CNN-BASED LUMINANCE AND COLOR CORRECTION FOR ILL-EXPOSED IMAGES

C. R. Steffens, V. Huttner, L. R. V. Messias,
P. L. J. Drews-Jr and S. S. C. Botelho*

Computer Science Center
Federal Univ. of Rio Grande - FURG
Rio Grande, Brazil

R. S. Guerra†

Dept. of Electrical Eng. Processing
Federal Univ.of Santa Maria - UFSM
Santa Maria, Brazil

ABSTRACT

Image restoration and image enhancement are critical image processing tasks since good image quality is mandatory for many image applications. We are particularly interested in the restoration of ill-exposed images. These effects are caused by sensor limitation or optical arrangement. They prevent the details of the scene from being adequately represented in the captured image. We proposed a deep neural network model due to the number of uncontrolled variables that impact the acquisition. The proposed network can converge in a representative model from the training data, loss, optimization and activation functions. The obtained results are evaluated using several image quality index which indicate that the proposed network is able to improve images damaged by heterogeneous exposure. Furthermore, our method offers a significant gain over the state-of-the-art methods both in simulated data and real data.

***Index Terms*—** Saturation, Image restoration, Image enhancement, Neural networks, Clipping

1. INTRODUCTION

During image acquisition, clipping is a common type of distortion which limits the signal representation once it exceeds a threshold. Clipping occurs when a signal is recorded by a sensor that presents limited acquisition range of light intensity. This phenomenon interferes in bright regions, resulting in overexposure, and dark regions of the scene, resulting in underexposure. In general, overexposure and underexposure are caused by poorly adjusted camera aperture, exposure time, or gain. Especially in high contrast scenes, adjusting these settings is not trivial, thus being prone to error.

Clipped pixels contain less information about the scene than other pixels. While non-saturated pixels can be linked to the incident irradiance by reversing the radiometric response function, saturated pixels only provide a lower limit to irradiance [1]. Large saturated regions pose a challenge for many classic computer vision algorithms, including popular edge detector, primitive descriptors, and feature extractors. Fourier-based reconstruction approaches also suffer from severe artifacts [2] in the case of sensor saturation.

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. With the advantage of semantic information and prior knowledge extracted from a large set of images, a neural network is able to improve the image restoration results.

We present a Convolutional Neural Network (CNN) approach for single-shot contrast enhancement and image reconstruction for ill-exposed RGB images. Our neural architecture takes inspiration on the works of [3] and [4]. The network is designed to widen the receptive field while keeps a small number of trainable parameters when compared to other state-of-the-art networks with similar purposes. The main contributions of this paper are summarized as follows: i) We present a fast and small exposure correction CNN which is able to synthesize substantial clipped parts in high-resolution image; ii) We design a custom content-based objective function to maximize restoration and reconstruction on clipped regions; and iii) We provide quantitative and qualitative results on both under and over-exposed images, outperforming recently proposed methods.

2. RELATED WORKS

The problem of luminance and color correction for ill-exposed images mixes aspects from signal restoration, denoising, contrast enhancement, color correction and tone mapping, image completion, and feature interpolation. Thus, the literature includes histogram equalization [5], shadow compensation [6], dehaze-based contrast enhancement [7],

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This work was also partly supported by CNPQ and FAPERGS.

†R. S. Guerra performed the work while at Dept. of Electrical Eng., National Taiwan Normal Univ., Taiwan.

Retinex based contrast enhancement [8], camera response based models [9], and exposure fusion based models [10].

The previous approaches rely only on the signal that is properly captured on the input image. While contrast enhancement is a task effortlessly performed by the aforementioned methods, our approach also benefits from the knowledge learned from the training data. This allows our model to generalize based on examples presented during the model adjustment to better interpolate on regions where the signal is reduced or lost.

3. DATASETS

This work intends to minimize the effects of saturation and underexposure on natural images. Therefore, we made use of two datasets (one real, one simulated) to train and evaluate the network. This enables us to further explore the behavior of our model under both real and simulated conditions.

3.1. A6300 Multi-Exposure Dataset (real)

The A6300 Multi-Exposure Dataset, previously introduced by us in [11], is composed of sets of four images for each scene: a properly exposed using a single shot, an underexposed, an overexposed, and an Image Composition of the aforementioned using [12] Tone Mapping method. The overexposed and underexposed images are obtained through exposure compensation using aperture priority, with exposure value (EV) ranging from EV -0.7 up to EV +0.7.

The dataset is acquired using a Sony A6300 camera and images are stored using JPEG compression. The lossy JPEG compression introduces additional challenges since: i) general loss of sharpness and oscillations around high-contrast edges, due to approximating intensity transitions with cosines functions; ii) blocking structure, as each image is processed separately for every 8x8 block, block edges become visible at high compression ratios; and iii) loss of color details due to compression on chromaticity channels. Although the downsides of JPEG compression are well known, this format is still the prevailing storage format for images. Thus, being a great baseline for real applications.

