J. Electrical Systems 20-10s (2024): 7921-7936

Aseel al'naeimee[1],*

Amir Lakizadeh[1]

A Method for Face Hallucination Based on Generative Adversarial Network by Extending ESRGAN Architecture

Absrtact : Deep learning has emerged as a pivotal approach across many scientific and industrial applications, primarily due to rapid advancements in computational power. Face hallucination, defined as enhancing the resolution of facial images, plays a significant role in computer vision applications such as facial recognition, facial feature analysis, and human identity analysis. Recently, deep generative models like Generative Adversarial Networks (GANs) have been extensively used for this purpose. However, further improvement in the precision and quality of the results is necessary. We present a novel GAN-based approach for face hallucination by expanding the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) architecture. This research introduces a customized version of ESRGAN using the pre-trained version of VGG16 architecture but in a reduced size to achieve a good trade-off between the accuracy of the final output image and the complexity requirements. It has been demonstrated during the experiments that the suggested changes enhance the face hallucination accuracy, reaching a training data classification accuracy with PSNR: 30.30, SSIM: 0.8757 and LPIPS: 0. 0817. This performance exceeds state-of-the-art approaches, emphasizing the significance of the modifications undertaken.

Keywords: low-resolution (LR), face hallucination, ESRGAN architecture, super-resolution (SR)

1. INTRODUCTION

Facial hallucination (FH) developed a set of methodologies like Face Super-resolution generic face images from low-resolution input images even when inferring height variants like age or gender. It was developed to enhance the resolution of facial images [1-3]. However, it can be derived that FH activates several tasks like face detection, face recognition and face identification; it can be claimed that one of the reasons for the deemed of great importance is the where image quality is low (this quality comes from due to practical terms such as security cameras). These functions also cut across complex interpolation, statistical information, and estimation or input-output models. The image-based approach enables effective mapping between the input and target images of a location within spatial coordinates in an even manner. Generation adversarial networks (GNA) can cover such cash by attributing shooting and science ideological perception where advanced formats prevail. Pixel-wise loss has always been more scintillating in achieving its primary goal since, over the years, they have been harnessing the potential locked up in GANs – getting and rendering high-resolution images in the form of Convolutional Neural Networks (CNNs) resources [3–4].

Strides, GANs primarily focus on generating higher resolution images alongside the aim of identifying similar features present in facial structures; thus, upsampling/diverging 4x-8x seems effective yet proves to be a venture to escape from due to adding in unnecessary/false facial elements into the imagery formed, by doing this one particularly places themselves in an abnormal state [5]. However, by employing various algorithms, the upsampling can assist with sharpening edges present or prominent within the image, thereby recognizably changing the pixels of the image into something more or less different, in simple terms, modifying the image.

[1]Computer Engineering Department, University of Qom, Qom, Iran
*Corresponding author

Increasing image quality and eliminating defects is a great challenge in medicine and the digital arts. Alongside the intricacies of facial reconstruction, face hallucination is particularly problematic because enriching facial photos is next to impossible due to the complexity of the face geometry [8].

Blurred images represent a great problem in enhancing the quality of face photos [9], for distortions render the task of information extraction and understanding severely compromised. Therefore, face hallucination represents a scientific challenge in image processing that calls for original and creative ideas [10].

The natural and artificial evolution of image enhancement has raised several algorithms and models that help solve, in particular, the lack of detailed input data or very low-resolution images due to the extension of deep neural network usage. One such network is ESRGAN, which is used to enhance the sharpness and details of the picture.

Recent studies showed that ESRGAN should be developed further as many issues need addressing, such as producing consistently high-quality images. The key aspect of the problem involves correctly interpreting images of peoples' faces during hallucinations; the understanding of the images containing facial features is completely lost during the hallucination. Some researchers have suggested extra layers to be put on top of the ESRGAN algorithm to help improve the clarity of the facial images during hallucinations. With these additional layers, the algorithm was expected to restore the detailed features of the picture and correct the hallucination concerning facial features. Numerous studies have been conducted to address hallucinations, initially defined as generating high-resolution images from low-resolution counterparts.

This technology is critical for security, facial recognition, and enhancement tasks [12–14]. Various deep learning-based techniques for FH have been utilized, including autoencoders, deep convolutional neural networks, and deconvolutional networks. Conventional deep neural networks have been employed to generate super-resolution images, aiming to minimize mean square error and maximize PSNR between low-resolution or high-resolution images and the resulting super-resolution image. The resulting image often exhibits blurred effects and unrealistic textures. This network incorporates multiple enhancements, such as implementing residual blocks [15], applying the Laplacian pyramid [16], using the residual dense network [17], and adopting recursive learning [18–19].

Cascade architectures have been explored to enhance super-resolution performance and achieve picture fidelity with increased precision. In [1, 21], we examined cascade techniques extensively. These structures operated independently or in conjunction with face component heatmaps and attribute data [1,2,22]. We developed a perceptual loss function to augment the super-resolution visual effect by reducing the perceived resemblance between source high-resolution (HR) and low-resolution (LR) images, thus producing super-resolution (SR) images [1, 3].

The primary objective of this study is to improve the perceptual quality of SR. In this phase, we outline and analyze the proposed network architecture enhancements addressing discriminative and perceptual loss.

2. RELATED WORKS

The study of face hallucinations using a better ESRGAN architecture looks at how picture super-resolution and deep learning techniques have improved. Researchers have explored diverse methodologies to augment facial features and elevate perceived quality. In their study, Hernandez, Rowell M., et al. [23] pioneered Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) for face hallucination, producing significant advancements in facial image improvement. Wang, Xintao, et al. [24] enhanced the ESRGAN framework, improving its capacity to reconstruct high-resolution face features accurately. Building on previous research, Li and Mingda [25] developed innovative methods to enhance the ESRGAN model's ability to recover facial details, pushing the boundaries of face hallucination. This section presents an overview of the pertinent literature.

2.1 Enhancements of Network Architecture

In conclusion, each article, in one way or another, succeeds in image zooming and aspects in its way to conquer approaches on how to create more realistic images. Sun et al. [26] proposed a new approach based on the sparse coding inside a cascaded end-to-end framework model. In another study [27], authors presented an advanced

ProGanSR system incorporating a GAN model application to enhance the realistic representation of images. The researchers proved that the gan models improved the reconstruction quality, especially when the upsampling factor is high, such as 8 fold. They proposed a pyramid-asymmetric structure model with custom-designed layers to reduce the negative impact caused by the increased data passing through the layers on the computational cost, focusing on the lowest levels.

