LLRNET: A MULTISCALE SUBBAND LEARNING APPROACH FOR LOW LIGHT IMAGE RESTORATION

Sameer Malik and Rajiv Soundararajan

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore-560012
Email:(sameer,rajivs)@iisc.ac.in

ABSTRACT

We consider the problem of low light image restoration through joint contrast enhancement and denoising. Deep convolutional neural networks (CNNs) based on residual learning have been successful in achieving state of the art performance in image denoising. However, their application to joint contrast enhancement and denoising poses challenges owing to the nature of the distortion process involving both loss of details and noise. Thus, we propose a multiscale learning approach by learning the subbands obtained in a Laplacian pyramid decomposition through a subband CNN (SCNN). The enhanced subbands at multiple scales are then combined to obtain the final restored image using a recomposition CNN (ReCNN). We refer to the overall network involving SCNN and ReCNN as low light restoration network (LLRNet). We show through extensive experiments based on the 'See in the Dark' Dataset that our approach produces better quality restored images when compared to other contrast enhancement techniques and CNN based approaches.

Index Terms— Contrast enhancement, low light enhancement, denoising, CNN. Laplacian pyramid.

1. INTRODUCTION

The performance of a camera in low light scenarios is becoming increasingly important in consumer electronic devices such as smartphones. Variations in camera hardware settings such as aperture size, sensor sensitivity (ISO) and shutter speed may result in distortions such as noise, motion blur and shallow depth of field. Moreover, slow shutter speeds may not be feasible during capture of highly dynamic scenes in low light situations. While using a flash might be helpful, it causes unwanted shadows and color distortions leading to unnatural images. Thus there is a need to study the problem of enhancing the quality of low light images through image processing techniques. We focus on the problem of joint contrast enhancement and denoising in grey scale low light images.

This work was supported by a grant from the Robert Bosch Center for Cyberphysical Systems, Indian Institute of Science, Bangalore, India.

Several contrast enhancement algorithms for low light enhancement exist in literature. These can be classified as histogram modification approaches [1, 2, 3, 4] or retinex based approaches [5, 6, 7]. Other approaches achieve contrast enhancement by enhancing the bandspass subbands [8, 9]. The hazy appearance of inverted low light images has led to the use of dehazing methods [10] for contrast enhancement. The above approaches do not account for the noise in the low light image.

The noise in low light images is commonly addressed by performing post enhancement denoising using algorithms such as BM3D [11]. However, the non-linearity of the contrast enhancement algorithms, might distort the Gaussian noise assumption in denoising. Alternately, the low light image can be first denoised and then enhanced [12]. However, the low light conditions affect the assumption of natural scene statistics of the clean image in denoising methods such as [11]. Thus the problem of joint enhancement and denoising appears to be challenging. Given the recent success of convolutional neural networks (CNNs) in image restoration [13], we explore the use of CNNs for this complex task.

Recently, DnCNN has been successfully applied to image denoising by exploiting the benefits of residual learning [13]. However, this approach poses challenges in the low light scenario due to the lack of knowledge of the residual noise. In particular, the low light image can be modeled to account for distortions due to the loss of details and additive noise. Thus, given a low light image and well lit image pair, estimating these distortion parameters is an ill posed problem. Since residual learning appears difficult here, we seek to investigate other approaches that do not directly learn the pixel domain image. Note that the formulation we consider is different from the problem of mapping raw sensor data to sRGB images considered in [14] using architectures such as UNet [15] and CAN [16].

Our main contribution is in the design of a CNN architecture, LLRNet (or low light restoration network), to perform joint contrast enhancement and denoising of low light images. We adopt a multiscale learning approach where we first decompose an image using a Laplacian pyramid and learn to jointly enhance and denoise the subbands. We believe this approach to be effective due to the well behaved statistical properties of the band pass coffieicients. The restored subbands are then recomposed using a Recomposition CNN (ReCNN) to produce a high quality restored image. We show through experiments on the See in the Dark Dataset [14] that our architecture improves upon other existing enhancement algorithms and CNN architectures for restoration.

