Reflash Dropout in Image Super-Resolution
Supplementary Materials

Xiangtao Kong[1,2,4*] Xina Liu[1,2*] Jinjin Gu[3,1,4] Yu Qiao[1,4] Chao Dong[1,4 †]
[1]ShenZhen Key Lab of Computer Vision and Pattern Recognition, SIAT-SenseTime Joint Lab,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]The University of Sydney [4]Shanghai AI Laboratory, Shanghai, China
{xt.kong, xn.liu, yu.qiao, chao.dong}@siat.ac.cn, jinjin.gu@sydney.edu.au

In this supplementary file, we first apply the proposed dropout method to SwinIR [3], which is a transformer-based SR backbone network. The experimental results show dropout is also helpful for transformer-based SR networks. Second, we provide more results of using different dropout probabilities and dropout positions under multi-degradation setting. Then, we show some training curves to illustrate that dropout does not change the convergence trend. Finally, we show more qualitative results to show the effectiveness of dropout.

A. Applying Dropout in SwinIR

SwinIR [3] is a newly proposed SR backbone network using the transformer mechanism. This model achieves state-of-the-art performance in many restoration tasks. We also apply the dropout method to this model to demonstrate that dropout is also helpful for transformer-based SR models.

We apply the dropout layer before the output convolutional layer (from 64 channels to 3 channels, `last-conv`). SwinIR also has this structure. We use the same training and testing data as Real-SRResNet and Real-RRDB for Real-SwinIR. The original setting of SwinIR that the $\times 4$ model is finetuned from the $\times 2$ model needs a too long training time. Therefore, we follow the reproduction [6] to train the models from scratch and also show the results of 250K iteration just like this reproduction. Note that, we only train the model with dropout ($p = 0.5$) to make a simple verification. This training setting and dropout probability may not be the most appropriate for SwinIR but are enough to illustrate dropout is also helpful.

The results are shown in Table A.1. When trained with dropout, Real-SwinIR obtains better PNSR performance on most of the five datasets with the tested degradations. The maximal improvement on PSNR is 0.46 dB.

B. Ablation Study on Dropout Positions and Probabilities

We propose to apply channel-wise dropout before the last convolution layer under multi-degradation in main experiments. Beside, we also provide experiments on different dropout positions and probabilities under multi-degradation.

Positions. We show the performance of Real-SRResNet with different dropout positions in Table B.3. Most dropout methods can improve the performance except `half-part` and `all-part` methods. The `last-conv` method obtains most of the best results (red text). So we chose `last-conv` method in main paper. This position is a safe and general choice, which can maintain the network capacity and improve the generalization ability. Besides, this position can be easily applied to different network structures, including Transformer, while Dropblock cannot. This simple and straightforward method can already lead to meaningful and robust results.

Probabilities. We show the performance difference of using different dropout probabilities in Table A.2. The results of Real-SRResNet with dropout probabilities from 10% to 90% are better than the results without dropout. We select $p = 0.7$ for Real-SRResNet and $p = 0.5$ for Real-RRDB. Nevertheless, other dropout probabilities are also useful. These results demonstrate that dropout methods can improve the generalization ability of SR networks stably.

*Equal contributions
†Corresponding author (e-mail: chao.dong@siat.ac.cn)

