SRGAN Enhancement through Autoencoder-Pretrained U-Net with Residual Blocks for Improved Image Super-Resolution

Amirreza Rouhbakhshmeghrazi[1•], Bo Li[1], Shayan Nalbandian[2], Chao Song[1], Ghazal Alizadeh[3] and Mohammad Reza Hassannezhad [3]

***Abstract*-- Super-resolution is a crucial task in image processing, enhancing the resolution of low-quality images for applications such as surveillance, remote sensing, and autonomous systems. Traditional methods often struggle to preserve fine details, leading to artifacts and reduced visual fidelity. This study introduces the Pretrained RU-SRGAN, an enhanced Super-Resolution Generative Adversarial Network (SRGAN) that incorporates U-Net architecture, residual learning, and autoencoder pretraining to improve both image quality and computational efficiency, particularly in resource-constrained environments like UAVs. The goal of this research is to evaluate how these architectural modifications can enhance super-resolution performance with limited data. Autoencoder pretraining enables the generator to leverage learned features from low-resolution images, accelerating convergence and improving high-resolution reconstructions. Experimental results show that Pretrained RU-SRGAN outperforms baseline models, achieving a PSNR of 25.7 dB and an SSIM of 0.83. These results highlight the model's ability to preserve fine details and structural integrity, making it particularly effective for real-time image enhancement in UAV applications. The Pretrained RU-SRGAN provides a robust solution for super-resolution tasks, balancing high-quality image reconstruction with computational efficiency, and is well-suited for practical deployment in dynamic, resource-limited environments.**

***Index Terms*- Transfer Learning, UAV, Image Reconstruction, RU-Net, SSIM, PSNR**

I. INTRODUCTION

A significant quantity of data generated in our modern world consists of images, including photos captured by individuals and high-resolution images collected by satellites and security cameras. However, resolution loss is frequently encountered, making it very challenging to interpret and utilize data. Resolution loss refers to the decrease in sharpness and level of detail caused by a decline in the quality of an image. This decline could be attributed to various factors, such as compression techniques for storage, limitations in software or hardware, or environmental conditions during image capture. The most effective method to address this problem is through image reconstruction. Enhancing the visualization effect, clarity, and details of images using high-quality super-resolution reconstruction can lead to improved accuracy and reliability in target recognition [1]. Initially, methods such as interpolation [2], frequency-domain techniques [3], wavelet transform [4], [5], neighbor embedding [6], and sparse representation [7], [8] were employed to enhance the resolution in image processing. Later, machine learning techniques, such as random forest, were utilized to analyze low-quality images and predict high-quality characteristics [9].

There has been a rise in the number of super-resolution reconstruction models in the field of deep learning, incorporating techniques like convolutional neural networks (CNN) for image super-resolution. For instance, the super-resolution convolutional neural network (SRCNN) [10], very deep super-resolution (VDSR) [11], and the super-resolution generative adversarial network, known as SRGAN [12]. Generative adversarial network (GAN) models have been extensively used in the context of super-resolution reconstruction problems. SRGAN is a sophisticated GAN model that enhances image clarity and realism by training a discriminator network and a generator network adversarially.

SRGAN shows great promise in various practical applications by enhancing image quality and recovering missing details [13], [14]. Recently, researchers have utilized SRGAN to improve the quality of input images for classification purposes. The increased resolution from SRGAN enhances the quality of features available for classification models, resulting in better accuracy in tasks such as medical image classification [15], facial recognition [16], and scene classification [17]. Moreover, SRGAN has been employed to enhance images in the initial phase of object detection projects. By increasing the clarity of input pictures, SRGAN improves the feature maps used in detectors, leading to enhanced performance in object detection tasks, particularly in low-quality images for detecting small objects, which in turn boosts the precision of object detection algorithms like YOLO and Faster R-CNN [18]. Furthermore, SRGAN is applied in tasks such as semantic segmentation and image generation within image-to-image translation projects. The improved resolution helps capture intricate details needed for these endeavors. Finally, SRGAN has been studied in self-driving systems to enhance low-resolution images from cameras, increasing the accuracy of object recognition and traffic sign detection [19].

1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an, Shaanxi, China.
• amirreza.rouhbakhshmeghrazi@gmail.com
2. School of Software Engineering, Northwest Polytechnic University, Xi'an, China.
3. School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi, China.

While the SRGAN model boasts impressive capabilities in super-resolution tasks, it is not without its limitations. Common challenges include noticeable distortions, visual artifacts, and a lack of critical fine details, which can compromise its effectiveness in certain applications. To address these shortcomings, ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) builds upon the original SRGAN framework by introducing a sophisticated loss function that prioritizes perceptual quality over mere pixel-wise accuracy [20]. A key innovation in ESRGAN is the integration of Residual-in-Residual Dense Blocks (RRDB), which significantly enhance feature extraction. These blocks not only boost image fidelity but also effectively minimize artifacts, resulting in more natural and realistic high-resolution images.

Taking this progression further, Real-ESRGAN (Real Enhanced Super-Resolution Generative Adversarial Networks) elevates the model's performance by focusing specifically on the challenges of real-world image degradation. Through an advanced training process and optimized architecture, Real-ESRGAN adeptly handles common image artifacts, delivering enhanced visuals that are not only strikingly realistic but also free from typical GAN-induced imperfections [21].

Despite the advancements in SRGAN and its enhanced variants, several persistent challenges remain unaddressed, limiting their effectiveness. Notable issues include mode collapse and vanishing gradients during training, arising from the adversarial nature of GANs. These problems stem from the delicate balancing act between the generator and discriminator networks; if one overpowers the other, the training process converges to suboptimal results, often yielding poor-quality outputs. Moreover, GAN-based models are notoriously data-hungry, requiring extensive datasets to achieve robust performance, which can be a significant limitation when working with specialized or resource-constrained applications.

To overcome these challenges, we propose an upgraded SRGAN model that incorporates a series of innovative enhancements aimed at preserving critical image characteristics while delivering superior results.

1) Redesigning the Generator Architecture with U-Net: We replace the original generator's residual blocks with a U-Net architecture, known for its exceptional ability to capture intricate details and spatial relationships. This modification enhances spatial resolution recovery while reducing the model's overall complexity. By streamlining the network, we mitigate the risk of overfitting and enable the model to focus on fundamental low-level features essential for accurate image reconstruction, particularly in challenging domains like facial reconstruction.

2) Integrating Residual Blocks within the U-Net: Our model incorporates residual blocks into both the encoder and decoder paths of the U-Net, creating a hybrid design that balances the benefits of U-Net and residual learning. This structure enhances the model's perceptual capabilities, allowing it to retain hierarchical feature details while stabilizing the network. The residual blocks also help to reduce visual artifacts and distortions, resulting in more natural and realistic high-resolution images.

3) Leveraging Transfer Learning with an Autoencoder Pretraining Step: To address the data hunger issue and improve the efficiency of training, we employ transfer learning. An autoencoder is initially trained on low-resolution (LR) images to learn a compressed representation of the LR feature space. The learned weights from the autoencoder's encoder are then transferred to the U-Net's encoder path. This initialization strategy offers several advantages:

- It accelerates convergence by starting the training process with a network already familiar with the distribution of LR image features.
- It boosts performance, particularly on datasets with limited training samples or significant variability, such as UAV-captured imagery.
- It ensures that key characteristics of LR images are effectively preserved, enhancing the model's reconstruction accuracy.

The novelty of our approach lies not only in the combination of U-Net with residual learning and transfer learning but also in how these techniques are synergistically applied to super-resolution tasks for datasets with limited resources or variability, particularly UAV imagery. This hybrid architecture is introduced for the first time in our work and addresses key challenges in the super-resolution domain. The design's streamlined architecture, coupled with its ability to achieve fast convergence, makes it particularly suitable for real-time applications such as UAVs and autonomous vehicles, where computational efficiency and speed are critical.

Our approach also effectively addresses the issue of low dataset availability by leveraging autoencoder pretraining to extract salient features from low-resolution images. This pretraining step not only boosts performance on datasets with limited training samples but also enhances the model's ability to generalize to highly variable data, ensuring robustness in real-world applications.

We evaluated the effectiveness of our proposed enhancements using a dataset of UAV-captured images. UAV datasets often pose unique challenges, including small object sizes and inherent low-resolution qualities, making them an ideal testbed for super-resolution models. The experimental results demonstrate that our upgraded SRGAN outperforms the original SRGAN in terms of image fidelity and feature preservation. The enhanced clarity and detail provided by our model lead to significant improvements in downstream tasks, such as object detection, where high-quality inputs are critical.