Lai, Zhibing et al. [28] adopted Laplacian ideas in parallel developments of pyramid architectures. This facilitated the effort of progressively training and forecasting at multiple levels, hence improving output quality. To fix problems with GAN-generated textures, Liu et al. [29] created a structured method with three main parts: the reconstruction module, the structure model, and the photorealism model. This framework sought to reduce the production of artificial textures in GAN-generated pictures, improving their realism.

In their further investigation of GAN networks, Kim et al. [30] observed difficulties generating realistic face shapes, especially with increasing resolution. Their methodology utilised facial landmarks and progressive training procedures, enhanced by facial attention loss values and improvements in facial alignment networks, to enhance facial detail reconstruction.

2.2 Enhancements of the Loss Function

Iqbal et al. [31] employed a diminutive thumbnail to generate the super-resolution FH. The input pictures were too tiny, measuring 16x16. We subjected them to a multi-stage progressive upsampling and inpainting GAN, which involved an eightfold increase in scale. Similar to earlier studies, this technique relies on a progressive GAN strategy. We employed multiple network architectures for various goals. We employed the Pro-UInet to generate heat maps of face landmarks. We employed the Local-D and Global-D modules to generate authentic face landmarks.

Liu et al. [32] employed recurrent conventional neural networks for upsampling. The results fluctuated due to the input provided to the network, attributable to the vanishing gradient inherent in recurrent networks.

Table (1) comprehensively compares the network and the suggested enhancements to the loss function. We carefully categorize each enhancement into two primary sections: "Enhancements of Network Architecture" and "Enhancements of Loss Function." Each enhancement specifies the methodology, planned contributions, and impacts, providing a comprehensive overview of the progress in network design and loss function techniques.

Table 1: Comparative Summary of Techniques, Contributions, and Effects in Relevant Studies

Model	Method	Contribution	Effect
VDSR [33]		Augments the number of convolutional network layers to 20	Substantially enhances PSNR value
EDSR [34]	Enhancements in Network Architecture	Employs residual blocks and eliminates superfluous batch normalization layers	Facilitates network convergence and enhances performance
RRDB [35]		Incorporates residual structure derived from density	Significantly enhances the network's reasoning capability.
Adder SR [36]		Develops an adder filter to	Minimises computational complexity while maintaining network performance
SRGAN [37]		Integrates perceptual loss, adversarial loss, and content loss into the model.	Minimises computational complexity while maintaining network performance
ESRGAN		Employs the attributes prior to	Possesses the capability to produce

[35]	Enhancements of Loss Function	activation to formulate the perceptual loss	intricate texture details of the image, significantly enhancing visual quality.
SROBB [38]		Develops a specific perceptual loss utilising object, backdrop, and border labels.	Enhances the visual quality of photographs and reconstructs intricate textures.

3. PROPOSED METHOD

The Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) follows the basic ResNet architecture of the Super Resolution Generative Adversarial Networks (SRGAN), as shown in Section 1, and is described in the literature. It replaces the residual block with a Residual-in-Residual Dense Block (RRDB). The DenseNet design serves as the basis for the RRDB block. It establishes direct connections among all levels inside the residual block.

The SRGAN employs a convolutional neural network as its generative network. It enhances its efficacy by incorporating a discriminative network during training to generate detailed images and achieve superior visual results. On the other hand, the residual network (ResNet) can be described as a deep learning model in computer vision. This Convolutional Neural Network (CNN) model was designed and assembled to have several hundreds or thousands of convolutional layers. Fusing multiple levels of residual networks with dense connections, the r esidual in residual density block (RRDB). It is also based on the assumption that more structural depth will improve performance and that the suggested RRDB is a more complicated and deeper structure than the previous residual block.

As a result, the ESRGAN architecture also looks familiar to SRGAN, though there have been changes. The multi-level residual network with dense connections not incorporating batch normalization is encoded in a dense residual block termed RRDB. It was advocated that convolutional layers be incorporated into the basic block of the RRDB approach to make these contributions.

Figure (1) serves as a schematic representation of the architecture of the proposed pre-trained neural network model. The diagram depicts the model's architecture, showcasing the layers, connections, and pre-trained components utilized. This layout guides understanding of the model's design, indicating the data flow and integration points for pre-trained parameters that contribute to the model’s performance and efficiency.

Convolutional layers execute convolution procedures, which mathematically multiply the pixel spectra of high-quality images to enhance the clarity among the components of the learnt picture data.

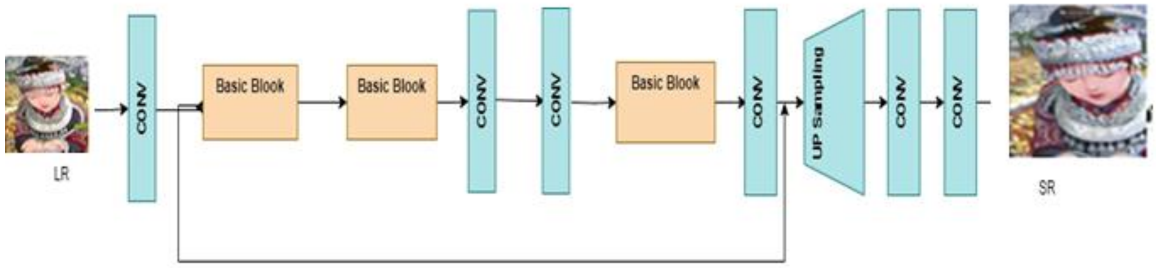

Figure 1: The proposed model uses some pre-trained neural networks

The network has several characteristics, indicating that increased features correlate with enhanced prediction outcomes. This network consists of two primary components: the generator and the discriminator networks. These networks must be trained to get optimal performance in optimisation.

3.1 loss functions

Instead of pixel-wise discrepancies, a pre-trained neural network calculates perceptual loss by contrasting high-level feature representations. VGG 16—characterized by its lightweight architecture and inverted residuals—is very effective for these tasks. By taking feature maps from the middle layers, perceptual loss measures the difference between produced and reference pictures at several levels of abstraction. This loss can be expressed as:

$$L_{perceptual} = \sum_{i=1}^{N} \lambda i \cdot \|\phi i(I_{SR}) - \phi i(I_{HR})\|_2^2 \qquad \ldots\ldots\ldots\ldots(1)$$

Where $\phi i(I)$ represents the feature map from the i-th layer and I_{SR} and I_{HR} IHR are the super-resolution and high-resolution images, respectively [39].

