Learning with Privileged Information for Efficient Image Super-Resolution Supplement

Wonkyung Lee*, Junghyup Lee*, Dohyung Kim*, and Bumsub Ham†

Yonsei University

In this supplementary material, we present an analysis of different sizes of the scale maps and the imitation loss in Sec. 1 and 2, respectively, and an ablation study on our framework including comparisons on a large dataset in Sec. 3. We also show more qualitative results of the student networks which have the same architectures as various SR methods in Sec. 4.

Table A: Average PSNR of student networks (FSRCNN [4]) with different sizes of the scale maps on the Set5 [3] dataset (2×).

Dimension of scale parameter	PSNR
$\mathbb{R}^{1\times1\times1}$	37.29 (+0.14)
$\mathbb{R}^{1\times H'\times W'}$	37.29 (+0.14)
$\mathbb{R}^{C\times1\times1}$	37.32 (+0.17)
$\mathbb{R}^{C\times H'\times W'}$	**37.33** (+0.18)

Table B: Average PSNR of student networks (FSRCNN [4]) trained with different teacher networks using various balance parameters $\lambda^{\mathcal{T}}$ on the Set5 [3] dataset (2×).

$\lambda^{\mathcal{T}}$	PSNR
0	37.23 (+0.05)
10^{-6}	37.25 (+0.10)
10^{-4}	**37.33** (+0.18)
10^{-2}	37.27 (+0.12)
1	37.25 (+0.10)

1 The size of scale map

We show in Table A performance comparisons of student networks having the same architecture as FSRCNN [4] for different sizes of scale maps in terms of the average PSNR. It shows that we can achieve the best performance, when the sizes of the scale map and a feature map in the teacher are the same. This adaptively controls the extent of distillation on each element of the feature map in the teacher network.

2 Imitation loss

We show in Table B performance comparisons of student networks having the same architecture as FSRCNN [4]. They are trained with different teacher networks that use different parameters $\lambda^{\mathcal{T}}$ for the imitation loss. The imitation loss encourages compact features $\hat{X}^{\mathcal{T}}$ extracted from HR images to be close to the LR counterparts, facilitating the initialization of a student network. We can see from the first row that the performance gain decreases without the imitation loss, suggesting that this loss is crucial for the weight transfer in our framework.

* equal contribution † corresponding author (bumsub.ham@yonsei.ac.kr)

Table C: Average PSNR of student networks, trained with variants of our framework, on the B100 [11] dataset.

Hourglass architecture	Weight transfer	$L_{\text{im}}^{\mathcal{T}}$	$L_{\text{distill}}^{\mathcal{S}}$	Student PSNR
-	-	-	-	31.58 (baseline)
✗	-	-	MAE	31.60 (+0.02)
✓	✗	✗	MAE	31.61 (+0.04)
✓	✓	✗	MAE	31.63 (+0.05)
✓	✓	✓	MAE	31.65 (+0.07)
✓	✓	✓	VID_G [2]	**31.66** (+0.08)
✓	✓	✓	VID_L [2]	31.65 (+0.07)

Table D: Average PSNR of student networks with different distillation frameworks and losses on the Set5 [3] and B100 [11] datasets.

Type	Teacher		$L_{\text{distill}}^{\mathcal{S}}$	Student PSNR	
	Model	Input		Set5 [3]	B100 [11]
Knowledge Distillaton	FSRCNN-L	LR	MAE	37.20 (+0.05)	**31.61** (+0.03)
	FSRCNN-L	LR	FitNet [12]	37.16 (+0.01)	31.59 (+0.01)
	FSRCNN-L	LR	AT [14]	37.21 (+0.06)	**31.61** (+0.03)
	FSRCNN-L	LR	SRKD [6]	37.21 (+0.06)	31.60 (+0.02)
	FSRCNN-L	LR	VID_L [2]	**37.23** (+0.08)	**31.61** (+0.03)
Proposed Distillaton	Ours	HR	MAE	37.27 (+0.12)	**31.65** (+0.07)
	Ours	HR	FitNet [12]	37.31 (+0.16)	31.64 (+0.06)
	Ours	HR	AT [14]	37.31 (+0.16)	**31.65** (+0.07)
	Ours	HR	SRKD [6]	37.29 (+0.14)	31.64 (+0.06)
	Ours	HR	VID_L [2]	**37.33** (+0.18)	**31.65** (+0.07)

If the parameter $\lambda^{\mathcal{T}}$ becomes too large (*e.g.*, 1 in the fifth row), the imitation loss forces the compact features to be identical to the LR images, which is relatively easy to achieve. In this case, our framework can be viewed as a self-distillation method [5], which however does not benefit from privilege information and degrades the performance (*e.g.*, a 0.08dB decrease in the fifth row compared to the result in the third row). Overall, the parameter $\lambda^{\mathcal{T}}$ of 10^{-4} gives a good compromise between imitated and privileged features, allowing the student network to achieve the best result.

