.74	27.57	0.7728	58.30	28.96	0.8195	200.37
	RFDN-M(conv)	739K	26.14	0.7872	29.92	27.61	0.7739	27.70	29.02	0.8209	195.96
	RFDN-M(D-conv)	740K	26.17	0.7878	29.94	27.61	0.7741	57.75	29.03	0.8210	196.15
	SRResNet-M(conv)	2.01M	26.03	0.7839	124.01	27.57	0.7734	249.14	28.92	0.8197	871.64
	SRResNet-M(D-conv)	2.01M	26.07	0.7850	124.04	27.57	0.7735	249.20	28.92	0.8197	871.89
	RCAN-M(conv)	16.04M	26.90	0.8106	728.49	27.86	0.7835	1457.45	29.31	0.8294	5084.20
	RCAN-M(D-conv)	16.04M	26.96	0.8119	728.52	27.86	0.7837	1457.51	29.31	0.8295	5084.44

3 The Importance of Dynamic Convolution in Our MP Module

The local patches of an image are spatially varying, as well as the relationship between the central pixel and its surrounding pixels. It is essential to employ dynamic conv to adaptively tackle the weight matrices for different local patches. The comparison results of our MP module with standard convolution or dynamic convolution on accelerating SR networks are listed in Table 2. One can see that the proposed MP module with dynamic convolution is indeed more helpful to our MGA scheme on accelerating the five popular SR networks, on the Urban100, Test2K, and Test4K datasets.

4 Detail Comparison of Our MGA Scheme with ClassSR

We denote our Mask-Guided variant as M-Net. From Tabel 1 of our main paper, Table 6, Table 3, and Table 7, we observe that All-Net (described in **Line328** of our main paper) achieves a better performance than D-Net (described in **Line322** of our main paper). The reason is that, processing patch-wise features further in the last layers of the accelerated SR network enables it to well reconstruct the local details of LR image, and thus boosting the SR performance.

Table 3: **Results of different SR networks accelerated by ClassSR and our MGA scheme** on the Set5, Set14, B100 and Manga109 datasets. "-D": D-Net. "-A": ALL-Net. "-M": network accelerated by our MGA scheme. "-C": network accelerated by ClassSR.

Scale	Method	# Params	Set5		Set14		B100		Manga109	
			PSNR(RGB)	FLOPs(G)	PSNR(RGB)	FLOPs(G)	PSNR(RGB)	FLOPs(G)	PSNR(RGB)	FLOPs(G)
×4	FSRCNN-D	42K	28.59	3.15	25.75	6.42	25.57	4.32	25.61	26.74
	FSRCNN-A	42K	28.69	3.17	25.83	6.47	25.61	4.35	25.74	26.94
	FSRCNN-C	113K	28.58	2.58	25.75	4.47	25.57	3.34	25.60	19.28
	FSRCNN-M	43K	28.69	2.47	25.83	4.13	25.60	3.06	25.74	17.21
	SRResNet-D	1.67M	30.25	23.22	26.78	47.38	26.23	31.86	28.40	197.37
	SRResNet-A	1.67M	30.26	26.45	26.79	53.96	26.24	36.29	28.40	224.80
	SRResNet-C	3.06M	30.23	20.75	26.75	38.54	26.20	26.01	28.40	190.67
	SRResNet-M	2.01M	30.25	20.85	26.79	35.14	26.24	25.96	28.40	146.38
	RCAN-D	15.70M	30.65	117.62	27.11	239.93	26.47	161.34	29.49	999.53
	RCAN-A	15.70M	30.66	157.97	27.12	322.26	26.48	216.70	29.59	1342.50
	RCAN-C	30.11M	30.64	110.42	27.09	205.02	26.44	147.56	29.50	938.28
	RCAN-M	16.04M	30.66	106.82	27.11	206.11	26.48	152.66	29.57	858.64

Table 4: **Comparison of compressed and swift-designed methods with similar or higher computational costs.** We test all the models on one NVIDIA Tesla V100 GPU.

Scale	Method	Urban100			Test2K			Test4K		
		PSNR (dB)	SSIM	FLOPs (G)	PSNR (dB)	SSIM	FLOPs (G)	PSNR (dB)	SSIM	FLOPs (G)
×4	CARN-Mobile	25.63	0.7701	**27.28**	27.44	0.7684	63.46	28.80	0.8156	243.47
	IMDN	26.03	0.7837	34.19	27.58	0.7734	79.55	28.98	0.8201	305.18
	RFDN-M	**26.17**	**0.7878**	29.94	**27.61**	**0.7741**	**57.75**	**29.03**	**0.8210**	**196.15**
	D-DBPN	26.38	0.7946	4352.32	27.71	0.7790	10124.85		OOM	
	RDN	26.61	0.8028	1093.45	27.69	0.7781	2543.70	29.11	0.8243	9759.12
	RCAN-M	**26.96**	**0.8119**	**728.52**	**27.86**	**0.7837**	**1457.51**	**29.31**	**0.8295**	**5084.44**

To this end, the Refine-Net (described in **Line221** of our main paper) of M-Net further processes the feature patches of the most under-SR areas selected by our MP module. Our MP module boosts the performance of M-Net to be close to the performance of All-Net. This is very different from CLassSR, since it achieves the performance close to that of D-Net. Therefore, the performance of our MGA scheme on accelerating SR networks can be potentially better than that of ClassSR.

5 Comparison of our MGA Accelerated SR Networks and Swift-Designed SR Networks with Similar FLOPs

The results in Table 4 show that the RFDN-M or RCAN-M accelerated by our MGA scheme outperforms the corresponding state-of-the-art SR networks in terms of PSNR and SSIM, while with similar or higher computational costs in terms of FLOPs.

6 An Implementation Detail of Our MGA Scheme

Here, we present more details about our proposed MGA scheme. We first get the coarse feature $\boldsymbol{F}_c \in \mathbb{R}^{C\times H\times W}$ via the Base-Net. Then we feed the coarse feature $\boldsymbol{F}_c$ into the Mask Prediction (MP) module. The MP module contains two 3×3 convolutional layers. To obtain a D-Conv of kernel size 5×5, through a 3×3 convolutional layer, our MGA scheme obtains from $\boldsymbol{F}_c$ the weight of the D-conv, which is of size $\mathbb{R}^{5\times 5\times H\times W}$. The size of feature patches obtained by the Feature Patch Selection is $p\times p$. To make the cut edges smoother, we crop the feature patches of size $p+pa$, where $pa=1$ is an additional padding size. After one convolutional layer, this padding size can be compensated, and we obtain feature patches of size $p\times p$. In order to keep the parameters consistent, we do not change the overall network structure. The padding size of the first convolutional layer in the Refine-Net is set as $pa=0$. The upsampling operation to produce the coarse SR image after the Base-Net is exactly the same as that in the original SR network. In our MGA scheme, a skip connection could be used within the Base-Net and Refine-Net, respectively, possibly with a little performance drop. We will define the compatible types of our MGA scheme in the revision.

