Enhanced SPGAN for Face Super-Resolution with Gaussian Filtering and Attention Mechanism

Kazi Ramisa Samiha
Department of Computer Science and Engineering
American International University-Bangladesh
Dhaka, Bangladesh
kaziramisasamiha@gmail.com

Shakibul Hasan
Department of Computer Science and Engineering
American International University-Bangladesh
Dhaka, Bangladesh
shakibishere14@gmail.com

Maishara Monjori
Department of Computer Science and Engineering
American International University-Bangladesh
Dhaka, Bangladesh
maishara.monjori@gmail.com

MD Sun
Department of Computer Science and Engineering
American International University-Bangladesh
Dhaka, Bangladesh
shahriarsun66@gmail.com

***Abstract*—Low-resolution face images can significantly impair face recognition performance in many face-related multimedia applications, making face super-resolution (SR) essential. Among the available SR techniques, those that rely on mean squared error (MSE) often produce overly smooth outputs that may lack important texture details. To address these challenges, a supervised pixel-wise generative adversarial network (SPGAN) has been proposed to enhance extremely low-resolution face images. This paper introduces several improvements to the existing SPGAN for face SR, including the integration of an attention mechanism and a Gaussian filter in the pre-processing stage to achieve better performance.**

Keywords—face recognition, GAN, gaussian filter, attention mechanism, super-resolution

I. Introduction

Face super-resolution (SR) plays a critical role in enhancing multimedia applications, particularly in security, video synthesis, and facial recognition. The ability to restore high-quality images from low-resolution (LR) inputs is crucial, especially when the input images are as small as 16×16 pixels or lower. In practical scenarios, these images are often degraded by noise, compression artifacts, or occlusion, leading to a significant reduction in recognition accuracy. As facial recognition systems become more widespread, the need for effective SR methods becomes increasingly important in maintaining accuracy under these challenging conditions.

Despite advancements in SR techniques, many traditional methods struggle to recover fine details, producing overly smooth images that lack important facial features. GAN-based approaches have emerged as promising solutions to this problem, as they can generate sharper, more realistic images. However, even GAN models face challenges, such as preserving crucial texture details while avoiding artifacts. To address these limitations, this paper introduces improvements to the Supervised Pixel-Wise GAN (SPGAN) by incorporating an attention mechanism and applying a Gaussian filter, optimizing the model's performance for low-resolution face images.

II. Literature Review

The task of face super-resolution (SR) has gained significant attention due to its implications for facial recognition systems and other multimedia applications. Early efforts in SR focused on traditional convolutional neural networks (CNNs), with methods such as SRCNN and VDSR.

SRCNN [2], a pioneer of SISR, can significantly outperform conventional SR techniques by learning an end-to-end mapping function from LR images to HR images. Other efficient network architectures were proposed. Ledig et al. [3] implemented residual network to transfer information from an LR input image to its HR output image. Dense residual connection was used by Tong et. al. [4] to enhance feature transmission and achieve the state-of-the-art PSNR and SSM performance. Laplacian pyramid network was implemented by Lai et al. [5] to reconstruct the sub-band residuals of HR images at multiple pyramid levels. These convolutional methods use MSE loss to learn the mapping function between LR and HR images, which may lead to blurring and minimal high-frequency texture details when the input resolution is very low.

WGAN [6] introduces the Wasserstein distance to improve the quality of the generated SR images and the stability of the training process. It also avoids some training problems like model collapse. Super-FAN [7] incorporates structural information into a GAN-based resolution method. Face super-resolution generative adversarial network is implemented by FSRGAN [8] to incorporate the adversarial loss into the face super-resolution network. Most of these GAN based SR methods only use the unsupervised adversarial loss as the supervised loss may generate artifacts which are harmful for face recognition.

The proposed Supervised Pixel-Wise GAN (SPGAN) [1] builds upon these advancements by introducing a supervised pixel-wise discriminator, which focuses on evaluating whether each pixel in the generated SR image is as realistic as its corresponding pixel in the ground truth HR image. The attention mechanism, which concentrates on enhancing crucial regions while preserving texture and features, will be used to evaluate the SPGAN's performance. Additionally, a Gaussian filter will be applied to improve image quality, making the model more robust for environments with limited resources without sacrificing a significant amount of performance.