3 Ablation studies

We present an analysis on each component of our framework using the Set5 [3] and B100 [11] datasets. We show in Table C, corresponding to Table 1 in the paper, the average PSNR on B100 [11] for student networks trained with variants of our framework. We can see that the average PSNR gradually increases by adding each component of our framework, which coincide with the findings on Set5 [3]. Table D compares the performance of the student networks using different distillation methods and losses. The first five and the last five rows show the average PSNR of the conventional knowledge distillation framework for network

compression and our framework, respectively, with different distillation losses. We can see our framework shows better performance than the knowledge distillation framework consistently in terms of PSNR on the Set5 [3] and B100 [11] datasets. This demonstrates that distilling features from the privileged information (*i.e.*, ground-truth HR images) is important, regardless of loss functions, and the proposed framework is more effective for the SISR task, boosting the performance of the student by a large margin. Note that the losses we used are typically leveraged for classification tasks to transfer 'dark knowledge' [7]. To our knowledge, there are no distillation losses effective to regression tasks. We do believe that the performance of our framework can be further boosted by the loss function specially-designed for our framework or at least for the regression tasks.

4 Qualitative results

Figures A, B, C, D, E show reconstruction examples on Set14 [15], B100 [11], and Urban100 [8] datasets using student networks, adopting the architectures of FSRCNN-L, FSRCNN [4], IDN [9], CARN [1], and VDSR [10], respectively. We can clearly see that the student networks trained with our framework consistently show better visual results compared with the original ones. Especially, our student networks accurately reconstruct sharp boundaries (*e.g.*, the alphabet in the last row in Fig. A and the marina floor in the first row in Fig. C), small-scale structures (*e.g.*, windows in the second and third rows in Fig. B, and the iron railings in the third row in Fig. C), textures (*e.g.*, the patterns of the zebra in the third row in Fig. D), and straight lines (*e.g.*, the ceiling of the bus stop in the second row in Fig. C and the bridge in the third row in Fig. E). These results indicate that our framework is generally applicable to the other SR methods, such as IDN, CARN, VDSR, FSRCNN, and the variant of FSRCNN.

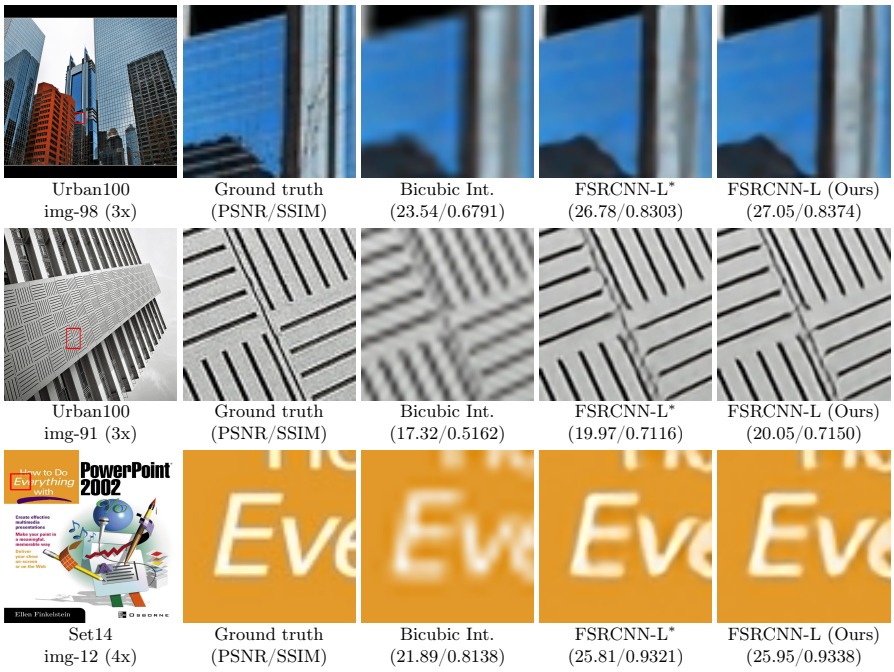

Fig. A: Visual comparison of reconstructed HR images (3× and 4×) on Set14 [15] and Urban100 [8]. We report the average PSNR/SSIM in the parentheses. Compared to the FSRCNN-L, our model reconstructs straight lines and object boundaries more accurately. *: models reproduced by ourselves using the DIV2K dataset [13] without distillation; Ours: student networks of our framework. (Best viewed in color.)