7 More Details on Network Decomposition

Our MGA scheme has different decomposing ways for different SR networks. Specifically, 1) for FSRCNN [2], the Base-Net has 3 basic blocks while the Refine-Net has 1 basic block; 2) for PAN [12], the Base-Net has 12 basic blocks while the Refine-Net has 4 basic blocks; 3) for RFDN [6], the Base-Net has 4 basic blocks while the Refine-Net has 2 basic blocks; 4) for SRResNet [5], the Base-Net has 12 basic blocks while the Refine-Net has 4 basic blocks; 5) for RCAN [11], the Base-Net has 7 basic blocks while the Refine-Net has 3 basic blocks. We use four Nvidia 2080Ti GPUs for training. The duration for FSRCNN, RFDN, PAN, SRResNet and RCAN are about 48, 72, 56, 96 and 168 hours, respectively. We will explain these details in our revision.

8 More Quantitative Results

In Table 6, we summarize the quantitative results of five SR networks on the three datasets (i.e., Urban100 [4], Test2K [3], and Test4K [3]) on the $\times 2$ SR task. In Table 7, we summarize the quantitative results of five SR networks on the rest four datasets (i.e., Set5 [1], Set14 [10], B100 [7], and Manga109 [8]) for the $\times 2, \times 3$, and $\times 4$ SR tasks.

9 More Ablation Study Results

More details on our Ablation Study 1) in the main paper. In Table 8, we summarize the quantitative results of two SR networks (FSRCNN [2] and

PAN [12]) on five datasets (i.e., B100 [7], Urban100 [4], Manga109 [8], Test2K [3], and Test4K [3]) for the $\times 2$ and $\times 4$ SR tasks. These results validate the necessity to design a two-channel spatial feature map $\boldsymbol{F}_s$ in our MGA scheme.

More details on our Ablation Study 2) in the main paper. In Table 9, we summarize the quantitative results of FSRCNN [2] and PAN [12] on five datasets (i.e., B100 [7], Urban100 [4], Manga109 [8], Test2K [3], and Test4K [3]) for the $\times 2$ and $\times 4$ SR tasks. This validates the advantages of our training strategy over the two variant strategies of "E2E" and "w/o S", for our MGA scheme.

More details on our Ablation Study 4) in the main paper. In Table 10, we summarize the quantitative results of two SR networks (FSRCNN [2] and PAN [12]) on five datasets (i.e., B100 [7], Urban100 [4], Manga109 [8], Test2K [3], and Test4K [3]) for the $\times 2$ and $\times 4$ SR tasks. The results show that $p = 4$ is better than the other choices to trade-off the performance of our MGA scheme between the PSNR/SSIM results and amounts of FLOPs.

How does the value of K influence our MGA on the SR performance? The K is an important factor to affect the FLOPs and SR capability of the mask-guided SR networks. Here, we compare the RFDN-M with different Ks and list the results in Table 5. One can see that the RFDN-M with larger K enjoys better PSNR and SSIM results, but suffers from higher FLOPs and Activations. By varying the value of K, our MGA scheme can provide a flexible trade-off between the SR capability and network efficiency for the RFDN.

Table 5: **Results of RFDN-M with different numbers (K) of elements selected from the error mask** on the Urban100 dataset. The image size is $376 \times 512 \times 3$. "All": selecting all errors for each image. The scale factor is 2.

Method	RFDN-B	RFDN-M(0)	RFDN-M(1000)	RFDN-M(2000)	RFDN-M(4000)	RFDN-M(6000)	RFDN-M(8000)	RFDN-M(All)
FLOPs(G)	76.4988	80.8363	89.5428	98.2492	115.6621	133.0749	150.4878	181.2546
Activations(M)	322.2028	331.8405	365.5445	399.2485	466.6565	534.0645	601.4725	727.7294
PSNR	32.00	31.99	32.26	32.30	32.34	32.36	32.36	32.36
SSIM	0.9268	0.9265	0.9296	0.9299	0.9302	0.9304	0.9305	0.9306

10 More Visual Comparisons

In Figures 1–14, we visualize the SR results of all five representative SR networks and the corresponding accelerated variants by our MGA scheme on seven test datasets for $\times 3$, and $\times 4$ SR tasks.

Table 6: Comparison of parameters (Params), PSNR (dB), SSIM [9], and FLOPs (G) results by different SR networks on the Urban100, Test2K and Test4K datasets. "-B": Base-Net. "-D": D-Net. "-A": ALL-Net. "-M": network accelerated by our MGA scheme.

Scale	Method	Params	Urban100			Test2K			Test4K		
			PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs
×2	FSRCNN-B	23K	29.15	0.8913	12.38	31.52	0.9107	28.65	33.32	0.9326	109.92
	FSRCNN-D	42K	29.38	0.8951	23.80	31.62	0.9122	55.08	33.43	0.9337	211.33
	FSRCNN-A	42K	29.59	0.8980	24.43	31.68	0.9129	56.55	33.50	0.9343	216.95
	FSRCNN-M	43K	29.56	0.8973	13.63	31.67	0.9128	30.23	33.49	0.9343	113.15
	PAN-B	204K	31.58	0.9234	48.26	32.37	0.9232	111.70	34.37	0.9428	428.56
	PAN-D	291K	31.86	0.9259	73.91	32.46	0.9242	171.09	34.48	0.9436	656.39
	PAN-A	291K	31.95	0.9267	97.67	32.49	0.9247	226.09	34.52	0.9440	867.40
	PAN-M	306K	31.81	0.9252	55.14	32.45	0.9240	122.25	34.45	0.9433	457.42
	RFDN-B	417K	32.00	0.9268	76.82	32.49	0.9247	177.82	34.53	0.9441	682.26
	RFDN-D	684K	32.16	0.9285	126.85	32.58	0.9255	293.62	34.61	0.9446	1126.54
	RFDN-A	684K	32.36	0.9306	182.02	32.62	0.9262	421.33	34.66	0.9451	1616.47
	RFDN-M	740K	32.26	0.9296	89.88	32.57	0.9254	196.61	34.59	0.9445	729.65
	SRResNet-B	1.10M	30.45	0.9162	221.92	30.84	0.9115	513.67	30.91	0.9199	1970.75
	SRResNet-D	1.67M	30.80	0.9186	340.95	31.23	0.9139	789.21	31.43	0.9238	3027.88
	SRResNet-A	1.67M	31.12	0.9222	430.12	31.37	0.9168	995.59	31.81	0.9275	3819.68
	SRResNet-M	1.71M	30.87	0.9199	246.28	31.13	0.9140	547.42	31.27	0.9214	2051.34
	RCAN-B	10.87M	32.58	0.9326	2084.39	32.70	0.9273	4824.77	34.72	0.9458	18510.62
	RCAN-D	15.70M	33.30	0.9385	3012.95	32.96	0.9302	6974.12	34.98	0.9476	26756.77
	RCAN-A	15.70M	33.29	0.9381	4127.81	32.95	0.9301	9554.71	34.97	0.9476	36657.38
	RCAN-M	15.74M	33.26	0.9379	2260.64	32.94	0.9300	5010.41	34.96	0.9475	18743.10

Table 7: Comparison results of parameter number (Params), PSNR (dB), SSIM [9], and FLOPs (G) by different SR networks on the Set5, Set14, B100 and Manga109 datasets. "-B": Base-Net. "-D": D-Net. "-A": ALL-Net. "-M": network accelerated by our MGA scheme.

