Image Super-Resolution using Generative Adversarial Networks with EfficientNetV2

Saleh AlTakrouri[1], Norliza Mohd Noor[2], Norulhusna Ahmad[3], Taghreed Justinia[4], Sahnius Usman[5]

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia
Kuala Lumpur Campus, Jalan Sultan Yahya Petra, Malaysia[1,2,3,5]
King Saud bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia[4]

Abstract—The image super-resolution is utilized for the image transformation from low resolution to higher resolution to obtain more detailed information to identify the targets. The super-resolution has potential applications in various domains, such as medical image processing, crime investigation, remote sensing, and other image-processing application domains. The goal of the super-resolution is to obtain the image with minimal mean square error with improved perceptual quality. Therefore, this study introduces the perceptual loss minimization technique through efficient learning criteria. The proposed image reconstruction technique uses the image super-resolution generative adversarial network (ISRGAN), in which the learning of the discriminator in the ISRGAN is performed using the EfficientNet-v2 to obtain a better image quality. The proposed ISRGAN with the EfficientNet-v2 achieved a minimal loss of 0.02, 0.1, and 0.015 at the generator, discriminator, and self-supervised learning, respectively, with a batch size of 32. The minimal mean square error and mean absolute error are 0.001025 and 0.00225, and the maximal peak signal-to-noise ratio and structural similarity index measure obtained are 45.56985 and 0.9997, respectively.

Keywords—*Single image super-resolution (SISR); generative adversarial networks (GAN); convolutional neural networks (CNN); EfficientNetv2*

I. INTRODUCTION

With the rapid development of information technology (IT) along with the boom of Internet technology, information processing based on image and signal is widely utilized by an enormous population, in which image processing is a crucial component of information processing. Here, the role of image super-resolution (SR) is significant when considering image processing-based applications [1], [2]. Image SR is the image transformation from low resolution (LR) to high resolution (HR) for obtaining an enhanced quality image. Medicine, agriculture, industry, and military applications utilize the SR technique due to its high practicability [3], [4]. While considering artificial intelligence, the role of SR is crucial for performing various processes [5], like public security, remote sensing imaging, medical imaging, image compression, and so on using the single image SR criteria [6], [7]. The image resolution enhancement using the up-sampling process lacks texture details. The image transformed into the HR provides enormous information with finer details [8]. For example, the crime scene image offers plenty of evidence for investigating crime. Likewise, an image acquired from the satellite image undergoes various processing, like resource detection, object detection, and several other processing using the HR image [9]. While considering the medical application domain, the disease diagnosis is employed based on Magnetic Resonance Imaging

(MRI) and Computed Tomography (CT) Scan images with better resolution for providing accurate medication. Thus, the role of SR is crucial in image-processing application domains.

The SR of the image is obtained from the LR image using three various categories: 1) learning-based approach, 2) reconstruction approach, and 3) interpolation approach [10]. Image resolution enhancement using interpolation is the earliest method most researchers utilized and is easy to implement. Some of the interpolation techniques used to enhance the image's quality are non-uniform sampling interpolation, Bicubic Interpolation, Bilinear Interpolation, and Nearest Neighbor Interpolation. In these approaches, the higher frequency details can be reconstructed through the linear characteristics of the approaches. The image SR using the interpolation approach provides a better outcome; still, the performance degrades with the scaling factor's elevation [11]. Reconstruction based SR approach transforms the LR image by gathering the non-redundant details. The approaches with non-negativity, energy boundedness, support boundedness, and smoothness-based hypothetical constraints using the Projection onto Convex Set are one of the methods utilized for SR reconstruction. The slower convergence rate is the limitation of the reconstruction-based criteria and has many solutions [12], [13]. In addition, the reconstruction solution acquired at the final stage hang-on on the initial evaluation. Also, the performance is limited while considering reliable robustness and real-time modeling due to the inefficiency in handling the noise level reconstruction [14]. Finally, the third approach is the learning-based image transformation approach that enhances the quality of the image using machine learning and deep learning algorithms [15]. Nowadays, learning-based image quality enhancement is the widely utilized approach by researchers [16] due to better image perception.

The learning of the network is employed in the learning-based approach of image SR for providing a high-quality reconstructed image. Here, for network learning, the high representation of the samples is utilized with variation in data for generalizing [17]. Thus, enormous data is acquired from various sources to obtain the required solution. While considering the image taken from the remote sensing domain, the information collection is a challenging task due to the variations of image based on the factors like different sensors and locations along with the difference in the objects [18], [19], [20]. Thus, the network learned with the limited samples affects the model's performance due to the poor generalization capability. Nowadays, the advent of deep learning models based on non-linear operation in artificial intelligence accumulates enormous

samples that can be utilized for several computer vision-based applications [21]. Hence, the generalization capability of the deep learning models is elevated through the enormous learning of data samples. The convolutional neural network (CNN) based image resolution enhancement technique was first suggested by Dong et al. [22]. The method acquired a better signal-to-noise ratio (SNR) with a high sampling rate, but the information about the image is lost due to the too smoothness of the reconstructed image [23], [24]. In the deep learning approaches, the error concerning the mean squared error (MSE) reduction of 2.98 and the elevation of the peak signal-to-noise ratio (PSNR) value of 21.15dB are obtained through the optimization target for acquiring the high-quality image from the ill-posed nature. The variation between the reconstructed and the original image based on the perception is not photorealistic by some existing approaches. Generative adversarial network (GAN) based methods are also designed by several researchers to overcome the issue concerning information loss in image reconstruction, in which the CNN is replaced with the GAN network [3]. The GAN-based image SR techniques gather the non-linear high-dimensional features from enormous data [25], [26] and make the generalization more effective for providing a better outcome.

This study aims to reconstruct the SR image from the LR image by minimizing the perceptual loss to obtain a perceptually good image with detailed information that gratifies the human eye. This study proposes ISRGAN with EfficientNet-v2 to enhance reconstruction accuracy with minimal distortions. The ISRGAN is designed for image reconstruction by minimizing the perceptual loss comprising of content and adversarial loss with EfficientNet-v2 is utilized for learning the discriminator of the ISRGAN, in which the perceptual loss concerning the image is minimized for the acquisition of a better quality image.

