2023 IEEE International Conference on Image Processing (ICIP) | 978-1-7281-9835-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIP49359.2023.10223088

JPEG INFORMATION REGULARIZED DEEP IMAGE PRIOR FOR DENOISING

Tsukasa Takagi[⋆], *Shinya Ishizaki*[†], *Shin-ichi Maeda*[⋆]

[⋆]Preferred Networks, Inc.
[†]Graduate School of Informatics, Kyoto University

ABSTRACT

Image denoising is a representative image restoration task in computer vision. Recent progress of image denoising from only noisy images has attracted much attention. Deep image prior (DIP) demonstrated successful image denoising from only a noisy image by inductive bias of convolutional neural network architectures without any pre-training. The major challenge of DIP based image denoising is that DIP would completely recover the original noisy image unless applying early stopping. For early stopping without a ground-truth clean image, we propose to monitor JPEG file size of the recovered image during optimization as a proxy metric of noise levels in the recovered image. Our experiments show that the compressed image file size works as an effective metric for early stopping.

Index Terms— Image Denoising, Deep Image Prior, JPEG Compression, Early Stopping

1. INTRODUCTION

Image denoising is one of the important tasks in computer vision. Recent attractive image denoising methods tackle the case in which only noisy images are given. In this setting, we cannot apply supervised learning which usually needs a lot of pairs of clean and noisy images for training.

Deep image prior (DIP) [1, 2] is an image denoising method applicable to this setting. The key idea of DIP is to make use of the implicit regularization brought by the convolutional neural network architectures. DIP could estimate a clean image only from a single noisy image without training the network beforehand. In DIP, the network parameters are just randomly initialized and optimized so as to reconstruct the given noisy image. To prevent reconstructing the original noisy image, several follow-up DIP works have proposed to incorporate systematic early-stopping (ES) [3, 4]. However, existing ES methods struggle to compute a stable metric because there is no absolute aesthetic criterion in denoising. Also, feasible denoising during optimization depends on the noise level and the image content. For example, the best result could be that noises can remain in the DIP-based denoised image at some high noise level.

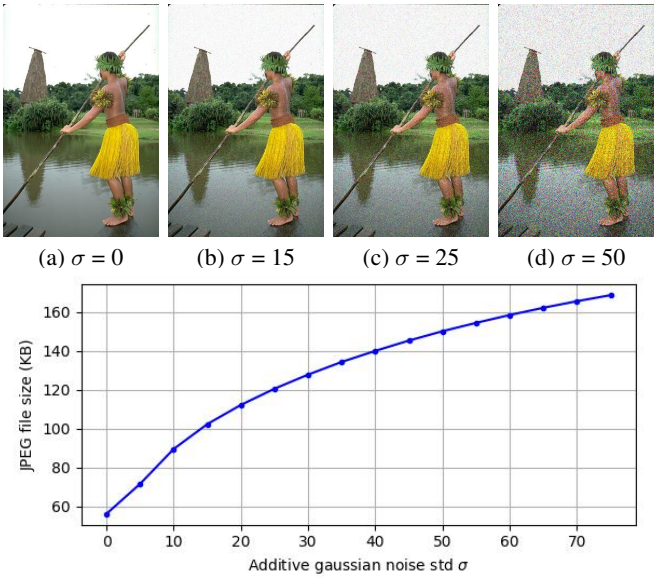

Fig. 1: **Relationship between different additive gaussian noise levels and JPEG file sizes.** The JPEG file sizes increase according to the noise levels in the images.

In this study, we propose a simple, but effective heuristic to determine the ES. We propose to use compressed image file size (CIFS), in particular, JPEG compressed file size of the reconstructed image to determine the termination. Since the target of the image compression is mainly clean images, JPEG file size tends to increase as the noise level increases. We preliminarily examined the relationship between the additive gaussian noise levels and their JPEG file sizes as shown in Fig. 1. As we can see, the JPEG file size monotonically increases according to the noise level increases. Our ES utilizes this relationship as the proxy indicator of the reconstructed noise in the denoised image. In our ES criterion, the optimization should be stopped where the CIFS increases although the image content would be reconstructed.

2. RELATED WORK

Since image denoising is one of the long standing problems in computer vision, there are many related works. For the sake of relevance to this paper, we overview only deep image

prior (DIP) and its subsequent works employing ES to prevent overfitting.