3.2 Dataset

For this study, we chose the DIV2K dataset [40], which contains 1000 images. The standard approach divides the database into three subsets: the training set, comprising 800 photographs; the validation set, containing 100 images; and the test set, containing 100 images. During the training phase, we use the training set to train our model and the validation set to check our results. When testing the SR, the test set measures how well it works. Quantitative super-resolution measures include industry-standard picture quality metrics, such as PSNR, SSIM, and LPIPS. The initial phase of picture preparation entails reading datasets. We divide the dataset into two segments.

Each training and testing dataset consists of two files. One file has low-resolution photos, while the other provides high-resolution photographs. We partition the low-resolution photos into patches and upscale them to create the high-resolution model. We then catalogue these photos in lists with their corresponding labels. The datasets must possess identical names during training, adhering to a designated naming convention.

Each patch must remain uncomplicated and unflattened to facilitate thresholding. Obtain arbitrary coordinates for the top left corner in LR space and multiply by the scaling factor. The subsequent objective is to get HR coordinates. This process entails determining the batch size and including potential locations for it. We only accept the batch if the standard deviation of pixel intensities surpasses a specified threshold or if we have already rejected n patches. We execute the square crops using the specified patch size derived from the chosen top left corners.

A transformation may be used to invert the input picture based on the selected transformation; these patches will be stored in a designated route for testing and validation datasets after that.

3.3 Training Process

Establishing and executing the training falls under a specialized class object. This class requires a generator capable of producing SR pictures for input. It may optionally incorporate a discriminator network and a feature extractor to formulate the elements of the perceptual loss.

When a discriminator is included, the model(s) and training procedures are GAN-like when we include a discriminator. Alternatively, the training adheres to a conventional ISR methodology. This stage includes numerous essential components, such as the generator, a Kera model with super-scaling capability, or a generating network. Furthermore, the discriminator, another Keras model, functions as the adversarial element of the perceived loss. The third crucial element is the feature extractor, a Keras model engineered to discern deep characteristics and enhance the perceptual loss function. It needs training and validation datasets, designated learning rates, and loss weights as inputs. This component encompasses the 'generator' for the generator loss component and may incorporate the 'discriminator' and 'feature extractor'. Many consecutive networks require post-combination synchronization, particularly for their weights. A specific building method for the integrated model included the generator network, discriminator, and gesture extractor.

4. RESULTS

4.1 Prediction Process

Upon completion of network training, the weights remain unchanged; these weights are preserved for utilisation with a predictor. The predictor class manages predictions based on an input model. Upon loading the photos in the input directory, the training of the specified model is conducted, and the results are stored in the output directory. It can accept a path for the weights or let the user navigate the weights directory to select the necessary weights. Subsequently, input the validation path or a designated picture for its reconstruction.

This research employs a model with a standard layer for input and a ReLU activation function. The primary function of this layer is to extract initial key characteristics and augment the incoming image. The subsequent component employed was the B residual blocks, including three structures of three primary blocks: convolution layer, Batch Normalisation, and ReLU. Each block is designated as an input for the subsequent block.

Subsequently, data is sent to the up-sampling layer with a sample value of two. The up-sampling occurs in two stages, each at a factor of 2x, resulting in a cumulative up-sampling of 4x for the entire system. The discriminator stage comprises an input conventional layer, followed by two sets of three layers each, which include convolution, activation, and batch normalisation. Subsequently, there are four phases of convolution and batch normalisation. This stage concludes with a flattened layer and a thick layer. Figure (2) delineates the structures employed in the suggested methodology.

4.2 Effect of Parallel Computing on the Proposed Method

Implementing a parallel structure in the discriminator stage is challenging since it consists of a series of sequential phases where each step is contingent upon the preceding one. Three RDS blocks in the generator enabled each stage to operate on a single core utilising threads. The training duration for the dataset improved and increased by around 30%. Figure (2) illustrates the data stream and the upsampling process. The diagram visually represents how data flows through the system and highlights the steps in increasing the sample rate. It emphasizes transforming the input data into a higher-resolution output, making it easier to analyze and interpret. Key components of the process are clearly labeled, allowing for a better understanding of how upsampling enhances data quality. The standard system needed 4 hours and 34 minutes, whereas the parallel structure took 3 and 22 minutes. The training assignment was conducted on a core-i7 CPU (3.4 GHz), supplemented by a GPU (4G) and a maximum memory capacity of 16GB.

4.3 Performance Metrics

Metrics such as PSNR, SSIM, and LPIPS are widely used to evaluate the quality of super-resolution images. PSNR is a traditional metric that measures the pixel-level similarity between the generated image and the ground truth, focusing on the overall error but often failing to capture perceptual quality [41]

$$PSNR = 10\ log_{10}(\frac{(L-1)^2}{MSE}) \quad \text{.......} \quad (2)$$

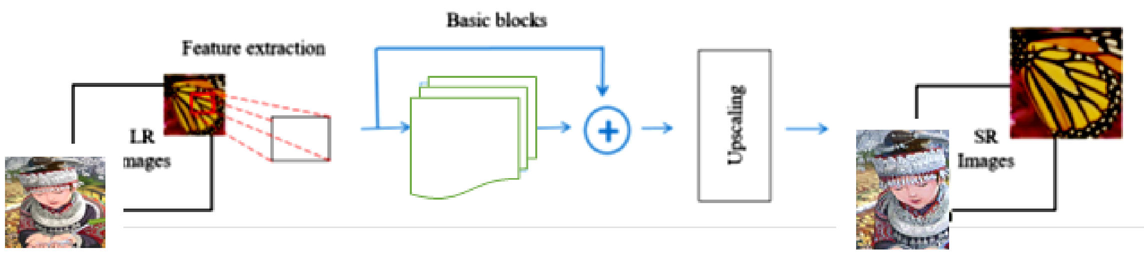

Figure 2: Diagram of the data stream and the upsampling process.

Where L represents the maximum intensity of the picture, often 255 due to using 8-bit images, SSIM improves upon PSNR by considering luminance, contrast, and structural information, offering a better understanding of image degradation in human visual perception [41]. However, both metrics can struggle with perceptual similarity.

$MSE = \frac{1}{MN}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(o(i,j) - D(i,j))^2$ ……………………………………………(3)

Where O represent the restored picture, D denotes the super-resolution image, and m and n signify the width and height of each image, respectively. Consequently, we can ascertain that an increased PSNR correlates with superior quality in compressed or reconstructed images.

LPIPS is a newer metric that addresses this limitation by comparing the feature representations from a pre-trained network, such as AlexNet or VGG, making it more aligned with the human judgment of visual quality [42]. These metrics complement one another in assessing image quality across different dimensions.