The significant contributions of the research are:

- Design of ISRGAN: Image Super-Resolution Generative Adversarial Network (ISRGAN) is designed for image reconstruction by minimizing the perceptual loss comprising of content and adversarial loss.

- EfficientNet-v2: EfficientNet-v2 is utilized for learning the discriminator of the ISRGAN, in which the perceptual loss concerning the image is minimized for the acquisition of a better quality image through the fast learning criteria.

The manuscript is organized as follows: The primary methods review is enclosed in Section II with its problem definition. The detailed proposed ISRGAN based on EfficientNet-v2 is portrayed in Section IV, Section V for implementation, and Section VI discusses the results and analysis of the study. Finally, Section VII concludes the work.

II. RELATED WORK

The primary methods concerning the image SR are detailed in this section. The image SR using the GAN concerning the quality loss was devised by Zhu et al. [27] to solve the issue of instability of the GAN in image reconstruction. The method addressed the loss issue by optimizing the pre-trained network. The gradient magnitude similarity deviation was utilized for the discriminator training to obtain a better performance. In addition, batch normalization, computation overhead, and memory reduction make learning more efficient and obtain better visual appeal. Image SR using GAN with the fused attentive network to extract global features was designed by Jiang et al. [28] using the scanning time reduction technique. In this study, the attributes at various scales are taken out based on the attention criteria to acquire the most relevant attributes and enhance reconstruction accuracy. In addition, the stability of the learning is accomplished through the spectral normalization approach. The analysis based on the distance measure and the similarity index offered a better outcome.

Shahidi [29] designed the wide attention-based GAN to stabilize the learning process using the Wasserstein with a Gradient penalty. Two different normalizations, such as weight and batch normalization, elevate the similarity index of the image by considering the texture and color restoration. Besides, the inclusion of the self-attention layers and the residual blocks assures a high-quality image by learning past information. The self-attention layer is utilized in the discriminator to extract the attributes from the image patches to correct the generator more effectively. The loss function was evaluated using the VGG-19 network. An enlightened GAN was developed by Gong [30] using the self-supervised hierarchical perception loss to acquire enhanced image reconstruction performance through network induced convergence. The enlighten blocks were introduced to accomplish a better gradient using the improved generalization capability. Besides, the occurrence of seam lines in the reconstructed image was eliminated through the clipping and merging approach-based learning criteria using the batch internal inconsistency loss. The image quality assessment of the method provided a superior outcome by solving the merging issue that depicts the realistic criteria of the developed method. GAN with cascading residual network was designed by hybridizing the neural network (NN) with the GAN by Ahn et al. [31] to acquire the network's short connection and multi-level representation for improving the reconstruction performance. The balanced distortion and the perception criteria of the designed method are made using the multi-scale discriminator and GAN-based learning. Here, the usage of the multi-scale discriminator gathers fine information from the image concerning the resolution to reduce the losses in the network. The visual outcome of the method provided a realistic and sharp image.

Residual-in-Residual Dense Block-based GAN was devised by Song et al. [32], in which the optimal network with transfer learning was employed to acquire a high-quality image with better perception and low distortion. The slow convergence rate and the unstable learning capability were enhanced through the inclusion of the dense block with enhanced perception and minimal distortion in the reconstructed image. In addition, the feature reuse and propagation were improved along with the vanishing gradient issue solution, which was accomplished through the dense connection establishment with minimal parameters. A conditional GAN-based image reconstruction was presented by Sun et al. [33], in which the color space transformation was initially employed based on the variance and mean of the channel for the channel normalization. The better visual representation of the image was accomplished through the colorization training criteria, which converge faster than the traditional image reconstruction approach. In addition,

curriculum learning was utilized to solve the issue of large resolution differences in the learning phase. Color normalization through self-supervised learning offers better generalization capability to obtain better performance by eliminating color variance. A weighted Multi-Scale Residual Block-based image reconstruction strategy was introduced by Li et al. [34] for image reconstruction through the weighted feature representation approach. In this, the low-level attributes were taken out from the image. Then, using the global residual learning, the high-frequency information from the features is mapped using the non-linear mapping criteria. The reconstruction of the image was accomplished from the reconstruction subnet of the network.The benefits and the challenges of the reviewed techniques is depicted in Table I.

A. Research Gap

The acquisition of high resolution image from the low resolution input image is the image super resolution technique. The learning based methods of image super resolution are widely utilized for reconstructing the image to accomplish the better performance through the non-linear learning capability. The goal behind the image reconstruction is to obtain the better perceptual image to identify the target solution in medical imaging and satellite imaging applications. Many of the prior researchers devised the GAN based image reconstruction technique for obtaining better performance. Still, the loss function estimation and the network training with minimal loss is a challenging task in most of the methods based on the GAN. In addition, the un-pleased image reconstruction with higher distortions and the over-smoothing of the image degrades the better reconstruction. Hence in order to overcome the above mentioned challenges, an ISRGAN is introduced for learning the discriminator with higher learning speed with minimal information loss, which enhances the generalization capability of the network and obtains the better image.