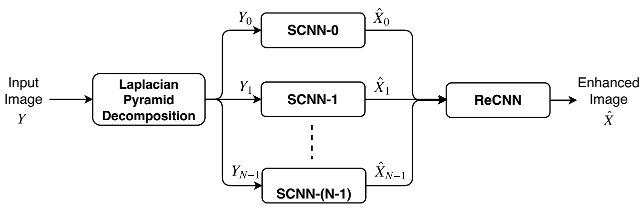

Fig. 1. Architecture of LLRNet

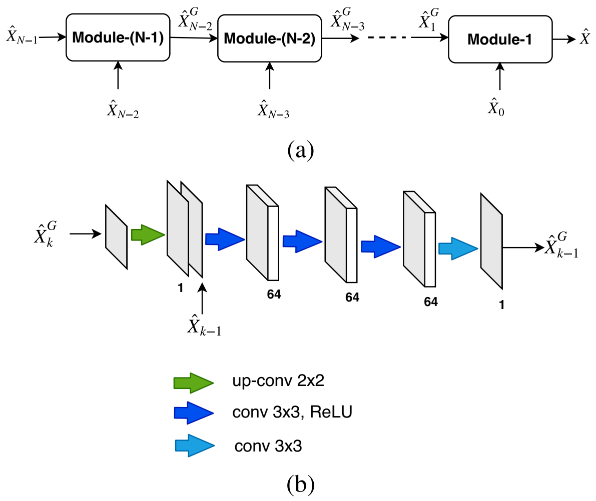

Fig. 2. (a) ReCNN; (b) Architecture of the k^{th} module of ReCNN

2. PROPOSED METHOD

Motivated by the success of DnCNN in image denoising and super resolution, we wish to apply DnCNN to the low light image restoration problem. A low light image can be modeled as

$$Y(i,j) = t(i,j)X(i,j) + Z(i,j), \tag{1}$$

where $X(i,j)$ is the input image at location (i,j), $Y(i,j)$ is the observed low light image, $t(i,j)$ is a detail loss coefficient, and $Z(i,j)$ represents the noise. The challenge in applying the residual noise learning approach of DnCNN for the above model is that both the noise term $Z(i,j)$ and the detail loss coefficient are unknown. Therefore, we adopt a multiscale approach for learning subbands of a Laplacian pyramid decomposition based on DnCNN.

2.1. Architecture

We apply a Laplacian pyramid to the low light image to obtain multiple subbands at different scales. Bandpass subband coefficients, when compared to the pixel domain image, have well behaved statistics [17, 18]. We believe that this regularity in statistics makes it easier for the CNN to learn a mapping from the distorted subband to the ground truth subband. Therefore, we train CNNs to restore the bandpass subbands. The low pass subband at the coarsest scale is potentially easier to learn than the original image in the pixel domain. Thus we apply the DnCNN to directly restore the low pass subband.

We use N subbands in the Laplacian pyramid decomposition and train N CNNs, referred to as SCNN (or subband CNN), one for each subband. In Figure 1, we show the proposed architecture, where SCNN-0 to SCNN-(N-2) operate on the bandpass subbands of the finest to the coarsest scales respectively. SCNN-(N-1) operates on the low pass subband. For each of the SCNNs, we use the same architecture as [13] without residual learning. In Figure 1, $\hat{X}_k$ denotes the output of SCNN-k, where $k \in \{0, 1, \ldots, N-1\}$. To recompose the enhanced subbands and obtain the final restored image, we train another CNN, which we refer to as ReCNN. We use ReCNN instead of the default Laplacian recomposition to learn how to combine the enhanced subbands to yield the restored low light image.

