Image Deblurring using IRC-Conditional Generative Adversarial Network

Gouri. A. V\textsuperscript{1}, G. Preethi Ravisankar\textsuperscript{2}, Gowri H. Bhagavathi\textsuperscript{3}, Bhadra Maya Anilkumar\textsuperscript{4}, Anoop K. Johnson\textsuperscript{5}

B.Tech. Student\textsuperscript{1, 2, 3, 4}, Assistant Professor\textsuperscript{5}

Department of Electronics and Communication Engineering
Mar Baselios College of Engineering and Technology, Thiruvananthapuram, Kerala, India

Abstract:
Image deblurring is a method of image processing in which the quality of the blurry image is restored. It has gone a long way in the past decade. While clicking a picture, we want the captured image to be a perfect visualization that we see through our naked eyes. But every image is more or less blurry. The blurriness in the captured images may be due to motion blur, improper holding of the camera, dirty lens, wrong focusing and shutter button delay, etc. Thus, deblurring of a blurry image is a fundamental process in transforming those images into structurally sharp, defined and useful ones. Convolutional Neural Network (CNN) is used widely for deblurring of images as it can find the blur kernel of an image. Recently, Generative Adversarial Network (GAN) is performing well and hence generating images in an end-to-end trend. Therefore, we propose an Inception-ResNet-ConvLSTM Conditional Generative Adversarial Network (IRC-GAN) based deblurring algorithm in which the blurred image is passed through a modified generator which is a combination of Inception block and Res-ConvLSTM blocks that extracts the spatial-temporal features of the blurred image. Furthermore, it is reconstructed into sharp images as a coarse-to-fine mode. A DenseNet is also used in this generator network to prevent over-fitting. To determine the generator performance, a conditional discriminator is proposed through which blurry images are taken in as blur conditions and the reconstructed images are differentiated from real sharp images. A combined loss function is proposed to train the entire network.

Keywords: Image deblurring, Generative Adversarial Network, Generator, Inception, ResNet, ConvLSTM, DenseNet, Conditional Discriminator, Single image deblurring.

I. INTRODUCTION

There are trillions of photographs taken on each and every day in various places all around the globe but only a very few turns out to be the most perfect ones. The images captured can be degraded due to various reasons such as motion blurring, improper holding of camera, dirty lens, shutter button delay and depth of field. However, the captured photographic images in the real scenes have the most complicated blurry outcomes in many cases. The blurry structures can be classified locally and globally based on certain conditions. The local blurry structures arise due to motion in the natural scene and global blurry structures are caused due to improper holding of camera or variation in screen depth. These structures affect the quality of the captured image as well as the texture and gradients of the image. Thus, image restoration is necessary in removing various blurry artifacts and deblurring the image in best possible ways. Image deblurring is studied in the applications of image processing and computer vision field. It is the process of restoring blurry image into a structurally sharp image and retrieving lost information of image due to blur artifacts. Image deblurring, thus can be classified in to “non-blind deblurring and blind deblurring”. In blind image deblurring, the information of blur kernel is not available whereas in non-blind deblurring, blur kernel known assuming that it is uniform.

Thus, the image degradation model of deconvolution can be defined as,

\[ b = k * s + n \tag{1} \]

where \( b \) is the ‘blurry image’, \( k \) is the ‘blur kernel’, \( * \) is the ‘convolution operator’, \( s \) is the ‘sharp image’, and \( n \) is the ‘noise component’. If the blur kernel is not uniform then while capturing the photograph, both salient object and background moves in their own directions at that instant. Different trajectories lead to this type of blur known as motion blur. Thus, we are dealing with motion blurred images such as images caught from video sequence and natural scenes. Therefore, in this paper, we propose a “Inception-ResNet-ConvLSTM Conditional Generative Adversarial Network” (IRC-GAN) for reconstructing sharp images from the blurry image which is passed through a modified generator network. It is represented by IRC-G which is a combination of Inception blocks, Residual Network and ConvLSTM blocks that extracts the spatial and temporal features of the motion blurred image and to reconstruct structurally sharp image in a coarse-to-fine mode. Hence to prevent over-fitting, a DenseNet is used. A conditional discriminator (CD) is used to assess the performance of deblurring in local and global fields. At the same time, blurry images are given as conditions to the discriminator which helps in distinguishing between TRUE and FAKE images. A combined loss function is used by taking in to account the losses of generator and discriminator to avoid gradient loss while training the whole network.

