2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Blind microscopy image denoising with a deep residual and multiscale encoder/decoder network.

Fabio Hernan Gil Zuluaga[1,2], Francesco Bardozzo[1], Jorge Ivan Rios Patino[2], Roberto Tagliaferri[1], *Senior Member, IEEE*

***Abstract*— In computer-aided diagnosis (CAD) focused on microscopy, denoising improves the quality of image analysis. In general, the accuracy of this process may depend both on the experience of the microscopist and on the equipment sensitivity and specificity. A medical image could be corrupted by several perturbations during image acquisition. Nowadays, CAD deep learning applications pre-process images with image denoising models to reinforce learning and prediction. In this work, an innovative and lightweight deep multiscale convolutional encoder-decoder neural network is proposed. Specifically, the encoder uses deterministic mapping to map features into a hidden representation. Then, the latent representation is rebuilt to generate the reconstructed denoised image. Residual learning strategies are used to improve and accelerate the training process using skip connections in bridging across convolutional and deconvolutional layers. The proposed model reaches on average 38.38 of PSNR and 0.98 of SSIM on a test set of 57458 images overcoming state-of-the-art models in the same application domain.**

***Clinical relevance* - Encoder-decoder based denoiser enables industry experts to provide more accurate and reliable medical interpretation and diagnosis in a variety of fields, from microscopy to surgery, with the benefit of real-time processing.**

I. Introduction

Medical image denoising is a well-known ill-posed inverse problem that has been extensively studied in the past decades (traditional models) and recently improved with deep learning approaches. It is possible to divide the traditional models into four main typologies: **(i)** *Spatial Domain Filtering* approaches, with Least Mean, Non-Local mean (NLM) and K-Means Singular Value Decomposition (K-SVD). **(ii)** *Transform Domain Filtering* with Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) and Block Matching 3-D (BM3D). **(iii)** Other domains covered by *Markov Random Fields* (MRF), Maximum a posteriori probability estimator (MAP), **(iv)** *Sparse Representations* with learned simultaneous sparse coding (LSSC), Convolution Sparse Representation (CSR). However, for a complete survey please refer to [1]. Modern applications for medical image denoising are mainly developed with deep learning models. In [2], for example, ad-hoc convolutional denoising autoencoders (CDAE) are used to denoise medical images corrupted with different noise types. Later in [3], an encoder-decoder neural network is designed to handle different noise levels by introducing skip connections. In the following year, in [4], a very deep convolutional neural network faced the problem of denoising using residual learning [5] and batch normalization [6]. Moreover, in [7], a variable splitting technique is used for denoising. In [8], a different approach with reversible downsampling operation and tunable noise map is proved to be an effective denoising method. For example, [9], [4] improve the chest radiographs reconstruction quality with slight modifications from the previous cited models. In [10], a dynamic residual attention network with noise gate is introduced to denoise medical images of different typologies. With respect to previous models, our work introduces a lightweight convolutional neural network, making possible to transfer the trained networks on Lab-On-Chip applications. Furthermore, the introduced model obtains better results with respect to state-of-the-art models with a relevant generalization power. In fact, our model is able to deal with unknown noise characteristics (blind denoising) in a wide range of σ ($\sigma \in [0, 50]$). In detail, noise can be generated due several issues in scanning procedures [11], [12] as well as improper staining [13]. Our model is able to reduce artifacts leveraging its multi-scale layered architecture (see also Fig **??**). Important details of the tissues and/or cell bodies or nuclei are preserved both by the architectural design and by rigid pixel-by-pixel based losses (i.e. mean absolute error). The paper is organized as follows: in section II the model architecture and the dataset are described. In section III experiments and results are presented and discussed, followed by section IV with the conclusions.

II. Methods

In section II-A the dataset is provided, while in section II-B the procedure for preprocessing is shown. Finally, the model architecture is explained in section II-C.

A. Dataset

Our deep learning model is trained and tested on a large collection of microscopy images from Histopathologic Detection Dataset[1]. In total the dataset contains 220025 training microscopy images and 57458 test images with size 96×96 pixels on three channels (RGB). In detail, Fig. 2 shows a sample set, illustrating various degrees of luminance, contrast and structure.

