FLASHLIGHT CNN IMAGE DENOISING

Pham Huu Thanh Binh [1], *Cristóvão Cruz* [1], *Karen Egiazarian* [1,2]

[1] Noiseless Imaging Ltd, Tampere, Finland,
[2] Faculty of Information Technology and Communication Sciences, Tampere University, Finland

ABSTRACT

This paper proposes a learning-based denoising method called FlashLight CNN (FLCNN) that implements a deep neural network for image denoising. The proposed approach is based on deep residual networks and inception networks and it is able to leverage many more parameters than residual networks alone for denoising grayscale images corrupted by additive white Gaussian noise (AWGN). FlashLight CNN demonstrates state of the art performance when compared quantitatively and visually with the current state of the art image denoising methods.

Index Terms— Image Denoising, Convolutional Neural Networks, Inception, Residual Learning, Gaussian Noise.

I. INTRODUCTION

Image denoising is a fundamental problem in image processing aiming at reconstructing an image from its noisy measurement. Since 2005 for a decade this field has been dominated by non-local transform domain patch based methods, such as BM3D [1], [2], and its modifications, BM3D-SAPCA [3] and WNNM [4]. Most notably, BM3D has been the state of the art method until the recent rapid advancement of Machine Learning (ML) based approaches, more specifically, Deep Neural Networks (DNN) based approaches. In contrast with the *traditional* approaches, DNN based solutions employ a vast dataset of examples and learn how to invert the degradation function [5]. These methods saw initial success in the field of computer vision [6], [7], [8], [9], but it was quickly realized that they could also be used in image denoising and other image restoration tasks. While in the field of computer vision the network learns a mapping between input image and image label [7], when applied as a denoiser it instead learns a mapping between a degraded and a clean image [10]. The introduction of techniques such as residual learning [6] and batch normalization [11] allowed the depth of these networks to increase over time along with their performance, also in the field of image denoising [10]. Currently, the state of the art in image denoising is dominated by Denoising convolutional neural network (DnCNN) based methods [10] and its modifications, FFDNet [12], IRCNN [13], HRLNet [14] and others. Despite their recent success, DNN based approaches suffer from diminishing feature reuse and are unable to take advantage of an increased number of parameters, be it either by increasing the number of layers or using wider kernels per layer. Furthermore, they have been shown to exhibit a narrow receptive field [15] which limits their ability to take advantage of long range correlations.

This paper proposes a convolutional neural network (CNN) a network called FlashLight CNN inspired by DnCNN [10] and Inception-ResNet [16] architectures that solves the above-mentioned issues. The main goal of the proposed network is to overcome the dimishing feature reuse by use of inception layers in such a way that an increase in the number of parameters of the network leads to increased performance. Additionally, by using layers with much wider support, we are able to effectively increase the receptive field of the network and its ability to restore image content. The proposed approach demonstrates state of the art performance when compared to current image denoising methods.

II. BACKGROUND

With the increased availability of computational resources, CNNs have the opportunity to grow and employ more and more parameters [17]. However, naïve approaches to increase the number of parameters of a CNNs by increasing the number of layers has resulted in decreased performance. This effect has been blamed in issues such as diminishing feature reuse and narrow receptive fields [6]. Several solutions have been proposed for these issues, which include the use of skip connections [18] and wide-residual layers [6].

Skip connections allow a network to learn a, so called, residual mapping. They consist of an identity mapping placed between two non-adjacent layers[18]. [6] showed that when these connections are used in every layer, increasing the network depth translates into performance gains, as opposed to performance loss observed when these connections are not used. We also know from DnCNN [10] that even when considering shallower networks, using just one skip connection between the input and the output of the whole network improves performance. These networks are called residual networks.

