Can fully convolutional networks perform well for general image restoration problems?

Subhajit Chaudhury
Takatsu, Kawasaki City
Kanagawa Prefecture, Japan
subhajit.ju4u@gmail.com

Hiya Roy
The University of Tokyo
JAXA, Sagamihara, Japan
hiya.roy@ac.jaxa.jp

Abstract

We present a fully convolutional network(FCN) based approach for color image restoration. FCNs have recently shown remarkable performance for high-level vision problem like semantic segmentation. In this paper, we investigate if FCN models can show promising performance for low-level problems like image restoration as well. We propose a fully convolutional model, that learns a direct end-to-end mapping between the corrupted images as input and the desired clean images as output. Our proposed method takes inspiration from domain transformation techniques but presents a data-driven task specific approach where filters for novel basis projection, task dependent coefficient alterations, and image reconstruction are represented as convolutional networks. Experimental results show that our FCN model outperforms traditional sparse coding based methods and demonstrates competitive performance compared to the state-of-the-art methods for image denoising. We further show that our proposed model can solve the difficult problem of blind image inpainting and can produce reconstructed images of impressive visual quality.

1 Introduction

Image restoration is the technique to convert a noisy image into a clean, original one. Common image restoration problems include image denoising and image inpainting. Image denoising is the method of removing the external noise (usually modeled as additive white Gaussian noise) to obtain the original uncorrupted image. Another form of corruption for image signal occurs in the form of missing pixel values. Image inpainting is used for predicting such missing pixel values or removing sophisticated patterns like superimposed texts from images and preserve the original image information. In this paper, we focus on the problems of image denoising and blind image inpainting.

Prominent techniques in image denoising perform modifications in the image domain itself. Notable methods in this category include total variation based image denoising [1], denoising by learning global image priors[2] etc. Additionally, sparse coding-based image denoising is shown to produce an impressive performance which can also be extended to solve other image restoration tasks. Carefully engineered algorithms such as BM3D[3] and its color variant CBM3D[4], which exploit similarity in appearance of different patches constitute the current state-of-the-art in image denoising.

Image inpainting can be broadly classified into two categories: non-blind inpainting and blind inpainting. While in non-blind inpainting, the algorithm is provided the prior knowledge of the spatial locations of the image with missing pixels or superimposed patterns that need to be restored, blind inpainting methods aim to solve a much more challenging problem of simultaneously identifying and restoring the corrupted pixels. In the field of non-blind image inpainting, region filling method[5], the sparse coding-based K-SVD[6] model etc. have been proposed, however blind inpainting is a less mature field of study with limited implementations. To the best of our knowledge, SSDA based blind inpainting[7] is the most notable work in blind image inpainting.

Our proposed method is inspired from classical domain transformation methods, where the image domain signal is converted to a new representation[8] and coefficients are altered in the transformed domain to finally reconstruct the clean image. The proposed method is application specific and fully data-driven with no requirements of human designed filters which is the reason for superior restoration performance. We use a similar idea to that of Dong et al. [9] for deep convolutional networks based image super-resolution and extend it to show that similar architectures can be used for image denoising and blind image inpainting, which is one of our major contributions in this paper. Moreover, our proposed solution is very simple to implement and consists of only convolutional layer (no pooling), which enables easy hardware implementation with fast image restoration performance.

Autoencoders based image restoration techniques(like SSDA[7]), compress the input image patch to a low-dimensional representation before decoding it to produce the final image reconstruction, which might lead to loss of information causing poor image restoration performance. In contrast, we maintain equal hidden unit dimension to the input image size throughout the network and perform the intermediate operations by filtering using convolution kernels. Since our proposed fully convolutional network does not compress input data, we believe that it is possible to perform better image restoration using the proposed model. Results in image denoising demonstrate that the proposed method is competitive with the state of the art methods. For image inpainting, although our model performs a much more difficult task of blind restoration, it demonstrates comparable visual reconstruction quality at par with non-blind inpainting methods. The capability of our model for blind inpainting of complex superimposed patterns is also a major contribution of this paper.

