IAES International Journal of Robotics and Automation (IJRA)
Vol. 12, No. 2, June 2023, pp. 179~183
ISSN: 2722-2586, DOI: 10.11591/ijra.v12i2.pp179-183

Development of image super-resolution framework

Sanvi Shekar, Kakarla Deepthi, D. Rahul, Dhawan K. S., Vinay V. Hegde
Department of Computer Science, R V College of Engineering, Bengaluru, India

Article Info

Article history:

Received Jul 19, 2022
Revised Oct 3, 2022
Accepted Dec 27, 2022

Keywords:

Generative adversarial network
Mean squared error
Peak signal to noise ratio
Satellite images
Structural similarity index
Super-resolution framework

ABSTRACT

There are some scenarios where the images taken are of low resolution and it is hard to judge the features from them, resulting in the need for enhancement. Super-resolution is a technique to produce a high-resolution image from a lower-resolution image. The intention here is to develop a system that enhances images of faces and satellite images by integrating these models and providing an interface to access this model. There have been various ways of achieving super-resolution using different techniques. Throughout the years, techniques involving deep learning methods, interpolation techniques, and recursive networks have been explored. We find it promising to use generative adversarial networks (GANs). The system has been deployed through Google Collaborate, Python libraries, and the TensorFlow framework. To assess the developed system, which consists of images, three metrics have been calculated. namely, peak signal-to-noise ratio, mean squared error, and structural similarity index. The model successfully demonstrated the capability of GANs by efficiently generating a high-resolution image from a low-resolution image for the given cases. The model would then be run on a standalone server for free Internet access for users to use super-resolution facial images and satellite images.

Corresponding Author:

Dr. Vinay V Hegde
Department of Computer Science, R V College of Engineering
Bengaluru, India
Email: vinayvhegde@rvce.edu.in

1. INTRODUCTION

Image super resolution is a method that is used to produce a higher-resolution image from a lower-resolution image. The up-sampled, higher-resolution image will have a higher pixel density, and therefore, more detailed features will be visible when compared to the original image. Image super resolution is used in a variety of domains, but most notably in computer vision applications to extract meaningful information from digital images. Sometimes high-resolution images are not readily available; applications like surveillance, forensics, and satellite imaging require images to be zoomed in on a specific area of interest in an image. Here, the Image Super Resolution technique can be used because it is less expensive than high resolution imaging setups, which are more expensive, and both serve the same purpose. Furthermore, existing low-resolution devices can use Super Resolution to convert the images captured in the system to produce a higher-quality image. There have been various ways of achieving super-resolution using different techniques. Throughout the years, techniques involving deep learning methods, interpolation techniques, and recursive networks have been explored. The networks here optimize the pixel difference between the predicted and output high-resolution image. Generative models attempt to optimize perceptual quality in order to generate images that are pleasing to the human eye. Thus, it is promising to use generative adversarial networks (GANs) to overcome the overhead [1]–[10].

GAN is a generative model in machine learning that discovers and learns the regularities or patterns in the input data and generates the output that plausibly could have been drawn from the original dataset. GANs consist of two sub-models: the generator model to train to generate new examples and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated). GANs find their use across a range of problem domains, notably in image-image translation [11]–[15].

A method was proposed for single image super-resolution [11], providing a direct end-to-end mapping between both the low- and high-resolution images. This method makes use of a deep convolutional neural network that takes the low-resolution image as input and outputs the high-resolution one. The method explored various network structures and parameters to achieve fast speed and state-of-the-art restoration quality. The method could handle three color channels at the same time and optimize all layers at the same time. Existing research includes the proposal of a new framework for estimating generative models [1] with the introduction of an adversarial process in which two models were trained separately: a generative model that trained from the data distribution and a discriminative model that estimated the probability that a sample came from the training data rather than from the generative model, which brought a breakthrough in how generative models could be made use of in terms of speed and accuracy [16]–[25].

Despite the breakthrough in accuracy and speed of super-resolution, one central problem existed, i.e., how to recover the finer texture details when super-resolution was done on a larger scale. A new method was proposed to improve the quality of the image largely by introducing the perceptual loss function [2], which consisted of adversarial loss and content loss. The overall enhancement increased, and later this new method was known as super-resolution GAN (SRGAN). The SRGAN [3] sometimes consists of unpleasant artifacts with hallucinated details. Enhanced SRGAN (ESRGAN) introduced residual-in-residual dense block, which provided stronger supervision for brightness consistency and texture recovery, thereby giving better visual quality with more realistic and natural textures when compared to the previous GANs.

2. METHOD

Similar to GAN architectures, the models considered here also contain two parts: a generator and a discriminator, where the generator produces some data based on the probability distribution and the discriminator tries to guess whether the data comes from the input dataset or the generator. The generator then tries to optimize the generated data so that it can fool the discriminator. Each of these generators and discriminators is a neural network consisting of intermediate convolutional neural network (CNN) layers that are essential to extracting features from the image.