Our proposed model not only addresses the common limitations of SRGAN and its variants but also introduces a novel hybrid architecture that bridges the gap between data

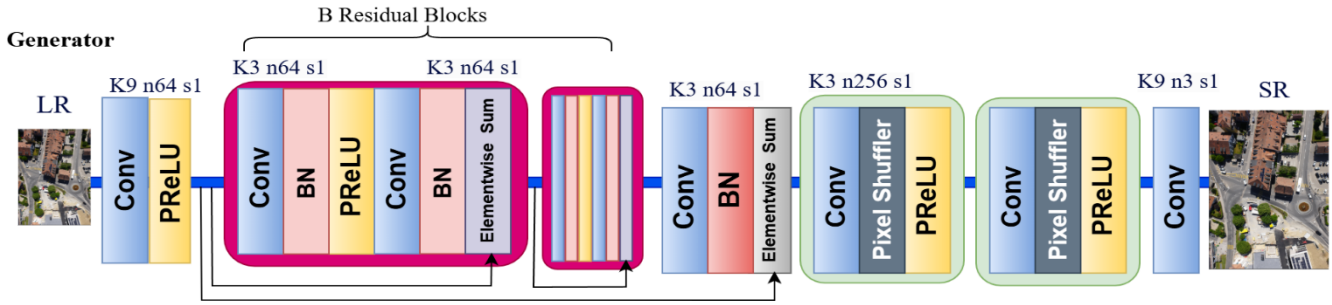

Fig. 1. The original architecture of generator in the SRGAN

efficiency and performance. By combining U-Net's detailed feature extraction capabilities with residual learning and leveraging transfer learning, we provide a robust and scalable solution for super-resolution tasks, especially in resource-constrained or high-variability datasets like UAV imagery. These advancements set a new benchmark for practical super-resolution applications in fields ranging from surveillance and remote sensing to medical imaging and autonomous systems.

II. LITERATURE REVIEW

Ever since its inception in 2017, the SRGAN [22] model has been widely used in numerous research projects due to its ability to enhance image quality for various purposes [23], [24], [25]. The primary structure of SRGAN, which consists of a generator and a discriminator, has proven successful in generating high-quality super-resolution images. In this section, we will explore the core design principles of SRGAN's generator and discriminator, providing detailed insight into its fundamental structure. GANs, a deep learning framework, comprise a generator and a discriminator as its two main components. Both of these neural networks are trained simultaneously in a competitive manner. The generator network is trained to produce fake data that resembles the input data distribution, while the discriminator network functions as a binary classifier to differentiate between real and generated data. Fig. 1. illustrates the design of the generator, which includes three primary elements: Convolutional layers, Residual blocks, and Upsampling blocks. Initially, the LR image is processed by the network and subsequently passes through a convolutional layer to create a feature map. It then goes through a PReLU (parametric rectifier linear unit) activation function [26]. The parametric ReLU allows negative values to have a negative slope instead of being set to zero. This enables the network to gather information from both positive and negative values, allowing for better feature extraction.

Following this, the picture passes through 16 residual blocks, which include two convolutional layers, batch normalization, PReLU, and skip connections. Residual blocks help the network capture image details, while skip connections preserve feature flow between blocks, unaffected by vanishing gradients. After passing through the residual blocks, the input undergoes a convolutional layer, batch normalization, and an Elementwise sum block before moving on to two upsampling sections, which lead to an output image that is four times higher in resolution compared to the input image. The progressive enhancement of resolution is essential for improving image quality and producing high-quality images with fewer parameters. Each upscaling module consists of a convolutional layer, PixelShuffler, and PReLU. The PixelShuffler [27] is responsible for increasing the LR size by two times in every block and then by four times before sending it to the last convolutional layer. The resulting image should resemble an accurate depiction of the high-resolution image.

SRGAN employs a traditional discriminator network, with the design of SRGAN's discriminator network illustrated in Fig. 2. The primary elements of this network consist of convolutional layers, Leaky ReLU activation functions, and Batch Normalization (BN). The discriminator function is trained to distinguish between super-resolution images and real images. The input contains a high-quality image produced by the network or an image taken from the training dataset. Like the generator, it contains several convolutional layers for extracting features. The first step includes a convolutional layer followed by a Leaky ReLU activation. The next step consists of eight groups of three layers, with a convolutional layer, BN, and Leaky ReLU activation in each. The Leaky ReLU function has a parameter of 0.2, and Batch Normalization layers are intentionally added to accelerate network training and enhance overall generalization abilities. A dense layer is necessary to transform the multi-dimensional feature maps of the input image into a single-dimensional vector for classification. Next,

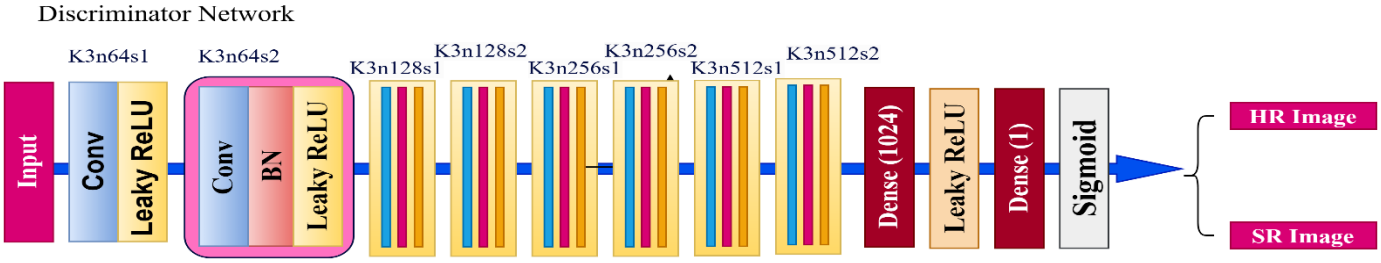

Fig. 2. The basic design of discriminator in SRGAN

the vector goes through an activation layer and is then passed through another dense layer to convert the 1024-sized one-dimensional vector into a single one. The Sigmoid function is employed to transform the input value into 0 or 1, denoting binary categories. The discriminator's results range from 0 to 1, with values closer to 0 indicating fake images and values closer to 1 indicating real images.

The loss of information in SRGAN is calculated based on feature maps from a previously trained deep neural network, usually the VGG network. Content loss, in contrast to mean squared error (MSE), is based on high-level image representations rather than pixel intensity differences, making it less reliant on exact pixel values. These depictions are derived from middle stages of the VGG model, which are better able to handle spatial differences. Therefore, the loss of content can more accurately represent the quality of an image by emphasizing the important high-level features that matter most to human visual perception. In SRGAN, the overall loss function is a combination of two parts: content loss for generator guidance, and adversarial loss for discriminator utilization. The adversarial loss plays a key role in differentiating genuine high-resolution images from those produced by the model, ultimately motivating the generator to produce increasingly convincing images that fool the discriminator.

Content loss, known as VGG loss, is a crucial element in the SRGAN, as it is calculated from the output of the generator. Content loss determines image quality by measuring perceptual similarity. This method depends on comparing high-level features of the generated image with the ground truth, ensuring that the generated image retains crucial structural and textural details similar to how humans perceive visual information. To determine content loss, a pre-trained deep neural network, usually VGG or ResNet, processes both the generated image and the ground truth image, producing feature maps. These maps represent images at a high level, capturing edges, textures, and object structures. Content loss is described as the Euclidean distance between the feature maps of the generated and ground truth images, measuring the discrepancy of these higher-order features.

Although different pre-trained networks can be utilized to extract these feature maps, VGG19 is frequently used, as it has been pre-trained on the ImageNet dataset. Nevertheless, studies indicate that better outcomes are achieved with deeper network layers, as they pay more attention to the small and detailed aspects of the image [28]. Using features from deeper layers like VGG54 usually results in superior perceptual quality compared to using earlier layers such as VGG19. The benefit of the deeper layers is that they are able to capture more complex and abstract image representations, resulting in sharper and more true-to-life images. When employing VGG19 as a pre-trained model, content loss is commonly calculated using a group of middle layers, often omitting the final max pooling layer. This occurs because the innermost layers of VGG retain a high-level semantic understanding without significantly decreasing the spatial resolution.

A. Related Works

In this section, we will discuss new advancements and changes in SRGAN and its variations that have led to improved visual quality and image clarity. Ren and Li [29] in 2021 introduced methods that include a parallel generative adversarial network structure using attention mechanism and multi-scale feature fusion within the SRGAN framework. This method combines a dual generator and discriminator model with an attention module in order to understand multi-scale features and incorporate high-frequency data across various scales in the residual network. The results from the experiment show that this technique greatly enhances the restoration of image details, as indicated by its performance on the benchmark datasets Set5, Set14, and BSD100. The suggested approach demonstrates a significant enhancement in visual and quantitative metrics, obtaining increased PSNR and SSIM values in contrast to conventional approaches such as Bicubic, SRCNN, SRResNet, and SRGAN.

In 2021, Yin Wang presented a new image super-resolution model named U-Net SRGAN, designed to enhance the perceptual quality of high-resolution images produced from low-resolution inputs. This model improves upon existing methods such as SRGAN and ESRGAN with several important upgrades. The Residual-in-Residual Self-Calibrated Convolution with Pixel Attention (RRSCPA) block in the generator is more efficient and effective at capturing details than prior architectures. The discriminator employs a U-Net design, offering feedback at a per-pixel level to assist the generator in creating more lifelike images. The conventional VGG-based perceptual loss is replaced with LPIPS (Learned Perceptual Image Patch Similarity) loss, enhancing alignment with human visual perception and elevating the quality of the generated images. The U-Net SRGAN outperforms other models, including SRGAN and ESRGAN, in visual quality and LPIPS scores, as demonstrated in experimental results [30].

