EXAGGERATED LEARNING FOR CLEAN-AND-SHARP IMAGE RESTORATION

Chang Liu[⋆§], *Qifan Gao*[⋆§], *Xiaolin Wu*[⋆†]

⋆ Shanghai Jiao Tong University
† McMaster University

ABSTRACT

Deep learning has become a methodology of choice for image restoration tasks, including denoising, super-resolution, deblurring, exposure correction, etc., because of its superiority to traditional methods in reconstruction quality. However, the published deep learning methods still have not solve the old dilemma between low noise level and detail sharpness. We propose a new CNN design strategy, called exaggerated deep learning, to reconcile two mutually conflicting objectives: noise free and detail sharpness. The idea is to deliberately overshoot for the desired attributes in the CNN optimization objective function; the cleanness or sharpness is overemphasized according to different semantic contexts. The exaggerated learning approach is experimented on the restoration tasks of super-resolution and low light correction. Its effectiveness and advantages have been empirically affirmed.

Index Terms— Exaggerated deep learning, image restoration.

1. INTRODUCTION

In addition to its prowess in recognition and classification tasks based on visual signals, deep convolutional neural networks (DCNN) have also had great successes in a wide range of image restoration applications, including super-resolution[1, 2], inpainting[3, 4], low light correction[5], etc. In general, the published DCNN image restoration methods outperform their counterparts of traditional image processing in terms of image quality, thanks to the use of large data and much increased computational power. However, still in overcoming one technical dilemma, the deep learning based image restoration methods have not yet made meaningful improvement over their predecessors; that is, the mutual conflict between two perceptual attributes associated with high image quality: the absence of noises and the clarity of details.

Different objective functions of the image restoration DCNNs guide them to generate different restoration results. Just like traditional methods most of these DCNN methods use signal distortion or fidelity measures, such as MSE and SSIM, as their objective functions [5, 1, 3]. As these metrics measure image quality in some average sense and favor the averages of plausible solutions, the so-designed networks tend to generate overly-smooth images. A popular technique to overcome the smoothing problem is to incorporate a generative adversary neural network (GAN) architecture into the network [6], and add an adversarial loss term to the objective function[2, 4]. The adversarial loss is not a sample-wise fidelity measure but a diversity measure between the manifolds of the restored images and the referenced images; as such, it penalizes unnatural smoothness and and produces relatively sharp details. But the GAN approach has an unwanted side effect: contaminating originally smooth areas with high frequency artifacts, as shown in Fig. 1.

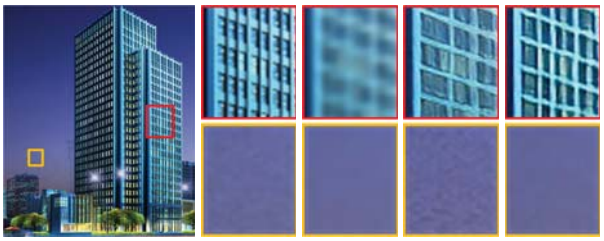


Fig. 1: The GAN-based SR method ESRGAN[2] generates noises in smooth areas and false edges. In contrast, the proposed method produces a clean and sharp image.

In this research, we take on the challenge and develop new techniques to make restored images simultaneously clean and sharp, or what we call the CAS property. To counter the smoothing and noising effects of the existing learning methods, we propose an overshoot strategy that exaggerates the desired features of the ground truth image in the training stage. This is a CAS preprocessing to idealize the ground truth image X and generate a cartoon like CAS version $X_{\sqcup}$. In $X_{\sqcup}$ we boost the contrast in regions on and near semantically significant edges; at the same time we denoise smooth regions and make them extra clean. In other words, the enhanced ground truth image $X_{\sqcup}$ is purposely polarized, sharper than X in high activity regions and smoother than X in smooth regions. In the CAS preprocessing, the CAS sharpening and smoothing are discriminatingly carried out according to image semantics. The

§These authors contribute equally to this work.

semantic adaptation is realized by combining edge analysis and Wiener-type filtering. The preprocessed CAS image $X_{\sqcup}$ is used as the ground truth to train the restoration network so that it can generate clean and sharp results. The proposed CAS approach is general, and it can be applied to improve any existing CNNs for image restoration, regardless of the specific restoration tasks.

2. METHODOLOGY

2.1. CAS preprocessing

In order to cancel the smoothing effects of using Euclidean sample-wise metric in image restoration, we would like to exaggerate the high frequency features in the target image (i.e., the so-called ground truth image for machine learning methods). However, operators that boost high frequency image features such as edges and textures also magnify noises. This is a dilemma for all image enhancement and restoration algorithms, because unless synthesized by computers, noises are ever present in acquired digital images.

