Super-Resolution of Biomedical Images with Generative Adversarial Networks and posterior Tumor Segmentation

João Luís Carrilho Guerreiro
joao.l.carrilho.guerreiro@tecnico.ulisboa.pt
Instituto Superior Técnico, Universidade de Lisboa
Lisboa, Portugal

ABSTRACT

Magnetic Resonance Imaging (MRI) is an expensive medical imaging technique typically associated with long scanning times. MRI acquisition can be potentially accelerated by decreasing the spatial coverage and reducing the number of measured slices. However, this results in a lower MRI resolution and can eventually lead to misleading medical interpretations. An alternative solution comes from recent breakthroughs in Machine Learning, which have shown that high-resolution images can be recovered via super-resolution, particularly through Generative Adversarial Networks. This thesis conducts a review on GAN-based SR methods, exhibiting the immersive ability of GANs on upscaling MRIs by a ×4 scale factor while at the same time maintaining trustworthy and high-frequency details. Despite quantitative results suggesting SRResCycGAN outperforms other popular deep learning methods in recovering ×4 downgraded images, qualitative results show Beby-GAN holds the best perceptual quality and proves GAN-based methods hold the capacity to reduce medical costs and enable MRI applications where it is currently too slow or expensive. Additionally, Tumor Segmentation is utilized to validate the proficiency of GANs in the MRI reconstruction task. Tumor Segmentation of the synthesized images advocates marginal dissimilarities, thus there is a window for improvement. Furthermore, this thesis suggests that a chain of processes for a faster diagnosis can be conceived by merging both Super-Resolution and Tumor Segmentation. Essentially, tumor segmentation algorithms benefit from the improved spatial resolution derived from super-resolution. The diagnosis process is accelerated by acquiring low-resolution MRIs and subsequently upscaling them (via super-resolution) to detect tumors.

KEYWORDS

Computer Vision, Medical Imaging, Super-Resolution, MRI Acceleration, Generative Adversarial Networks, Tumor Segmentation

1 INTRODUCTION

Magnetic resonance imaging (MRI) is a medical imaging technique that is predominantly necessary across patient diagnoses and medical tracking of ongoing diseases. The detailed information of organs, soft tissues, and bones extracted from an MRI scan allows physicians to effectively evaluate, adjust and control treatments. A relevant problem that arises is the prolonged MRI acquisition time. Moreover, a slight movement from the patient can ruin the scan, requiring retesting. Hence, patients have to lie still in the scanners and even hold their breath for thoracic or abdominal imaging [6] since even the slightest movement of breathing can ruin the results. Therefore, the slow acquisition of MRI scans manifests discomfort among subjects and presents inconvenience in healthcare.

The desired image quality also impacts the acquisition time. The decrease in acquisition time is proportional to the spatial resolution reduction. If an MRI is acquired with half the resolution, then the acquisition time is practically halved [5] (excluding scanning preparation and/or pre-scanning time). Therefore, the ability to infer a high-resolution (HR) image from a low-resolution (LR) image yields a massive impact on the performance of image analysis and MRI acceleration. A convenient concept in Machine Learning was introduced, called Image Super-Resolution (SR), referred to as the task responsible for the reconstruction of an image from low to high resolution. After running an MRI scan faster and gathering less raw data, an SR method can be exploited to reconstruct the MRI. Since collecting that data is what makes MRI so slow, this concept can speed up the scanning process significantly.

In general, SR methods are based on Generative Adversarial Networks (GANs), which were introduced in 2014 by Goodfellow *et al.* [7] and have recently gained a lot of attention. GANs introduce an alternative way of conceiving models capable of generating data, entitled generative models, and recently they have been used for several image-based applications. To complement Super-Resolution and provide a more sophisticated and fast diagnoses, automated tumor segmentation can be considered. MRIs have been widely utilized to detect and evaluate brain tumors. However, the amount of detailed information present in MRIs poses a significant problem, as it prevents manual segmentation in a reasonable time. Distinct tumoral subregions can be perceived, and accurately detecting these regions within the MRI is consequential. Similarly to super-resolution, there is a concept in Machine Learning entitled Semantic Segmentation (SS). It is the process dedicated to associating each pixel of an image with a class label.

Succeeding a rigorous analysis of the state-of-the-art, several GAN-based models were selected based on a comprehensive selection criteria that took into consideration several key aspects, such as the performance under multiple applications and the publication date. The performance of these models is evaluated over FastMRI [30]. Furthermore, super-resolved MRIs having similar results as ground-truth MRIs in the tumor segmentation task is suggestive of an accurate recovery of the details inherent to high-resolution MRIs. Therefore, tumor segmentation algorithms and techniques were considered to employ a task-based evaluation intended to assess the GAN-based super-resolution performance. Essentially, the super-resolution performance is estimated by assessing the tumor segmentation performance of the super-resolved brain MRIs. Additionally, through the application of tumor segmentation methods, not only the reconstruction quality of GANs is exposed, but it

also supports the idealized pipeline for a faster diagnosis, where Super-Resolution and Tumor Segmentation are consolidated.

2 RELATED WORK

2.1 Super-Resolution

2.1.1 ***Interpolation-based Upsampling Methods.*** Image Interpolation is the task of resizing images from one pixel grid to another by estimating the pixel intensities of the interpolated points. Interpolation algorithms, such as the Nearest Neighbor, Bilinear, and Bicubic Interpolation, can be very efficient and easy to implement. However, despite being the simplest way to upscale an image, these interpolation methods oversimplify the SR problem and in most cases attain solutions with excessively smooth textures [4].

2.1.2 ***Deep Learning Methods.*** In practice, super-resolution is a problem of missing data. Lost data cannot be recovered by further processing, i.e, information that is not present cannot be inferred. This is where Neural Networks manifest significant value, considering they can learn to conceive details based on some prior information they have extracted from a large training sample.

GANs employ a clever strategy to train a generative model by posing the super-resolution task as a supervised learning problem. They consist of two adversarial Neural Networks that compete with each other. The first network, denoted as Generator, captures the data distribution, while the second one, named Discriminator, estimates the probabilities of samples being real or fake. A generalized application of GANs applied on the SR task is shown in Figure 1.

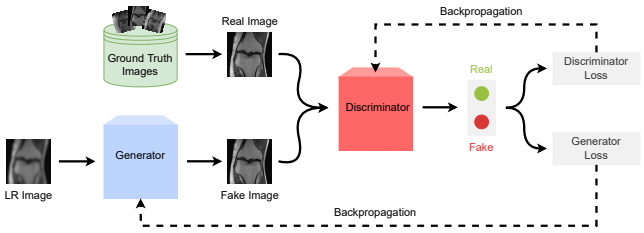

Figure 1: Main concept behind GANs.

