Hybrid Training of Denoising Networks to Improve the Texture Acutance of Digital Cameras

Raphaël Achddou$^{(\boxtimes)}$, Yann Gousseau, and Saïd Ladjal

LTCI, Telecom Paris, Institut Polytechnique de Paris, Palaiseau, France
`raphael.achddou@telecom-paris.fr`

Abstract. In order to evaluate the capacity of a camera to render textures properly, the standard practice, used by classical scoring protocols, is to compute the frequential response to a dead leaves image target, from which is built a *texture acutance metric*. In this work, we propose a mixed training procedure for image restoration neural networks, relying on both natural and synthetic images, that yields a strong improvement of this acutance metric without impairing fidelity terms. The feasibility of the approach is demonstrated both on the denoising of RGB images and the full development of RAW images, opening the path to a systematic improvement of the texture acutance of real imaging devices.

Keywords: Image denoising · Deep learning · Image quality assessment

1 Introduction

In order to correctly visualize a photograph, its corresponding RAW image undergoes a complex sequence of development operations including white balancing, demosaicking, tone mapping, and image restoration operations such as deblurring and denoising. Camera manufacturers implement proprietary algorithms fine-tuned for each setting of each camera. As a result, the overall image quality is a combination of hardware characteristics (quality of the lens, size of the sensor) and software performances. In order to fairly assess the quality of an imaging device, independent agencies have defined standard tests and ISO protocols. Each of these tests focus on a specific characteristic such as chromatic aberrations, noise reduction, or texture rendering.

Recently, with the increase in computational power and the advent of deep learning for image processing, more and more digital image processing stages can be replaced by learned neural networks [17]. Recent works already aim at completely replacing the full image development pipeline with a single neural network, producing impressive results in standard conditions [20] or extremely low-light conditions [13]. Moreover, light neural network architectures can now be integrated in embedded systems, e.g. on smartphone devices. Neural methods

present another key advantage: one can easily optimize their response to specific test images, by including them in training databases.

For the specific task of texture rendering evaluation, Cao et al. [10] first presented a protocol quantifying the ability of an imaging pipeline to preserve texture information. This is obtained through the frequential response of the system to dead leaves images with a specific perceptual metric called *texture acutance*. These images are known for their invariance properties, as well as statistical properties making them close to natural images (non Gaussianity, scaling property, distribution of the spectrum and gradient), as studied in [4, 18, 23]. This quality evaluation protocol later became an ISO standard to measure the preservation of textures [21] and is now used by classical camera scoring protocols. In a different direction, Achddou et al. [3] showed that image restoration networks could be trained from synthetic images only, using databases of dead leaves images.

Inspired by these results, we propose, in this paper, to train a denoising neural network on natural and dead leaves images, to jointly optimize a new metric derived from the texture acutance and the classic data fidelity metrics on natural images. After presenting some related works on image restoration in Sect. 2, we first introduce in Sect. 3 the texture acutance metric and the corresponding perceptual loss for image restoration networks. We then show in Sect. 4 that we can strongly improve the texture acutance metric without impairing performances on natural images, first for the task of Additive White Gaussian Noise removal (AWGN) and then for the development of RAW images. These results open the path to an automatic improvement of standard quality evaluation tests.

2   Related Works

The goal of image restoration is to retrieve a clean image from distorted observations. In many cases, the distortion process can be modeled as follows: $y = Ax + n$, where x is the theoretically perfect image, y the distorted observation, A is a linear operator and n is some noise.

In order to solve this problem, a first class of methods are based on prior hypotheses on the distribution of natural images. These methods try to impose regularity properties on the restored solutions. For instance, wavelet shrinkage methods [15, 16] or DCT-filtering methods [31] reconstruct an image assuming that the targeted images can be well approximated by a sparse decomposition. In turn, variational methods based on the total variation [12, 29] assume that the image gradient follows a Laplacian distribution. Based on the assumption of self-similarity, non-local methods leverage the redundancy in the image content. This is either done by weighted averaging (Non Local Means [9] Non Local Bayes [22]) or by collaborative filtering (BM3D [14]).