Scale	Method	Params	Set5			Set14			B100			Manga109		
			PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs
×2	FSRCNN-B	23K	36.59	0.9546	1.81	32.36	0.9068	3.67	31.25	0.8872	2.46	35.51	0.9670	15.38
	FSRCNN-D	42K	36.82	0.9556	3.48	32.52	0.9080	7.06	31.35	0.8886	4.73	36.03	0.9690	29.57
	FSRCNN-A	42K	36.97	0.9560	3.58	32.61	0.9088	7.24	31.42	0.8894	4.85	36.27	0.9701	30.35
	FSRCNN-M	43K	36.97	0.9561	2.09	32.60	0.9087	3.98	31.42	0.8895	2.75	36.22	0.9697	16.69
	PAN-B	204K	37.87	0.9601	7.07	33.41	0.9168	14.31	32.08	0.8987	9.59	38.28	0.9766	59.96
	PAN-D	291K	37.93	0.9603	10.82	33.51	0.9174	21.91	32.15	0.8992	14.68	38.57	0.9771	91.84
	PAN-A	291K	37.98	0.9605	14.30	33.56	0.9178	28.96	32.17	0.8996	19.40	38.64	0.9773	121.36
	PAN-M	306K	37.95	0.9603	8.46	33.52	0.9173	16.12	32.15	0.8991	11.12	38.50	0.9769	67.52
	RFDN-B	417K	37.95	0.9603	11.25	33.57	0.9183	22.78	32.16	0.8995	15.26	38.75	0.9773	95.45
	RFDN-D	684K	38.02	0.9606	18.57	33.64	0.9185	37.61	32.19	0.8998	25.19	38.95	0.9777	157.61
	RFDN-A	684K	38.06	0.9607	26.65	33.72	0.9191	53.96	32.24	0.9005	36.15	39.09	0.9781	266.16
	RFDN-M	740K	38.03	0.9606	13.97	33.67	0.9186	26.16	32.23	0.9003	18.21	38.92	0.9776	109.57
	SRResNet-B	1.10M	37.61	0.9595	32.49	33.05	0.9155	65.80	31.87	0.8979	44.08	35.89	0.9708	275.74
	SRResNet-D	1.67M	37.75	0.9599	49.92	33.19	0.9161	101.09	32.01	0.8981	67.72	36.31	0.9723	423.65
	SRResNet-A	1.67M	37.85	0.9601	62.97	33.36	0.9168	127.52	32.02	0.8985	85.43	37.30	0.9747	534.43
	SRResNet-M	1.71M	37.80	0.9600	37.67	33.27	0.9165	72.05	31.99	0.8985	49.63	36.98	0.9739	301.83
	RCAN-B	10.87M	38.11	0.9610	305.17	33.79	0.9198	617.99	32.27	0.9011	414.02	39.07	0.9782	2589.92
	RCAN-D	15.70M	38.30	0.9617	441.12	34.19	0.9224	893.30	32.41	0.9026	598.47	39.68	0.9794	3743.68
	RCAN-A	15.70M	38.31	0.9618	604.34	34.17	0.9222	1223.84	32.40	0.9026	819.91	39.68	0.9795	5128.93
	RCAN-M	15.74M	38.31	0.9618	346.80	34.15	0.9221	660.70	32.40	0.9026	456.03	39.62	0.9793	2767.90
×3	FSRCNN-B	23K	32.53	0.9077	1.62	29.16	0.8204	3.31	28.25	0.7848	2.20	29.84	0.9070	13.91
	FSRCNN-D	42K	32.75	0.9111	3.19	29.31	0.8232	6.49	28.36	0.7877	4.31	30.39	0.9147	27.31
	FSRCNN-A	42K	32.89	0.9129	3.23	29.41	0.8247	6.58	28.43	0.7889	4.37	30.71	0.9180	27.65
	FSRCNN-M	43K	32.89	0.9129	2.14	29.40	0.8244	3.84	28.43	0.7886	2.72	30.68	0.9173	16.14
	PAN-B	204K	34.21	0.9262	4.00	30.18	0.8400	8.15	29.01	0.8045	5.41	33.13	0.9421	34.25
	PAN-D	291K	34.29	0.9266	7.13	30.28	0.8408	14.54	29.06	0.8048	9.65	33.39	0.9434	61.13
	PAN-A	291K	34.32	0.9271	9.70	30.33	0.8424	19.76	29.09	0.8060	13.12	33.47	0.9442	83.09
	PAN-M	306K	34.30	0.9267	5.95	30.30	0.8416	10.29	29.07	0.8053	7.43	33.36	0.9429	43.19
	RFDN-B	417K	34.25	0.9263	4.97	30.26	0.8408	10.13	29.05	0.8048	6.72	33.36	0.9428	42.58
	RFDN-D	684K	34.37	0.9273	8.23	30.27	0.8411	16.78	29.09	0.8057	11.14	33.57	0.9443	70.55
	RFDN-A	684K	34.41	0.9278	11.74	30.37	0.8428	23.93	29.13	0.8070	15.89	33.80	0.9456	100.62
	RFDN-M	740K	34.41	0.9276	7.36	30.34	0.8422	12.81	29.12	0.8066	9.21	33.69	0.9445	53.77
	SRResNet-B	1.10M	34.17	0.9260	17.34	30.07	0.8390	35.34	28.98	0.8048	23.47	32.74	0.9395	148.60
	SRResNet-D	1.67M	34.39	0.9275	28.14	30.28	0.8416	57.35	29.10	0.8064	38.08	33.38	0.9433	241.12
	SRResNet-A	1.67M	34.38	0.9274	33.81	30.30	0.8416	68.90	29.10	0.8065	45.75	33.44	0.9435	289.72
	SRResNet-M	1.71M	34.37	0.9272	22.94	30.28	0.8411	41.41	29.10	0.8060	29.22	33.36	0.9425	173.91
	RCAN-B	10.87M	34.53	0.9289	134.93	30.46	0.8450	275.02	29.20	0.8094	182.62	34.03	0.9474	1156.36
	RCAN-D	15.70M	34.80	0.9303	196.35	30.66	0.8485	400.22	29.33	0.8122	265.76	34.70	0.9508	1682.77
	RCAN-A	15.70M	34.78	0.9305	267.25	30.68	0.8485	544.74	29.33	0.8124	361.72	34.73	0.9510	2290.39
	RCAN-M	15.74M	34.78	0.9304	176.71	30.66	0.8490	317.28	29.33	0.8123	224.57	34.67	0.9504	1332.46
×4	FSRCNN-B	23K	30.25	0.8625	1.59	27.39	0.7536	3.25	26.85	0.7115	2.18	27.19	0.8519	13.52
	FSRCNN-D	42K	30.47	0.8660	3.15	27.54	0.7559	6.42	26.91	0.7128	4.32	27.41	0.8552	26.74
	FSRCNN-A	42K	30.58	0.8684	3.17	27.62	0.7578	6.47	26.95	0.7141	4.35	27.65	0.8605	26.94
	FSRCNN-M	43K	30.58	0.8684	2.47	27.61	0.7575	4.13	26.94	0.7140	3.06	27.64	0.8598	17.21
	PAN-B	204K	31.92	0.8936	3.03	28.50	0.7802	6.19	27.51	0.7339	4.16	30.18	0.9052	25.77
	PAN-D	291K	32.06	0.8949	5.25	28.54	0.7807	10.70	27.56	0.7349	7.20	30.33	0.9071	44.58
	PAN-A	291K	32.09	0.8952	6.63	28.57	0.7815	13.52	27.57	0.7356	9.09	30.41	0.9080	56.31
	PAN-M	306K	32.09	0.8951	5.10	28.56	0.7810	8.36	27.57	0.7353	6.27	30.37	0.9070	34.84
	RFDN-B	417K	32.08	0.8939	2.89	28.57	0.7817	5.90	27.54	0.7340	3.97	30.38	0.9073	24.60
	RFDN-D	684K	32.15	0.8956	4.82	28.62	0.7824	9.83	27.59	0.7364	6.61	30.57	0.9093	40.95
	RFDN-A	684K	32.21	0.8963	6.81	28.65	0.7827	13.90	27.61	0.7367	9.35	30.69	0.9104	57.92
	RFDN-M	740K	32.20	0.8962	4.13	28.64	0.7823	8.38	27.61	0.7364	6.34	30.63	0.9089	34.90
	SRResNet-B	1.10M	31.96	0.8933	13.47	28.43	0.7792	27.49	27.50	0.7339	18.48	29.84	0.9025	114.51
	SRResNet-D	1.67M	32.20	0.8960	23.22	28.59	0.7822	47.38	27.58	0.7364	31.86	30.33	0.9073	197.37
	SRResNet-A	1.67M	32.17	0.8955	26.45	28.60	0.7822	53.96	27.59	0.7367	36.29	30.31	0.9068	224.80
	SRResNet-M	1.71M	32.17	0.8955	20.85	28.59	0.7821	35.14	27.59	0.7366	25.96	30.30	0.9064	146.38
	RCAN-B	10.87M	32.37	0.8980	79.73	28.78	0.7862	162.65	27.70	0.7400	109.37	31.07	0.9148	677.55
	RCAN-D	15.70M	32.59	0.9003	117.62	28.93	0.7902	239.93	27.81	0.7440	161.34	31.50	0.9181	999.53
	RCAN-A	15.70M	32.58	0.9007	157.97	28.94	0.7906	322.26	27.82	0.7444	216.70	31.60	0.9197	1342.50
	RCAN-M	15.74M	32.57	0.9007	106.82	28.93	0.7903	206.11	27.82	0.7444	152.66	31.56	0.9189	858.64

