International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Image Enhancement using Generative Adversarial Networks

Yash Prakash, Bhavesh Phumbhra

Abstract: ***Mobile Photography has been brought to a significantly new level in the last several years. The quality of images taken by the compact lenses of a smartphone have now appreciably increased. Now, even some of the low- end phones of the market spectrum are able to take exceedingly good photos in suitable availability of lighting, due to the advancement in numerous software methods for processing the images post capture. However, despite these tools, these cam- eras still fall behind the aesthetic capabilities of their DSLR counterparts. In the quest to achieve high quality images through a smartphone camera, various image semantics are inadvertently ignored leading to a less artistic image quality than a pro- fessional camera. Although numerous techniques for manual as well as computerized image en- hancement tasks do exist, they are generally only focused on brightness or contrast and other such global parameters of the image and does not go on to improve the content or texture of the image and neither do they take the various semantics of the image into account. Moreover, they are usually based on a predetermined set of rules that never considers the actual device specifics that is capturing the image — the smartphone camera. For our enhancement, we have endeavored to use a unique deep learning technique to transform lower quality images from a smartphone camera into DSLR-quality images. To enhance the image sharpness, we have used an error function that combines the three losses - the content, texture and color loss from the given image. By training on the large-scale DSLR Photo Enhancement Dataset, we have optimized the loss function using Generative Adversarial Networks. The end results produced after testing on a number of smartphone images yield enhanced quality images comparable to the DSLR images with an average SSIM score of approximately 0.95.***

Index terms - image enhancement neural networks generative adversarial networks gans deep convolutional neural-networks generator loss vgg loss multi-component loss function ssim score comparison

I. INTRODUCTION

Smartphone photography has seen an exponen- tial rise in the quality of images captured in the last several years. The decade we live in has been marked with extensive development in mobile phone cameras that are continually remodeled by the manufacturing companies to inch closer to capturing the perfect shot. When the very idea of low-light photography was met with incredulity only a decade ago, it is now a reality in a majority of smartphones in the market.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.

* Correspondence Author

Yash Prakash*, Department of Computer Science and Engineering student at SRM Institute of Science and Technology, Tamil Nadu, India.

Bhavesh Phumbhra, Department of Computer Science and Engineering student at SRM Institute of Science and Technology, Tamil Nadu, India.

But when compared to the quality of an image taken through profes- sional cameras like the digital single-lens reflex devices, they trail far behind. Larger sensors and high-aperture optics on the professional cameras produce better resolution for the image, color retainment and significantly less noise levels,

and their additional sensors also help in fine-tuning the image capture parameters. These major physical differences of the professional cameras from the smartphone cameras result in some major difficulties, making the professional camera quality not feasible for the smartphones. Therefore, the main approach for post processing of the image is still based on different photo editing softwares.

We propose a deep learning approach for the photo enhancement task. Using the concept of Generative Adversarial Networks from computer vision, we have developed a model to artistically transform a lower quality image captured by a smartphone camera into its high quality coun- terpart, the one taken by a DSLR. Through the implementations of the model we envision, aver- age consumers will be granted the power to take the perfect shot without being forced to indulge in larger, less-portable and significantly expensive professional cameras.

II. RELATED WORKS

The area of Computer Vision has not yet wit- nessed a detailed and comprehensive study of the problems associated with the automatic image quality enhancement, and a number of similarly tasked problems have been successfully imple- mented through various methods involving deep learning. The numerous related works done usu- ally deal with image to image translation problems with techniques like the following:

1. Image Modification:

Modification of image color, texture and contrast have been previously done in works such as:

a. A technique presented in [1] is the automated correction of exposure that can be used to predict the best photo specific non-linear tone curve, along with the best correction possible when the best exposure is known for every region of the image. Like the other automatic approaches, this method does not take into consideration preference of the user.

b. In [2] a new method is introduced that can generate visually diverse styles for an image with an unsupervised learning technique. Here, images with similar content were taken and the desired images (target) was generated with their style.

c. In [3] a different method is introduced that uses the image pixels and their local descriptions for reproducing various styles. This method helps for better modifications and image corrections when compared with some popular photo-editing softwares.

2. Image Deblurring:

a. In [4] the authors aim to recover the color and visibility features of the photo from the given hazy images. They use total variation loss minimization to remove the hazing from images. In total, there are four types of fog in images, according to [5]

— those being uniform, heterogeneous extinction,

By restoring the desired contrast, this method is applicable to work with all four kinds of fog situations in the images.