III. METHODOLOGY

The ISRGAN is designed by modifying the loss function of the traditional GAN to enhance the image quality that is more pleasant to the human eyes and to improve the accuracy of image reconstruction. The GAN for the super image resolution of the single image is termed ISRGAN for transforming the LR image into the SR image. The ISRGAN comprises of discriminator and generator, which denotes $A_{\theta A}$ the discriminator model. The min-max issue of the ISRGAN is solved by optimizing $Q_{\theta Q}$ alternatively and $A_{\theta A}$. The formulation for the min-max issue is expressed as,

$$\min_{\theta \infty} \max_{\theta 1} \mathrm{E}_{S^R \sim p-\mathrm{train}\left(S^R\right)} \left[\log A_{\theta A}\left(S^R\right)\right]$$
$$+ \mathrm{E}_{S^L \sim p-G\left(S^L\right)} | \left[\log\left(1 - A_{\theta A}\left(D_{\theta D}\left(S^R\right)\right)\right)\right] \quad (1)$$

where, the high-resolution image is notated as S^R, the low-resolution image is indicated as S^L, the discriminator is notated as A, and the generator is indicated as Q. Here, the role of the generator is to mislead the discriminator, in which the discriminator module of the ISRGAN is learned to identify the difference between the real image and the SR image. Similarly,

TABLE I. REVIEWED TECHNIQUES: BENEFITS AND CHALLENGES

Technique	Benefits	Challenges
GAN [27]	The enhanced adversarial learning along with the loss function estimation using the image quality assessment provided a high quality image.	The failure case of the method was higher and provides an unsatisfactory result. Besides, the imperfection of the network portrayed diagonal lines in the blocks of the outcome.
Fused Attentive GAN [28]	The assessment with various GAN-based image super-resolution acquired better performance in terms of signal-to-noise ratio using various datasets. Besides, the reconstructed image is near to the real image with high resolution.	The reduction of the loss function with a higher number of iterations elevates the computation overhead.
Wide-Attention GAN [29]	Two normalizations like weights and batch normalization along with the self-attention layer provide the method a small range of gradient clipping that elevates the accuracy of image reconstruction.	The analysis without the pre-processing step limits the performance of the model.
Enlighten-GAN [30]	The convergence stabilization along with the network intensification provided a better outcome with an artifact-free reconstructed image by maintaining the hue and shape.	The outcome accomplished using the method still has the issue of an unclear outline due to the object fusing technique. Hence, for the complex image, the developed method provided poor performance.
Cascading Residual Network-GAN [31]	The improved performance is accomplished for the image translation technique by balancing distortion and perception.	The complexity analysis based on the execution time depicts the minimal time complexity compared to some baseline techniques; still, acquired poor performance due to the memory fragmentation criteria that limit the parallelism of the approach.
Residual-in-Residual Dense Block-GAN [32]	The visual outcome of the method was better and was applicable for the applications that don't require the details concerning the place and obtained better running time.	The signal-to-noise ratio evaluated by the method was not pleasant to the human eye.
Weighted Multi-Scale Residual Block [33]	The weight-based image reconstruction obtained better performance using the high-frequency attributes acquired from the low-resolution image.	The method failed to analyze the accuracy.
Conditional GAN with color normalization [34]	The augmentation of the features helps to acquire various degradations degrees, which assures the capability of learning from the de-blurred image that enhances the convergence rate and boosts the performance of the method.	For the images with severe information loss, the method offered an unsatisfactory outcome.

the generator module is trained to produce the SR image close to the real image, making the discriminator's detection capability a tedious task. As mentioned above, the generator and the discriminator processes provide a more appropriate perceptual solution. The maximization issue is solved using the discriminator with minimal MSE. The architecture of the ISRGAN is portrayed in Fig. 1.

IV. PROPOSED WORK

The proposed ISRGAN with EfficientNet-v2 incorporates the EfficientNet-v2 to evaluate the loss function and train the discriminator based on the loss to enhance the quality of image reconstruction. The evaluation of the loss function and the
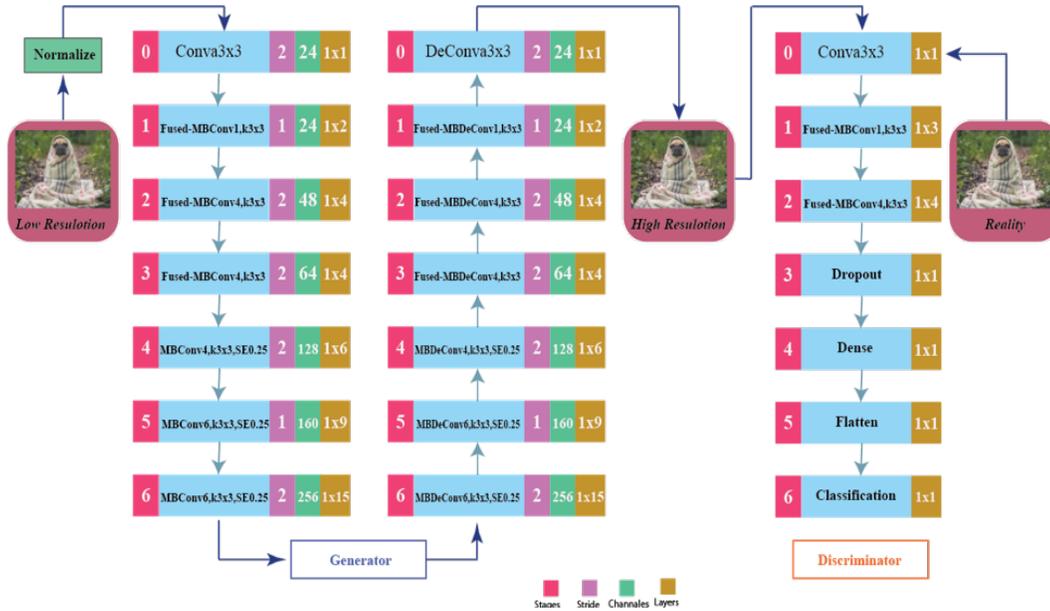
Fig. 1. The architecture of ISRGAN

learning based on the loss function using the EfficientNet-v2 is detailed below.

A. Loss Function

1) Perceptual loss : The perceptual loss function of the ISRGAN is considered a crucial factor that elevates the perceptual characteristics of the image based on the loss function [4]. The adversarial and content loss function of the ISRGAN constitutes the perceptual loss function and is formulated as,

$$h^H = h_B^H + 10^{-3} h_G^H \tag{2}$$

where, the adversarial loss of the ISRGAN is referred to as $10^{-3} h_G^H$, the content loss of the ISRGAN is indicated as h_B^H and the perceptual loss evaluated by the VGG-based network is indicated as h^H.