The major contributions made in this paper are as follows:

1) The blurry images are passed through a combination of Inception blocks and Res-ConvLSTM Generator (IRC-G) network which reconstruct the blurry image in to sharp images by extracting spatial and temporal features. A DenseNet is also used to for flattening and hence avoiding over-fitting.

2) A Conditional Discriminator (CD) is used to provide blurred images as conditions to the discriminator that helps in differentiating reconstructed images from the real sharp images.
It also helps in evaluating the quality of image by analyzing both local and global scopes.

3) Further training of generator is done with the output of discriminator such that, when back propagation happens through all the layers of IRC-G and CD, the gradients are vanished. Hence, a combined loss function is proposed for gradient enhancing and helps in the training of generator.

II. RELATED WORKS

Deep Learning is nowadays widely used in image deblurring. ‘Convolutional Neural Network’ (CNN) is the widely used technique in the process of image deblurring. It is understood as an effective method for ‘denoising, restoration and super resolution’ of images [2]. It is capable of performing image processing in huge range and hence removing noise and blurs. CNN always tries to keep up with algorithm. Most of the deblurring algorithms have achieved a promising performance. Some learning methods and traditional energy-optimization based methods [8] depends on the blur kernel traits as complicated natural blurs can hardly be parameterized. Seungjun Nah [5] used a multiscale CNN for deblurring a blind image that reconstructs structurally sharp images back-to-back. Also, ‘deblurring and super resolution’ techniques are both used in low resolution images. Rozenn Dahyot [6] developed a deblurring and super resolution CNN which learns to provide an end-to-end mapping that takes in the blurry image and directly provides deblurred high resolution reconstructed image. Kupyn et al. [3] use Generative Adversarial Network (GAN) [3] for deblurring purpose. Noroozi et al. [11] proposed Deblur Network (DeblurNet), a multi-scale CNN architecture which uses blurred picture for directly predicting a sharp image. Isola et al. [12] proposed “conditional GAN” (cGAN), defined in the form of ‘pix2pix’, for producing the yy image by learning a mapping function from a random noise vector zz and an observed xx image. That is,

\[ G_{\theta}(xx, zz) \rightarrow yy \] (2)

Ramakrishnan et al.[13] presented a novel GAN architecture-based deep filter combined with Dense Network architecture and global skip link for resolving the issue in relative speed among the target in 3D space and the camera causing a blurring effect which is spatially varied throughout whole image. Kupyn et al. [3] proposed “DeblurGAN”, a process of deblurring focused on cGAN and Residual Network architecture. A scale recurrent network (SRN) was put forth by Xin Tao [4] which reduces the number of trainable parameters significantly. This SRN can incorporate recurrent modules, in which the hidden state takes over important information and helps in restoration across scales. The already existing algorithms are optimized by perceptual loss or pixel-wise computation[4]. These calculations can hardly assess local similarities and ignoring the global structural features of the image. In order to overcome such a condition, the discriminator is equipped with recurrent receptive field to determine the reconstructed sharp image in local and global regions [1].

III. METHODOLOGY

The proposed algorithm is an IRC-GAN which consists of a modified Generator that is a combination of Inception blocks, Residual Network, ConvLSTM blocks (IRC-G) and a Conditional discriminator (CD) in which blurry images are provided as conditions to the discriminator. The generator (IRC-G) takes in images which are blurry (B) as input. The output from the generator (IRC-G) are reconstructed sharp images R(S). These sharp images R(S) or ground truth/original images (O) are given into the conditional discriminator as inputs and providing blurry images (B) as conditions. Both IRC-G and CD are trained with an effective combined loss function.

