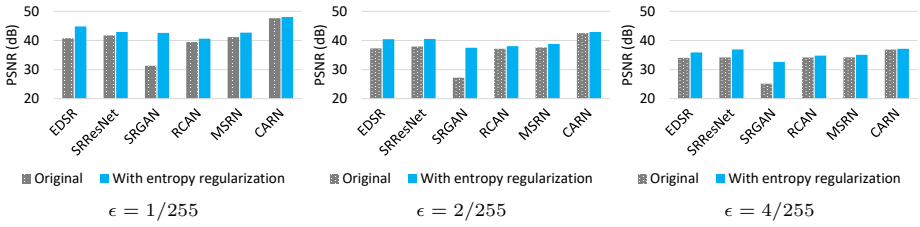
Adversarially Robust Deep Image Super-Resolution using Entropy Regularization

– Supplementary Material –

This supplementary material provides additional results of the experiments in the main paper. First, we show the attack results obtained from additional gradient-based attack methods (i.e., FGSM and PGD) in Fig. 1 similarly to those in Section 5.1 of the main paper. In addition, we show example super-resolved images obtained from the EDSR models used in our abalation study (Section 5.2 of the main paper) in Fig. 2. Finally, we show the results obtained for additional datasets including Set5 [1] and Set14 [3]. Figs. 3 and 4 compare the performance of the super-resolution methods for the FDA and I-FGSM attacks and Figs. 5 and 6 show the intermediate activation patterns with the output images (regarding Section 5.1). Figs. 7 and 8 show the ablation study results (regarding Section 5.2). Figs. 9 and 10 show the results obtained from the super-resolution models trained with different amounts of contribution of the entropy regularization (i.e., λ in the main paper, regarding Section 5.3). Figs. 11 and 12 show the results when we change the target layer for entropy regularization (regarding Section 5.4). Figs. 13 and 14 show the results obtained from the super-resolution models trained with different δ values (regarding Section 5.5). Figs. 15 and 16 compare the performance of the models trained with different defense methods (regarding Section 5.6). Overall, the results obtained on these datasets show similar observations to the results shown in the main paper for the BSD100 dataset [2].

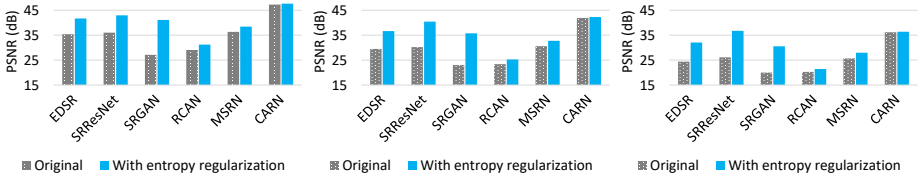
References

1. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of the British Machine Vision Conference. pp. 1–10 (2012)
2. 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: Proceedings of the IEEE International Conference on Computer Vision. pp. 416–423 (2001)
3. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Proceedings of the International Conference on Curves and Surfaces. pp. 711–730 (2010)



$\epsilon = 1/255$ $\epsilon = 2/255$ $\epsilon = 4/255$

(a) FGSM



$\alpha = 1/255$ $\alpha = 2/255$ $\alpha = 4/255$

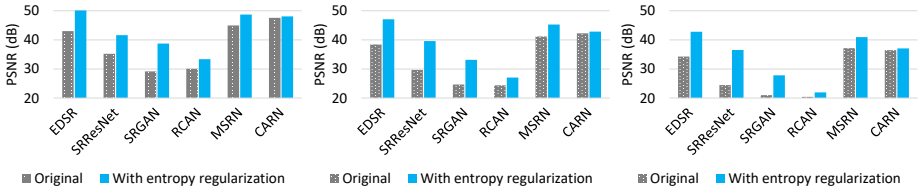
(b) PGD

Fig. 1. PSNR values of the super-resolved images for different models trained only with the original reconstruction loss (gray colors) and with both the reconstruction and entropy regularization losses (blue colors). A larger PSNR indicates better robustness. The results are obtained on BSD100 [2].