Table 8: **Comparison results of PSNR (dB), SSIM [9], and FLOPs (G) by the FSRCNN-M and PAN-M with the spatial feature map $\boldsymbol{F}_s$ of one channel or two channels in our MP module** on B100, Urban100, Manga109, Test2K, and Test4K. "-B": Base-Net. "-D": D-Net. "-M": network accelerated by our MGA scheme. "1C": the spatial feature $\boldsymbol{F}_s$ is of one channel.

Scale	Method	B100			Urban100			Manga109			Test2K			Test4K		
		PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs
×2	FSRCNN-B	31.25	0.8872	2.46	29.15	0.8913	12.38	35.51	0.9670	15.38	31.52	0.9107	28.65	33.32	0.9326	109.92
	FSRCNN-M(1C)	31.41	0.8894	2.75	29.54	0.8972	13.63	36.18	0.9696	16.69	31.66	0.9128	30.23	33.49	0.9342	113.15
	FSRCNN-M	31.42	0.8895	2.75	29.56	0.8973	13.63	36.22	0.9697	16.69	31.67	0.9128	30.23	33.49	0.9343	113.15
	FSRCNN-D	31.35	0.8886	4.73	29.38	0.8951	23.80	36.03	0.9690	29.57	31.62	0.9122	55.08	33.43	0.9337	211.33
	PAN-B	32.08	0.8987	9.59	31.58	0.9234	48.26	38.28	0.9766	59.96	32.37	0.9232	111.70	34.37	0.9428	428.56
	PAN-M(1C)	32.12	0.8990	11.12	31.75	0.9250	55.14	38.39	0.9768	67.52	32.44	0.9240	122.25	34.44	0.9433	457.42
	PAN-M	32.15	0.8991	11.12	31.81	0.9252	55.14	38.50	0.9769	67.52	32.45	0.9240	122.25	34.45	0.9433	457.42
	PAN-D	32.15	0.8992	14.68	31.86	0.9259	73.91	38.57	0.9771	91.84	32.46	0.9242	171.09	34.48	0.9436	656.39
×4	FSRCNN-B	26.85	0.7115	2.18	24.32	0.7192	10.93	27.19	0.8519	13.52	27.01	0.7499	25.43	28.22	0.7984	97.57
	FSRCNN-M(1C)	26.92	0.7136	3.06	24.46	0.7251	14.60	27.57	0.8591	17.21	27 07	0.7520	29.19	28.30	0.8003	101.74
	FSRCNN-M	26.94	0.7140	3.06	24.50	0.7261	14.60	27.64	0.8598	17.21	27.09	0.7523	29.19	28.31	0.8004	101.74
	FSRCNN-D	26.91	0.7128	4.32	24.43	0.7230	21.63	27.41	0.8552	26.74	27.06	0.7512	50.31	28.29	0.7997	193.01
	PAN-B	27.51	0.7339	4.16	25.87	0.7780	20.84	30.18	0.9052	25.77	27.52	0.7716	48.47	28.91	0.8187	185.96
	PAN-M(1C)	27.54	0.7344	6.27	25.96	0.7812	29.74	30.22	0.9060	34.84	27.55	0.7725	58.30	28.94	0.8193	200.37
	PAN-M	27.57	0.7353	6.27	26.02	0.7825	29.74	30.37	0.9070	34.84	27.57	0.7728	58.30	28.96	0.8195	200.37
	PAN-D	27.56	0.7349	7.20	26.01	0.7828	36.05	30.33	0.9071	44.58	27.56	0.7725	83.86	28.96	0.8195	321.74

