Fast Bayesian Uncertainty Estimation and Reduction of Batch Normalized Single Image Super-Resolution Network
(Supplementary Material)

Aupendu Kar, Prabir Kumar Biswas
Indian Institute of Technology Kharagpur, WB, India
Webpage: aupendu.github.io/sr-uncertainty
mailtoaupendu@gmail.com, pkb@ece.iitkgp.ac.in

1. Adversarial Attack Mechanism

The modified version of the Iterative Fast Gradient Sign Method (I-FGSM) is used to perform the adversarial attack on the super-resolution model [3]. The main goal of the adversarial attack on the super-resolution model is to perturb the low-resolution (LR) image so that the super-resolution model fails to perform and creates unwanted artifacts. Following the paper, we use two types of attacks: Basic Attack and Partial Attack. We adversarially perturb the whole LR image in the Basic Attack and a small portion of the LR image for Partial Attack.

Basic Attack: Let $f(\cdot)$ is the super-resolution model. I_{L_0} is the LR image, and I_L is the adversarially perturbed LR image. $f(I_{L_0})$ and $f(I_L)$ are the super-resolved images. The main objective is to maximize the difference between $f(I_{L_0})$ and $f(I_L)$. It can be defined as:

$$\mathcal{L}(I_L, I_{L_0}) = \| f(I_L) - f(I_{L_0}) \|_2 \quad (1)$$

I_L can be iteratively updated by using the I-FGSM update rule by:

$$\widetilde{I}_L(N+1) = clip_{0,1}(I_L(N) + \tfrac{\alpha}{T} sign(\nabla \mathcal{L}(I_L(N), I_{L_0}))) \quad (2)$$

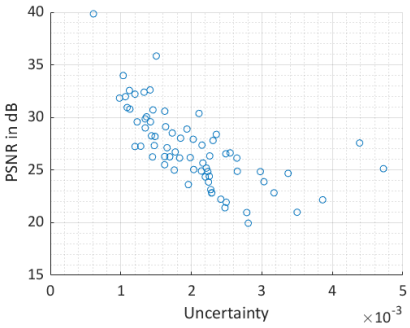
$$I_L(N+1) = clip_{-\alpha,\alpha}(\widetilde{I}_L(N+1) - I_{L_0}) + I_{L_0} \quad (3)$$

Here, T is the number of iterations, $sign(\nabla \mathcal{L}(I_L(N), I_{L_0}))$ is the sign of the gradient, $clip_{0,1}$ clip the pixel values in between 0 and 1, and α is the adversarial noise level.

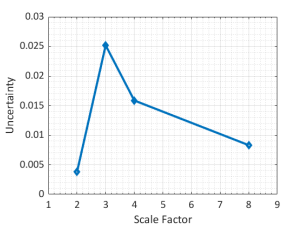
Partial Attack In case of a partial attack, we restrict the attack region and add adversarial noise in that specific region. Let M is the masked region, where the area that will be attacked is set to 1, and the rest is 0. Therefore, the new I-FGSM update rule is defined as,

$$\widetilde{I}_L(N+1) = clip_{0,1}(I_L(N) + \tfrac{\alpha}{T} sign(\nabla \mathcal{L}(I_L(N), I_{L_0})) \circ M) \quad (4)$$

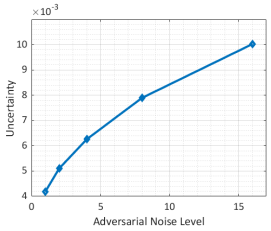
where $\circ$ is the pixel-wise multiplication.



(a) PSNR vs Uncertainty (UN)



(b) UN vs Scale



(c) UN vs Adv. Noise Level

Figure 1. (a) presents the relation between PSNR and Bayesian uncertainty. (b) presents the influence of uncertainty for different scale factor models. (c) presents impact of adversarial noise levels with uncertainty. (zoom for the best view.)

2. Additional Experiments: Residual Network

We perform similar experiments on another deep neural architecture like VDSR, as described in the main paper. For this purpose, we use the popular SR-ResNet architecture [7], which consists of consecutive residual blocks.

2.1. Network Architecture

The residual network is one of the most popular architecture for image classification. SR-ResNet [7] uses the consecutive residual blocks for super-resolution. We adopt

