ATTENTION-BASED NEURAL NETWORK FOR ILL-EXPOSED IMAGE CORRECTION

Lucas R. V. Messias, Paulo L. J. Drews-Jr and Silvia S. C. Botelho

Federal University of Rio Grande - FURG
Computer Science Center
Rio Grande, Brazil

ABSTRACT

The present work presents an artificial neural network architecture for the restoration of images damaged by underexposure and overexposure. The problem is relevant in computer vision applications that are applied in conditions where the limitation of the sensor prevent the scene details from being adequately represented in the captured image. This research presents an attention-based architecture composed of two convolutional neural networks, where one performs a preprocessing of the input image, while the other performs the restoration and enhancement of the degraded image. Regarding the evaluation of research results, a broad range of image quality metrics is used to assess the quality of the results produced by the model. The obtained results indicate that the proposed architecture is able to enhance images damaged by exposure heterogeneity, offering gains over state-of-art models in real data.

Index Terms— Deep Learning, Image Processing, Enhancement, Attention

1. INTRODUCTION

Digital image processing is primarily aimed at improving visual information for human interpretation, as well as processing image data for storage, transmission and representation, considering automatic perception through visual computation [1]. When it comes to image processing, a system has been defined where its input and output are an image, Scenes with a wide range of intensities values represent a challenge for image acquisition systems, directly impacting the final result. Images acquired from conventional cameras, which operate in the visible light spectrum, are commonly affected by artifacts and distortions resulting from excess or lack of light. Scene radiance outside the limits of the acquisition system results in underexposure and overexposure [2].

In image processing, underexposure is a phenomenon that occurs when the camera sensor is unable to capture differences between the darkest parts of the image, thus causing only the details located in the brightest regions of a photographed scene to be seen. Underexposure can be caused by a number of factors, including insufficient lighting, exposure time that is too short, or the lens iris aperture being too small. On the other hand, overexposure occurs when the sensor receives too much light, being a phenomenon caused by parameters wrongly configured in the camera.

Estimating the irradiance of an improperly exposed image requires restoration and enhancement of the non-clipped pixels to maximize visibility and color accuracy, as much as it requires reconstruction strategies for regions where the signal has been clipped. In this sense, deep learning models overcome the limitations of classical image enhancement methods by being able to learn objects, textures, and patterns from examples.

The present work proposes a neural architecture for single-shot contrast enhancement and image reconstruction for poorly exposed color images, composed of two networks: one dedicated to generating an exposure map of the degraded image; and the other for the purpose of restoring lost information and highlighting mismatched information. The main contributions of this work are summarized as follows:

- A new fast and small-scale deep learning architecture for image enhancement is presented.
- A degradation weighting function is projected to indicate where the information was most affected by ill-exposure degradation.
- Improved quantitative performance for ill-exposed images compared with state-of-the-art.

2. RELATED WORKS

For ill-exposed images, luminance and color correction incorporate elements from distinct areas of image processing such as contrast enhancement, signal reconstruction, noise suppression, tone mapping, and image completion. Thus, the literature includes histogram equalization [3], dehaze-based contrast enhancement [4], Retinex based contrast enhancement [5], camera response based models [6], as well as exposure fusion based models [7].

All of the previous methods do not use any neural network approach. Deep learning based image processing has

978-1-6654-9620-9/22/$31.00 ©2022 IEEE

gained a lot of attention in recent years. Its applications include super-resolution [8], inpainting [9], as well as general image enhancement [10, 11], low light image enhancement [12], and sRGB ill-exposure correction [13, 14, 15].

3. METHOD

3.1. Network Architecture

This section presents a proposed solution based on convolutional neural networks using an internal and external attention approach of the image restoration model. A small and efficient architecture is modeled, to produce qualitatively and quantitatively better results and that uses little computational resources, when compared to current deep learning architectures for image restoration tasks. Figure 1 presents a detailed view of the idealized architecture.

Attention Network. The attention network aims to generate an attention map, where the degradation resulting from inappropriate exposure has been highlighted. The main idea is to direct the restoration network to the places in the image where more focus should be given to restore the information present in the image. The attention network is represented by Figure 1b based on [11]. In the proposed model, the degraded image is submitted to seven convolutional blocks to obtain the exposure map (EM). The expansion rate of the convolutions of each block is defined by $\delta_{dilat} = 2^{n-1}$, where $n = [1, 7]$ is the number of the convolutional block. In the end, the image is processed with two more convolutional layers to generate an attention map.

