BACHELOR'S THESIS / BIOMEDICAL ENGINEERING 2022

SUPER-RESOLUTION OF MAGNETIC RESONANCE FETAL IMAGES USING SEQUENTIAL GENERATIVE ADVERSARIAL NETWORKS

Manuel Obelleiro Liz

upf. Universitat Pompeu Fabra Barcelona

SUPER-RESOLUTION OF MAGNETIC RESONANCE FETAL IMAGES USING SEQUENTIAL GENERATIVE ADVERSARIAL NETWORKS

Manuel Obelleiro Liz

Bachelor's thesis UPF 2021/2022

Thesis supervisor(s): Benjamin Lalande Chatain

Department of Information and Communication Technologies

Universitat Pompeu Fabra Barcelona

Acknowledgments

I would like to thank my research supervisor Benjamin Lalande for the support and feedback provided during the whole year. Without his assistance and involvement throughout the project, this thesis would not have been possible. I would also like to thank my family and friends for their patience and support during the last 4 years of the degree.

Abstract

Magnetic Resonance imaging is a really important imaging technique distinguished by the lack of ionizing radiation, great soft-tissue contrast and remarkable resolution. The main drawback of performing a Magnetic Resonance Imaging (MRI) scan is time consumption. As a result, many MRI scans are performed in a shorter time sacrificing overall image quality. Thus, there is a need for algorithms able to convert low-resolution images, obtained from faster scans, to super-resolution images comparable to full-time scans.

Generative Adversarial Networks (GANs) have the ability to generate new content, mainly images and data distributions. This kind of network can be divided into 2 units that act in competition, a generator and a discriminator. The generator receives an input and must return, an output which follows the target distribution. The discriminator acts as the judge of the process, determining which image is real and which was created by the generator.

In this bachelor thesis, a two-stage training process is developed using the ESRGAN framework. First, a Peak Signal to Noise Ratio (PSNR) oriented model is developed without adversarial training, performing similar to standard deep networks. In the second stage, the discriminator is introduced and trained together with the PSNR-oriented generator. The second stage allows the model to learn realistic texture representations, measured with the perceptual index (P.I.) metric.

The obtained results demonstrate the potential application of these networks to medical imaging. The reconstructed MRI images were objectively more similar to the original High Resolution (HR) ground truth when compared to other conventional image reconstruction methods. The PSNR trained model yielded better PSNR and structural similarity (SSIM) performance but with a lower perceptual index. The GAN model showed greater P.I. by being able to replicate image textures and creating overall more visually appealing images.

Keywords

MRI, fetal images, image processing, computer-aided diagnosis (CAD), single image super-resolution, artificial intelligence (AI), deep learning (DL), deep convolutional neural networks, generative adversarial networks (GANs).

Prologue

The motivation behind the proposed bachelor's thesis started from my interest in developing a project related to machine learning and its applications in biomedical image processing. During the last years of the degree in Biomedical Engineering, we were introduced to imaging systems, computational modeling as well as signal and image analysis. By developing small projects, I found a profound interest in topics such as image processing, computer-aided diagnosis and deep learning.

These experiences stimulated my pursuit of a bachelor thesis on a Generative Adversarial Network (GAN) for image super-resolution of MRI fetal images. GANs are relatively new and they have meant a revolution in machine learning, this was one of the reasons I chose this topic, as it is a new field with great potential. The proposed project consists of building, training and validating a neural network able to transform low-resolution images obtained through magnetic resonance (MRI) of a fetus and return a higher resolution version of the former. The main problems of MRI are the time of acquisition of images and the necessity of remaining still during the process, thus, although the lack of radiation may be perfect for fetal imaging, images obtained can often be blurry and lack resolution. This limitation could be solved by the generation of a high-resolution image from its low-resolution counterpart, reducing time, costs and improving image quality.

For this thesis, I am supervised by one of the assistant professors who introduced us to image and signal processing as well as machine learning, Benjamin Lalande Chatain, a Ph.D. student with a predoctoral contract at the Departament de Tecnologies de la Informació i les Comunicacions in Universitat Pompeu Fabra (UPF).