LPIPS $(x, y) = \sum_{l} wl \cdot \|\phi lx) - \phi i(y) \|_2^2$ ………………………………………………… (4)

Where x and y are the original and generated images, respectively, ϕl represents the deep network feature activations at layer *l*, w*l* is a learned weight for each layer *l* , $\|\cdot\|_2^2$ is the squared Euclidean distance between the feature representations of the two images at layer *l*.

4.4 Experiments

A deep learning model as advanced as ESRGAN is designed by [24]. In the field of image, the role of ESRGAN is quite significant in the area of super-resolution. The aim of this model is to upscale images of low resolution by focusing on the key elements and the textures, which end up adding more to the enlargement. Applying advanced concepts from deep learning and GANs together with perceptual loss functions enables ESRGAN to produce convincing outputs which include finer image details.

This is useful in many areas, including improving images such as photography, medicine, satellites, and surveillance cameras. The ability of the ESRGAN to enhance the quality of images by increasing their resolution has made it a useful tool for image enhancement in many situations.

The generated image turned out impeccably accurate with fewer prediction value errors. This followed the recommended framework and the outcomes of the Picture Ultra High-Resolution Adam Becker system. Tables (2) and Figure (2) present the analysis results of the various super-resolution techniques based on PSNR, SSIM, and LPIPS metrics. In this table, a comparison is made between the present results and studies conducted in the past. It indicates the relative performance of each technique. This evaluation addresses the overall picture of the ability of super-resolution methods to improve the quality of images as progression is made across these well-known and widely used metrics.

Table 2: Comparative Analysis of Super-resolution Techniques Using PSNR, SSIM, and LPIPS Metrics Limiting findings of us to considering previous related works only.

Method	GAFH-RIDN [17]	VDSR [43]	SRCNN[44]	SelfEx [45]	RFL [22]	A+ [46]	Ours
PSNR	24.28	25.18	24.52	26.44	25.86	26.03	**26.83**
SSIM	0.71	0.75	0.722	0.80	0.79	79.73	**0.8108**
LPIPS	_	_	_	_	_	_	**30.30**

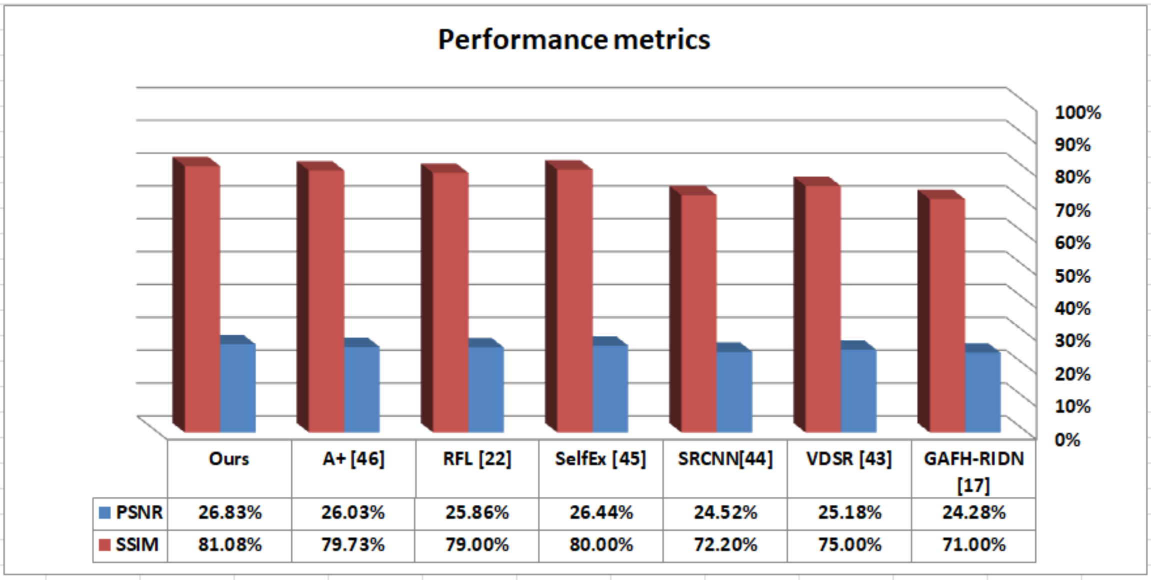

Figure 3: The comparison of results mentioned in the proposed method with the others.

As shown in Figures 4-6, the model's performance has been evaluated over the training period of 100 epochs on three selected image quality measures alongside noise ratio. Each Figure corresponds to a specific loss value: Figure 4 shows the average loss of PSNR, which measures how much input signal has been preserved. Figure 5 displays the average loss of SSIM where requirements on the image structure are of interest. Figure 6 illustrates the average loss of LPIPS regardless of its fashion. Over the training period, these metrics demonstrate the model’s performance and gradual improvement.