The generated exposure map is defined by the equations 1 (underexposure), and 2 (overexposure), generating an array with values *range* $[0, 1]$,

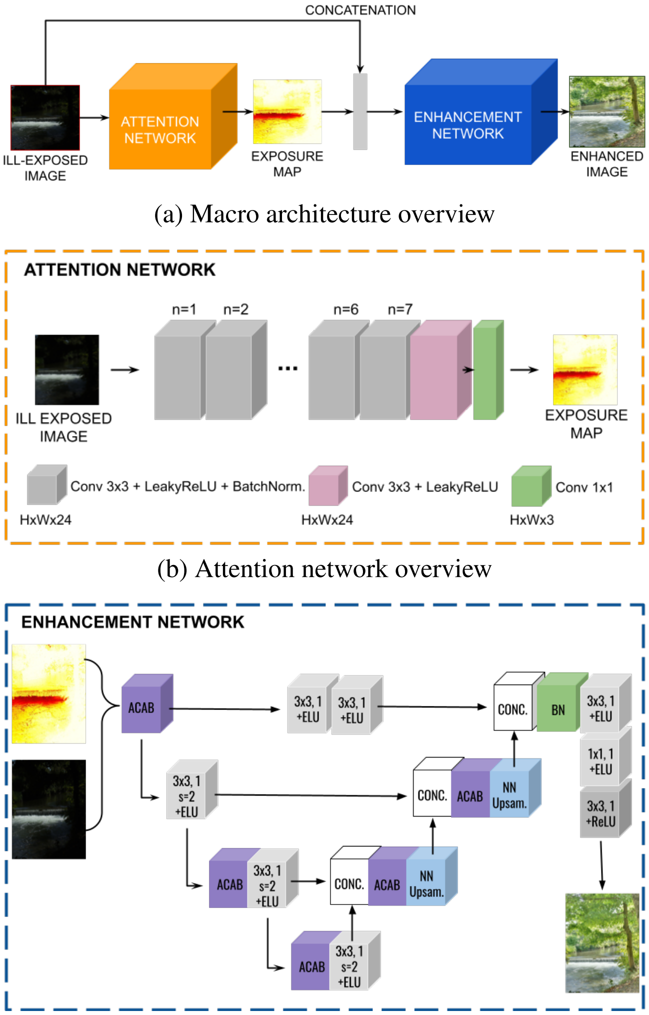
$$ME_{sub} = \frac{\mid max(\tilde{I}) - max(I) \mid}{max(\tilde{I})}, \quad (1)$$

$$ME_{sup} = \frac{\mid min(\tilde{I}) - min(I) \mid}{max(inver(\tilde{I}))}, \quad (2)$$

where $\tilde{I}$ and I represent respectively the reference image and the input image; the $max()$ and $min()$ functions return respectively an array with the maximum and minimum value of each pixel per channel; and the $inver(\cdot)$ function returns the absolute value of subtracting the input image from the highest-valued pixel.

Enhancement Network. The proposed restoration network aims to restore the poorly exposed image, focusing on the regions most affected by degradation 1c. It combines properties of the U-Net [16] and CAN (Context Aggregation Network) [11] structures. Its become a compact model with a small number of weights when a encoder-decoder model, combined with dilated convolutions.

Trainable convolutional layers were chosen to perform the down-sampling of the encoder instead of pooling alternatives. Convolutional layers are flexible and converge to a better combination for extracting features at the expense of training, whereas pooling layers have fixed characteristics. The decoder uses up-sampling by nearest neighbors followed by a convolution layer. The encoder-decoder is connected by skip-connections where the uppermost one has a block with two layers of convolutions in order to aggregate the structural information of the input. The output of this block is combined with the lower flow of the network through a convolutional block that acts as a fusion mechanism for the restored information with the structural information present in the input image.



(a) Macro architecture overview

(b) Attention network overview

(c) Enhancement network overview

Fig. 1. Overview of the proposed architecture.

3.2. Attention and Context Aggregation Block (ACAB)

The Attention and Context Aggregation Block (ACAB) (figure 2) is a combination of layers in order to work with a large receptive field and to have the ability to pay attention to important features. The main idea is to create a block with the task of directing and restoring the ”most lost” information due to ill-exposure.

First, ACAB starts with a layer of four parallel dilated convolutions. The rate of expansion is described in the form 2^{n-1}, where n defines the block and $n = \{x \in Z_+^* \mid 1 \leq x \leq 4\}$. Then, the information pass through a channel attention sub-block (AC) and a spatial attention sub-block (AS) to generate the output. the AC block is composed of two pooling functions (maximum and average) of dimension $1 \times 1 \times C$ connected in parallel followed by an MLP (Multilayer Perceptron). Its two outputs are added and applied to a sigmoid function, returning an attention map for the channel. The AS block is composed of a maximum pooling function followed by an average pooling (both of dimension $H \times W \times 1$), with a convolution layer followed by a sigmoid function generation the spatial attention map.