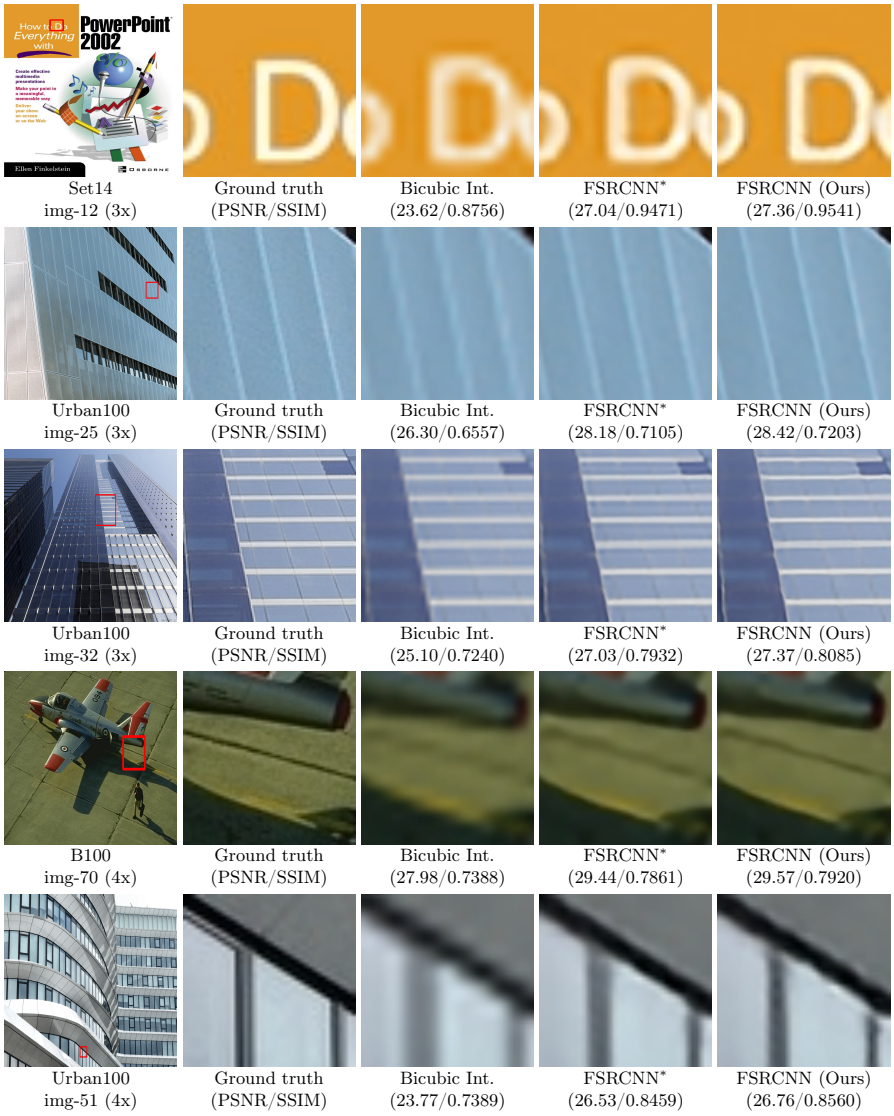

Fig. B: Visual comparison of reconstructed HR images (3× and 4×) on Set14 [15], Urban100 [8], and B100 [11]. We report the average PSNR/SSIM in the parentheses. Compared to the baseline FSRCNN, our model reconstructs small-scale structures, straight lines, and object boundaries more accurately. (Best viewed in color.)