[1]Link: Histopathologic Cancer Detection Dataset

*This work was supported by Università degli Studi di Salerno

[1,2] Fabio Hernan Gil Zuluaga is with Faculty of Engineering, Universidad Tecnológica de Pereira and Neuronelab, DISA-MIS, Università degli Studi di Salerno. fhgil@utp.edu.co

[1] Francesco Bardozzo and Roberto Tagliaferri are with Neuronelab, DISA-MIS, Università degli Studi di Salerno {fbardozzo, robtag}@unisa.it

[2] Jorge Ivan Rios Patino is with Faculty of Engineering, Universidad Tecnológica de Pereira jirios@utp.edu.co

B. Image pre-processing and experimental setup

A synthetic generated Additive white Gaussian noise (AWGN) is added to the microscopy images. AWGN follows the standard assumption that there is no prior information of the type of noise perturbation. This is also according to [14] where real-world noise can be approximated as locally AWGN. The noise generator is built with the numpy library [15]. The 220025 images were corrupted with standard deviation in the range between 0 and 50 ($\sigma \in [0, 50]$). The perturbations are equally distributed over the total of the images, obtaining sets of 4314 images each one belonging to a σ level (e.g 4314 images for $\sigma = 1$, 4314 images for $\sigma = 2$, and so on). In this way we obtained a training set of images divided into 51 subsets each one with a different value of σ from 0 to 50. The proposed model is trained by using all these subsets. Moreover, despite the work in [8], [7] and [9], in which the training dataset was generated with fixed σ, we used a multi sigma training set; in fact, these works trained different networks for each specific sigma while we trained a single network capable of handling noise levels, ranging from 0 to 50. In detail, this means that the network is able to perform the so called *blind denoising procedure* after training. In other words, our network can deal with noisy images without knowing its characteristic perturbations (i.e. noise intensity, distribution, standard deviation, etc...). The different noise levels are generated with a fixed seed to ensure fair comparison and experiment reproducibility. In detail, all the noise maps are created with mean (μ = 0), $\sigma \in [0, 50]$. It is performed a pixel-wise 8-bit quantization in range $[0, 255]$ (please for more details refer to our online repository [2])

C. Model architecture

Our model architecture is inspired to the unsupervised denoising autoencoders provided by [16], [2]. However, it is not an autoencoder (in the strict sense of the term) but, more precisely, an encoder/decoder network. The whole model architecture is described in Figure 1. Given x the clean image and x^* the noised one, the objective is to learn a mapping from x^* (noisy image) to its denoised representation z (reconstructed image). Formally, the model m can be represented as $m(x \mid x^*; \theta)$; with θ parameters to be learned. Initially, as it is shown in Figure 3, the original microscopy image can be represented as a *d*-dimensional space with pixel intensities normalized between 0 and 1 ($x \in [0, 1]^d$). Then, this space is corrupted by means of a stochastic mapping $x^* \sim q_d(x^* \mid x)$ where x^* is a corrupted version of x (see also Section II-B). In the encoder f_θ, the corrupted x^* is mapped into a hidden representation $y = f_\theta(x^*) = \delta_{W,b}(x^*)$. The activation function δ is the Rectified Linear Unit (ReLU). While, the learnable parameter θ is equal to {W,b}, with *W* the weight matrices and *b* the biases. The decoder $g_{\theta'}$ reconstruct the original image from the latent space ($z = g_{\theta'}(y) = \delta^{'}_{W',b'}(y)$. The parameters $W, b, W^{'}, b^{'}$ are obtained by minimizing the reconstruction error between the original image (*x*) and the reconstructed one (*z*) (see also Fig 3). The mean absolute error (MAE) is the loss function that our optimizer tries to minimize. In detail, the model architecture is designed leveraging the interplay between two *inception blocks* [17] (Figure 1 - Box **(b)** and **(c)**). According to the denoise model of [18], the multiscale configuration is adopted because it performs better on difficult image microscopy areas (edges and homogeneous textures). To reduce the vanishing gradient problem [19], the network architecture is designed wider rather than deep with a strategic positioning of *skip connections*. In detail, two different types of *skip connections* (by layer concatenation) are designed to provide an alternative gradient path in back-propagation. The first type of skip connections are positioned between the encoder and the decoder (see Figure 1 - Box **(a)**). In detail, they are typically adopted to avoid information loss (see also [3], [20]). The second type of skip connections are suited inside the two *inception blocks* and named *shortcut connections* (see also Figure 1 - Box **(b)** and **(c)**). In some situations, *shortcut connections* increase model accuracy by leveraging residual learning approaches [4], [7]

[2]GitHub Repository - IRUNet

III. Experiments and Discussion

In section III-A, training process and prediction time are presented. In section III-B our model results are shown in comparison with traditional [21] and state-of-the-art deep learning models [4], [9], [10].