Index

List of figures

List of tables

1. INTRODUCTION

1.1 Low-resolution MRI images

Magnetic Resonance Imaging (MRI) is a widely used non-invasive imaging technique based on the nuclear resonance phenomena for the acquisition of information regarding the structure or composition of the imaged sample. The imaged subject is under the effects of a stationary magnetic field $\vec{B}$. Hydrogen protons, very abundant especially in water-rich tissues, orient themselves following the externally applied magnetic field, magnetizing the sample. With the addition of a periodic radiofrequency wave (RF), the atoms perform a 90º degree flip. When the RF wave stops the atoms perform a physical phenomenon called relaxation, which is recorded by the receiving RF coils of the MRI scanner. The data acquired from an MRI scan is then used to fill the frequency domain, also known as the k-space or Fourier space [1]. We can define a convention regarding the spatial dimensions of an MRI scan as seen in Figure 1; the in-plane dimension is referred to as the phase and frequency encoding of the 2D plane, the positional information is related to each point in the k space through phase and frequency. The third dimension determines the slice of the 3D space and is linked to a gradient in the fixed magnetic field allowing for a slice selective RF pulse. Once the frequency domain is acquired, a reconstruction algorithm is applied to obtain the representation of the imaged sample in the pixel domain. The time between phase encodes is the most time-consuming part of an MRI scan [2]; scan resolution increases with an improvement of spatial resolution along the in-plane dimension, consequently increasing the acquisition time of the MRI scan.

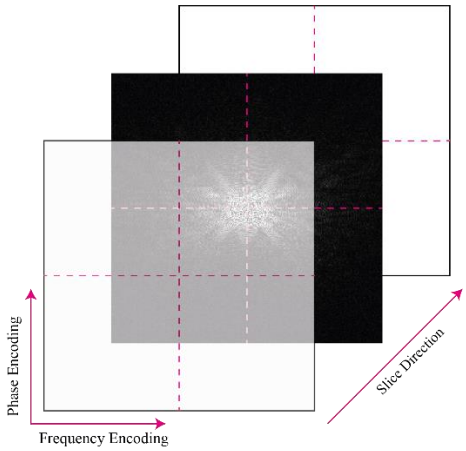

Figure 1. Dimension convention for the Fourier space in an MRI scan

Nowadays, MRI is used in brain disease, spinal disorders or angiography among other applications. It is best suited for the study of soft tissues (great water content) thanks to its outstanding contrast [3] and it allows functional imaging (FMRI: imaging of metabolic function). Scans can be performed in any given direction and it is non-invasive due to the lack of ionizing radiation. On the downside, MRI requires a longer acquisition time compared to other techniques such as Computed Tomography (CT) and some patients may experience claustrophobia whilst inside the scanner. To produce high resolution (HR) and quality images, patients must stay still throughout the scan; this problem also applies to respiratory or internal organ motion, typically with cardiac imaging, fetal MRI, functional MRI or uncooperative patients, the scanning times must be remarkably fast to avoid motion artifacts [2]. These are the main reasons behind the use of high-speed MRI

scans. The high demand for HR images both in research and in clinical applications caused an increase in popularity of the use of image processing techniques to improve the image quality of scans performed with better patient comfort, lower costs and fewer artifacts caused by motion [4]. High-quality HR images are important in many different clinical and research applications, they are necessary for radiologists to perform diagnoses, for surgeons in image-guided interventions and for researchers during the training of machine learning algorithms.

The resolution of an MRI scan can be improved in several ways, we can distinguish between hardware improvements and post-acquisition image processing techniques. Increasing the number of coil receiver channels or using a magnetic field of higher strength causes an increase in image resolution [5]. MRI scanners used for medical purposes have a magnetic field strength between 1.5 and 3T, obtaining image resolutions of $1.5 \times 1.5 \times 4$ mm^3; in research, as higher resolution images are required, an ultra-high magnetic field scanner with a magnetic flux density of 11.7 T can produce images with resolutions of $80 \times 80 \times 200$ μm^3 [5].

In this study, we will be interested in regular, low magnetic flux density clinical MRI scanners and the use of post-acquisition image processing to improve image quality. In most magnetic resonance scanners, a basic interpolation option is available to increase image size; interpolation introduces problems such as blurring or loss of contrast as there is an approximation to pixel intensity based on neighboring pixels and there is no external information or learning process involved.