2 Proposed Method

We map noisy images at the input to their corresponding clean image version by image domain transformation method. This mapping conceptually consists of three operations- (1) Basis projection i.e. projecting image patches onto learn dictionaries which is a novel representation of noisy images, (2) non-linear transformation for mapping the coefficients onto a new domain for representations of clean images and (3) reconstruction of clean image using weighted averaging on overlapping patches. Although the proposed concept is similar to image denoising using domain transformation, our method benefits from the unique feature of the ability to learn from data in an end-to-end fashion. Similar to the image super-resolution model presented in Dong et al. [9], we find that these three operations are similar to multidimensional filtering operations and can be performed by convolution operations. Hence, the mapping described above can be represented as a fully convolutional network.

2.1 Model description

IRCNN **(5-5-1-5-5-5)**: For solving the tasks of image denoising and blind image inpainting, we propose a 6 layered Image restoration CNN model (IRCNN) consisting of only convolution layers. Figure 1 shows details on filter weights and number of convolution parameters for each layer. First two convolution layers of filter size 5×5 perform basic projection, next convolution layer (1×1 filter size) performs pixel-wise co-efficient alterations, and finally, last three convolution layers are responsible for converting the clean image representations to clean image.

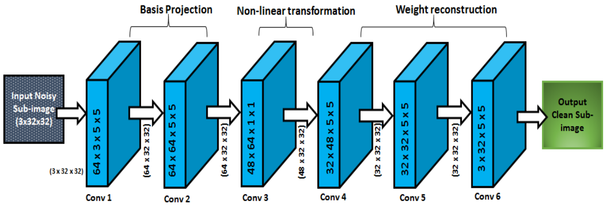


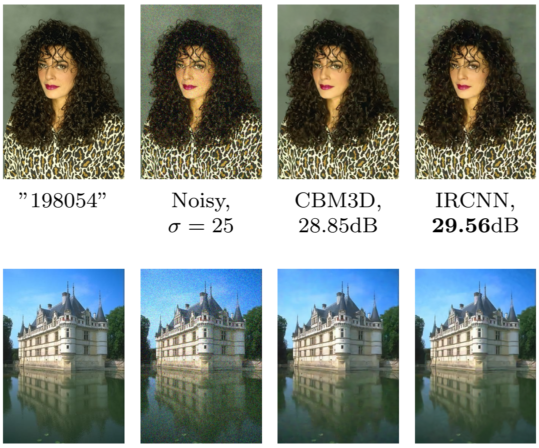
Figure 1: Proposed Image restoration convolutional neural network

2.2 Training

We optimize the network parameters $\Theta = \{W_i, B_i\}$, $i = \{1, 2, ..., l\}$ by minimizing the loss between the set of clean images $\{\boldsymbol{Y}_i\}$ and images predicted $\{\hat{\boldsymbol{Y}}_i\}$ from the noisy image set $\{\boldsymbol{X}_i\}$. Let us define this overall mapping as $\vec{\hat{Y}_i} = F(\vec{X_i}, \Theta)$. Then the optimal parameters are obtained as,

$$\hat{\Theta} = \arg\min_{\Theta} \frac{1}{n} \sum_{i=1}^{n} \|F(\boldsymbol{X}_i, \Theta) - \boldsymbol{Y}_i\|_2^2 \qquad (1)$$

where n is the number of images used for training the network. Minimizing the mean squared error between the clean and predicted image is performed by randomly sampling some smaller images from the clean/noisy images. Some pre-processing is done on these "sub-images" in the form of mean subtraction and normalization. In order for the size of the input and output sub-image to be same, we perform padded convolution in each layer. In our implementation, we used $3 \times 32 \times 32$ sized sub-images. For each kind of noise we produce the noisy image from the clean image and sample the same spatial location on each of these image pairs to produce a clean/noisy sub-image pairs.Training is done following standard mini-batch gradient descent approach(batch-size=256) with momentum update.