In 2023, Hrishikesh et al. presented a new deep network structure incorporating a V-SRGAN based on Relativistic Average GAN (RaGAN) and VGG19 architecture. The method emphasizes improving visual clarity and recovering texture in high-resolution images. Important advancements consist of incorporating multi-scale receptive field blocks (RFBs) in the generator for capturing more intricate texture details and a lightweight design with smaller kernels to decrease computational complexity. They obtained better outcomes than current cutting-edge techniques, as indicated by PSNR and LPIPS metrics [31].

In 2023, Wu unveiled an upgraded model of SRGAN designed to enhance image quality. Major changes involve eliminating Batch Normalization layers, adding a novel Residual Block with attention mechanisms for better feature extraction, and streamlining loss functions to only two components - Adversarial Loss and L1 Loss. These improvements result in superior feature extraction, enhanced face restoration, and artifact removal, producing images of higher visual quality that better match human perceptual standards, even with a slight decrease in PSNR. The findings provide important perspectives for progressing research on image super-resolution [32].

Finally, Sun et al. introduced TESRGAN, an innovative model for image super-resolution that integrates Transformer architecture with CNNs. It uses a Residual in Residual Dense Block Network (RRDBNet) for feature extraction and a Dense Residual Transformer module to capture global dependencies

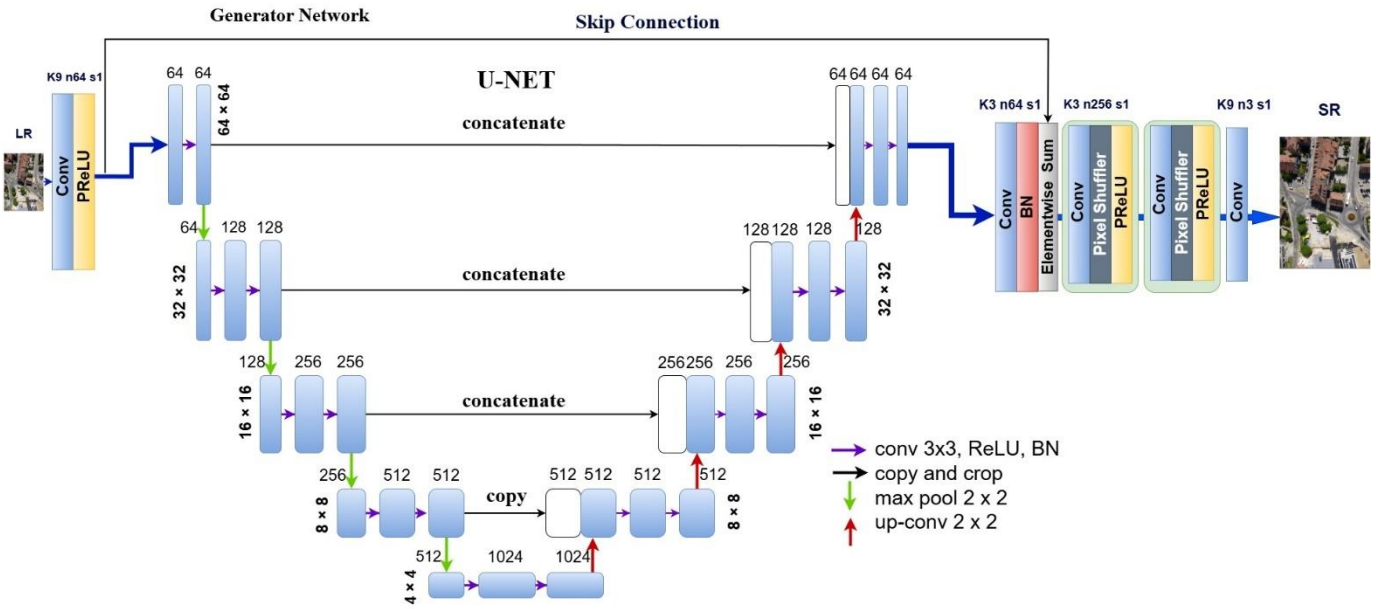

Fig. 3. The Generator in U-SRGAN: The residual blocks of the SRGAN architecture are replaced by a U-Net structure, resulting in an enhanced design.

and complex textures. TESRGAN demonstrates superior performance with a Peak Signal-to-Noise Ratio (PSNR) of 31.67 dB and a Structural Similarity Index Measure (SSIM) of 0.88, outperforming ESRGAN in convergence rates and stability. The study emphasizes the potential of hybrid deep learning architectures in enhancing image processing tasks [33].

III. METHODOLOGY

To enhance the efficiency of SRGAN, we proposed a series of modifications to the generator design while retaining the original discriminator architecture. The discriminator's binary classification task was deemed well-suited to the super-resolution context, and thus, its architecture remained unchanged throughout our experiments. Our primary focus was on testing various generator designs to assess their individual contributions to performance improvements. In this study, we systematically introduced and evaluated three distinct models to facilitate an ablation analysis, isolating the impact of each design change:

1) U-SRGAN: The residual blocks in the generator were replaced with a U-Net architecture. This modification aimed to improve the model's ability to capture intricate details and enhance spatial resolution recovery.

2) RU-SRGAN: Building upon U-SRGAN, residual blocks were incorporated within the U-Net architecture to leverage the strengths of both residual learning and U-Net's detailed feature extraction capabilities. This hybrid design was intended to stabilize the network and further improve image fidelity.

3) Pretrained RU-SRGAN: To address the data-hungry nature of GAN-based models and enhance performance, the U-Net generator in RU-SRGAN was pretrained using an autoencoder on low-resolution images. This pretrained encoder provided a robust initialization, enabling faster convergence and more accurate reconstruction, particularly in datasets with limited or challenging samples.

These models were introduced and systematically evaluated to provide a comprehensive understanding of how each architectural enhancement contributes to the overall performance. The results of this ablation study are detailed in the following sections.

A. Model Description

1) U-SRGAN

Originally created for biomedical image segmentation [34], [35], the U-Net architecture has gained popularity across various computer vision tasks for its effective handling of segmentation issues and its ability to capture local and global features. U-Net was developed with a focus on situations where there is a lack of training data. We chose U-Net due to limited data, as it efficiently utilizes information through augmentation and network structure design. Skip connections enable U-Net to merge low-level features, such as edges and textures from early layers, with high-level abstract features like shapes and patterns from deeper layers. This guarantees that intricate elements are preserved in the final segmentation, even in complex images. The structure of U-Net is symmetrical, as the encoder captures features from the input image while the decoder reconstructs the segmentation map. This balance allows it to effectively learn both local and global contexts simultaneously, which is essential for pixel-wise segmentation tasks. In our research, we removed the segmentation head from the U-Net and solely utilized the feature extraction capabilities of the network. Fig. 3. shows how we implemented U-Net in the generator architecture.

The revised SRGAN's generator network utilizes a U-Net-style design specifically designed for super-resolution purposes, increasing the input image size by a factor of 4. The network takes in a small 64×64×3 image and transforms it into

a larger 256×256×3 image through gradual processing. The design combines hierarchical feature extraction and skip connections for successful image super-resolution. The generator starts with a starting convolutional layer (9x9 kernel, 64 filters, stride 1) and then uses a Parametric ReLU (PReLU) activation function to identify basic features. The encoder section of the U-Net gradually reduces the spatial dimensions while simultaneously increasing the feature map size. Every encoder block comprises of convolutional layers with a 3x3 kernel, batch normalization, and ReLU activation, then followed by 2x2 max-pooling for downsampling. The encoder captures image features at different scales and resolutions, ranging from 32×32 to 4×4, while increasing the number of feature maps from 64 to 512.

During the bottleneck phase, 512 feature maps are generated by convolutional layers to extract deep features. The decoder mirrors the encoder, gradually increasing the spatial dimensions until they reach their original size. Up-convolutions of size 2x2 are utilized for upsampling in the transposed convolutional layers, and connections from the encoder are combined with the decoder at every level. These shortcuts maintain spatial information and help regain lost features from downsampling, enhancing the quality of the reconstruction. The decoder's output is further refined through residual learning blocks and pixel shufflers. The pixel shuffler layers upscale images using a subpixel technique to enhance image quality by preserving smooth transitions and detailed features in the super-resolved image. The high-resolution image is produced by the last convolutional layer with a 9x9 kernel. Residual learning is applied in certain stages of the network through element-wise addition to prioritize the reconstruction of intricate details such as textures and edges. Overall, the design effectively merges U-Net's hierarchical feature extraction and skip connections with super-resolution-specific enhancements like pixel shuffling and residual learning, enabling it to generate high - quality super-resolved images.