Table 9: **Comparison of PSNR (dB), SSIM [9], and FLOPs (G) results by the FSRCNN-M and PAN-M with different training strategies for our Mask Prediction (MP) module** on the B100, Urban100, Manga109, Test2K, and Test4K datasets. "E2E": end-to-end train the mask-guided network with our MP module from scratch. "w/o S": train our MP module separately in **Step 3** without supervision.

Scale	Method	B100			Urban100			Manga109			Test2K			Test4K		
		PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs
×2	FSRCNN-B	31.25	0.8872	2.46	29.15	0.8913	12.38	35.51	0.9670	15.38	31.52	0.9107	28.65	33.32	0.9326	109.92
	FSRCNN-M(E2E)	31.22	0.8874	2.75	29.10	0.8910	13.63	35.40	0.9667	16.69	31.50	0.9110	30.23	33.27	0.9326	113.15
	FSRCNN-M(w/o S)	31.38	0.8890	2.75	29.50	0.8970	13.63	36.13	0.9693	16.69	31.64	0.9124	30.23	33.47	0.9341	113.15
	FSRCNN-M	31.42	0.8895	2.75	29.56	0.8973	13.63	36.22	0.9697	16.69	31.67	0.9128	30.23	33.49	0.9343	113.15
	FSRCNN-D	31.35	0.8886	4.73	29.38	0.8951	23.80	36.03	0.9690	29.57	31.62	0.9122	55.08	33.43	0.9337	211.33
	PAN-B	32.08	0.8987	9.59	31.58	0.9234	48.26	38.28	0.9766	59.96	32.37	0.9232	111.70	34.37	0.9428	428.56
	PAN-M(E2E)	32.02	0.8979	11.12	31.40	0.9212	55.14	38.10	0.9761	67.52	32.32	0.9224	122.25	34.29	0.9420	457.42
	PAN-M(w/o S)	32.11	0.8990	11.12	31.72	0.9248	55.14	38.37	0.9767	67.52	32.43	0.9239	122.25	34.43	0.9432	457.42
	PAN-M	32.15	0.8991	11.12	31.81	0.9252	55.14	38.50	0.9769	67.52	32.45	0.9240	122.25	34.45	0.9433	457.42
	PAN-D	32.15	0.8992	14.68	31.86	0.9259	73.91	38.57	0.9771	91.84	32.46	0.9242	171.09	34.48	0.9436	656.39
×4	FSRCNN-B	26.85	0.7115	2.18	24.32	0.7192	10.93	27.19	0.8519	13.52	27.01	0.7499	25.43	28.22	0.7984	97.57
	FSRCNN-M(E2E)	26.81	0.7096	3.06	24.25	0.7155	14.60	26.95	0.8446	17.21	26.99	0.7486	29.19	28.19	0.7971	101.74
	FSRCNN-M(w/o S)	26.90	0.7127	3.06	24.44	0.7240	14.60	27.50	0.8574	17.21	27.06	0.7514	29.19	28.28	0.7997	101.74
	FSRCNN-M	26.94	0.7140	3.06	24.50	0.7261	14.60	27.64	0.8598	17.21	27.09	0.7523	29.19	28.31	0.8004	101.74
	FSRCNN-D	26.91	0.7128	4.32	24.43	0.7230	21.63	27.41	0.8552	26.74	27.06	0.7512	50.31	28.29	0.7997	193.01
	PAN-B	27.51	0.7339	4.16	25.87	0.7780	20.84	30.18	0.9052	25.77	27.52	0.7716	48.47	28.91	0.8187	185.96
	PAN-M(E2E)	27.47	0.7319	6.27	25.73	0.7736	29.74	29.93	0.9019	34.84	27.49	0.7697	58.30	28.84	0.8165	200.37
	PAN-M(w/o S)	27.53	0.7343	6.27	25.91	0.7798	29.74	30.17	0.9055	34.84	27.54	0.7720	58.30	28.92	0.8189	200.37
	PAN-M	27.57	0.7353	6.27	26.02	0.7825	29.74	30.37	0.9070	34.84	27.57	0.7728	58.30	28.96	0.8195	200.37
	PAN-D	27.56	0.7349	7.20	26.01	0.7828	36.05	30.33	0.9071	44.58	27.56	0.7725	83.86	28.96	0.8195	321.74

Table 10: **Comparison of PSNR (dB), SSIM [9], and FLOPs (G) results by the FSRCNN-M and PAN-M with different sizes of feature patches (selected between mask prediction and Refine-Net)** on B100, Urban100, Manga109, Test2K, and Test4K. "2": 2×2. "4": 4×4. "8": 8×8.