Dataset	Attack Level	Uncertainty Scale Factor $\times 3$/ $\times 4$/ $\times 8$		PSNR Scale Factor $\times 3$/ $\times 4$/ $\times 8$	
		No Defense	After Defense	No Defense	After Defense
BSDS100	No Attack	0.0252/ 0.0159/ 0.0083	0.0019/ 0.0016/ 0.0018	27.57/ 26.61/ 24.56	27.88/ 27.23/ 24.56
	1	0.0225/ 0.0153/ 0.0075	0.0026/ 0.0019/ 0.0021	26.04/ 25.53/ 23.80	27.60/ 26.96/ 24.37
	2	0.0197/ 0.0146/ 0.0075	0.0028/ 0.0023/ 0.0024	23.78/ 24.08/ 22.66	27.45/ 26.70/ 24.28
	4	0.0282/ 0.0202/ 0.0103	0.0031/ 0.0027/ 0.0029	20.27/ 21.25/ 20.50	26.88/ 26.24/ 24.07
	8	0.0270/ 0.0217/ 0.0121	0.0036/ 0.0033/ 0.0035	17.25/ 18.21/ 18.03	26.14/ 25.50/ 23.66
	16	0.0308/ 0.0290/ 0.0156	0.0048/ 0.0045/ 0.0045	14.97/ 15.28/ 15.50	24.77/ 24.17/ 22.85
Urban100	No Attack	0.0257/ 0.0325/ 0.0119	0.0031/ 0.0030/ 0.0034	26.21/ 24.17/ 22.02	25.95/ 25.03/ 21.94
	1	0.0276/ 0.0344/ 0.0168	0.0038/ 0.0032/ 0.0038	24.66/ 23.02/ 20.94	25.54/ 24.68/ 21.63
	2	0.0291/ 0.0369/ 0.0156	0.0039/ 0.0036/ 0.0042	22.67/ 21.51/ 19.82	25.34/ 24.17/ 21.45
	4	0.0331/ 0.0392/ 0.0201	0.0040/ 0.0038/ 0.0047	19.71/ 19.26/ 17.98	24.93/ 23.99/ 21.23
	8	0.0303/ 0.0503/ 0.0192	0.0043/ 0.0046/ 0.0051	16.73/ 16.69/ 16.00	24.41/ 23.27/ 20.93
	16	0.0320/ 0.0523/ 0.0222	0.0056/ 0.0057/ 0.0061	14.31/ 14.20/ 14.02	23.24/ 22.25/ 20.37
Manga109	No Attack	0.0170/ 0.0123/ 0.0081	0.0024/ 0.0025/ 0.0030	31.26/ 28.89/ 24.02	31.50/ 29.29/ 23.78
	1	0.0142/ 0.0118/ 0.0085	0.0030/ 0.0028/ 0.0032	29.44/ 27.68/ 23.39	30.59/ 28.87/ 23.59
	2	0.0165/ 0.0143/ 0.0104	0.0036/ 0.0033/ 0.0034	26.85/ 25.87/ 22.41	29.36/ 27.88/ 23.26
	4	0.0258/ 0.0160/ 0.0097	0.0044/ 0.0038/ 0.0039	22.94/ 22.94/ 20.80	28.15/ 26.89/ 22.84
	8	0.0294/ 0.0185/ 0.0124	0.0051/ 0.0046/ 0.0046	19.09/ 19.50/ 18.47	26.77/ 25.71/ 22.24
	16	0.0267/ 0.0285/ 0.0147	0.0059/ 0.0058/ 0.0057	16.11/ 16.22/ 15.95	24.88/ 23.89/ 21.31

Table 1. Quantitative evaluation of our proposed Bayesian uncertainty reduction technique based adversarial defense mechanism.

the SR-ResNet architecture with some minor modifications. In our architecture, the residual block consists of two consecutive convolution layers and each followed by ReLU and batch-normalization. There are 16 consecutive residual blocks in the network.

2.2. Training Details

We use the DIV2K dataset [1, 10] for training, which contains 800 training images and 100 images for validation. Five standard benchmark testing datasets, namely Set5 [2], Set14 [11], BSD100 [8], Urban100 [4], Manga109 [9] are used for performance analysis. We randomly extract patches of size 32×32 from each LR image during training for a batch update. Each batch contains 16 patches. We augment the patches by horizontal flip, vertical flip, and 90-degree rotation and randomly choose each augmentation with a 50% probability. Each input patch is normalized into $[0, 1]$ and $[0.4488, 0.4371, 0.4040]$ is subtracted channel-wise before feeding to the network. We train each model with the PyTorch framework for 1000 epochs, where a single epoch constitutes 1000 batch updates. Adam optimizer [6] is used to update the weights. The learning rate is initialized to 10^{-4} and reduced to half after every 200 iterations. We use mean-squared error to optimize model parameters.

2.3. Behaviour of Uncertainty

Like VDSR architecture [5], as described in the main paper, the Figure 1(a) shows a strong relationship between reconstructed image quality and uncertainty in reconstruction. BSDS100 dataset is used for this experiment, and the experiment is performed for scale factor $\times 4$. We observe that images with lower PSNR, which represents the image quality, give higher uncertainty. Figure 1(b) shows the relation between scale factor and uncertainty. Unlike VDSR, we do not observe any relationship between those. In Figure 1(c), We witness that images with more adversarial perturbation lead to higher uncertainty and experiment is performed for scale factor $\times 2$.

2.4. Defense Against Adversarial Attack

The performance of our proposed adversarial defense mechanism on the SR-ResNet model is shown in Table 1 using the three most popular testing datasets, namely BSDS100, Urban100, and Manga109. Like the main paper, we use the five different adversarial attack levels, as shown in [3], to show the efficacy of the defense mechanism. We also show the performance of the defense mechanism on the images which are not perturbed adversarially. The performances are shown for three different scale factors, that is, $\times 3$, $\times 4$, $\times 8$.

Like VDSR architecture in the main paper, Table 1 shows similar performance in preventing adversarial attacks while we use SR-ResNet architecture and several datasets like BSDS100, Manga109, and Urban100. Our proposed defense mechanism performs well to improve the reconstruction performance in adversarially attacked images. As the attack level increases, the performance drops drastically. Our proposed method successfully prevents that performance drop.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017. 2

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

[3] Jun-Ho Choi, Huan Zhang, Jun-Hyuk Kim, Cho-Jui Hsieh, and Jong-Seok Lee. Evaluating robustness of deep image super-resolution against adversarial attacks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 303–311, 2019. 1, 2

[4] 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

[5] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 2

[6] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *ICLR*, 2014. 2

[7] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1

[8] D Martin, C Fowlkes, D Tal, and J 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

[9] 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

[10] Radu Timofte, Shuhang Gu, Jiqing Wu, Luc Van Gool, Lei Zhang, Ming-Hsuan Yang, Muhammad Haris, et al. Ntire 2018 challenge on single image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 2

[11] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010. 2