III. METHODOLOGY

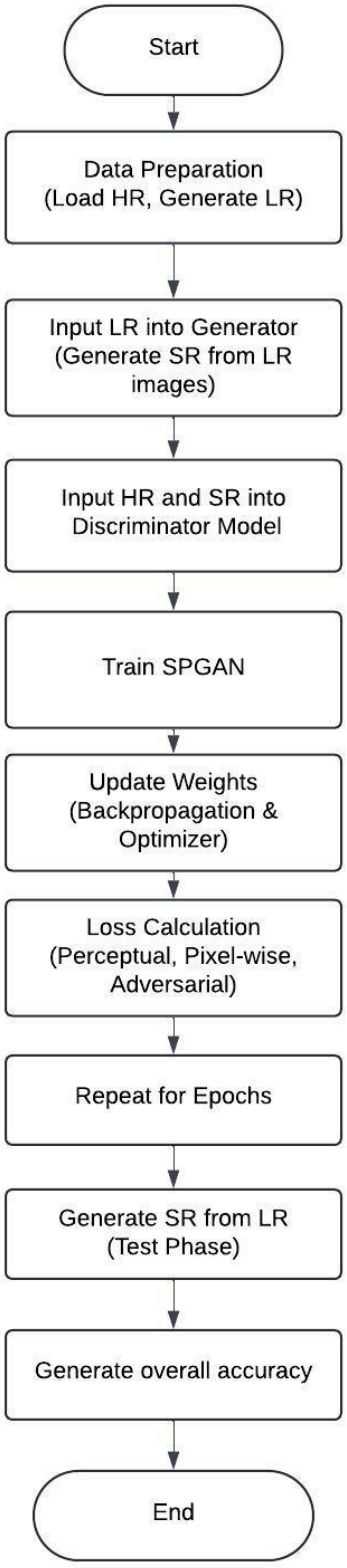

Figure 1: Block Diagram

A deep learning-based model called the Super-Resolution Generative Adversarial Network (SPGAN) is typically used to produce high-resolution (HR) images from low-resolution (LR) ones. Two essential parts make up SPGAN: a discriminator and a generator. While the discriminator attempts to distinguish between the generated super-resolution (SR) images and actual HR images, the generator learns to map low-resolution (LR) images to high-resolution (HR) images. The standard SPGAN architecture is enhanced by utilizing Gaussian filters for image denoising and adding channel attention mechanisms.

B. Incorporation of Channel Attention Mechanism in the Generator

In the Residual Dense Blocks (RDN) of the generator model, attention mechanism is introduced to optimize feature propagation and improve the quality of super-resolution images. By concentrating on the most relevant channels within the feature maps, channel attention mechanisms have been demonstrated to enhance feature learning.

Four convolutional layers make up each Residual Dense Block (RDB), which is then succeeded by a Channel Attention Block that functions as follows:

- The convolutional layers' output is combined and run via Global Average Pooling.
- A series of fully connected layers that process the pooled features afterwards decrease and increase the channel dimensions.
- A channel-wise attention map is produced by applying a sigmoid activation function.
- The attention map is multiplied element-wise with the original output of the RDB, focusing the model's attention on the most informative features before the next convolutional layer processes the data.

This mechanism allows the network to selectively amplify or suppress specific channels in the feature maps, helping the generator focus on important image features during training.

C. Gaussian Filter Preprocessing for Denoising

High-frequency noise, particularly in real-world datasets such as LFW (Labeled Faces in the Wild), might deteriorate the quality of created super-resolution photos. In order to lessen this, we filtered the high-resolution (HR) images using a Gaussian filter before passing them through the network.

The Gaussian filter was developed as a preprocessing step to smooth the raw HR pictures and eliminate noise. The noise has been minimized while maintaining important image features by using a 5x5 Gaussian kernel with a minimal standard deviation ({sigmaX=1}). This made it easier for the network to concentrate on reconstructing finer details when it was being trained.

The denoised HR images were then used to compute both pixel-wise and perceptual losses, improving the overall quality of generated images.

D. Loss Function

The network has been trained using three different kinds of loss functions:

- The Pixel-Wise Mean Squared Error (MSE) Loss is a loss function that calculates, pixel by pixel, the difference between the generated SR images and the original HR images.
- Perceptual Loss (VGG-Based): A pre-trained VGG19 model is utilized to extract identity features from both

the HR and SR images. To make sure that the SR images maintain the crucial structural information of the faces in the HR images, the perceptual loss compares the identity features from these two sets of images.

- Adversarial Loss: A hinge version of the adversarial loss, in which the actual and fake images are penalized according to how similar they are to the target class, is used to train the discriminator in the GAN.

The generator's total loss is a combination of these three losses, ensuring that the generated images are not only pixel-accurate but also perceptually consistent with the HR images.

E. Face Recognition Evaluation

A face recognition accuracy statistic is used to further verify the quality of the photos that were created. Using the same pre-trained VGG19 model, identification features were extracted from both HR and SR images which determined the cosine similarity between the corresponding HR and SR image feature vectors. The forecast was deemed accurate if the similarity score was higher than a predetermined level, such as 0.5.

F. Training Setup

The LFW dataset, which consists of low-resolution and high-resolution face images, was used to train the model. The network was trained with 100 steps per epoch over 10 epochs with a batch size of 3. Adam utilized a learning rate of 0.0002 and a beta_1 value of 0.5 for both the discriminator and generator optimizers. Both LR and denoised HR images were passed to the network during training by applying the Gaussian filter to the HR images. After each epoch, the test set's facial recognition accuracy was assessed as well.