2.2 Tumor Segmentation

2.2.1 ***Conventional Methods for Semantic Segmentation.*** Classical methods were usually based on pixel value comparisons between regions. These methods perceive image features locally while considering variations and gradients of pixel values. They are divided into three main categories: threshold-based [11], edge-based [3] and region-based [26].

2.2.2 ***Deep Learning Methods for Semantic Segmentation.*** Semantic Segmentation based on Neural Networks (NN) is feasible due to the unfolding of large medical datasets and the reduction of computing requirements necessary to process them. Furthermore, developments in the deep learning field have greatly advanced the performance of these state-of-the-art visual recognition systems, thus leading neural networks to surpass the hard work of traditional machine learning models.

Among the diverse CNN-based models, Fully Convolutional Networks (FCNs) [15] introduced a novel strategy to solve semantic segmentation (see Section 5.2). They received a lot of attention by exhibiting that convolutional networks can be trained to accommodate pixel-level classification in an end-to-end manner.

Furthermore, built upon the concept of FCNs, U-Net was proposed by Ronneberger *et al.* [21]. It was designed for biomedical image segmentation. However, it has proven to be generalizable for practically any semantic segmentation task.

The main difference compared to FCN is that U-Net network is symmetric, i.e, the second half of this architecture, regarded as the decoder, is the mirror version of the first half, the encoder. Reasoning, the popularity of encoder-decoder architectures for semantic segmentation was solidified with the onset of works U-Net.

3 SUPER-RESOLUTION GENERATIVE MODELS

3.1 SRGAN

Most methods reviewed in this work were inspired by SRGAN [14], which was a novel super-resolution approach using the GAN concept. Before SRGAN, the most relevant work had largely focused on minimizing the mean squared reconstruction error (MSE), however the resulting estimates failed to match the fidelity present at the high resolution domain. To cope with this issue, SRGAN introduces a new GAN architecture (see Figure 2) and diverges from MSE as the single target for optimization.

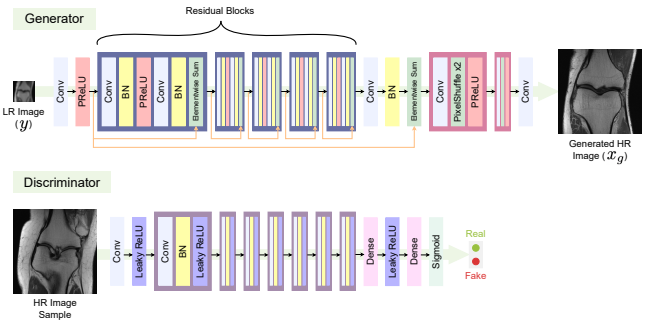

Figure 2: Basic architecture of SRResNet (SRGAN). Figure adapted from [14].

3.2 ESRGAN

Based on the SRGAN pioneer work [14], a model named Enhanced SRGAN (ESRGAN) [28] was introduced to reduce unpleasant artifacts present in the SRGAN generated data. ESRGAN revisits three key components to improve the previous approach: network architecture, adversarial loss and perceptual loss.

The original SRGAN model is built with residual blocks [8] and optimized using a perceptual loss in a GAN framework. Meanwhile, ESRGAN improves the generator structure by removing Batch Normalization (BN) layers and introducing the Residual-in-Residual Dense Block (RRDB), which is of higher capacity and easier to train (see Figure 3).

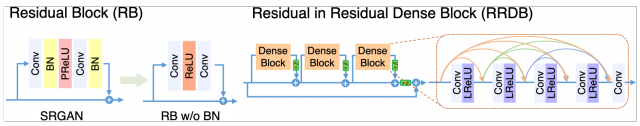

Figure 3: Batch normalization removal on the left. On the right, Residual in Residual Dense Block is embedded in the model. Figure adapted from [28].

3.3 RankSRGAN

Perceptual quality can be assessed by perceptual metrics, such as Perceptual Index (PI) [2], Natural Image Quality Evaluator (NIQE) [19], and Ma [16], which are highly correlated with human perception. However, existing methods cannot directly optimize these metrics. Therefore, to optimize a network in the direction of these perceptual metrics an approach was proposed consisting of a GAN with a Ranker, named RankSRGAN [32]. The Ranker adopts a Siamese architecture to learn the behaviour of perceptual metrics as depicted in the middle section of Figure 4.

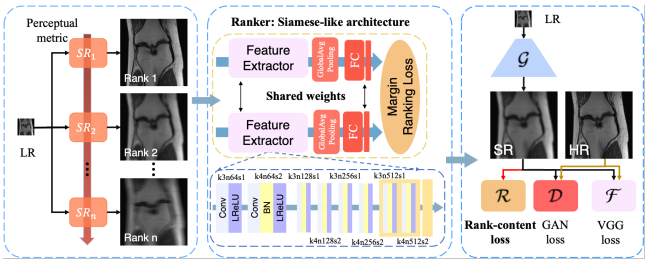

Figure 4: Overview of RankSRGAN. Essentially, RankSRGAN consists of a generator ($\mathcal{G}$), a discriminator ($\mathcal{D}$), a fixed feature extractor ($\mathcal{F}$) and a ranker ($\mathcal{R}$). Figure adapted from [32].

3.4 SRResCycGAN

Inspired by the success of CycleGAN [33] in image-to-image translation applications, a new deep cyclic network structure was proposed, named SRResCycGAN [25]. In essence, a GAN is trained to achieve LR to HR translation in an end-to-end manner. MRI acquisition can contain a significant amount of noise caused by operator performance, patient motion, equipment or environment, leading to unpleasant results. SRResCycGAN overcomes this challenge and maintains the domain consistency between the LR and HR data distributions by following the CycleGAN structure, as shown in Figure 5.

3.5 BSRGAN

Single Image Super-Resolution (SISR) methods would not perform well if the assumed degradation model deviates from those in real images. Therefore, a model named BSRGAN [31] was proposed along with a degradation model. A deep blind ESRGAN is trained based on the degradation model, which consists of randomly shuffled blur, downsampling and noise degradations as shown in Figure 6.

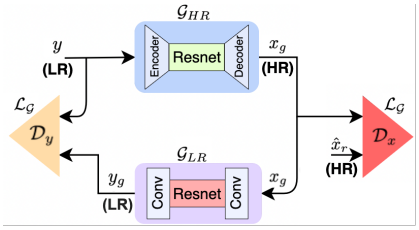

Figure 5: SRResCycGAN structure. Figure adapted from [25].