2) Content loss: Here, the goal of the ISRGAN is to replace the perceptual loss using the VGG feature map of the network. The loss function of the pixel-wise MSE, evaluated by VGG, is formulated as,

$$h_E^H = \frac{1}{s^2 KN} \sum_{m=1}^{sK} \sum_{n=1}^{sN} \left(S_{m,n}^R - Q_{\theta Q}(S^L)_{m,n} \right)^2 \tag{3}$$

where, the MSE error estimated by the VGG is indicated as h_E^H, and the height and width of the low-resolution image are indicated as N and K, respectively. For the image points *(m, n)*, the gray level value is indicated as $(S^L)_{m,n}$. Then, the characteristic loss of the VGG network is formulated as,

$$h_{VGG/u,v}^H = \frac{1}{K_{u,v} N_{u,v}} \sum_{m=1}^{K_{u,v}} \sum_{n=1}^{N_{u,v}} \left(\alpha_{u,v}(S^R)_{m,n} \right.$$
$$\left. - \alpha_{u,v}(Q_{\theta Q}(S^L))_{m,n} \right)^2 \tag{4}$$

where, H refers to the image SR, R refers to the high-resolution image, $h_{VGG/u,v}^H$ refers to the low-resolution image, L refers to the characteristic loss for the VGG network, the reference image is indicated as (S^R), and the reconstructed image is indicated as $Q_{\theta Q}.u^t h$ Max-pooling layer with $v^t h$ convolution corresponding to the feature map is indicated as $\alpha_{u,v}$. The dimensions of the feature maps are indicated as $N_{u,v}$ and $K_{u,v}$, respectively. Here, the characteristics loss function is evaluated based on the activation function of VGG loss, which is nothing but the Euclidean distance between the reference and reconstructed image. The combination of the characteristic loss and the MSE error constitutes the content loss of the ISRGAN.

3) Adversarial loss: The loss generated by the generative component of the ISRGAN for providing the most appropriate solutions by tricking the discriminator of the network is termed Adversarial loss, which is formulated as,

$$h_G^H = \sum_{c=1}^{C} - \log A_{\theta A}(Q_{\theta Q}(S^L)) \tag{5}$$

where, the natural high-resolution image is indicated as $Q_{\theta Q}(S^L)$, and the reconstructed image's probability is notated as $A_{\theta A}(Q_{\theta Q}(S^L))$. Here, the adversarial loss function needs to be minimized for the acquisition of better gradient behavior.

B. EfficientNet-v2 based Perceptual Loss Function

The proposed improved ISRGAN utilizes EfficientNet-v2 to estimate the loss function to enhance the quality of the reconstructed image perceptually. The ISRGAN has the issue of gradient disappearance that makes information losses due to poor generalization capability. Thus, the EfficientNet-2 is utilized for learning the discriminator based on the perceptual loss function through the higher learning rate and the parameter efficiency. Here, the role of the EfficientNet-v2 is to detect the difference between the ground truth and the reconstructed image for measuring the loss function to train the discriminator to make the reconstruction more effective. The loss functions like content loss and the adversarial losses that constitute the perceptual loss are evaluated using the EfficientNet-v2 to enhance the efficiency of the image SR.

EfficientNet-v2 for estimating the loss function utilizes smaller kernel sizes with minimal memory access overhead. Besides, the last stride-1 is eliminated to reduce the memory access overhead and size of the parameter. The runtime overhead is minimized through network capacity elevation, and the training overhead, along with higher memory, is minimized by restricting the interference of the image. The learning capability of the EfficientNet-v2 is higher by making the original interference size for learning, which depicts the scaling characteristics of the loss function estimator. The building blocks of the EfficientNet-v2 are Fused MBConv along with the mobile inverted bottleneck MBConv and are portrayed in Fig. 2.

C. ISRGAN Architecture

The ISRGAN comprises two different modules: the discriminator and the generator for producing the perceptually quality image.

1) Generator: The transformation of the input image to obtain the SR image is employed in the generator module of the proposed ISRGAN. Initially, the encoding of the input image is performed to acquire the high-dimensional information. Then, the decoding is employed in the hidden layer of the generator to provide high-resolution images. Here, the image dimensions are scaled using the scaling algorithm to acquire the feature resolution, which is performed as convolutional and deconvolutional operations in the generator. Thus, the CNN is utilized in the generator module of the ISRGAN to extract the relevant features with a single dimension, which leads to the failure in acquiring the granular patterns of the image. Thus, the compound scaling factor is introduced in the ISRGAN to obtain the image's finer granular patterns. Hence, the generator module is designed with 42 layers of convolutional and deconvolutional, along with the Leaky rectified linear unit (Leaky ReLU) activation function and batch normalization. Besides, training aware Neural Architecture Search (NAS) and scaling are utilized for designing the generator module.

2) Discriminator: The difference between the high-resolution image generated by the generator module and the real image is identified by the discriminator module using the EfficientNet-v2. Here also, finer granular patterns are extracted from the image. Seven Fused-MBConv, one convolutional layer, and eight hidden layers constitute the discriminator module along with the EfficientNet-v2. The sigmoid function
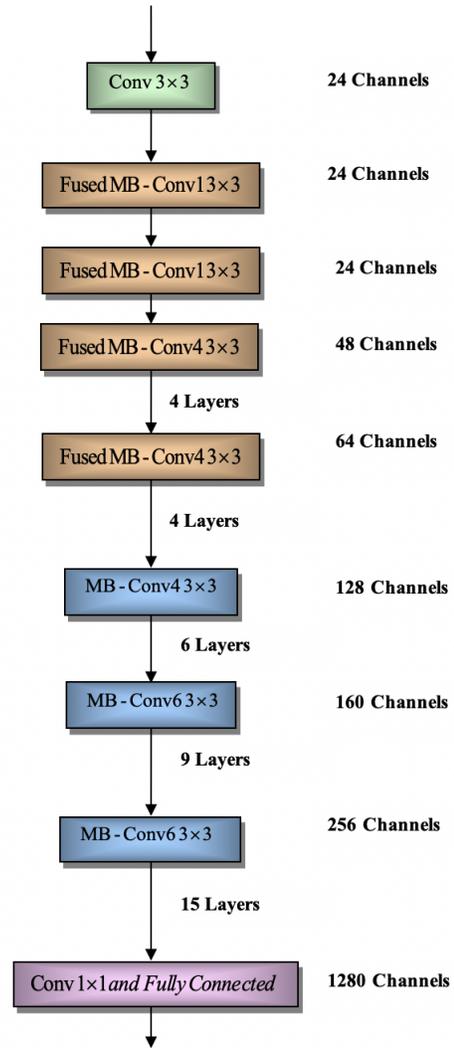