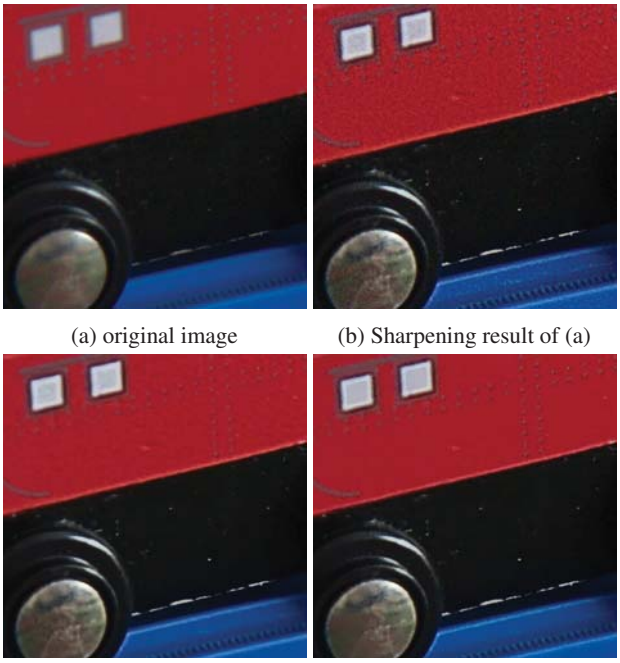
Sharpening is high-pass filtering in nature, whereas denoising is low-pass. It is difficult to achieve the two goals simultaneously in one operation. Instead, we propose a two-step CAS process. Before being used to train the restoration CNN, the ground truth images are sharpened by a matured image processing tool (e.g., the smart sharpen filter of Photoshop). The sharpening is performed on the luminance channel of X only, generating the over-sharpened image X_s. As the sharpening operation also magnifies noises, we need to remove noises in image X_s. But the denoising cannot blur semantically and perceptually important edges, otherwise the sharpening effects would be canceled. To this end, we propose a Weiner-type denoising filter that adjusts the strength of its low-pass filtering according to a soft edge map M. The map M is generated as follows. We use the Canny edge detector $\mathbb{E}$ to construct a binary edge map $\mathbb{E}(X)$ [7], and then compute M by convoluting the binary image $\mathbb{E}(X)$ with a Gaussian kernel h:

$$M = h * \mathbb{E}(X). \tag{1}$$

The resulting soft edge map M has pixel value $M(i)$ that decays from 1 to 0 as position i moves away from the edge locations. Based on M, we apply an edge-aware Wiener-type denoiser to the sharpened image X_s to compute the CAS image:

$$X_{\sqcup}(i) = X_s(i) - \frac{n}{max(n, \sigma_i^2 M(i))}(X_s(i) - \bar{X}(i)). \tag{2}$$

Where σ_i^2 is the local variance of the region centered at pixel i in image X, n is the estimated noise energy, and $\bar{X}$ is the low-pass filtered X. The local signal energy σ_i^2 and the soft edge weight $M(i)$ balance the degree that pixel i is to be sharpened or smoothed. The effects on the CAS image $X_{\sqcup}$ are that in smooth regions away from edges $X_{\sqcup}$ is close to the clean image $\bar{X}$, while in high activity regions $X_{\sqcup}$ is close to the sharpened image X_s.



(a) original image (b) Sharpening result of (a)

(c) Denosing result of (b) (d) CAS preprocessing result

Fig. 2: Effects of the CAS preprocessing. The sharpening operation magnifies noises (b), and the recent learning-based denoising method [8] blurs the edges (c). In contrast, the CAS preprocessing produces a clean and sharp image (d).

Figure 2 illustrates the design objective of the proposed CAS preprocess; the CAS exaggeration effects can be clearly seen in comparison with enhancement and denoising results. Although the CAS image is somewhat unnatural in reference to the original image, it is more suited as the "ground truth" image when training the restoration CNNs and GANs because overshooting the target can counter the side effects of these deep learning methods as outlined in the introduction.

2.2. Learning under overshoot criterion

Setting the learning target to be the CAS preprocessed image rather than the original image is a general overshooting strategy. It is applicable to any existing image restoration CNN techniques regardless of the network architectures. For instance, widely used ℓ_1 norm in [9, 10, 5] can be changed to $\mathcal{L}_1 = \|\hat{X} - X_{\sqcup}\|_1$, where $\hat{X} = G(Y)$, G is the mapping function of restoration network and Y is the degraded input.