Image Super-Resolution (SR) techniques can be divided into Single Image Super-Resolution (SISR) and Multi-Frame Image Super-Resolution (MISR) [6]. MISR is a process based on the correlation of multiple LR images from the same sample, it requires more data in the form of multiple images of the same view in a slightly different position and even different time frames. SISR creates an HR image from a single LR counterpart; it is less demanding in terms of the amount of data and is one of the main methods used for improving image quality in post-processing both in medical and natural images. The most recent SISR methods are related to deep learning-based algorithms, these methods can reach state-of-the-art results on different super-resolution benchmarks [7]. They can have different architectures, learning methods and strategies, ranging from classical Convolutional Neural Networks (CNN) to revolutionary and ground-breaking Generative Adversarial Networks (GAN), which will be the main focus of the present thesis.

1.2 Deep Learning

With the recent revolution in artificial intelligence, machine learning methods have improved in terms of performance, obtaining outstanding results in image analysis challenges [8], [9]. These advancements were also applied to the clinical framework; deep learning techniques were introduced to medical image processing and analysis in applications such as computerized prognosis, organ segmentation, lesion detection and classification [10], [11], [12], [13].

Deep learning is a subfield of machine learning made up of algorithms called neural networks that artificially model the functioning of the human brain with a computer. Neural networks are made up of layers containing neurons, also called nodes or perceptron. A neural network is composed of an input and output layer with a certain

number of intermediate "hidden" layers. Each neuron (see Figure 2) receives a linear combination of weighted inputs from nodes in the previous layer. The weighted sum of inputs that the neuron receives is passed through a non-linear activation function which determines the output of that node.

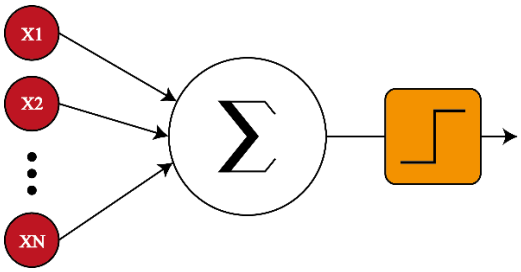

Figure 2. Perceptron model

Convolutional neural networks (CNNs) are one of the most popular deep learning algorithms. These models are able to receive an image as input and detect important features without human supervision [14]. CNNs have been applied to many different applications such as computer vision, speech processing and facial recognition. Their architecture is inspired by the organization of neurons in the visual cortex of the human brain [15], [16]. They are often divided into 2 sections, feature learning and classification. During the first stage, the CNN is made up of convolutional and pooling layers where an image is passed through a series of weighted kernels. The output of each layer is obtained by calculating the dot product between the input data and the weights of the kernel, which are adjusted at each training iteration, allowing the algorithm to learn how to extract significant features [14]. The objective of the convolutional operation is to extract features such as edges or textures and ultimately have the image reduced into a representation that is easier to process. During the second stage, a network of neurons receives the result of the convolutional layers and is often trained to perform image classification tasks.

Each convolutional operation is mainly defined by kernel size, stride and number of feature maps. It follows the expression shown in Equation 1, where $*$ is the 2D cross-correlation operator, N is the batch size, C denotes the number of channels, H is the height of the image in pixels and W is the width.

$$out\left(N_i, C_{out_j}\right) = bias\left(C_{out_j}\right) + \sum_{k=0}^{C_{in}-1} weight(C_{out_j}, k) * input(N_i, k)$$

Equation 1. 2D convolutional operation

In each convolutional layer, a kernel 'K' (of size 3x3 in Figure 3) is used to perform the convolution with the input data I, resulting in the output matrix I∗K. The stride controls how the filter moves along the input image, a stride of 2 would mean a shift of 2 pixels every time the filter is moved along the image. The number of feature maps defines the number of convolutional kernels that are applied at a given layer.