3.2. FiveK-based Multi-Exposure Dataset (simulated)

The MIT-Adobe FiveK Dataset [13] features 5,000 images taken with SLR cameras by a set of different photographers. The images are made available in DNG file format, a lossless raw image format. Therefore, all the information recorded by the camera sensor is preserved. This dataset does not present multi-exposure image pairs, therefore requiring preprocessing in order to be used for training.

For training, we prepare our image pairs by converting the DNG file to RGB representation and clipping the values according to a percentile, as shown in Eq. 1, where I is the reference image, $\hat{I}$ is the saturated image, P_{LT} and P_{HT} are the percentile values which define lower and upper bounds for the hard clipping.

$$C_{ij} = \begin{cases} P_{LT}, & I_{ij} \leq P_{LT} \\ I_{ij}, & P_{LT} \leq I_{ij} \leq P_{HT} \\ P_{HT}, & I_{ij} \geq P_{HT} \end{cases} \tag{1}$$

The clipping operation is followed by a min-max normalization to extend the value range over the entire representation interval:

$$\hat{I} = \frac{C_{ij} - min(C)}{max(C) - min(C)}. \tag{2}$$

Although this approach lacks in reproducing some of the saturation effects, such as blooming [1], it results in severe image damage. Thus, posing a considerable challenge in terms of restoration and reconstruction.

4. REEXPOSE-NET CNN MODEL

U-Nets [3] have recently shown great potential for image-to-image tasks [14]. Nonetheless, one significant drawback which affects all U-Net inspired models is the amount of required memory to store the partial results in the intermediate layers. This occurs due to the intensive use of skip connections, which require the early layers of the processing (encoder) to be stored as an input for the decoder layers.

We propose a new architecture based on ideas presented in [3], [11], and [15]. This new model minimizes the memory requirements of the network and improves the results for image exposure correction. The network uses atrous convolutions and trainable down-scaling and up-scaling layers. Thus, we significantly improve the prediction accuracy and increase the size of each batch during the training stage. Fig. 1 shows an abstraction of the network architecture.

4.1. Atrous convolutions

Image-to-image reconstruction, as other dense prediction tasks, calls for multiscale contextual reasoning in combination with full-resolution output [4]. In this sense, dilated convolutions can provide very large receptive fields without requiring large kernels nor deep stacks of convolutional layers. Our convolutional block includes four 3×3 parallel atrous convolutional layers, with dilatation rates ranging from 2^0 up to 2^3. Thus, each convolutional block is able to cover 19 features in the input space using only nine trainable weights for each filter. Atrous convolutions provide context aggregation for each pixel by allowing the model to access a large region in the neighborhood, allowing us to reduce the amount of scaling layers in the network.

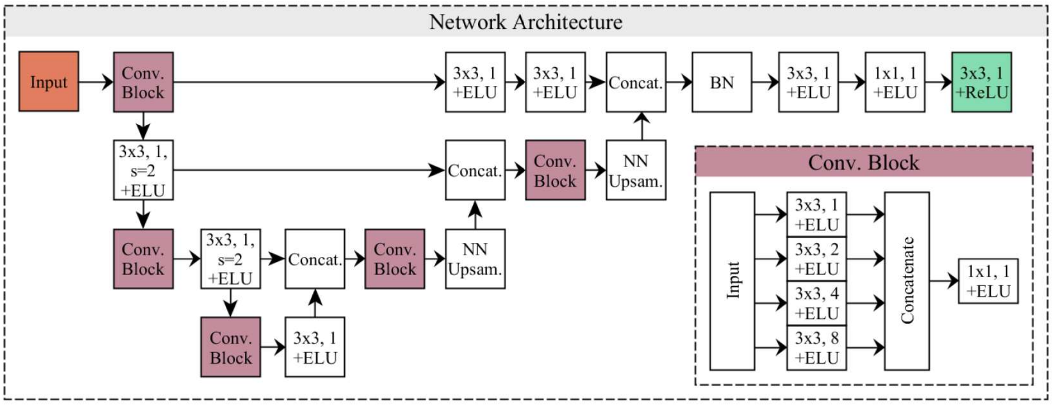

Fig. 1. The ReExpose-net architecture.

4.2. Down-sampling and Up-sampling

Spatial down-scaling is an essential component of our model leading to lower memory usage and wider receptive fields. Through experimental evaluation, we found that learned down-scaling through strided convolutions perform better than polling layers, especially in large saturated regions. Spatial up-scaling becomes necessary in this context since feature maps need to be re-scaled to the original input image size. Further exploration has shown that nearest neighbor interpolation followed by convolution result in fewer checkerboard artifacts on the output images. This fact has also been reported by [16].