In [6] an end to end deblurring technique for in- motion image deblurring is introduced, based upon learning from a conditional GAN and as a special case of image translation task. This approach uses a multi-component loss function composed of – the loss from the adversarial network and the con- tent loss are used as the network to compare and hence calculate the distance between the enhanced images output by the generator network of the GAN and enhanced images. Whereas in [7], a bi- channel CNN is used with Mean Square Error loss. To recover original colors from the original source image, a technique like in [8] or [9] is used, to find the new values for the pixels based on their local description in the image. In [8], conditional generative adversarial networks deliver

better performance.

3. Image Super-Resolution:

The source (origi- nal) image is obtained from its low-quality version in these works:

a. In [10], a convolutional network and MSE loss are used to learn the low resolution to high resolution transformation for the image. This map- ping outputs a high resolution photo as output from an input of a lower resolution to a deep network. This is a CNN technique for a singular super resolution of the image, and is similar with the other method like in [11].

b. In [12], training of feed-forward network are performed for image style transfer tasks based on a perpetual loss function. This loss replaces the pixel to pixel loss that has been typically used in earlier methods. One of the best works of super- resolution is in [13] with GANs and a VGG based loss function.

III. METHOD

Our task comprises: converting a lower quality image I_s (a source image from a smartphone) into the target high-quality image I_t . The transforma- tion function is used to minimize the weight W given by:

$$W = arg\ min\ [\frac{1}{} \cdot \lambda(G(I^i), I^i)]$$

both heterogeneous extinction and luminance fog. Here, λ is our proposed combined loss function.

A. The Combined Loss Function

Our loss function is built on the premise that an image has three components that need trans- formation simultaneously for producing a higher quality image. These are namely, the color, the texture and the content components. Separate loss functions for each component is used to match our source and target images closely (pixel to pixel).

1) Color Loss: Since this is an independent component of the combined loss function, we need to eliminate the effects of content and texture losses. We do so by applying a Gaussian Blur on the two images, enabling us to compute the difference in color parameter between the given source and target blurred images.

An image when passed through the Gaussian Blur:

$$I_1(i, j) = \sum_{x,y} I(i + x, j + y) \quad G(x, y)$$

Here, G is defined as a two-dimensional Gaussian Blur function.

2) Texture Loss: Using Generative Adversarial Networks from [14], a log loss function is defined to predict the shift-invariant textural difference between the two images. Grayscale images are used so as to focus only on the textural features of the image, which are then fed into the Discrim- inator network which then predicts if the given image is a fake image(a generated image) or a real(original) one. The network is trained on image pairs of smartphone and DSLR and is thus used to minimize the texture loss.

3) Content Loss: The *relu2_2 and relu3_3* from the pre-trained network of VGG-16 are used, alike to [12] is used to produce feature maps of the pair of images which describe their content quality specifically. The Euclidean distance (between the enhanced images output from the generator net- work and the real images) is then minimized via the content loss function.

B. Our proposed Generative Adversarial Network

The two major parts of the GAN are the Generator and the Discriminator Network. Our Generator network is the network that produces the enhanced images and is entirely convolutional with 32 channels, with the activation function as *ReLU* for all layers except the last one which has *tanh* for the output. The Discriminator network consists of seven convolutional layers and *LeakyReLU* is used as activation function, barring the last layer, the dense layer which has *sigmoid* for the output. This layer is responsible for producing the enhanced image that is similar to the image from the actual DSLR camera.

C. Calculation of total loss

The total loss is calculated from the combined loss function which is described as:

$$C.L. = \alpha \times L_{content} + \beta \times L_{texture} + \gamma \times L_{color}$$

Based on various experiments on the training dataset and some tweaks in the model, we have α from the produced features from *ReLU* layers of the VGG-16 network, β as 0.3 and γ as 0.2.

The Losses are described as:

$$L_{color} = \| B_a - B_b \|_2$$

The B_a and B_b being the blurred images

$$L_{texture} = - \sum \log D(G(I_s), I_t)$$

here, D is the discriminator model and G is the generator model from our network.

$$L = \frac{1}{\ } \|\psi(F(I)) - \psi(I)\|$$

Here, n consists of the number of feature- representations for the images, W and H represent the width and height of the features, and $F(I_s)$ is the enhanced image from the generator network while ψ representing the feature maps.