Models	Set5 [1]		Set14 [7]		BSD100 [4]		Manga109 [5]		Urban100 [2]	
	clean	blur	clean	blur	clean	blur	clean	blur	clean	blur
Real-SwinIR (p=0)	25.58	25.50	23.89	23.68	24.43	24.23	23.80	23.53	21.73	21.57
Real-SwinIR (p=0.5)	26.04	25.78	23.97	23.69	24.44	24.19	23.88	23.55	21.86	21.67
Improvement	**+0.46**	**+0.29**	**+0.08**	**+0.01**	**+0.01**	**-0.04**	**+0.08**	**+0.03**	**+0.12**	**+0.10**
	noise	jepg	noise	jepg	noise	jepg	noise	jepg	noise	jepg
Real-SwinIR (p=0)	24.40	24.03	22.97	22.71	23.40	23.34	22.83	22.27	21.20	20.95
Real-SwinIR (p=0.5)	24.64	24.32	23.10	22.86	23.42	23.40	22.79	22.34	21.35	21.11
Improvement	**+0.24**	**+0.30**	**+0.13**	**+0.15**	**+0.03**	**+0.06**	**-0.03**	**+0.07**	**+0.15**	**+0.16**
	b+n	b+j	b+n	b+j	b+n	b+j	b+n	b+j	b+n	b+j
Real-SwinIR (p=0)	23.64	23.67	22.48	22.43	22.94	23.08	22.11	21.72	20.71	20.59
Real-SwinIR (p=0.5)	23.80	23.84	22.59	22.54	22.89	23.10	22.01	21.77	20.77	20.71
Improvement	**+0.17**	**+0.17**	**+0.11**	**+0.11**	**-0.05**	**+0.02**	**-0.10**	**+0.04**	**+0.06**	**+0.12**
	n+j	b+n+j	n+j	b+n+j	n+j	b+n+j	n+j	b+n+j	n+j	b+n+j
Real-SwinIR (p=0)	23.45	22.91	22.29	21.96	22.86	22.53	21.80	21.17	20.67	20.28
Real-SwinIR (p=0.5)	23.67	23.10	22.44	22.08	22.89	22.51	21.73	21.11	20.81	20.35
Improvement	**+0.22**	**+0.19**	**+0.14**	**+0.12**	**+0.03**	**-0.02**	**-0.07**	**-0.06**	**+0.14**	**+0.07**

Table A.1. The PSNR (dB) results of Real-SwinIR with $\times 4$. Each of two columns gives a test set with 8 types of degradations. We apply bicubic, blur, noise and jpeg to generate the degradation, e.g. clean means only bicubic, noise means bicubic $\rightarrow$ noise, b+n+j means blur $\rightarrow$ bicubic $\rightarrow$ noise $\rightarrow$ jpeg.

Prob.	p=0	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
Set1	22.15	22.31	22.35	22.51	22.57	22.31
Set2	20.82	20.85	20.88	20.97	20.94	20.64

Table A.2. The performance of using different dropout probabilities for Real-SRResNet with $\times 4$. Set1 is Manga109 with noise and Set2 is Urban100 with noise (standard deviation is 20). Red/Blue text: best/second-best PSNR (dB).

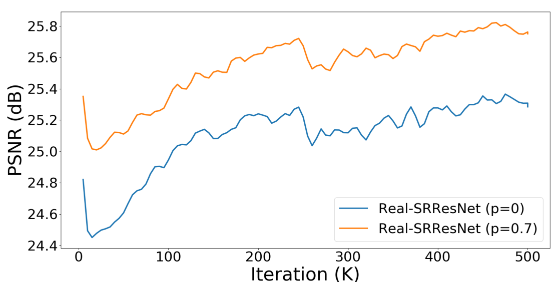


Figure C.1. Training curves of Real-SRResNet. The validation set

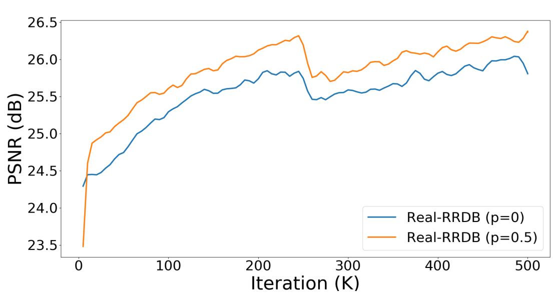


Figure C.2. Training curves of Real-RRDB. The validation set is Set5 [1] (clean).

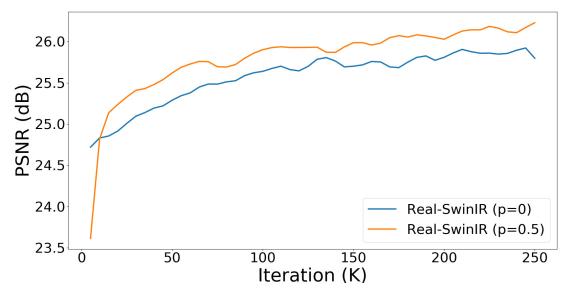


Figure C.3. Training curves of Real-SwinIR. The validation set is Set5 [1] (clean).

C. Training Curves of Models

Is the improvement in performance on account of dropout changes the convergence characteristics of networks? We visualize the training curves of Real-SRResNet (Figure C.1), Real-RRDB (Figure C.2) and SwinIR(Figure C.3). As shown in Figure C.1,C.2 and C.3, dropout does not change the convergence characteristics of the networks. During the training process, a PSNR comparison of Set5 (clean) shows that the models (both SRResNet, RRDB and SwinIR) with dropout consistently perform better than the normal models. However, they have convergence curves that are almost exactly the same.