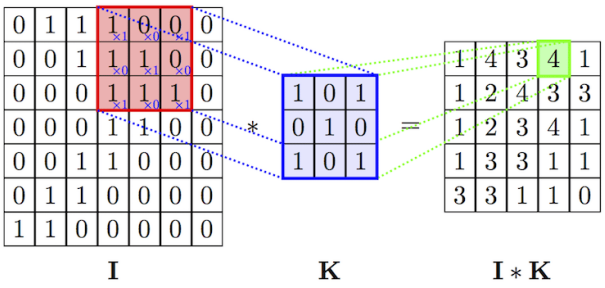

Figure 3. 2D convolutional operation [17]

In CNNs there is no need to learn additional weights for each neuron, this makes them a great solution for image processing tasks [14]. Weight sharing in CNNs causes a reduction in the number of trainable parameters and helps the network to generalize, avoid overfitting, use less memory and optimize the overall training time.

1.3 Generative Adversarial Networks

GANs were first introduced in 2014 by Ian Goodfellow [18] as a novel generative model able to learn feature representations from complex data distributions. In the proposed framework, 2 models are trained simultaneously; a generator G which is able to create artificial data distributions, and a discriminator D which determines the probability that a given sample comes from the real data distribution and not from G, as seen on Figure 4. Competition in the proposed two-player minimax game drives both sides to improve until the generated samples are indistinguishable from the real data distribution. When these networks are applied to image processing in tasks such as domain transfer or artificial super-resolution, the generator and discriminator models are constituted as CNNs.

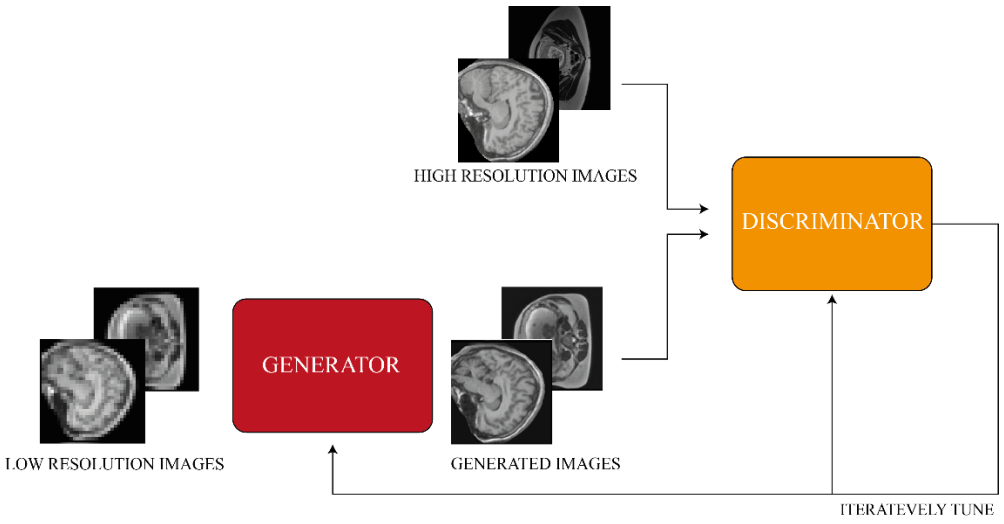

Figure 4. Overview of a generative adversarial network framework

GANs have many advantages but also come with important drawbacks. The main disadvantages are that there is no explicit representation of the generators distribution $p_g(x)$, and that the learning process must be synchronized between G and D. The generator must not be trained too much without the update of the discriminator as it could cause the

model to collapse, a scenario where G would not produce a wide variety of outputs. If the generator produces an output which completely fools the discriminator, it may tend to only generate this same output, with the discriminator stuck and unable to learn how to correctly classify the image. In order to force G to create new outputs and prevent it from collapsing, there are different loss functions that can solve the problem. This problem goes both ways as a discriminator that is too good may cause vanishing gradients as it stops giving useful information for the generator to make progress. The Wassertein loss [19] prevents from vanishing gradients and mode collapse by rejecting outputs that the generator stabilizes on. GANs may also fail to converge, there are different regularizations that can be attempted to improve convergence [20], [21].

Traditional deep learning models are mostly PSNR-oriented networks. They tend to produce unrealistic textures and a lack of attention to fine details, leaving perceptually blurred impressions [22]. This is the main reason behind the introduction of GAN frameworks with perceptual quality oriented loss functions, which allow the models to generate realistic textures and details. A perceptual index metric, described later on, was developed [23] in order to measure the perceptual quality of the reconstructed images and avoid using pixel-wise scores such as PSNR.