Deep image prior (DIP) [1, 2] demonstrates that randomly initialized neural networks can be used as an image prior for standard inverse problems including denoising. Although DIP can optimize the network parameters when only given a noisy image in denoising, the performance would degrade because of overfitting to the noisy image. One of the typical approaches to avoid overfitting is ES, in which some criterion is used as a proxy to measure the difference between the network output and the unknown clean image.

Several subsequent ES works of DIP have been proposed. DIP-SURE [5] introduce Stein's unbiased risk estimator (SURE) to compute approximately unbiased estimate for gaussian noise. DIP-denoising [6] extends DIP-SURE for poisson noise. Self-validation [3] and ES-WMV [4] are different directional ES methods which do not assume noise types and levels in a noisy image. Self-validation trains an autoencoder from windowed consecutive reconstructed images and the autoencoder evaluates the next reconstructed image quality as the ES criterion. ES-WMV computes windowed moving variance (WMV) as the ES criterion.

Our proposed ES method also does not assume noise types and levels. Our ES criterion is based on JPEG compressed image file size (CIFS).

3. PROPOSED METHOD

3.1. Preliminaries: Deep Image Prior (DIP)

Before describing our proposed ES method, we describe deep image prior (DIP). DIP is proposed for image restoration problems including denoising. Image restoration algorithms aim to recover an unknown clean image $\boldsymbol{x}$ given a corrupted image $\boldsymbol{x}_0$. DIP argued that a neural network architecture behaves as the general image prior, and the clean image $\boldsymbol{x}$ can be recovered from the corrupted image $\boldsymbol{x}_0$ without additional regularizations:

$$\min_{\boldsymbol{\theta}} \mathcal{L}(f_{\boldsymbol{\theta}}(\boldsymbol{z}), \boldsymbol{x}_0), \tag{1}$$

where $\boldsymbol{z}$ is randomly initialized and fixed noise, $f_{\boldsymbol{\theta}}$ is the neural network parameterized by $\boldsymbol{\theta}$, and $\mathcal{L}$ is the loss function such as mean squared error.

3.2. Compressed Image File Size based ES

As described in Sec. 1, compressed image file size (CIFS) can be used as a proxy metric of degrees of noise. As the network parameters fit to the corrupted image, the loss function $\mathcal{L}$ decreases almost consistently. On the contrary, the network gradually outputs a noisy image which would have a larger compressed image file size. Based on this insight, our ES considers finding trade-off between the loss function $\mathcal{L}$ and CIFS based regularizer $\mathcal{R}$ formulated as:

$$E(\lambda, t; \boldsymbol{z}) := \lambda \mathcal{L}(f_{\boldsymbol{\theta}_t}(\boldsymbol{z}), \boldsymbol{x}_0) + \mathcal{R}(C(f_{\boldsymbol{\theta}_t}(\boldsymbol{z}))), \tag{2}$$

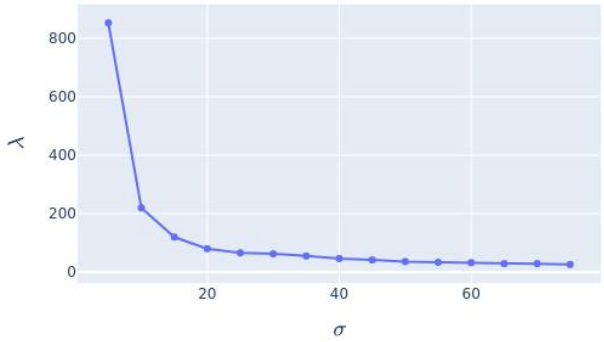

Fig. 2: **The gaussian noise standard deviation σ and the estimated λ in Eq. (2).**

where C is a function which takes an image as input and outputs CIFS, $t \in \mathbb{N}$ is an epoch number during optimization, and $\lambda \in \mathbb{R}_{\geq 0}$ is a balancing weight between two terms. In our experiments, we adopt mean squared error as the loss function $\mathcal{L}$ same as DIP, JPEG file size output function as C, and squared JPEG file size averaged over image size as $\mathcal{R}$. Specifically, $\mathcal{R}$ is

$$\mathcal{R}(L) = \frac{L^2}{HW}, \tag{3}$$

where L is CIFS and (H, W) are image height and width, respectively. Since CIFS depends on its image size, L is averaged over its image size HW. Moreover, we also use squared file size L^2 so that regularization works strongly as the file size increases.