4.3. Activation Function

All hidden convolutional layers in the CNN model are followed by an Exponential Linear Unit (ELU). Through experiments, we have found it significantly speeds up the learning process and leads to overall higher image quality. Firstly introduced in [17], an ELU activation is given by:

$$f(x) = \begin{cases} x, & x \geq 0 \\ \alpha\left(e^{x} - 1\right), & x < 0 \end{cases} \tag{3}$$

ELUs ability to result in negative values allow them to push the mean unit activation closer to zero, speeding up learning because they bring the gradient closer to the unit natural gradient [11]. We also use a ReLU non-linear activation function on the output layer to avoid negative outputs. ReLU is defined as:

$$g(x) = max(x, 0). \tag{4}$$

4.4. Loss Function

We adopted a custom loss function that emphasizes the image regions with values closer to the sensor limits to maximize the reconstruction and signal restoration accuracy. Our objective function combines Structural Dissimilarity (DSSIM) and Pixel-wise Euclidean Distance (L_2). DSSIM is based on SSIM [18], a similarity index calculated on various 3×3 window of an image.

Although DSSIM provides great insight on the image quality, the index is unable to assess the pixel values in the exact position. Therefore, our objective function also considers the pixel-value in the reference image. Via element-wise multiplication, we reinforce the importance of the pixels that are more likely to be affected by ill-exposure conditions. Assuming images a and b are in the representation interval $[0, 1]$, with an empirical constant $\lambda = 0.2$, the final loss function is given by:

$$\mathcal{L}(a, b) = \lambda|0.5 - b| \circ L_2(a, b) + (1 - \lambda)DSSIM(a, b). \tag{5}$$

4.5. Training and Early Stopping Criteria

Initially, we reserve 30% of each dataset for the testing stage to ensure that our supervised learning model is able to generalize. All trainable weights are initialized using the Glorot uniform initializer [19]. The model is trained using the Adam optimization algorithm with its default parameters. Training is stopped once 300 batches are processed without yielding an improvement larger than 10^{-5} on the validation loss. The same stopping criterion is also applied to the baseline models.

5. COMPARATIVE RESULTS

In this section, we compare our method with image enhancement methods from the literature on the test data. Our quantitative evaluation includes several image quality measurements, including classic Peak Signal-to-Noise ratio (PSNR), Structural Similarity (SSIM) [18], Mean Absolute Error (MAE), as well as less popular Gradient Magnitude Similarity deviation (GMSD) [21], Sobel intersection over union, and histogram difference. Tab. 1 shows that our approach outperforms histogram based image enhancement

FiveK Dataset [13] with hard clipping							
	PSNR	MAE	SSIM	Sobel IoU	Canny IoU	Hist. Diff	GMSD
Ours	2.359E+01	6.234E-02	9.142E-01	7.991E-01	6.229E-01	3.246E-03	2.025E-05
U-net [3]	2.186E+01	7.563E-02	8.559E-01	6.777E-01	5.150E-01	3.417E-03	3.303E-05
Can24 [15]	1.922E+01	1.235E-01	8.316E-01	6.923E-01	4.687E-01	5.007E-03	4.292E-05
DHE [5]	1.529E+01	1.581E-01	7.300E-01	5.548E-01	3.005E-01	4.890E-03	8.671E-05
Ying [9]	1.503E+01	1.887E-01	7.240E-01	6.087E-01	3.797E-01	5.580E-03	6.751E-05
Fu [20]	1.579E+01	1.673E-01	7.544E-01	6.278E-01	3.486E-01	5.113E-03	6.275E-05
None	1.853E+01	1.463E-01	7.776E-01	7.450E-01	5.938E-01	4.569E-03	4.674E-05
A6300 Dataset [11] with multiexposure bracketing							
Ours	1.780E+01	1.233E-01	8.628E-01	6.148E-01	3.903E-01	5.754E-03	3.707E-05
U-net [3]	1.640E+01	1.481E-01	8.332E-01	5.551E-01	3.536E-01	6.674E-03	4.376E-05
Can24 [15]	1.404E+01	2.045E-01	7.987E-01	5.287E-01	3.321E-01	7.493E-03	5.310E-05
DHE [5]	1.436E+01	1.836E-01	7.810E-01	5.314E-01	3.006E-01	6.941E-03	9.072E-05
Ying [9]	1.312E+01	2.595E-01	7.756E-01	5.915E-01	3.509E-01	7.846E-03	8.189E-05
Fu [20]	1.153E+01	2.946E-01	7.241E-01	5.387E-01	3.273E-01	8.534E-03	9.466E-05
None	9.666E+00	3.419E-01	5.722E-01	4.584E-01	2.213E-01	9.365E-03	1.164E-04

Table 1. Quantitative results show our method outperforms prior approaches in terms of color and gradient enhancement.