"198054" | Noisy, $\sigma = 25$ | CBM3D, 28.85dB | IRCNN, **29.56**dB

"Castle" | Noisy, $\sigma = 25$ | CBM3D, **32.24**dB | IRCNN, 32.17dB

Figure 2: Image denoising results(PSNR) on Berkeley segmentation dataset

3 Experimental Results

3.1 Image denoising

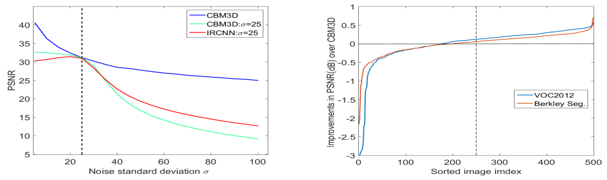
Noisy images are created by corrupting clean images with additive white Gaussian noise. For our experiments we trained our network IRCNN for noise levels of $\sigma = 25$ and $\sigma = 50$. For training we extract sub-image pairs from original clean/noisy image pairs and train our network on these sub-image pairs.

For training our network we use data from two datasets:(1) Image-Net[10] (2) MSCOCO [11]. We randomly choose 6000 images from each of the two datasets and corrupt each image with additive white Gaussian noise. From each such image pair, we choose 16 random samples of size $3 \times 32 \times 32$, giving a total of 192,000 sub-image pairs. It took 4 days to train the network on a modern GPU, during which time 4000 passes over all the 192,000 sub-images were performed for IRCNN network. However, for testing purpose we used two test datasets (1)Berkeley segmentation dataset[12] and (2)Pascal VOC 2012[13] for evaluating our performance. Testing is performed by sliding window technique and averaging overlapping reconstructions.

Images from the Berkeley segmentation dataset, used in [6], were used to compare the performance of IRCNN with baseline method K-SVD[6] and CBM3D[4], a state-of-the-art color image denoising method. For each image, experiments were performed 10 times and the average PSNR value was reported. We used PSNR values reported by the authors in [6] for the comparison. For CBM3D, we used the Matlab code provided by the authors for our evaluations. Table 1 shows the comparison of performance for image denoising with $\sigma = 25$ and $\sigma = 50$. On this small testing dataset, IRCNN produces a superior performance for 3 out of 5 images(for both $\sigma = 25$ and $\sigma = 50$) and has the best overall performance out-performing both sparse coding-based KSVD method and CBM3D method. We also tested with Convolutional Autoencoders on 96000 image patches for 1000 epochs. The average PSNR for denoising task on the 5 images in Table 1 are 27.36dB and 25.06db for sigma=25 and 50 respectively. Since it is our own implementation and we believe that these CNN autoencoder results can be

Table 1: Image denoising performance for Berkeley segmentation dataset images

Image	$\sigma = 25$			$\sigma = 50$	
	KSVD	CBM3D	IRCNN	CBM3D	IRCNN
Castle	31.19	**32.24**	32.17	**28.67**	28.66
Mushroom	30.26	**31.20**	30.92	**27.77**	27.60
Horse	29.81	30.67	**30.83**	27.59	**27.84**
Kangaroo	28.39	29.19	**29.30**	26.37	**26.45**
Train	28.16	28.72	**28.88**	24.52	**25.06**
Average	29.56	30.40	**30.42**	26.98	**27.12**



(a) (b)

Figure 3: (a) IRCNN trained at $\sigma = 25$ vs CBM3D. (b) Improvements of IRCNN compared to CBM3D.

slightly improved by hyper-parameter tuning and more training with larger datasets, we do not report it in Table 1. Figure 2 show the qualitative comparison for image denoising.

For a more comprehensive comparison with CBM3D method, we tested the performance of both the methods on two large datasets of 500 images from Berkeley segmentation dataset[12] and Pascal VOC 2012[13] dataset. For each image in the dataset, experiments were performed 5 times and the average value was used. Improvements in PSNR achieved by our method, compared to CBM3D for $\sigma = 25$ on both datasets is shown in figure 3(b). The comparisons between the CBM3D and IRCNN is shown in Table 1. These quantitative results demonstrate that the proposed IRCNN model performs at par with(often better than) state-of-the-art denoising methods.