(a) (b)

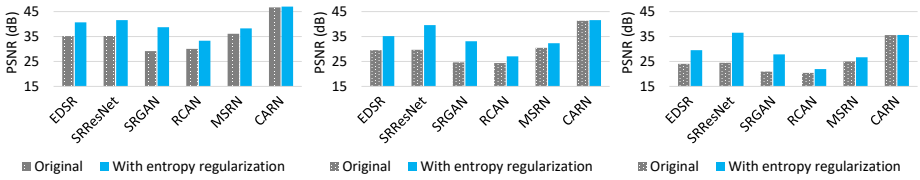
(c) (d)

Fig. 2. Images obtained from the EDSR models trained (a) without any defense, (b) with probability estimation, (c) with random noise, and (d) both. FDA ($\epsilon = 8/255$) is employed as the attack method.



$\epsilon = 1/255$ $\epsilon = 2/255$ $\epsilon = 4/255$

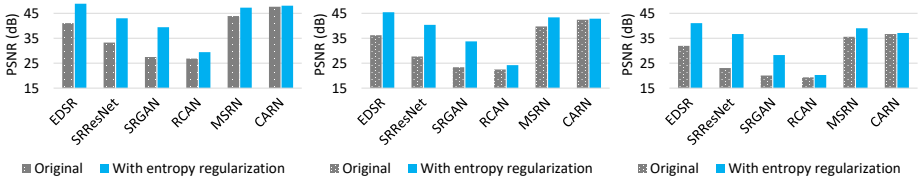
(a) FDA



$\alpha = 1/255$ $\alpha = 2/255$ $\alpha = 4/255$

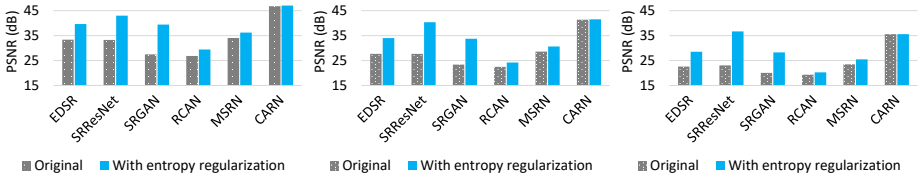
(b) I-FGSM

Fig. 3. PSNR values of the super-resolved images for different models trained only with the original reconstruction loss (gray colors) and with both the reconstruction and entropy regularization losses (blue colors). A larger PSNR indicates better robustness. The results are obtained on Set5 [1].



$\epsilon = 1/255$ $\epsilon = 2/255$ $\epsilon = 4/255$

(a) FDA



$\alpha = 1/255$ $\alpha = 2/255$ $\alpha = 4/255$

(b) I-FGSM

Fig. 4. PSNR values of the super-resolved images for different models trained only with the original reconstruction loss (gray colors) and with both the reconstruction and entropy regularization losses (blue colors). A larger PSNR indicates better robustness. The results are obtained on Set14 [3].

