Single image HDR synthesis using a Densely Connected Dilated ConvNet

Akhil K A and Jiji C V
College of Engineering Trivandrum
Trivandrum, India
akhilkashok@cet.ac.in, jijicv@cet.ac.in

Abstract

Visual representations using high dynamic range (HDR) images have become increasingly popular because of their high quality and expressive ability. HDR images are expected to be used in a broad range of applications, including digital cinema, photography, and broadcast. The generation of a HDR image from a single exposure Low Dynamic Range (LDR) image is a challenging task where one must make up for missing data due to underexposure or overexposure and color quantization. In this paper, we propose a deep convolutional neural network (CNN) model with a stack of dilated convolutional blocks for reconstructing a HDR image from a single LDR image. Within each dilation block, the dilation rate of the convolution layer is three and progressively decreases to one. Multiple dilation convolution blocks are further connected densely to improve the representation capacity of the network. As the network is trained in a supervised manner, the additional information is reconstructed from learned features. Our experimental results show that the model effectively captures missing information that was lost from the original image.

1. Introduction

Image restoration is one of the classical problems in computer vision. The method of reconstructing a clean, original image from a corrupted or noisy image is known as image restoration. The image is usually corrupted during the acquisition process or transmission because of noise, low-light, pixel value errors, out-of-focus blurring, or camera motion blurring, etc. Unlike image enhancement, image restoration attempts to retrieve a latent clean image x from a degraded image y by properly understanding the degradation phenomenon given by

$$y = H\,x + d \tag{1}$$

where H is the degradation function and d is additive noise.

Recent advancements in deep learning have brought significant growth in the field of image restoration. As a function approximator, it estimates the unknown mapping between the input and output image sets. Deep learning has advanced to state-of-the-art efficiency in image restoration tasks like deblurring[25], denoising[26], super-resolution[2, 9], etc. Similar attempts have been made to restore an original image having much richer contrasts like those found in the physical world, referred to as the high dynamic range (HDR) imaging.

Camera sensors cannot capture the wide range of luminance in a natural scene. This will result in pixel information being lost in underexposed and overexposed areas of an image, resulting in a low dynamic range (LDR) image. Furthermore, since a narrow-range image is enlarged to a wide-range image, finding an appropriate relationship between the two spaces of different ranges is challenging. HDR images must be generated to recover missing information and reflect a wide range of illuminance in an image. Few advanced hardware systems have been proposed to directly generate HDR images[20, 16], but they are usually too costly to be broadly accepted. As a result, computational HDR imaging techniques have gotten a lot of attention.

There has been active research going on in the area of deep learning for HDR image reconstruction. The introduction of convolutional neural networks (CNN) brought significant growth in the field of HDR image reconstruction. Taking a sequence of LDR images at various exposures and then combining them into an HDR image is the most common technique [1, 12, 14, 18, 23]. In these methods, the scene and the camera should be kept static to produce a high-quality HDR image. When there is motion between these multiple exposure images, there will be ghosting and blurring artifacts. Recovering a HDR image from a single exposure LDR image is a more challenging problem. The methods in [3, 13, 5, 24, 11] above convert a single exposure LDR image to an HDR image. HDRCNN[4] suggested a deep auto-encoder for HDR image restoration that recovers only the overexposed regions of an LDR image using a weighted mask. In DRTMO[6], a framework consisting of two networks for creating up-exposure and down-exposure