D. More Qualitative Results

In this section, we provide additional qualitative results on different degradations to clearly show the effectiveness of dropout (see Figure D.4 to Figure D.11). Following the

	Set5		Set14		BSD100		Manga109		Urban100	
	clean	blur	clean	blur	clean	blur	clean	blur	clean	blur
Real-SRResNet (p=0)	24.89	24.76	23.24	23.04	23.89	23.67	22.97	22.59	21.23	21.06
last (p=0.7)	25.67	25.34	23.74	23.44	24.18	23.89	23.58	22.98	21.58	21.31
B4 (p=0.7)	25.68	24.93	23.95	23.40	24.25	23.74	23.49	22.61	21.66	21.08
B8 (p=0.7)	26.59	25.42	24.38	23.62	24.61	23.95	23.80	22.54	21.98	21.24
B12 (p=0.7)	26.78	25.59	24.44	23.70	24.64	24.00	23.81	22.54	22.04	21.28
B16 (p=0.7)	26.75	25.48	24.48	23.66	24.64	23.97	23.77	22.49	22.01	21.20
quarter (p=0.7)	25.10	24.66	23.22	22.87	23.64	23.32	22.77	22.07	21.04	20.74
half (p=0.7)	25.84	24.25	23.90	22.78	24.20	23.31	22.56	21.18	21.25	20.30
all (p=0.7)	25.82	24.24	23.89	22.77	24.19	23.31	22.56	21.18	21.24	20.30
	noise	jepg	noise	jepg	noise	jepg	noise	jepg	noise	jepg
Real-SRResNet (p=0)	23.75	23.70	22.51	22.31	23.01	23.03	22.15	21.75	20.82	20.59
last (p=0.7)	24.14	24.06	22.70	22.64	23.02	23.24	22.57	22.03	20.94	20.89
B4 (p=0.7)	24.00	24.05	22.73	22.84	22.93	23.36	22.35	21.92	20.83	20.96
B8 (p=0.7)	23.73	24.33	22.40	22.96	22.47	23.45	21.99	21.79	20.60	21.08
B12 (p=0.7)	23.91	24.38	22.52	22.98	22.61	23.47	22.11	21.76	20.70	21.12
B16 (p=0.7)	23.91	24.37	22.53	22.97	22.61	23.44	22.06	21.73	20.69	21.08
quarter (p=0.7)	22.97	23.50	21.71	22.15	21.90	22.85	21.47	21.09	20.06	20.43
half (p=0.7)	22.65	23.54	21.56	22.40	21.70	23.02	20.80	20.61	19.77	20.40
all (p=0.7)	22.64	23.53	21.56	22.40	21.69	23.02	20.79	20.60	19.77	20.40
	b+n	b+j	b+n	b+j	b+n	b+j	b+n	b+j	b+n	b+j
Real-SRResNet (p=0)	23.20	23.44	22.19	22.06	22.65	22.78	21.56	21.25	20.46	20.29
last (p=0.7)	23.47	23.69	22.26	22.38	22.60	22.97	21.81	21.45	20.47	20.53
B4 (p=0.7)	23.29	23.48	22.20	22.44	22.45	22.98	21.50	21.24	20.23	20.43
B8 (p=0.7)	22.84	23.57	21.73	22.46	21.92	23.00	20.96	20.95	19.88	20.43
B12 (p=0.7)	23.03	23.65	21.89	22.50	22.09	23.04	21.07	20.94	20.00	20.47
B16 (p=0.7)	23.01	23.59	21.88	22.46	22.09	23.00	21.05	20.92	19.98	20.42
quarter (p=0.7)	22.49	23.17	21.39	21.91	21.60	22.60	20.83	20.63	19.71	20.12
half (p=0.7)	21.76	22.71	20.86	21.75	21.17	22.49	19.84	19.83	19.08	19.70
all (p=0.7)	21.76	22.71	20.86	21.74	21.16	22.48	19.83	19.83	19.07	19.69
	n+j	b+n+j	n+j	b+n+j	n+j	b+n+j	n+j	b+n+j	n+j	b+n+j
Real-SRResNet (p=0)	23.17	22.75	22.01	21.74	22.67	22.39	21.37	20.82	20.41	20.09
last (p=0.7)	23.53	23.04	22.26	21.97	22.81	22.51	21.65	21.03	20.63	20.22
B4 (p=0.7)	23.50	22.95	22.39	21.98	22.85	22.46	21.42	20.79	20.61	20.09
B8 (p=0.7)	23.52	22.83	22.33	21.84	22.75	22.32	21.17	20.42	20.59	19.97
B12 (p=0.7)	23.59	22.91	22.37	21.90	22.79	22.37	21.22	20.47	20.64	20.03
B16 (p=0.7)	23.57	22.86	22.34	21.86	22.74	22.32	21.16	20.43	20.60	19.98
quarter (p=0.7)	22.97	22.68	21.85	21.64	22.46	22.21	20.77	20.33	20.21	19.89
half (p=0.7)	22.58	21.92	21.64	21.12	22.14	21.73	20.02	19.38	19.88	19.28
all (p=0.7)	22.58	21.92	21.63	21.12	22.13	21.72	20.01	19.37	19.87	19.28