Increasing the depth of the networks also slows down training and can lead to diminishing feature reuse [19]. So on top of using skip connections, [19] propose that layers should also be made wider and thicker by adding more feature

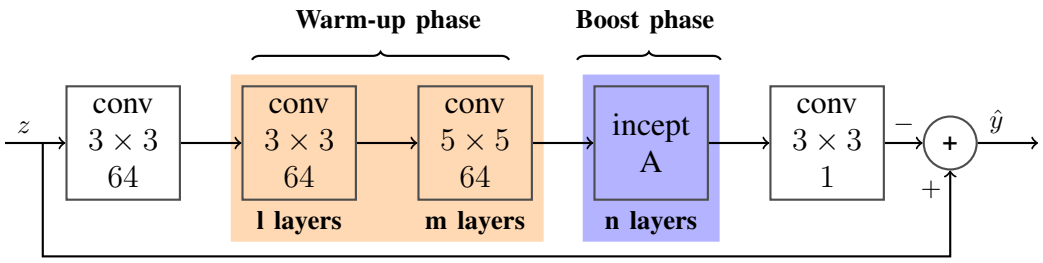

Fig. 1. The proposed Flash Light CNN architecture for denoising, with noisy input z and estimate $\hat{y}$. The orange background marks the *warmup* phase while the blue background marks the *boost* phase. Batch normalization and *relu* units were omitted for sake of clarity.

planes and more convolution kernels per network layer, so each network layer would contain many more parameters, leading to the wide residual layers. Networks equipped with these wide residual layers performed better than networks with the *regular* layers, when the number of parameters remained constant. An added bonus was that these shallower networks train much faster. Bottom line, if one can afford more parameters, one should not add more layers, but make them wider and thicker instead.

One notable network that successfully combined several of these techniques is the Inception Network [16]. The Inception Network has gained in popularity since 2014 when it achieved the top position in the ILSVRC 2014 competition [20]. There are different versions of it, with the latest ones being Inception-v4 and Inception-Resnet [16]. On top of the use of skip connections and wide-residual layers, Inception-Resnet also uses cascades of small kernels as opposed to big kernels in an attempt to reduce the overall number of parameters while maintaining the depth of the networks [17].

III. PROPOSED METHOD: FLASHLIGHT CNN

We propose FlashLight CNN, a network architecture that combines elements from DnCNN and Inception-Resnet to combat diminishing feature reuse and successfully leverage a significantly increased number of parameters for the task of denoising grayscale images corrupted by AWGN.

FlashLight CNN is made up two phases: *warmup* and *boost*, with a residual skip connection between the input and the output, as shown in Fig. 1. The *warmup* phase uses only *conventional* convolutional layers and resembles a *typical* CNN. The *boost* phase on the other hand, uses much wider residual inception layers that rapidly increase the number of parameters of the network while avoiding the diminishing feature reuse that would ensue if only conventional convolutional layers would be employed. The inception layers used in this network, shown on Fig. 2, were based on the work of [16]. They employ input dimensionality reduction as a way to reduce the computational complexity. They also use cascades of smaller filter banks instead of a

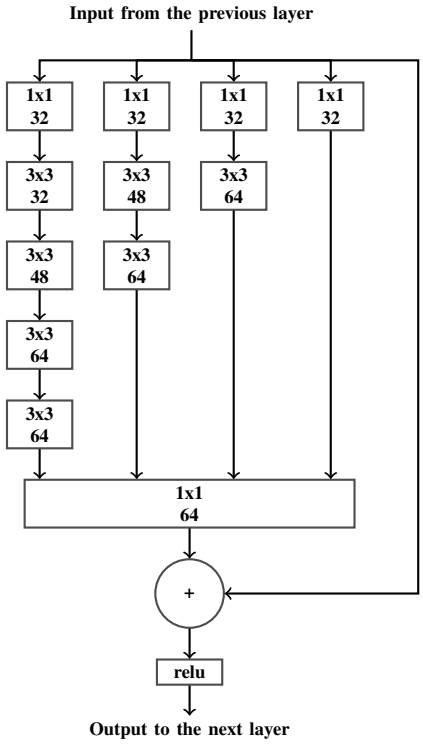

Fig. 2. Inception layer used in the proposed architecture.

single big filter in order to reduce the number of parameters required by each layer. Finally, they sport a residual skip connection that has been shown to be an effective way of avoiding the diminishing feature reuse that comes with the increase of the number of parameters in the network.