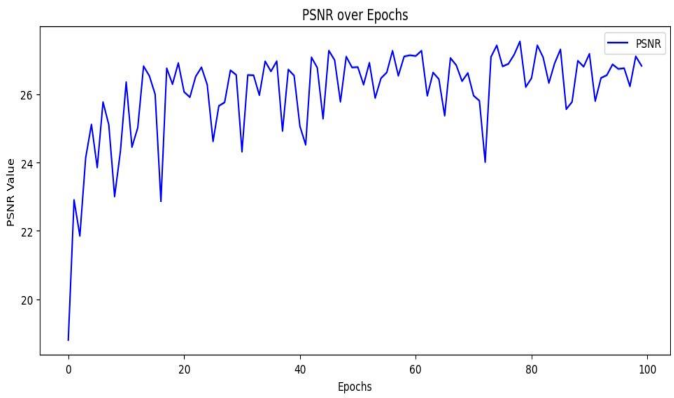

Figure 4: Performance metric across 100 epochs for PSNR average loss value

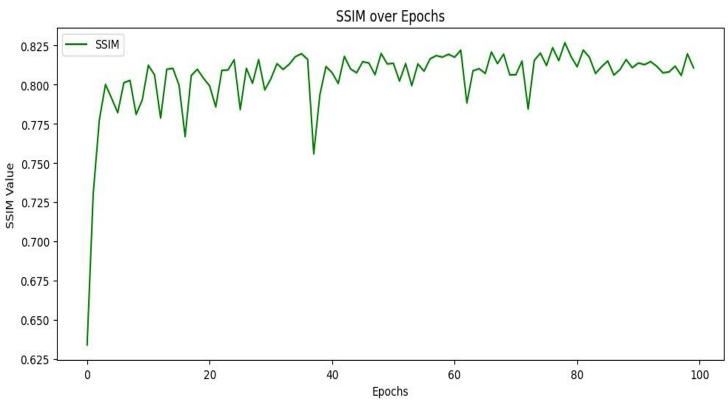

Figure 5: Performance metric across 100 epochs for SSIM average loss value

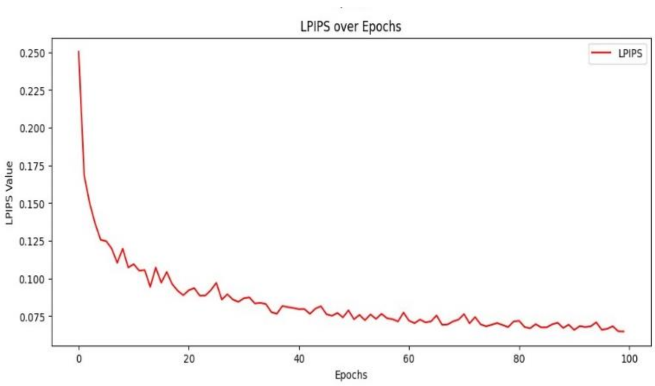

Figure 6: Performance metric across 100 epochs for LPIPS average loss value

4.5 The New Contributions

As said, operating face super-resolution systems necessitates increased hardware complexity and time investment. The primary objective of the research was to achieve an appropriate balance between the output image's accuracy and the network's complexity. Minimising the network size will lead to diminished performance but with the advantage of reduced computational resource demands and decreased time consumption. Initially, we reduce the network's complexity by constructing a smaller network based on a pre-trained VGG16 architecture.

In Figures 7-8, the architecture of the proposed ESRGAN-based method is detailed, focusing on both the Generator and Discriminator networks. Figure 7 illustrates the layered design of the Generator network, emphasizing components like residual dense blocks that enhance feature extraction and image upscaling. Figure 8 displays the structure of the Discriminator network, which aims to distinguish generated high-resolution images from real images, ensuring quality and realism in the super-resolution results. These structures collectively drive the method's effectiveness in producing high-quality images.

Modefing increased hardware complexity and time investment. The primary objective of the research was to achieve an appropriate balance between the output image's accuracy and the network's complexity.

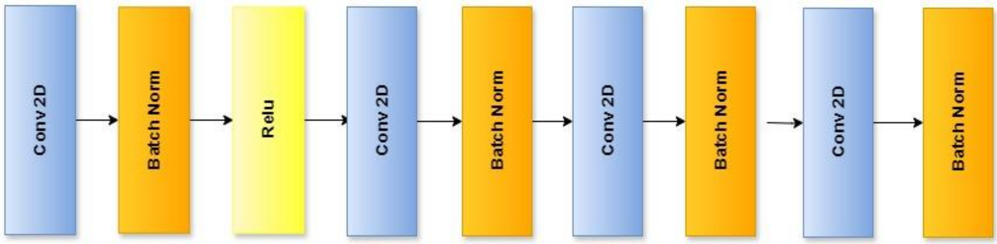

Figure 7: The structure of the Generator network in the proposed method

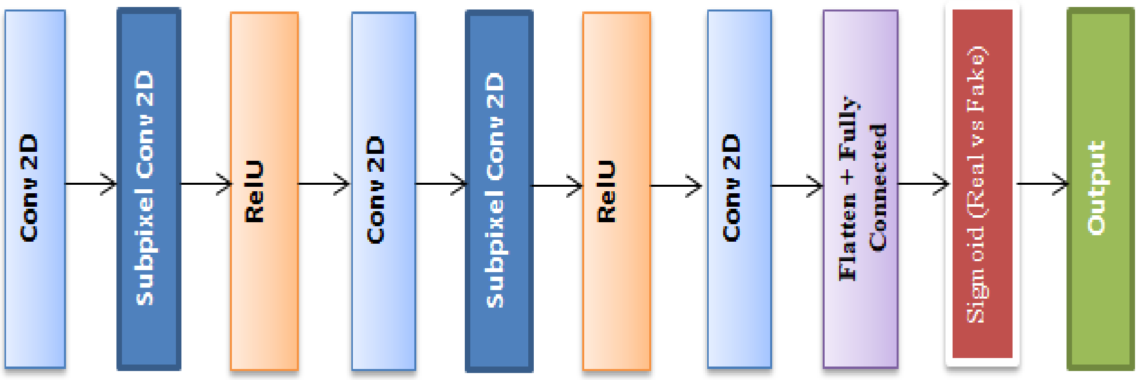

Figure 8: The structure of the Discriminator network in the proposed method

The most important features of this research include:

1) Network Architecture Optimization: The study proposes a modified Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) that utilizes a smaller version of the pre-trained VGG16 architecture to balance output image accuracy with computational resource requirements.
2) Enhanced Feature Mapping: Incorporation of a multi-scale refine block for 4x and 8x up-sampling, which preserves existing features at various resolutions while generating new features with high accuracy, leading to improved image quality.
3) Improved Loss Function: Utilization of a combination of loss functions, including Mean Square Error (MSE), VGG loss, and GAN loss, to enhance the training process and achieve better perceptual quality in the generated images.
4) Performance Metrics: The proposed model has method provides high level of performance metrics with a PSNR of 30.30, SSIM of 0.8757, and LPIPS of 0.0817 which is high and quality than existing methods
5) Parallel Computing Implementation: The study also investigates the impact of parallel computing on the time taken to train, accomplishing an enhancement of about 30% relative to normal systems in terms of calculation time.
6) Focus on Facial Hallucination: The objective of this study is limited to the improvement of facial features in images and also the role of the techniques in ameliorating face hallucination difficulties integrating them to enhance realism of the images generated
7) Comprehensive Dataset Utilization The standard DIV2K dataset, which contains resolution as low as 1000 images, is used in the research to enable quality training and validation of the model.

These features collectively contribute to advancing the field of image super-resolution more so the enhancement of facial images.

5. DISCUSSION

Discussion Figure (9) presents output samples generated by the proposed method of face hallucination, a detailed exploration of pixels that makes it possible to recreate a high final resolution face image starting with a lower one with the samples showing reconstruction details. The applicability of the techniques above on images with fine features and the expected result, which is the enhancement effect that improves the sharper images of the face vividly, is achieved. And the figure shows that the improved method details face image characteristics by utilizing better upscaling and fine details recreation functions, especially for applications that require clear recognition of the face characteristic details.