Datasets for the proposed application are taken from Kaggle. Each of the models considered must be given two different datasets. Face enhancement model with images consisting of faces and satellite image enhancement with images containing satellite imagery. The Flickr-Faces-HQ (FFHQ) dataset is considered for faces. Extraction of the Deep Globe Road dataset for satellite images the FFHQ dataset consists of 17,000 images at 1024×1024 pixels. Deep Globe Road Extraction consists of 1024×1024 images specifically meant for road masking. Though these datasets are not meant for this purpose, they are sufficient to achieve our goals.

Using the images can be hectic for the training process. Training is done using TFRecords. The datasets for each of these datasets are divided into certain types of records. Records are the binary records for the TensorFlow framework. Each of these records holds images that are stored in batches of 1024×1024 resolution. However, operating on images of that resolution can slow down the process. Thus, random cropping of 256×256 is done on input, and a corresponding low-res counterpart is chosen to be a 64×64 image patch, aiming to achieve 4 times upscaling. In the data flow diagram in Figure 1, the user chooses the type of image depending on the use case (aerial image or face image), which then leads to the display module where the result is obtained and again gets trained by the GAN module.

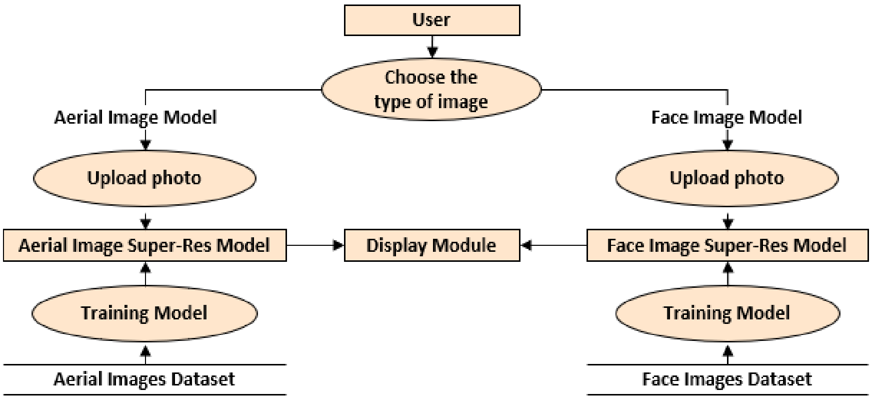

Figure 1. Data flow diagram

3. RESULTS AND DISCUSSION

Evaluation metrics are principles for testing diverse procedures and the behavior of the algorithms or strategies can be resolved utilizing metrics. A few strategies fulfill a portion of the metrics. In this, the yields that are obtained from the distinctive inputs specified to the scheme are contrasted with output according to requirements to check whether these metrics are fulfilled. This system was assessed in light of what number of the expectations worked consummately in given circumstances. The confidence score tells how confident and the probability, the model is about the object predicted. Two of the important metrics used for the proposed system are peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM) which are described in detail below.

PSNR is an expression representing the ratio between the maximum possible value of the power of a signal and the power of distorted noise affecting the quality of its representation. PSNR is given by

$$PSNR = 20\, log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right)$$

where the mean squared error (MSE) is

$$MSE = \frac{1}{mn}\sum_{0}^{m-1}\sum_{0}^{n-1}\left|\left|f(i,j)\ -\ g(i,j)\right|\right|^2$$

However, PSNR is not a reliable metric. Hence, further SSIM has been introduced.

The SSIM between 2 given images assigns a value between -1 and +1. *A value of +1* indicates that the 2 given images are very similar or the same while a *value of -1* indicates the 2 given images are very different. These values are adjusted to be in the range [0, 1], where the extremes hold the same meaning. SSIM calculates this value based on three features of an image, which are luminance (measured by averaging over all the pixel values), contrast (measured by taking standard deviation of all pixel values, and structure (the ratio between the input signal and standard deviation of the image). SSIM for two images can then be given as

$$SSIM(x,y)\ = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μx is luminance of image x, σx is contrast for image x.

For face super-resolution model, the model is trained with 10,000+ images.

- Dataset: FFHQ Dataset
- Loss Function - Discriminator + MSE
- Number of Epochs – 1,600

For satellite-imagery super-resolution, the model is trained with 1,900+ images.

- Dataset: Deep Globe Road Extraction Dataset
- Loss Function – Discriminator + MSE
- Number of Epochs – 1,600

Figure 2 represents values for d-real, d-fake, and generator loss values varying across 400 epochs of training. Calculating the accuracies is complicated for the images. The ground truth cannot be compared through mean square errors as they are obsolete and not a qualitative approach. Thus, MSE, PSNR, and SSIM metrics are taken to ensure image enhancement with better quality. According to this, an enhanced image must have a lower MSE value, higher PSNR value, and SSIM values closer to 1. Table 1 shows these values for 10 example images. The value from the low versus enhanced resolution is shown in Figure 3.