We also test the IRCNN model trained at $\sigma = 25$ for various other noise levels and plot the PSNR performance. The plot at various noise levels for the image "mushroom" from Berkeley segmentation dataset is shown in figure 3(a) which shows that our learned model produces competitive performance compared to CMB3D at $\sigma = 25$ but performance deteriorates for other noise levels. To compare with similar effects in CBM3D, we fixed the input parameter to $\sigma = 25$ for CBM3D. Similar performance is seen for CBM3D algorithm with knowledge of $\sigma = 25$, although our learned network performs slightly better at higher noise levels. CBM3D provided with correct noise information produces a superior performance which is understandable.

3.2 Blind image inpainting

We perform image inpainting task for images corrupted with (1) uniformly distributed missing pixels (2) complicated patterns like text. The training data for blind inpainting is same as that for image denoising. We make no attempt to change the network architecture for this task and perform training on IRCNN.

3.2.1 Missing pixel inpainting

Noisy images were created by randomly assigning 80% of the pixel values in each channel as zeros and then 192,000 sub-images(similar to denoise case) were created by randomly sampling 16 images from each clean/noisy image pair. The training procedure is similar to the image denoising case.

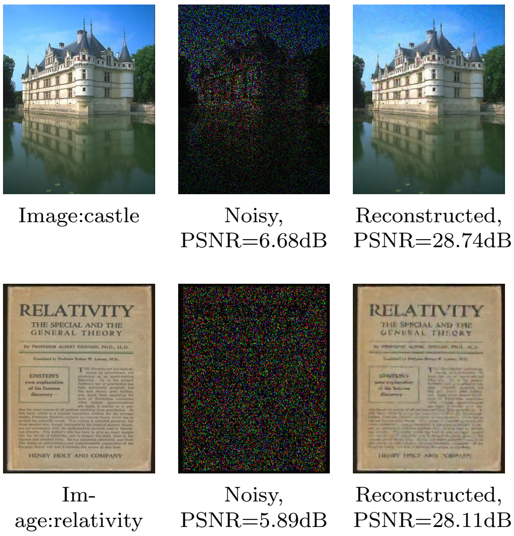


Image:castle — Noisy, PSNR=6.68dB — Reconstructed, PSNR=28.74dB

Image:relativity — Noisy, PSNR=5.89dB — Reconstructed, PSNR=28.11dB

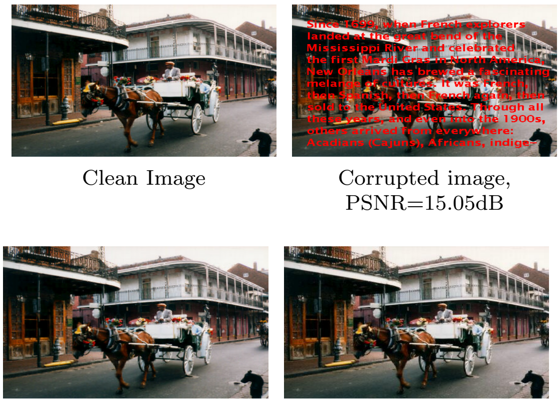
Figure 4: Missing pixel inpainting results on various images by IRCNN

For 80% missing pixel case we obtain a PSNR performance of 28.74dB for the image "castle" from Berkeley segmentation dataset. The best reconstruction performance of 29.65dB reported in [6] by non-blind K-SVD inpainting technique. Our model has a lower PSNR performance compared to K-SVD because we solve a more difficult task of blind inpainting where the location of the missing pixels are unknown compared to the non-blind case where the information about the location of the missing pixel simplifies the inpainting problem to a large extent.

Qualitatively our model shows good reconstruction quality, as seen from the results in figure 4. For the castle image, the reconstructed image is visually similar to the original clean image. For the "relativity" image, we observe that, while the text in the noisy image is not at all clearly visible, the image predicted by our model successfully restores readability for moderately large text. These qualitative and quantitative results demonstrate the effectiveness of our model for missing pixel restoration.