1.4 State of the art

There are many GAN architectures, each with its own name and purpose [24] and they are used in a wide range of applications, evolving and adapting to different problems but keeping the adversarial principles introduced by Goodfellow [18]. It is a field that is growing rapidly due to the outstanding results obtained with deep learning-based methods, specially SISR based on GANs which can achieve the most advanced visual performance results [25]. Another factor that could influence the increase of the number of deep learning models is the progressive improvement of computational power, following Moore's Law, a prediction made by the American engineer and Intel founder Gordon Moore in 1965, which stated that the number of transistors per silicon chip would double every year, and with them, the computational capabilities of computers [26]. Also, these algorithms would not perform without Big Data [27], great volumes of data that are growing thanks to the digitalization of the world. Thanks to the mentioned improvements deep learning algorithms can train properly and achieve unprecedented performance in their fields of application.

There is a wide variety of approaches, the present thesis will be focused on the ones that are more relevant to the proposed MRI super-resolution problem. Christian Ledig et al. [28] introduced the SRGAN (Super-Resolution GAN) network, the first framework capable of inferring photo-realistic images for 4× upscaling factors. They introduce a perceptual loss term to the training process which features a perceptual similarity measurement instead of similarity in pixel space. They were able to recover photo-realistic textures from heavily downsampled non-medical images on public benchmarks. As a continuation of the mentioned work, Xintao Wang et al. [22] created the ESRGAN (Enhanced SRGAN) network which introduced changes in the perceptual loss, modified the architecture of the generator and proposed a relativistic discriminator to further enhance the visual quality of the upsampled images. For the present thesis, the mentioned ESRGAN framework is applied for biomedical image super-resolution, seeking to generate realistic textures and details in fetal MRI images.

Similarly, D. Mahapatra et al. [29] proposed a progressive generative adversarial network where a low-resolution input image is enhanced progressively (as seen in Figure 5) using the generator and discriminator architectures introduced in the SRGAN [28] paper. They present a loss term named Triplet Loss which ensures that the GAN always improves towards the desired high-resolution image.

$$\sum_{i}^{N} [\|f(I_2^{SR}) - f(I^{HR})\|_2^2 - \|f(I_2^{SR}) - f(I_1^{SR})\|_2^2]$$

Equation 2. Triplet loss proposed by Mahapatra et al. [29]

This loss term makes sure that a generated image I_2^{SR} is as far away as possible from the image generated in the previous step I_1^{SR}, and as close as possible to the HR reference image that it tries to replicate I^{HR}. This simple term forces the image reconstruction process to keep improving rather than collapsing and producing similar results.

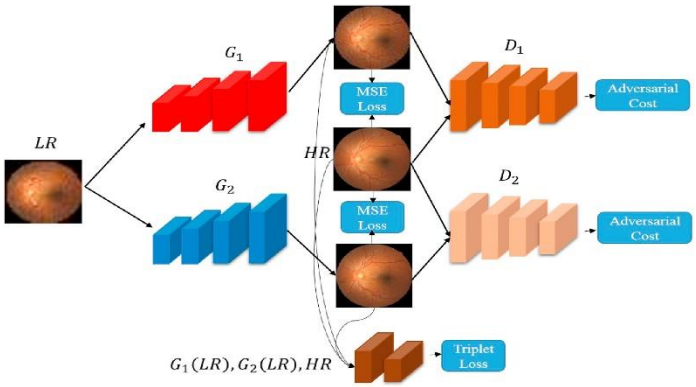

Figure 5. Depiction of progressive GAN architecture [18]

Lastly, there are approaches to image super-resolution which utilize cycle GANS. Cyclic adversarial networks are mostly used in domain transfer problems [30]. Classical cycle GAN approaches assume that input and output images are of the same size, thus they need to be adapted for super-resolution purposes [31]. Yuan Y. et al. proposed CinCGAN [24], a framework where the cyclic architecture is applied for 2 domain transferring processes. First, they perform image de-noising, going from noisy LR images to clean LR. Once the LR image is free of noise it is up-sampled with traditional methods, yielding a noise-free up-sampled version of the original LR input. This image is then fed through a second cycle GAN in charge of performing domain transfer from low to high resolution. They obtained worse results in terms of pixel-wise metrics but the images had improved perceptual quality.