Over the past decade, learning-based approaches for image restoration have developed drastically. After the success of neural networks for high-level computer vision tasks [19], these methods have been adapted to image restoration through the use of generative models [32, 33]. Rather than using prior hypotheses, the parameters of the neural networks are tuned in a long optimization process to directly minimize the reconstruction error in a black box manner. For

the training to succeed, these methods require large databases of pairs of distorted and clean images. Even though they are hard to interpret, they surpassed prior-based methods on most image restoration benchmarks by a large margin for a wide variety of tasks such as image denoising [33], demosaicking [17] etc.

Following these initial works, recent papers extended the use of deep learning methods to real-world problems of image restoration such as RAW image denoising [5,24]. Ignatov et al. [20] and Chen et al. [13] also propose to fully replace the image development pipeline by a learned neural network, producing surprisingly good results. However, acquiring datasets of real-world pairs of distorted and clean RAW images is a cumbersome task [1,13], which often requires complex post-processing algorithms. In order to ease the training process, a more restrained approach consists in modeling the distortion process accurately, and to synthesize them accordingly [30].

Going further, Achddou et al. [2,3] proposed to train image restoration neural networks on generated dead leaves images in order to completely circumvent the data acquisition process, reaching performances close to the networks trained on real images, for various image restoration tasks. These images indeed exhibit statistical properties close to those of natural images [4,18,23] even though they depend from few parameters. Following [3], similar synthetic databases were also used to pre-train image classification networks [7] and disparity map estimators [25]. Prior to these works, dead leaves images were used to assess the capacity of cameras to render textures properly. This idea was first presented in 2009 by Cao et al. [10], which was later improved in the following references [6,11]. We will present in detail these works in the following section.

3   Texture Acutance: A Frequential Loss Assessing Texture Preservation
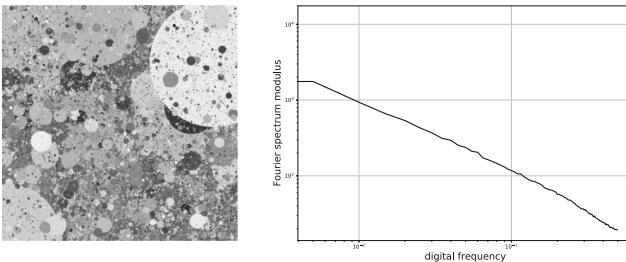
3.1   Dead Leaves Images

Fig. 1. Grey level dead leaves image and its associated digital spectrum in logarthmic scales. The theoretical value is a straight line. (Color figure online)

Dead leaves images were first introduced by Matheron in 1975 [26], with the aim of modeling porous media. It was later shown that if object sizes fulfill some

scaling property, this model accounts for many statistics of natural images [4, 23]. To generate such images, shapes of random size, color and position are superimposed on top of each other until the whole image plan is covered. In the simplest set-up, these shapes are disks of random radius. An example of a dead leaves image is given in Fig. 1, along with its spectrum. A precise mathematical formulation of dead leaves images is given in [8].