Table B.3. The PSNR (dB) results of Real-SRResNet with different dropout positions. Each of two columns gives a test set with 8 types of degradations. We apply bicubic, blur, noise and jpeg to generate the degradation, e.g. clean means only bicubic, noise means bicubic $\rightarrow$ noise, b+n+j means blur $\rightarrow$ bicubic $\rightarrow$ noise $\rightarrow$ jpeg. Red/Blue text: best/second-best PSNR (dB).

testing setting, we select Gaussian blur with kernel size 21 and standard deviation 2 (denoted by "Blur"), bicubic downsampling (denoted by "Clean"), Gaussian noise with a standard deviation 20 (denoted by "Noise") and JPEG compression with quality 50 (denoted by "JPEG") as degradations to show the qualitative results. We also include complex mixed degradations that are combined by the above component. For these mixed degradations, we synthesize them in the same order as the training method.

References

[1] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 2

[2] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 2

[3] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration us-

ing swin transformer. In *IEEE International Conference on Computer Vision Workshops*, 2021. 1

[4] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 2

[5] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017. 2

[6] Xintao Wang, Ke Yu, Kelvin C.K. Chan, Chao Dong, and Chen Change Loy. BasicSR: Open source image and video restoration toolbox. https://github.com/xinntao/BasicSR, 2020. 1

[7] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 2

GT

LR (Clean)

Real-SRResNet
17.57 dB

Real-SRResNet w/
18.22 dB

Figure D.4. Visual results of “Clean”. We use “w/” to represent the model with dropout. (Zoom in for best view)