Modified generator

The basic idea in the working of generator is to reconstruct images that are blurred, which is given as an input. Then, these reconstructed sharp images are given as input to CD. Firstly, a blurred image of size 256x256x3 is input to the generator and the next convolutional layer’s output should be 256x256x64 feature maps with 64 kernels of size 7x7x3. The dimensions are preserved for prevention of checkerboard artifacts which often arise in networks made of pyramidical architecture. The modified generator is a combination of Inception blocks, Residual Network and Dense Network along with ConvLSTM blocks. The entire network consists of 2 convolutional blocks with stride 2, 9 IRC blocks with DenseNet blocks, 2 transposed convolutional blocks, 2 transition blocks and 1 ‘global skip connection’. The constant number of feature maps is defined with a parameter N. In this case, N = 64. Each and every convolutional block consists of a 3x3 layer of convolution, a layer for normalization, and a Rectified Linear Unit (ReLU) as activation function. There are 2 convolutional blocks. The output of first one consists of feature maps at a range of 128 (2N) and second output is that of 256 (4N). Our IRC block is broadly divided into 3 parts consisting of a Bottleneck layer, Inception, and concatenation part of ResNet along with ConvLSTM. The ‘bottleneck layer’ contains less neurons in that layer above and below it. This part has a layer for normalization, ReLU part, and a 1x1 layer of conv2D.
Conditional Discriminator
The main task of a discriminator aims at distinguishing the original sharp image and reconstructed sharp output. It checks whether the reconstructed image has sharp edges or not. If the reconstructed sharp image R(S) is similar to the original one O, the reconstructed image is discriminated as FAKE and that of original image as TRUE and hence we can say that IRC-GAN is working properly. When input images R(S) to the discriminator have sharp images, they are likely to be recognized as TRUE image which is the original images under certain conditions. However, the blurred pictures taken from nature scenes may not be due to shake of camera but also due to object movement. When we set camera location and just object moves, a significant risk is seen that the images taken will consist of both blurred and sharp gradients. Under such a condition, if these outputs are given to the discriminator as input, they are tending to be identified as the original ones due to the sharp edge structures in them. Thus, the discriminator will be confused and it is hard for it to distinguish because the such images with both sheer and rough gradients are trained and represented as ‘TRUE’ and ‘FALSE’ within unalike instants. Therefore, to avoid such a condition, the resultant blurry images (B) are given to the discriminator in the form of conditions as shown in Fig. 1, so as to provide the information of blurry area. This indeed helps to aim at the regions of blur by the discriminator.

Combined loss function
Our combined loss function consists of content, adversarial and perception loss. Adversarial loss pays attention on restoring texture features. The deblurgan converges but produces smooth and blurry images, when the it is trained without GAN components, [3]. The content loss function i.e., MSE loss at raw pixels and MAE loss are used as individual targets for optimization which leads to generated images having blur component features because of average pixel-wise probable rate of solutions to the regional area of pixels. Thus, perceptual loss is adopted which is the MSE loss. It mainly focuses on restoring the normal features. Training without the perceptual loss on pixel leads to a meaningless state. So, now the loss function is a mixture containing ‘Wasserstein GAN (WGAN) loss, perceptual loss, and /1-loss’ thus trains our IRC-G generator.

\[
L = L_{\text{WGAN}} + \lambda_1 L_{\text{per}} + \lambda_2 L_{\text{11}} \tag{4}
\]

Where WGAN is

\[
L_{\text{WGAN}} = -\mathbb{D}_{\theta_\mathcal{G}}(G_{\theta_\mathcal{G}}(I_B)) \tag{5}
\]