Cleanness and sharpness have distinctive characteristics in frequency domain. The former corresponds to zero high frequency energy, while the latter to strong high frequency energy. In light of this, we introduce a CAS-tuned adversary neural network (denoted by CAS-GAN) that demands statisti-

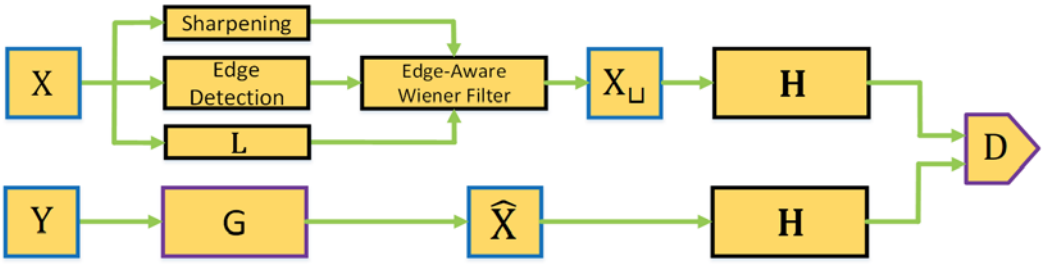


Fig. 3: The proposed deep learning system for CAS image restoration. Symbols **L** and **H** denote low pass filter and high pass filter respectively.

cal indistinguishability between the restored image $\hat{X}$ and the CAS image $X_{\sqcup}$ in high frequency aspect when optimizing the network. CAS-GAN uses a high frequency loss term:

$$\mathcal{L}_{adv} = -\log(1 - D(\mathbf{H}(\hat{X}))). \tag{3}$$

where D is the discriminator and **H** is a high pass filter module.

To further emphasize on the perceptual quality, we also include a perceptual loss term:

$$\mathcal{L}_p = \|\phi(\hat{X}) - \phi(X_{\sqcup})\|_1. \tag{4}$$

where $\phi(\hat{X})$ and $\phi(X_{\sqcup})$ are the feature maps of $\hat{X}$ and $X_{\sqcup}$ generated by the 19-layer VGG network [11]. Putting all the above loss terms together, we have the total objective function for the restoration task:

$$\mathcal{L}_G = \mathcal{L}_p + \alpha\mathcal{L}_{adv} + \beta\mathcal{L}_1 \tag{5}$$

where α and β are weights.

The proposed deep learning system for CAS image restoration is schematically presented in Figure 3. Since the network architecture requires no modification, the CAS training process can be expedited by initializing G with the available pre-trained models.

3. EXPERIMENTS

We have implemented the CAS-GAN and tested it on two restoration tasks: image super-resolution and low light correction (LLC). The network architecture is based on the state-of-the-art method ESRGAN[2], and the performance evaluations are conducted on widely used benchmark image sets.

3.1. Datasets and Training Details

In image super-resolution task, the training data are from the DIV2K dataset [12], which contains 800 images of rich details in diverse scenes. LR images are obtained through bicubic downsampling with scaling factor 4. In the low light correction task, we use the SID dataset [5], which contains pairs of insufficient exposure and corresponding correctly exposed images for the purpose of supervised learning for the LLC task. For both learning tasks random crop and flip are adopted for data augmentation.

We use Adam optimizer [13] to train the network. The learning rate is set to 1×10^{-4} and decayed by a factor of 2 every 200k iterations. Restoration network and the discriminator are alternately optimized for about 1000k iterations. The weights α and β in Eq.5 are set to1×10^{-3} and 3×10^{-2}, empirically.

3.2. Evaluations

We conduct comparison experiments of the proposed CAS-GAN versus the state-of-the-art methods: ESRGAN [2] and RCAN [14] for image super-resolution, SID [5] for low light correction. Since SID is trained with only fidelity term, we implement a GAN-based method by retraining ESRGAN with the SID training set, which is denoted by LLC-GAN.

We compare the quantitative performances of the above algorithms on the benchmark image sets Set14[15], DIV2K test set[12], PIRM test set[16], OST300[17] and Urban100[18] for super-resolution, and SID-test[5] for low light correction. The widely used image quality assessment metric, NIQE [19] is adopted. A lower NIQE represents a better perceptual quality. Table 1 and Table 2 are the NIQE results of the two tasks respectively. The methods RCAN and SID, which only use a fidelity measure(e.g., L_1), perform the worst. The GAN-based methods ESRGAN and LLC-GAN generate images with better perceptual quality. Standing out against other methods, CAS-GAN achieves the best results on widely used benchmark image sets for both restoration tasks.