Scale	Method	B100			Urban100			Manga109			Test2K			Test4K		
		PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs	PSNR	SSIM	FLOPs
×2	FSRCNN-B	31.25	0.8872	2.46	29.15	0.8913	12.38	35.51	0.9670	15.38	31.52	0.9107	28.65	33.32	0.9326	109.92
	FSRCNN-M(2)	31.41	0.8892	2.77	29.53	0.8969	13.70	36.18	0.9696	16.77	31.66	0.9126	30.31	33.47	0.9341	113.27
	FSRCNN-M(4)	31.42	0.8895	2.75	29.56	0.8973	13.63	36.22	0.9697	16.69	31.67	0.9128	30.23	33.49	0.9343	113.15
	FSRCNN-M(8)	31.37	0.8888	2.74	29.44	0.8957	13.53	36.01	0.9691	16.51	31.63	0.9123	30.20	33.44	0.9338	113.11
	FSRCNN-D	31.35	0.8886	4.73	29.38	0.8951	23.80	36.03	0.9690	29.57	31.62	0.9122	55.08	33.43	0.9337	211.33
	PAN-B	32.08	0.8987	9.59	31.58	0.9234	48.26	38.28	0.9766	59.96	32.37	0.9232	111.70	34.37	0.9428	428.56
	PAN-M(2)	32.13	0.8990	11.76	31.73	0.9244	57.81	38.45	0.9767	70.20	32.43	0.9238	124.94	34.43	0.9432	460.21
	PAN-M(4)	32.15	0.8991	11.12	31.81	0.9252	55.14	38.50	0.9769	67.52	32.45	0.9240	122.25	34.45	0.9433	457.42
	PAN-M(8)	32.09	0.8987	10.86	31.66	0.9241	53.81	38.28	0.9765	65.83	32.41	0.9236	121.17	34.40	0.9430	456.31
	PAN-D	32.15	0.8992	14.68	31.86	0.9259	73.91	38.57	0.9771	91.84	32.46	0.9242	171.09	34.48	0.9436	656.39
×4	FSRCNN-B	26.85	0.7115	2.18	24.32	0.7192	10.93	27.19	0.8519	13.52	27.01	0.7499	25.43	28.22	0.7984	97.57
	FSRCNN-M(2)	26.95	0.7144	3.09	24.50	0.7264	14.73	27.56	0.8585	17.76	27.08	0.7527	29.26	28.30	0.8003	101.82
	FSRCNN-M(4)	26.94	0.7140	3.06	24.50	0.7261	14.60	27.64	0.8598	17.21	27.09	0.7523	29.19	28.31	0.8004	101.74
	FSRCNN-M(8)	26.92	0.7131	3.04	24.48	0.7226	14.52	27.55	0.8573	16.73	27.07	0.7517	28.64	28.29	0.7999	101.71
	FSRCNN-D	26.91	0.7128	4.32	24.43	0.7230	21.63	27.41	0.8552	26.74	27.06	0.7512	50.31	28.29	0.7997	193.01
	PAN-B	27.51	0.7339	4.16	25.87	0.7780	20.84	30.18	0.9052	25.77	27.52	0.7716	48.47	28.91	0.8187	185.96
	PAN-M(2)	27.56	0.7348	7.25	25.98	0.7808	33.82	30.33	0.9063	38.92	27.56	0.7722	62.38	28.94	0.8190	204.48
	PAN-M(4)	27.57	0.7353	6.27	26.02	0.7825	29.74	30.37	0.9070	34.84	27.57	0.7728	58.30	28.96	0.8195	200.37
	PAN-M(8)	27.54	0.7345	5.86	25.92	0.7800	27.95	30.13	0.9049	32.27	27.53	0.7724	55.59	28.92	0.8188	198.67
	PAN-D	27.56	0.7349	7.20	26.01	0.7828	36.05	30.33	0.9071	44.58	27.56	0.7725	83.86	28.96	0.8195	321.74

References

1. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012)
2. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Eur. Conf. Comput. Vis. pp. 391–407. Springer (2016)
3. Gu, S., Lugmayr, A., Danelljan, M., Fritsche, M., Lamour, J., Timofte, R.: Div8k: Diverse 8k resolution image dataset. In: Int. Conf. Comput. Vis. Worksh. pp. 3512–3516. IEEE (2019)
4. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 5197–5206 (2015)
5. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 4681–4690 (2017)
6. Liu, J., Tang, J., Wu, G.: Residual feature distillation network for lightweight image super-resolution. In: Eur. Conf. Comput. Vis. pp. 41–55. Springer (2020)
7. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Int. Conf. Comput. Vis. vol. 2, pp. 416–423. IEEE (2001)
8. Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications **76**(20), 21811–21838 (2017)
9. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
10. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: International Conference on Curves and Surfaces. pp. 711–730. Springer (2010)
11. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Eur. Conf. Comput. Vis. pp. 286–301 (2018)
12. Zhao, H., Kong, X., He, J., Qiao, Y., Dong, C.: Efficient image super-resolution using pixel attention. In: Eur. Conf. Comput. Vis. pp. 56–72. Springer (2020)

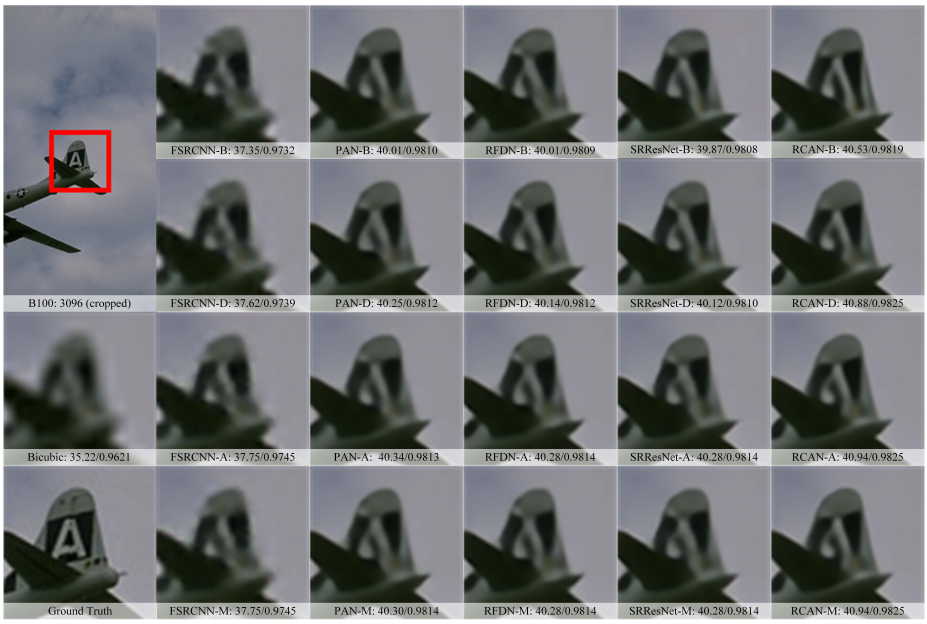

Fig. 1: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of five SR networks (i.e., FSRCNN [2], PAN [12], RFDN [6], SRResNet [5], and RCAN [11]) on B100 [7] at a scale factor of 4.

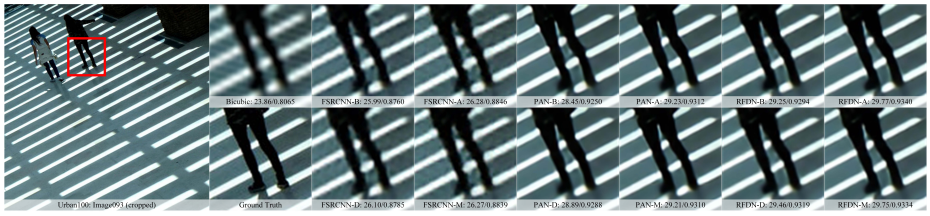

Fig. 2: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., FSRCNN [2], PAN [12], and RFDN [6]) on Urban100 [4] at a scale factor of 4.