Dead leaves images were first used for camera evaluation in 2009 by Cao et al. [10]. The proposed idea is to measure the response of a camera to a specific image target. Because of their invariances and statistical properties, the dead leaves model was chosen by the authors as the generation algorithm for the target. Among the desired properties, scale invariance is achieved when the disks radii follows a power law with $\alpha = 3$. The dead leaves target is therefore generated with this parameter. Note that to ensure the convergence of the algorithm, bounding parameters r_{min}, r_{max} are required [18].
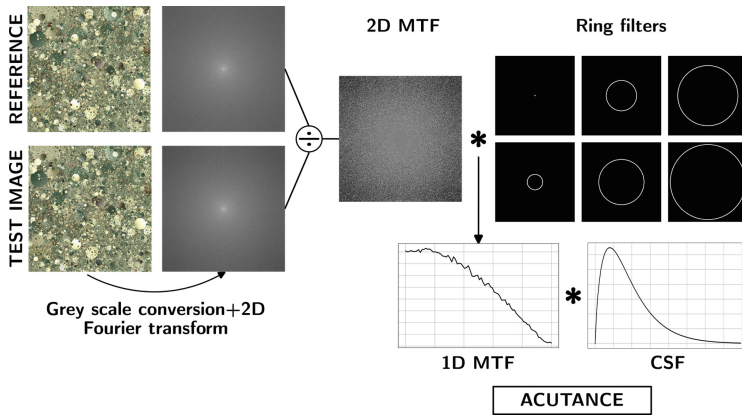
3.2 Texture Acutance

Fig. 2. Diagram explaining the computation of the acutance metric

In [10], the authors evaluate the response of a camera to the dead leaves target by computing the ratio of the power spectra, resulting in a Modulation Transfer Function (MTF). At each position (m, n) for an (N, N) image:

$$\text{MTF}_{2D}(m, n) = \frac{|\hat{Y}(m, n)|}{|\hat{X}(m, n)|},$$

where $\hat{Y}$ is the spectrum of the obtained image and $\hat{X}$ is the ground truth spectrum. In all that follows, we compute the image spectra on a greyscale version of the color image, obtained by the standard linear combination

Grey $= 0.2126R + 0.7152G + 0.0722B$. The classical idea behind the MTF is that the ratio of the power spectra corresponds to the Fourier coefficients of the blur kernel induced by the camera, excluding non linear transforms often involved in the image development, as well as the impact of noise (Fig. 2).

In order to account for the impact of noise in the estimation of the MTF, Artmann first proposed a corrected version of the MTF, by subtracting an estimate of the noise spectrum. However, the latter was computed by taking a photograph of a uniform grey surface, assuming an additive and signal independent noise model, which is far from reality. Moreover, some image development pipelines include a nonlinear noise reduction operation, which affect the estimation of the real MTF.

In the same paper [6], Artmann proposes a new computation trying to correct these issues. Here, we consider the complex spectrum of a reference digital dead leaves target $\hat{X}$, rather than the estimate of the power spectrum $|\hat{X}|$ in the spatial domain. In the previous version, the phase information was lost. This is necessary in the context of camera calibration since phase information is reliable only if a registration algorithm is applied. In the context of training a denoising algorithm registration between noisy and restored image is supposed perfect and dealt with by the MSE-loss. Relying only on the amplitude of the spectrum meant that we could not differentiate frequencies which were already in the target and information that was added by the imaging device. Therefore noise and non linear functions had an impact in previous computations.

The proposed method, which we call MTF_{cross} uses the cross power density between the target and the obtained image $\phi_{XY}(m,n)$, and the auto power density $\phi_{XX}(m,n)$. More precisely,

$$\phi_{XY}(m,n) = \hat{Y}(m,n)\hat{X}^*(m,n) \quad \text{and} \quad \phi_{XX}(m,n) = \hat{X}(m,n)\hat{X}^*(m,n).$$

Given these quantities, the MTF becomes :

$$\text{MTF}_{cross}(m,n) = \left| \frac{\phi_{XY}(m,n)}{\phi_{XX}(m,n)} \right|. \tag{1}$$

Since the dead leaves target is rotationally invariant, so is its spectrum. We therefore express the MTF as a 1D function by averaging it on concentric rings of width 1. The MTF becomes :

$$\text{MTF}_{1D}(k) = \frac{\mathbb{1}_{C_k}}{\#C_k} \times \text{MTF}_{cross},$$

where $C_k = \{(i,j) \in [-N/2, N/2]^2 | (k-1)^2 \leq |i^2 + j^2| < k^2\}$ corresponds to a ring of radius k and $\#C_k$ is its cardinal.

