

CTsGAN - Image Enhancement of Portable CT Scanner using ESRGAN

Amit Thakur^{1*}, Nilamadhab Mishra^{2*}

^{1,2} School of Computing Science & Engineering, VIT Bhopal University, Madhya Pradesh, India

*Corresponding Author: Amit Thakur

Abstract

This research introduces CTsGAN, a novel Generative Adversarial Network tailored for enhancing the perceptual quality of CT scan images obtained from portable scanners. Building upon the foundations of super-resolution GANs and inspired by the Enhanced SRGAN (ESRGAN), CTsGAN addresses the increasing demand for high-quality medical imaging in the context of portable CT scanners. The model employs a Kth Order Degradation Module to simulate real-world multi-level degradation, enhancing the robustness of the training process. The network architecture incorporates a modified Residual-in-Residual Dense Block with Batch Normalization for efficiency, a U-Net Discriminator with Spectral Normalization for discriminative accuracy, and a Perceptual Loss function for improved visual fidelity. The research also delves into the challenges posed by the mobile CT scanning environment, exploring a second-order degradation model to bridge the gap between synthetic and realistic degraded images. Additionally, the paper details the training process, web application deployment, and comparative evaluations against state-of-the-art methods, showcasing CTsGAN's superior performance in terms of sharpness, brightness, and detail preservation. The user-friendly web application, implemented using the Streamlit library, provides a seamless experience for medical professionals to enhance CT scan images and make informed decisions. As a contribution to the field of medical image enhancement, CTsGAN demonstrates potential applications in real-time scenarios and remains open for further development and contributions.

Keywords: Super-Resolution Generative Adversarial Network (SRGAN), CT-Scan images, COVID-19 CT Scan Images, Image Enhancement methods, COVID Pandemic Era

1. Introduction

In this work, the researchers build CTsGAN by working on SRGAN and modifying the architecture inspired by the techniques followed in designing Enhanced SRGAN (ESRGAN). The researchers use the same Residual-in-Residual Dense Block as a building unit of the network. However, they reintroduce the Batch Normalization (BN) Layer as our resulting network architecture is smaller and has a lower computational cost (since we only deal with grayscale images). They use the ESRGAN's Relativistic average GAN (RaGAN) for the discriminator to obtain relative realness value instead of absolute value. They also employ Spectral Normalization to stabilize training and suppress overshooting artifacts. This helps the generator in recovering more realistic details. For sharper edges and more visually pleasing results, they employ a modified version of

perceptual loss. Here, they use the VGG features before activation instead of after activation (in SRGAN).

The pandemic era has resulted in an increased demand for portable CT Scanners. Manufacturing companies generally achieve portability with a small statistically significant compromise in subjective quality compared to stationary CT Scanners [1]. With the increase in remote appointments (online) and second opinions, the images undergo complex natural degradation and compression while shared via various social media platforms. Sharing the images multiple times significantly diminishes the quality on the receiver end, making it all the more difficult for doctors and other medical professionals to find vital information and analyze the image effectively. This problem lies under the broad spectrum of Single Image super-resolution (SISR), a fundamental low-level vision problem. SISR aims to recover a high-resolution (HR) image from a single low-resolution (LR) counterpart. Many AI companies have achieved prosperous development. Various network architecture designs and training strategies have emerged, continuously improving SR performance, especially the Peak signal-to-noise ratio (PSNR) value [2] [3]. The major drawback is that the PSNR metric fundamentally disagrees with the subjective evaluation of human observers. It tends to over-smoothed results without sufficient high-frequency details.

Generative Adversarial Networks were introduced to Super Resolution to favor more natural-looking images [4]. SRGAN is a milestone in obtaining visually pleasing images. The journey has seen the introduction to perceptual loss, proposed to optimize super-resolution models in a feature space instead of pixel space. The Basic SRGAN model is built with residual blocks and optimized using perceptual loss in a GAN framework [5]. Thus, SRGAN is a significant improvement over PSNR models in terms of the overall visual quality of reconstruction. However, there still exists a clear gap between SRGAN results and the ground-truth (GT) images.

In this research, the researchers visit the commonly used Image Enhancement Techniques in Medical Image Processing, explore the critical components of SRGAN, and improve the model. The common approach to image reconstruction follows a classical pattern: reducing and removing unwanted noise with filters, cropping and resampling data for faster processing, using specified anatomical segmentation tools, and applying statistical tools to quantify the parts of the image. In recent years, some situation-specific enhanced processing methods such as Sparse Reconstruction of CT Images with low-dose projection data by the authors of [6] proposed a wavelet frame-based regularization method. For efficient enhancement of noisy optical coherence tomography (OCT) images, the authors of [7] proposed a collaborative shock filtering to enhance details and layered structures better. Single Image super-resolution (SISR) is a fundamental low-level vision problem. SISR aims to recover a high-resolution (HR) image from a single low-resolution (LR) counterpart. Here, we want to recover high-quality CT Scan Images from degraded low-quality images.

Major Highlights:

The researchers aim to build CTsGAN, a low-resolution to high-resolution image enhancer for portable CT Scanner using Super Resolution. Single Image Super Resolution aims at reconstructing a high-resolution (HR) image from its low-resolution (LR) counterpart. The researchers implement the project by developing and improving the following modules:

- Kth Order Degradation Module: To mimic real-world multi-level degradation, we employ a recursive degradation model. The low-resolution and noisy data is paired with high-resolution data to create training pairs.
- Network Modeling: Simplify the network architecture of the generator by modifying the composition of the Residual-in-Residual Dense Block (RRDB) and boost the training by reintroducing batch normalization. Introduce Relativistic average Discriminator with U-Net architecture and Spectral Normalization.
- Perceptual Loss: Improve the loss function by constraining features before activation and introduce voxel factor to adversarial loss.