The λ is affected by degrees of noise. We show estimated λ for additive gaussian noise with different standard deviation σ in Fig. 2. The λ estimation is based on our preliminary experiment on CBSD500 [7]. We assume that we can observe 400 clean images on CBSD500. Under this assumption, we could search λ by the following steps: (1) adding a gaussian noise to a clean image to create a noisy image synthetically, (2) denoising the noisy image by DIP and monitoring values $\mathcal{L}, \mathcal{R}$ in Eq. (2), (3) and finding the optimized λ as maximizing the average PSNR between the clean image and the denoised image at the ES criterion minimized epoch over the observed 400 images. Specifically, the final step is formulated as:

$$\max_{\lambda} \mathbb{E}_{\boldsymbol{z} \sim U_z(\boldsymbol{z}), \boldsymbol{x} \sim U_x(\boldsymbol{x})}[PSNR(\boldsymbol{x}, f_{\boldsymbol{\theta}_{\hat{t}}}(\boldsymbol{z}))]$$
$$\text{where } \hat{t} = \arg\min_{t} E(\lambda, t; \boldsymbol{z}),$$

where $U_{\boldsymbol{z}}$ is a pixel-wise uniform distribution on $[0, 1]$, and $U_{\boldsymbol{x}}$ is a uniform distribution on the 400 observed images of CBSD500.

Table 1: **PSNR comparison on CBSD68 [8] and Kodak24 [9].** Each value in the table denotes mean ± std of PSNR for the corresponding method and dataset. σ denotes the std of the gaussian noise. "No ES" indicates the PSNR at the last epoch ($T = 20$k epoch) and "Peak" indicates the maximum PSNR between a clean image and a denoised image in $T = 20$k epochs.

Dataset	σ	BRISQUE [10]	NIQE [11]	ES-WMV [4]	CIFS (Ours)	No ES	Peak
CBSD68 [8]	15	27.750±2.108	27.484±4.172	28.640±3.877	**29.223±2.618**	26.464±0.535	30.463±2.066
	25	24.754±2.740	26.642±2.759	26.375±3.577	**27.338±2.634**	21.740±0.410	27.767±2.206
	50	21.078±2.183	23.782±2.369	23.289±2.826	**23.803±2.321**	15.843±0.416	24.388±2.241
Kodak24 [9]	15	29.685±2.211	29.247±3.768	29.779±3.905	**31.263±1.771**	28.567±0.533	31.585±1.730
	25	26.341±4.906	27.498±3.490	27.686±3.272	**28.804±1.947**	24.188±1.651	29.040±1.853
	50	22.759±3.426	**25.224±1.807**	24.640±2.322	25.197±1.759	17.385±0.328	25.668±1.844

Fig. 3: **PSNR performance gap comparison on CBSD68 [8] for additive gaussian noise with different standard deviation σ.**

4. EXPERIMENTS

4.1. Datasets

We adopted CBSD68 [8] and Kodak24 [9] as image denoising benchmark datasets. CBSD68 is a different split not overlapping the split used in λ estimation described in Sec. 3.2.

4.2. Evaluation process

Non-reference image quality assessment (NR-IQA) has tackled to score image quality without a reference image. Since NR-IQA methods are expected to provide good image criteria, these methods can be employed as ES. Following the recent ES method for DIP, ES-WMV [4], we compared to BRISQUE [10] and NIQE [11] as the NR-IQA methods and ES-WMV [4] as the ES method to validate whether our proposed method can provide a good ES criterion. We used models and code provided by OpenCV[1] [12] and LIVE[2] for BRISQUE and NIQE, respectively.

Let T denote the number of training epochs. First, we optimized the neural network parameters with respect to Eq. (1) reaching to T epochs. Second, we obtained the ES detected epoch t_* such that each ES criterion is satisfied. Following ES-WMV ES detection, our and other ES criteria find ES candidate epochs $\mathcal{T}$ and adopt the minimum epoch in $\mathcal{T}$ as the ES detected epoch t_*. The candidate epoch $t \in \mathcal{T}$ satisfies that the ES criterion outputs a smaller value than the next consecutive S epochs. Specifically,

$$\mathcal{T} = \{t \mid t = \underset{\tau \in [t, t+S]}{\arg\min} M\left(f_{\boldsymbol{\theta}_\tau}(\boldsymbol{z})\right)\}, \tag{4}$$

where M is a specific criterion function for each ES method. If an ES method cannot find any candidate epochs (i.e., $\mathcal{T} = \emptyset$), T is adopted as an alternative to the ES detected epoch. Finally, PSNR is computed between the clean image $\boldsymbol{x}$ and the denoised image $f_{\boldsymbol{\theta}_{t_*}}(\boldsymbol{z})$ as the ES criterion evaluation.