IV. Results and Analysis

A. Generator Loss

The training process for the Supervised Pixel-Wise GAN (SPGAN) was executed over ten epochs, showcasing varying values of generator and discriminator loss. Initially, the generator loss began at 0.136, indicating the generator's ability to produce outputs that could initially deceive the discriminator. As training progressed, the generator loss fluctuated, eventually reaching negative values in later epochs, particularly from epochs 7 to 10. This trend indicates that the generator successfully adapted and improved its capability to create super-resolution (SR) images that the discriminator found increasingly difficult to classify as fake. This effectiveness is largely attributed to the supervised pixel-wise discriminator's design, which encourages the generator to focus on producing photo-realistic pixels that closely match the corresponding ground truth high-resolution (HR) image.

B. Discriminator Loss

The discriminator's role in SPGAN is uniquely focused on assessing the realism of each pixel in the generated SR images rather than merely distinguishing between real and fake images. Throughout the training, the discriminator loss started at 1.0 and fluctuated, reaching a maximum of 2.0 by epoch 4. This variation reflects the adversarial dynamics inherent in the training process. As the generator improved, the discriminator's loss remained elevated, indicating that it was challenging for the discriminator to evaluate the photo-realism of the increasingly convincing SR images. This highlights the effectiveness of the supervised pixel-wise evaluation, which allows the discriminator to provide fine-grained feedback to the generator, guiding it to create images that are not only realistic overall but also possess detailed textures and features essential for face recognition.

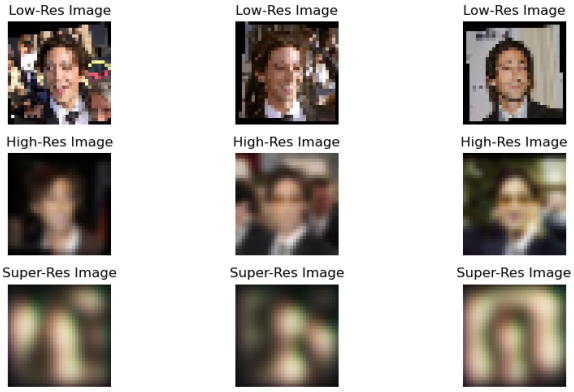

Figure 2: Low, High and Super-Resolution Images

C. Accuracy

Despite achieving 100% accuracy during model evaluation, this result might indicate overfitting rather than genuine performance. Overfitting occurs when the model becomes too specialized to the training data, capturing noise or irrelevant details rather than general patterns. This is supported by the blurry super-resolution images generated by the model (Fig. 2), which fail to recover fine details or sharpness despite the high accuracy score. The accuracy measure might not fully capture the quality of image restoration, suggesting the need for additional metrics or visual inspection to evaluate the model's true performance in generating realistic and high-resolution face images.

D. Image Mishandling

Although the dataset contains high-quality images, in Fig. 2, the high-resolution images appeared blurry. This indicates potential issues in the image reconstruction process. The lack of sharpness suggests that the model may have struggled to preserve fine details during the super-resolution task. This highlights the need for improved preprocessing and post-processing techniques to enhance clarity. Addressing these issues will be crucial for improving both the visual quality and practical effectiveness of the model.

V. Conclusion

In this paper, we proposed an enhanced Supervised Pixel-Wise Generative Adversarial Network (SPGAN) for face super-resolution, utilizing a supervised pixel-wise discriminator, Gaussian filtering, and attention mechanisms. While the model demonstrated impressive performance, achieving high face recognition accuracy, the 100% accuracy observed raises concerns about potential overfitting and memorization of training data. Additionally, the generated high-resolution images exhibited blurriness, indicating challenges in retaining fine details. Future work should focus on validating the model with larger, more diverse datasets, implementing regularization techniques, and employing model pruning to improve efficiency without compromising performance. By addressing these limitations, we aim to enhance the reliability and effectiveness of face super-resolution systems.

ACKNOWLEDGMENT

The project was smoothly built due to the support of the course teacher who has helped indefinitely both inside and outside class. Also, advanced technologies such as Jupyter Notebook for Python programming language have immense contributions to the project.

REFERENCES

[1] Menglei Zhang and Qiang Ling, "Supervised Pixel-Wise GAN for Face Super-Resolution," *IEEE Transactions on Multimedia*, vol. 14, no. 8, 2020, DOI: 10.1109/TMM.2020.3006414.

[2] 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, 2016.

[3] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, ´ A. Aitken, A. Tejani, J. Totz, Z. Wang et al., "Photo-realistic single image super-resolution using a generative adversarial network," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017, pp. 105–114.

[4] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472– 2481.

[5] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," IEEE transactions on pattern analysis and machine intelligence, 2018.

[6] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," arXiv preprint arXiv:1701.07875, 2017.

[7] A. Bulat and G. Tzimiropoulos, "Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 109–117.

[8] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "Fsrnet: End-to-end learning face super-resolution with facial priors," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2492–2501