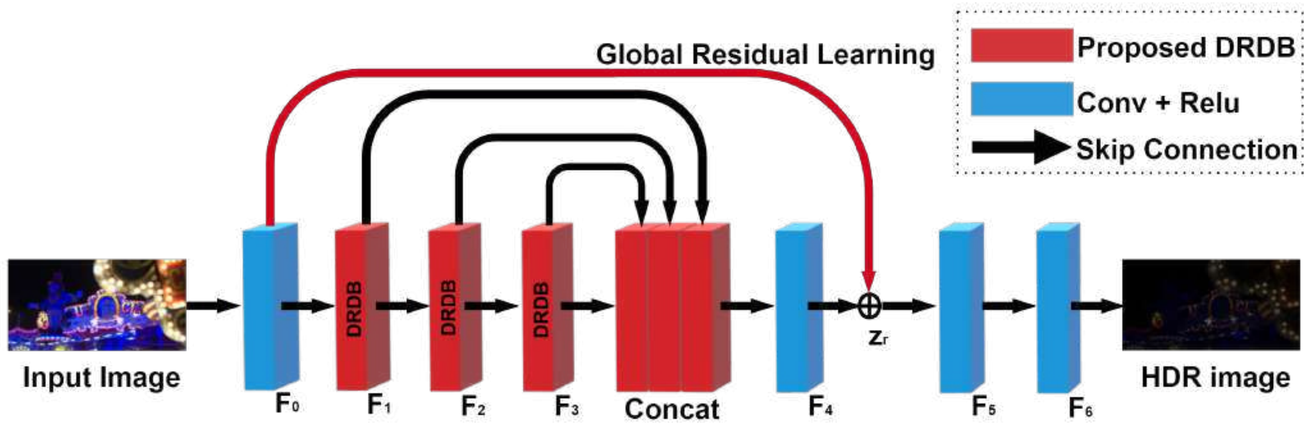


Figure 1: Proposed DCDCNet architecture. The network uses DRDB for efficient feature extraction.

LDR images, are combined to form an HDR image.

In this work to synthesis HDR images, we propose a fully connected convolutional neural network namely Densely Connected Dilated ConvNet (DCDCNet) inspired by [22]. Here, the feature extraction in the network has been improved by integrating the dilated residual dense blocks (DRDBs).

The rest of the paper is organized as follows: In Section 2, we describe the methodologies that we propose for HDR image synthesis. Section 3 details our experiments and in Section 4, we present our result analysis. Finally, Section 5 concludes the work.

2. Proposed Method

In this section, we describe the proposed network used for the synthesis of HDR images from a single exposure LDR image. The input LDR image and generated HDR image are RGB images with three channels. Before feeding the LDR image, to generate a corresponding set of HDR images, we first map the input LDR images I to the HDR domain using gamma correction[21, 8].

$$X = I^{\gamma}/t \tag{2}$$

where $\gamma > 1$ denotes the gamma correction parameter and t denotes the exposure time of the image. In this work we use $\gamma = 2.24$. Given X as input, the proposed densely connected dilated ConvNet (DCDCNet) obtain the HDR image by

$$\widehat{X} = f\left(X, \theta\right) \tag{3}$$

where $f\left(.\right)$ denotes the proposed DCDCNet, and θ is the network parameters.

2.1. Overview of the DCDCNet Architecture

The proposed densely connected dilated ConvNet (DCDCNet) as shown in Figure 1, takes the gamma-corrected image as input and generates the corresponding HDR image using a sequence of dilated residual dense blocks (DRDBs) using a global residual learning(GRL) technique. The DRDBs and GRL help in the effective use of image features and the generation of an HDR image with plausible details. The network consists of several convolution layers, dilated residual dense blocks, and skip connections, as shown in Figure 1. $F_j, j = 0, 1, \cdots, 6.$ are the feature maps at different layers.

Given the gamma-corrected image, the DCDCNet first obtains a feature map of 64 channels by a convolution layer. Then this is fed to series of three DRDBs. As a result, three function maps F_1, F_2, and F_3 are generated. In DCDCNet, we use residual dense blocks (RDBs) with dilated convolution for HDR imaging instead of the RDB proposed in [27]. The feature map from three DRDB (ie F_1, F_2, F_3) are concatenated and then fed to a convolution layer of filter size 3 x 3 to generate another feature map F_4.

Inspired by super-resolution methods [10, 28], we use a global residual learning strategy to obtain feature maps before reconstructing the HDR image from F_4.

$$Z_r = F_4 + F_0 \tag{4}$$

The less processed information from the reference image is found in the feature map F_0. Thus the feature map Z_r