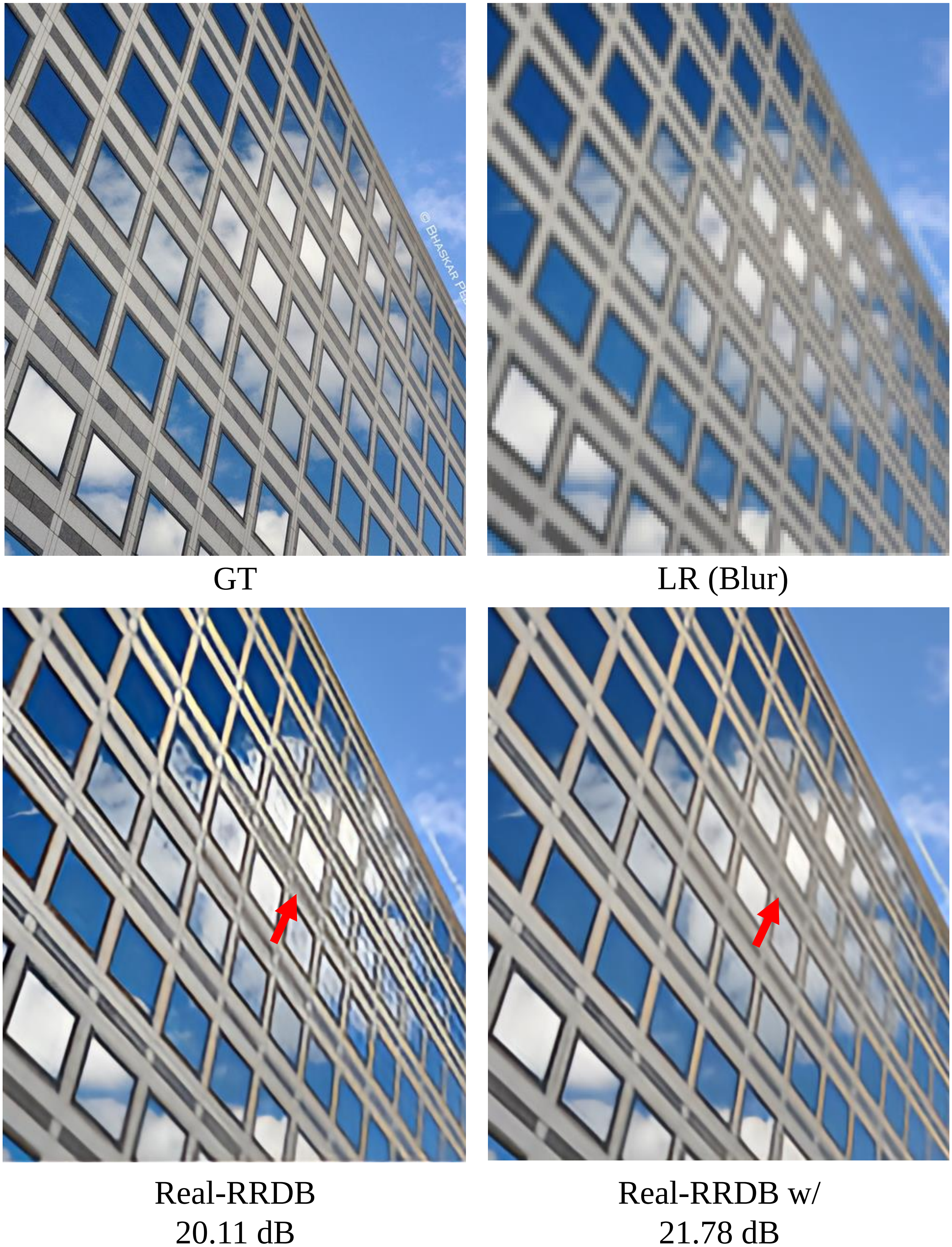
