Aerial Imagery Super-Resolution Using a Super Resolution Generative Adversarial Network

Abbas khan
Department of Computer Science
Islamia College Peshawar, Peshawar, Pakistan
abbaskhaliil1997@gmail.com

Altaf Hussain
Department of Computer Science
Islamia College Peshawar, Peshawar, Pakistan
develper.altaf@gmail.com

Noreen Fayyaz khan
Department of Computer Science
Islamia College Peshawar, Peshawar, Pakistan
noreen.fayyaz@fu.edu.pk

Sultanat Shah
Department of Computer Science
Islamia College Peshawar, Peshawar, Pakistan
sultanatshah2@gmail.com

Abstract—Aerial images play a crucial role in many facets of worldly life such as smart agriculture, aerial post disaster monitoring and land use classification. Images obtained for performing these tasks are usually low quality due to certain limitations of image capturing devices. In order to address these challenges properly, high-resolution images are fundamentally very desirable which is not ubiquitously available. For this purpose, we use modern image spatial resolution enhancement techniques. In this paper, we aim to enhance the spatial resolution of the images and then comparing it with the traditional image upscaling techniques. Different image assessment techniques are used to assess the reconstruction quality of the reconstructed images. All results show the superior performance of the proposed technique by achieving the average value of 31.5 and 0.81 for PSNR and SSIM respectively. Moreover, higher MOS score is also recorded for the proposed method, which also fortifies the effectiveness of the proposed method.

Keywords— Aerial imagery, Deep learning, GAN, Image Super Resolution.

I. Introduction

High-resolution aerial imagery is considered very desirable in solution of challenges posed by the modern times in the fields such as surveillance[1], smart agriculture[2], and disaster monitoring[3], object tracking[4] and land use classification[5]. Image Super Resolution (SR) is one of the most attractive areas of research due to its application in a variety of real world problems such as medical imaging, surveillance and security and many more. Image Super Resolution adds extra pixels to the input Low-Resolution image to make its High-Resolution counterpart. Image Super Resolution techniques[6] help to overcome the limitations of low-cost vision sensors. It is quite difficult to obtain high-resolution images from low cost devices that are commercially used. Hence, High-Resolution images are obtained from their low-resolution counterparts by using super resolution techniques. Numerous SR techniques have been developed over time to enhance the quality of the image and to increase its resolution. Traditional image Super Resolution techniques have been observed to suffer from lacking high frequency details and blurring effects [7] . These methods have also been reported to suffer from distortions. Broadly, this is an intrinsically ill-posed problem due to the possibility of multiple HR outputs associated with a single LR input. To avoid these problems, we resort to some modern techniques that promise to mitigate the aforementioned problems up to much extent. Deep learning has been actively used for image quality enhancement since the inception of some of its modern techniques that are very robust to the problems persistent in the traditional techniques. We use some modern techniques as proposed by ledig et al[8] and compared it with the traditional techniques. This method improves the quality of HR reconstructed image by minimizing distortions during reconstruction.

Before deep learning dictionary based approaches has been widely used for image super resolution[9]. After the recrudescence of modern deep learning techniques, a lot of research has been carried out which yielded incredible results and solved many problems posed by modern times and limitations. Some of the famous techniques from Convolutional Neural Networks (CNN) based methods [10] to the most emerging and auspicious Generative Adversarial Nets based approaches (GAN) [8]. The gist of all this work done is that High Resolution imagery is considered desirable in every field. In the field of drone imagery, it is very necessary to obtain high-resolution data so that to use it for autonomous classification of semantically important areas and autonomous assessment of disaster affected area. We proposed in this work, the conversion of low-resolution drone imagery to high resolution based on modern deep learning technique i.e. SRGAN[8]. We get plausible high-resolution images for its low-resolution inputs by training the adversarial model. Evaluating the results quantitatively and qualitatively, our proposed method got much encouraging complements.

II. Proposed Methodology

Our proposed method aims to upscale the low-resolution images to high-resolution images using a scale factor s of 4. We apply SRGAN technique, which increases the pixel density of the input image and contend the fine-grained details in the input image. In our proposed method, we aim to predict a high-resolution image, Super Resolved image I^{SR} provided low-resolution image I^{LR} High-resolution I^{HR} images are made available only during training. For an image with width of W, height H and color channels C the resultant image I^{SR} we obtained is sW x sH x C.

We train our network with the aim to develop a Generating Model **G** parameterized by θ_G which convert the I^{LR} to I^{SR}. θ_G is the combination of several convolutional layers and bias layers.