2. Related Works

In recent years, the medical industry has observed a rise in demand for CT Scanners leading to the innovation of mobile/portable CT Scanners. There are many advantages of using mobile CT Scanners, provided that they do not compromise image quality. H. Andersson et al. [8] found a small statistically significant difference in subjective quality rating between portable CT Scanners and stationary CT Scanners. For overall image quality, 14% of portable CT images were rated grade 1 (poor quality) compared to 4% of high-end stationary CT Scanners. S. Hermena et al. [9] explain the software conversion of data into images. Each pixel represents a two-dimensional projection of a three-dimensional volume, termed a voxel. Each voxel and pixel represent a number reflecting the amount of photon energy absorbed and measured by the detector. The image processor can retrieve and manipulate the values of every possible voxel as defined by the imaging software. The number of voxels represents the imaged tissue at different resolutions. A row of voxels that forms a line from one side of the image space to the other is called an attenuation profile [10]. The researchers introduce the voxel factor that aids in image reconstruction.

The image super-resolution field [11-15] has witnessed a variety of development since the pioneering work on SRCNN [16-18]. To achieve visually pleasing results with accurate details, generative adversarial networks [19] are usually employed as loss supervision. The classical degradation model is widely adopted in blind SR methods. Gong et al. [20] proposed a flexible higher-order degradation model to synthesize more practical degradation. Dong et al. [21] proposed SRCNN to learn the mapping from LR to HR images. Zhang et al. [22] proposed to use effective residual dense blocks in SR, and they further explored a deeper network with channel attention, achieving state-of-the-art PSNR performance. He et al. [23] proposed a robust initialization method for the VGG-style network. The researchers in this paper develop a compact and effective RRDB with a Batch Normalization Layer to aid in training a deeper network.

A lot of work has been done in blind similarity retrieval, which is an important application for medical imaging. The researchers propose a general framework that can address blind similarity retrieval and be applied to voxel-based and volume-based images. Their methods are based on generating training pairs as close to real data as possible and then training a unified network on these pairs with cycle consistency loss. The degradations are constrained to mobile CT Scanners and cannot be well generalized with every image.

Perceptual-driven approaches have also been proposed to improve the visual quality of SR results. Based on the idea of perceptual similarity, perceptual loss is proposed to enhance the visual quality by minimizing the error in a feature space instead

of a pixel space. Contextual loss is developed to generate images with natural image statistics by using an objective that focuses on the feature distribution rather than merely comparing the appearance. Rad et al. [24] proposed the SRGAN model that used perceptual loss and adversarial loss to favor outputs residing on the manifold of natural images. Sajjadi et al. [25] developed a similar approach and further explored the local texture-matching loss. Based on these works, Wang et al. [20] proposed spatial feature transformation to effectively incorporate semantics into an image and improve the recovered image.

In the literature, photo-realism is usually associated with adversarial training, where a generator is pitted against samples from a real dataset to find where the two match. Several methods have been proposed to stabilize the training of deep models. For example, the residual path is developed to stabilize the training and improve performance. The residual scaling strategy was introduced by Szegedy et al. [26]. A relativistic discriminator can be used to increase the probability that generated data simultaneously is real and decrease the probability that real data are real by increasing the number of samples from which it learns. The researchers enhance SRGAN by employing a more effective Realistic average Discriminator with U-Net architecture and spectral normalization for stabilized training.

With the increase in degradation space, the training becomes challenging. The discriminator requires a more powerful capability to discriminate realness from complex training outputs, while the gradient feedback from the discriminator needs to be more accurate for local detail enhancement. Thus, the researchers improve the VGG-style discriminator to a U-Net design. The U-Net structure and degradations also increase the training instability. Hence, the researchers employ spectral normalization regularization to stabilize the training dynamics.

PSNR and SSIM are used to evaluate SR algorithms. However, these metrics fundamentally disagree with the subjective evaluation of human observers [27]. Ma's score [28] and NIQE [29] calculate the perceptual index. Blau et al. [30] find that distortion and perceptual quality are typical tradeoffs.

3. Materials and Methodology

The main aim of this research is to restore and enhance the overall perceptual quality of CT Scan Images. In this section, the network architecture will be discussed, as the adaptation of the RaGAN discriminator, and the training method used to incorporate modified perceptual loss. The methodology includes modeling the network architecture, training the model, and testing it on different self-validation datasets.

A. Kth Order Degradation Module

The classical degradation model is usually adopted to synthesize the low-resolution input. Generally, the ground truth image I , is first convolved with blur kernel k . Then, a down-sampling operation with scale factor r is performed. The low-resolution x is obtained by adding noise n . Finally, JPEG compression is also adopted as it is widely used in real-world images.

$$x = D(I) = [(I \otimes k) \downarrow r + n]_{JPEG}$$

eq 3.1.1

where D denotes the degradation process.

The researchers adopt the above classical degradation model to synthesize training pairs. They found the trained model could handle some real samples. However, it still cannot resolve some complicated degradations in the real world, especially unknown noise, and complex artifacts. This is because the synthetic low-resolution images still have a large gap with realistic degraded images. The researchers thus extend the classical model to a high-order degradation process to model more practical degradations.

For instance, if we want to restore a low-quality CT Scan image of a brain tumor downloaded from the Internet, its underlying degradation involves a complicated combination of the different degradation processes. Specifically, the original image of the CT Scan image was taken through a cell phone camera, which inevitably contains degradations such as camera blur, sensor noise, low resolution, and JPEG compression. The image was then edited with sharpening and resizing operations, bringing in overshoot and blurred artifacts. After that, it was uploaded to some social media applications (like WhatsApp for a second opinion), which introduced further compression and unpredictable noises. As digital transmission will also bring artifacts, this process becomes more complicated when the image spreads several times on the Internet.

A k -order model involves a repeated degradation process, where each degradation process adopts the classical degradation model (eq 3.1.1) with the same procedure but different hyperparameters. To keep the image resolution in a reasonable range, the down-sampling operation in eq 3.1.1 is replaced with a random resize operation.

Empirically, we adopt a second-order degradation process, as it could resolve most real cases while keeping simplicity.

$$x = D(I) = {}^n(Dn \circ \dots \circ D2 \circ D1)(I)$$

eq 3.1.2

It is worth noting that the improved high-order degradation process is imperfect and could not cover the whole degradation space in the real world. Instead, it merely extends the solvable degradation boundary of previous blind SR methods by modifying the data synthesis process.