2) RU-SRGAN

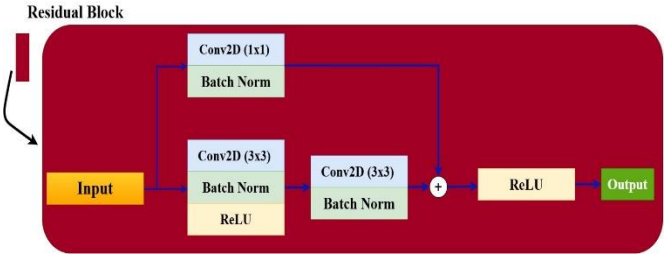

Fig. 4. A residual block used in the architecture of our RU-SRGAN

Utilizing residual blocks in U-Net enhances performance by addressing issues of vanishing gradients and feature recycling. Residual connections allow the transfer of low-level and high-level features between layers, facilitating the efficient learning of complex patterns within the network. This is particularly beneficial in U-Net, as skip connections already retain spatial information, while residual blocks further improve fine detail reconstruction by focusing on learning the residual rather than the complete mapping. This leads to quicker convergence, enhanced gradient flow, and more accurate reconstructions, especially for high-frequency details in tasks such as super-resolution. Adding more layers in deep CNNs can result in a decrease in training accuracy due to vanishing gradients. Residual connections solve this problem by passing the input of each block through a one-by-one convolutional layer and occasionally a BN layer to the output, followed by ReLU activation. This enables faster propagation of each block's input through the residual connections, producing a neural network with fewer parameters while maintaining similar or improved performance. RU-Net, a U-Net that incorporates residual blocks, has achieved success in various tasks [36], [37], [38].

Fig. 4. shows the residual block in the U-Net design to improve gradient flow and enhance learning efficiency in our

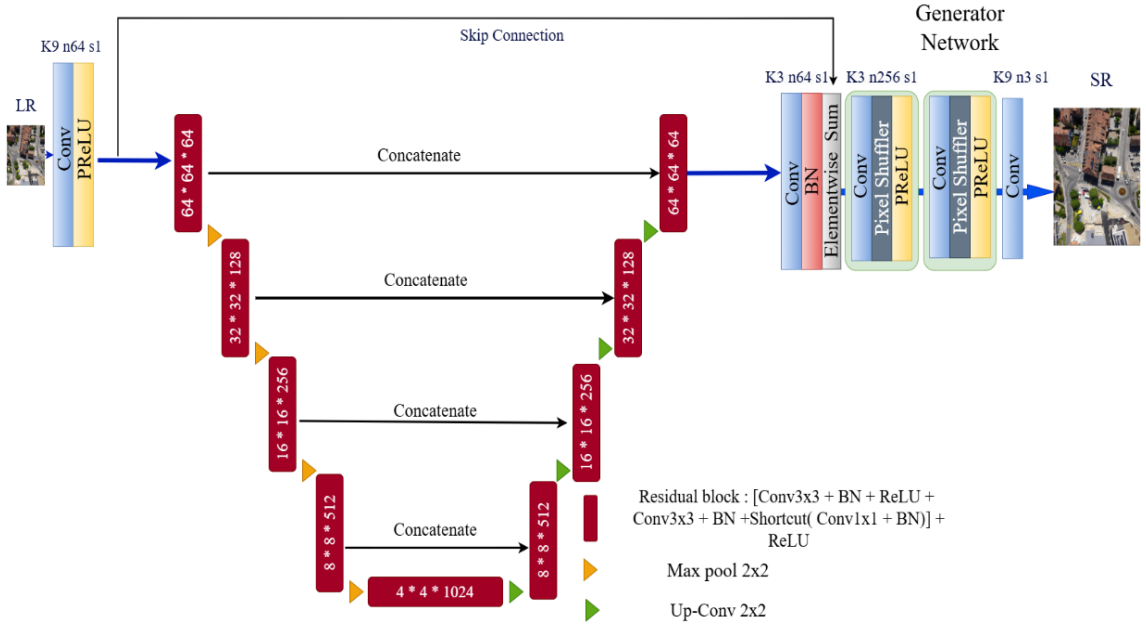

Fig. 5. The Generator of RU-SRGAN

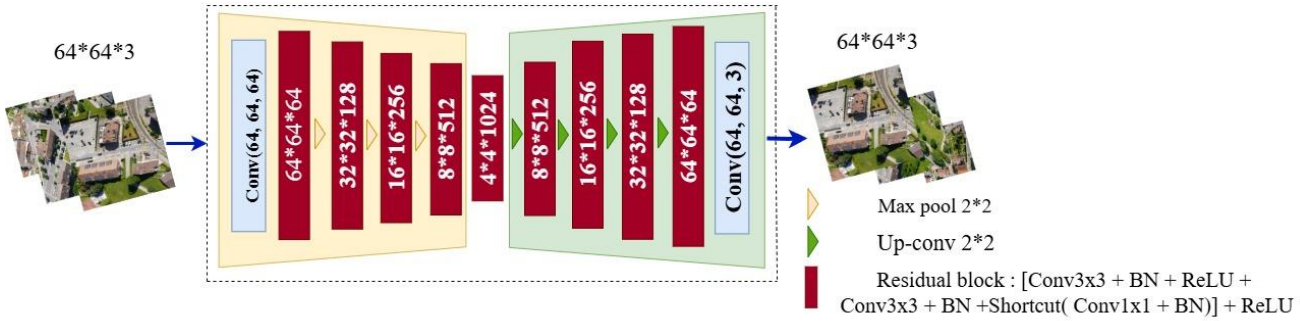

Fig. 6. The architecture of proposed autoencoder designed to reconstruct the low-resolution images of our dataset

generator architecture. The block is made up of two primary paths: a sequential path containing convolutional layers and a shortcut connection that skips the convolutions. The consecutive journey starts with a 1x1 convolutional layer, succeeded by batch normalization. The aim of the 1x1 convolution is to decrease or align the size of the input feature maps, guaranteeing they can work with the residual connection. This is succeeded by a 3x3 convolutional layer, which handles spatial details, followed by another layer for batch normalization. After the 3x3 convolution, a ReLU activation function is used to bring in non-linearity, helping the block understand intricate patterns and connections within the data.

Fig. 5. illustrates the modified U-Net-based architecture with residual blocks proposed for RU-SRGAN model. The system processes a LR picture as an input and produces a HR result, enlarging the input by a factor of 4. The structure merges U-Net's hierarchical feature extraction abilities with the effective gradient flow and learning stability offered by residual connections. The combination of U-Net's skip connections with residual blocks and GAN training objectives creates a highly effective RU-SRGAN for generating detailed and visually realistic super-resolved images. Similar to the previous version, U-SRGAN, we kept the discriminator of SRGAN unchanged (see Fig. 2).

3) Pretrained RU-Net

The Autoencoder and U-Net fusion is a new deep learning approach put forward for segmentation and classification. The U-Net model architecture and Autoencoders were selected after analyzing various deep learning-based model architectures [39], [40], [41]. The design of the model offers a hopeful approach for capturing important characteristics, particularly in tasks like image segmentation, with improved effectiveness, reduced overfitting, and increased ability to generalize. An unsupervised artificial neural network known as an Autoencoder consists of two parts: an encoder and a decoder. The encoder in autoencoder learns to compress and encode data effectively, while the decoder part learns to reconstruct the data from the encoded representation to match the original input as closely as possible. The primary concept when constructing an

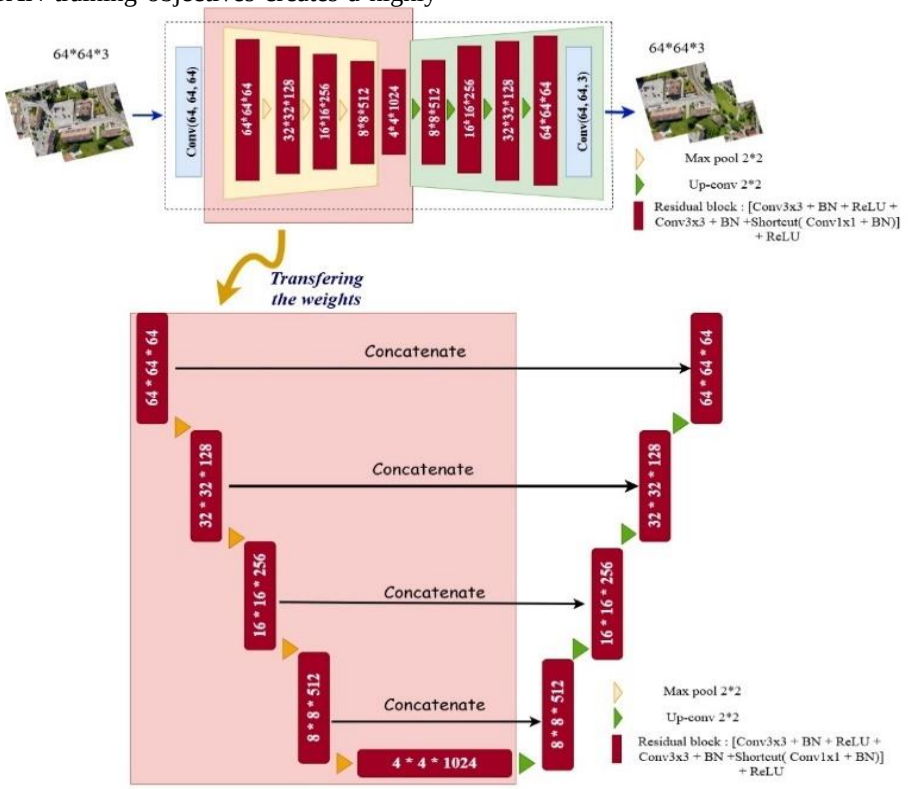

Fig. 7. Once the autoencoder is trained individually, the encoder block weights will be transferred to the corresponding blocks in the RU-net integrated in the SRGAN generator.

autoencoder is that the encoder block maintains identical structure in both the autoencoder and RU-Net models as illustrated in Fig. 6. Even though skip connections are not included in the decoder of the autoencoder, the weights obtained from training the autoencoder on our prepared dataset were utilized as initial weights for the RU-Net encoder block in the model, as shown in Fig. 7. This approach of training the U-Net with pretrained encoder weights can significantly accelerate the training.