4.3. Implementation details

CIFS utilizes JPEG as image compression method. JPEG can specify a quality value $Q \in [0, 100]$ (higher is better image quality) which affects the saved image file size. Q is set to 95 which is the default value in OpenCV.

The following other implementation details are shared with CIFS and all comparative methods. Pixel-wise gaussian noise is independently sampled from $\mathcal{N}(0, \sigma^2)$ where we set $\sigma \in \{5, 10, \ldots, 75\}$ for pixel value range $[0, 255]$. The network architecture is same as DIP and Adam [13] with a learning rate 0.01 is used. The input noise $\boldsymbol{z} \in \mathbb{R}^{D \times H \times W}$ to the network is randomly initialized and fixed during optimization where D is set to 32. We adopted the $\boldsymbol{z}$ perturbation technique from DIP [1, 2] where we perturb $\boldsymbol{z}$ at each iteration. The number of training epochs T is set to 20k. S is set to 1k, which is same as ES-WMV experimental setting.

4.4. Analysis

Table 1 shows PSNR comparisons of our and other comparative ES methods for $\sigma \in \{15, 25, 50\}$ on CBSD68 and Kodak24. Fig. 3 depicts PSNR performance gap comparison for $\sigma \in \{5, 10, \ldots, 75\}$ on CBSD68 as the broader noise level range evaluation. DIP without ES ("No ES") is significantly

[1] https://github.com/opencv/opencv_contrib
[2] https://github.com/utlive/live_python_qa