1.5 Hypothesis

Generative adversarial network architectures have demonstrated outstanding results in image super-resolution tasks on natural [28] and medical [29] images. These networks excel at generating realistic data distributions with complex textures but require a great amount of high and low-resolution image pairs. The training of said network will be performed using a brain MRI dataset consisting of over 30 full-brain scans with over 5000

high-resolution images and a smaller dataset consisting of 12 abdominal MRI scans of pregnant women containing over 1000 images. By applying the proposed models, the trained network on the brain and fetal MRI images should be able to improve image resolution while keeping realistic-looking textures and details, significantly improving the results obtained with standard image interpolation or PSNR-oriented models.

1.6 Objectives

The objective of the present thesis is to develop a GAN architecture capable of generating super-resolution fetal MRI images after being trained on the brain and fetal MRI datasets, attempting to achieve high visual performance results measured by both pixel-wise (PSNR, SSIM) and perceptual quality metrics.

2. METHODS

The present project aims to perform fetal and brain MRI image estimations from low-resolution input images. In order to achieve this goal, a generator (G) and discriminator (D) network are trained using high-low resolution image pairs consisting of MRI slices from brain and abdominal scans. The ultimate objective of said networks is to yield high resolution and high perceptual quality image estimations that are indistinguishable from real HR scans.

To learn how to estimate data distributions over the input data 'x', the generator creates a mapping to the data space as G (z; θ_g) where θ_g are the multilayered perceptron G parameters [18], [25]. D (x; θ_d) outputs a scalar representing the probability that x belongs to the real data distribution rather than to the artificially generated one. The training is performed on G and D trying to maximize the probability of correctly classifying x (training of D) and minimizing log(1 − D(G(z))), training G by learning to fool D. In the end it all can be summarized by a minimax game with the value function V (G, D):

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

Equation 3. Minimax algorithm

2.1 Network architecture

The proposed generator network receives LR images which are obtained by down-sampling the original HR data and forwarding it through the first convolutional layer. This operation is performed with a convolutional filter of kernel size 9x9, stride=1 and 64 feature maps followed by a leaky ReLu activation function and no batch normalization (BN) layers. The obtained output of the first stage is then fed through B (B=23) identical Residual-in-Residual dense blocks pictured as 'RRDB Blocks' in Figure 6. Each of the RRDB blocks combines multi-level residual networks and dense connections, as depicted in the zoomed-in sections of the figure. As seen in different studies [32], [33], [34] more layers and connections boost performance significantly, the RRDB block is composed of a deeper and more complex architecture when compared to the basic block proposed in the SRGAN paper. The RRDB blocks also contain a residual scaling factor β which scales

down the residuals by multiplying them with a constant (β=0.2) improving network training stability [22].

Removing BN layers from the generator has proven to reduce computational complexity and boost performance. BN can introduce artifacts and limit the generalization when there are significant statistical differences between training and testing datasets [22].