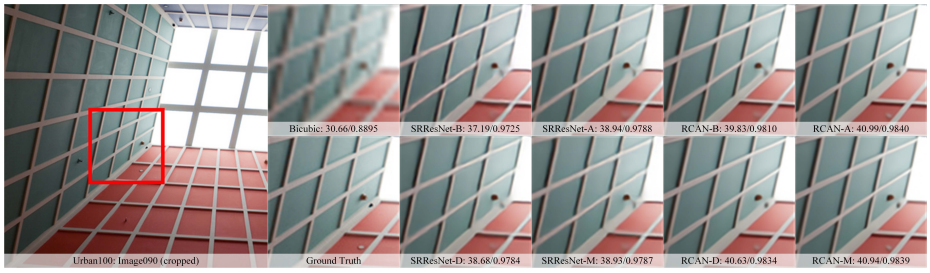

Fig. 3: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., SRResNet [5] and RCAN [11]) on Urban100 [4] at a scale factor of 4.

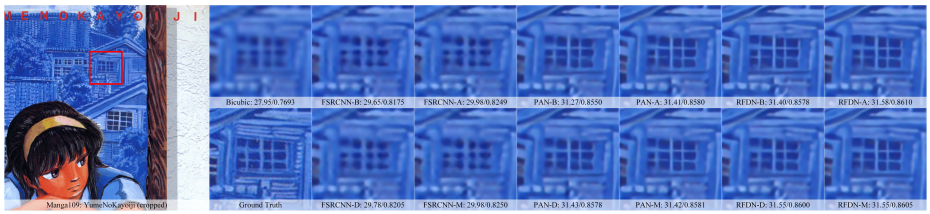

Fig. 4: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., FSRCNN [2], PAN [12], and RFDN [6]) on Manga109 [8] at a scale factor of 4.

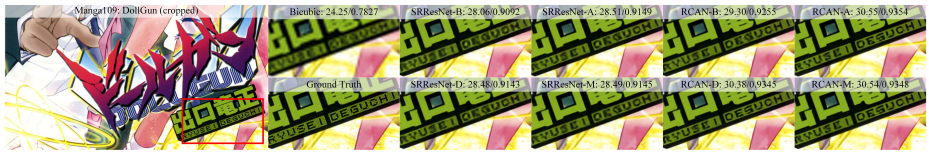

Fig. 5: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., SRResNet [5] and RCAN [11]) on Manga109 [8] at a scale factor of 4.

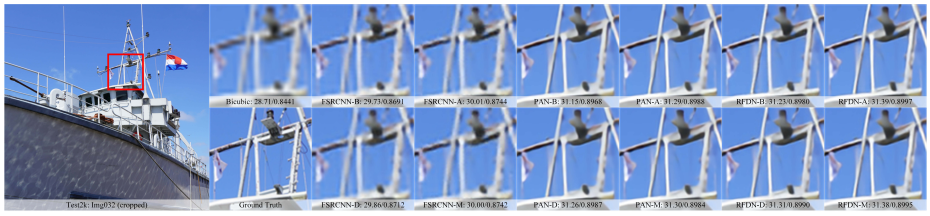

Fig. 6: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., FSRCNN [2], PAN [12], and RFDN [6]) on Test2K [3] at a scale factor of 4.

Fig. 7: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., SRResNet [5] and RCAN [11]) on Test2K [3] at a scale factor of 4.

Fig. 8: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., SRResNet [5] and RCAN [11]) on Test4K [3] at a scale factor of 4.

Fig. 9: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of existing SR networks (i.e., SRResNet [5] and RCAN [11]) on Test4K [3] at a scale factor of 4.

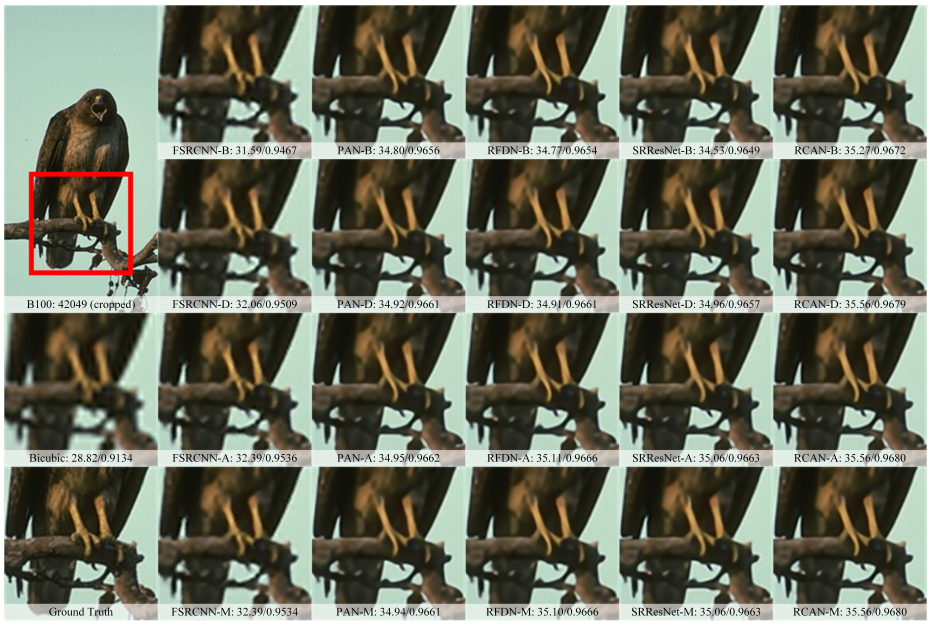

Fig. 10: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of five SR networks (i.e., FSRCNN [2], PAN [12], RFDN [6], SRResNet [5], and RCAN [11]) on B100 [7] at a scale factor of 3.

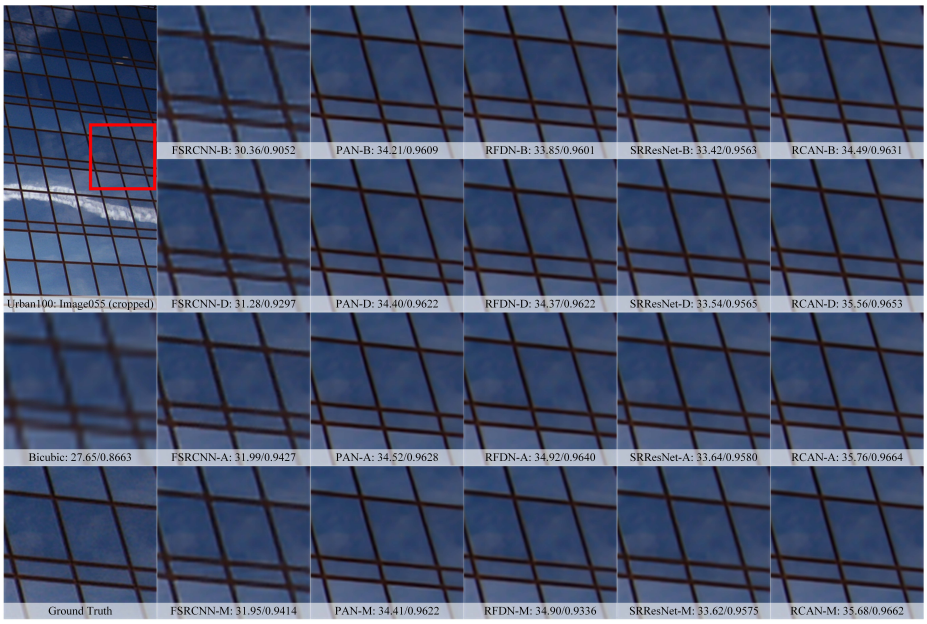

Fig. 11: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of five SR networks (i.e., FSRCNN [2], PAN [12], RFDN [6], SRResNet [5], and RCAN [11]) on Urban100 [4] at a scale factor of 3.