Ringling artifacts often appear as spurious edges near sharp transitions in an image. They visually look like bands or ghosts near the edges. Overshoot artifacts are usually combined with ringling artifacts, which manifest themselves as an increased jump at the edge transition. The main cause of these artifacts is that the signal is bandlimited without high frequencies. These artifacts are common and usually produced by a sharpening algorithm, JPEG compression, etc.

The researchers employ the *sinc* filter, an idealized filter that cuts off high frequencies, to synthesize ringing and overshoot artifacts for training pairs. The sinc filter kernel can be expressed as

$$k(i, j) = [\omega c / (2\pi\sqrt{i^2 + j^2})] J_1(\omega c\sqrt{i^2 + j^2})$$

eq 3.1.3

Where (i, j) is the kernel coordinate; ω_c is the cutoff frequency and J_1 is the first-order Bessel function of the first kind. It can synthesize ringing and overshooting artefacts (especially introduced by over-sharp effects).

The researchers adopt SINC filters in two places: the blurring process and the last step of the synthesis. The order of the last SINC filter and JPEG compression is randomly exchanged to cover a larger degradation space, as some images may be first over-sharpened and then have JPEG compression while some images may first do JPEG compression followed by sharpening operation.

B. Network Architecture

SRGAN [2] is a significant milestone in solving the single-image SR problem. Wang et al. [31] proposed a modification to the building blocks of SRGAN architecture. They introduced Residual-in-Residual Dense Block (RRDB), which combined multi-level residual networks and dense connections to make networks more robust. We propose to reintroduce the Batch Normalization Layer in the RRDB structure to enable a higher learning rate and faster training. Due to the sheer nature of software in the CT Scanners that converts the data into images, we can reduce the number of RRDB units required in the network.

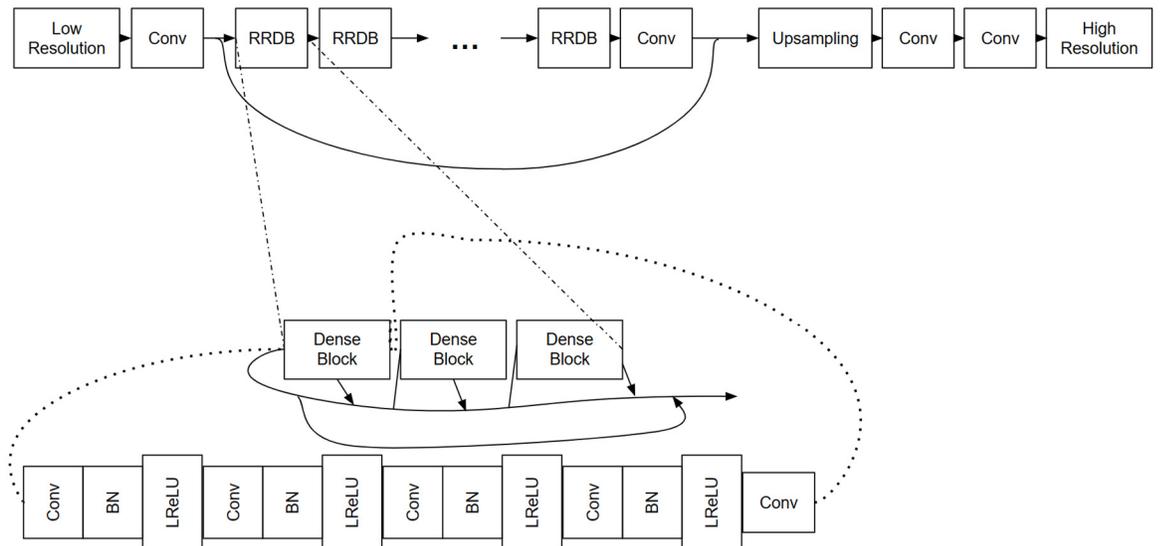


Figure 1. Generator architecture with Batch normalization layer added in RRDB structure.

In addition to the improved architecture, we also exploit several techniques to facilitate training a very deep network. First, residual scaling, i.e., scaling down the residuals by multiplying a constant between 0 and 1 before adding them to the main path to prevent instability. Second, smaller initialization, as we empirically find residual architecture is easier to train when the initial parameter variance becomes smaller.

C. U-Net Discriminator with Spectral Normalization

The standard discriminator D in SRGAN [2] estimates the probability of input image I being real or fake. This results in a vanishing gradient problem. In this situation, the discriminator is unable to provide enough feedback for the generator to learn as the error for the discriminator eventually tends to 0 (vanishing gradient). The researchers replace the standard discriminator D with the Relativistic average Discriminator, D_{Ra} [21]. The Relativistic average Discriminator tries to predict the probability that a real image I_r is relatively more realistic than a fake one I_f .

The standard discriminator can be expressed as

$$D(I) = \sigma(C(I))$$

where σ is the sigmoidal function and $C(I)$ is the non-transformed discriminator output. Thus, D_{Ra} is formulated as $D_{Ra}(I_r, I_f) = \sigma(C(I_r) - E_{I_f}[C(I_f)])$, where $E_{I_f}[\cdot]$ represents the operation of taking the average for all fake data in the mini-batch.