A. Model configurations

The network is trained with Adam optimizer for a total of 123.379 trainable parameters with $b_1 = 0.9$, $b_2 = 0.999$ and ϵ equal to $1 * 10^{-7}$. The learning rate is of $1 * 10^{-4}$. The hyperparameters tuning comes through a grid search on filter selection, learning rate monitoring, skip connection positioning and several cost functions testing. The evaluation of predictions and model performances are based on PSNR evaluations. The model was trained with Nvidia GeForce GTX 1080, processor Intel® Xeon(R) CPU E5-2630 v4 @ 2.20GHz × 20, employing Tensorflow v2.2, Cuda and Cudnn v10.1 with Python v3.8. Regarding computation performance, the average prediction time of the network is approximately 0.03 seconds per image.

B. Model performances

As it is shown in Table I, the proposed architecture outperforms the other denoising methods using as reconstruction measures PSNR and SSIM; for PSNR at $\sigma = 10$ the difference between the proposed method and the second and third top methods are $2.84dB$ and $5.27dB$, respectively; at $\sigma = 25$ the gap increases to $9.66dB$ with respect to DRAN and $10.41dB$ to Residual MID. When $\sigma = 25$ it is obtained the biggest improvement over the three σ evaluations. In the last comparison, with $\sigma = 50$, the differences with the second and third best evaluations were $5.25dB$ and $11.66dB$, respectively. Similar results can be seen for SSIM, when our network reached the highest values in all three σ evaluations, being the only network with values over 0.96. In Fig.