The *warmup* phase is composed of two stages of layers, with the first stage employing only 3×3 kernels and the second stage employing bigger 5×5 kernels. Both of the stages extract 64 features per layer. The purpose of this phase is to extract low level features from the input image that can then be processed by the much more capable *boost* phase. The use of wider kernels in the second stage allows

Dataset	σ	BM3D [1] PSNR	BM3D [1] SSIM	DnCNN [10] PSNR	DnCNN [10] SSIM	FFDnet [12] PSNR	FFDnet [12] SSIM	IRCNN [13] PSNR	IRCNN [13] SSIM	HRLNet [14] PSNR	HRLNet [14] SSIM	FLCNN PSNR	FLCNN SSIM
set12	15	32.41	0.8959	32.87	0.9030	32.77	0.9033	32.77	0.9009	—	—	32.97	0.9053
	25	30.00	0.8505	30.44	0.8616	30.45	0.8639	30.38	0.8597	30.46	0.8368	30.66	0.8673
	50	26.76	0.7660	27.19	0.7822	27.33	0.7896	27.14	0.7795	27.29	0.7369	27.51	0.7955
bsd68	15	31.13	0.8741	31.74	0.8908	31.62	0.8902	31.63	0.8881	—	—	31.78	0.8928
	25	28.61	0.8024	29.23	0.8279	29.19	0.8290	29.14	0.8247	29.14	0.8238	29.33	0.8326
	50	25.69	0.6881	26.23	0.7183	26.30	0.7242	26.18	0.7162	26.16	0.7143	26.40	0.7291
urban100	15	32.40	0.9232	32.68	0.9250	32.44	0.9277	32.49	0.9244	—	—	33.02	0.9323
	25	29.77	0.8790	29.97	0.8789	29.95	0.8895	29.82	0.8839	—	—	30.53	0.8962
	50	26.08	0.7797	26.28	0.7864	26.55	0.8060	26.24	0.7927	—	—	27.05	0.8183

Table I. Performance comparison in terms of PSNR and SSIM on Set12, BSD68 and Urban100 with noise levels of 15, 25, 50. The unavailable values are replaced by "—".

the gradual widening of the receptive field before the boost phase. The *boost* phase in turn uses residual *inception* layers, the model of which is shown in Fig. 2. The combination of all these features leads to a network that uses more processing as the level of abstraction increases, that is, as we move away from the pixel domain in the input. The network is also progressively wider allowing longer range connections to be established as the abstraction level increases towards the output. The progressive widening of the layer's receptive field inspired us to name the network: FlashLight CNN.

After having defined the overall architecture of the network, the only remaining free parameters are the number of layers in each stage, identified in Fig. 1 by l, m, n. We used exhaustive search to find the best set of parameters, but in order to keep this search tractable, we set constraints on the values taken by each parameter based on our expectation of the network behaviour. We fixed $l = 5$ and set the search space for the other parameters as $m \in \{3, 4, 5\}, n \in \{3, 4, 5, 6, 7\}$.

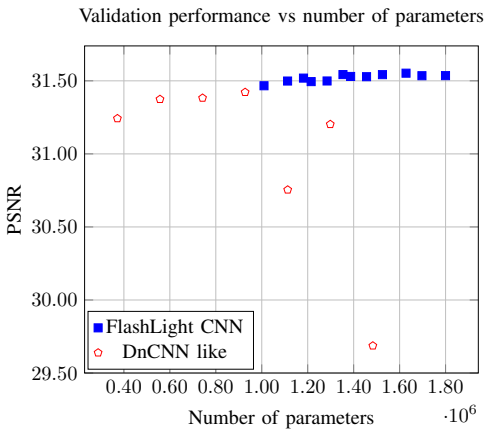

Fig. 3. Validation performances vs number of parameters, when the number of parameters of NN increases from one to about two million parameters. The performance of *DnCNN* like network decreases drastically, while FlashLight CNN sees increased performance.

During the search for the best architecture, we observed that an increased number of parameters translated, up to a certain point, to increased performance, as can be observed in Fig. 3. This behaviour confirms that the proposed architecture is indeed able to leverage the extra parameters that are made available to the network. Furthermore, we also infer from the comparison with a *DnCNN* like network, which corresponds roughly to our *warmup* phase, that the *boost* phase is essential to the increased performance.

Based on the validation performance of these experiments presented in Fig. 3, our final configuration is defined by: $l = 5, m = 4, n = 6$, to a total of 15 layers and 1627905 trainable parameters.