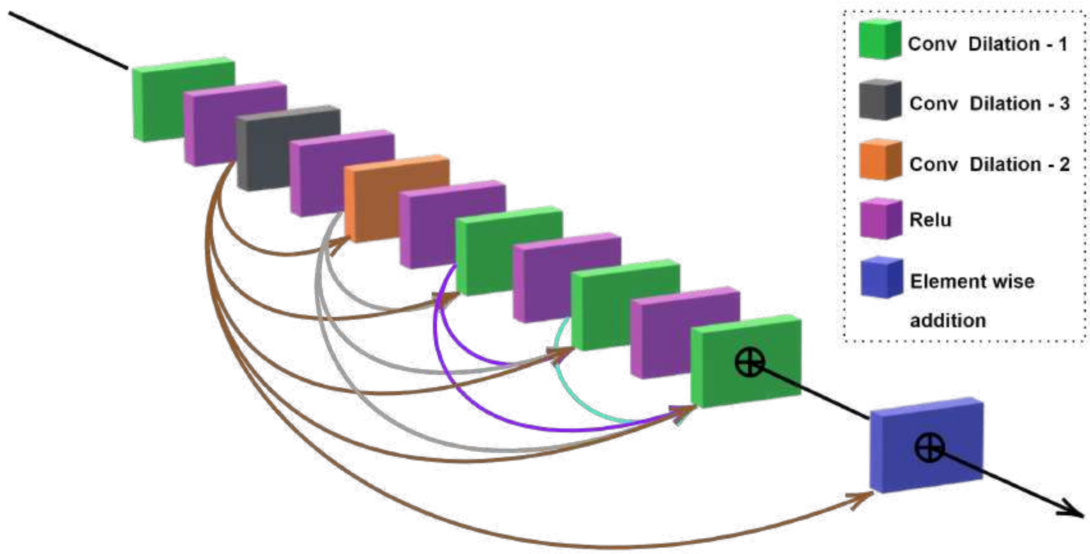


Figure 2: The DRDB sub-net used in DCDCNet. The block has a series of densely connected convolutional layers with varying dilation rates for feature extraction.

contains enough information to reconstruct the HDR image. We estimate the HDR image $\widehat{X}$ in the HDR domain after two convolution layers with ReLU activation.

2.2. Dilated Residual Dense Block

A convolutional filter with a broad receptive field is needed, to capture degradations with a wide spatial extension. However, this would theoretically increase network complexity and would necessitate a large dataset to prevent overfitting. Another alternative is to use a series of convolutional layers with small filter sizes, resulting in a larger receptive field with less learnable parameters. But this would increase the number of trainable parameters, which could lead to deep network vanishing gradient problems. These limitations can be solved using dilated convolutions with large dilation factors. However, since the aggregated features in two adjacent pixels come from a completely non-overlapping input feature set[15], high dilation rates can result in information loss. Additionally, the restored images can have a gridding effect as a result of these. To tackle this problem, Dilated residual dense block (DRDB)[19] is used for feature extraction. As shown in Figure 2, DRDB is a residual densely [7] connected convolutional network with varying dilatation rates in different convolution layers. A sequence of convolution layers is accompanied by ReLU activations and dense concatenation-based skip-connections in the proposed dilated residual dense block (DRDB). As input, each convolution layer takes the concatenation of all the previous layer's feature maps. This ensures the smooth flow of gradients to shallow layers of the network while minimizing the feature redundancy in deeper layers. Each DRDB implementation has six convolution layers. As a result, the network achieves a large receptive field without information loss.

2.3. Loss Function

In this method, the network is trained based on the mean squared error (MSE) and the MSE loss is the mean squared error between the ground-truth HDR images and the predicted images. It is incorporated to generate high fidelity images and is formulated as :