Though the full MTF_{1D} is a good indicator of the camera's capacity to render textures, it is more helpful to compute a single score. To that end, the *texture acutance* [10] is defined as a weighted sum of the MTF_{1D}, with weights defined by a contrast sensitivity function (CSF), inspired by the slanted edge Spatial Frequency Response (SFR), used to evaluate the sharpness of a camera. Our visual system is indeed more sensitive to some frequencies than others. In that

regard, the CSF models the sensitivity of the visual system to spatial frequencies expressed in cycle/degree.

Based on the physiological analysis of the contrast sensitivity of infants and monkeys led by Movshon and Kiorpes [27], the chosen formula to model the CSF is :

$\mathrm{CSF}(\nu) = a.\nu^c.e^{-b\nu}$, where ν is a spatial frequency expressed in cylces/degree, parameters are fixed as $b = 0.2$, $c = 0.8$, and a is a normalizing parameter so that $\int_0^{\mathrm{Nyquist}} \mathrm{CSF}(\nu)d\nu = 1$. Given this formula, the texture acutance score can be written as :

$$A = \int_0^{\mathrm{Nyquist}} \mathrm{CSF}(\nu).\mathrm{MTF}_{1D}(\nu)d\nu.$$

Note that we need to convert spatial frequencies in cycles/degree to a digital frequency in cycles/pixel for homogeneity. To do so we use the following formula: $f_{spatial} = \frac{1}{\alpha} f_{digital}$, where α is the viewing angle. The latter depends on viewing conditions with the equality $\alpha = \frac{180}{\pi}\arctan(\frac{P}{D})$, where P is the pixel size and D is the viewing distance, assumed to be equal to 0.2 mm and 1m respectively. This corresponds to a maximal spatial frequency of 40 cycles/degree which is approximately the limit of the human visual system.

The perfect MTF corresponds to a constant function equal to 1, meaning that the frequential content has been perfectly restored by the camera for every frequency. This leads to an acutance $A = 1$. An acutance greater than 1 indicates that some frequential content was added to the image, probably because of noise or sharpening. An acutance lower than 1 indicates that some frequencies have been lost.

3.3   Acutance Loss for Image Restoration CNNs

In [3], the authors showed that models trained on mixed databases (natural and synthetic images) perform on par with models trained on natural images only, while improving results on dead leaves image targets. We believe we can improve the frequential response of models trained on mixed sets, by using the acutance score in a loss function.

In the context of AWGN removal for color RGB images, the noisy image corresponds to $Y = X + n$ where X is a ground truth dead leaves image of size $(N, N, 3)$. The denoising network f_θ produces an estimate of the clean image $Z = f_\theta(Y)$. For our restoration problem, we can consider that the denoising network is analogous to the camera which acquires the dead leaves target. We can compute MTF_{cross} for the denoising network using Formula (1), based on the computation of the digital spectrum of both X and Z.

The obtained MTF_{cross} is turned into a 1D signal as described above. For faster computation, concentric ring masks are stored in GPU so that the computation of MTF_{1D} can be accelerated with parallel computing. Since the best possible acutance is 1, we define the acutance loss function as :

$$\mathcal{L}_{acutance}(Y, X) = |1 - A(f_\theta(Y), X)|,$$

which penalizes both adding or removing frequential information. In order to get a complete loss function, we add to it the $\mathcal{L}_2$ loss, the initial fidelity term of the network. Indeed, the acutance loss $\mathcal{L}_{acut}$ is computed solely on an aggregation of the Fourier spectrum and is therefore blind to the spatial organisation of the image and can not replace an MSE-loss. When training on dead leaves images the loss is therefore

$$\mathcal{L} = \mathcal{L}_2 + \lambda.\mathcal{L}_{acut},$$

where λ is a weighting parameter.