During the training process we used the mean squared error (MSE) loss function, 55 epochs, epoch length 4096 and batch size 64. The network weights are initialized by orthogonal method [57] and we use the Adam optimizer. We set the initial learning rate to 1×10^{-3} and modulate the learning rate using a step function that drops to 1×10^{-4} after 30 epochs. We use batch normalization [11] before every Rectified linear unit (ReLu) activation function with exception of the first and last layers. We trained and validated using the DIV2K dataset training and validation splits respectively [21].

IV. EXPERIMENTAL EVALUATION

Our introduced FlashLight CNN is evaluated over three common datasets: Set12, BSD68 and Urban100 with AWGN noise levels of $\sigma \in 15, 25, 50$ and compared with state-of-the-art methods, namely, BM3D [22] , DnCNN [10], FFDNet [12], IRCNN [13], and HRLNet [14]. The results of evaluation in terms of PSNR and SSIM metric values are depicted in Table I. Our proposed method exhibits better performance than the other methods in the comparison. Notably it performs significantly better than DnCNN for all noise levels and datasets.

Fig. 4 shows several examples to demonstrate the visual performances of the proposed solution. The proposed method

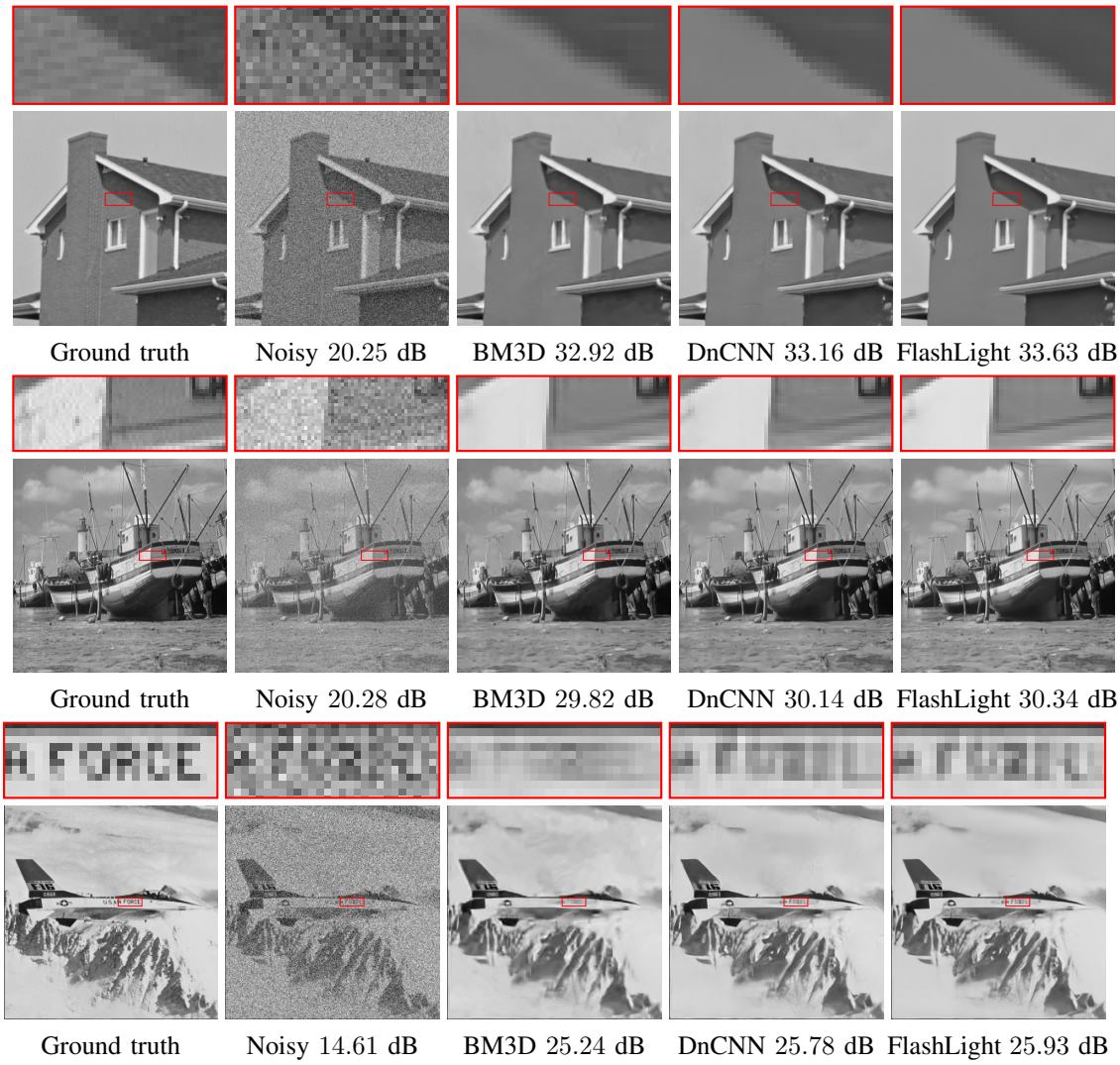

Fig. 4. Visual results with corresponding PSNR for the house and boat with $\sigma = 25$ and the plane with $\sigma = 50$ on *Set12*.

recovers better the edge patterns than other competing methods. In the *house* and *boat* images, the line shadow and the line below the text are more effectively recovered. For the stronger noise level 50 used in the *plane* our proposal exhibits superior performance recovering the text.

The code and models used for this evaluation can be downloaded in https://github.com/binhpht/flashlightCNN.

V. DISCUSSION AND CONCLUSION

We presented FlashLight CNN, a deep neural network that is able to leverage more parameters than residual networks for the task of image denoising. We showed that the performance of the proposed network increases as the number of parameters is increased. However, the increased number of parameters come with an increased computational cost. While a DnCNN like network with 557057 parameters takes 1.55 seconds to process all images in *Set12*, FlashLight CNN, with 1361649 parameters, takes 4.48 seconds, on an NVIDIA GeForce GTX 1080 Ti. Furthermore, experimental results showed that the performance of FlashLight CNN stops increasing after roughly 1.6 million parameters. It would be worth investigating how to overcome this barrier. Finally, the proposed solution has the potential to be successfully applied to other image processing tasks, such as multispectral image denoising, image super-resolution or deblurring, with minimal modifications.