The findings stemming from the research which improved on the ESRGAN architecture for face hallucination show that the quality of the generated images in detail has been enhanced significantly. The model's performance, as measured by the metrics PSNR and SSIM together achieved the following values: PSNR: 30.30, SSIM: 0.8757 and LPIPS: 0.0817. These numerical measurements cover an understanding of the high quality of the images in terms of details and how they are perceived. In order to contextualize these figures better, it is crucial to compare them with the earlier attempts that have carried out such methods.

Figure 9: The output samples for face hallucination using the proposed method

Compared to our results, the original ESRGAN model by [24], employed for general super-resolution tasks, achieved a PSNR of 29.40 and SSIM of 0.845 on the benchmark dataset. While ESRGAN excels in preserving finer details, our proposed method surpasses this performance, particularly in tasks requiring facial image enhancement. The improvement can be attributed to our model’s modifications, including additional residual blocks and parallel computing in the discriminator phase. Additional residual blocks and parallel computing in the discriminator phase. These modifications improved the resolution and reduced artefacts, as evident from our superior LPIPS score.

Another relevant study [45], one of the earlier GAN-based architectures for image super-resolution. Their model yielded a PSNR of 29.08 and SSIM of 0.815 on the same benchmark dataset. While SRGAN brought GANs into the spotlight for super-resolution, it struggled with balancing pixel-level accuracy and perceptual quality. Our model’s LPIPS score of 0.0817 indicates a significant enhancement in perceptual quality. LPIPS is designed to better capture human-perceived image quality by comparing feature maps rather than pixel-wise differences. In contrast, SRGAN reported a higher LPIPS value, indicating less accurate perceptual similarity to human vision.

Additionally, [47] introduced the LPIPS metric to measure perceptual similarity more effectively. Their studies showed that even high PSNR values do not necessarily correlate with good perceptual quality, a problem that our approach addresses by lowering the LPIPS score to 0.0817. This demonstrates the utility of integrating perceptual loss into the training process to enhance the realism of generated images. The lower LPIPS score from our model demonstrates that our face hallucination results are closer to what a human observer would expect from high-resolution images.

Furthermore, our study utilized the VGG network to compute perceptual loss, combining it with traditional losses such as mean square error (MSE) and adversarial loss to optimize the network. This approach allowed for a more comprehensive loss function that balances pixel-level accuracy and feature-level perceptual quality. Prior work by [48], which also employed perceptual loss based on VGG features, demonstrated its effectiveness in generating visually appealing super-resolution images. However, their model primarily focused on general image super-resolution tasks. In contrast, our study specifically fine-tuned the ESRGAN architecture to handle face hallucination tasks. Adding perceptual loss in our work helped achieve a higher PSNR and SSIM. It significantly improved perceptual similarity, as reflected in our LPIPS score.

Our parallel computing approach in the discriminator phase also plays a crucial role in the improvements observed. In our case, however, the images employed in the model's training were generated within a shorter period, enabling us to split the training time for the dataset by approximately 30%, though the picture quality was not compromised in the end. On the other hand, studies that did not use parallelism had longer training periods attached to achieving similar images. For example, Kim et al. [43], on their VDSR model, which is based on a set of residual blocks for super-resolution, achieved a peak signal-to-noise ratio (PSNR) of 28.03 on the faces dataset but spent a lot of time training due to a lack of parallel processing. It was shown that parallelism increases performance and saves time, making training a viable solution for large-scale usage.

Using RRDB provided us with possibilities concerning the extraction and reconstruction of features during the super-resolution process. This approach enabled the network to learn important features of the face and thus enhance the small features improved in the final output images, which include the eyes, nose and lips. Some works, for example, [49], which look at how DenseNet is used for super-resolution, reported similar achievements using dense connections but had problems with instability during training because of vanishing gradients in the deeper parts of the network.

Our implementation overcame these issues using a more stable training process facilitated by RRDBs and adaptive learning rates.

6. CONCLUSION

This paper presents a modified GAN network architecture for facial hallucination. It specifically formulates a mapping structure that directly calculates facial parsing from LR. It subsequently uses the retrieved parsing map to produce the SR image by forecasting the final facial pictures. A multi-scale refine block with 4x and 8x up-sampling structures is included to preserve prior features at different resolutions and to generate new features with exceptional accuracy. This method results in a significant performance improvement. The new map relies on the interrelationship among pixels, perhaps disregarding the formation of the new image, and generating new pixels results in less similarity in certain regions, such as the eyes, mouth, or other features. Ultimately, the research juxtaposes the outcomes of other GAN architectures to demonstrate the efficacy of the suggested technique.

Including three further layers to the ESRGAN algorithm enhances its capacity to improve picture clarity. These layers improve the capacity to recover intricate details and provide ever-sharper outcomes. Minimise noise and blur By adding supplementary layers. The trained model can comprehend and produce clearer and sharper pictures, improving the quality of the resultant image and diminishing the blur in the original photographs. Enhance intricate details: Incorporating further layers improves the model's capacity to recover intricate information in the picture, including edges and tiny structures, enhancing overall image clarity. The executive summary recommends applying Face Recognition Technology to places such as Nebraska, while improving image quality in low- lighting environments. Additional layers improve image quality in difficult lighting conditions or situations where the model has been trained. They can select extra information data to resolve images and utilize it for further clarification. To reduce noise in the image, expect the algorithm readers to expect close supervision from the staff members, which assists in the clear generation of images through the enhancement of photos while Improving the algorithm with additional layers enhances its performance across different photos, including those of lower quality and those characterized with being blurred or emoting noise

Glossary

PSNR: Peak Signal to Noise Ratio

SSIM: Structural Similarity Index Measure

LPIPS: Learned Perceptual Image Patch Similarity

VDSR: Very Deep Super Resolution

RRDB: Residual-in-Residual Dense Block

VGG: Visual Geometry Group

GAFH-RIDN: Generative Adversarial Face

Hallucination

Residual in Dense Network

SelfEx: Self-Example-based Super-Resolution

RFL: Random Forest Learning

CBN: Conditional Batch Normalization

Data Availability

The datasets generated during and/or analyzed during the current study can be downloaded from

https://www.kaggle.com/datasets/soumikrakshit/div2k-high-resolution-images

ACKNOWLEDGEMENTS

Profound gratitude is extended to all individuals who contributed to the completion of this research, particularly the faculty and administrative colleagues in the Department of Computer Engineering at Qom University, Qom, Iran, for their outstanding support that facilitated the author's acquisition of essential knowledge and insights for the study.