Fig. 2. Architecture of efficientNetv2

is utilized to determine the likelihood of SR images by considering the fake and real images. The perceptual loss function training is employed alternatively in both the discriminator and generator to obtain perceptually better image SR.

V. IMPLEMENTATION

The experimentation setup along with the interpretation of the introduced ISRGAN is detailed in this section.

A. Experimental Setup

The implementation of the proposed method is employed in Google Colaboratory (Colab) Notebooks tool. The total number of iterations performed is 100, with batch sizes of 8 and 32.

1) Image dataset: The dataset utilized for the analysis of the proposed ISRGAN is the CIFAR-10 dataset [35] comprises 10000 test images and 50,000 training images with 10 different classes. Thus, a total of 60,000 images with the size of is utilized. The type of image utilized for the evaluation is synthesis.

2) Network learning: The generator module of the ISR-GAN is learned with the loss function, expressed as,

$$\text{Loss}_Q = h^H + \text{Loss}_{\text{texture}} \qquad (6)$$

where, Loss_Q refers to the loss utilized to learn the generator, h^H notates the perceptual loss, which comprises adversarial loss and content loss and the texture loss is indicated as $\text{Loss}_{\text{texture}}$. Here, the Adam optimizer is utilized to solve the optimization issue, and the learning rate of $1e^{-4}$ is initially utilized. EfficientNet-v2 is utilized for extracting the finer granular patterns from the image.

3) Loss function: The empirical risk of the proposed IS-RGAN in reconstructing the HR image is evaluated based on the loss function and is enunciated as,

$$\text{Loss} = \frac{1}{W} \sum_{i=1}^{W} \left(O_i - \hat{O}_i \right)^2 \qquad (7)$$

where, the total reconstructed samples is indicated as W, the HR reconstructed image by the proposed ISRGAN is O_i and the targeted outcome is indicated as O_i.

4) Network parameter setting: The setting for the network parameters of the Generator and Discriminator in the ISRGAN architecture are displayed in Table II.

TABLE II. ISRGAN PARAMETER SETTINGS

Parameters	Values
Generator	
Dense Layer	256x4x4
Conv-1	128 filters, 3x3 kernel, 2x2 stride
Conv-2	128 filters, 3x3 kernel, 2x2 stride
Conv-3	128 filters, 3x3 kernel, 2x2 stride
Conv-4	3 filters, 3x3 kernel
Discriminator	
Input Layer	32x32x3
Conv-2	64 filters, 3x3 kernel, 2x2 stride
Conv-3	128 filters, 3x3 kernel, 2x2 stride
Conv-4	128 filters, 3x3 kernel, 2x2 stride
Conv-5	256 filters, 3x3 kernel, 2x2 stride

VI. RESULTS AND ANALYSIS

A. Qualitative Analysis

The experimental outcome concerning the input LR image and the output HR image is shown in Fig. 3.

The outcome acquired by the introduced ISRGAN shown in Fig. 3(b) is visually pleasing to the human eye by reconstructing the LR image 3(a). The discriminator learning based on the perception loss estimated by the generator using the EfficientNet-v2 elevates the accuracy of image reconstruction by minimizing the MSE.

B. Quantitative Analysis

The quantitative analysis based on the loss function evaluated by the proposed method and the error analysis is detailed in this section.

Fig. 3. Outcome of ISRGAN: (a) input LR image and (b) output HR image

1) Analysis based on loss function: The quantitative analysis of the introduced ISRGAN is analyzed based on the loss function measure and is portrayed in Fig. 4. The Table III presents the batch sizes, number of iterations, and outcomes of generator, discriminator, and the self-supervised. From the table, we can observe that the increase in batch size and iteration minimizes the loss. Here, the increase in the batch size of the method reduces the noise in the gradients, which assures enhanced accuracy in image reconstruction and reduces the loss. Similarly, the increase in iteration helps the algorithm obtain a closer solution to the target solution, which reduces the loss and enhances image reconstruction accuracy. The ISRGAN with the EfficientNet-v2 acquires minimal loss and can be applied for the media, medical, and surveillance-related application domains by detecting various objects in the image. Besides, the loss function evaluation based on the perceptual similarity loss assists in capturing the high-level information that provides a more accurate solution.
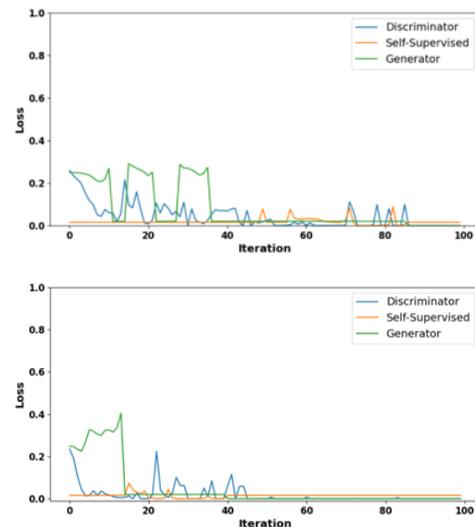

Fig. 4. Analysis based on loss function with batch size 8 (top) and batch size 32 (bottom)

TABLE III. Loss Function Analysis

Batch size	Iteration	Generator	Discriminator	Self supervised
8	20	0.251	0.012	0.0154
32	20	0.02	0.000057	0.037
8	100	3.778	3.916	0.0154
32	100	0	9.130	0.015

2) Error and similarity analysis: The analysis is executed based on MSE, PSNR, structural similarity index measure (SSIM), and mean absolute error (MAE).