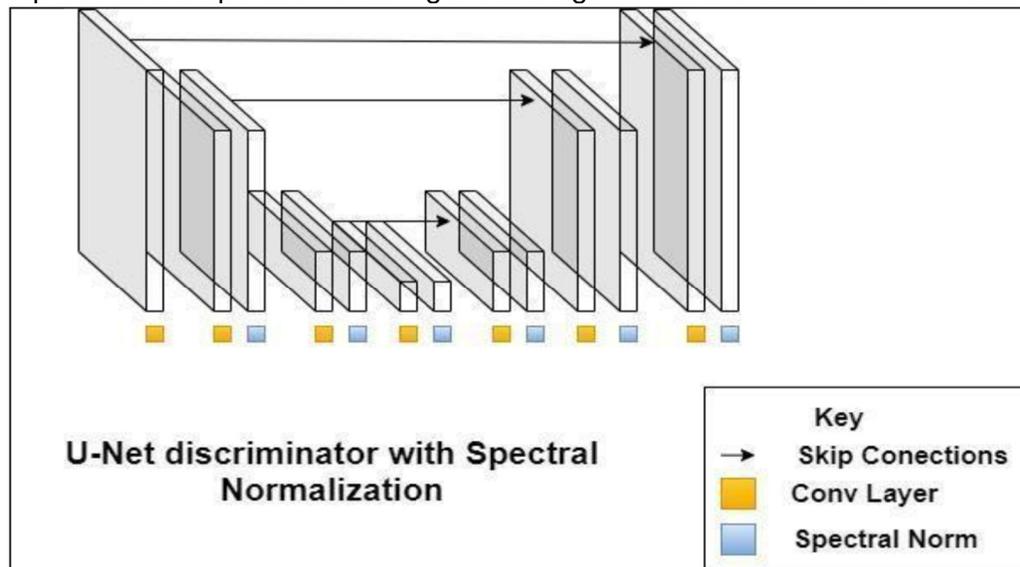


Figure 2. Network Architecture of U-Net Discriminator with Spectral Normalization

As the researchers aim to address a large degradation space, the discriminator requires a greater discriminative power for complex training outputs. Instead of discriminating global styles, it also needs to produce accurate gradient feedback for local texture. Inspired by [21], the researchers also improved the VGG-style discriminator to a U-Net design with skip connections (Fig 3.3.1). The U-Net structure outputs the realness value for each pixel and can provide detailed per-pixel feedback to the generator. The researchers employ spectral normalization

regularization to stabilize the training dynamics. It is also beneficial to alleviate the over-sharp and annoying artifacts generated during GAN training. Thus, the researchers can establish a good balance between local details enhancement and artefact suppression.

D. Perceptual Loss

The researchers improve on the perceptual loss L_{percep} by constraining on features before activation as introduced in [32]. Wand et al. [31] proposed to tackle two drawbacks of the conventional method. First, by using the features before activation, the features do not become sparse. The sparse activation provides weak supervision and leads to inferior performance. Second, using features before activation helps with tackling inconsistent reconstructed brightness.

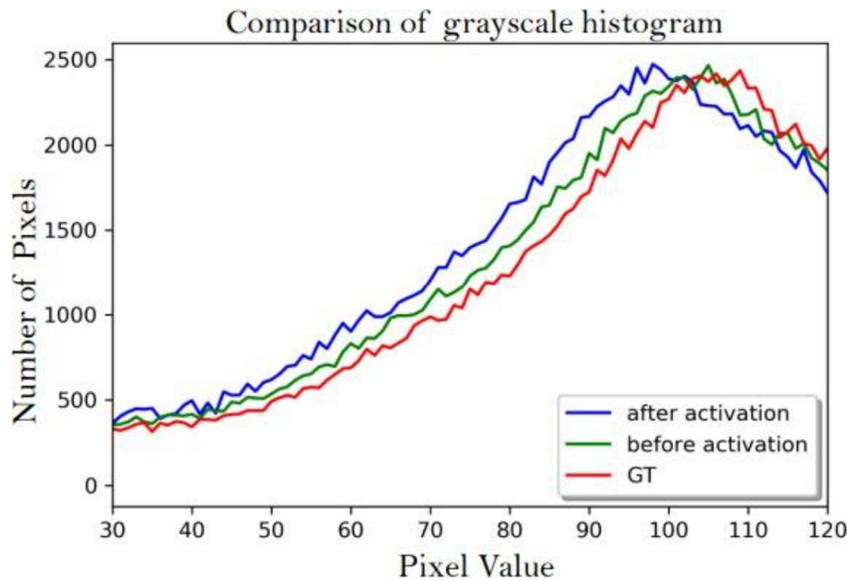


Figure 3. Grayscale Histogram Comparison between before activation and after activation in terms of brightness.

Hence, the total loss for the generator is:

$$L_G = L_1 + \lambda L_{percep} + \eta L_{v}$$

eq 3.4.1

where $L_1 = E_{x_i} ||G(x_i) - y||_1$ is the content loss that evaluates the 1-norm distance between the generated image $G(x_i)$ and the ground truth, and λ, η are the coefficients to balance the different loss terms.

The quality of the generated image is directly proportional to the number of voxels provided to the reconstruction algorithm. Thus the researchers introduce the voxel factor, v_i , in the generator loss function to get:

$$L_{Gv} = L_1 + \lambda L_{percep} + \eta L_v$$

eq 3.4.2

where $L_v = E_{x_i} ||E_{v_i}[G(x_i) - y]||$. The introduction of the voxel factor determines the order of reconstruction and reads the linear attenuation profile of the image (from

metadata). It is based on a fine-tuned VGG network for material recognition, which focuses on texture rather than the object.

E. Training Process

Before the researchers jump into the training process, the degradation system needs to be discussed. The researchers try to mimic real-world multi-level degradation by following the steps for k iterations. First, they typically apply Gaussian blur, followed by resizing the image (down sampling). Then they introduce random noise to the channel to imitate unpleasant artefacts followed by JPEG compression with a Sinc kernel as employed by many social media platforms. They repeat this process k times (where k is a different random constant for every 100 image batches). Now that they have the low-resolution dataset available, they can now start training the model.