Every encoder block in our autoencoder aims to increase the quantity of feature maps while reducing the spatial size of the input data. Each encoder block utilizes a residual block, as shown in Fig. 4, followed by a max pooling layer to decrease the dimensionality of the input data while preserving key features. On the contrary, every decoder block carries out the opposite process of its encoder block counterpart. It converts the reduced representation from the encoder block to the original spatial dimensions, with a reduction in the number of feature maps. This is done by adding upsampling layers to each decoder block in order to increase the spatial dimensions of the feature maps. Initially, we inputted the LR images into an autoencoder and then proceeded to train it to reconstruct the original images. We utilized an Adam optimizer and mean square error loss function for training our autoencoder model for 3000 epochs until achieving over 90% accuracy. Afterwards, we transferred the encoder block weights to the RU-Net incorporated in the SRGAN generator.

B. Loss Function

In SRGAN, the total loss function is not just one function, as it is made up of two loss components with varying weights.

$$L_{SRGAN} = \alpha\ L_{content} + \beta\ L_{adversarial} \tag{1}$$

In Equation (1), α and β represent coefficients that balance various loss components. Content Loss assesses how similar the generated high-resolution image is to the original high-resolution image, helping the model accurately depict the core components and layout of the original image. Two ways to measure content loss in SRGAN are Pixel-wise Mean Squared Error (MSE), a standard method that can lead to images appearing too smooth without fine details, and Perceptual Loss (VGG Loss), which uses features from a pre-trained VGG network to evaluate similarity. This approach better captures important sensory details. The MSE loss does not effectively represent the variations in image texture at the pixel level. A high-resolution image pixel consists of various combinations, and the MSE loss typically averages these combinations, resulting in an unrealistic comparison to the ground truth. In order to address this problem, the initial SRGAN model used a VGG loss which was calculated by comparing the feature maps generated by the VGG-19 model using Euclidean distance. Equation (2) represents the perceptual loss function.

$$L_{per} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\Phi_{i,j}(I^{HR})_{x,y} - \Phi_{i,j}(G(I^{LR}))_{x,y}\right)^2 \tag{2}$$

$W_{i,j}$ represents the width dimension of feature maps in the VGG network. $H_{i,j}$ represents the height dimensions of the feature maps in the VGG network. $\Phi_{i,j}$ denotes the feature map produced by the j-th convolutional layer preceding the i-th convolutional layer in the network. I^{HR} represents a high-resolution picture. I^{LR} stands for a low-quality picture. Therefore, $Lcontent$ can be replaced by $Lper$ in (1).

An often-used choice for adversarial loss in SRGAN is the Binary Cross Entropy (BCE) loss, employed to train the discriminator and distinguish between real high-resolution images and generated images.

$$L_{adv} = -E_{real}[\log(D(\mathrm{HR}))] - E_{fake}\left[\log\left(l - D(\mathrm{G(LR)})\right)\right] \tag{3}$$

Equation (3) defines E_{real} and E_{fake} as expectations for real and fake image distributions, respectively, with D as the Discriminator network, HR as the Real high-resolution image, and G(LR) as the high-resolution image generated by the Generator network from a low-resolution input.

C. Evaluation Metrics

1) PSNR

The numerical measurement commonly used for evaluating image quality is known as PSNR, which stands for Peak Signal-to-Noise Ratio. It evaluates the level of distortion by comparing the initial signal with the compressed or reconstructed one. The PSNR quantifies the difference between the maximum possible signal power and the power of the signal that has been altered. Greater PSNR values, measured in decibels (dB), suggest improved image quality with reduced distortion, whereas lower values indicate inferior quality and increased distortion. This measurement utilizes Mean Squared Error (MSE) to compute the average squared discrepancy between matching pixels in the initial and altered signals. The PSNR scale is calculated using a logarithmic function applied to the MSE, as in (4).

$$PSNR = 20\log_{10} \frac{MaxI}{\sqrt{MSE}} \tag{4}$$

MaxI is the highest attainable pixel value found in an image. MSE is the average of the squared discrepancies between matching pixels in two images.

$$MSE = \frac{1}{M*N} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(f_{ij} - f'_{ij}\right)^2 \tag{5}$$

Equation (6) illustrates how f_{ij} represents the pixel values of the original high-resolution image and f'_{ij} represents the pixel values of the image after reconstruction. M represents the width of the image while N represents the height.

2) SSIM

SSIM, also known as Structural Similarity Index, is a commonly used metric that evaluates the similarity of images based on their structural intricacies and pixel values to gauge their perceived quality. It assesses luminance, contrast, and structure by comparing nearby pixel groups in both original and modified images. Brightness similarity is evaluated by luminance, contrast levels are measured by contrast, and pattern and texture similarity are determined by structure. Combining the calculated component scores creates an index ranging from 0 to 1, with 1 indicating full similarity and 0 indicating complete dissimilarity. SSIM provides a more meaningful evaluation by taking into account human visual perception and image structure, unlike conventional metrics like PSNR.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (6)$$

In Equation (6), x represents the original high-resolution image. y is the reconstructed image. The mean values of the images are denoted by μ_x and μ_y, while the variances are σ_y^2 and σ_x^2. The covariance between the two images is σ_{xy}.

IV. EXPERIMENTAL RESULTS

A. Dataset Collection and Preprocessing

In this study, we utilized the Pix4Dmatic dataset, which comprises high-resolution images captured by unmanned aerial vehicles (UAVs). These images are in .JPG format with resolutions ranging from 12 to 20 megapixels per image, depending on the camera specifications. The dataset encompasses diverse scenarios, including industrial, agricultural, and urban settings, with some images captured using multi-spectral sensors while others use RGB sensors. These high-quality UAV-captured images are particularly well-suited for super-resolution tasks due to their rich details and variety of textures and features. For our research, we selected a subset of 250 images from the dataset. These images were specifically chosen from urban environments to ensure consistent content and representation. The original resolution of the images was 6000 × 4000 pixels, reflecting the high fidelity of UAV imagery.

However, due to computational constraints, the images were resized to a standardized resolution of 256 × 256 pixels. This resizing step was necessary to ensure feasible memory usage and processing speed while maintaining sufficient details for super-resolution training. The dataset was divided into training and testing subsets to facilitate the training and evaluation of our model. Specifically, 80% of the images were allocated for training, while the remaining 20% were set aside for testing. This division ensures a robust framework for model evaluation and generalization to unseen data. To prepare the dataset for training the SRGAN model, a systematic preprocessing pipeline was implemented to create paired datasets of low-resolution (LR) and high-resolution (HR) images. The availability of such paired data is essential for supervised learning tasks in super-resolution. The first step involved resizing the original images to a standardized high-resolution size of 256 × 256 pixels. This resizing step balanced computational efficiency and detail retention, ensuring the HR images maintained sufficient quality for effective model training.

Following the resizing, the HR images were downsampled by a factor of 4 using bicubic interpolation to generate corresponding LR images. The resulting LR images, with dimensions of 64 × 64 pixels, simulate real-world conditions of low-quality image acquisition. This step ensures that the SRGAN learns to effectively map degraded LR inputs to their corresponding HR counterparts during training. To ensure compatibility with the generator's output layer, which utilizes a tanh activation function, both the LR and HR images were normalized to the range [-1, 1]. This normalization step is critical as it improves the numerical stability of the model and facilitates faster convergence during training. By standardizing the pixel values, the model can focus on learning meaningful features rather than being affected by scale differences in the data.

The Python programming language, along with the TensorFlow framework, was used to implement SRGAN for this experiment. The model was trained on a machine operating Windows 10, equipped with an Nvidia GeForce RTX 2060 GPU with 6 GB of memory. The Adam optimizer was employed with β1=0.9 and a learning rate of 1×10^{-4}, and finally the number of epochs was set to 5000.

The standard SRGAN, which relies heavily on deep residual blocks, exhibited a peak GPU memory usage of approximately 4.8 GB during training. The proposed U-Net-based SRGAN (U-SRGAN) design, which replaces residual blocks with a U-Net encoder-decoder structure, resulted in a reduction in memory usage to 4.2 GB, attributed to the more structured feature extraction process. However, the RU-Net and pre-trained RU-Net models exhibited slightly higher memory consumption at 4.6 GB and 4.7 GB, respectively, due to the additional residual connections and the incorporation of pre-trained encoder weights.

Regarding runtime efficiency, a full training session of the baseline SRGAN took approximately 48 hours for 5000 epochs, while the U-SRGAN variant reduced this to 42 hours due to improved feature reuse. The RU-Net model required 44 hours, whereas the pre-trained RU-Net variant demonstrated the best efficiency, completing training in 37 hours. This reduction is attributed to the transfer learning strategy, which provided a well-initialized feature extraction process, allowing faster convergence.