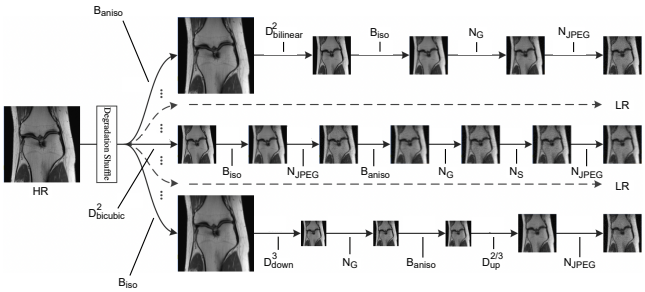

Figure 6: Proposed BSRGAN degradation model for a scale factor of 2. For scale a factor of 4, an additional bilinear or bicubic downscaling is applied. The type of blur employed is denoted by B_{type} and N_{type} is the type of noise. Meanwhile, D^{scale}_{type} stands for the downsampling applied under a defined scale. Figure adapted from [31].

3.6 Real-ESRGAN

The previous ESRGAN approach is extended to achieve superior visual performance on various datasets. Real-ESRGAN [27] aims to restore general real-world LR images by synthesizing training pairs with a more practical degradation process. In essence, starts by improving the VGG-style discriminator in ESRGAN to a U-Net design [22]. Then, employs the Spectral Normalization (SN) regularization [20] to stabilize the training process, since the U-Net structure and complicate degradations also increase the training instability. Real-ESRGAN uses a synthetic data generation process as depicted in Figure 7. Consequently, Real-ESRGAN robustness is improved and is capable of restoring more realistic textures for real-world samples, while other methods either fail to remove degradations or add unnatural textures.

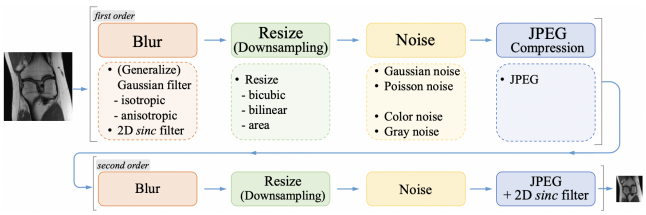

Figure 7: High-order Degradation Model. Figure adapted from [27].

3.7 Learning Strategies

This section discusses learning strategies utilized in super-resolution. Furthermore, a concise comparison of the numbers of parameters and generator losses from each GAN model regarded in this work is given in Table 1.

Table 1: Comparison of GAN-based SR models. $\mathcal{L}_{\mathcal{P}}$ represents the perceptual loss, $\mathcal{L}_{\mathcal{G}}$ the adversarial loss, $\mathcal{L}_{\mathcal{R}}$ the rank-content loss, $\mathcal{L}_{\mathcal{CYC}}$ the cyclic loss, $\mathcal{L}_{\mathcal{BB}}$ the best-buddy loss, $\mathcal{L}_{\mathcal{TV}}$ the total-variation loss and $\mathcal{L}_1$ the content loss. Moreover, λ, η, θ and ϕ are coefficients to balance the different loss components.

Method	Parameters	Loss
SRGAN	16.7M	$\mathcal{L}_{\mathcal{P}} + \lambda\mathcal{L}_{\mathcal{G}}$
ESRGAN	16.7M	$\mathcal{L}_{\mathcal{P}} + \lambda\mathcal{L}_{\mathcal{G}} + \eta\mathcal{L}_1$
RankSRGAN	1.55M	$\mathcal{L}_{\mathcal{P}} + \lambda\mathcal{L}_{\mathcal{G}} + \eta\mathcal{L}_{\mathcal{R}}$
SRResCycGAN	380k	$\mathcal{L}_{\mathcal{P}} + \mathcal{L}_{\mathcal{G}} + \mathcal{L}_{\mathcal{TV}} + \lambda\mathcal{L}_1 + \eta\mathcal{L}_{\mathcal{CYC}}$
BSRGAN	16.7M	$\mathcal{L}_{\mathcal{P}} + \lambda\mathcal{L}_{\mathcal{G}} + \eta\mathcal{L}_1$
Beby-GAN	16.7M	$\lambda\mathcal{L}_{\mathcal{BB}} + \eta\mathcal{L}_{\mathcal{BP}} + \theta\mathcal{L}_{\mathcal{P}} + \phi\mathcal{L}_{\mathcal{G}}$
Real-ESRGAN	16.7M	$\mathcal{L}_{\mathcal{P}} + \lambda\mathcal{L}_{\mathcal{G}} + \eta\mathcal{L}_1$

3.7.1 ***Perceptual Loss*** *($\mathcal{L}_{\mathcal{P}}$).* Proposed by Johnson *et al.* [10] to measure the perceptual similarity between two images and enhance the visual quality by minimizing the error in a feature space rather than pixel space. Perceptual loss can be expressed in the equation below:

$$\mathcal{L}_{\mathcal{P}} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{\mathcal{VGG}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\phi\left(\hat{x}_{r_i}\right) - \phi\left(x_{g_i}\right)\right\|_2^2, \tag{1}$$

where x_{g_i} represents the generated HR image and $\hat{x}_{r_i}$ is the corresponding ground truth image. Moreover, N represents the number of training samples and $\phi(\cdot)$ denotes the image feature maps obtained by some convolution layer within the VGG19 network [23].

3.7.2 ***Adversarial Loss*** *($\mathcal{L}_{\mathcal{G}}$).* The standard GAN loss function introduced by Goodfellow *et al.* [7] corresponds to a min-max game approach, therefore it is also known as the min-max loss. The generator tries to minimize the following function while the discriminator tries to maximize it:

$$\min_{\theta_{\mathcal{G}}}\max_{\theta_{\mathcal{D}}} \mathbb{E}_{x_r}[\log(\mathcal{D}_{\theta_{\mathcal{D}}}(x_r))] + \mathbb{E}_y[\log(1 - \mathcal{D}_{\theta_{\mathcal{D}}}(\mathcal{G}_{\theta_{\mathcal{G}}}(y)))], \tag{2}$$

where x_r denotes a real image and $x_g = \mathcal{G}_{\theta_{\mathcal{G}}}(y)$ represents a generated HR image when given input LR image y. Additionally, $\mathbb{E}_{x_r}$ corresponds to the expected value over all real data instances and $\mathcal{D}_{\theta_{\mathcal{D}}}(x_r)$ is the discriminator's estimate of the probability that a real data instance x_r is real. Meanwhile, $\mathbb{E}_y$ is the expected value over all input LR instances y and, in consequence, the expected value over all generated fake instances x_g. In addition, $\mathcal{D}_{\theta_{\mathcal{D}}}(\mathcal{G}_{\theta_{\mathcal{G}}}(y)))$ is the discriminator's estimate of the probability that a generated image is real. Moreover, $\theta_{\mathcal{G}}$ and $\theta_{\mathcal{D}}$ denote the weights and biases that parameterize the generator network $\mathcal{G}$ and discriminator network $\mathcal{D}$, respectively.