MSE: The analysis of the risk function concerning the target solution in reconstructing the image is measured through the MSE and is formulated as,

$$\text{ISRGAN}_{\text{MSE}} = \frac{1}{T_{\text{samples}}} \sum_{i=1}^{T_{\text{samples}}} \left(I_{\text{outcome}} - I_{\text{Target}} \right)^2 \quad (8)$$

Here, the targeted outcome is indicated as I_{Target}, the outcome acquired by the proposed ISRGAN is indicated as I_{outcome}, the total number of samples is indicated as T_{samples} and the MSE is notated as $\text{ISRGAN}_{\text{MSE}}$. The MSE evaluated by the proposed ISRGAN and the comparative methods like Wide-Attention GAN [29], Enlighten GAN [30], and Fused Attentive GAN [28] is portrayed in Fig. 5. The MSE evaluated by the introduced ISRGAN with the EfficientNet-v2 is minimal compared to the other conventional methods.
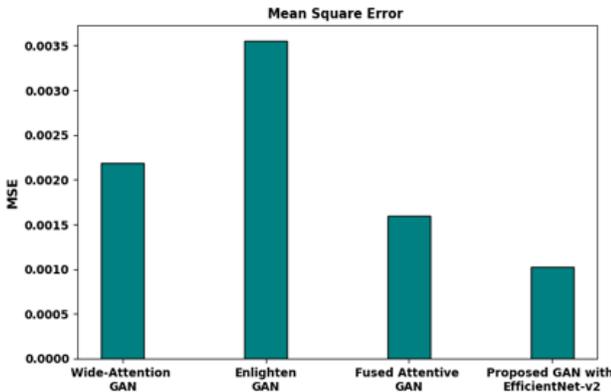

Fig. 5. Analysis based on MSE

MAE: The absolute difference between the obtained and the targeted outcome of the super image resolution is measured through the MAE and is formulated as,

$$\text{ISRGAN}_{MAE} = \frac{1}{T_{\text{samples}}} \sum_{i=1}^{T_{\text{sanples}}} \left| I_{\text{outcome}} - I_{T\text{ arget}} \right| \quad (9)$$

where, the MAE measure is indicated as ISRGAN_{MAE}. The outcome of the proposed ISRGAN, along with the comparative methods, is portrayed in Fig. 6. The MAE evaluated by the proposed method is minimal compared to the other state-of-the-art techniques.
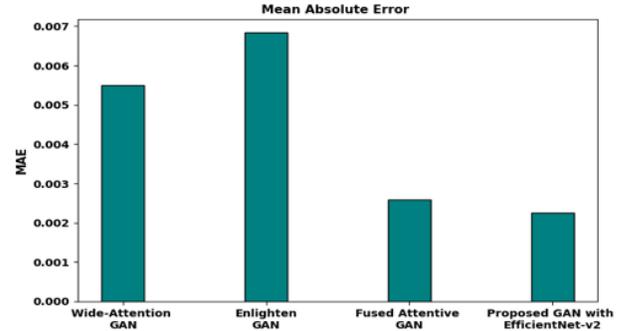

Fig. 6. Analysis based on MAE

PSNR: The quality of the image reconstructed from the low-resolution image is measured through the PSNR. The greater value of PSNR depicts better reconstruction. The formulation of the PSNR of the reconstructed image is expressed as,

$$\text{ISRGAN}_{PSNR} = 10 \log_{10} \left(\frac{R^2}{ISRGAN_{MSE}} \right) \quad (10)$$

where, the PSNR measure is indicated as ISRGAN_{PSNR}, and the fluctuation of the reconstructed image compared to the original image is denoted as R. The PSNR estimated by the proposed and the comparative methods is portrayed in Fig. 7. The PSNR acquired by the introduced method is higher and depicts a better image reconstruction quality.
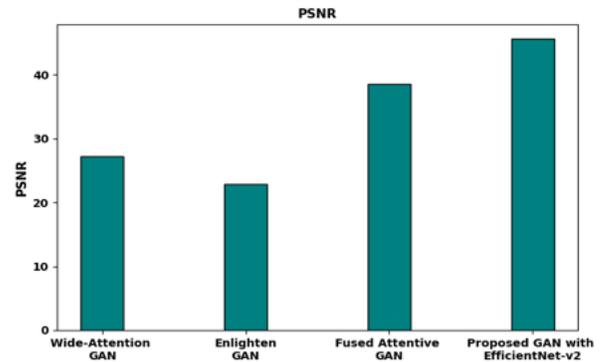

Fig. 7. Analysis based on PSNR

SSIM: The similarity between the reconstructed and the original image is measured through the SSIM that depicts the perceptual quality of the image. The mathematical formulation of the SSIM is expressed as,

$$\text{ISRGAN}_{SSIM} = \frac{\left(2\mu_p\mu_q + s_1 \right)\left(2\sigma_{pq} + s_2 \right)}{\left(\mu_p^2 + \mu_q^2 + s_1 \right)\left(\sigma_p^2 + \sigma_q^2 + s_2 \right)} \quad (11)$$

where, the SSIM is indicated as ISRGAN_{SSIM}. The mean and the variance of the reconstructed image are indicated

as μ_p and σ_p^2, and the mean and variance of the original image are indicated as μ_q and σ_q^2. The covariance between the original and reconstructed image is notated as σ_{pq}, and the variable utilized for the stabilization is indicated as s_1 and s_2, respectively.
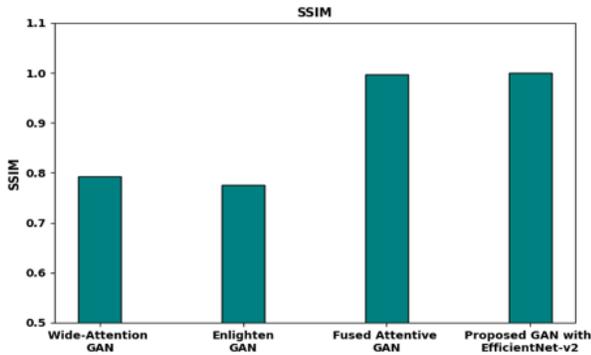

Fig. 8. Analysis based on SSIM

The outcome of the proposed method, along with its comparative methods, is portrayed in Fig. 8. The higher SSIM evaluated by the proposed method depicts the similarity of the reconstructed super resolution image with the original image.