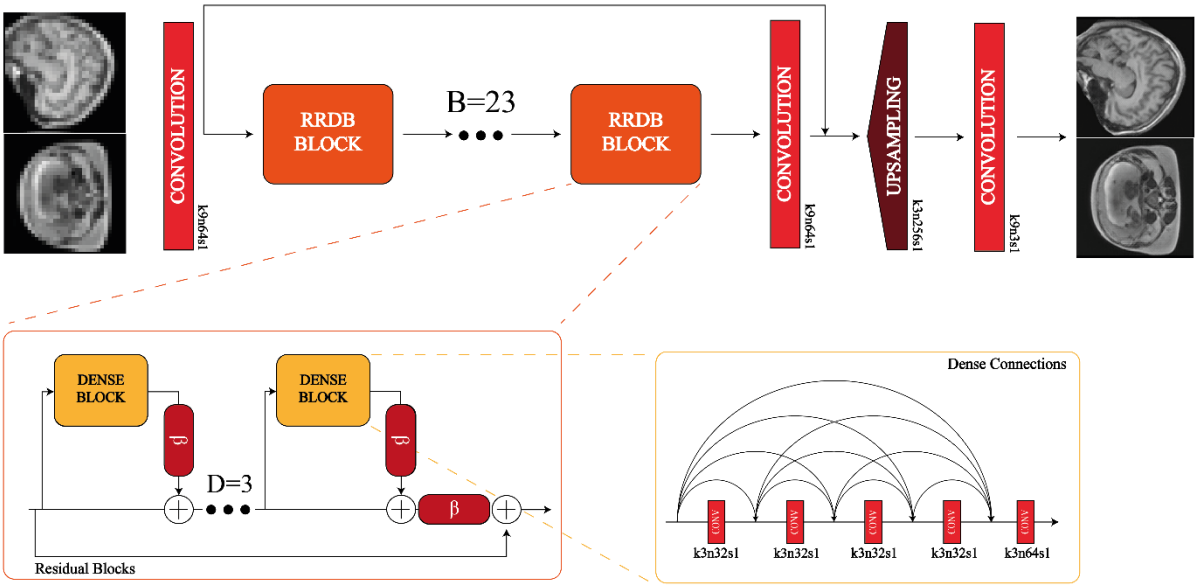

Figure 6. ESRGAN Generator network

The discriminator (D), as previously mentioned, classifies between real HR images and generated SR data. The basic model follows the architecture proposed by Ledig et al. [28] but it is enhanced through the introduction of relativistic properties inherited from the Relativistic GAN [35] framework. The discriminator is made up of 8 convolutional layers with 3x3 kernel filters of increasing feature maps from 64 to 512 as seen on the VGG network [36] as seen in Figure 7. The strided convolutions 's2' reduce the data structure size as the number of features is doubled. In the discriminator model there are also BN and leaky ReLu (α=0.2) layers after every convolution.

In the present thesis, the relativistic average discriminator (D_{Ra}) is implemented, it seeks to solve the missing property of the standard GAN framework, that is, allowing the generator to influence the discriminator just as D is relevant during the training of G.

$$D(x_r) = \sigma(C(x_r)) \rightarrow 1 \quad ; D(x_f) = \sigma\left(C(x_f)\right) \rightarrow 0$$

Equation 4. Traditional discriminator model

$$D_{Ra}(x_r, x_f) = \sigma\left(C(x_r) - E_{x_f}[C(x_f)]\right) \rightarrow 1 \quad ; D_{Ra}(x_f, x_r) = \sigma(C(x_f) - E_{x_r}[C(x_r)]) \rightarrow 0$$

Equation 5. Relativistic average discriminator

In the relativistic model, D_{Ra} determines the probability that a real image x_r is relatively more realistic than a fake one x_f, instead of simply estimating the probability that a given image x is real or fake (eq. 4.1 and 4.2). As seen in equations 5.1, and 5.2, the relativistic discriminator performs the sigmoid function σ over the discriminator output and the average ('E[C(x)]') of a batch made up of all fake or real data. The results published [35] suggest that the relativistic properties of the proposed discriminator improve the image quality and stability of the system with no associated computational overhead.

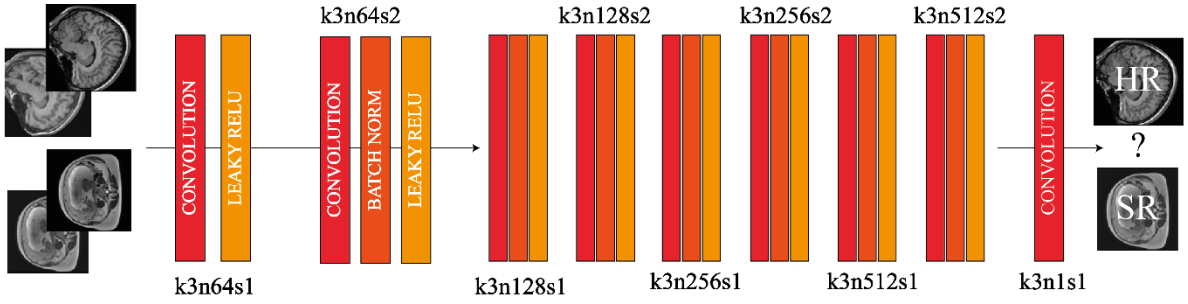

Figure 7. ESRGAN Discriminator network

2.2 Loss functions

In order to train the generator and discriminator models, a loss function needs to be defined to guide the training of the algorithm. A loss function evaluates the performance of a model at a given task. During the training process, with the help of an optimization function, the algorithm progressively improves and obtains results that yield a better score.