Since we train the image denoiser on both natural images and dead leaves images, we compute the acutance loss only on the dead leaves images in a mini-batch D of size K and the $\mathcal{L}_2$ loss for all images. The formation of minibatches during training indeed randomly samples images from the mixed set. Thus, the loss in a batch becomes:

$$\mathcal{L}_{batch} = \frac{1}{K} \sum_{i=0}^{K} ||x_i - f_\theta(x_i + n_i)||_2^2 + \frac{\lambda}{m^T \mathbf{1}} \sum_{i=0}^{K} m_i.\mathcal{L}_{acut}(f_\theta(x_i + n_i), x_i), \quad (2)$$

where m is a masking vector of size K such that $m_i = 1$ if x_i is a dead leaves image, or $m_i = 0$ otherwise. In order to count the number of dead leaves images we sum this masking vector which is given by $m^T \mathbf{1}$.

4   Image Denoising Results with FFDNet

We choose to train the FFDNet network [33] to illustrate the impact of the perceptual loss we presented in the previous section. We adapt the training scheme of the network to the present problem as follows. First, we increase the size of the training patches from $(50, 50, 3)$ to $(100, 100, 3)$. The reason for this is that the estimation of the 1D-MTF on a small patch is not sufficiently accurate. Keeping the same rings' width would result in fewer estimates for the 1D-MTF. On the other hand, decreasing the rings' width would lead to noisier estimates. Therefore, we perform the training with larger patches. Second, we reduce the batch size from 64 to 32 during training to decrease the memory footprint. We use 150000 samples, made of 100000 natural image patches and 50000 synthesized dead leaves patches. The other training hyper-parameters remain unchanged, such as the number of epochs or the learning rate decaying schedule.

4.1   Quantitative Evaluation

In order to show that the proposed scheme indeed has the potential to improve the texture acutance without impairing the usual PSNR evaluation of the performances on natural images, we compute both these metrics for various values of λ, the weighting parameter in Eq. (2). We consider values of $\lambda \in [0, 2, 5, 10, 20, 50, 100, 200, 500]$. Moreover, we also compute the classical SSIM metric and the perceptual metric PieAPP recently introduced in [28].

The models are evaluated numerically on two datasets. First, we evaluate the data fidelity by computing the PSNR, SSIM and PieAPP metrics on the Kodak24 dataset, a benchmark test set of 24 natural images. Second, we evaluated the acutance metric on a test set of synthesized dead leaves images.

We report, in Table 1, the numerical evaluation of the trained models. We observe a similar behaviour for the tested noise levels $\sigma = 25$ and $\sigma = 50$. In both cases, we notice that the standard evaluation metrics, i.e., the PSNR and SSIM, are not affected by the increase of the weighting parameter λ until $\lambda = 20$. For values greater than $\lambda = 100$ these metrics decrease rapidly. On the other hand, the acutance metric keeps improving until $\lambda = 100$ and then reaches a plateau. This table shows that we can optimize the texture acutance without impairing classic denoising evaluation. The perceptual evaluation with the PieAPP metric suggests that, for high noise values, the perceptual image quality is slightly enhanced by the addition of the acutance loss. Some results can be visualized in Fig. 3 (please zoom in the electronic version of this document). The result with and without using the acutance loss appear quite close, despite the strong improvement of the texture acutance measurement. Nonetheless, we can notice some improvements in the preservation of low-contrast details in the first row. Moreover, the contrast is also rendered better when training with the acutance loss. Finally, on the third row, details on the dead leaves images are better preserved using the acutance loss. On the second and third row, we see that the network trained with natural images sometimes hallucinates details, which are removed when training with dead leaves images.

Table 1. Denoising results of FFDNet trained with different weighting coefficients of the acutance loss for two noise levels. Each cell contains the PSNR, SSIM, PieAPP evaluated on Kodak 24 , and the Acutance metric evaluated on a test set of dead leaves images. Best results in blue, second results in red.