Figure 4 exhibits the output images of the competing methods on super-resolution. As shown in the figure, CAS-GAN results have high contrast and are noise free at the same time. Other methods are inferior in terms of the CAS criterion. The results of MSE-based method RCAN look blurred in texture regions. The results of the GAN-based method ESRGAN are plagued with disturbing noises. Figure 5 presents more challenging cases for low light correction where the original input images suffer from extreme low signal-to-noise ratio. CAS-GAN obtains superior restoration results consistently. The high frequency details are recovered naturally such as the bar

(a) (b) (c) (d) (e)

Fig. 4: Sample results of the compared techniques on super-resolution. (a) HR images, (b) Bicubic, (c) RCAN, (d) ESRGAN, (e) CAS-GAN.

codes and the printed letters, and at the same time the sensor noises in smooth region are suppressed. In comparison, other techniques fail to reproduce clean and sharp images.

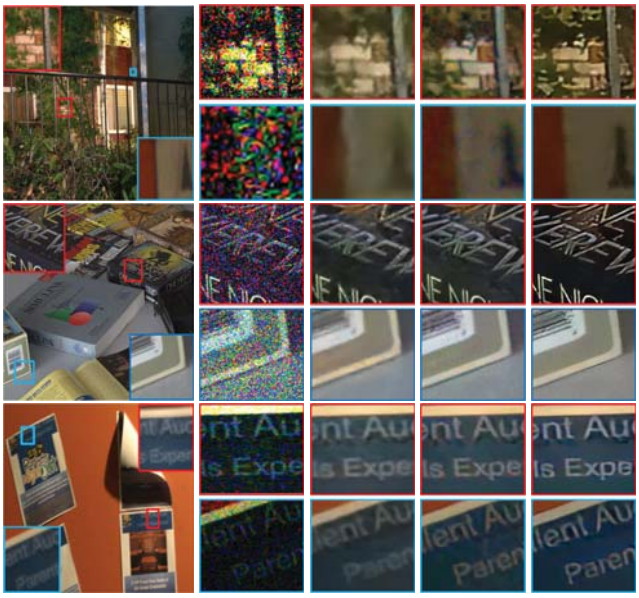
Table 1: NIQE of different methods on super-resolution.

	RCAN	ESRGAN	CAS-GAN
Set14	5.500	3.497	**3.402**
DIV2K test	4.435	2.768	**2.570**
PIRM test	5.042	3.448	**2.866**
OST300	4.878	3.237	**3.153**
urban100	4.4.753	3.879	**3.418**

Table 2: NIQE of different methods on low light correction.

	SID	LLC-GAN	CAS-GAN
SID test	5.217	4.261	**3.402**

For more comprehensive evaluations, we also compare CAS-GAN with the method of sharpening the ESRGAN restored images by the Photoshop smart sharpen filter. As shown in Figure 6, the super-resolution result of ESRGAN has relatively low contrast. The sharpening postprocessing makes the ESRGAN image quite noisy. In contrast, the CAS-GAN generated image is sharp but without much noise.



(a) (b) (c) (d) (e)

Fig. 5: Sample results of the compared techniques on low light correction. (a) correctly exposed images (ground truth), (b) Low light input images, (c) SID, (d) LLC-GAN, (e) CAS-GAN. The Low light images are dynamic range stretched for better visualization.

(a) (b) (c) (d)

Fig. 6: Comparison with post-sharpen algorithm. (a) HR image, (b) ESRGAN, (c) ESRGAN + sharpening, (d) CAS-GAN.

4. CONCLUSION

We proposed a new CNN design approach, called exaggerated deep learning, to simultaneously achieve noise free smooth regions and sharp edges/textures in image restoration, which are two mutually conflicting requirements for the current methods. The idea is to deliberately overshoot for the desired attributes in the CNN optimization objective function; the cleanness or sharpness is overemphasized in different semantic contexts. Experimental results demonstrate that the new approach is able to significantly improve the perceptual quality of existing image restoration CNNs.

5. REFERENCES

[1] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144.

[2] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[3] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 85–100.

[4] Raymond A Yeh, Chen Chen, Teck Yian Lim, Alexander G Schwing, Mark Hasegawa-Johnson, and Minh N Do, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5485–5493.

[5] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun, "Learning to see in the dark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3291–3300.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[7] Rafael C Gonzalez and RE Woods, "Digital image processing: Pearson education india," 2009.

[8] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang, "Toward convolutional blind denoising of real photographs," *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[10] 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*, 2017, pp. 4681–4690.

[11] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[12] 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*, 2017, pp. 126–135.

[13] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[14] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286–301.

[15] Roman Zeyde, Michael Elad, and Matan Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711–730.

[16] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor, "The 2018 pirm challenge on perceptual image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[17] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 606–615.

[18] 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*, 2015, pp. 5197–5206.

[19] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.