The generator and discriminator are jointly optimized with the objective given in function (2). Looking at it as a min-max game, this formulation of the loss enables the function above to be categorized into two equations formulating the Discriminator and Generator losses. Accordingly, the generator loss $\mathcal{L}_{\mathcal{G}}$ is defined based on the discriminator's output and only affects the right term of the expression (2), the term that reflects the distribution of the generated data. Therefore, during the generator's training the left term is dropped, since it only reflects the distribution of the real data. In essence, the adversarial loss for the generator can be represented as follows:

$$\mathcal{L}_{\mathcal{G}} = \frac{1}{N}\sum_{i=1}^{N} -\log(\mathcal{D}_{\theta_{\mathcal{D}}}(\mathcal{G}_{\theta_{\mathcal{G}}}(y_i))), \tag{3}$$

where N represents the number of LR training samples and y_i is a input LR image.

3.7.3 ***Content Loss*** *($\mathcal{L}_1$ and $\mathcal{L}_2$).* Reasonably the most used optimization target in SR applications due to its simplicity and decent results. From this class of loss functions many variants are formulated, such as $\mathcal{L}_1$ and $\mathcal{L}_2$. These loss functions are in charge of optimizing the error between pixel values corresponding to the generated and ground truth images. Reducing the distance between pixels can effectively ensure the quality of the reconstructed image and therefore hold a higher peak signal to noise ratio value.

Regarding $\mathcal{L}_1$, also known as Mean Absolute Error (MAE), it is computed by averaging the sum of the absolute differences between predictions and actual observations:

$$\mathcal{L}_1 = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathcal{G}(y_i) - \hat{x}_{r_i}\right\|_1, \tag{4}$$

where $\mathcal{G}(y_i)$ represents a generated HR image x_{g_i} when given an LR image y_i and $\hat{x}_{r_i}$ is the corresponding ground truth image.

Concerning $\mathcal{L}_2$, also known as Mean Square Error (MSE) or quadratic loss, it is computed by averaging the sum of the squared differences between generated and real images:

$$\mathcal{L}_2 = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{G}(y_i) - \hat{x}_{r_i})^2, \tag{5}$$

Due to the squaring operation, the predictions that are far away from the actual values are heavily penalized in comparison to those less deviated.

3.8 Implementation Details

3.8.1 ***ESRGAN.*** The ESRGAN model training is performed with mini-batch size set to 16. The generator is trained with a learning rate of 1×10^{-4} and decayed every 1×10^5 mini-batch updates by a rate of 2. The optimizer employed is Adam [13] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

3.8.2 ***RankSRGAN.*** The ranker is trained over DIV2K [1] and Flickr2K [24] datasets. For optimization, the Adam optimizer [13] is used with a weight decay of 1×10^{-4}. The learning rate is set to 1×10^{-3} and is decayed with a factor of 2 every 1×10^5 iterations. Concerning the pre-trained RankSRGAN network, the training is

carried out with a mini-batch size of 8. The optimization target is defined in Table 1, where $\lambda = 5 \times 10^{-3}$ and $\eta = 3 \times 10^{-2}$. To optimize the network, the Adam optimizer [13] is employed with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Both generator and discriminator learning rates are initialized to 1×10^{-4} and halved every 1×10^5 iterations.

3.8.3 ***SRResCycGAN.*** The training phase is carried out with a batch size of 16 over 51×10^3 iterations. For optimization, the Adam optimizer [13] is employed with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and no weight decay. The learning rate is initialized to 1×10^{-4} and decayed with a factor of 2 every 10^4 iterations.

3.8.4 ***BSRGAN.*** Pre-trained with batch size of 48 over a unified dataset including DIV2K [1], Flick2K [24], WED [17] and FFHQ [12]. BSRGAN is trained by minimizing a weighted combination of losses, as shown in Table 1, where $\lambda = 0.1$ and $\eta = 1$. For optimization, the Adam optimizer [13] is employed with a fixed learning rate of 1×10^{-5}.

3.8.5 ***Real-ESRGAN.*** Since the same generator architecture from ESRGAN [28] is adopted, then initially a network from ESRGAN is finetuned for faster convergence. Both the generator and discriminator of Real-ESRGAN model are trained for 4×10^5 iterations with Adam [13] as optimizer. The learning rate is set to 1×10^{-4} with $\beta_1 = 0.9$, $\beta_2 = 0.99$ and no weight decay. For optimization, the equation in Table 1 is minimized, where $\lambda = 0.1$ and $\eta = 1$.

4 SUPER-RESOLUTION EXPERIMENTS

4.1 Data

4.1.1 ***FastMRI Dataset.*** To test every GAN method mentioned in section 5 the FastMRI dataset [30] was employed. FastMRI is a large-scale release of raw MRI data. It consists of two collections: knee MRIs and brain MRIs. Each collection is split into training, validation, and downsampled/masked test sets. Considering both collections and all splits, FastMRI contains a total of 8344 MRI volumes, corresponding to 167.375 slices, where each slice corresponds to one 2D image. In this work only the knee collection is considered, for instance, 973 volumes were used from the single-coil knee training set.

4.1.2 ***Image Preprocessing.*** The training set yields, for each slice, the k-space data and the corresponding ground truth. To evaluate the super-resolution performance it is necessary to formulate LR-HR image pairs. Consequently, a preprocessing step is employed to simulate the degradation inherent to MRI acquisition under few measurements. At the beginning of the test phase each k-space data of every MRI slice is downsampled through bicubic interpolation with a downscale factor of ×4, resulting in LR-HR pairs holding the downsampled k-space data and the ground truth (reconstructed from fully-sampled multi-coil acquisitions using the simple root-sum-of-squares method).

4.2 Image Quality Metrics

Several Image Quality Metrics (IQMs) are used to evaluate models' performances quantitatively.

4.2.1 ***Mean Squared Error (MSE).*** Among the many IQM used to evaluate the HR image quality, Mean Squared Error (MSE) is the most popular metric. It is computed by averaging the pixel-wise squared differences between the generated HR image and the corresponding ground truth. The MSE between two images is given as follows:

$$MSE = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} (\hat{x}_r(i,j) - x_g(i,j))^2, \tag{6}$$

where W denotes the image width and H the image height. Moreover, (i, j) define the pixel position, while $\hat{x}_r$ and x_g represent the ground truth and generated HR images, respectively. Evidently, both images must share the same size.