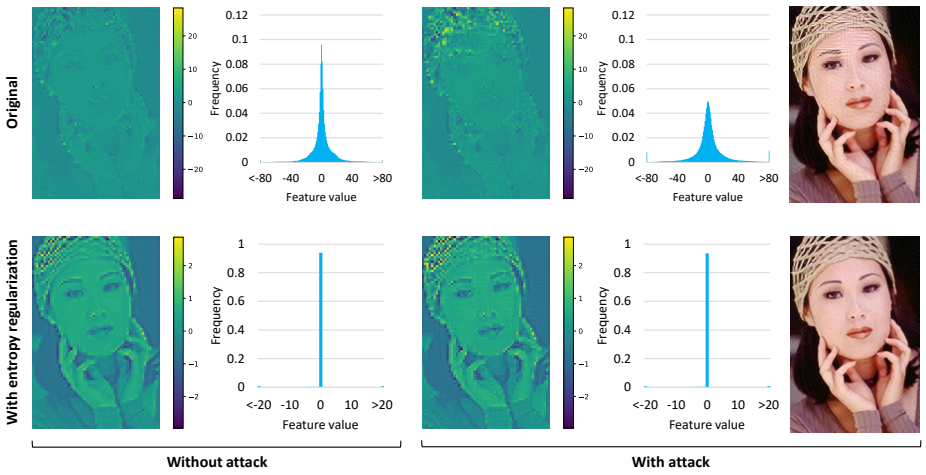


Fig. 5. Intermediate features, histograms of the intermediate feature values, and output images of the EDSR models trained without and with entropy regularization. FDA with $\epsilon = 8/255$ is employed as the attack method. The input image is from Set5 [1].

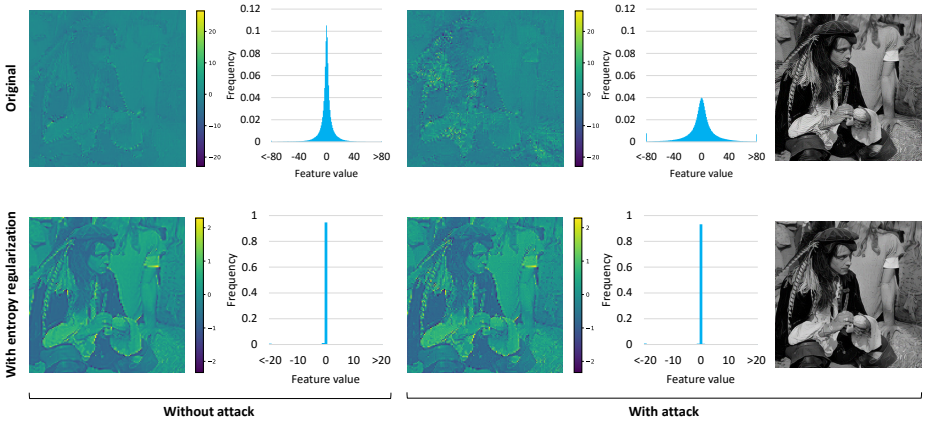
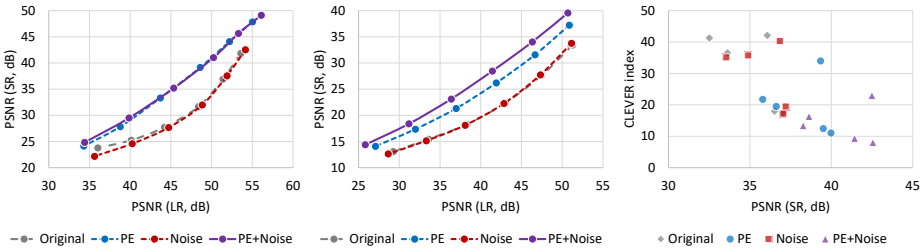
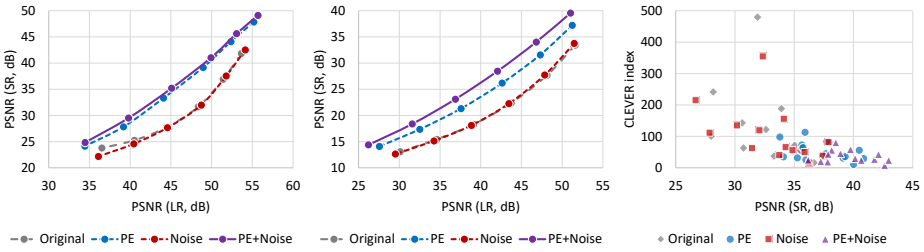


Fig. 6. Intermediate features, histograms of the intermediate feature values, and output images of the EDSR models trained without and with entropy regularization. FDA with $\epsilon = 8/255$ is employed as the attack method. The input image is from Set14 [3].