IV. EXPERIMENTS

The dataset used for training our network is DPED - "DSLR Photo Enhancement Dataset" from [15]. It consists of images taken from three smartphone cameras simultaneously (iPhone, Sony and Blackberry) and one professional DSLR-camera. The smartphone cameras being the low quality image source and the DSLR images being their high-quality coun- terparts. For our experiments, we have used only the iPhone and Sony smartphone data for training along with their respective high-quality images.

Due to difference in aperture size and other factors of DSLR cameras from the smartphone ones, we used a cropping algorithm from [16] to get the overlapping regions from image pairs and then perform crop from the intersection to get images of same resolutions.

A. Training

The training of the image transformation net- work was performed on GPU *Nvidia 1050Ti* with 30k iterations with 30 as the batch size. Our training time was approximately 19 hours.

B. The Results:

The iterative progression of the Generator net- work can be seen from the images in Figures 1 and 2:

Figure 1 represents the generated image with respect to the input (source) image in the first iteration of the training.

Whereas, Figure 2 shows the generated image after 30k iterations of training. This method is repeated for both the datasets used for the training

— iPhone image dataset and the Sony image dataset.

Some examples of photos after application of our network:

1. Sony model: The Figure 3 shows the en- hancement done with the model trained on the Sony dataset.
2. iPhone model: The Figure 4 shows the en- hancement done with model trained on the iPhone dataset.

C. Comparison Methods

The two established methods are used: 1. From Dong et al. in [10] which performs image super- resolution through image translation. 2. From Kim et al. in [11] which uses deep convolutional net- work for single image super-resolution and a VGG network based loss function.

V. CONCLUSION

Our work is focused primarily on producing professional camera quality camera images like that of the DSLRs from user smartphone images. We use the combined loss function to improve upon the classic image enhancement methods. Using the popular metric of SSIM, we compare our network with two others and conclude that our method yielded a better score on average for both datasets. The method we devise can be applied to various smartphone cameras in real life to get high quality, DSLR like photos.

There are flaws that come with our method such as, amplification of high-frequency noise components in images due to the very nature of the GANs. In some cases, the artificial nature of the enhanced images can't be discarded for reasons like excessive color deviations and contrasts.

REFERENCES

1. L. Yuan and J. Sun, "Automatic Exposure Correction of Consumer Photographs," in *Computer Vision - ECCV 2012. ECCV 2012*, F. A. ., L. S. ., P. P. ., S. Y. ., and S. C. ., Eds., vol. 7575. Springer, 2012.
2. J.-Y. Lee, "Automatic Content-Aware Color and Tone Stylization," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2470–2478, 2015.
3. Z. Cheng, Q. Yang, and B. Sheng, "Deep Colorization," 12 2015, pp. 415–423. [Online]. Available: 10.1109/ ICCV.2015.55
4. L. Li, W. Feng, and J. Zhang, "Contrast Enhancement based Single Image Dehazing via TV-l1 Minimization," in *Proceedings - IEEE International Conference on Multimedia and Expo*, vol. 2014, 07 2014. [Online]. Available: 10.1109/ICME.2014.6890277
5. N. Hautiere, J. P. Tarel, and D. Aubert, "Mitigation of vis- ibility loss for advanced camera based driver assistances," *Trans. on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 474–484, 2010.
6. O. Kupyn, "DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks," *IEEE/CVF Confer- ence on Computer Vision and Pattern Recognition*, pp. 8183–8192, 2017.
7. W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single Image Dehazing via Multi-scale Convo- lutional Neural Networks," in *Computer Vision - ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 154– 169.
8. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image- to-Image Translation with Conditional Adversarial Net- works," *arxiv*, 2016.
9. R. Zhang, P. Isola, and A. A. Efros, "Colorful Image Colorization," in *ECCV*, 2016.
10. C. Dong, C. C. Loy, K. He, and X. Tang, "Learn- ing a Deep Convolutional Network for Image Super- Resolution," pp. 184–199, 2014.
11. Kim, J. . Lee, J. . Lee, and Kyoung, "Accurate im- age super-resolution using very deep convolutional net- works," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), and others, Ed., 2016, pp. 1646–1654.
12. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," pp. 694–711, 2016.
13. C. Ledig, L. Theis, F. Huszar, J. Caballero Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 07 2017, pp. 105–114. [Online]. Available: 10.1109/CVPR.2017.19
14. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," pp. 2672–2680, 2014.
15. A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. V. Gool, "DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks," in Proceedings of the IEEE international conference on computer vision, 2017.
16. D. G. Lowe, "Distinctive Image Features from Scale- Invariant Keypoints," Int. J. Comput. Vision, vol. 60, no. 2, pp. 91–110, November 2004. [Online]. Available: 10.1023/B:VISI.0000029664.99615.94