In Figure 2(a), we show the architecture of ReCNN, the design of which is based on the Laplacian recomposition algorithm. ReCNN is composed of $N-1$ modules, where each module takes as input, the image at a lower scale, and produces an image at the higher scale by using the lower scale image upsampled by a factor of 2 and the band pass details. We denote $\hat{X}_k^G$ as the image output by Module-$(k+1)$ based on the input $\hat{X}_k$ and $\hat{X}_{k+1}^G$. $\hat{X}_k^G$ has double the resolution of $\hat{X}_k$. Note that Module-$(N-1)$ alone takes as input $\hat{X}_{N-1}$ and $\hat{X}_{N-2}$. Further, $\hat{X}_0^G$ is the same as $\hat{X}$, the restored image. In Figure 2(b), we show the architecture of a module of the ReCNN.

2.2. Training Methodology

We use the recently released See In the Dark dataset (SID) [14] for training. This dataset has low light and well-lit image pairs of 424 scenes from two cameras, SONY and FUJI. The well-lit image for each scene has been collected using

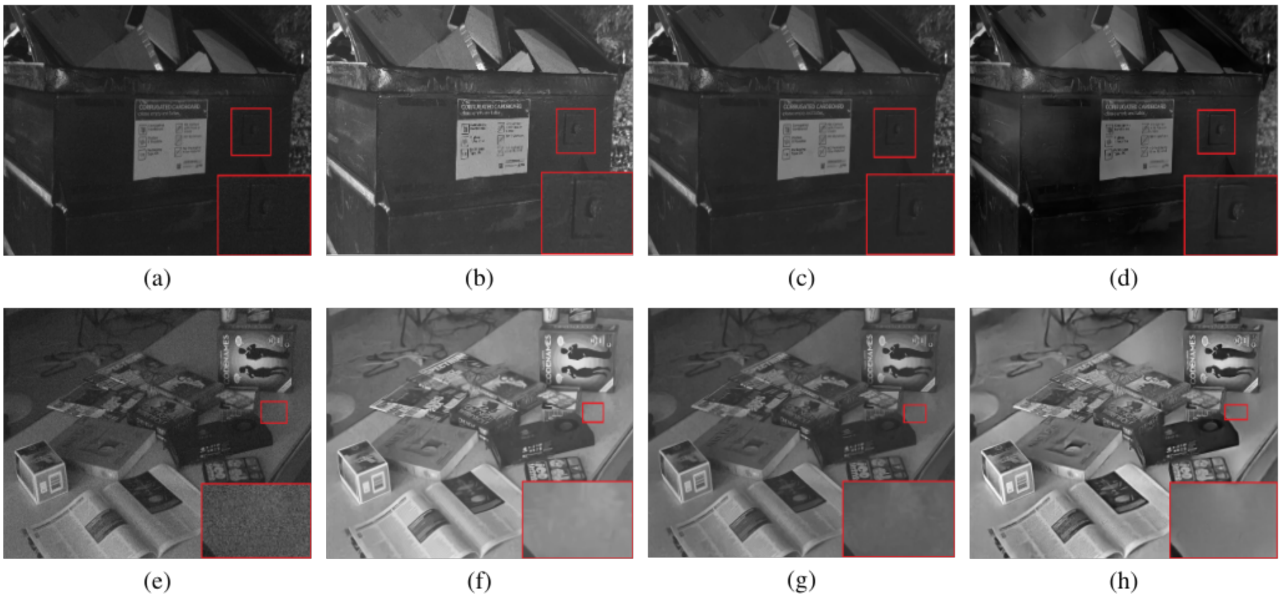

Fig. 3. (a) A low light image from the SID dataset enhanced using; (b) LIME [7]; (c) FUSE [19]; (d) LLRNet (proposed); (e) A low light image from SID datset enhanced using; (f) LIME; (g) FUSE; (h) LLRNet

Fig. 4. A low light image from the SID dataset enhanced using; (a) CAN [16]; (b) DnCNN [13]; (c) U-Net [15]; (d) LLRNet

low ISO level and high exposure time of atleast 10 seconds. These well lit images serve as ground truth images. There are 2697 and 2396 low light images for SONY and FUJI camera respectively at different camera settings of exposure and ISO levels. The images are raw in bayer format. We use a python based library *libraw* to convert the raw images to sRGB format. The sRGB images under both low light and well lit conditions are then converted to gray scale. We train two different models, one for camera. We use 1865 and 1654 of the image pairs from SONY and FUJI respectively, for training as mentioned in the released dataset.