4.2.2 ***Peak Signal-to-Noise Ratio (PSNR).*** It is commonly used to measure the reconstruction quality, and is inversely proportional to the logarithm of the MSE between the ground truth and the HR generated image. PSNR is expressed in the following equation:

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX_I}{RMSE(\hat{x}_r, x_g)} \right), \tag{7}$$

where MAX_I corresponds to the maximum possible pixel value, for instance, 255 regarding 8-bit images.

4.2.3 ***Structural Similarity Index Measure (SSIM).*** Structural Similarity Index Measure (SSIM) is based on luminance, contrast, and changes in structural information. The key idea behind considering structural information changes is that pixels are strongly correlated especially when they are spatially close. SSIM can be defined as follows:

$$SSIM = \frac{\left(2\mu_{\hat{x}_r}\mu_{x_g} + c_1\right)\left(2\sigma_{\hat{x}_r x_g} + c_2\right)}{\left(\mu_{\hat{x}_r}^2 + \mu_{x_g}^2 + c_1\right)\left(\sigma_{\hat{x}_r}^2 + \sigma_{x_g}^2 + c_2\right)}, \tag{8}$$

where $\mu_{\hat{x}_r}$ and μ_{x_g} represent the means of the ground truth and the generated HR image, respectively. Accordingly, $\sigma_{\hat{x}_r}$ and σ_{x_g} are the standard deviations of $\hat{x}_r$ and x_g. Moreover, $\sigma_{\hat{x}_r x_g}$ denotes the covariance between both images, while c_1 and c_2 are constants set to avoid instability.

4.3 Quantitative Results

All experiments were conducted on Google Colab using an Intel Xeon CPU with 2.20GHz and 13GB of RAM. Results can be seen in Table 2. For every method the LR images were obtained with bicubic downsampling and a scaling factor of ×4. Time (ms) column shows the average time in milliseconds spent to reconstruct an 80×80 degraded MRI slice into a HR one with size 320×320. Moreover, the scale column denotes the upscaling factor.

As can be seen in Table 2, MSE, PSNR, and SSIM suggest SRResCycGAN outperforms every other GAN-based method in recovering ×4 downgraded images.

4.4 Qualitative Results

Ultimately, despite quantitative results suggesting SRResCycGAN outperforms other popular deep learning methods in recovering ×4 downgraded images, qualitative results (see Figure 8) show Beby-GAN holds the best perceptual quality and proves GAN-based methods hold the capacity to reduce medical costs, distress patients and even enable new MRI applications where it is currently too slow or expensive.

Table 2: Results Comparison. Red color indicates the worst performance overall and Green color the best. Gray color stands for the additional time derived from the denoise step.

Method	Input	Scale	Optimizer	Datasets	MSE	PSNR	SSIM	Time (ms)
ESRGAN	Bicubic	×4	Adam	DIV2K, Flickr2K	297.46	24.47	0.5939	4417
RankSRGAN	Bicubic	×4	Adam	DIV2K, Flickr2K	266.94	24.99	0.6319	651
SRResCycGAN	Bicubic	×4	Adam	AIM2020 RISR	228.00	25.94	0.7456	2602
BSRGAN	Bicubic	×4	Adam	DIV2K, Flick2K, WED, FFHQ	254.11	25.33	0.7157	3652
Beby-GAN	Bicubic	×4	Adam	DIV2K, Flickr2K	264.76	25.11	0.6493	3819
Real-ESRGAN	Bicubic	×4	Adam	DIV2K, Flickr2K, OutdoorSceneTraining	274.40	24.99	0.7137	3715

Figure 8: Super-Resolution qualitative results.

4.5 Discussion

Using perceptual loss as a term in the loss function will encourage natural and perceptually pleasing results. However, this can be misleading in the medical imaging context, for instance MRI, since the reconstructed MRI may look natural and real, but not equal to the ground truth. This dissimilarity due to artifacts inclusion or omissions of relevant details can lead to erroneous conclusions. The same occurs in adversarial training with GANs, usually used to attain photo-realism. The discriminator predicts relative realness instead of the absolute value. Consequently, realistic fake patterns can be wrongly conjectured as real even if they are far from the ground truth. However, the function that perfectly recovers the target image might be impossible to estimate, since the reconstruction problem is inherently ill-posed, i.e., for any distorted image there can be multiple plausible solutions that would be perceptually pleasing. Therefore, GANs remain a solid candidate to spatially resolve MRIs and accelerate their acquisition.

Additionally, optimizing to the content loss usually leads to unnatural and overly smooth reconstructions with low perceptual quality. In contrast, the distortion-based performance is improved, since they focus on minimizing pixel-wise errors. Alternatively, focusing on the adversarial loss leads to a perceptually better reconstruction, but as aforementioned it tends to decrease the distortion-based quality. Therefore, finding a balance between both optimization targets is the best option. Nonetheless, it is evident that the ideal loss function depends on the application where super-resolution is employed. For example, approaches that hallucinate finer detail might be less suited for medical applications or surveillance.

Ultimately, despite quantitative results suggesting SRResCycGAN outperforms other popular deep learning methods in recovering ×4 downgraded images, qualitative results show Beby-GAN holds the best perceptual quality and proves GAN-based methods hold the capacity to reduce medical costs, distress patients and even enable new MRI applications where it is currently too slow or expensive.

5 MODELS FOR TUMOR SEGMENTATION

5.1 Traditional Machine Learning Methods

A clever strategy to perform semantic segmentation is by merging all the conventional techniques to extract a set of features. Afterwards, a traditional machine learning algorithm, such as Random Forest, Support Vector Machine, or XGBoost, can be trained with these features to recognize patterns and make pixel-level predictions.

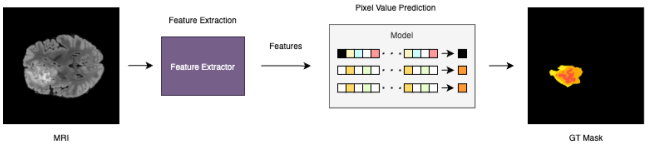

Figure 9: Main concept behind Semantic Segmentation with Traditional Machine Learning algorithms.

5.2 Deep Learning Methods

5.2.1 ***Fully Convolutional Network (FCN).*** Fully Convolutional Networks (FCNs) were a pioneer CNN architecture designed to solve spatially dense prediction tasks. Long *et al.* [15] suggested the removal of the fully connected layers from CNNs, as these layers can be thought of as doing 1×1 convolutions. From this work three distinct architectures were proposed (see Figure 10).