REFERENCES

[1] Grm, K., Scheirer, W. J., & Štruc, V. (2019). Face hallucination using cascaded super-resolution and identity priors. IEEE Transactions on Image Processing, 29, 2150-2165.

[2] Jiang, J., Wang, C., Liu, X., & Ma, J. (2021). Deep learning-based face super-resolution: A survey. ACM Computing Surveys (CSUR), 55(1), 1-36.

[3] Bayramli, B., Ali, U., Qi, T., & Lu, H. (2019). FH-GAN: Face hallucination and recognition using generative adversarial network. In Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part I 26 (pp. 3-15). Springer International Publishing.

[4] Ursuleanu, T. F., Luca, A. R., Gheorghe, L., Grigorovici, R., Iancu, S., Hlusneac, M., ... & Grigorovici, A. (2021). Deep learning application for analyzing of constituents and their correlations in the interpretations of medical images. Diagnostics, 11(8), 1373.

[5] Lu, Y., Chen, D., Olaniyi, E., & Huang, Y. (2022). Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review. Computers and Electronics in Agriculture, 200, 107208.

[6] Sajjadi, S. M. M. (2024). Enhancement and Evaluation of Deep Generative Networks with Applications in Super-Resolution and Image Generation (Doctoral dissertation, Universität Tübingen).

[7] Abello, A. A. (2021). Two studies on Convolutional Neural Networks sensibility to resolution (Doctoral dissertation, Universidade de São Paulo).

[8] Hayat, M., & Aramvith, S. (2024). Saliency-Aware Deep Learning Approach for Enhanced Endoscopic Image Super-Resolution. IEEE Access.

[9] Chandler, D. M. (2013). Seven challenges in image quality assessment: past, present, and future research. International Scholarly Research Notices, 2013(1), 905685.

[10] García, A. S., Fernández-Sotos, P., Vicente-Querol, M. A., Sánchez-Reolid, R., Rodriguez-Jimenez, R., & Fernández-Caballero, A. (2021). Co-design of avatars to embody auditory hallucinations of patients with schizophrenia: A study on patients' feeling of satisfaction and psychiatrists' intention to adopt the technology. Virtual Reality, 1-16.

[11] Pavan Kumar, M. R., & Jayagopal, P. (2021). Generative adversarial networks: a survey on applications and challenges. International Journal of Multimedia Information Retrieval, 10(1), 1-24.

[12] Zhang, Y., Tsang, I. W., Li, J., Liu, P., Lu, X., & Yu, X. (2021). Face hallucination with finishing touches. IEEE Transactions on Image Processing, 30, 1728-1743.

[13] Li, P., Prieto, L., Mery, D., & Flynn, P. J. (2019). On low-resolution face recognition in the wild: Comparisons and new techniques. IEEE Transactions on Information Forensics and Security, 14(8), 2000-2012.

[14] Jiang, J., Hu, R., Wang, Z., & Han, Z. (2014). Noise robust face hallucination via locality-constrained representation. IEEE Transactions on Multimedia, 16(5), 1268-1281.

[15] Lu, T., Hao, X., Zhang, Y., Liu, K., & Xiong, Z. (2019). Parallel region-based deep residual networks for face hallucination. IEEE Access, 7, 81266-81278.
[16] Wang, C., Jiang, J., & Liu, X. (2021, July). Heatmap-aware pyramid face hallucination. In 2021 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.
[17] Wang, X., Yang, Y., Pang, Q., Tang, Q., & Du, S. (2021, August). End-to-end generative adversarial face hallucination through residual in internal dense network. In 2021 29th European Signal Processing Conference (EUSIPCO) (pp. 676-680). IEEE.
[18] Lu, T., Wang, J., Jiang, J., & Zhang, Y. (2020). Global-local fusion network for face super-resolution. Neurocomputing, 387, 309-320.
[19] Fang, C., Li, G., Han, X., & Yu, Y. (2019). Self-enhanced convolutional network for facial video hallucination. IEEE transactions on image processing, 29, 3078-3090.
[20] Gao, J., Tang, N., & Zhang, D. (2023). A multi-scale deep back-projection backbone for face super-resolution with diffusion models. Applied Sciences, 13(14), 8110.
[21] Favorskaya, M. N., & Pakhirka, A. I. (2023). SF-SRGAN: Progressive GAN-based face hallucination. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 47-52.
[22] Schulter, S., Leistner, C., & Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3791-3799).
[23] Wang, S., Hussien, A. G., Kumar, S., AlShourbaji, I., & Hashim, F. A. (2023). A modified smell agent optimization for global optimization and industrial engineering design problems. Journal of Computational Design and Engineering, 10(6), 2147-2176.
[24] Alijoyo, F. A., Prabha, B., Aarif, M., Fatma, G., & Rao, V. S. (2024, July). Blockchain-Based Secure Data Sharing Algorithms for Cognitive Decision Management. In 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET (pp. 1-6). IEEE.
[25] Ramachandran, R., & Sujathamalini, J. (2024). Promoting Diversity And Inclusion In Higher Education: Strategies And Best Practices. Educational Administration: Theory and Practice, 30(4), 6997-7007.
[26] Shourbaji, I. A., & AlAmeer, R. (2013). Wireless intrusion detection systems (WIDS). arXiv preprint arXiv:1302.6274.
[27] F. A. Alijoyo, B. Prabha, M. Aarif, G. Fatma, V. S. Rao and P. Valavan M, "Blockchain-Based Secure Data Sharing Algorithms for Cognitive Decision Management," 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET, Sydney, Australia, 2024, pp. 1-6, doi: 10.1109/ICECET61485.2024.10698611.
[28] Al-Janabi, Samaher & Al-Shourbaji, Ibrahim. (2016). A smart and effective method for digital video compression. 532-538. 10.1109/SETIT.2016.7939927.
[29] Kalpurniya, S., Ramachandran, R., & Chandramohan, N. (2023). A Study on Stress Level, Happiness, Challenges, and Emotional Bonds of Parents having Children with Disabilities Availing Services at
[30] NIEPMD, Chennai. Integrated Journal for Research in Arts and Humanities, 3(5), 72-88.
[31] Shourbaji, I. A., & AlAmeer, R. (2013). Wireless intrusion detection systems (WIDS). arXiv preprint arXiv:1302.6274.
[32] Band, S. S., Ardabili, S., Danesh, A. S., Mansor, Z., AlShourbaji, I., & Mosavi, A. (2022). Colonial competitive evolutionary Rao algorithm for optimal engineering design. Alexandria Engineering Journal, 61(12), 11537-11563.
[33] Singh, A., & Ramachandran, R. (2014). Study on the effectiveness of smart board technology in improving the psychological processes of students with learning disability. Sai Om Journal of Arts & Education, 1(4), 1-6.
[34] Katrawi, A. H., Abdullah, R., Anbar, M., AlShourbaji, I., & Abasi, A. K. (2021). Straggler handling approaches in mapreduce framework: a comparative study. International Journal of Electrical & Computer Engineering (2088-8708), 11(1).
[35] Shiju, K. K., Breja, M., Mohanty, N., Ramachandran, R., & Patra, I. (2023). Importance of Special Education and Early Childhood General Education Teachers' Attitudes toward Culturally Linguistically Diverse People. Journal for ReAttach Therapy and Developmental Diversities, 6(9s (2)), 1544-1549.
[36] Band, S. S., Ardabili, S., Danesh, A. S., Mansor, Z., AlShourbaji, I., & Mosavi, A. (2022). Colonial competitive evolutionary Rao algorithm for optimal engineering design. Alexandria Engineering Journal, 61(12), 11537-11563.