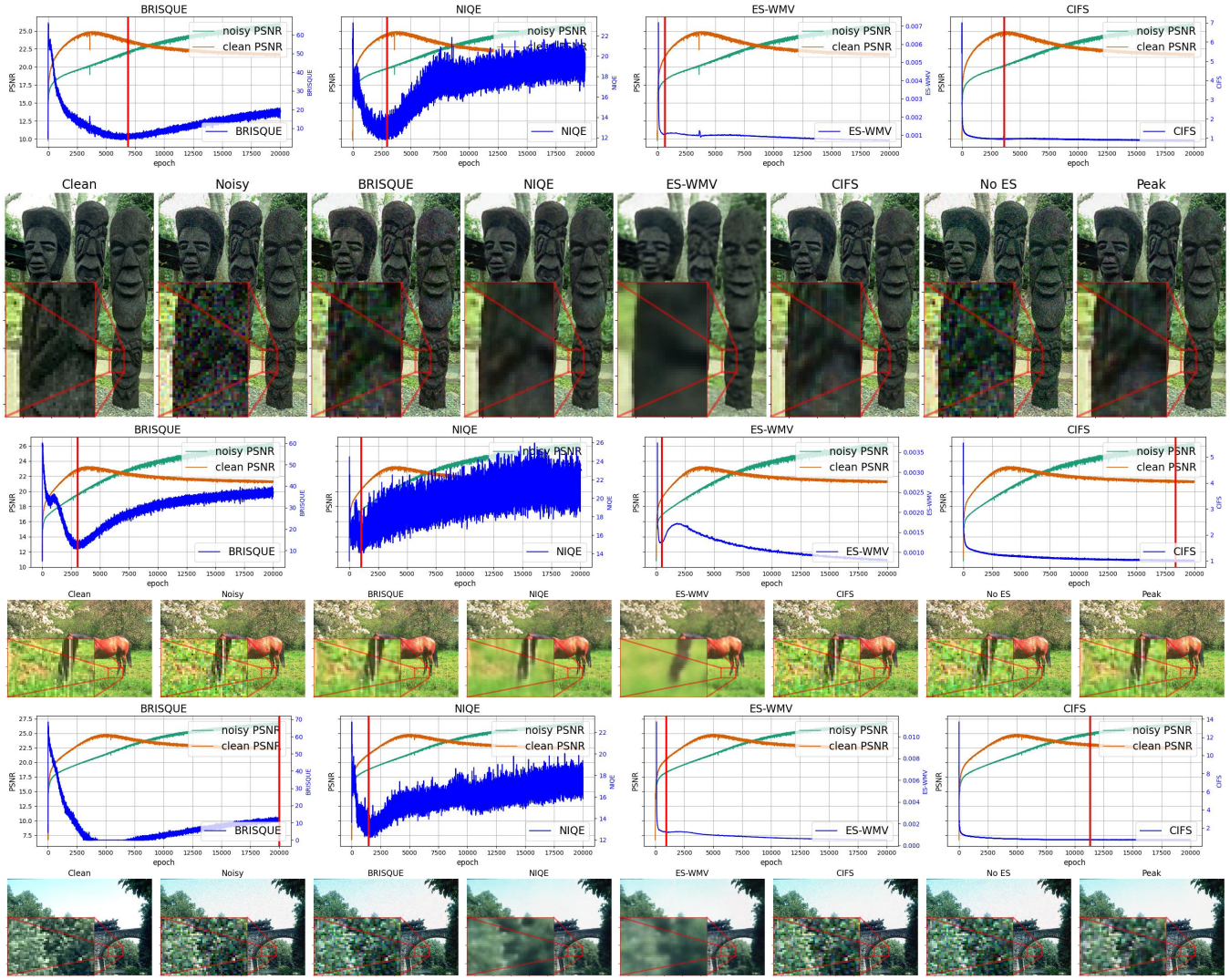

Fig. 4: **Qualitative comparison for $\sigma = 25$ on CBSD68 [8].** CIFS almost always found good ES epochs whereas the comparative methods sometimes failed to find good ES epochs.

lower than the maximum achievable PSNR ("Peak") regardless of gaussian noise levels. CIFS provides good ES criterion and almost outperforms other comparative methods for both datasets. Moreover, the standard deviations of PSNR for different noise levels are relatively small compared to other ES methods. ES criteria and qualitative comparison for $\sigma = 25$ are depicted in Fig. 4. Note the "Peak" PSNR is same for all ES methods since all ES metrics are computed in the same optimization process. BRISQUE and NIQE, which are NR-IQA methods, have an issue with the magnitude of variance of their metrics. ES-WMV is stable, however, it keeps a long variance sequence since the metric is windowed moving variance. CIFS can compute the metric independently at each iteration and still provide a stable metric.

5. CONCLUSION

We propose a novel ES for DIP designed for image denoising. Our ES employs JPEG compressed image file size as a proxy metric to measure degrees of noise for denoised images. Our method provides a good ES criterion and outperforms many of the other comparative methods. Moreover, our ES method can compute the metric independently at each iteration and still provide a stable metric.

Acknowledgement

We thank Sol Cummings for helpful feedback. This research work was financially supported by the Ministry of Internal Affairs and Communications of Japan with a scheme of "Research and development of advanced technologies for a user-adaptive remote sensing data platform" (JPMI00316).

6. REFERENCES

[1] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep image prior," in *CVPR*, 2018. 1, 2, 3

[2] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep Image Prior," *IJCV*, 2020. 1, 2, 3

[3] Taihui Li, Zhong Zhuang, Hengyue Liang, Le Peng, Hengkang Wang, and Ju Sun, "Self-validation: Early stopping for single-instance deep generative priors," in *32nd British Machine Vision Conference 2021, BMVC 2021, Online, November 22-25, 2021*. 2021, p. 108, BMVA Press. 1, 2

[4] Hengkang Wang, Taihui Li, Zhong Zhuang, Tiancong Chen, Hengyue Liang, and Ju Sun, "Early stopping for deep image prior," 2023. 1, 2, 3

[5] Christopher A. Metzler, Ali Mousavi, Reinhard Heckel, and Richard G. Baraniuk, "Unsupervised learning with stein's unbiased risk estimator with applications to denoising and compressed sensing," in *International Biomedical and Astronomical Signal Processing Frontiers Workshop (BASP)*, 2019. 2

[6] Yeonsik Jo, Se Young Chun, and Jonghyun Choi, "Rethinking deep image prior for denoising," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 5087–5096. 2

[7] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011. 2

[8] S. Roth and M.J. Black, "Fields of experts: a framework for learning image priors," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 2, pp. 860–867 vol. 2. 3, 4

[9] Rich Franzen, "Kodak lossless true color image suite," 1999. 3

[10] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012. 3

[11] 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, 2013. 3

[12] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000. 3

[13] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun, Eds., 2015. 3