For training we subsampled the images to 832×1248. We train using patches of size 160×160, cropped randomly at each iteration from the training set of images. Further, for data augmentation, the patches are subjected to random flipping and rotation at each iteration. The patches from both the low light as well as the ground truth images are decomposed into 5 subbands including the low pass subband. We then use these subband patches to train each of the SCNNs. We train the SCNNs by minimizing the L2 loss using Adam optimizer for 1200 epochs with a learning rate of 10^{-3}, decayed by a factor of 10^{-1} after every 400 epochs.

To train the ReCNN, we use the subband patches enhanced using the SCNNs and the corresponding ground truth pixel domain image patch. We find that end to end training of ReCNN from scratch gives poor results. We therefore first pre-train each of the modules separately as follows. We pre-train Module-k of ReCNN by minimizing the L_2 loss between its output $\hat{X}_k^G$ and X_k^G, where X_k^G corresponds to the ground truth patch downsampled to the resolution of $\hat{X}_k^G$. We pre-train each of the modules for 30 epochs following which we perform end to end training of ReCNN for 30 epochs.

3. RESULTS

We now evaluate the performance of LLRNet and provide comparisons with popular contrast enhancement algorithms on images from both the cameras in the SID dataset [14]. The test set of this dataset consists of 597 and 523 low light images from SONY and FUJI camera respectively. We also evaluate the performance of some of the popular CNN architectures that can be deployed for low light image restoration. We carry out all the tests on images subsampled to 832×1248.

3.1. Performance Analysis

We first compare the performance of LLRNet with popular contrast enhancement approaches such as a fusion based method (FUSE) [19] and a retinex based method (LIME) [7]. Since these approaches do not have any noise reduction mechanisms, we denoise the images enhanced by these methods using BM3D [11]. In Figure 3, we show two low light images, enhanced using FUSE, LIME and LLRNet. As can be seen, the image enhanced by LLRNet has significantly better contrast and lower noise. In Table 1, we report the SSIM [20] and peak signal to noise ratio (PSNR) scores of all these methods on both sets of camera images. LLRNet is trained separately for each set of camera images. As can be observed, the scores for LLRNet are better on SONY images while lags behind FUSE on FUJI images. Low light images from SONY suffer from faint scan line distortion due to high ISO level camera setting. Presence of such distortions in addition to the Gaussian noise is often the case in very low light images. LLRNet, when compared to FUSE and LIME, effectively mitigates them and thus has significantly better scores on SONY images.

(a) (b)

Fig. 5. A low light image from the SID dataset enhanced using LLRNet (a) without ReCNN; (b) with ReCNN;

We now discuss the performance of other popular CNN architectures for low light image restoration. Specifically, we analyze the performance of U-Net [15], CAN [16] and DnCNN [13]. In Figure 4, we show a low light image enhanced using U-Net, CAN, DnCNN and LLRNet. As can be seen, U-Net, CAN and DnCNN give poor quality enhanced images when compared to the LLRNet. This also implies that simply using DnCNN to learn to directly predict the high quality image does not work well. In Table 2, we report the SSIM and PSNR scores, which lead to similar conclusions.

In order to understand the benefit of ReCNN, we replace it with the default Laplacian reconstruction algorithm and compare the corresponding output images in Figure 5. As can be seen, the ReCNN helps reduce the artifacts that are observed while using the Laplacian recomposition thus resulting in smoother image. We also see quantitative benefit in terms of objective quality scores in Table 3.