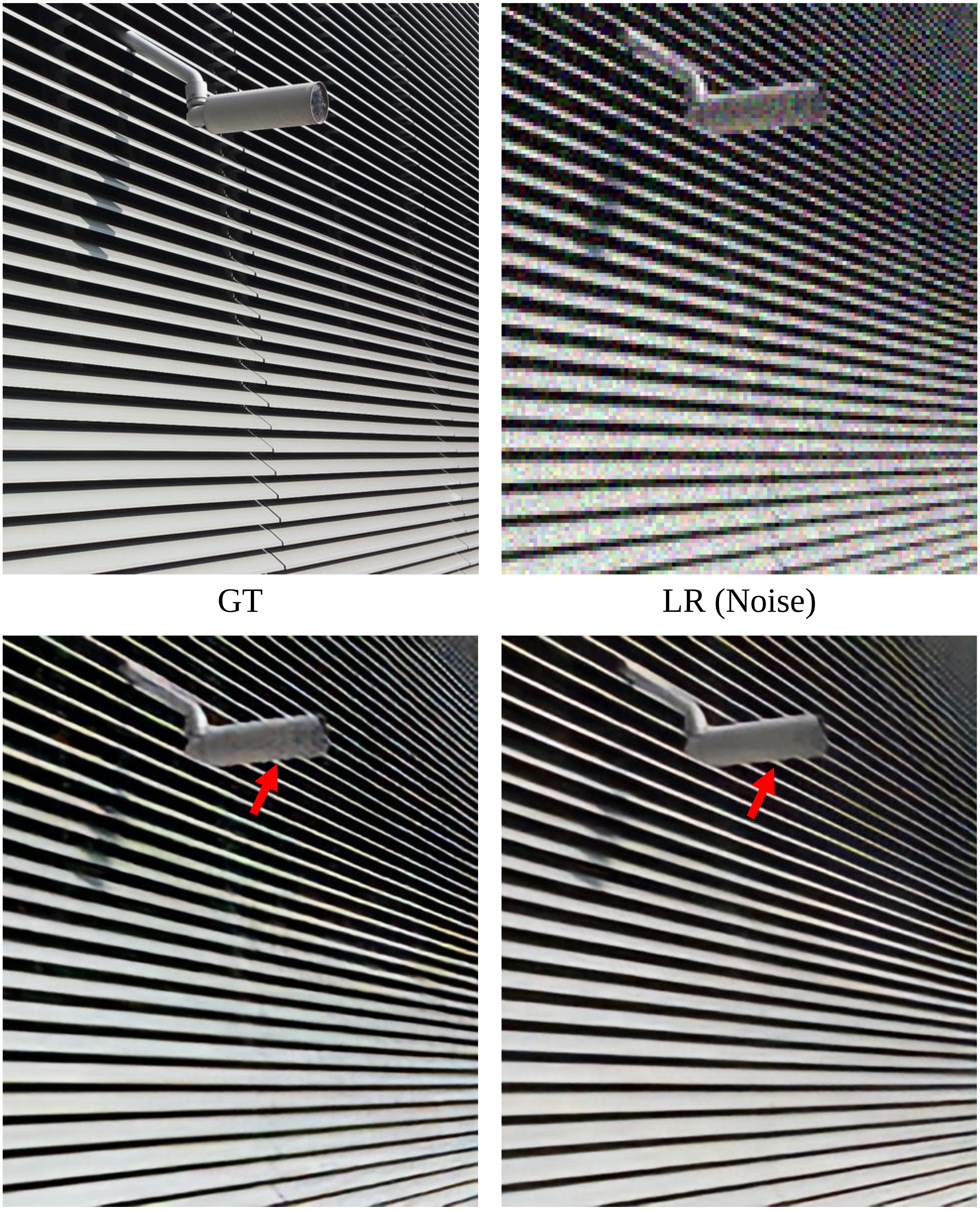