3) Comparative discussion: The comparative analysis of the proposed method, along with the conventional image reconstruction techniques, is portrayed in Table IV. The MSE acquired by the proposed method is 0.001025, which is 52.93%, 70.99%, and 35.54% minimum compared to the traditional Wide-Attention GAN, Enlighten GAN, and Fused Attentive GAN methods. Similarly, the proposed method accomplished the minimal MAE of 0.00259, which is 59.09%, 67.08%, and 13.13% minimum compared to the traditional Wide-Attention GAN, Enlighten GAN, and Fused Attentive GAN methods. In contrast, the proposed method acquired the maximal PSNR of 45.5698, which is 40.26%, 49.89%, and 15.34% better compared to conventional Wide-Attention GAN, Enlighten GAN, and Fused Attentive GAN methods. Likewise, the maximal SSIM measured by the proposed method is 0.9997, which is 20.62%, 22.32%, and 0.18% superior compared to state-of-the-art Wide-Attention GAN [29], Enlighten GAN [30], and Fused Attentive GAN [28] methods.

TABLE IV. COMPARATIVE ANALYSIS

Method	MSE	PSNR	SSIM	MAE
Wide-Attention GAN [29]	0.002188	27.2254	0.7936	0.0055
Enlighten-GAN [30]	0.00355	22.834	0.77656	0.006835
Fused Attentive GAN [28]	0.001598	38.58	0.9979	0.00259
Proposed Method	0.001025	45.56985	0.9997	0.00225

The proposed method incorporates the EfficientNet-v2 with the traditional GAN for enhancing the image super resolution quality by minimizing the perceptual loss. For this, the EfficientNet-v2 is utilized at the discriminator that minimizes the information loss with fast learning rate. Thus

The Fused GAN was devised by integrating the attention module in the traditional GAN for enhancing the performance of the image super resolution technique. The quality assessment based on the similarity measure was closer to the proposed method; still, the higher time complexity and failure in maintaining the system parallelism mechanism limits the method. Thus, the existing Fused GAN acquired little minimal performance compared to the proposed GAN with EfficientNet-v2. Followed by, Wide Attention GAN devised by integrating the self-attention module. Due to the inclusion of the attention mechanism, the introduced Wide Attention GAN also suffered with higher time complexity. Finally, the Enlighten-GAN that was designed by incorporating enlighten blocks into the traditional GAN. Here, the introduced method accomplished better quality image as outcome by solving the unstable convergence issue through the enlighten block. However, method is inefficient in processing the image with complex background. Thus, the analysis based on the error estimation and the similarity analysis shows that the proposed method accomplished better performance than other super image resolution methods. The training of the discriminator using the EfficientNet-v2 provides fast learning and the least loss function based on the perception enhancing the image's visual quality.

VII. CONCLUSION

The SR image reconstruction from the LR image for the acquisition of fine-grained information from the image, along with the visual quality, is attained using ISRGAN with EfficientNet-v2 in this paper. The proposed image reconstruction technique elevates the quality of the image by minimizing the perception loss present in the GAN network. The information loss is minimized through the efficient learning of the discriminator using the EfficientNet-v2, which has the probability of a fast learning rate with more accurate learning based on the perception loss generated by the generator module of the ISRGAN. The learning capability of the EfficientNet-v2 is higher by making the original interference size for learning that depicts the scaling characteristics of the loss function estimator. Thus, the proposed method provides a promising outcome based on the error analysis and acquires the minimal values of loss of 0.02 at the generator, 0.1 at the discriminator, and 0.015 at the self-supervised learning, respectively, with a batch size of 32. Besides, the minimal MSE and MAE accomplished by the proposed method are 0.001025 and 0.00225, respectively. Likewise, the maximal PSNR and SSIM acquired by the proposed method are 45.56985 and 0.9997, respectively. However, the loss of the method further needs to be reduced for real-time processing.

In the future, we plan to introduce a novel hybrid meta-heuristic optimization-based deep learning approach to overcome the challenges the proposed model encountered in this study.

REFERENCES

[1] J. Xiu, X. Qu, and H. Yu, "Face super-resolution using recurrent generative adversarial network," in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, vol. 6. IEEE, 2022, pp. 1169–1174.

[2] X. Lin, X. Zhou, T. Tong, X. Nie, L. Wang, H. Zheng, J. Li, E. Xue, S. Chen, M. Zheng *et al.*, "A super-resolution guided network for improving automated thyroid nodule segmentation," *Computer Methods and Programs in Biomedicine*, p. 107186, 2022.

[3] A. Abedjooy and M. Ebrahimi, "Multi-modality image super-resolution using generative adversarial networks," *arXiv preprint arXiv:2206.09193*, 2022.

[4] S. Altakrouri, S. B. Usman, N. B. Ahmad, T. Justinia, and N. M. Noor, "Image to image translation networks using perceptual adversarial loss function," in *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2021, pp. 89–94.