A. Network Architecture:-

This work is primarily based on the most emerging technique of deep learning known as GAN[11]. GAN is a tool of deep learning which is formed by aligning two deep learning model in adversarial manner and trained with the problem specific loss function. In this work, a loss function known as perceptual loss is used along with the original adversarial loss to get the result. However, the adversarial models is briefly described in the subsections of the paper.

$$\min_{\theta_G} \max_{\theta_D} \mathrm{E}_{I^{HR} \sim p_{train}(I^{HR})} \left[\log D_{\theta_D}\left(I^{HR}\right) \right] + \mathrm{E}_{I^{LR} \sim p_G(I^{LR})} \left[\log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(I^{HR}\right)\right)\right) \right] \quad (1)$$

This equation reflects the complete overview of the SRGAN network. In equation (1) discriminator maximizes its output and Generator minimizes the output eventually reaching a converging point.

B. Generator:-

In order to get super resolved I^{SR} images we used adversarial model as proposed by , which use residual blocks[12] in the generator model and a plain network as discriminator. The Generator model constitutes 4 identical residual blocks R, two upsampling layers (i.e. Transposed Convolution) which upsamples the input image effectively. ReLU activation is used in the Generator. Super Resolution specific loss function is used in the network known as perceptual loss, which is the combination of content loss and adversarial loss. Content loss is the difference calculated between the original image whereas adversarial loss is the same loss as proposed in the original paper of GAN.

C. Discriminator

As Discriminator, we use a plain network with 4 convolutional layers and two dense layers which differentiates between the real and generated image. Leaky ReLU is employed in the Discriminator model D. The Discriminator model also constitute two dense layers at the end. Second last dense layer used in the model is often variable whereas the last layer is of fixed size having two nodes for discriminating real from fake images.

D. Objective function

For every deep learning model, we need a loss function, which we try to minimize for efficient training of the model. The loss function for SRGAN is a perceptual loss. Perceptual loss is the weighted combination of content loss and adversarial loss. Content loss is the difference calculated between the real image and generated image. Usually, VGG model is used to calculate the content loss between two images. Adversarial loss is the probability of discriminator network over the real and generated images. Both of these losses combine to form final loss known as perceptual loss. SRGAN model is always trained on the basis perceptual loss.

$$\mathcal{L}^{perceptual} = \mathcal{L}^{content} + \mathcal{L}^{adversarial} \quad (2)$$

The above equation (2) shows the combination of both losses, such that content loss is combined with adversarial to form a single loss function known as perceptual loss, which is our ultimate goal to minimize.

III. Experimental Results

A. Training dataset

We use publically available UCMerced landuse dataset for training our network. Originally, this dataset consist of 21 classes each having 100 images. For efficient training and avoid overfitting we use data augmentation techniques which increased the number of images to 1000 per class. Each image has a resolution of 256x256. During training Generator model, we reduce the size of each image to 64x64, which reconstruct it to high-resolution. The Discriminator is trained on real images as well as generated images.

B. Evaluation metrics

We evaluate the reconstructed images both qualitatively and quantitatively. For quantitative analysis, we use the standard evaluation metrics i.e. PSNR and SSIM[13] . For Qualitative analysis, we use MOS test[14], which is the human rating technique and is widely used for image quality assessment. All evaluation metrics have their own importance in image quality assessment however MOS score is also widely used for reconstructed image quality.