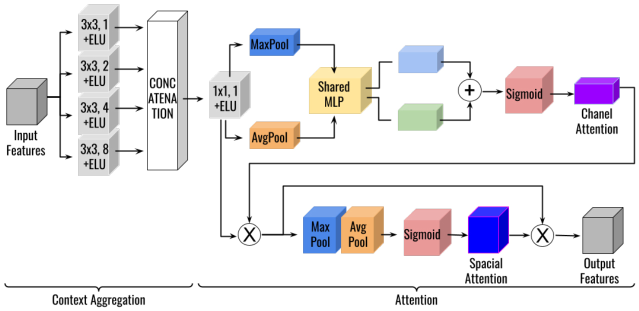


Fig. 2. ACAB architecture overview.

3.3. Loss Function

Attention Loss. The loss funcion for Attention Network is the Mean Square Error (MSE). The MSE is a function that corresponds to the expected value of the squared error loss. Equation 3 defines attention loss.

$$\mathcal{L}_{aten} = \text{MSE}(I, \tilde{I}). \tag{3}$$

Enhancement Loss. The loss function for Enhancement Network aims to emphasize regions of the image closer to the edge, where they are more likely to suffer the adverse effects of inadequate exposure. This objective function combines structural dissimilarity (DSSIM) and pixel-by-pixel absolute weighted error. Equation 4 defines enhancement loss.

$$\mathcal{L}_{enh}(I, \tilde{I}) = \lambda DSSIM(I, \tilde{I}) + (1 - \lambda)|\tilde{I} - 0,5 \mid \cdot \mid I - \tilde{I} \mid, \tag{4}$$

where $\lambda = 0.76$ is an empirical constant that was found to be more reasonable during training phase.

4. EXPERIMENTS

4.1. Datasets

We used four sets of images with ill-exposure degradation. Half of them are simulated making use of the equation given by:

$$I = f(\tilde{I}, \alpha, \beta, \gamma) = \beta \cdot (\alpha \cdot \tilde{I})^{\gamma}, \tag{5}$$

where $\tilde{I}$ is the reference image, I is the resulting degraded image, $f(\cdot)$ denotes the exposure degradation function, α, β and γ are constants generated from a uniform distribution presented earlier for underexposure in [18], and extended in this work to overexposure. The simulate datasets are generated with $\alpha = [0.9, \frac{1}{0.9}]$; $\beta = [0.5, \frac{1}{0.5}]$; and $\gamma = [\frac{1}{1.5}, 1.5]$.

Simulated: FiveK-based and HDR+ Burst. The MIT-Adobe FiveK Dataset [19] contains 5,000 photographs shot with SLR cameras from a variety of photographers. The HDR+ Burst Photography Dataset, initially presented by Hasinoff et al. [20], comprises sequences of images in different exposures by smartphone cameras.

Real: A6300 and Cai Multi-Exposure Datasets. Proposed by Steffens *et al.*[17], the A6300 dataset is composed of sets of 4 images for each scene: an appropriately exposed image using a single photograph, an underexposed image, an overexposed image, and a composition of the previous ones using the Tone Mapping method. The Cai dataset, presented in [13], it consists of 589 image sets with separate exposure settings for each scene and a tone-mapped composition.

4.2. Implementation Details

The proposed model is adjusted and tested on three different sets of images (FiveK, HDR+ Burst and Cai Multi-Exposure). In all cases, 70% of the dataset is used for training and the remainder for testing. The A6300 dataset is used only in testing phase with a state-of-art method. The samples used for each stage are selected at random. The Adam optimizer [21] is used with the standard hyper-parameters. Weights are updated in mini-batches of 8 images with varying resolution. All data used for training is paired.

Training for underexposed and overexposed images is carried out separately, resulting in a specific model for restoring underexposed images and a specific model for overexposed images. Training is terminated once 300 batches of images are processed without making improvements larger than 10^{-5}.

4.3. Main Results

This section presents a quantitative comparison of the proposed method with some image enhancement methods in the literature. The results presented are with data reserved for testing. Numerical evaluation includes several image quality measurements, including the classic signal-to-noise

Table 1. Comparison results of mean value of presented metrics .