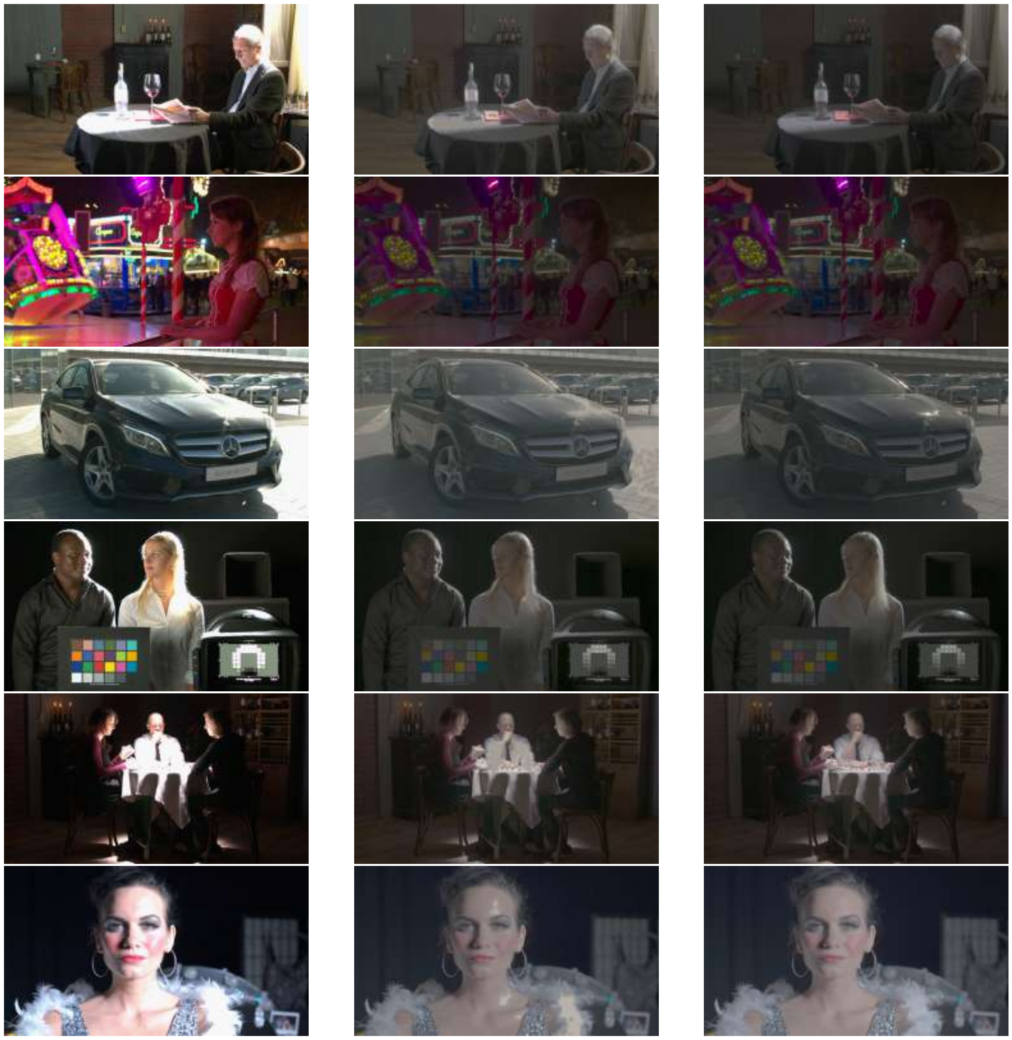
$$L_{MSE} = \frac{1}{W \times H \times C} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} \sum_{k=0}^{C-1} \left(\mathbf{X_{i,j,k}} - \widehat{\mathbf{X}}_{\mathbf{i,j,k}} \right)^2 \tag{5}$$

where, W, H, and C are the width, height, and the number of channels of the output, Y is the ground truth HDR image and $\widehat{X}$ is the predicted image.

3. Experiment

3.1. Dataset

The dataset used in the experiment is the challenge dataset provided by the organizers of NTIRE workshop 2021 for the HDR image synthesis for LDR images. The dataset consists of a low dynamic range (LDR) medium exposed image as input and a corresponding high dynamic range (HDR) ground-truth image. The dataset contains 1494 pairs of LDR medium exposed images and their corresponding HDR ground truth images for training, 60 pairs



(a) (b) (c)

Figure 3: Sample results of the proposed DCDCNet for (a) input LDR medium exposed image, (b) images generated using DCDCNet shows high fidelity with (c) the HDR ground-truth image.

for validation, and 201 pairs for testing. Each image is of resolution 1060 × 1900 × 3.

3.2. Training

The DCDCNet is trained using image patch size 512 × 512 × 3 and batch size 10. Data augmentation is done using random cropping, flip, and rotation. Adam optimizer with β_1 = 0.9, β_2 = 0.99 and learning rate $2e^{-1}$ was used in networks. The learning rate is decayed after every 5000 epochs with a decay rate of 0.5 in network. The network is trained for a total of 11000 epochs using a Tesla P100 GPU card with 16 Gib memory.

3.3. Evaluation Metrics

For the quantitative evaluation of the proposed method, we use the standard Peak Signal To Noise Ratio (PSNR). The PSNR score is calculated in dB between the µ-law tonemapped ground truth and generated images.

4. Result Analysis

Densely Connected Dilated ConvNet (DCDCNet) is a competing entry in Track 1 Single frame NTIRE 2021 High Dynamic Range challenge [17]. Table 1 shows the performance comparison based on muPSNR and PSNR of the proposed DCDCNet with other competing entries in the challenge. It is evident from the table that, the proposed DCDCNet obtained comparable results with the challenge winners

Teams	muPSNR	PSNR
Ours	**33.93(4)**	**32.00(4)**
Team 1	34.87(1)	32.84(2)
Team 2	34.77(2)	32.27(3)
Team 3	34.49(3)	33.47(1)
Team 4	32.83(5)	30.98(5)
Team 5	16.42(6)	26.94(6)

Table 1: Comparative study of the results from NTIRE 2021 High Dynamic Range Challenge - Track 1 Single Frame

in terms of muPSNR.