Figure D.5. Visual results of “Blur”. We use “w/” to represent the model with dropout. (Zoom in for best view)



Real-RRDB
19.92 dB

Real-RRDB w/
20.58 dB

Figure D.6. Visual results of “Noise”. We use “w/” to represent the model with dropout. (Zoom in for best view)

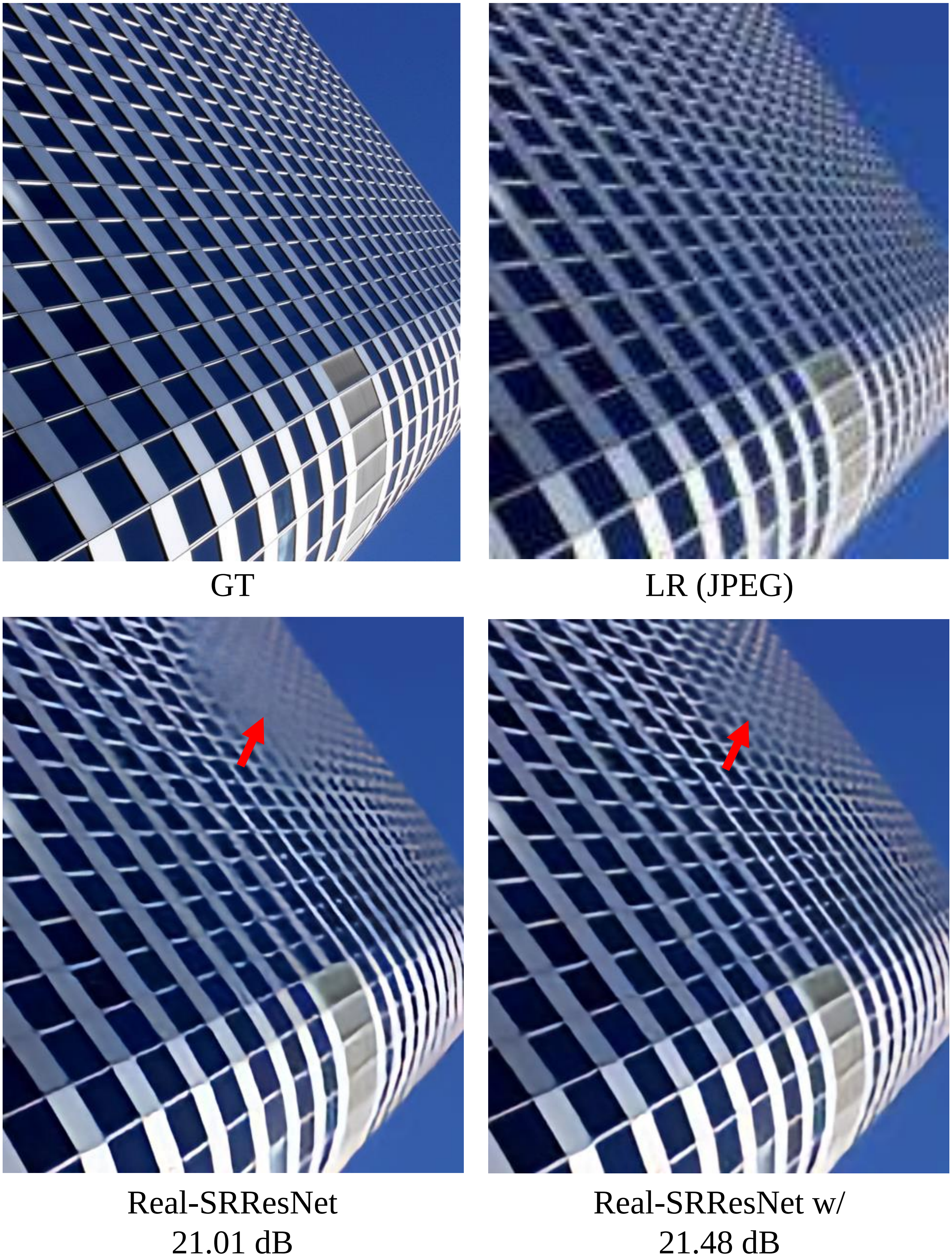
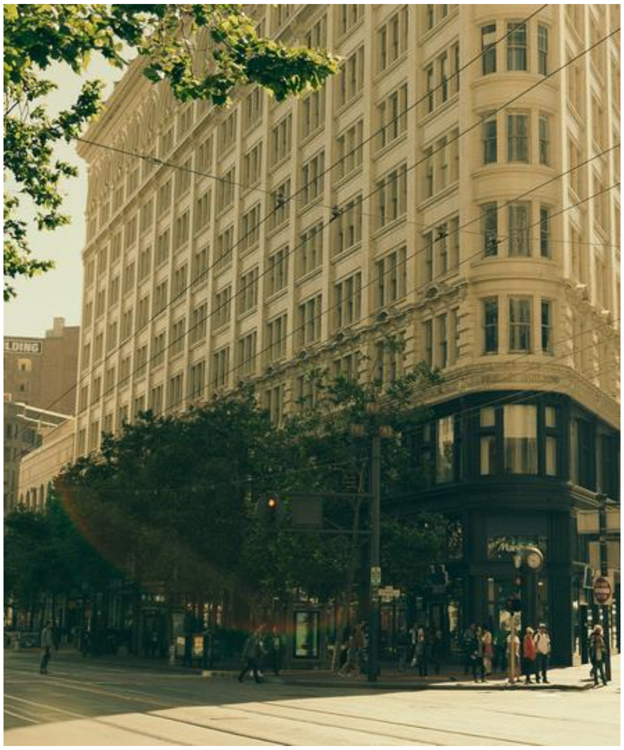


