Revolutionizing Image Quality: An In-depth Investigation into Super-Resolution via Generative Adversarial Networks

1st Maidul Islam
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
maidul.islam@g.bracu.ac.bd

2nd Farjana Alam
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
farjana.alam@g.bracu.ac.bd

3rd Md Sabbir Hossain
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
md.sabbir.hossain1@g.bracu.ac.bd

4th Farah Binta Haque
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
farah.binta.haque@g.bracu.ac.bd

5th Annajiat Alim Rasel
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
annajiat@gmail.com

***Abstract*—Low-resolution (LR) images are often blurry and lack detail, limiting their usefulness in various applications. This paper investigates the potential of Generative Adversarial Networks (GANs) for super-resolution, specifically focusing on the effectiveness of the VGG19 model as a perceptual loss function. We propose a novel GAN architecture utilizing VGG19 features to guide the generator towards producing high-resolution (HR) images that are not only faithful to the original content but also visually appealing. The model includes A powerful generator with residual blocks and upsampling layers to enhance the resolution and refine details and a discriminative network based on the VGG19 architecture to distinguish between real and generated HR images, further improving the perceptual quality. This study includes combined training with both adversarial loss and content loss derived from VGG19 features, ensuring both high fidelity and visual realism.**

***Index Terms*—Super-resolution, Generative Adversarial Networks, VGG19, Image quality, GAN architecture**

I. Introduction

Low-resolution (LR) images pose a significant challenge in various fields, limiting their utility and hindering our ability to extract valuable information. Traditional interpolation methods often struggle to overcome these limitations, introducing artifacts and failing to accurately reconstruct details. This paper presents a novel data-driven approach to super-resolution using Generative Adversarial Networks (GANs) guided by the perceptual loss of a pre-trained VGG19 model. We propose a robust GAN architecture featuring a detail-enhancing generator and a VGG19-powered discriminator for discerning real from generated high-resolution (HR) images. Our framework leverages the perceptual expertise of VGG19 to ensure not only fidelity but also visual appeal in the reconstructed images. Through extensive experiments on a large dataset, we demonstrate significant improvements in sharpness, content accuracy, and visual quality compared to existing methods. This data-driven approach paves the way for further advancements in image super-resolution, unlocking the potential of countless LR images across various domains.

II. Literature review

In order to improve the super-resolution of medical images, this research [1] offers the Deep Residual Feature Distillation Channel Attention Network (DRFDCAN). The OASIS, BraTS, ACDC, and COVID medical image datasets were the four used by the authors to assess DRFDCAN's performance. The study's datasets include: The ACDC dataset contains cardiac MR pictures, the COVID dataset presents chest CT scans, the OASIS dataset offers single-modality brain MR images, and the BraTS dataset offers multimodal brain MR images.The suggested model enhances computational effectiveness and image quality through the employment of a channel attention mechanism and a residual-within-residual design. Two measures were used by the authors to assess the effectiveness of DRFDCAN: the structural similarity index (SSIM) and the peak signal-to-noise ratio (PSNR). The suggested model achieved the maximum PSNR of 29.51 dB and an SSIM of 91.56%, outperforming other cutting-edge models in terms of image quality.

The study [2] offers a novel method for super-resolution of multimodal medical images utilizing a multi-attention network and wavelet transform. For training and transfer learning, the authors used the high-resolution DIV2K dataset, and then assessed their approach using a variety of Shenzhen Hospital datasets. The rebuilt images have better texture details and edge sharpness because they trained a perceptual loss function using generative adversarial networks (GANs). Improved texture information and more precise reconstruction were made possible by the incorporation of wavelet transformations. The outcomes showed the possibility for better medical picture interpretation and diagnosis, with greater performance in terms of visual, quantitative, adversarial, and perceptual loss.

The study [3] presents MISNet, an ultralight-weight super-resolution network that makes use of residual framework feature interpolation at the multi-resolution level. When tested on benchmark datasets, such as Urban100, the network performs better than current cutting-edge networks. MISNet improves image quality by combining residual and interpolated features, which makes it appropriate for low-power and storage-capable applications. When compared to existing ultralight-weight networks, the suggested network performs better at retrieving high-frequency features and edges, especially in the Urban100 dataset.