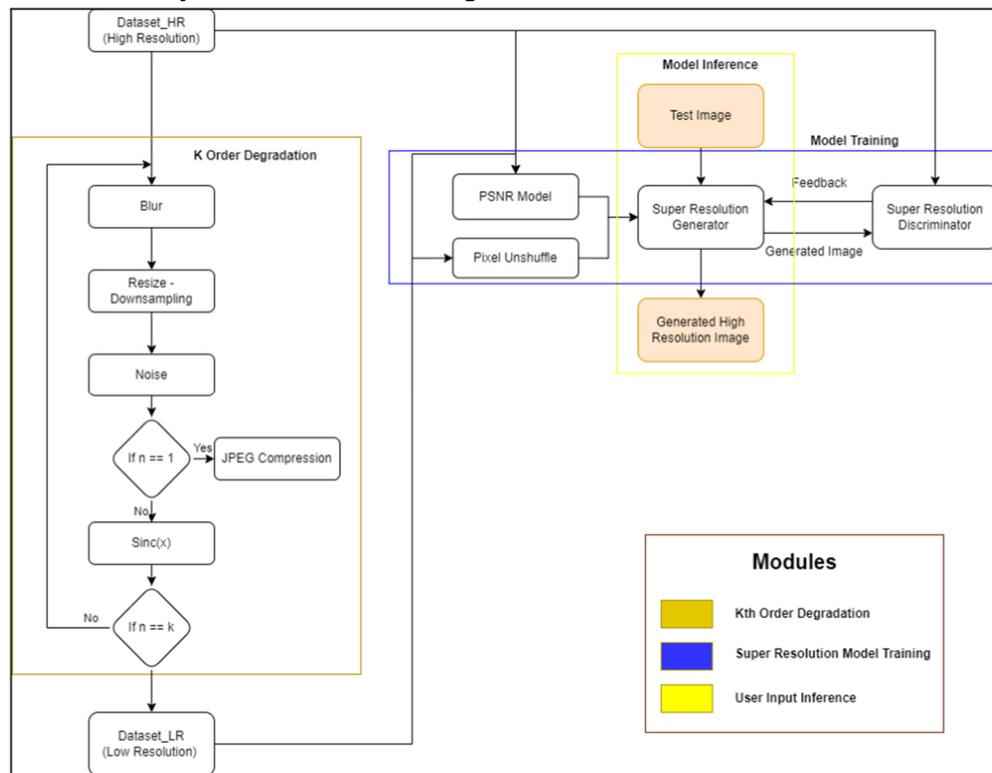


Figure 4. System flow diagram showcasing the three major modules of CTsGAN Training - Kth order degradation, Model Training, and User input interface.

The training process is divided into two stages. The researchers train a PSNR-oriented model with L_1 loss. They initialize the learning rate as 1×10^{-4} and decay by a factor of 4 every 4×10^5 of the mini-batch update. This trained PSNR model serves as a starting point for the generator. The generator is trained with a combination of perceptual and adversarial loss L_{GV} with $\alpha = 1 \times 10^{-2}$, $\beta = 2 \times 10^{-2}$, and $\gamma = 2$. Here, the learning rate is set to 2×10^{-4} and decays with a factor of 2 every 1000 iterations (for 10,000 iterations). Using a pre-trained PSNR-based model with L_1 loss helps avoid undesired local optima. The discriminator now receives relatively better-generated images instead of extremely fake ones, thus helping to focus more on texture discrimination. The researchers use the Adam

optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. As every GAN is trained, they alternate between updating the generator and discriminator until the model converges. Instead of the 16-residual and 23-RRDB blocks of ESRGAN, they employ the generator with 16 RRDB blocks only.

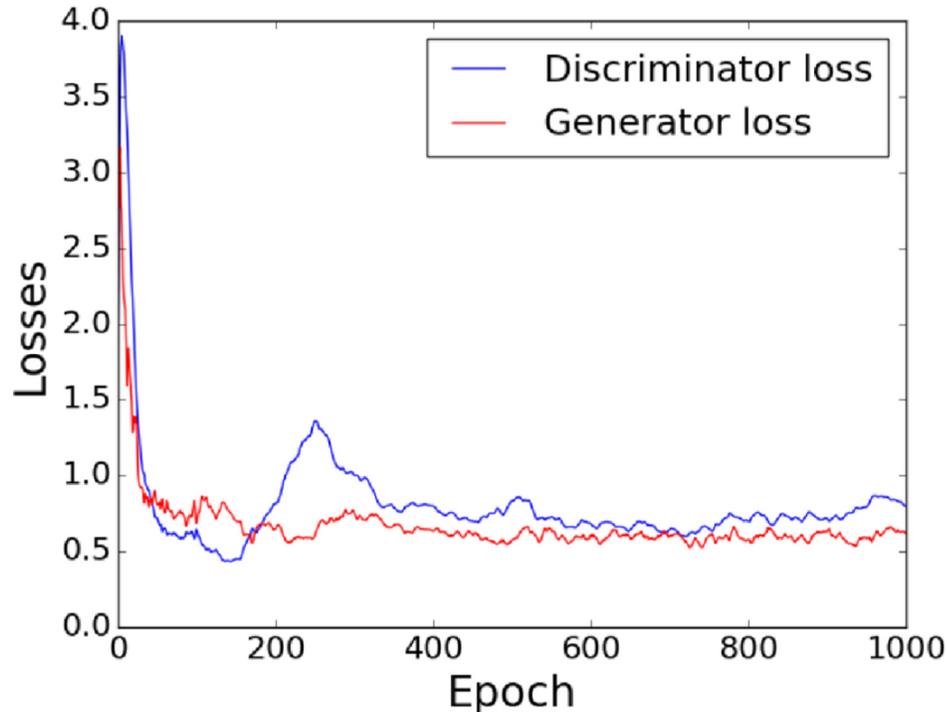


Figure 5. Conversion graph for Loss function of Discriminator and Generator.

The researchers implement their models with the PyTorch framework and deploy CTsGAN on a web application using the Streamlit library. For training, they primarily use the **CT Medical Images** and **COVID-19 CT Scans dataset**, which provides high-quality (2k resolution) images. The total number of images including both datasets is 800. The researchers have found that using this large dataset with richer textures helps the generator produce more natural results. The researchers train their models in the grayscale channel and augment the dataset with random horizontal flips and 90-degree rotations. They evaluate their model on a grayscale converted PRIM self-validation dataset.