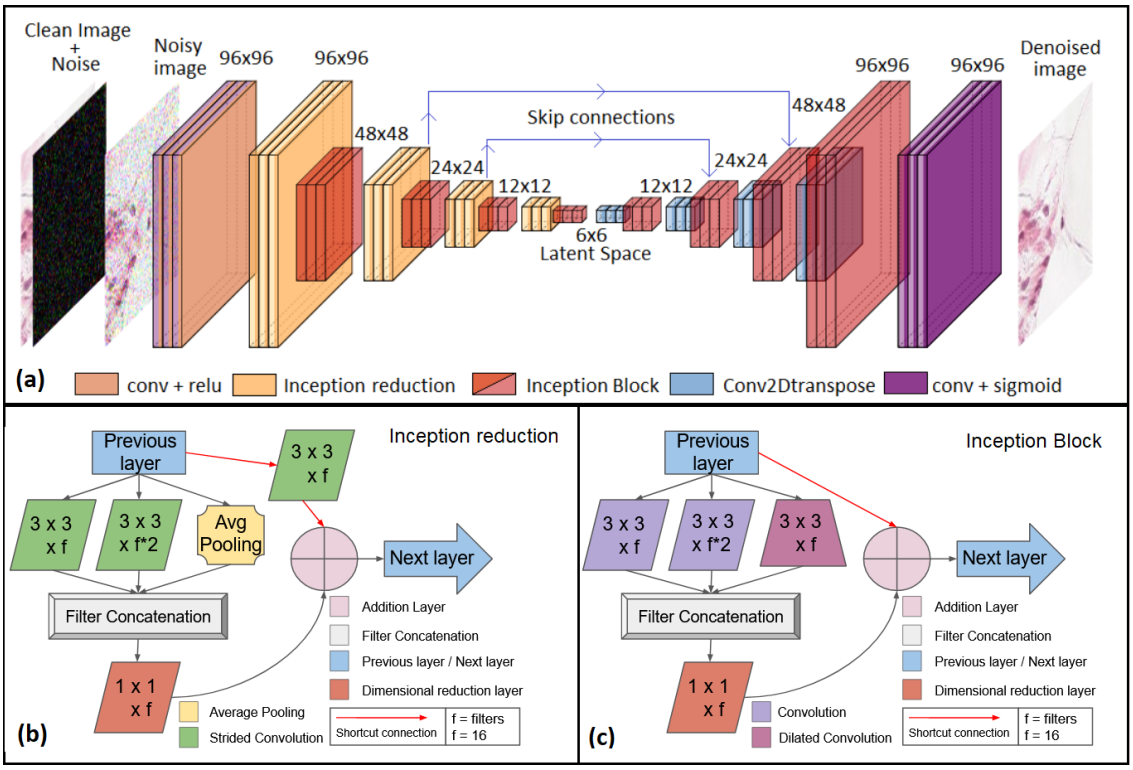

Fig. 1. Figure 1 - Box **(a)** shows the proposed network architecture: the noised image x^* is given as input, the first layer (conv + relu) maps the initial features followed by four *inception reduction* and *inception blocks* building the latent space y. The image reconstruction is composed of four transposed convolutions and inceptions blocks, the last layer is a convolutional layer with a sigmoid activation function (conv + sigmoid). Figure 1 Box **(b)** shows the proposed *inception reduction block*, the main branch has two strided convolution and average pooling that are merged in the concatenation layer. They are followed by a dimensional reduction layer. The previous layer uses the *shortcut connections* for residual learning by executing a strided convolution to match the spatial reduction occurred in the main path due to strided convolutions and average pooling. The addition layer, at the end, sum the weights and passes the output to the next layer. Fig. 1 - Box **(c)** shows the proposed *inception block*: the main path has two convolutions and a dilated convolution that are merged in the concatenation layer. They are followed by a dimensional reduction layer. The previous layer uses the *shortcut connections* for residual learning. Finally, the addition layer sum the weights and passes the output to the next layer.

4, the reconstruction quality can be evaluated considering homogeneous areas, edges and image borders.

IV. Conclusion

We presented a novel light weight CNN, that compares well with state-of-the-art methodologies both classical and deep neural networks. Our model takes advantages both from its architecture and from the learning of multi-σ images. Given the reduced number of learned parameters, the trained network can work on Lab-On-Chip applications. Future work includes new medical image typologies and higher degrees of noise map spatial distributions to increase the generalization power. Future investigations will include more robust control mechanisms that will be tested on larger datasets. For exam-

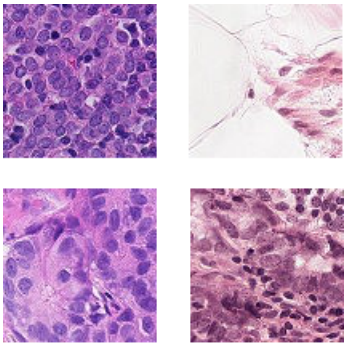

Fig. 2. The figure shows four microscopy image tissues from the Histopathologic Cancer Detection dataset

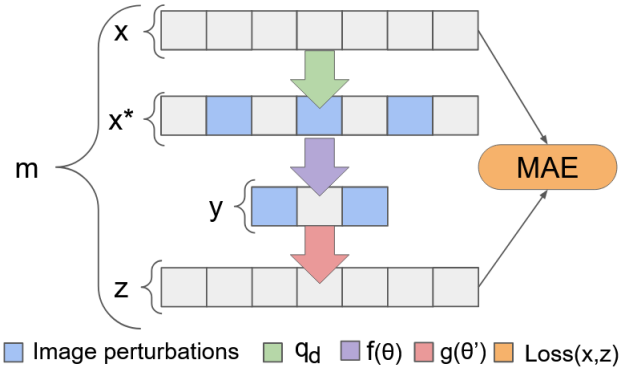

Fig. 3. Encoder-decoder pipeline: the perturbation of the clean image x is done by q_d obtaining the noisy image x^*. The encoder f_θ maps into a latent space y. The decoder $g_{\theta'}$ takes the latent space as input and outputs an approximation of x, producing z. Finally, the model tries to minimize during the epochs the reconstruction error between x and z (loss(x,z)).

TABLE I

RESULTS AND COMPARISONS

Model	σ	PSNR	SSIM
BMD3* [21]	10	28.19	0.6670
DnCNN [4]		35.26	0.8119
Residual MID [9]		36.93	0.8769
DRAN [10]		39.36	0.9735
IRUNet (Proposed)		**42.20**	**0.9977**
BMD3* [21]	25	25.02	0.5042
DnCNN [4]		26.70	0.7976
Residual MID [9]		29.23	0.8518
DRAN [10]		29.98	0.8993
IRUNet (Proposed)		**39.64**	**0.9925**
BMD3* [21]	50	20.14	0.4248
DnCNN [4]		21.49	0.5046
Residual MID [9]		21.65	0.5652
DRAN [10]		28.06	0.8198
IRUNet (Proposed)		**33.31**	**0.9655**

- Note: The traditional model is shown with *.

ple, any deformations induced by an incorrect reconstruction of tissues or cellular details could be controlled and corrected both with deep learning based approaches [22], [23] or accurate thresholding techniques [24].