B. Background

We created several designs to conduct an ablation comparison, aiming to evaluate the effectiveness of incorporating each component into the generator of the SRGAN. Training GAN networks poses a major challenge, as it necessitates extensive

datasets containing thousands of images. Yet, the dataset we have consists of only 250 images, which is not enough for effectively training a strong GAN model. In order to overcome this drawback, we integrated designs such as U-Net, RU-Net, and pretrained RU-Net into the generator by substituting the original residual blocks. Our suggested networks were selected for their capacity to excel in image reconstruction tasks even when trained on limited datasets. U-Net is famous for its powerful feature extraction and accurate localization abilities, which make it highly effective in situations with limited data. RU-Net enhances U-Net by incorporating residual connections to improve gradient flow and address vanishing gradients in training. Ultimately, utilizing a pre-trained RU-Net takes advantage of transfer learning, paving the way for improved generalization even with restricted training data. Our goal in incorporating these structures was to enhance convergence, combat overfitting, and ultimately enhance the quality of the produced high-resolution images. This method enabled us to assess how various changes to the architecture impact the GAN's performance in data-limited scenarios.

The pre-trained RU-Net utilized transfer learning by first training an autoencoder to rebuild LR images. During the pretraining stage, low resolution images were fed into the autoencoder and trained for 3000 epochs until the accuracy was

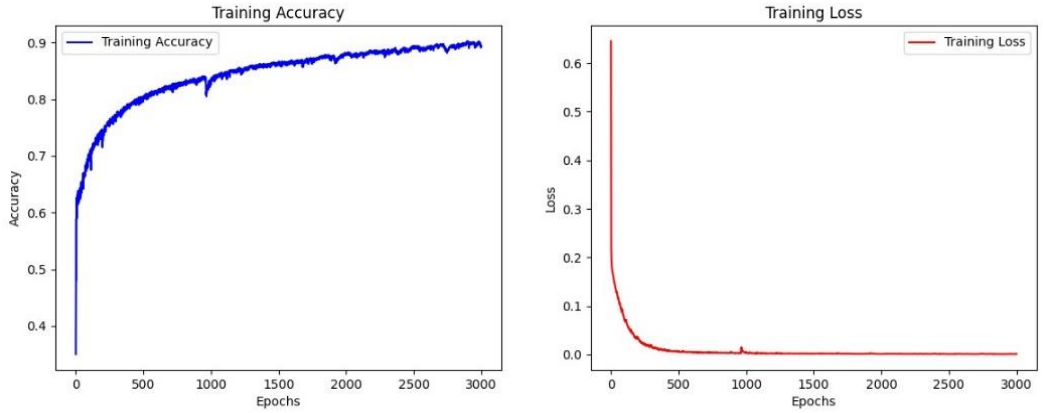

Fig. 8. The accuracy and loss figures for autoencoder. We trained the autoencoder 3000 epochs to reconstruct the LR images so that we can transfer the weights to the RU-Net

around 90%. In Fig. 8, the accuracy increased consistently as the loss decreased during training. After finishing the training of the autoencoder, the encoder's pretrained weights were moved to the encoder part of the RU-Net which is included in the generator of SRGAN. This method of transfer learning gave the generator a powerful starting point, speeding up its progress and improving its capacity to identify important characteristics in low-resolution inputs. Through utilizing the pretraining phase, RU-Net was more prepared to address the difficulties presented by small datasets and the inherent volatility of GAN training. This procedure enhanced both the generator's ability to reconstruct high-resolution images and its learning efficiency, leading to more stable results and reducing overfitting while improving overall quality.

C. Ablation Comparison

TABLE I
Final PSNR and SSIM Values for Each Model

Model	PSNR (dB) ↑	SSIM ↑
SRGAN	23.877	0.727
U-SRGAN	24.843	0.759
RU-SRGAN	**25.807**	0.795
Pretrained RU-SRGAN	25.795	**0.832**

The performance of the four models—SRGAN, U-SRGAN, RU-SRGAN, and Pretrained RU-SRGAN—was evaluated using PSNR and SSIM, two widely accepted metrics for assessing image quality and structural fidelity. The quantitative results are summarized in Table I, while Fig. 9 and Fig. 10 illustrate the progression of PSNR and SSIM across epochs. The Pretrained RU-SRGAN achieved the second highest PSNR (25.795 dB) and highest SSIM (0.832), demonstrating its superior capability in generating high-quality images with fine details and structural integrity. This model's performance highlights the advantages of integrating transfer learning with RU-Net architecture in the generator. We have assumed that using the pretrained weights only helps the faster convergence. RU-SRGAN also performed well, achieving a PSNR of 25.807 dB and SSIM of 0.795, slightly below the pretrained variant. These results emphasize the benefits of the residual U-Net structure for improving image reconstruction. The U-SRGAN achieved intermediate results, with a PSNR of 24.843 dB and SSIM of 0.759, indicating an improvement over the baseline SRGAN. The baseline SRGAN exhibited the lowest performance, achieving a PSNR of 23.877 dB and SSIM of 0.727. This result underscores the limitations of the standard SRGAN generator and the need for architectural enhancements like U-Net or RU-Net to achieve higher-quality outputs.

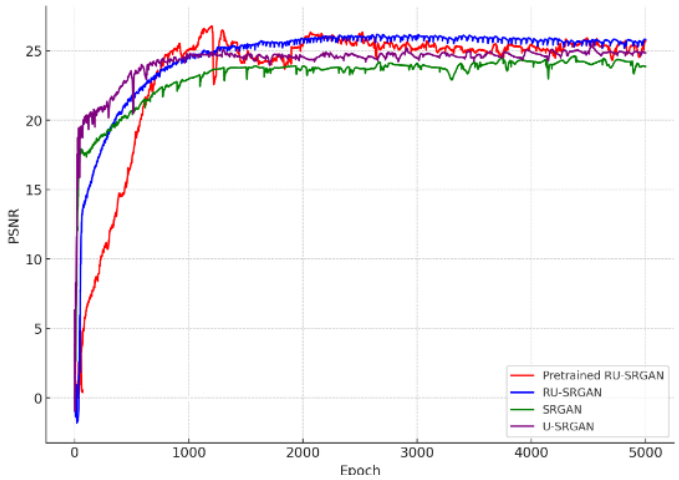

Fig. 9. PSNR progression across epochs for all models

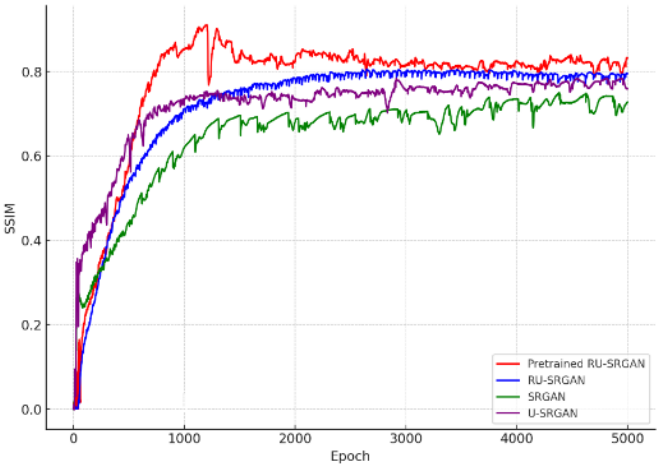

Fig. 10. SSIM progression across epochs for all models

Fig. 11. Ablation comparisons on our proposed model through created samples

In the PSNR plot, Fig. 9, all models show a consistent increase during the initial training epochs, reflecting successful learning and improved reconstruction accuracy. The Pretrained RU-SRGAN consistently outperforms other models across all epochs, demonstrating the effectiveness of pretraining in accelerating convergence and enhancing performance. Similarly, in the SSIM plot, Fig. 10, the Pretrained RU-SRGAN achieves the best structural similarity, with a noticeable performance gap compared to the other models, especially after convergence. These results highlight the effectiveness of combining advanced network architectures with transfer learning to achieve state-of-the-art performance in super-resolution tasks.

The findings showcased the effectiveness of different SRGAN models in enlarging small LR images measuring 64×64 by 4 times. The regular SRGAN and U-SRGAN models show they can increase the size of low-resolution images and also bring back some details. Nonetheless, these models produce visible defects that impact the visual appearance of the generated images. The origins of the artifacts could be due to restrictions in the structure or loss functions utilized, since these models do not have particular improvements to adequately retain spatial information or address inconsistencies in reconstruction. However, RU-SRGAN and Pretrained RU-SRGAN show superior results compared to SRGAN and U-SRGAN. These models rely on the utilization of residual blocks, which are essential for their exceptional performance. Residual blocks enhance the model's capacity to capture and transmit advanced features, making the training process more stable and improving the reconstruction of intricate details. Therefore, RU-SRGAN produces more defined results with less imperfections, closely mirroring the high-quality HR images. The pre-trained RU-SRGAN also gains advantages from its initial training on a vast dataset, which helps it perform well on new data and generate visually appealing outcomes with enhanced texture and edge sharpness. Overall, the use of residual blocks in RU-SRGAN and Pretrained RU-SRGAN models demonstrates their effectiveness in super-resolution tasks. While SRGAN and U-SRGAN struggle with artifacts and limited detail reconstruction, the advanced architecture of RU-SRGAN models achieves a notable improvement in both fidelity and perceptual quality, setting a strong benchmark for super-resolution performance. Fig. 11. shows some examples created by our models studied.