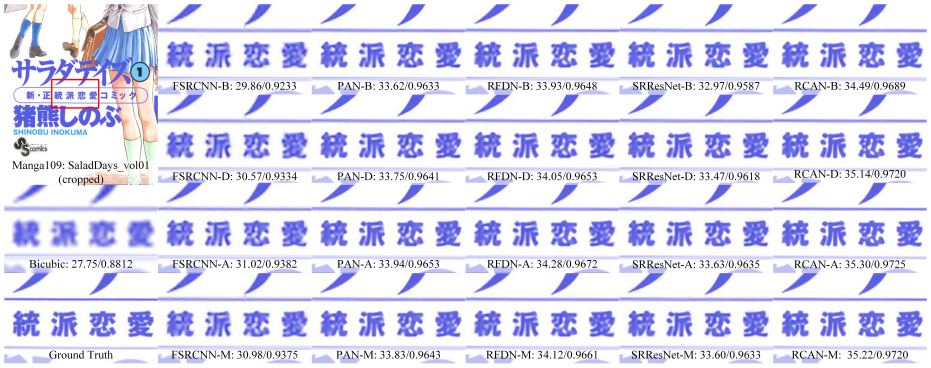

Fig. 12: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of five SR networks (i.e., FSRCNN [2], PAN [12], RFDN [6], SRResNet [5], and RCAN [11]) on Manga109 [8] at a scale factor of 3.

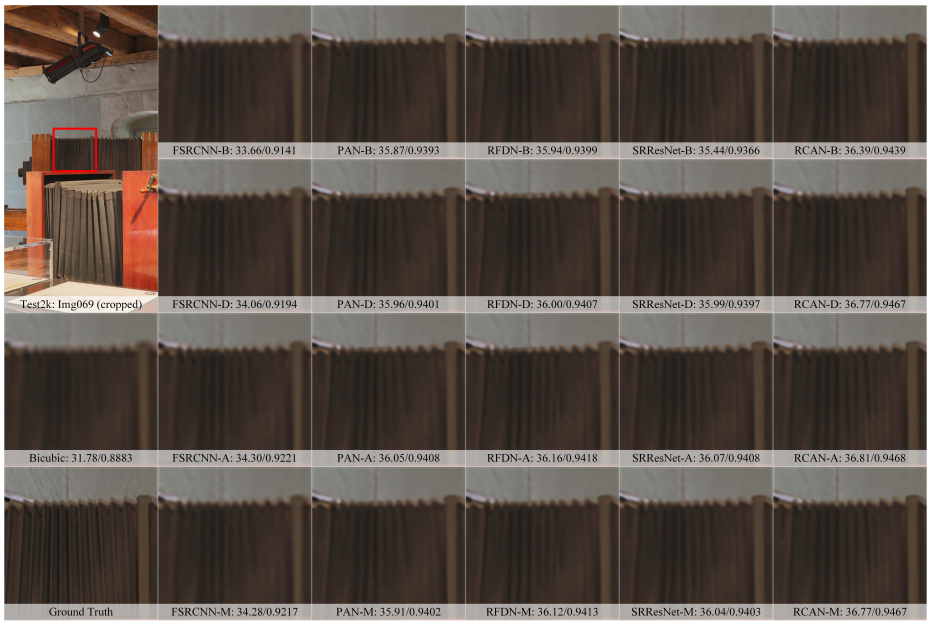

Fig. 13: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of five SR networks (i.e., FSRCNN [2], PAN [12], RFDN [6], SRResNet [5], and RCAN [11]) on Test2K [3] at a scale factor of 3.

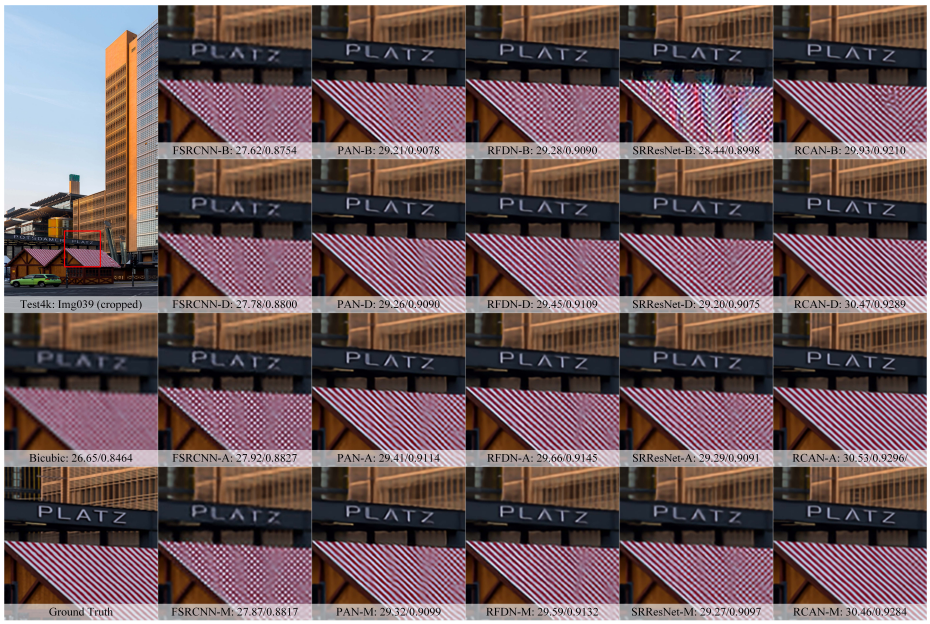

Fig. 14: **Comparison of visual quality and PSNR (dB)/SSIM results by different variants** of five SR networks (i.e., FSRCNN [2], PAN [12], RFDN [6], SRResNet [5], and RCAN [11]) on Test4K [3] at a scale factor of 3.