REFERENCES

[1] B. Goyal, A. Dogra, S. Agrawal, B. Sohi, and A. Sharma, "Image denoising review: From classical to state-of-the-art approaches," *Information fusion*, vol. 55, pp. 220–244, 2020.

[2] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2016, pp. 241–246.

[3] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," *arXiv preprint arXiv:1606.08921*, 2016.

[4] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[6] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[7] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3929–3938.

[8] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[9] W. Jifara, F. Jiang, S. Rho, M. Cheng, and S. Liu, "Medical image denoising using convolutional neural network: a residual learning approach," *The Journal of Supercomputing*, vol. 75, no. 2, pp. 704–718, 2019.

[10] S. Sharif, R. A. Naqvi, and M. Biswas, "Learning medical image denoising with deep dynamic residual attention network," *Mathematics*, vol. 8, no. 12, p. 2192, 2020.

[11] S. M. Ayyad, M. Shehata, A. Shalaby, A. El-Ghar, M. Ghazal, M. El-Melegy, N. B. Abdel-Hamid, L. M. Labib, H. A. Ali, A. El-Baz *et al.*, "Role of ai and histopathological images in detecting prostate cancer: A survey," *Sensors*, vol. 21, no. 8, p. 2586, 2021.

[12] N. Elazab, H. Soliman, S. El-Sappagh, S. Islam, and M. Elmogy, "Objective diagnosis for histopathological images based on machine learning techniques: Classical approaches and new trends," *Mathematics*, vol. 8, no. 11, p. 1863, 2020.

[13] Y. Kurmi, V. Chaurasia, and N. Kapoor, "Histopathology image segmentation and classification for cancer revelation," *Signal, Image and Video Processing*, pp. 1–9, 2021.

[14] J.-S. Lee, "Refined filtering of image noise using local statistics," *Computer graphics and image processing*, vol. 15, no. 4, pp. 380–389, 1981.

[15] T. E. Oliphant, *A guide to NumPy*. Trelgol Publishing USA, 2006, vol. 1.

[16] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1096–1103.

[17] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[18] Y. Song, Y. Zhu, and X. Du, "Grouped multi-scale network for real-world image denoising," *IEEE Signal Processing Letters*, vol. 27, pp. 2124–2128, 2020.

[19] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[22] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural Networks*, 2020.

[23] G. Ciaparrone, F. Bardozzo, M. D. Priscoli, J. L. Kallewaard, M. R. Zuluaga, and R. Tagliaferri, "A comparative analysis of multi-backbone mask r-cnn for surgical tools detection," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[24] F. Bardozzo, B. De La Osa, L. Horanská, J. Fumanal-Idocin, M. delli Priscoli, L. Troiano, R. Tagliaferri, J. Fernandez, and H. Bustince, "Sugeno integral generalization applied to improve adaptive image binarization," *Information Fusion*, vol. 68, pp. 37–45, 2021.

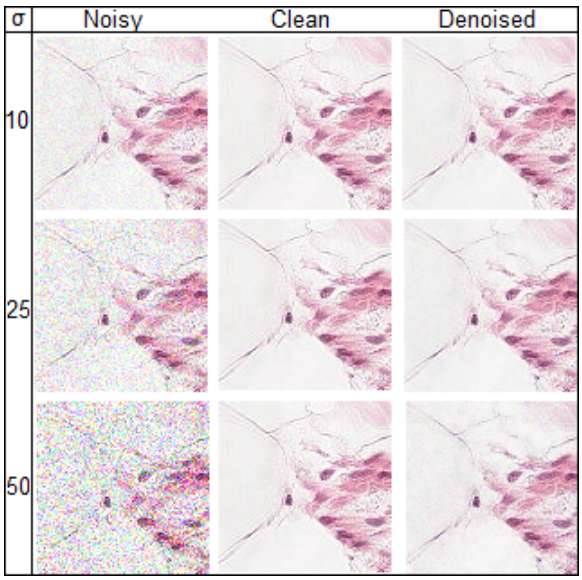

Fig. 4. In Figure 4 three samples perturbed by three $\sigma = [10, 25, 50]$ noise levels are shown. In detail, in the first coloumn (*Noisy*) the noisy images are depicted, while the ground truth is labelled as *Clean*, finally, in the third coloumn, the denoised images are shown. As it is described in Section III-B, our model is able to remove the various levels of noise over the homogeneous areas and along the edges.