This research [4] presents a Deep Spiking Neural Network (SNN) based Progressive Tandem Learning (PTL) framework for pattern identification. Using the MNIST, Cifar-10, and ImageNet-12 datasets, the authors validate their method and obtain competitive classification accuracy. The technique uses an adaptive training scheduler for layer-wise learning and spike counts for network conversion. The results showcase the efficacy and efficiency of the proposed PTL framework for SNN-based pattern recognition, showing state-of-the-art performance in tasks such as object identification, time-domain voice separation, and image reconstruction.

Islam and Zhang present a GAN-based method in this study [5] for creating artificial brain PET images for the diagnosis of Alzheimer's disease. For their study, they made use of the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. With a high degree of closeness to actual brain PET images, the suggested model was able to produce brain PET images for three distinct stages of Alzheimer's disease. Using axial, coronal, and sagittal slices from the produced PET data, the scientists additionally created a 2D-CNN model, which they used to achieve 71.45% classification accuracy for CN/AD classification. The study shows how creating synthetic medical images can be a low-cost method of creating automated diagnostic equipment.

III. Methodology

A. Data collection and Preprocessing

The dataset was collected from kaggle which consists of human face images. As we had limited hardware resources, we tried to implement the Super Resolution GAN on only human faces rather than focusing on other classes. The dataset contained 3500 high resolution along with 3500 low resolution images paired with each oher in the training dataset and 1000 high resolution along with 1000 low resolution images paired with each oher.

B. Implemented Algorithms

Existing GAN-based SR approaches primarily focus on minimizing adversarial loss, which can lead to visually appealing but content-inaccurate reconstructions. Our proposed architecture addresses this limitation by incorporating the perceptual expertise of VGG19. The network utilizes a series of residual and upsampling blocks to progressively refine the features and increase the resolution of the LR image, ultimately generating an HR counterpart. VGG19-powered discriminator leverages a pre-trained VGG19 model to extract perceptual features from HR images. The discriminator then utilizes these features to distinguish between real and generated HR images, providing feedback to the generator for improvement. Weighted loss function is a combined loss function which guides the training process. This loss includes an adversarial loss term that encourages the generator to produce realistic HR images, and a VGG19-based perceptual loss term that ensures the generated images retain the essential visual characteristics of the original content. The balance between these two terms can be adjusted through weights to prioritize either perceptual quality or content fidelity. The proposed VGG19-guided GAN architecture for super-resolution consists of three key components: a detail-enhancing generator, a VGG19-powered discriminator, and a combined loss function.

The VGG19-guided GAN architecture proposed for super-resolution comprises three pivotal components: a detail-enhancing generator, a VGG19-powered discriminator, and a unified loss function. The detail-enhancing generator begins with an input layer accepting low-resolution images, followed by residual blocks containing convolutional layers, batch normalization, and PReLU activation. Upsampling blocks further enhance resolution through convolutional and UpSampling2D layers, ultimately generating a high-resolution output matching the ground truth image. The VGG19-powered discriminator, designed for discriminating high-resolution images, utilizes convolutional blocks with increasing complexity, including

convolutional layers, batch normalization, LeakyReLU activation, and dense layers. The combined loss function integrates adversarial and perceptual losses, leveraging binary cross-entropy for adversarial alignment and mean squared error for VGG19-based perceptual alignment. Weighted combinations of these losses strike a balance between perceptual quality and content fidelity, elucidating how the architecture achieves superior super-resolution results.

REFERENCES

[1] Umirzakova, S., Mardieva, S., Muksimova, S., Ahmad, S., & Whangbo, T. (2023). Enhancing the Super-Resolution of Medical Images: Introducing the Deep Residual Feature Distillation Channel Attention Network for Optimized Performance and Efficiency. Bioengineering, 10(11), 1332.

[2] Dharejo, F. A., Zawish, M., Deeba, F., Zhou, Y., Dev, K., Khowaja, S. A., & Qureshi, N. M. F. (2022). Multimodal-boost: Multimodal medical image super-resolution using multi-attention network with wavelet transform. IEEE/ACM Transactions on Computational Biology and Bioinformatics.

[3] Esmaeilzehi, A., Ahmad, M. O., & Swamy, M. N. S. (2021, May). MISNet: Multi-Resolution Level Feature Interpolating Ultralight-Weight Residual Image Super Resolution Network. In 2021 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1-5). IEEE.

[4] Wu, J., Xu, C., Han, X., Zhou, D., Zhang, M., Li, H., & Tan, K. C. (2021). Progressive tandem learning for pattern recognition with deep spiking neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7824-7840.

[5] Islam, J., & Zhang, Y. (2020). GAN-based synthetic brain PET image generation. Brain informatics, 7, 1-12.