5. Conclusion

In this paper, we proposed a fully connected convolutional neural network to reconstruct a high dynamic range (HDR) image from a single exposure low dynamic range (LDR) image. Densely Connected Dilated ConvNet (DCDCNet) for Single image HDR synthesis has obtained state-of-the-art restoration performance in terms of standard evaluation metrics muPSNR and PSNR. The dilated residual dense blocks in the network enable feature reuse, thus learning robust representations with minimum parameters. Local and global feedback connections boost learning ability by guiding low-level features from higher-level features. Further improvement and additional studies about the network structure will be addressed in future work.

References

[1] Debevec, E Paul, Malik, and Jitendra. Recovering high dynamic range radiance maps from photographs. *In Conference on Computer Graphics Interactive Techniques*, pages 369–378, 1997.

[2] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. *In European Conference on Computer Vision*, 1:184–199, 2014.

[3] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics (TOG)*, 36(6):178, 2017.

[4] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics (TOG)*, 36(6):178, 2017.

[5] Y. Endo, Y. Kanamori, and J. Mitani. Deep reverse tone mapping. *In ACM Trans*, 36(6):177, 2017.

[6] Y. Endo, Y. Kanamori, and J. Mitani. Deep reverse tone mapping. *ACM Transactions on Graphics (TOG)*, 36(6):177, 2017.

[7] Huang G., Liu Z., Van Der Maaten L., and Weinberger K.Q. Densely connected convolutional networks. *In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[8] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):1–12, 2017.

[9] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1:1646–1654, 2016.

[10] Christian Ledig, Lucas Theis, Ferenc Huszarand Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, and Zehan Wang. Photorealistic single image super-resolution using a generative adversarial network. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1451–1460, 2017.

[11] S. Lee, G. H. An, and S.-J. Kang. Deep chain hdri: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access*, 6:49913–49924, 2018.

[12] Mann, Picard, S. Mann, and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. *In Proceedings of IST*, pages 442–448, 1995.

[13] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *In Computer Graphics Forum*, 37:37–49, 2018.

[14] Granados Miguel, Ajdin Boris, Wand Michael, Theobalt Christian, Seidel Hans-Peter, and P. A. Lensch Hendrik. Optimal hdr reconstruction with linear digital cameras. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 215–222, 2010.

[15] Liu P., Zhang H., Zhang K., Lin L., and Zuo W. Multi-level wavelet-cnn for image restoration. *In IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 886–88609, 2018.

[16] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[17] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Aleš Leonardis, Radu Timofte, et al. NTIRE 2021 challenge on high dynamic range imaging: Dataset, methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.

[18] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul E Debevec. High dynamic range imaging, acquisition, display, and image-based lighting. *Princeton University Press*, 2005.

[19] Hrishikesh P. S., Densen Puthussery, Melvin Kuriakose, and Jiji C. V au2. Transform domain pyramidal dilated convolution networks for restoration of under display camera images, 2020.

[20] Jack Tumblin, Amit Agrawal, and Ramesh Raskar. Why i want a gradient camera. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 103–110, 2005.

[21] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. *In European Conference on Computer Vision (ECCV)*, 2018.

[22] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian D. Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. *CoRR*, abs/1904.10293, 2019.

[23] Qingsen Yan, Jinqiu Sun, Haisen Li, Yu Zhu, and Yanning Zhang. High dynamic range imaging by sparse representation. *Neurocomputing*, pages 160–169.

[24] X. Yang, K. Xu, Q. Zhang Y. Song, X. Wei, and R. W. Lau. Image correction via deep reciprocating hdr transformation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1798–1807, 2018.

[25] Zhang, H. Dai, Y. Li, and H. Koni. Deep stacked hierarchical multi-patch network for image deblurring. *In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5971–5979, 2019.

[26] K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep cnn denoiser prior for image restoration, 2017. `arXiv preprint arXiv:1704.03264`.

[27] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2472–2481, 2018.

[28] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 36(4):2472–2481, 2018.