[37] Ramachandran, R., & Singh, A. (2014). The Effect of Hindustani Classical Instrumental Music Santoor in improving writing skills of students with Learning Disability. International Journal of Humanities and Social Science Invention, 3(6), 55-60.

[38] Al-Shourbaji, I., & Duraibi, S. (2023). IWQP4Net: An Efficient Convolution Neural Network for Irrigation Water Quality Prediction. Water, 15(9), 1657.

[39] Sudarsanan, S., Ramkumar Thirumal, H. D. K., Shaikh, S., & Ramachandran, R. (2023). Identifying the Scope of Reattach Therapy for Social Rehabilitation for Children with Autism. Journal for ReAttach Therapy and Developmental Diversities, 6(10s), 681-686.

[40] AlShourbaji, I., Helian, N., Sun, Y., & Alhameed, M. (2021). Customer churn prediction in telecom sector: A survey and way a head. International Journal of Scientific & Technology Research (IJSTR).

[41] Hernandez, R. M., Ragasa, A. A., Macatangay, L., Palad, L., Mendez, C., & Mendez, C. (2023, October). Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) on Melanoma Skin Cancer Detection. In Proceedings of the 8th International Conference on Sustainable Information Engineering and Technology (pp. 137-142).

[42] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).

[43] Mosaher, M., & Alam, E. K. A Joint Framework of GFP-GAN and Real-ESRGAN for Real-World Image Restoration.

[44] Sun, X., Nasrabadi, N. M., & Tran, T. D. (2019). Supervised deep sparse coding networks for image classification. IEEE Transactions on Image Processing, 29, 405-418.

[45] Wang, Y., Perazzi, F., McWilliams, B., Sorkine-Hornung, A., Sorkine-Hornung, O., & Schroers, C. (2018). A fully progressive approach to single-image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 864-873).

[46] Lai, Z., Chen, L., Liu, Z., & Yang, X. (2022). Gradient guided pyramidal convolution residual network with interactive connections for pan-sharpening. International Journal of Remote Sensing, 43(15-16), 5572-5602.

[47] Liu, J., Xu, S., He, P., Wu, S., Luo, X., Deng, Y., & Huang, H. (2024). VSG-GAN: A high-fidelity image synthesis method with semantic manipulation in retinal fundus image. Biophysical Journal.

[48] Kim, J., Li, G., Yun, I., Jung, C., & Kim, J. (2021). Edge and identity preserving network for face super-resolution. Neurocomputing, 446, 11-22.

[49] Iqbal, S., Khan, T. M., Naqvi, S. S., Naveed, A., Usman, M., Khan, H. A., & Razzak, I. (2023). Ldmres-Net: a lightweight neural network for efficient medical image segmentation on iot and edge devices. IEEE Journal of Biomedical and Health Informatics.

[50] Liu, B., Tang, J., Huang, H., & Lu, X. Y. (2020). Deep learning methods for super-resolution reconstruction of turbulent flows. Physics of fluids, 32(2).

[51] Ejaz, M. S., Islam, M. R., Sifatullah, M., & Sarker, A. (2019, May). Implementation of principal component analysis on masked and non-masked face recognition. In 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT) (pp. 1-5). IEEE.

[52] Girshick, R. (2015). Fast r-cnn. arXiv preprint arXiv:1504.08083.

[53] Lin, K., Zhao, H., Lv, J., Li, C., Liu, X., Chen, R., & Zhao, R. (2020). Face Detection and Segmentation Based on Improved Mask R-CNN. Discrete dynamics in nature and society, 2020(1), 9242917.

[54] Qin, B., & Li, D. (2020). Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Sensors, 20(18), 5236.

[55] Dasgupta, M., Bandyopadhyay, O., & Chatterji, S. (2019, December). Automated helmet detection for multiple motorcycle riders using CNN. In 2019 IEEE Conference on Information and Communication Technology (pp. 1-4). IEEE.

[56] Tian, Z., Shen, C., Chen, H., & He, T. (2020). FCOS: A simple and strong anchor-free object detector. IEEE transactions on pattern analysis and machine intelligence, 44(4), 1922-1933.

[57] Wang, X. (2021). Deep learning based image super resolution (Doctoral dissertation).

[58] Agustsson, E., & Timofte, R. (2017). Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 126-135).

[59] Prodan, M., Vlăsceanu, G. V., & Boiangiu, C. A. (2023). Comprehensive evaluation of metrics for image resemblance. Journal of Information Systems & Operations Management, 17(1), 161-185.

[60] Rehman, A., Rostami, M., Wang, Z., Brunet, D., & Vrscay, E. R. (2012). SSIM-inspired image restoration using sparse representation. EURASIP Journal on Advances in Signal Processing, 2012, 1-12.

[61] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

[62] Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13 (pp. 184-199). Springer International Publishing.

[63] Huang, J. B., Singh, A., & Ahuja, N. Single image super-resolution using transformed self-exemplars. CVPR2015, 1-10.

[64] Timofte, R., De Smet, V., & Van Gool, L. (2015). A+: Adjusted anchored neighborhood regression for fast super-resolution. In Computer Vision--ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part IV 12 (pp. 111-126). Springer International Publishing.

[65] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 586-595).

[66] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14 (pp. 694-711). Springer International Publishing.

[67] Anwar, S., Khan, S., & Barnes, N. (2019). A deep journey into super-resolution: a survey. arXiv. arXiv preprint arXiv:1904.07523.