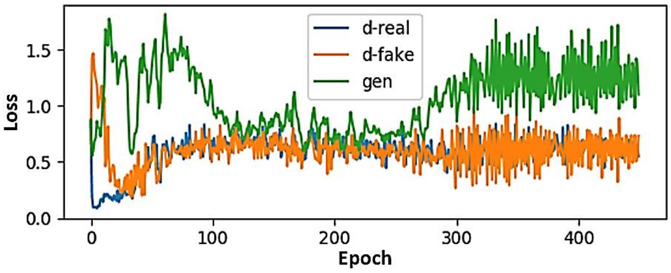

Figure 2. Loss graph (value loss vs epochs)

Table 1. MSE, PSNR and SSIM metric for low-res and enhanced image

Low-res image	Enhanced Image
1392_sat.jpg PSNR: 18.958832709705376 MSE: 2479.238712310791 SSIM: 0.41625687788079374	1392_sat.jpg PSNR: 24.23709487119858 MSE: 735.3466672897339 SSIM: 0.6419950600849048
14400_sat.jpg PSNR: 26.33920293360446 MSE: 453.19104766845703 SSIM: 0.5826171857016487	14400_sat.jpg PSNR: 31.75561014045952 MSE: 130.2088794708252 SSIM: 0.7779415941087989
14635_sat.jpg PSNR: 21.91884690563828 MSE: 1254.0559844970703 SSIM: 0.41149690344480333	14635_sat.jpg PSNR: 27.26519411957763 MSE: 366.1695308685303 SSIM: 0.6601726129794051
17150_sat.jpg PSNR: 26.63763220118849 MSE: 423.0954895019531 SSIM: 0.5837573443937609	17150_sat.jpg PSNR: 31.446718096919078 MSE: 139.8072862625122 SSIM: 0.7540682894551961
17945_sat.jpg PSNR: 24.675010509695184 MSE: 664.814624786377 SSIM: 0.502224560627941	17945_sat.jpg PSNR: 28.290148131829824 MSE: 289.19235134124756 SSIM: 0.6577874726561035
18111_sat.jpg PSNR: 25.257998539561967 MSE: 581.3019256591797 SSIM: 0.44272127237357023	18111_sat.jpg PSNR: 27.948711908254495 MSE: 312.84586906433105 SSIM: 0.580526048202101
18482_sat.jpg PSNR: 21.525676057611875 MSE: 1372.8845252990723 SSIM: 0.5178360781224681	18482_sat.jpg PSNR: 26.405759674704058 MSE: 446.29872703552246 SSIM: 0.672322812458153
18787_sat.jpg PSNR: 17.1552295412767 MSE: 3755.5942153930664 SSIM: 0.47852273923986727	18787_sat.jpg PSNR: 21.613539003802153 MSE: 1345.3885126113892 SSIM: 0.6438372541950352
19814_sat.jpg PSNR: 25.329004781043633 MSE: 571.8750343322754 SSIM: 0.5270071209364352	19814_sat.jpg PSNR: 30.952161076897056 MSE: 156.66987895965576 SSIM: 0.7435040797825035
22816_sat.jpg PSNR: 20.885275987820982 MSE: 1591.0140571594238 SSIM: 0.4892699743866129	22816_sat.jpg PSNR: 26.882423199238804 MSE: 399.9072666168213 SSIM: 0.7457541796358215

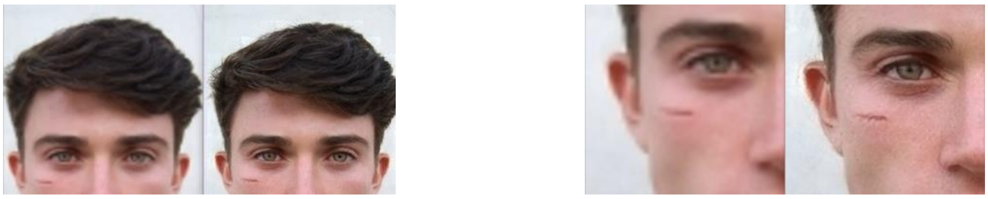

Figure 3. Low resolution vs enhanced

4. CONCLUSION

This project demonstrates the capability of GANs that can efficiently generate a high-resolution image from a lower-resolution image which offers a high pixel density that helps in identifying specific details about the original scene. The proposed system made use of GANs for image super-resolution. It is believed that this could be very beneficial to areas of surveillance specifically in closed-circuit television (CCTV) footage for facial enhancement of the footage and in satellite imagery where satellite images can be enhanced to mark out clearly the objects and boundaries in the image. Despite this, it is hoped that the work presented here is used for the benefit of humankind and is used to improve the quality of life and progress research in GAN systems. The system aims to improve the model to enhance more different kinds of images in different scenarios in the future.