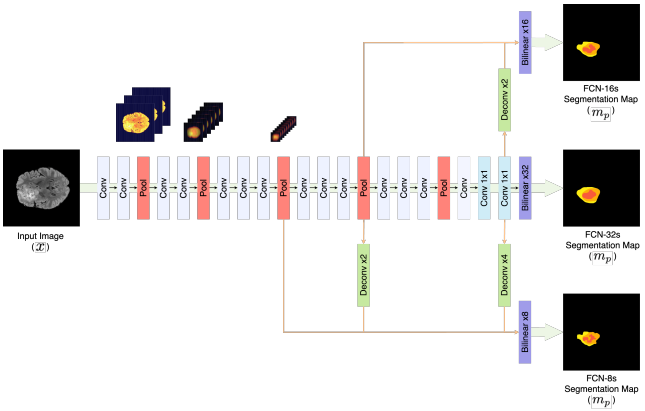

Figure 10: Architecture of FCN-32s, FCN-16s and FCN-8s. Feature maps are illustrated in a distinct colormap (not grayscale) for visualization purposes and ease of comprehension. Activation functions are omitted for the same reason.

5.2.2 ***U-Net.*** The FCN architecture is modified and extended by Ronneberger *et al.* [21] in order to excel with very few data and yield precise segmentations. From this work, an FCN-based semantic segmentation architecture entitled U-net was proposed. The name U-net comes from its peculiar U-shaped architecture (see Figure 11) and consists of an encoder that downsamples the input image to a feature map and a decoder that adversely upsamples back the feature map to the input image size using learned upsampling layers.

5.2.3 ***Open BraTS Solution.*** Henry *et al.* [9] trained multiple U-Net based neural networks to automate and standardize brain tumor segmentation. Two independent ensembles of models from different training pipelines were trained. In each pipeline, the execution of a model was repeated several times and at the end the saved weights were averaged, effectively creating a new self-ensembled model. Afterwards, both pipeline segmentation maps were merged, taking into account the performance of each ensemble for each

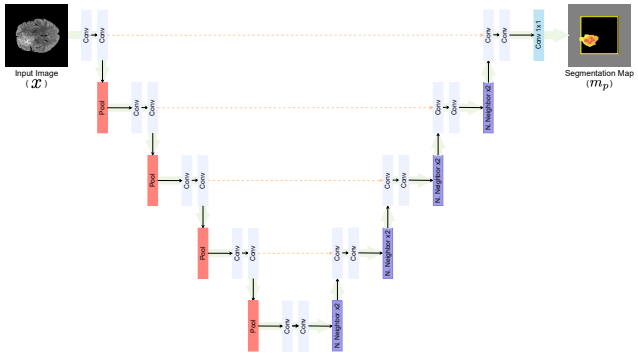

Figure 11: U-Net Architecture. Dashed orange lines denote skip connection and concatenation operations.

specific tumor subregion. The network employed follows a 3D U-Net architecture with convolutional and max pooling layers in the encoder part. Regarding the decoder, is consists of convolutional and upsampling layers. A $1 \times 1 \times 1$ convolution follows the last step of the upsampling to map each feature vector to the desired dimensionality (see Figure 12).

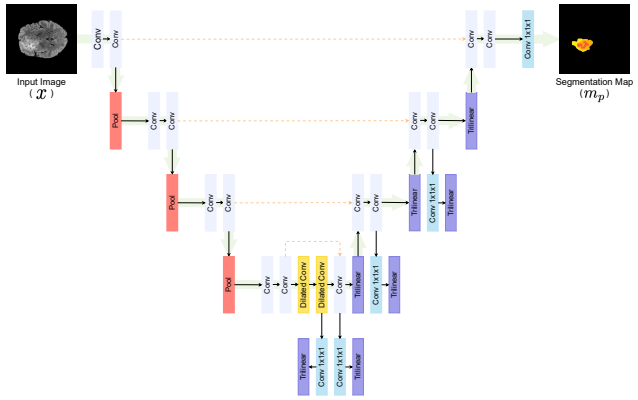

Figure 12: Open BraTS Solution Architecture. Dashed orange lines denote skip connection and concatenation operations.

5.3 Learning Strategies

5.3.1 ***Dice Loss*** ($\mathcal{L}_{Dice}$). This loss derives from the Dice Similarity Coefficient (see Section 6.2.1), a widely used metric in computer vision to estimate the similarity between two images or volumes, for instance, segmentation masks. High Dice scores translate to high-fidelity segmentations. Therefore, this loss can be defined, directly in terms of the Dice coefficient, as follows:

$$\mathcal{L}_{Dice} = \frac{1}{N} \sum_{n=1}^{N} 1 - Dice(\hat{m}_t, m_p), \tag{9}$$

where N denotes the number of training samples, $\hat{m}_t$ is the ground truth segmentation mask and m_p the predicted mask. Reasoning, the smaller the Dice score, the greater the loss.

5.3.2 ***Jaccard Loss*** **($\mathcal{L}_{Jaccard}$)**. The Jaccard loss is frequently referred to as the intersection-over-union (IoU) loss. Similar to Dice Loss and derived from the Jaccard Index (see Section 6.3), this loss is employed to optimize the segmentation task through minimization of the given equation:

$$\mathcal{L}_{Jaccard} = \frac{1}{N}\sum_{n=1}^{N} 1 - Jaccard(\hat{m}_t, m_p), \tag{10}$$

where N denotes the number of training samples, $\hat{m}_t$ is the ground truth segmentation mask and m_p the predicted mask. Implicitly, minimizing equation (10) implies maximizing the Jaccard Index, which consequently advocates a superior segmentation quality.

5.3.3 ***Cross-Entropy Loss*** **($\mathcal{L}_{CE}$)**. Cross-Entropy (CE) intends to measure the differences in information content between the ground truth and predicted segmentation maps. The segmentation output is required to be a distribution of probabilities. Essentially, each pixel has an estimated probability distribution representing the predicted probability for each class. CE punishes how close to zero is the predicted probability of the actual ground truth class. The penalty is logarithmic, yielding larger absolute values for probability estimations that are close to zero and smaller values for estimations tending to one. Cross-Entropy loss is defined as follows:

$$\begin{aligned} \mathcal{L}_{CE} &= \frac{1}{N}\sum_{n=1}^{N} CE\left(\hat{m}_{t_n}, m_{p_n}\right) = \\ &= -\frac{1}{NWH}\sum_{n=1}^{N}\sum_{i=1}^{W}\sum_{j=1}^{H}\sum_{c=1}^{C} P^c_{t_n}(i,j)\cdot\log(P^c_{p_n}(i,j)), \end{aligned} \tag{11}$$

where N denotes the number of training samples, C is the number of classes present, W is the width of the segmentation maps, and H is the height. Moreover, $\hat{m}_t$ is the ground truth segmentation mask, m_p is the predicted mask, and $P^c_t(i,j)$ is a binary signal that is equal to one if the pixel in position (i,j) of the ground truth segmentation map n has class c. Essentially, it emulates the true class probability distribution of a pixel from the mask $\hat{m}_t$. Furthermore, $P^c_p(i,j)$ represents the predicted probability of a pixel (i,j) being of class c.