	Under							Over						
Method	**PSNR**$\uparrow$	**MSE**$\downarrow$	**MAE**$\downarrow$	**SSIM**$\uparrow$	**Sobel IoU** $\uparrow$	**Canny IoU** $\uparrow$	**Hist. Diff.**$\downarrow$	**PSNR**$\uparrow$	**MSE**$\downarrow$	**MAE**$\downarrow$	**SSIM**$\uparrow$	**Sobel IoU** $\uparrow$	**Canny IoU** $\uparrow$	**Hist. Diff.**$\downarrow$
Proposed	**28.862**	**0.001**	**0.029**	**0.962**	**0.862**	**0.716**	**4.299**	**28.898**	**0.002**	**0.029**	**0.974**	**0.856**	**0.692**	**4.619**
[17]	22.865	0.009	0.072	0.895	0.766	0.629	5.608	20.058	0.012	0.077	0.886	0.704	0.492	5.443
[16]	22.537	0.008	0.067	0.845	0.625	0.537	5.568	18.332	0.016	0.092	0.811	0.568	0.403	5.394
[11]	21.201	0.011	0.079	0.866	0.678	0.531	6.514	19.022	0.014	0.086	0.874	0.675	0.438	4.967
[3]	17.109	0.022	0.116	0.759	0.606	0.345	6.889	12.004	0.066	0.210	0.739	0.560	0.295	6.438
[7]	17.581	0.025	0.121	0.778	0.615	0.421	6.380	9.745	0.110	0.298	0.697	0.561	0.312	8.159
[5]	16.706	0.037	0.148	0.686	0.640	0.470	6.461	12.427	0.063	0.204	0.767	0.577	0.355	6.524
[4]	16.194	0.030	0.134	0.711	0.542	0.264	6.808	13.481	0.102	0.287	0.681	0.506	0.284	7.972
[6]	15.746	0.027	0.137	0.753	0.575	0.351	6.958	9.276	0.120	0.325	0.669	0.538	0.294	8.354
None	16.150	0.069	0.189	0.610	0.617	0.523	7.291	11.431	0.081	0.235	0.777	0.601	0.381	7.308

ratio (PSNR), pixel-to-pixel mean absolute error (MAE), root mean square error (MSE) and structural similarity (SSIM), Sobel intersection over union, Canny intersection over union, and histogram difference.

Table 1 compares the proposed model with neural-based models [16, 11, 15]. It also compares to classic image enhancement approaches (which do not employ deep learning) [3, 4, 5, 7, 6]. The unprocessed images are also included in the comparison to allow an observation of the gain provided by the enhancement models. The proposed architecture outperforms all the compared models in all compared metrics.

The A6300 dataset [17] was used to compare with a state-of-art model. None of the methods had access to the data previously. Table 2 presents the mean values of the metrics for both methods. The proposed method outperforms [14] in all the quantitative metrics, demonstrating a better generalization of the problem of ill-exposed images. Table 3 represents a ablation study applied proving that the complete architecture is necessary to resolve the problem.

Table 2. Comparison with a state-of-art method.

Method	**PSNR**$\uparrow$	**MSE**$\downarrow$	**MAE**$\downarrow$	**SSIM**$\uparrow$	**Sobel IoU** $\uparrow$	**Canny IoU** $\uparrow$	**Hist. Diff.**$\downarrow$
[14]	18.669	0.019	0.111	0.819	0.614	0.378	6.652
Proposed	**28.028**	**0.002**	**0.034**	**0.954**	**0.777**	**0.645**	**4.497**

Table 3. Ablation study for the proposed method.

Configuration	**PSNR**$\uparrow$	**MSE**$\downarrow$	**MAE**$\downarrow$	**SSIM**$\uparrow$	**Sobel IoU** $\uparrow$	**Canny IoU** $\uparrow$	**Hist. Diff.**$\downarrow$
w/o AN and ACAB	16.349	0.023	0.130	0.769	0.541	0.394	7.132
w/o AN	21.026	0.008	0.071	0.888	0.678	0.412	6.049
w/o AB	25.214	0.003	0.041	0.960	0.830	0.621	4.628
Completed	**27.070**	**0.002**	**0.033**	**0.959**	**0.820**	**0.634**	**4.183**

5. CONCLUSION

We proposes a new fast and small-scale deep learning architecture for single-shot contrast enhancement and feature reconstruction of poorly exposed RGB images. Numeric comparison with others methods using four distinct datasets has shown our model significantly better in terms of brightness adjustment, contrast enhancement, image completion, and edge restoration. A ablation study confirms that the Attention Network with ACAB added on Enhancement Network brings satisfactory results for all scenarios. As future work, we plan to optimize the smoothness of recovered regions (de-blocking), the synthesis of texture, and the completion of broad clipping utilizing semantic characteristics.