Figure D.7. Visual results of "JPEG". We use "w/" to represent the model with dropout. (Zoom in for best view)

Figure D.8. Visual results of “Blur+Noise”. We use “w/” to represent the model with dropout. (Zoom in for best view)



GT

LR (Blur+JPEG)

Real-SRResNet
20.17 dB

Real-SRResNet w/
20.62 dB

Figure D.9. Visual results of “Blur+JPEG”. We use “w/” to represent the model with dropout. (Zoom in for best view)

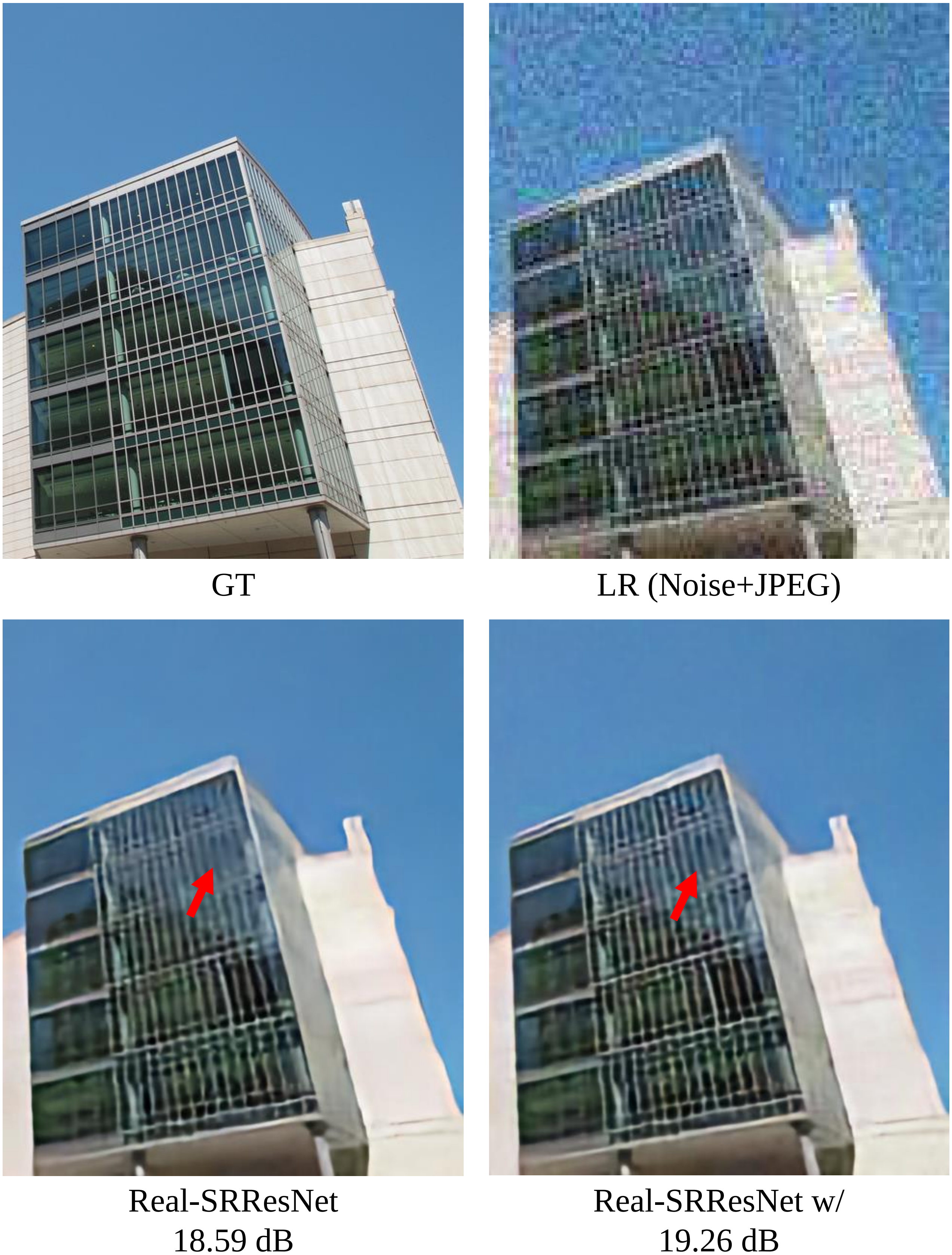
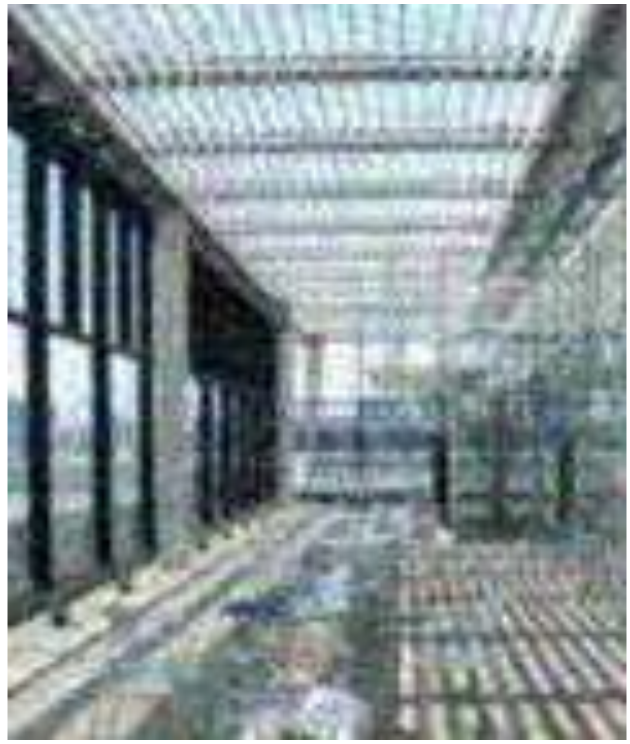


Figure D.10. Visual results of “Noise+JPEG”. We use “w/” to represent the model with dropout. (Zoom in for best view)

GT



LR (Blur+Noise+JPEG)

Real-RRDB
19.18 dB

Real-RRDB w/
19.53 dB

Figure D.11. Visual results of "Blur+Noise+JPEG". We use "w/" to represent the model with dropout. (Zoom in for best view)