5.4 Implementation Details

5.4.1 ***Tree-based Method***. A five-fold cross-validation technique was employed with a Random Forest classifier, training 175 trees per fold. The maximum depth of each tree was fixed at 60. Only 20 volumes were considered due to computation constraints. The function selected to measure the quality of a split was Gini impurity. Meanwhile, the optimization target was a multiclass Dice Score. Prior to training, a feature selection process was conducted on a five-fold cross-validation pipeline, where a Random Forest Classifier was trained with 100 trees per-fold.

5.4.2 ***Open BraTS Solution***. A U-Net based model is trained for 60 epochs over the BraTS dataset (see Section 6.1.1). For optimization, the Ranger optimizer [29] is used with a learning rate set to 2×10^{-4} and no rate decay. The batch size selected was 1. Additionally, the optimization target was simply a batch-wise Dice loss without weighting. None of the pipelines from [9] were considered, and the ensemble strategy was discarded. The base model from pipeline A of [9] was trained once.

6 TUMOR SEGMENTATION EXPERIMENTS

6.1 Data

6.1.1 ***BraTS Dataset***. Since FastMRI did not hold segmentation maps for each MRI scan, then to perform the semantic segmentation experiments, the BraTS dataset [18] was used. BraTS compromises a collection of volumetric brain MRIs that have tumoral regions. The training split provided for the BraTS2021 challenge included 1251 brain MRIs, along with the segmentation annotations of the tumorous regions.

6.1.2 ***Data Preprocessing***. To proceed with the evaluation of the tumor segmentation, a hold-out technique was employed. The training set was shuffled into two splits, training and testing. The new training split contains 80% of the data from the original split, while the testing split has the remaining 20%. Every volume was cropped to a central 128×128×128 region, thus improving data balancing and reducing computations required, as redundant background voxels (label 0) on the borders of each volume are cropped. Furthermore, two new datasets were consummated to validate the reconstruction quality of Super-Resolution GANs. From the original BraTS, a tricubic interpolation was used to downsample the whole training and testing splits generated by the hold-out technique. The downscaling factor adopted was 2. Thus a new dataset, entitled Low-Resolution BraTS (LRBraTS), was formulated. Simultaneously holding BraTS and LRBraTS means LR-HR pairs are present, thus the conditions to perform Super-Resolution are met. The Real-ESRGAN model, discussed in Section 3.6, was used to super-resolve the LRBraTS by an upscale factor of ×2, resulting in a new dataset, SRBraTS. The implementation details of Real-ESRGAN are described in Section 3.8.

6.2 Evaluation Metrics

6.2.1 ***Dice Similarity Coefficient***. A successful prediction intends to maximize the overlap between true and predicted labels. Dice similarity coefficient (DSC), also known as F1-score or Sørensen-Dice index, is a metric that aims to mathematically quantify how good this overlapping is. DSC is defined as:

$$\text{Dice}(A,B) = \frac{2\cdot\|A\cap B\|}{\|A\| + \|B\|}, \tag{12}$$

where A and B denote two binary segmentation masks for a given class, $\|A\|$ represents the norm of A, and $\|A\cap B\|$ corresponds to the overlap given by the intersection between both masks.

6.3 Jaccard Index

Similarly to DSC, it can be used to measure the similarity between two segmentation maps. It is also known as the Intersection over Union (IoU), as is defined as follows:

$$\text{IoU}(A,B) = \frac{\|A\cap B\|}{\|A\cup B\|}, \tag{13}$$

where A and B denote two binary segmentation masks for a given class.

6.4 Hausdorff Distance (95%)

The Hausdorff Distance (HD) is the maximum perpendicular distance between the closest points from the contours of two regions. Essentially, it is complementary to the DSC, as it measures the maximum distance between the margin of the two regions. It is computed as follows:

$$\mathrm{H}(A,B) = \max\left(\max_{a\in A}\min_{b\in B} d(a,b), \max_{b\in B}\min_{a\in A} d(b,a)\right), \qquad (14)$$

where $d(a,b)$ denotes the distance between two pixels, a and b, in the border of two region included in the segmentation masks, A and B, respectively.

6.5 Quantitative Results

Table 3: Tumor segmentation results comparison between the super-resolved and ground truth brain MRIs. Red color indicates the worst performance overall and Green color the best.

Method	Input	Scale	Optimization Target	DSC	IoU	HD95
Tree-based	SRBraTS	×2	Multiclass DSC	0.26	0.26	28.76
Tree-based	BraTS	-	Multiclass DSC	0.29	0.28	57.81
Open BraTS	SRBraTS	×2	Multiclass DSC	0.61	0.52	21.4
Open BraTS	BraTS	-	Multiclass DSC	0.82	0.75	8.35

All metrics suggest that tumor segmentation with the ground truth MRIs outperforms tumor segmentation performed over the super-resolved MRIs. Despite Super-Resolved MRIs exhibiting photo-realistic details, they did not manifest the best results for the segmentation of tumors. However, the algorithms were intensively trained with the ground truth images, which evidently may slightly benefit the tumor segmentation of the ground truth images. Although super-resolved images are reconstructions of the ground truths, the distributions between them can have minor dissimilarities. For instance, Super-Resolution algorithms can marginally change pixel value intensities in some regions, which can subsequently lead to an inferior segmentation. Furthermore, the Super-Resolution algorithm used (Real-ESRGAN) was not trained over the BraTS dataset, thus the Super-Resolution has the potential to be improved further. Nonetheless, super-resolving medical images is a complex task, and despite having all these constraints, the tumor segmentation still manifested satisfactory results over the super-resolved dataset.

6.6 Qualitative Results

Qualitative results advocate an adequate Super-Resolution and a non-optimal but decent tumor segmentation of the super-resolved MRIs. Looking at Figure 13 it is possible to see a few dissimilarities between the predicted segmentations of the super-resolved and the ground truth MRIs. The difference is not large despite IoU suggesting that the tumor segmentation over SRBraTS is inferior by some margin.