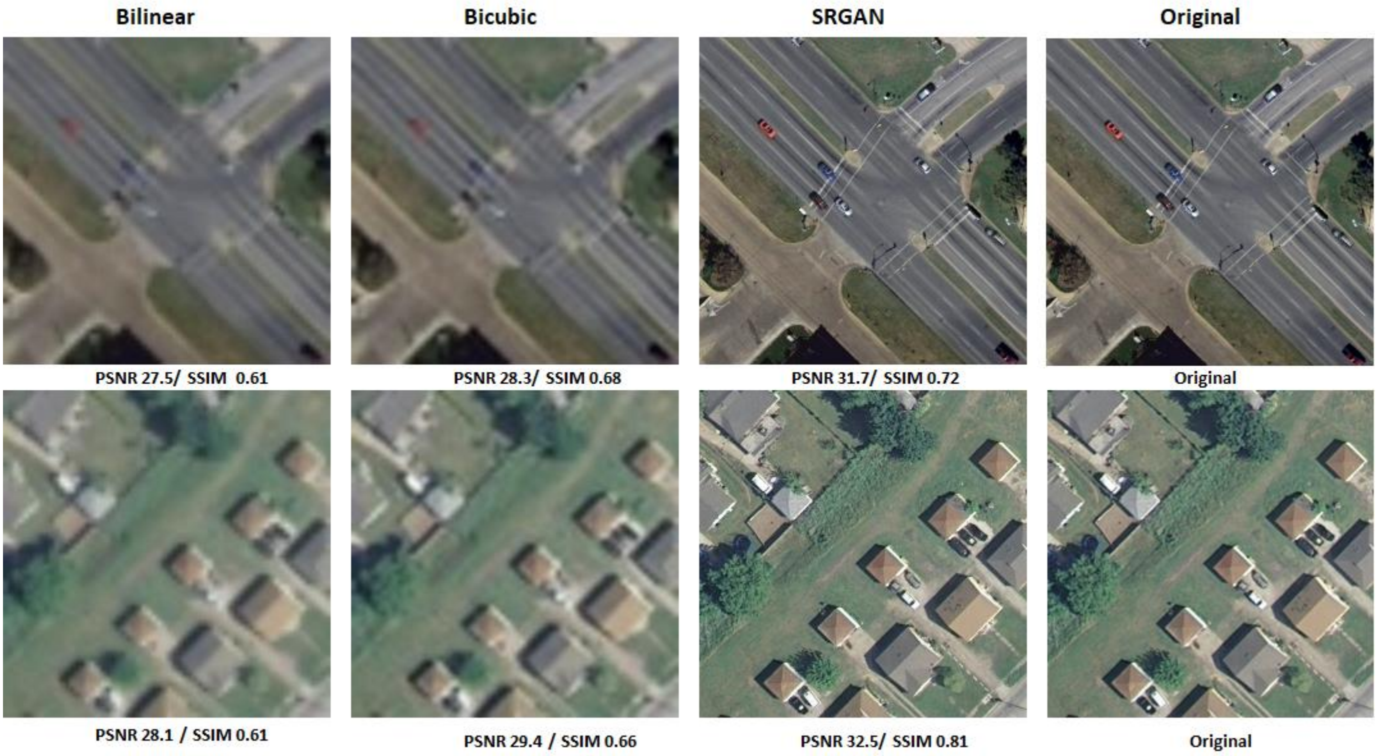

Figure 1 Qualitative results with corresponding PSNR and SSIM of image with the proposed and other reconstruction techniques.

From figure 1, it is easy to understand the efficiency of the proposed reconstruction technique with respect to qualitative and quantitative analysis. The highest PSNR, SSIM values and better visual quality of the reconstructed images suggest that the proposed technique produces plausible results much closer to the original images.

More than 100 testing images from different classes were reconstructed using the proposed method with Average PSNR and SSIM values were calculated 31.5db and 0.81 respectively. Moreover, all the reconstructed images were quite plausible to human raters and got an average rating of 4.2 much closer to excellent i.e. 5. The rating scale were defined between 0 and 5 whereas 0 for bad quality and 5 for excellent reconstruction quality of the images

IV. Conclusion

In this work, we presented a modern technique for reconstructing aerial images from low resolution to high-resolution images. Experimental results shown in the paper reflects the efficiency and effectiveness of the proposed technique for the problem. As aerial imagery plays most significant role in transitioning certain real life challenges from manual to automatic. This work can be extended to enhance low-resolution videos to high resolution for autonomous classification of different scenes that occurs most frequently in aerial imagery.

References

1. Kumar, R., et al., *Aerial video surveillance and exploitation.* Proceedings of the IEEE, 2001. **89**(10): p. 1518-1539.
2. Tripicchio, P., et al. *Towards smart farming and sustainable agriculture with drones*. in *2015 International Conference on Intelligent Environments*. 2015. IEEE.
3. Doshi, J., S. Basu, and G. Pang, *From satellite imagery to disaster insights.* arXiv preprint arXiv:1812.07033, 2018.
4. Bhattacharya, S., et al., *Moving object detection and tracking in forward looking infra-red aerial imagery*, in *Machine vision beyond visible spectrum*. 2011, Springer. p. 221-252.

5. Cleve, C., et al., *Classification of the wildland–urban interface: A comparison of pixel-and object-based classifications using high-resolution aerial photography.* Computers, Environment and Urban Systems, 2008. **32**(4): p. 317-326.
6. Yue, L., et al., *Image super-resolution: The techniques, applications, and future.* Signal Processing, 2016. **128**: p. 389-408.
7. Anbarjafari, G. and H. Demirel, *Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image.* ETRI journal, 2010. **32**(3): p. 390-394.
8. Ledig, C., et al. *Photo-realistic single image super-resolution using a generative adversarial network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
9. Yang, J., et al., *Coupled dictionary training for image super-resolution.* IEEE transactions on image processing, 2012. **21**(8): p. 3467-3478.
10. Dong, C., et al. *Learning a deep convolutional network for image super-resolution*. in *European conference on computer vision*. 2014. Springer.
11. Goodfellow, I., et al. *Generative adversarial nets*. in *Advances in neural information processing systems*. 2014.
12. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
13. Hore, A. and D. Ziou. *Image quality metrics: PSNR vs. SSIM*. in *2010 20th International Conference on Pattern Recognition*. 2010. IEEE.
14. Wang, Z., H.R. Sheikh, and A.C. Bovik. *No-reference perceptual quality assessment of JPEG compressed images*. in *Proceedings. International Conference on Image Processing*. 2002. IEEE.