Fig. 1. Iteration 0 (left: input image, right: generated image)

Fig. 2. Iteration 30k(left: input image, right: generated image)

Fig. 3. Before and after enhancement- Sony model

Fig. 4. Before and after enhancement- iPhone model

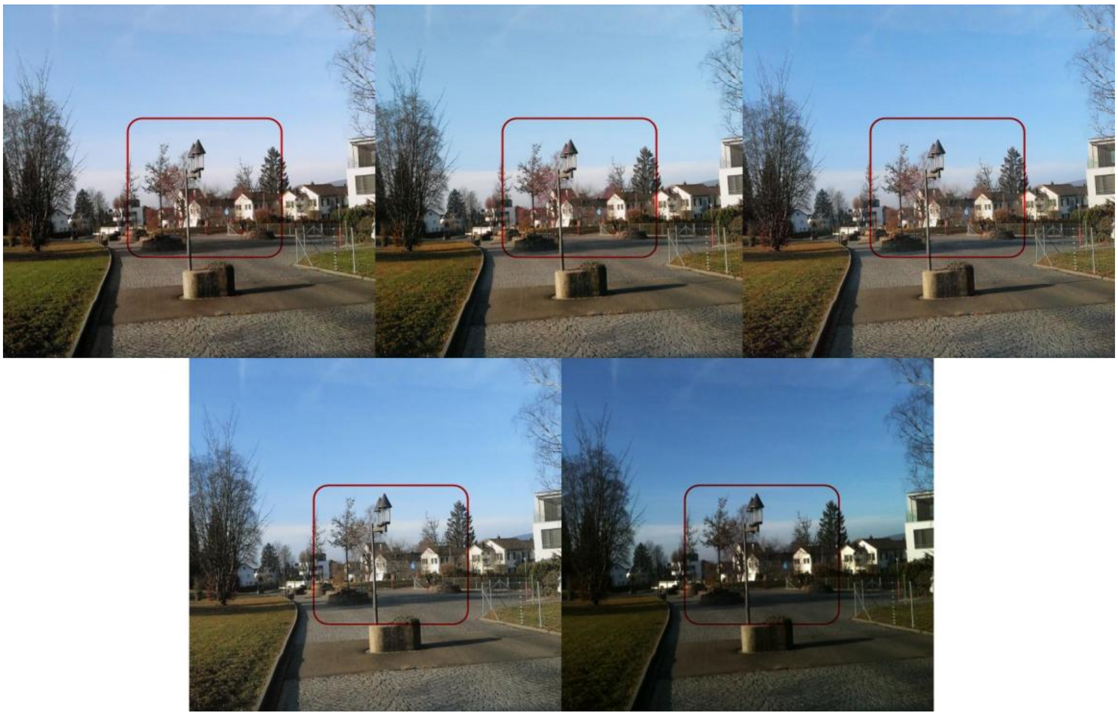

Fig. 5. Images from top to bottom and left to right include: Dong et al network, Kim et al network, our proposed method, The high-quality DSLR image from dataset, the low-quality smartphone image from the dataset.

Table I. Average Ssim Score Comparison

Phone image	Proposed	Dong et al.	Kim et al.
iPhone	0.9451	0.909	0.9356
Sony	0.9558	0.9303	0.9434

AUTHORS PROFILE

Yash Prakash is a Bachelor's Computer Science and Engineering student at SRM Institute of Science and Technology, Tamil Nadu, India. His research interests include Machine Learning algorithms, Computer Vision and Database Management Systems.

Bhavesh Phumbhra is a Bachelor's Computer Science and Engineering student at SRM Institute of Science and Technology, Tamil Nadu, India. His research interests include Machine Learning, Computer Networks and Natural Language Processing algorithms.