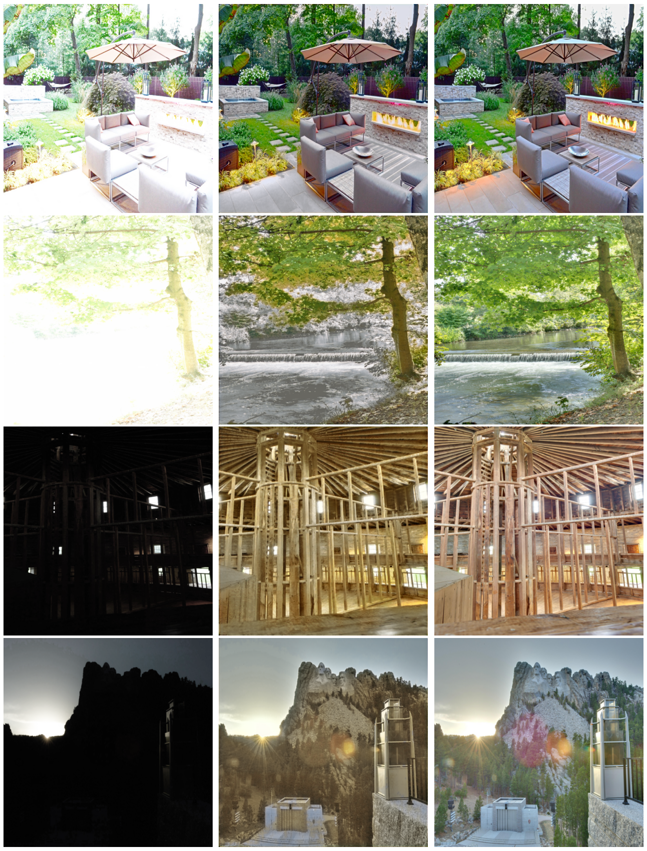


Fig. 3. Qualitative results of proposed method applied in real images. The first column is input image; the second one is the output image; and the third one is the reference image.

6. REFERENCES

[1] Rafael C Gonzalez and Richard C Woods, *Processamento digital de imagens .*, Pearson Educación, 2009.

[2] Cristiano Steffens, Paulo Drews-Jr, and Silvia Botelho, "Deep learning based exposure correction for image exposure correction with application in computer vision for robotics," in *Latin American Robotic Symposium and Brazilian Symposium on Robotics (LARS/SBR)*. IEEE, 2018, pp. 194–200.

[3] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. Akber Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 593–600, May 2007.

[4] Xuan Dong, Guan Wang, Yi Pang, Weixin Li, Jiangtao Wen, Wei Meng, and Yao Lu, "Fast efficient algorithm for enhancement of low lighting video," in *2011 IEEE International Conference on Multimedia and Expo*. IEEE, 2011, pp. 1–6.

[5] Ana Belén Petro, Catalina Sbert, and Jean-Michel Morel, "Multiscale retinex," *Image Processing On Line*, pp. 71–88, 2014.

[6] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang, "A new low-light image enhancement algorithm using camera response model," in *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3015–3022.

[7] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2017, pp. 36–46.

[8] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8798–8807.

[9] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.

[10] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool, "Wespe: weakly supervised photo enhancer for digital cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 691–700.

[11] Qifeng Chen, Jia Xu, and Vladlen Koltun, "Fast image processing with fully-convolutional networks," in *IEEE International Conference on Computer Vision*, 2017, vol. 9, pp. 2516–2525.

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

[13] Jianrui Cai, Shuhang Gu, and Lei Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.

[14] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown, "Learning multi-scale photo exposure correction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9157–9167.

[15] Cristiano R Steffens, Lucas RV Messias, Paulo JL Drews-Jr, and Silvia S d C Botelho, "Cnn based image restoration," *Journal of Intelligent & Robotic Systems*, pp. 1–19, 2020.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas 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.

[17] Cristiano R. Steffens, Lucas R. V. Messias, Paulo J. L. Drews-Jr, and Silvia S. d. C. Botelho, "Cnn based image restoration," *Journal of Intelligent & Robotic Systems*, vol. 99, no. 3, pp. 609–627, Sep 2020.

[18] Feifan Lv, Yu Li, and Feng Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," 2019.

[19] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[20] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, pp. 192, 2016.

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