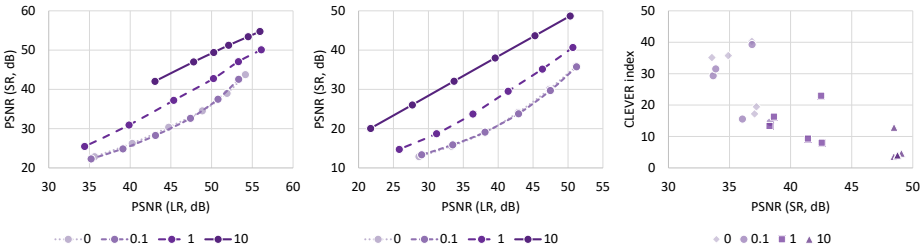
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 7. Performance comparison on Set5 [1] in terms of PSNR and CLEVER index values for the EDSR models trained without any defense (Original), with probability estimation (PE), with random noise (Noise), and both (PE+Noise).



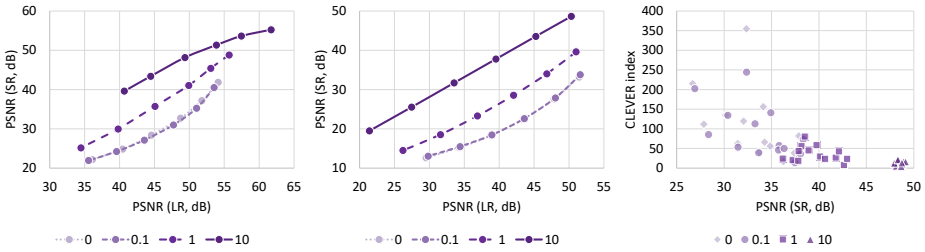
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 8. Performance comparison on Set14 [3] in terms of PSNR and CLEVER index values for the EDSR models trained without any defense (Original), with probability estimation (PE), with random noise (Noise), and both (PE+Noise).



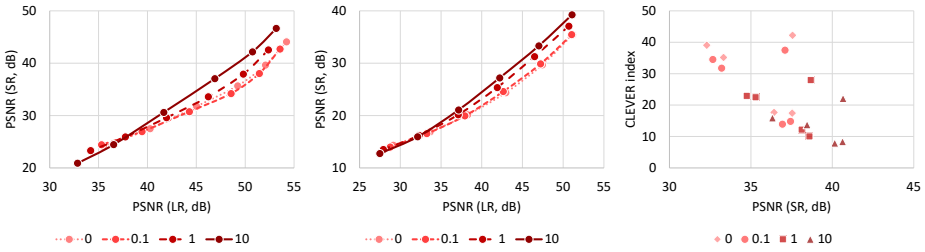
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 9. Performance comparison on Set5 [1] in terms of PSNR and CLEVER index values for the EDSR models trained with different values of λ.



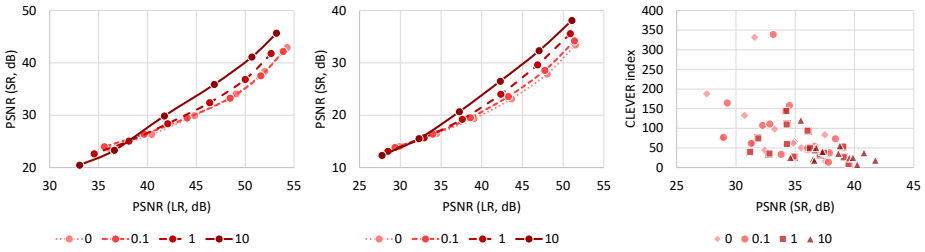
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 10. Performance comparison on Set14 [3] in terms of PSNR and CLEVER index values for the EDSR models trained with different values of λ.