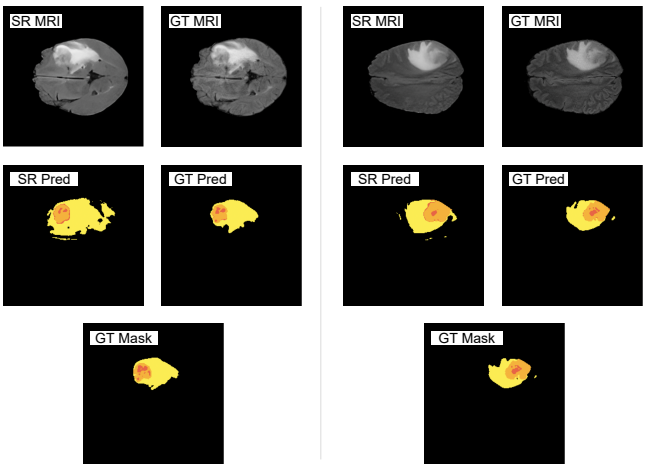

Figure 13: Tumor Segmentation results with BraTS and SR-BraTS. The first row exhibits the super-resolved MRIs and the corresponding ground truths. Below each MRI is the predicted segmentation map that was obtained from it.

6.7 Discussion

Matching the tumor segmentation performance over SRBraTS with the performance over BraTS can be suggestive that the super-resolution was reliable. Looking at Tables 3, it is possible to acknowledge that a few dissimilarities were present. A reason can be the usage of content loss as an optimization target. Accordingly, this leads to overly smooth results, as discussed in section 4.5, or marginal changes in the pixel value intensities of some areas, consequently confusing the segmentation model into interpreting some regions as tumors, which can cascade to a larger region and jeopardize the prediction. This explains the impact on the prediction of peritumoral edematous (ED — label 2), which was the region that suffered the highest impact. Since it contains the tumor border, the SR algorithm can get confused due to the transitions of tissue inherent to that region. Additionally, the Super-Resolution model employed was not the best model from the experiments of section 5. This suggests an additional extent for improvements. For instance, if the Super-Resolution algorithm was trained intensively with MRIs holding tumoral regions, then the SR algorithm could have learned patterns to better mimic the data distribution of the high-resolution images and reconstruct tumoral regions accordingly instead of interpreting them as downsampling artifacts. Furthermore, if the models for tumor segmentation were trained with the super-resolved images and their corresponding annotations, then the segmentation performance with SRBraTS would possibly be substantially higher.

7 CONCLUSIONS

A knee MRI scan usually takes 30 to 60 minutes but can take as long as 2 hours. Acquiring less amount of k-space data will reduce the acquisition time. However this results in MRIs with relatively low spatial resolution. Furthermore, medical imaging techniques often have low resolution, inherent noise, and lack of structural information. Therefore, making a correct diagnosis judgment in the

medical field becomes a significant challenge. This work has proven that high-frequency details can be recovered from Low-Resolution signals, and GAN-based super-resolution has the potential to quarter the acquisition time (not considering the negligible period of time to reconstruct the MRI, which does not affect the patient in any manner). Therefore, GAN-based techniques are promising CS-MRI reconstruction methods, enabling resolution improvements, zooming into images, and data acquisition acceleration. Additionally, denoising solutions led to performance boosts on the super-resolution task, with manifested reduction of the checkerboard pattern inherent to GAN synthesis.

Although the task-based evaluation showcases space for improvements in the performance of GANs, they still provide good perceptual quality. Tumor Segmentation of Super-Resolved images exhibited an inferior performance relative to tumor segmentation with ground truth images. However, several constraints coexisted that impacted these results. The tumor segmentation still manifested satisfactory results over the SRBraTS dataset. Furthermore, fully convolutional neural networks exhibited solid results in segmenting tumors, thus solidifying the proficiency of Deep Learning in the medical image context. Merging both Super-Resolution and Tumor Segmentation can provide an automatic pipeline for diagnoses that healthcare can substantially benefit from. Ultimately, two new datasets were formulated to use Tumor Segmentation to validate the Super-Resolution quality in medical image reconstruction.

REFERENCES

[1] Eirikur Agustsson and Radu Timofte. 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 126–135.

[2] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. 2018. The 2018 PIRM challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 0–0.

[3] John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), 679–698.

[4] W Knox Carey, Daniel B Chuang, and Sheila S Hemami. 1999. Regularity-preserving image interpolation. *IEEE transactions on image processing* 8, 9 (1999), 1293–1297.

[5] Yuhua Chen, Anthony G Christodoulou, Zhengwei Zhou, Feng Shi, Yibin Xie, and Debiao Li. 2020. MRI super-resolution with GAN and 3D multi-level DenseNet: smaller, faster, and better. *arXiv preprint arXiv:2003.01217* (2020).

[6] Eva Funk, Per Thunberg, and Agneta Anderzen-Carlsson. 2014. Patients' experiences in magnetic resonance imaging (MRI) and their experiences of breath holding techniques. *Journal of advanced nursing* 70, 8 (2014), 1880–1890.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[9] Theophraste Henry, Alexandre Carré, Marvin Lerousseau, Théo Estienne, Charlotte Robert, Nikos Paragios, and Eric Deutsch. 2020. Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution. In *International MICCAI Brainlesion Workshop*. Springer, 327–339.

[10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*. Springer, 694–711.

[11] Jagat Narain Kapur, Prasanna K Sahoo, and Andrew KC Wong. 1985. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer vision, graphics, and image processing* 29, 3 (1985), 273–285.

[12] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.

[13] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[14] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690.

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.

[16] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. 2017. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding* 158 (2017), 1–16.

[17] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. 2016. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing* 26, 2 (2016), 1004–1016.

[18] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. 2014. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging* 34, 10 (2014), 1993–2024.

[19] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters* 20, 3 (2012), 209–212.

[20] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. 2018. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957* (2018).

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

[22] Edgar Schonfeld, Bernt Schiele, and Anna Khoreva. 2020. A u-net based discriminator for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8207–8216.

[23] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[24] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. 2017. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 114–125.

[25] Rao Muhammad Umer and Christian Micheloni. 2020. Deep cyclic generative adversarial residual convolutional networks for real image super-resolution. In *European Conference on Computer Vision*. Springer, 484–498.

[26] Luc Vincent and Pierre Soille. 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 13, 06 (1991), 583–598.

[27] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. 2021. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1905–1914.

[28] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*. 0–0.

[29] Less Wright. 2019. Ranger - a synergistic optimizer. https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer.

[30] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839* (2018).

[31] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. 2021. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4791–4800.

[32] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. 2021. RankSRGAN: Super Resolution Generative Adversarial Networks with Learning to Rank. *arXiv preprint arXiv:2107.09427* (2021).

[33] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*. 2223–2232.