λ		0	2	5	10	20	50	100	200	500
$\sigma = 25$	PSNR ↑	31.88	31.87	31.88	31.87	31.88	31.85	31.77	31.65	31.56
	Acutance ↓	0.034	0.029	0.023	0.020	0.015	0.013	0.012	0.012	0.012
	SSIM ↑	0.877	0.876	0.875	0.875	0.877	0.876	0.873	0.872	0.869
	PieAPP ↓	0.568	0.596	0.598	0.602	0.586	0.587	0.591	0.62	0.612
$\sigma = 50$	PSNR ↑	28.81	28.79	28.80	28.80	28.80	28.76	28.66	28.58	28.42
	Acutance ↓	0.084	0.078	0.073	0.053	0.035	0.026	0.022	0.022	0.022
	SSIM ↑	0.791	0.788	0.789	0.789	0.790	0.789	0.786	0.783	0.779
	PieAPP ↓	0.932	0.956	0.952	0.948	0.912	0.925	0.934	0.940	0.953

4.2   Spectral Preservation

For mixed trainings of FFDNet, the texture acutance score is greatly improved when using the corresponding loss, which is expected. However, the acutance score only gives a partial information about the MTF of the trained network.
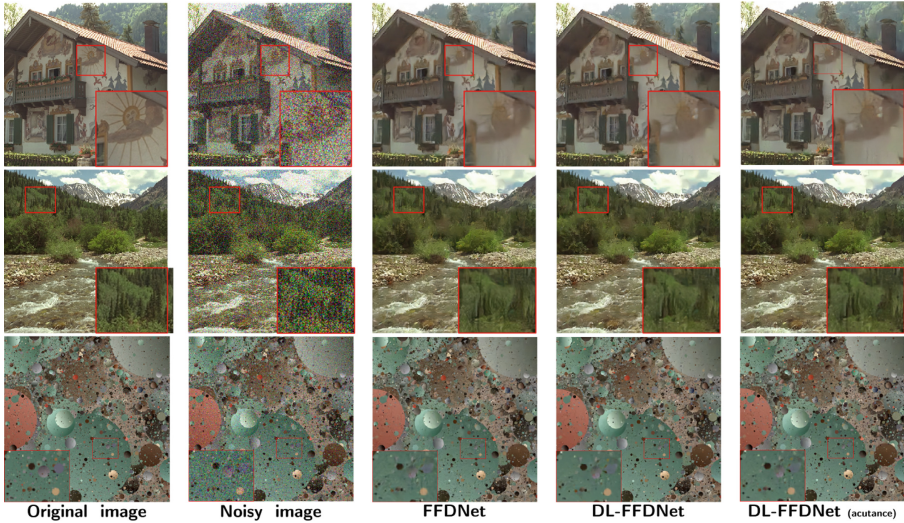
Fig. 3. Comparison of FFDNet results on two natural images and on a dead leaves image. From left to right: original image, noisy image, image denoised with standard FFDNet, image denoised with FFDNet trained on a mixed database without the acutance loss, and finally with the acutance loss.

In order to further understand the impact of the acutance loss on the spectral preservation ability of the network, we compute its MTF as described next. We compute the 1D-MTF from the denoised image and the original image for each dead leaves image of the synthetic test set.

Since the 1D-MTF depends on the image's content, which differs from image to image, we average the obtained MTF over the whole dataset. In Fig. 4, we report the MTF of FFDNet trained with and without the acutance loss (with $\lambda = 50$) for a noise level $\sigma = 25$. Recall that a perfect MTF should be equal to one. We can observe that for low to medium frequency, the MTF of the model trained with the accutance loss is much closer to one. Actually,

Fig. 4. Comparison of the MTF evaluated with FFDNet trained on a mixed database with or without the acutance loss, on the whole dead leaves image test set.