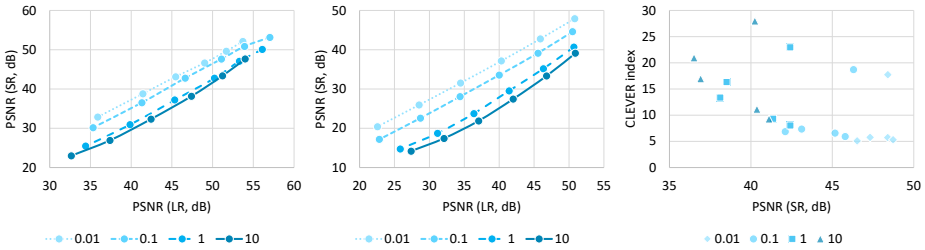
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 11. Performance comparison on Set5 [1] in terms of PSNR and CLEVER index values for the EDSR models trained with different values of λ, where the entropy regularization loss is applied to the first convolutional layer.



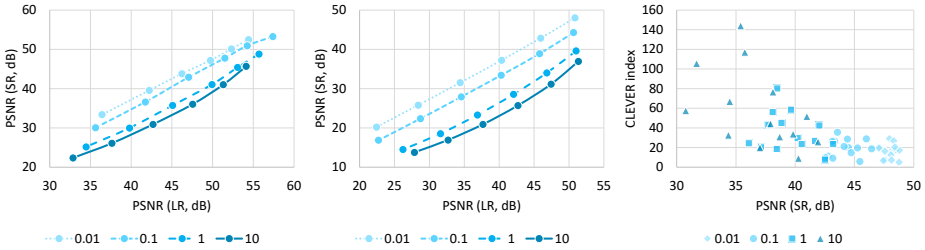
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 12. Performance comparison on Set14 [3] in terms of PSNR and CLEVER index values for the EDSR models trained with different values of λ, where the entropy regularization loss is applied to the first convolutional layer.



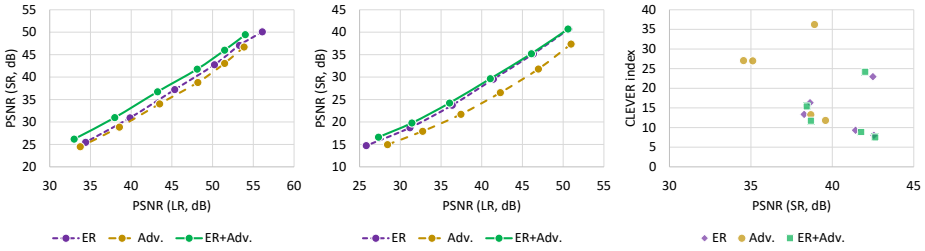
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 13. Performance comparison on Set5 [1] in terms of PSNR and CLEVER index values for the EDSR models trained with different values of δ.



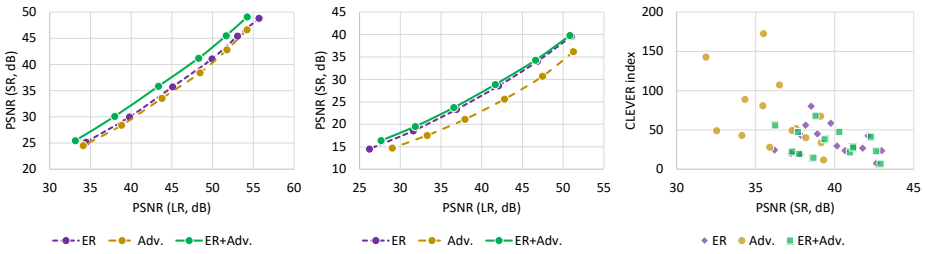
(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 14. Performance comparison on Set14 [3] in terms of PSNR and CLEVER index values for the EDSR models trained with different values of δ.



(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 15. Performance comparison on Set5 [1] in terms of PSNR and CLEVER index values for the EDSR models trained with entropy regularization (ER), adversarial training (Adv.) and both (ER+Adv.).



(a) PSNR (FDA) (b) PSNR (I-FGSM) (c) CLEVER

Fig. 16. Performance comparison on Set14 [3] in terms of PSNR and CLEVER index values for the EDSR models trained with entropy regularization (ER), adversarial training (Adv.) and both (ER+Adv.).