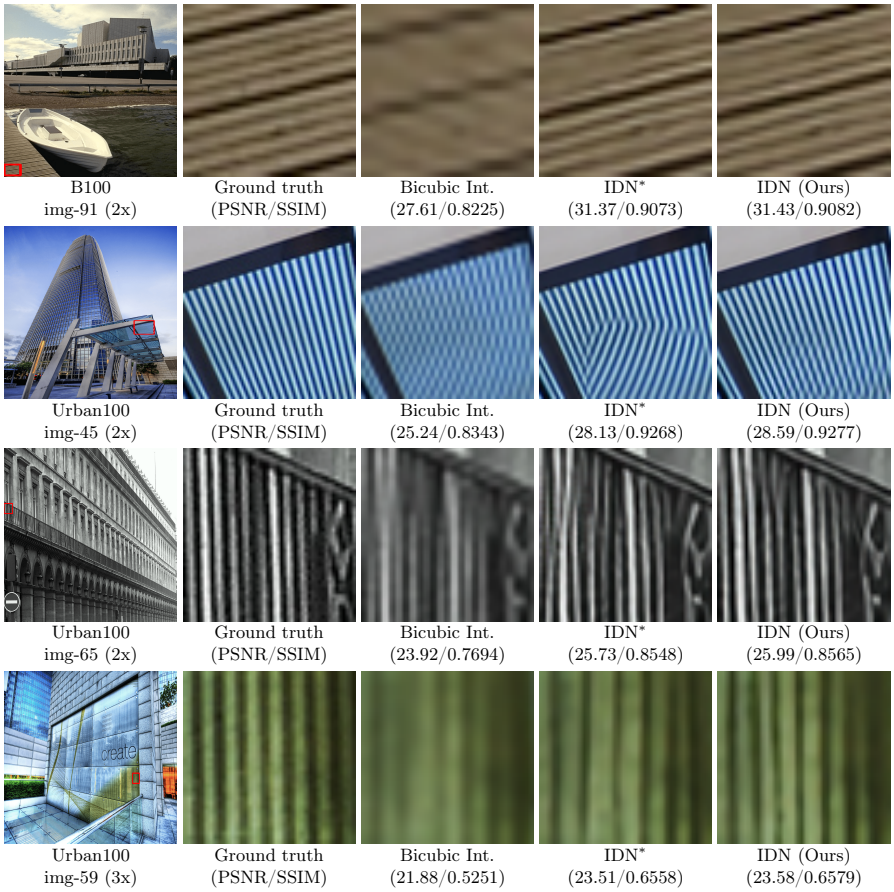

Fig. C: Visual comparison of reconstructed HR images (2× and 3×) on B100 [11] and Urban100 [8]. We report the average PSNR/SSIM in the parentheses. Compared to the baseline IDN, our model reconstructs repetitive patterns, straight lines, and object boundaries more accurately. (Best viewed in color.)

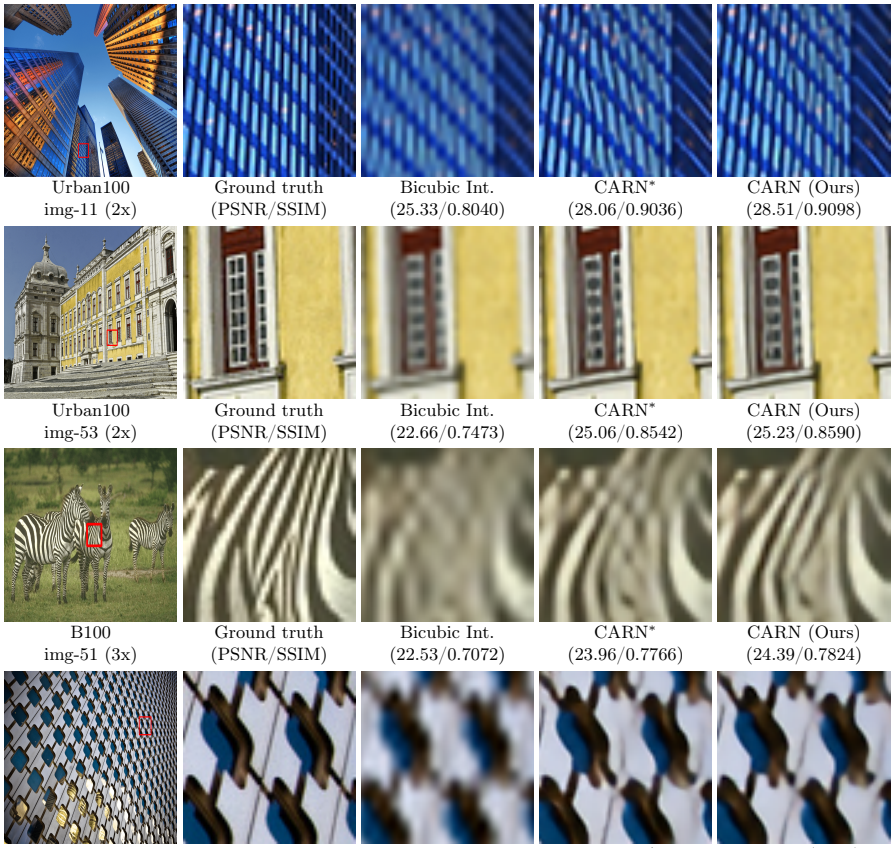

Fig. D: Visual comparison of reconstructed HR images (2×, 3×, and 4×) on Urban100 [8] and B100 [11]. We report the average PSNR/SSIM in the parentheses. Compared to the baseline CARN, our model reconstructs small-scale structures, repetitive patterns, and straight lines more accurately. (Best viewed in color.)