D. Benchmark Comparison

In this section, we focused on evaluating the performance of the pretrained RU-SRGAN model, which incorporates U-Net architecture, residual learning, and transfer learning through pretraining on autoencoders to improve the reconstruction of high-resolution images from low-resolution inputs. This model was compared against several benchmark models, including Bicubic, SRCNN, ESRGAN, EDSR, and EnhanceNet (see Table II). The pretrained RU-SRGAN outperformed these models in both PSNR and SSIM, demonstrating superior image quality and structural similarity.

While EDSR, enhanced deep super-resolution, performed well, particularly in PSNR, the pretrained RU-SRGAN excelled in SSIM, a metric more closely related to perceptual quality. This is largely due to the use of residual blocks, a feature shared with EDSR, which enhances the model's ability to recover high-frequency details. The incorporation of U-Net and autoencoder pretraining in our model provided a significant advantage in perceptual quality, highlighting the effectiveness of combining advanced architectures and transfer learning to achieve state-of-the-art results in super-resolution tasks. Fig. 12 illustrates qualitative comparisons between our models and other well-known SR models.

TABLE II
Comparison with State-of-the-art Model

Models	Metrics	Bicubic	SRCNN	ESRGAN	EDSR	EnhanceNet	Our model
Pix4Dmatic	PSNR (dB) ↑	21.0	23.5	24.5	**26.3**	25.5	25.7
	SSIM ↑	0.45	0.55	0.63	0.74	0.67	**0.83**
BraTS	PSNR (dB) ↑	28.42	30.48	31.35	32.15	32.12	**32.26**
	SSIM ↑	0.81	0.86	0.88	**0.89**	**0.89**	**0.89**

Fig. 12. Qualitative comparisons between SR models

E. Validation of Generalization Across Diverse Datasets

To substantiate the generalizability and robustness of the pretrained RU-SRGAN model, we conducted experiments on a brain imaging dataset, specifically using the BraTS (Brain Tumor Segmentation) dataset. The BraTS dataset is widely recognized for its high-quality multimodal MRI scans, providing a valuable benchmark for assessing the performance of super-resolution techniques in medical imaging. By demonstrating the effectiveness of our models on this dataset, we aim to establish that their performance is not restricted to UAV imagery and that their results are not coincidental.

The experimental results reveal that our model significantly outperforms traditional methods such as bicubic interpolation, SRCNN, and ESRGAN in reconstructing high-resolution brain images. The comparisons highlight that our model is capable of preserving fine anatomical details, enhancing structural clarity, and reducing artifacts, which are critical in applications like medical diagnostics. These findings, as shown in Table IIII, underscore the adaptability of our models to diverse datasets, reinforcing their potential for broader applications.

The inclusion of results on the BraTS dataset supports our claim that the proposed models are not dataset-specific and are generalizable to other domains. This experiment demonstrates the ability of our model to handle distinct image characteristics, further validating their robustness and versatility. Consequently, this enhances the applicability of pretrained RU-SRGAN across fields where high-resolution imagery is essential, such as in both aerial and medical imaging.

V. DISCUSSION

The integration of autoencoder pretraining into the Pretrained RU-SRGAN model significantly enhances both the efficiency and the practicality of the system for real-time applications. By transferring the fine details learned from low-resolution images directly into the GAN architecture, our model not only accelerates convergence but also improves the quality of high-resolution reconstructions. This pretraining strategy allows the generator to effectively use prior knowledge from the autoencoder, reducing the reliance on extensive training datasets and minimizing the time required to adapt to new or limited data.

This approach is particularly advantageous in UAV-based applications, where rapid image processing and computational efficiency are essential. UAVs often operate in environments with limited computing resources and variable data quality, such as low-resolution images captured under different environmental conditions. By leveraging the pretrained weights from the autoencoder, our model ensures that fine-grained image details are preserved and efficiently reconstructed, without incurring high computational costs. This makes the model well-suited for on-the-fly image enhancement and real-

time decision-making, as required in fields like surveillance, remote sensing, and autonomous navigation, where timely and accurate image processing is critical.

Thus, the pretrained RU-SRGAN model offers a robust solution for super-resolution tasks in UAV systems, balancing the trade-off between image quality and computational efficiency. This makes it an ideal choice for scenarios where high-quality image enhancement is needed within a limited time-frame, providing a scalable and resource-efficient approach to real-world image reconstruction challenges.

It is crucial to highlight that these results were not obtained by chance but through a thorough investigation of various architectures. In our initial tests, we evaluated the R2-U-Net

(Residual Recurrent U-Net) [42] as it has the capacity to include residual and recurrent connections, which we believed would improve feature extraction and reconstruction. Although it possesses strong theoretical foundations, R2-U-Net did not meet expectations due to its inability to provide the necessary level of detail and accuracy for superior resolution. This led us to investigate additional architectural improvements to enhance the results.

We also investigated incorporating attention gates [43] into both RU-Net and U-Net structures, aiming to direct the network's focus towards the most important parts of the images and enhance overall performance. Nevertheless, this method resulted in visible flaws in the final images, detracting from the overall visual appeal of the results. These artifacts could result from too much focus on specific attributes or instability during training caused by the attention mechanism. These experiments provided us with important knowledge that led us to incorporate residual blocks and pretraining strategies in our final model. The repeated cycle of testing, evaluating, and improving designs not only allowed us to find the best solution but also enhanced our understanding of the difficulties and trade-offs in super-resolution projects. This systematic approach emphasizes the thoroughness of our investigation and highlights the significance of considering a diverse array of options before settling on the best solution.

VI. CONCLUSION

This study advances the image super-resolution field by proposing innovative enhancements to the SRGAN design. By incorporating U-Net-inspired architectures, residual connections, and pretraining strategies, we effectively addressed major challenges in super-resolution, such as artifacts and loss of fine details, all while improving training stability and convergence speed. Integrating residual blocks into the U-Net structure facilitated the extraction of advanced features and promoted smooth gradient flow, ultimately resulting in improved visual quality and structural coherence in the generated images. Additionally, initializing the RU-Net with an autoencoder led to faster convergence and reduced overfitting, especially when data is scarce.

This work is significant because it helps fill gaps in current super-resolution techniques. While traditional SRGAN designs have been successful, the proposed model, pre-trained RU-SRGAN, demonstrated superior results in both quantitative metrics (PSNR and SSIM) and visual quality. These advancements are particularly crucial for tasks involving low-resolution images, such as data captured by UAVs, where accurate reconstructions are vital for activities like object detection, scene categorization, and remote sensing. Future research could focus on expanding these models to handle real-time super-resolution tasks and investigating their applicability in various fields, including medical imaging and autonomous systems.

While the pretrained RU-SRGAN model has demonstrated strong generalizability across UAV imagery and medical datasets, further validation on a wider range of datasets-such as low-light images, satellite imagery, and thermal imaging-could enhance its robustness for diverse real-world applications. Additionally, the model currently operates with a fixed upscaling factor (x4), which may not be optimal for all scenarios. Future research could explore adaptive super-resolution mechanisms, where the model dynamically adjusts the upscaling factor based on input quality and application requirements, improving its flexibility for various practical deployments.

REFERENCES

[1] Y. Wei, Y. Li, Z. Ding, Y. Wang, T. Zeng, and T. Long, "SAR Parametric Super-Resolution Image Reconstruction Methods Based on ADMM and Deep Neural Network," IEEE Trans. Geosci. Remote Sensing, vol. 59, no. 12, pp. 10197–10212, Dec. 2021, doi: 10.1109/TGRS.2021.3052793.

[2] D. Khaledyan, A. Amirany, K. Jafari, M. H. Moaiyeri, A. Z. Khuzani, and N. Mashhadi, "Low-Cost Implementation of Bilinear and Bicubic Image Interpolation for Real-Time Image Super-Resolution," in 2020 IEEE Global Humanitarian Technology Conference (GHTC), Oct. 2020, pp. 1–5. doi: 10.1109/GHTC46280.2020.9342625.

[3] A. M. John, K. Khanna, R. R. Prasad, and L. G. Pillai, "A Review on Application of Fourier Transform in Image Restoration," in 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India: IEEE, Oct. 2020, pp. 389–397. doi: 10.1109/I-SMAC49090.2020.9243510.

[4] Y. Yang, Z. Su, and L. Sun, "Medical image enhancement algorithm based on wavelet transform," Electronics Letters, vol. 46, no. 2, pp. 120–121, Jan. 2010, doi: 10.1049/el.2010.2063.

[5] W. Ren et al., "Wavelet Transform Based Network for Spectral Super-Resolution," in IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA: IEEE, Jul. 2023, pp. 7551–7554. doi: 10.1109/IGARSS52108.2023.10281411.

[6] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., Jun. 2004, p. I–I. doi: 10.1109/CVPR.2004.1315043.

[7] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image Super-Resolution Via Sparse Representation," IEEE Transactions on Image Processing, vol. 19, no. 11, pp. 2861–2873, Nov. 2010, doi: 10.1109/TIP.2010.2050625.

[8] S. Tang and N. Zhou, "Local Similarity Regularized Sparse Representation for Hyperspectral Image Super-Resolution," in IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia: IEEE, Jul. 2018, pp. 5120–5123. doi: 10.1109/IGARSS.2018.8518168.