The blurred image I_B is given as input to the generator G_{\theta_\mathcal{G}} which generates a latent image which is then evaluated by the discriminator. The perceptual loss term which fits our need is as follows:

\[
L_{\text{per}} = \frac{1}{w_{ij}h_{ij}} \sum_{x=1}^{w_{ij}} \sum_{y=1}^{h_{ij}} (\Theta_{ij}(I_S)_{x,y} - \Theta_{ij}(G_{\theta_\mathcal{G}}(I_B))_{x,y})^2 \tag{6}
\]

where and \(h_{ij}\) and \(w_{ij}\) shows the height and width of the feature maps. When image becomes more blurry, we add a L1 loss by R(S) and O, it is as follows:

\[
L_{\text{11}} = \|I_S - R_{\theta_\mathcal{G}}(I_B)\|_1 \tag{7}
\]

where I_S is the ground truth, R_{\theta_\mathcal{G}}(I_B) is the reconstructed sharp image obtained by the IRC-G generator G_{\theta_\mathcal{G}} of the given blurred image I_B.

IV. TRAINING DETAILS
With TensorFlow backend, we developed our model using Keras platform of deep learning. The model training took place on a single NVIDIA GeForce 940MX, 4GB Dedicated Graphics, CUDA Enabled GPU. The model was trained with 256 images from 1000 GoPro training dataset images on random crops of size256x256. The learning rate is initially set to 10^{-4} for both networks of generator and discriminator and expected training for minibatch size is set to 16. After the first 150 epochs over the next 150 epochs, we decay the rate linearly from 10^{-4} A batch size = 1 and a sum of 5,000 epochs were taught to the model.

V. EXPERIMENTS AND RESULTS
We implemented our IRC-GAN model with Generator network (IRC-G) with 4 layers of Inception block, 9 ResNets followed by ConvLSTM and DenseNet. The model is fully trained simultaneously using 2103 images of blurry and original images with a quality of 720p within the GOPRO dataset. After fully training the model, 1111 images for testing from the GOPRO dataset are considered and finally Peak Signal to Noise Ratio (PSNR) calculation is done for performance evaluation. The proposed model IRC-GAN is trained with an average rate of 5000 epochs and a learning parameter which is initially set to 10^{-4} for both IRC-G (generator) and CD (discriminator) and the expected training for minibatch size is set to 16. We linearly decay the rate from 10^{-4} after the first 150 epochs then to the next set of 150 epochs. The training of the entire model was done fully with a batch size of 16 and a total of 5000 epochs. The Adam optimizer is being used with hyper-parameters. In the proposed IRC-GAN, the hyper-parameters are set as \(\beta_1 = 0.9, \beta_2 = 0.999\) and \(\epsilon = 10^{-8}\). The experiments are implemented using Keras with TensorFlow backend on a single NVIDIA GeForce 940MX, 4GB Dedicated Graphics, CUDA enabled Graphics Processing Unit (GPU).
Performance evaluation (using GOPRO dataset)

The performance evaluation of our model IRC-GAN followed certain experiments in ‘energy-optimization-based algorithms’ such as RC GAN [1], C. Min et al. [2], DeblurGAN [3], SRN-Deblur [4] and T. H. Kim et al. [5] are conducted. These ‘state-of-the-art algorithms’ along with our model IRC-GAN are trained to be having exact testing images as that of ‘GOPRO’ and are evaluated with ‘Peak Signal to Noise Ratio’ (PSNR). On comparing with rest of the ‘state-of-the-art methods’ as shown in Table I, our model is having much better PSNR value. RC GAN (1) and our proposed model IRC-GAN has a difference range of 1.2065 in which the PSNR of our model exceeds a factor by 1.2 when compared with RC GAN [1]. Similarly, when compared with DeblurGAN [3], the PSNR rate of our model exceeds by a factor of 4.2818. After testing with the prior GOPRO dataset, we have selected 45 images from the test set of GOPRO, out of which the lowest PSNR so obtained is 28.7908 dB as shown in Fig. 2 and the highest PSNR obtained is 32.1116 dB as shown in Fig. 5. We have also considered a mid value ranges of PSNR as shown in Fig. 3 and Fig.4. The figures from 6 to 13 shows the various deblurred output we have obtained using GOPRO dataset test images. The average PSNR so obtained by our proposed model IRC-GAN is 31.4818 dB as shown in Table I. If the PSNR value is high, the image quality is better. The highest image quality ranges from 30 dB to 40 dB. Therefore, our proposed model IRC-GAN when compared with other state-of-the-art algorithms is much effective in image deblurring with a PSNR value of 31.4818 dB as shown below.