VI. ACKNOWLEDGEMENTS

This work is in part supported by the Business Finland (project 3418 - E!7632 ITEA3 COMPACT, 2017-2020)

VII. REFERENCES

[1] 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, aug 2007.

[2] Dmytro Rusanovskyy and Karen Egiazarian, "Video denoising algorithm in sliding 3d dct domain," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2005, pp. 618–625.

[3] Vladimir Katkovnik, Alessandro Foi, Karen O. Egiazarian, and Jaakko Astola, "From local kernel to nonlocal multiple-model image denoising," *International Journal of Computer Vision*, vol. 86, pp. 1–32, 2009.

[4] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng, "Weighted nuclear norm minimization with application to image denoising," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. jun 2014, IEEE.

[5] Viren Jain and Sebastian Seung, "Natural image denoising with convolutional networks," in *Advances in neural information processing systems*, 2009, pp. 769–776.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, may 2017.

[8] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. jun 2016, IEEE.

[9] 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.

[10] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei 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, jul 2017.

[11] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. 2015, ICML'15, p. 448–456, JMLR.org.

[12] Kai Zhang, Wangmeng Zuo, and Lei 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, sep 2018.

[13] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang, "Learning deep CNN denoiser prior for image restoration," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. jul 2017, IEEE.

[14] Wuzhen Shi, Feng Jiang, Shengping Zhang, Rui Wang, Debin Zhao, and Huiyu Zhou, "Hierarchical residual learning for image denoising," *Signal Processing: Image Communication*, vol. 76, pp. 243–251, aug 2019.

[15] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds., pp. 4898–4906. Curran Associates, Inc., 2016.

[16] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[17] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. jun 2015, IEEE.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[19] Sergey Zagoruyko and Nikos Komodakis, "Wide residual networks," in *Proceedings of the British Machine Vision Conference (BMVC)*, Edwin R. Hancock Richard C. Wilson and William A. P. Smith, Eds. September 2016, pp. 87.1–87.12, BMVA Press.

[20] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, apr 2015.

[21] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, Bee Lim, et al., "Ntire 2017 challenge on single image super-resolution: Methods and results," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[22] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*. International Society for Optics and Photonics, 2006, vol. 6064, p. 606414.