[9] H. Li, K.-M. Lam, and M. Wang, "Image super-resolution via feature-augmented random forest," Signal Processing: Image Communication, vol. 72, pp. 25–34, Mar. 2019, doi: 10.1016/j.image.2018.12.001.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[11] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.

[12] A. Rouhbakhshmeghrazi, B. Li, W. Iqbal, and G. Alizadeh, "Super-Resolution Reconstruction of UAV Images with GANs: Achievements and Challenges," in 2024 International Conference on Cyber-Physical Social Intelligence (ICCSI), Doha, Qatar: IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/ICCSI62669.2024.10799467.

[13] C. Mollière, J. Gottfriedsen, M. Langer, P. Massaro, C. Soraruf, and M. Schubert, "Multi-Spectral Super-Resolution of Thermal Infrared Data Products for Urban Heat Applications," in IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA: IEEE, Jul. 2023, pp. 4919–4922. doi: 10.1109/IGARSS52108.2023.10283339.
[14] W. Yang, Z. Ma, and Y. Shi, "SAR Image Super-Resolution based on Artificial Intelligence," in IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia: IEEE, Jul. 2022, pp. 4643–4646. doi: 10.1109/IGARSS46834.2022.9884456.
[15] J. Ferdousi, S. I. Lincoln, Md. K. Alom, and Md. Foysal, "A deep learning approach for white blood cells image generation and classification using SRGAN and VGG19," Telematics and Informatics Reports, vol. 16, p. 100163, Dec. 2024, doi: 10.1016/j.teler.2024.100163.
[16] M. Ullah, A. Hamza, I. Ahmad Taj, and M. Tahir, "Low Resolution Face Recognition using Enhanced SRGAN Generated Images," in 2021 16th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan: IEEE, Dec. 2021, pp. 1–6. doi: 10.1109/ICET54505.2021.9689885.
[17] Q. Zhu, X. Fan, Y. Zhong, Q. Guan, L. Zhang, and D. Li, "Super Resolution Generative Adversarial Network Based Image Augmentation for Scene Classification of Remote Sensing Images," in IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA: IEEE, Sep. 2020, pp. 573–576. doi: 10.1109/IGARSS39084.2020.9324043.
[18] Y. Wang, H. Wu, L. Shuai, C. Peng, and Z. Yang, "Detection of plane in remote sensing images using super-resolution," PLOS ONE.
[19] D. Rathgamage Don, R. Aygun, and M. Karakaya, "A Multistage Framework for Detection of Very Small Objects," in Proceedings of the 2023 6th International Conference on Machine Vision and Applications, Singapore Singapore: ACM, Mar. 2023, pp. 9–14. doi: 10.1145/3589572.3589574.
[20] Sub-r-paChayanon, FanMing-Zhong, and ChenRung-Ching, "Super-resolution for traffic signs: a comparative analysis of SRGAN and ESRGAN performance," IET Conference Proceedings, Dec. 2024, doi: 10.1049/icp.2024.4162.
[21] P. Nandal, S. Pahal, A. Khanna, and P. Rogério Pinheiro, "Super-Resolution of Medical Images Using Real ESRGAN," IEEE Access, vol. 12, pp. 176155–176170, 2024, doi: 10.1109/ACCESS.2024.3497002.
[22] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690.
[23] Y. Wang, Z. Xu, X. Wang, J. He, and X. Zhao, "An Improved SRGAN-Based Deblurring Model for Multiple Blurriness in Microscopy," IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1–13, 2024, doi: 10.1109/TIM.2024.3470059.
[24] "SRGAN-LSTM-Based Celestial Spectral Velocimetry Compensation Method With Solar Activity Images | IEEE Journals & Magazine | IEEE Xplore." Accessed: Nov. 15, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10526305
[25] S. Madhav, T. M. Nandhika, and M. K. Kavitha Devi, "Super Resolution of Medical Images Using SRGAN," in 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Feb. 2024, pp. 1–6. doi: 10.1109/ic-ETITE58242.2024.10493588.
[26] R. Duggal and A. Gupta, "P-TELU: Parametric Tan Hyperbolic Linear Unit Activation for Deep Neural Networks," presented at the Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 974–978. Accessed: Nov. 15, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2017_workshops/w18/html/Duggal_P-TELU_Parametric_Tan_ICCV_2017_paper.html
[27] S. Ji et al., "Super-resolution reconstruction of variable length infrared image sequences based on convolutional neural networks and pixel shuffling," in 2024 International Conference on Optoelectronic Information and Optical Engineering (OIOE 2024), H. Bin Ahmad and M. Jiang, Eds., Kunming, China: SPIE, Jun. 2024, p. 10. doi: 10.1117/12.3030372.
[28] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Apr. 10, 2015, arXiv: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.
[29] X. Ren and X. Li, "Research on image super-resolution based on attention mechanism and multi-scale," J. Phys.: Conf. Ser., vol. 1792, no. 1, p. 012025, Feb. 2021, doi: 10.1088/1742-6596/1792/1/012025.
[30] Y. Wang, "Single Image Super-Resolution with U-Net Generative Adversarial Networks," in 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China: IEEE, Jun. 2021, pp. 1835–1840. doi: 10.1109/IMCEC51613.2021.9482317.
[31] P. S. Hrishikesh, D. Puthussery, K. A. Akhil, and C. V. Jiji, "Relativistic GAN using Receptive Field Block for Single Image Super-Resolution with improved Perceptual Quality," in 2023 11th International Symposium on Electronic Systems Devices and Computing (ESDC), May 2023, pp. 1–6. doi: 10.1109/ESDC56251.2023.10149876.
[32] Z. Wu, "Research on Image Super-Resolution Using Attention Mechanisms based on Super-Resolution Generative Adversarial Network," FCIS, vol. 5, no. 2, pp. 131–134, Sep. 2023, doi: 10.54097/fcis.v5i2.13142.
[33] B. Sun, B. Chen, Y. Tian, and W. Chen, "TESRGAN: Transformer Enhanced Super-Resolution Generative Adversarial Networks," in 2024 4th International Conference on Neural Networks, Information and Communication (NNICE), Guangzhou, China: IEEE, Jan. 2024, pp. 137–141. doi: 10.1109/NNICE61279.2024.10498773.
[34] M. Hasan, M. Vijay, S. Sharanyaa, and V. S. D. Tejaswi, "ENSEMBLE MODEL WITH IMPROVED U-NET-BASED SEGMENTATION FOR LEUKEMIA DETECTION," Biomed. Eng. Appl. Basis Commun., vol. 36, no. 03, p. 2450011, Jun. 2024, doi: 10.4015/S101623722450011X.
[35] A. Rouhbakhshmeghrazi, B. Li, and W. Iqbal, "Color Image Segmentation of Dental Caries Using U-Net Enhanced with Residual Blocks and Attention Mechanisms," in 2024 International Conference on Cyber-Physical Social Intelligence (ICCSI), Doha, Qatar: IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/ICCSI62669.2024.10799395.
[36] H.-H. Chang, S.-J. Yeh, M.-C. Chiang, and S.-T. Hsieh, "RU-Net: skull stripping in rat brain MR images after ischemic stroke with rat U-Net," BMC Med Imaging, vol. 23, no. 1, p. 44, Mar. 2023, doi: 10.1186/s12880-023-00994-8.
[37] S. Leclerc et al., "RU-Net: A refining segmentation network for 2D echocardiography," in 2019 IEEE International Ultrasonics Symposium (IUS), Glasgow, United Kingdom: IEEE, Oct. 2019, pp. 1160–1163. doi: 10.1109/ULTSYM.2019.8926158.
[38] J. Zhang et al., "Multispectral Drone Imagery and SRGAN for Rapid Phenotypic Mapping of Individual Chinese Cabbage Plants," Plant Phenomics, vol. 2022, p. 0007, 2022, doi: 10.34133/plantphenomics.0007.
[39] S. Nasrin, M. Z. Alom, R. Burada, T. M. Taha, and V. K. Asari, "Medical Image Denoising with Recurrent Residual U-Net (R2U-Net) base Auto-Encoder," in 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA: IEEE, Jul. 2019, pp. 345–350. doi: 10.1109/NAECON46414.2019.9057834.
[40] A. K. Kakumani, L. P. Sree, C. S. Krishna, G. Uppalapati, G. S. S. Pavithra, and S. Harshini, "Semantic Segmentation of Cells in Microscopy Images via Pretrained Autoencoder and Attention U-Net," in 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS), Bhubaneswar, India: IEEE, Aug. 2022, pp. 94–99. doi: 10.1109/MLCSS57186.2022.00025.
[41] K. N. Alkhamaiseh, J. L. Grantner, S. Shebrain, and I. Abdel-Qader, "Towards reliable hepatocytic anatomy segmentation in laparoscopic cholecystectomy using U-Net with Auto-Encoder," Surg Endosc, vol. 37, no. 9, pp. 7358–7369, Sep. 2023, doi: 10.1007/s00464-023-10306-4.
[42] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," May 29, 2018, arXiv: arXiv:1802.06955. doi: 10.48550/arXiv.1802.06955.
[43] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention Gate ResU-Net for Automatic MRI Brain Tumor Segmentation," IEEE Access, vol. 8, pp. 58533–58545, 2020, doi: 10.1109/ACCESS.2020.2983075.