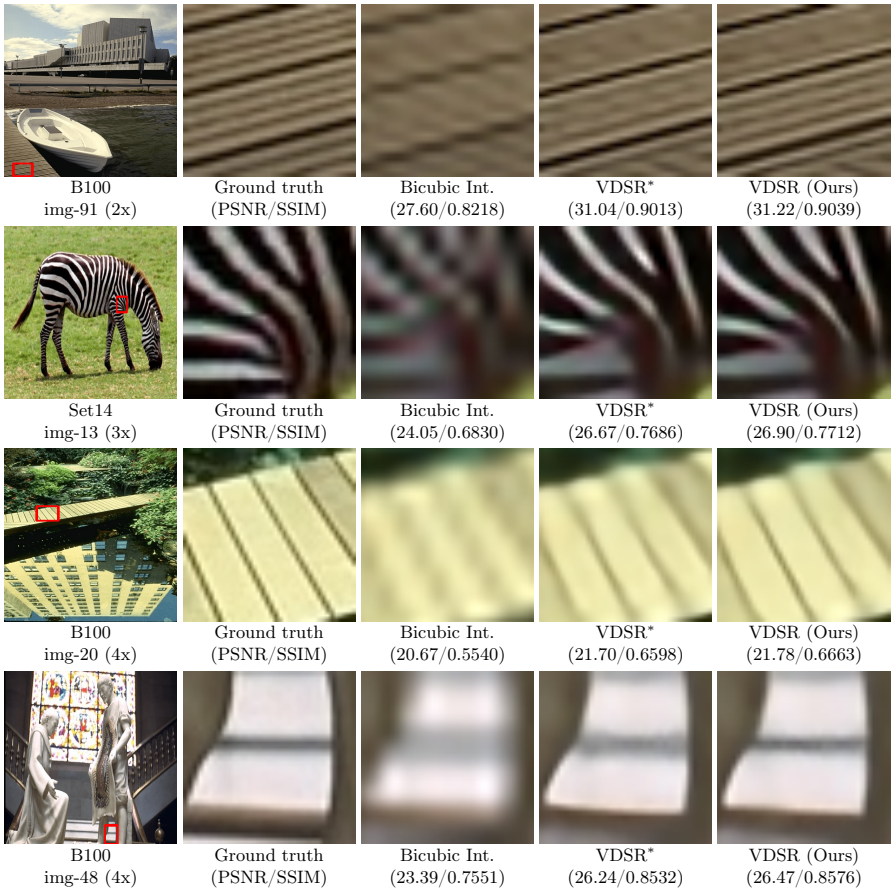

Fig. E: Visual comparison of reconstructed HR images (2×, 3×, and 4×) on B100 [11] and Ser14 [15]. We report the average PSNR/SSIM in the parentheses. Compared to the baseline VDSR, our model reconstructs small-scale structures, straight lines, and object boundaries more accurately. (Best viewed in color.)

References

1. Ahn, N., Kang, B., Sohn, K.A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: ECCV (2018)
2. Ahn, S., Hu, S.X., Damianou, A., Lawrence, N.D., Dai, Z.: Variational information distillation for knowledge transfer. In: CVPR (2019)
3. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: BMVC (2012)
4. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: ECCV (2016)
5. Furlanello, T., Lipton, Z.C., Tschannen, M., Itti, L., Anandkumar, A.: Born again neural networks. In: ICML (2018)
6. Gao, Q., Zhao, Y., Li, G., Tong, T.: Image super-resolution using knowledge distillation. In: ACCV (2018)
7. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. In: NIPS Workshop (2014)
8. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: CVPR (2015)
9. Hui, Z., Wang, X., Gao, X.: Fast and accurate single image super-resolution via information distillation network. In: CVPR (2018)
10. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: CVPR (2016)
11. Martin, D., Fowlkes, C., Tal, D., Malik, J., et al.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV (2001)
12. Romero, A., Ballas, N., Kahou, S.E., Chassang, A., Gatta, C., Bengio, Y.: FitNets: Hints for thin deep nets. In: ICLR (2015)
13. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: NTIRE 2017 challenge on single image super-resolution: Methods and results. In: CVPR Workshop (2017)
14. Zagoruyko, S., Komodakis, N.: Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. In: ICLR (2017)
15. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Curves and Surfaces (2010)