METHOD	SONY		FUJI	
	SSIM	PSNR	SSIM	PSNR
LIME	0.6635	17.53	0.6166	15.12
FUSE	0.7341	18.44	**0.7119**	**17.40**
LLRNet	**0.7632**	**19.87**	0.7042	16.12

Table 1. Performance comparison with other contrast enhancement methods

METHOD	SONY		FUJI	
	SSIM	PSNR	SSIM	PSNR
CAN	0.6435	18.32	0.4409	14.00
U-Net	0.5963	18.41	0.6226	14.49
DnCNN	0.6126	18.21	0.6440	**16.71**
LLRNet	**0.7632**	**19.87**	**0.7042**	16.12

Table 2. Comparison with other CNN approaches

4. CONCLUSION

We presented a new CNN architecture, LLRNet, to jointly enhance and denoise low light images. The architecture exploited the learning of subband coefficients instead of learning the original image. Further, we proposed an effective method of training ReCNN, which combines the enhanced subbands to output the restored image. We showed through experiments that this approach performs better than other contrast enhancement methods followed by denoising as well as other CNN based approaches. It will be interesting to explore the benefit of this subband learning approach in other image restoration tasks.

Method	SSIM	PSNR
w/o ReCNN	0.6978	15.83
With ReCNN	**0.7042**	**16.12**

Table 3. Comparisons with and without ReCNN on the FUJI images from the SID dataset

5. REFERENCES

[1] H. D. Cheng and X. J. Shi, “A simple and effective histogram equalization approach to image enhancement,” *Digital signal processing*, vol. 14, no. 2, pp. 158–170, 2004.

[2] Karel Zuiderveld, “Graphics gems iv,” chapter Contrast Limited Adaptive Histogram Equalization, pp. 474–485. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[3] T. Celik and T. Tjahjadi, “Contextual and variational contrast enhancement,” *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3431–3441, Dec 2011.

[4] C. Lee, C. Lee, and C. S. Kim, “Contrast enhancement based on layered difference representation of 2d histograms,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, Dec 2013.

[5] D. J. Jobson, Z. Rahman, and G. A. Woodell, “A multiscale retinex for bridging the gap between color images and the human observation of scenes,” *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, Jul 1997.

[6] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel, “A variational framework for retinex,” *International Journal of Computer Vision*, vol. 52, no. 1, pp. 7–23, Apr 2003.

[7] X. Guo, Y. Li, and H. Ling, “LIME: Low-Light image enhancement via illumination map estimation,” *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, Feb 2017.

[8] N. Bonnier and E. P. Simoncelli, “Locally adaptive multiscale contrast optimization,” in *IEEE International Conference on Image Processing 2005*, Sept 2005, vol. 1, pp. I–949–52.

[9] G. Deng, “A generalized unsharp masking algorithm,” *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1249–1261, May 2011.

[10] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, Dec 2011.

[11] 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.

[12] L. Li, R. Wang, W. Wang, and W. Gao, “A low-light image enhancement method for both denoising and contrast enlarging,” in *2015 IEEE International Conference on Image Processing (ICIP)*, Sept 2015, pp. 3730–3734.

[13] 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, July 2017.

[14] C. Chen, Q. Chen, J. Xu, and V. Koltun, “Learning to See in the Dark,” *ArXiv e-prints*, May 2018.

[15] 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.

[16] 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.

[17] M. J. Wainwright and E. P. Simoncelli, “Scale mixtures of Gaussians and the statistics of natural images,” in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA, 1999, NIPS’99, pp. 855–861, MIT Press.

[18] E P Simoncelli, “Statistical models for images: Compression, restoration and synthesis,” in *Proc 31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA, Nov 2-5 1997, vol. 1, pp. 673–678, IEEE Computer Society.

[19] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley, “A fusion-based enhancing method for weakly illuminated images,” *Signal Processing*, vol. 129, pp. 82 – 96, 2016.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.