F. Web Application

CTsGAN is a powerful Generative Adversarial Network that can be inferred through command line prompts. The inference script takes input such as the input directory, model name, output directory, out scaling factor, tiles, padding, and up sampler. This can be difficult to manage for many users who are not well-versed in the technology. Hence, the researchers create a user-friendly web application to deploy and host CTsGAN using the Streamlit library. Using Hydralit components, they create a multi-page navigation bar to connect the home app, references, and source code page. The home app hosts the basic functionality of CTsGAN. With the help of uploading widgets and sidebar selection boxes, they

call the inference script in the backend whenever an image is uploaded. The image is then processed and displayed in comparison with the original image. The user can also download the full-size output image to save and access it locally at any time. The researchers offer X2 and X4 scaling factor options for CTsGAN in our Web Application for users with various device compatibility, who can enhance the images according to their convenience.

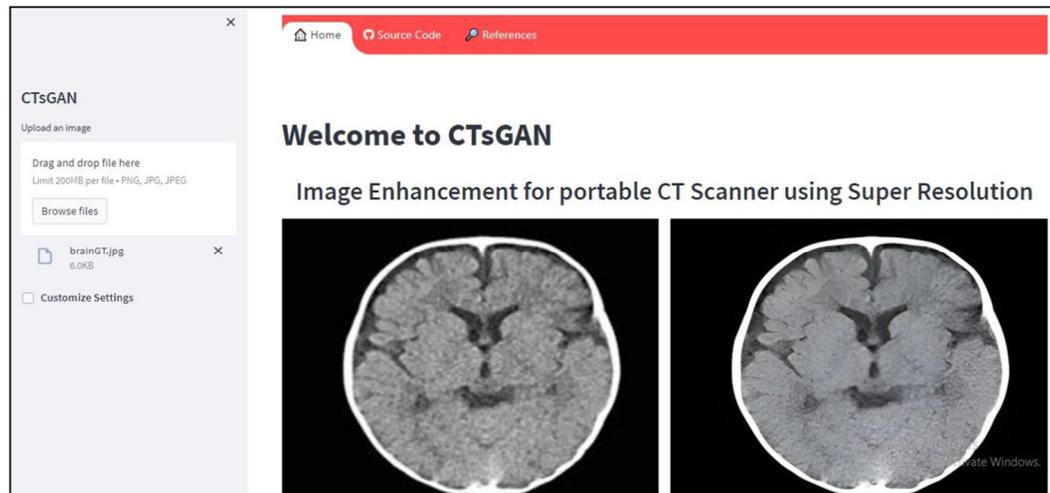


Figure 6. Project Snapshot of Home Page - Web Application for CTsGAN.

In Figure 6, the researchers provide a snapshot of the homepage of the web application for CTsGAN. As visible in Figure 6, the web application's homepage provides the option to upload the image on the left-hand side of the page, providing two alternatives - drag-and-drop image or browse image file from local machine. The researchers also mention the size limit and extension of image for users to choose image from.

In Figure 7 below, the researchers provide another snapshot from the web application where the customizing options for inferring CTsGAN are displayed for users to choose from. As visible in the figure and mentioned before, the researchers provide options in the form of radio buttons in the webpage for the two scaling factors. Also, a drop-down menu is also provided for users to input the restoration strength. After that, a slider is available for users to input the number of tiles.

The screenshot displays the CTsGAN web application interface. At the top, the title "CTsGAN" is centered. Below it, the instruction "Upload an image" is followed by a large white box containing the text "Drag and drop file here". To the right of this box, the upload limits are specified: "Limit 200MB per file • PNG, JPG, JPEG". A "Browse files" button is located below the upload area. Underneath, the "Customize Settings" option is checked with a red checkbox. The "Scaling Factor" is set to "CTSGAN_x4", indicated by a red radio button, with "CTSGAN_x2" as an alternative. The "Restoration Strength" is set to "Medium (default)" via a dropdown menu. At the bottom, the "Tiles" section shows a slider with a red dot at the beginning and a question mark icon to its right.

Figure 7. Customizing options for inferring CTsGAN Web Application.

4. Results & Discussions

The researchers compare their final model, CTsGANX4v2, on the Chest CT Scan Image test dataset with state-of-the-art methods including Blind SRGAN [33], Wavelet frame-based Sparse Reconstruction [34], and Enhanced SRGAN [31]. They present representative qualitative results since there is no effective and standard metric for perceptual quality.

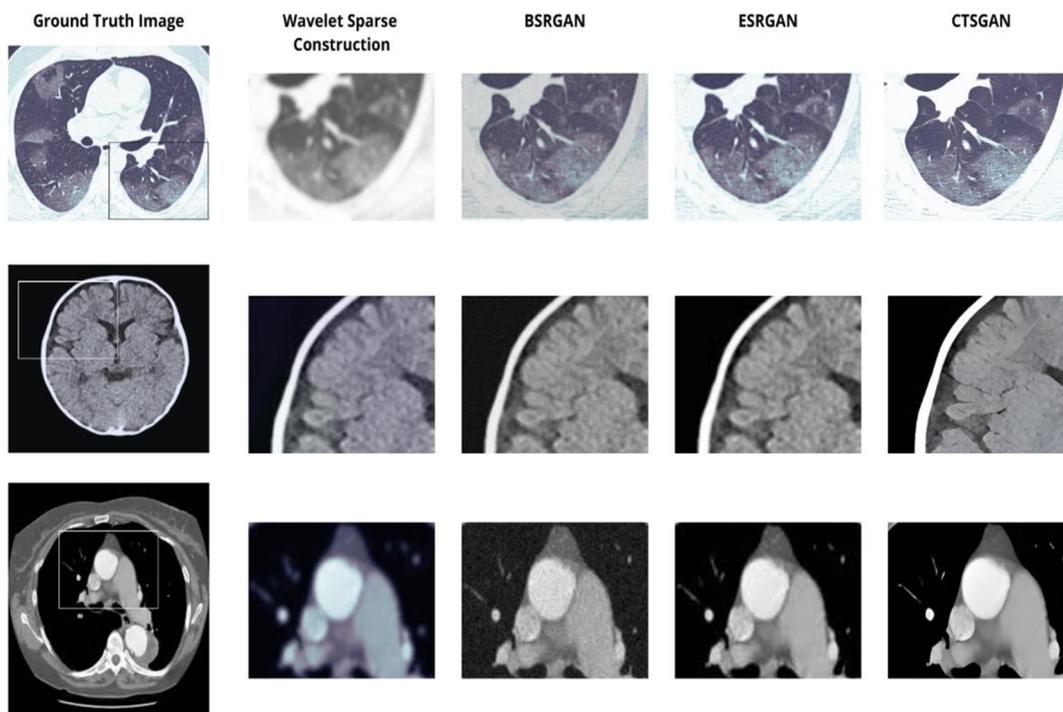


Fig 8: Qualitative comparisons on diverse samples with the upscaling factor of CTsGAN outperform previous approaches in removing artifacts, restoring texture details, and optimal brightness, further boosting visual sharpness. Other methods typically fail to remove complicated artifacts or fail to restore the natural texture and sharp edges where required.