Table.I: Average PSNR Results (in dB) Of The Proposed IRC-GAN and State-Of-The-Art-Algorithms On GOPRO Dataset

<table>
<thead>
<tr>
<th>State-of-the-art Algorithms</th>
<th>Average PSNR (in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. H. Kim et al. [5]</td>
<td>25.6000</td>
</tr>
<tr>
<td>C. Min et al. [2]</td>
<td>29.7800</td>
</tr>
<tr>
<td>SRN-Deblur [4]</td>
<td>30.1000</td>
</tr>
<tr>
<td>RC GAN [1]</td>
<td>30.2753</td>
</tr>
<tr>
<td>IRC-GAN (Ours)</td>
<td>31.4818</td>
</tr>
</tbody>
</table>

Figure.2. Lhs: blurred image; rhs: deblurred image with PSNR 28.7908 dB

Figure.3. Lhs: Blurred Image; Rhs: Deblurred Image With PSNR 30.1130 dB

Figure.4. Lhs: blurred image; rhs: deblurred image with PSNR 31.088dB

Figure.5. Lhs: Blurred Image; Rhs: Deblurred Image With PSNR 32.1116dB

Figure.6. Lhs: blurred image; rhs: deblurred image; PSNR 29.5005 dB

Figure.7. Lhs: blurred image; rhs: deblurred image; PSNR 29.8517dB
Single image analysis

Single images are those images which do not have a reference image or ground truth. Deblurring of such images are found to be a difficult task. There are various methods proposed for retrieving the latent sharp images under the motion of camera. But, our proposed model IRC-GAN is capable of restoring the sharp features in the single images. We have analyzed images captured using digital camera of ratio 1:1 with 1080 x 1080 resolution, as well as some CCTV footages (taken from internet). The deblurred output so obtained from these single images are found to be satisfactory.
The single image deblurred output so obtained are shown in the above set of figures. We have taken certain CCTV footages available in the internet as well as naturally blurred images from various scenes. As seen in those deblurred images, the lost gradients due to blurriness have been restored successfully to an extent. In Fig.14 and Fig.15, the semantic object is face and it is reconstructed into coarse to fine manner. In various CCTV footages as shown in Fig.16 and Fig.17, the letters and numbers can be read easily and are deblurred effectively. A CCTV footage of street view is shown in Fig.18 in which the salient objects are blurry in nature which are deblurred. The following figures, Fig.19 and Fig.20 are captured using digital camera with a ratio of 1:1 and they are deblurred in a coarse-to-fine mode. Since, these single images don’t have a reference image, the performance evaluation cannot be done. Instead, our model IRC-GAN can be used in various applications of deblurring CCTV-footages snipped during various crime scenes. Thus, providing a helping hand in various fields.

VI. CONCLUSION

Blurriness in image has been a major problem for the photographers as well as in various applications of image processing. So, image deblurring is a widely posed solution. The proposed IRC-GAN model will help in making the process more efficient. Our method can handle blind and non-blind blurriness for different kinds of blur. Also, an added advantage is that the method is based on deep learning. The different levels of blur observed in each image demonstrates that this architecture performs better in the removal. Also, the blur conditions given into the discriminator is capable of distinguishing true and fake images accurately. The proposed algorithm can be effectively applied in various applications and forms favourably, when compared to existing methods that have been specifically designed for each task. The overall accuracy of our proposed model came out to be much better and that meets our expected accuracy. Also, single image deblurring is made possible and is able to deblur images in an effective way.

VII. REFERENCES