the values for low frequency exceed one which is one way the system can improve the acutance and which indeed is a limitation of the approach. For high frequency, the gap between the two MTF is smaller, probably as a result of the profile of the CSF function, which quickly decreases for high frequency, see Fig. 1. This behavior, as well as the addition of low frequency, could be modified by considering alternative CSF functions and can be easily integrated into our frame-
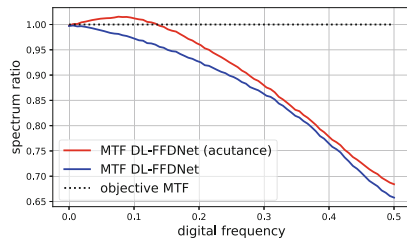
work. This could also further improve the preservation of details on examples such as those of Fig. 3. In this paper, we have decided to keep the original definition of the acutance, since our main goal is to show that this standard measure of the ability to preserve texture can be greatly improved without impairing the other aspects of image quality.

4.3 RAW Image Denoising

As a proof of concept, we extended our experiments to real-world image denoising on the SIDD benchmark [1] for cameraphones denoising. To that end, we trained the same denoising network with a U-Net architecture to denoise real RAW images and synthesized RAW dead leaves images. This network produces a RAW denoised image from a RAW input noisy image. To simulate RAW noise for dead leaves images, we used a Poisson-Gaussian model with realistic noise parameters. Unlike Gaussian noise removal, the loss is here a combination of the $\mathcal{L}_1$ loss and the acutance loss : $\mathcal{L} = \mathcal{L}_1 + \lambda \mathcal{L}_{acut}$. For RAW images, the acutance computation differs slightly. In order to convert a RAW image to a grey-scale image, we first pack the (H, W) image in a $(H/2, W/2, 4)$ RGGB tensor, then we average them in a single $(H/2, W/2)$ grey array, by weighting each channel with the white balance parameters. We ran the training for $\lambda \in [0, 10, 100]$. We report the numerical results obtained in Table 2. In comparison with $\lambda = 0$, the PSNR is still good for $\lambda = 10$, while the RAW acutance is largely improved. This improvement also translates in a better acutance in the RGB domain, which was not seen during training. This metric is computed on the denoised images developed with a standard ISP. This experiment shows that we can improve camera evaluation without impairing the image quality in the case of a full camera development pipeline. For $\lambda = 100$, the PSNR noticeably decreases while the RAW acutance reaches a plateau.

Table 2. Denoising results obtained by training a denoising Unet for real RAW images evaluated on the SIDD test set of cameraphone images [1]. We report the PSNR, Acutance RAW and acutance RGB metrics. Best results in **bold.**

λ	0	10	100
PSNR RAW	**51.31**	51.24	50.61
Acutance RAW	0.018	**0.011**	0.012
Acutance RGB	0.086	0.061	**0.049**

5 Conclusion

In this work, we have shown that a specific training of image restoration neural networks can greatly improve a standard evaluation metric quantifying the preservation of textures, without impairing classical performance evaluation criteria. As a proof of concept, we extended the use of the acutance loss for

real-world image denoising networks, showing that the proposed framework can improve a complete RAW images development pipeline. Considering that the texture acutance metric is routinely used to evaluate digital camera, this founding has potential important practical applications.