Although the working system is efficient enough as per requirements, there are a few cases of usage that will lead to loopholes in the functioning. Such limitations arise from various factors including the complexity of the problem and the lack of resources to achieve them as well as time constraints. Some of the limitations have been highlighted: the model may not produce better results for images that are not relevant to the model, i.e., the satellite image feature use case requires only satellite images, and the facial image feature use case requires only facial images for enhancement of the picture. Models may take more time for images of higher resolution as the room for enhancement will be very little. However, it will still enhance the image to 4 times its earlier resolution. Improving the estimation accuracy for the high-resolution image will enable the generator to forecast the output more accurately for a specific type of input image. Training the model on a more precise, larger dataset with a greater number of epochs, the model could be scaled properly to other specific domains.

To overcome the limits of the project, the following will be taken up in the future. Some improvements can be made to the existing system. The model would keep enhancing the network's architecture to enable GAN to handle images with robust repeated structures. The model would be hosted on a standalone server, for the service of internet users. The model would be optimized by putting an external database so that it can have a sole administrator and users. A large number of training photographs with rich textures and excellent quality would be put for enhancing the model.

REFERENCES

[1] I. J. Goodfellow *et al.*, "Generative adversarial networks," *Prepr. arXiv.1406.2661*, Jun. 2014.

[2] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," *Prepr. arXiv.1609.04802*, Sep. 2016.

[3] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," *Prepr. arXiv.1809.00219*, Sep. 2018.

[4] Z. Li, Z.-S. Kuang, Z.-L. Zhu, H.-P. Wang, and X.-L. Shao, "Wavelet-based texture reformation network for image super-resolution," *IEEE Transactions on Image Processing*, vol. 31, pp. 2647–2660, 2022, doi: 10.1109/TIP.2022.3160072.

[5] G. Shim, J. Park, and I. S. Kweon, "Robust reference-based super-resolution with similarity-aware deformable convolution," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 8422–8431. doi: 10.1109/CVPR42600.2020.00845.

[6] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 11057–11066. doi: 10.1109/CVPR.2019.01132.

[7] M.-Y. Liu, X. Huang, J. Yu, T.-C. Wang, and A. Mallya, "Generative adversarial networks for image and video synthesis: Algorithms and applications," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 839–862, May 2021, doi: 10.1109/JPROC.2021.3049196.

[8] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 3862–3871. doi: 10.1109/CVPR.2019.00399.

[9] W. Wen, W. Ren, Y. Shi, Y. Nie, J. Zhang, and X. Cao, "Video super-resolution via a spatio-temporal alignment network," *IEEE Transactions on Image Processing*, vol. 31, pp. 1761–1773, 2022, doi: 10.1109/TIP.2022.3146625.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[11] J. Wang, Z. Shao, T. Lu, X. Huang, R. Zhang, and Y. Wang, "Unsupervised remoting sensing super-resolution via migration image prior," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2021, pp. 1–6. doi: 10.1109/ICME51207.2021.9428093.

[12] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, 2016, pp. 469–477.

[13] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep Residual channel attention networks," in *Computer Vision – ECCV 2018*, 2018, pp. 294–310. doi: 10.1007/978-3-030-01234-2_18.

[14] D. Huang and J. Chen, "MESR: Multistage enhancement network for image super-resolution," *IEEE Access*, vol. 10, pp. 54599–54612, 2022, doi: 10.1109/ACCESS.2022.3176605.

[15] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "A new deep generative network for unsupervised remote sensing single-image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 11, pp. 6792–6810, Nov. 2018, doi: 10.1109/TGRS.2018.2843525.

[16] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Computer Vision and Image Understanding*, vol. 158, pp. 1–16, May 2017, doi: 10.1016/j.cviu.2016.12.009.

[17] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013, doi: 10.1109/LSP.2012.2227726.

[18] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers*, 2003, pp. 9–12.

[19] R. Timofte *et al.*, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jul. 2017, pp. 1110–1121. doi: 10.1109/CVPRW.2017.149.

[20] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 1664–1673. doi: 10.1109/CVPR.2018.00179.

[21] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 2472–2481. doi: 10.1109/CVPR.2018.00262.

[22] Lin Zhang, Lei Zhang, Xuanqin Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011, doi: 10.1109/TIP.2011.2109730.

[23] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 4501–4510. doi: 10.1109/ICCV.2017.481.

[24] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 1646–1654. doi: 10.1109/CVPR.2016.182.

[25] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010, doi: 10.1109/TPAMI.2010.25.