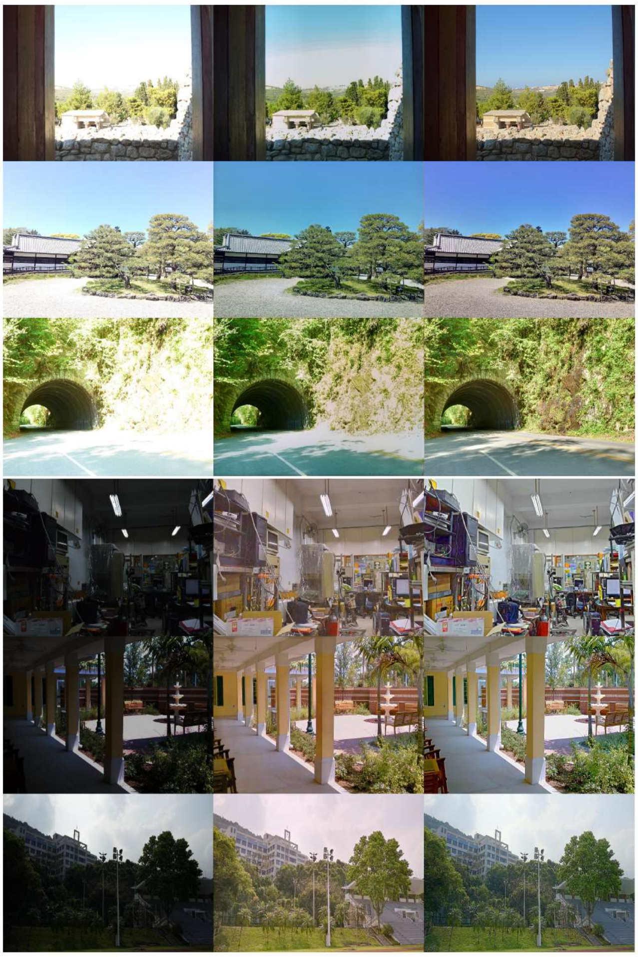

Fig. 2. Sample results for overexposed and underexposed images in the test set. Left to right: input, the restored image, and the reference.

methods [5], camera response based methods [9], probabilistic methods [20], and prior convolutional networks [3, 15] trained with the same conditions.

Deep Learning based restoration outperforms the deterministic algorithms on both datasets. Furthermore, we also notice a slight difference between the results for simulated data and real data. Results on the A6300 dataset show that all compared methods contributed to improving the PSNR and SSIM. For the FiveK simulated image set, we notice that [5, 9, 20] actually worsened the overall image condition, resulting in more noise and lower similarity with the reference image.

Qualitatively, Fig. 2 shows a few outputs of the network for severely over/underexposed images. We notice a significant improvement in element visibility, texture restoration, and re-colorization. In large regions where all three channels are clipped, we notice that our model is yet unable to restore the smoothness of the surface, resulting in images that preserve the block artifacts of the JPEG compression. Overall, a subjective evaluation highlights the robustness and effectiveness of the proposed method for challenging ill-exposure conditions. A more comprehensive qualitative comparison with baseline methods can be found in the supplementary material.

6. CONCLUSIONS AND FUTURE WORKS

We propose ReExpose-net, a new CNN-based model designed to maximize signal restoration and feature reconstruction of poorly exposed RGB images. Numeric and qualitative evaluation using two distinct datasets has shown our model significantly better in terms of brightness adjustment, contrast enhancement, image completion, and edge restoration. As future research, we intend to refine the smoothness of reconstructed regions (de-blocking), texture synthesis, and completion of large clipped using semantic features.

7. REFERENCES

[1] Samuel W Hasinoff, "Saturation (imaging)," in *Computer Vision*, pp. 699–701. Springer, 2014.

[2] Gordon Wetzstein, Ivo Ihrke, and Wolfgang Heidrich, "Sensor saturation in fourier multiplexed imaging," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 545–552.

[3] 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.

[4] Fisher Yu and Vladlen Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[5] 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.

[6] F. Albu, C. Vertan, C. Florea, and A. Drimbarean, "One scan shadow compensation and visual enhancement of color images," in *16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 3133–3136.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] Tom Mertens, Jan Kautz, and Frank Van Reeth, "Exposure fusion," in *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on*. IEEE, 2007, pp. 382–390.

[13] 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.

[14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1125–1134.

[15] 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.

[16] Augustus Odena, Vincent Dumoulin, and Chris Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016.

[17] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[18] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[19] Xavier Glorot and Yoshua Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[20] Xueyang Fu, Yinghao Liao, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[21] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.