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

[6] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[7] Y. Li, B. Sixou, and F. Peyrin, "A review of the deep learning methods for medical images super resolution problems," *IRBM*, vol. 42, no. 2, pp. 120–133, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1959031820301408

[8] B. Meng, L. Wang, Z. He, G. Jeon, Q. Dou, and X. Yang, "Gradient information distillation network for real-time single-image super-resolution," *J. Real-Time Image Process.*, vol. 18, no. 2, p. 333–344, apr 2021. [Online]. Available: https://doi.org/10.1007/s11554-021-01083-1

[9] J. Hou, Y. Si, and X. Yu, "A novel and effective image super-resolution reconstruction technique via fast global and local residual learning model," *Applied Sciences*, vol. 10, no. 5, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/5/1856

[10] Y. Sun, R. Wang, D. Cao, and R. Lee, "Who are social media influencers for luxury fashion consumption of the chinese gen z? categorisation and empirical examination," *Journal of Fashion Marketing and Management: An International Journal*, 2021.

[11] X. Liu, J. Li, T. Duan, J. Li, and Y. Wang, "Dsma: Reference-based image super-resolution method based on dual-view supervised learning and multi-attention mechanism," *IEEE Access*, 2022.

[12] Z. Lu and Y. Chen, "Single image super-resolution based on a modified u-net with mixed gradient loss," *signal, image and video processing*, vol. 16, no. 5, pp. 1143–1151, 2022.

[13] P. Andreini, G. Ciano, S. Bonechi, C. Graziani, V. Lachi, A. Mecocci, A. Sodi, F. Scarselli, and M. Bianchini, "A two-stage gan for high-resolution retinal image generation and segmentation," *Electronics*, vol. 11, no. 1, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/1/60

[14] K. M. Peddecord, "Behold the future: Implications of automated image analysis for the photographic arts," *The International Journal of the Image*, vol. 13, no. 1, p. 47–62, 2022.

[15] K. Varma, G. S. Reddy, and N. Subramanyam, "Face image super resolution using a generative adversarial network," in *2021 Smart Technologies, Communication and Robotics (STCR)*. IEEE, 2021, pp. 1–8.

[16] T. Shinohara, R. Ito, Y. Kobayashi, T. Satoh, Y. Shimazaki, and S. Nakamura, "Application of pre-trained real-world super resolution models to optical satellite image," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 167–173, 2022.

[17] T.-A. Song, S. R. Chowdhury, F. Yang, and J. Dutta, "Pet image super-resolution using generative adversarial networks," *Neural Networks*, vol. 125, p. 83–91, 2020.

[18] Y. Tao, S. J. Conway, J.-P. Muller, A. R. Putri, N. Thomas, and G. Cremonese, "Single image super-resolution restoration of tgo cassis colour images: Demonstration with perseverance rover landing site and mars science targets," *Remote sensing*, vol. 13, no. 9, p. 1777, 2021.

[19] Y. Kim and D. Son, "Noise conditional flow model for learning the super-resolution space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 424–432.

[20] H. Y. Park, H.-J. Bae, G.-S. Hong, M. Kim, J. Yun, S. Park, W. J. Chung, and N. Kim, "Realistic high-resolution body computed tomography image synthesis by using progressive growing generative adversarial network: visual turing test," *JMIR Medical Informatics*, vol. 9, no. 3, p. e23328, 2021.

[21] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine Learning*. PMLR, 2021, pp. 10 096–10 106.

[22] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 184–199.

[23] B. Liu and J. Chen, "A super resolution algorithm based on attention mechanism and srgan network," *IEEE Access*, vol. 9, pp. 139 138–139 145, 2021.

[24] S. Kim, D. Jun, B.-G. Kim, H. Lee, and E. Rhee, "Single image super-resolution method using cnn-based lightweight neural networks," *Applied Sciences*, vol. 11, no. 3, p. 1092, 2021.

[25] Y. Xiong, S. Guo, J. Chen, X. Deng, L. Sun, X. Zheng, and W. Xu, "Improved srgan for remote sensing image super-resolution across locations and sensors," *Remote Sensing*, vol. 12, no. 8, p. 1263, 2020.

[26] X. Liu, S. Chen, L. Song, M. Woźniak, and S. Liu, "Self-attention negative feedback network for real-time image super-resolution," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 6179–6186, 2022.

[27] X. Zhu, L. Zhang, L. Zhang, X. Liu, Y. Shen, and S. Zhao, "Gan-based image super-resolution with a novel quality loss," *Mathematical Problems in Engineering*, vol. 2020, 2020.

[28] M. Jiang, M. Zhi, L. Wei, X. Yang, J. Zhang, Y. Li, P. Wang, J. Huang, and G. Yang, "Fa-gan: Fused attentive generative adversarial networks for mri image super-resolution," *Computerized Medical Imaging and Graphics*, vol. 92, p. 101969, 2021.

[29] F. Shahidi, "Breast cancer histopathology image super-resolution using wide-attention gan with improved wasserstein gradient penalty and perceptual loss," *IEEE Access*, vol. 9, pp. 32 795–32 809, 2021.

[30] Y. Gong, P. Liao, X. Zhang, L. Zhang, G. Chen, K. Zhu, X. Tan, and Z. Lv, "Enlighten-gan for super resolution reconstruction in mid-resolution remote sensing images," *Remote Sensing*, vol. 13, no. 6, p. 1104, 2021.

[31] N. Ahn, B. Kang, and K.-A. Sohn, "Efficient deep neural network for photo-realistic image super-resolution," *Pattern Recognition*, vol. 127, p. 108649, 2022.

[32] H. Song, M. Wang, L. Zhang, Y. Li, Z. Jiang, and G. Yin, "S2rgan: sonar-image super-resolution based on generative adversarial network," *The Visual Computer*, vol. 37, no. 8, pp. 2285–2299, 2021.

[33] L. Sun, Z. Liu, X. Sun, L. Liu, R. Lan, and X. Luo, "Lightweight image super-resolution via weighted multi-scale residual network," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 7, pp. 1271–1280, 2021.

[34] B. Li, A. Keikhosravi, A. G. Loeffler, and K. W. Eliceiri, "Single image super-resolution for whole slide image using convolutional neural networks and self-supervised color normalization," *Medical Image Analysis*, vol. 68, p. 101938, 2021.

[35] "The cifar-10 dataset," (Date last accessed 20-November-2014). [Online]. Available: https://www.cs.toronto.edu/ kriz/cifar.html