It can be observed that the proposed CTsGAN outperforms previous approaches in sharpness, brightness, and details. For instance, CTsGAN can produce sharper and more dynamically contrasting images. It can generate finer details such as smaller tumor nodules with sharp boundaries where the Blind SRGAN failed to capture the node boundaries effectively. The currently popular methods like Sparse Reconstruction and shock filtering techniques cannot enlarge the image size and create a relatively less clear image when compared to PSNR-based models.

ESRGAN [31] outperforms every other super-resolution GAN in perceptual quality, and we can reach a similar level of perceptual quality with CTsGAN, with much faster performance (compared to ESRGAN). CTsGAN also gets rid of any unpleasant artefacts generated in Blind SRGAN and produces results with natural texture.

The researchers provide the non-reference image quality assessment - NIQE for reference. It's important to note that existing perceptual quality metrics cannot well reflect the actual human perceptual preferences on a fine-grained scale.

Table 1: NIQE scores on diverse datasets

(The lower the score, the better the resolution).

Data Sets	ESRGAN	CTsGAN
CT Medical Images	6.7715	4.5314

CoVID-19 CT Scan Images	6.7480	5.0247
Chest CT Scan Images (test)	3.5245	2.8191

The Table 1 provided by the researcher delves into the Natural Image Quality Evaluation to assess the quality of images from a human perspective. The researchers provide a comparison of the performance of the two Generative Adversarial Networks based on this metric i.e. the ESRGAN and CTsGAN. The table shows that the CTsGAN model produced better quality images on all four datasets according to the NIQE score. The NIQE score for ESRGAN on CT medical images is 6.7715, whereas the NIQE score for CTSGAN on CT medical images is 4.5314.

The researchers also provide an additional qualitative comparison with other super-resolution models like the DAN, Bicubic and BSRGAN respectively in Figure 9 below.

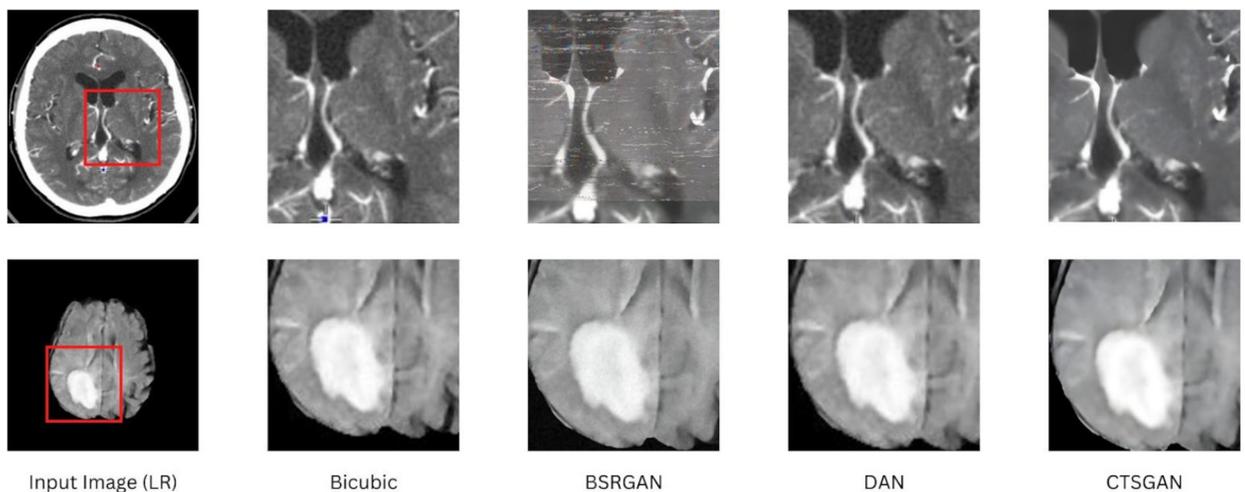


Figure 9: Additional qualitative comparison with previous super-resolution models

5. Conclusion and Future Work

In conclusion, the researchers have presented CTsGAN, a variation of the ESRGAN model, that achieves consistently better perceptual quality for naturally degraded images than previous Image Enhancement methods. Since they typically deal with grayscale images, the network architecture is simplified significantly (reduced computational complexity). They have reintroduced the Batch Normalization Layer to the RRDB Block to create a high-performance novel architecture.

The addition of the BN Layer helps in faster training with an increased learning rate. CTsGAN provides a stunning 2.622 PI, 15.47 RMSE on grayscale PRIM self-validation with modified Perceptual Loss. The researchers observe that CTsGAN can overcome the disadvantages of existing methods and generate high-resolution images free of unpleasant artefacts efficiently. The researchers have also deployed their models (scaling factors 2 and 4) on a user-friendly web application using the Streamlit library. Users can upload their images, compare the output with the input, and also download the full-size enlarged image via the web application. CTsGAN aids medical professionals with high-quality CT Scans which enables them to make important decisions effectively.

Some of the prospective future work on this project is to make CTsGAN compatible with video inputs and work in real-time scenarios. CTsGAN is open-sourced and we encourage the development of ideas and contributions.

6. Author Contributions

Amit Thakur and Nilamadhab Mishra designed the Models and carried out a Literature Review, Model Testing, Conceptualization, Methodology, Field study, Data curation, writing original draft preparation and Result Validation. Amit Thakur carried-out Model Testing, Investigation, Writing-viewing, Editing, Writing draft preparation, Field study, and Result Validation.