3.2.2 Text removal

For text removal problem, noisy images were created by superimposing random texts on the clean images from 15 different font styles and font-size varying from 15pix to 25pix. Following similar methodology as previous methods, we create 192,000 sub-images by randomly sampling 16 images from each clean/noisy pair and training is done following standard mini-batch gradient descent with similar parameters as mentioned for previous tasks. Interestingly, we observed our model does not differentiate between the various tasks(denoising or inpainting) it is learning and takes almost similar time for learning the direct mapping between input and output in each case.


Since 1699, when French explorers
landed at the great bend of the
Mississippi River and celebrated
the first Mardi Gras in North America,
New Orleans has brewed a fascinating
melange of cultures. It was French,
then Spanish, then French again, then
sold to the United States. Through all
these years, and even into the 1900s,
others arrived from everywhere:
Acadians (Cajuns), Africans, indige-

Clean Image

Corrupted image, PSNR=15.05dB

IRCNN, PSNR=30.95dB

FoE,PSNR=32.35dB

Figure 5: Comparison of superimposed text removal performance

We tested the performance of our algorithm of the classic image used in the original inpainting paper[14] for text removal. Quantitative evaluation on the data revealed that our model obtained a PSNR value of 30.95dB. For lack of blind inpainting methods, we compare our performance with non-blind inpainting method of Field-of-Experts(FoE) model[2] and K-SVD model[6]. For this image, FoE achieves PSNR value of 32.35dB while K-SVD(as reported in [6]) achieves 32.45dB. We used the Matlab code provided by the authors, for evaluating the performance using FOE model. The time required by FoE for text removal was 584 seconds (using 24 ,5 × 5 filters) while IRCNN took 5.6 seconds for the same task.The capability of our method for blind inpainting of complicated superimposed texts is a notable contribution of this paper.

4 Conclusion

We have presented a fully convolutional deep learning approach for image restoration of RGB images. The proposed approach learns an end-to-end mapping between noisy and clean image patches. Experimental evaluations on image denoising show that fully convolutional image denoising demonstrates competitive performance with the state-of-the-art methods. For image inpainting, our model solves the difficult problem of blind inpainting and successfully removes uniformly distributed impulse noise as well as sophisticated patterns like text with the impressive visual quality of reconstruction. These results show that proposed FCN model can indeed provide a good model for low-level image restoration problems. In addition to the demonstrated competitive accuracy, the proposed FCN based image restoration model is light-weight and feed-forward in structure which can be readily implemented in practical systems.

References

[1] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60** (1992) 259–268

[2] Roth, S., Black, M.J.: Fields of experts: a framework for learning image priors. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Volume 2. (2005) 860–867 vol. 2

[3] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing **16** (2007) 2080–2095

[4] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance chrominance space. In: 2007 IEEE International Conference on Image Processing. Volume 1. (2007) I – 313–I – 316

[5] Criminisi, A., Perez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing **13** (2004) 1200–1212

[6] Mairal, J., Elad, M., Sapiro, G.: Sparse representation for color image restoration. IEEE Transactions on Image Processing **17** (2008) 53–69

[7] Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., eds.: Advances in Neural Information Processing Systems 25. Curran Associates, Inc. (2012) 341–349

[8] Luisier, F., Blu, T., Unser, M.: A new sure approach to image denoising: Interscale orthonormal wavelet thresholding. IEEE Transactions on Image Processing **16** (2007) 593–606

[9] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **38** (2016) 295–307

[10] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09. (2009)

[11] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common Objects in Context. In: Computer Vision – ECCV 2014:, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V. Springer International Publishing, Cham (2014) 740–755

[12] Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision. Volume 2. (2001) 416–423

[13] Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes challenge: A retrospective. International Journal of Computer Vision **111** (2015) 98–136

[14] Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '00, NY, USA, ACM Press/Addison-Wesley Publishing Co. (2000) 417–424