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1692–1700 (2018)
2. Achddou, R., Gousseau, Y., Ladjal, S.: Fully synthetic training for image restoration tasks, January 2023. https://hal.science/hal-03940525. working paper or preprint
3. Achddou, R., Gousseau, Y., Ladjal, S.: Synthetic images as a regularity prior for image restoration neural networks. In: Elmoataz, A., Fadili, J., Quéau, Y., Rabin, J., Simon, L. (eds.) SSVM 2021. LNCS, vol. 12679, pp. 333–345. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75549-2_27
4. Alvarez, L., Gousseau, Y., Morel, J.M.: The size of objects in natural and artificial images. In: Advances in Imaging and Electron Physics, vol. 111, pp. 167–242. Elsevier (1999)
5. Anwar, S., Barnes, N.: Real image denoising with feature attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3155–3164 (2019)
6. Artmann, U.: Image quality assessment using the dead leaves target: experience with the latest approach and further investigations. In: Digital Photography XI, vol. 9404, pp. 130–144. SPIE (2015)
7. Baradad, M., Wulff, J., Wang, T., Isola, P., Torralba, A.: Learning to see by looking at noise. In: Advances in Neural Information Processing Systems, vol. 34 (2021)
8. Bordenave, C., Gousseau, Y., Roueff, F.: The dead leaves model: a general tessellation modeling occlusion. Adv. Appl. Probab. **38**(1), 31–46 (2006)
9. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005), vol. 2, pp. 60–65. IEEE (2005)
10. Cao, F., Guichard, F., Hornung, H.: Measuring texture sharpness of a digital camera. In: Digital Photography V, vol. 7250, p. 72500H. International Society for Optics and Photonics (2009)
11. Cao, F., Guichard, F., Hornung, H.: Dead leaves model for measuring texture quality on a digital camera. In: Digital Photography VI, vol. 7537, p. 75370E. International Society for Optics and Photonics (2010)
12. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imag. Vis. **20**(1), 89–97 (2004)
13. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3291–3300 (2018)
14. Dabov, K., Foi, A., Egiazarian, K.: Video denoising by sparse 3D transform-domain collaborative filtering. In: European Signal Processing Conference, vol. 16, no. 8, pp. 145–149 (2007)
15. Donoho, D., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage **81**(3), 425–455 (1994)

16. Donoho, D.L., Johnstone, I.M.: Minimax estimation via wavelet shrinkage. Ann. Stat. **26**(3), 879–921 (1998)
17. Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. ACM Trans. Graph. (ToG) **35**(6), 1–12 (2016)
18. Gousseau, Y., Roueff, F.: Modeling occlusion and scaling in natural images. Multiscale Model. Simulat. **6**(1), 105–134 (2007)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
20. Ignatov, A., Van Gool, L., Timofte, R.: Replacing mobile camera ISP with a single deep learning model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 536–537 (2020)
21. Photography − Digital cameras − Part 2: Texture analysis using stochastic pattern. Standard, International Organization for Standardization, Geneva, CH (2019)
22. Lebrun, M., Buades, A., Morel, J.M.: A nonlocal Bayesian image denoising algorithm. SIAM J. Imag. Sci. **6**(3), 1665–1688 (2013)
23. Lee, A.B., Mumford, D., Huang, J.: Occlusion models for natural images: a statistical study of a scale-invariant dead leaves model. Int. J. Comput. Vision **41**(1–2), 35–59 (2001)
24. Liu, Y., et al.: Invertible denoising network: a light solution for real noise removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13365–13374 (2021)
25. Madhusudana, P.C., Lee, S.J., Sheikh, H.R.: Revisiting dead leaves model: training with synthetic data. IEEE Signal Process. Lett. (2021)
26. Matheron, G.: Random sets and integral geometry (1975)
27. Movshon, J.A., Kiorpes, L.: Analysis of the development of spatial contrast sensitivity in monkey and human infants. JOSA A **5**(12), 2166–2172 (1988)
28. Prashnani, E., Cai, H., Mostofi, Y., Sen, P.: PieAPP: perceptual image-error assessment through pairwise preference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1808–1817 (2018)
29. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
30. Wei, K., Fu, Y., Zheng, Y., Yang, J.: Physics-based noise modeling for extreme low-light photography. IEEE Trans. Pattern Anal. Mach. Intell. (2021)
31. Yu, G., Sapiro, G.: DCT image denoising: a simple and effective image denoising algorithm. Image Process. Line **1**, 292–296 (2011). https://doi.org/10.5201/ipol.2011.ys-dct
32. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017)
33. Zhang, K., Zuo, W., Zhang, L.: FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans. Image Process. **27**(9), 4608–4622 (2018)