7. Data Availability Statement

The data that support the findings of this study are available at the GitHub repository given at https://github.com/Sidhved/CTSGAN_Repo.

8. Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Andersson, H., Tamaddon, A., Malekian, M., Ydström, K., Siemund, R., Ullberg, T. and Wasselius, J., 2023. Comparison of image quality between a novel mobile CT scanner and current generation stationary CT scanners. *Neuroradiology*, 65(3), pp.503-512.
- [2] Chen, H., He, X., Qing, L., Wu, Y., Ren, C., Sheriff, R.E. and Zhu, C., 2022. Real-world single image super-resolution: A brief review. *Information Fusion*, 79, pp.124-145.
- [3] Szczykutowicz, T.P., Toia, G.V., Dhanantwari, A. and Nett, B., 2022. A review of deep learning CT reconstruction: concepts, limitations, and promise in clinical practice. *Current Radiology Reports*, 10(9), pp.101-115.
- [4] Sun, Y., Yang, Z., Tao, B., Jiang, G., Hao, Z. and Chen, B., 2021. Multiscale generative adversarial network for real-world super-resolution. *Concurrency and Computation: Practice and Experience*, 33(21), p.e6430.
- [5] Li, X., Dong, W., Wu, J., Li, L. and Shi, G., 2023. Superresolution Image Reconstruction: Selective milestones and open problems. *IEEE Signal Processing Magazine*, 40(5), pp.54-66.
- [6] Zhou, W. and Xiang, H., 2017. Frame-Based CT Image Reconstruction via the Balanced Approach. *Journal of Healthcare Engineering*, 2017.
- [7] Liu, G., Wang, Z., Mu, G. and Li, P., 2018. Efficient OCT image enhancement based on collaborative shock filtering. *Journal of Healthcare Engineering*, 2018.
- [8] Andersson, H., Tamaddon, A., Malekian, M., Ydström, K., Siemund, R., Ullberg, T. and Wasselius, J., 2023. Comparison of image quality between a novel mobile CT scanner and current generation stationary CT scanners. *Neuroradiology*, 65(3), pp.503-512.
- [9] Hermena, S., Francis, M. and Francis, M.A., 2021. Clinical presentation, imaging features, and management of Müller–Weiss disease. *Cureus*, 13(10).
- [10] Hermena, S. and Young, M., 2021. CT-scan image production procedures. Study Guide from StatPearls Publishing, Treasure Island (FL).
- [11] Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H. and Zhang, L., 2016. Image super-resolution: The techniques, applications, and future. *Signal processing*, 128, pp.389-408.

- [12] Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.H. and Liao, Q., 2019. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 21(12), pp.3106-3121.
- [13] Li, K., Yang, S., Dong, R., Wang, X. and Huang, J., 2020. Survey of single image super-resolution reconstruction. *IET Image Processing*, 14(11), pp.2273-2290.
- [14] Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J. and Norouzi, M., 2022. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4), pp.4713-4726.
- [15] Lepcha, D.C., Goyal, B., Dogra, A. and Goyal, V., 2023. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, 91, pp.230-260.
- [16] El-Shafai, W., Aly, R., Taha, T.E. and Abd El-Samie, F.E., 2023. CNN framework for optical image super-resolution and fusion. *Journal of Optics*, pp.1-20.
- [17] Vb, S.K., 2020. Perceptual image super resolution using deep learning and super resolution convolution neural networks (SRCNN). *Intelligent Systems and Computer Technology*, 37(3).
- [18] Kim, J., Lee, J.K. and Lee, K.M., 2016. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).
- [19] Altakrouri, S., Noor, N.M., Ahmad, N., Justinia, T. and Usman, S., 2023. Image Super-Resolution using Generative Adversarial Networks with EfficientNetV2. *International Journal of Advanced Computer Science and Applications*, 14(2).
- [20] Gong, Z., Xu, Y. and Liu, H., 2023, February. Research on Image Super-resolution Restoration Algorithm Based on High Order Degradation Model. In *2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE)* (pp. 586-590). IEEE.
- [21] Dong, C., Loy, C.C., He, K. and Tang, X., 2015. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2), pp.295-307.
- [22] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 286-301).
- [23] He, K., Zhang, X., Ren, S. and Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
- [24] Rad, M.S., Bozorgtabar, B., Marti, U.V., Basler, M., Ekenel, H.K. and Thiran, J.P., 2019. Srobb: Targeted perceptual loss for single image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2710-2719).
- [25] Sajjadi, M.S., Scholkopf, B. and Hirsch, M., 2017. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE international conference on computer vision* (pp. 4491-4500).
- [26] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A., 2017, February. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1).
- [27] Kastrulin, S., Zakirov, J., Pezzotti, N. and Dyllov, D.V., 2023. Image quality assessment for magnetic resonance imaging. *IEEE Access*, 11, pp.14154-14168.
- [28] Ma, C., Yang, C.Y., Yang, X. and Yang, M.H., 2017. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158, pp.1-16.
- [29] Mittal, A., Soundararajan, R. and Bovik, A.C., 2012. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3), pp.209-212.

- [30] Blau, Y. and Michaeli, T., 2018. The perception-distortion tradeoff. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6228-6237).
- [31] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. and Change Loy, C., 2018. ESRGAN: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0).
- [32] Tej, A.R., Halder, S.S., Shandeelya, A.P. and Pankajakshan, V., 2020, July. Enhancing perceptual loss with adversarial feature matching for super-resolution. In *2020 International joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.
- [33] Zhang, W., Shi, G., Liu, Y., Dong, C. and Wu, X.M., 2022. A closer look at blind super-resolution: Degradation models, baselines, and performance upper bounds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 527-536).
- [34] Zhang, J., Mao, H., Wang, X., Guo, Y. and Wu, W., 2024. Wavelet-Inspired Multi-channel Score-based Model for Limited-angle CT Reconstruction. *IEEE Transactions on Medical Imaging*.