$$L_G = L_{percep} + \lambda \cdot L_{Adv} + \eta \cdot L_{content}$$

Equation 6. Generator Loss

As seen in Equation 6 the loss is composed of three different components. The adversarial loss term L_{Adv} encourages the network to favor solutions that reside on the manifold of natural images [28] by trying to fool the discriminator. This term is based on the discriminator's output over all training samples, defined as:

$$L_{Adv} = \sum_{n=1}^{N} -\log D_{\theta_D}^{Ra}(G_{\theta_G}(I^{LR}), I^{HR})$$

Equation 7. Adversarial loss term

Here, $D_{\theta_D}^{Ra}(G_{\theta_G}(I^{LR}), I^{HR})$ is the probability that generated data $G_{\theta_G}(I^{LR})$ is relatively less realistic than real data I^{HR}. For improved gradient behavior [28] the minimized term is $-\log D_{\theta_D}^{Ra}(G_{\theta_G}(I^{LR}), I^{HR})$ instead of $\log[1 - D_{\theta_D}^{Ra}(G_{\theta_G}(I^{LR}), I^{HR})]$.

The content loss term $L_{content}$ is defined as the L1 distance between all generated and real image pairs of the batch. L1 calculates the mean absolute error (MAE), which measures the distance between the real data and the generated images [37].

$$MAE = \frac{\sum_{n=1}^{N} abs(I^{HR} - G(I^{LR}))}{N}$$

Equation 8. Mean absolute error

The last term of the defined loss function is the perceptual loss L_{percep}. This loss term is responsible for obtaining results of closer perceptual similarity [22]. Training the network using only pixel-wise metrics such as the mentioned $L_{content}$ results in overly smooth textures and images that lack high-frequency details and are perceptually unsatisfying [28]. The perceptual loss is defined as the L1 distance between real and fake image features obtained using the pre-trained feature-extractor 19-layer VGG model [36]. Contrary to convention, the features are obtained before the activation layers of the network overcoming the fact that the features after activation are sparse. Also, using features after activation results in inconsistent reconstructed brightness when compared to ground truth images [22].

$$L_{percep} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}\left(G_{\theta_G}(I^{LR}) \right)_{x,y} \right)^2$$

Equation 9. Perceptual loss term

Here, $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the feature maps within the VGG network. $\phi_{i,j}$ indicates the feature map obtained after the j-th convolution (before activation) and before the i-th max-pooling layer within the VGG network.

2.3 Dataset and preprocessing

The dataset used to train the model belongs to the Hammers adult atlases dataset [38] consisting of 30 adult brain T1-weighted MRI scans. The original MRI files were preprocessed to obtain a dataset of individual horizontal-plane images. Fetal MRI data consisted of 12 abdominal scans of pregnant women. These files were preprocessed to generate the high-resolution ground-truth data once more, with the difference that the horizontal-plane images were pre-checked to avoid using data that lacked relevant information.

In order to train the models, these images were also degraded using the bicubic kernel function to create low-resolution image datasets used as input data to the network. Data is divided between training and validation datasets following an 80%/20% division.

2.4 Training details

The training was performed 3 times for each model, using a scaling factor of x2, x4 and x8 between the LR and HR images. The batch size was set to 16, following the recommendations of Xintao et al. [22] and the HR image size was set to 128x128 pixels due to limitations in computational power (consequently LR images were of size 64x64, 32x32 and 16x16).

The training process was divided into two stages. First, a PSNR-oriented model is trained using a learning rate of 2 x 10^{-4} and using a loss function with only the $L_{content}$ term. Once this first stage of training is complete, the trained PSNR-oriented generator is used as an initialization for the generator during the second stage. During the second part of training, the complete loss function described in Eq. 5 is used with $\lambda = 5 \cdot 10^{-3}$ and $\eta = 1 \cdot 10^{-2}$.

Performing pre-training with a pixel-wise oriented loss function helps the GAN method to obtain more visually pleasing results [22] and avoid local optima for the generator. Also, the discriminator starts to train receiving relatively good resolution images instead of black or noisy images (as can be seen in Figure 8) which allows it to focus on texture and detail discrimination from the beginning.

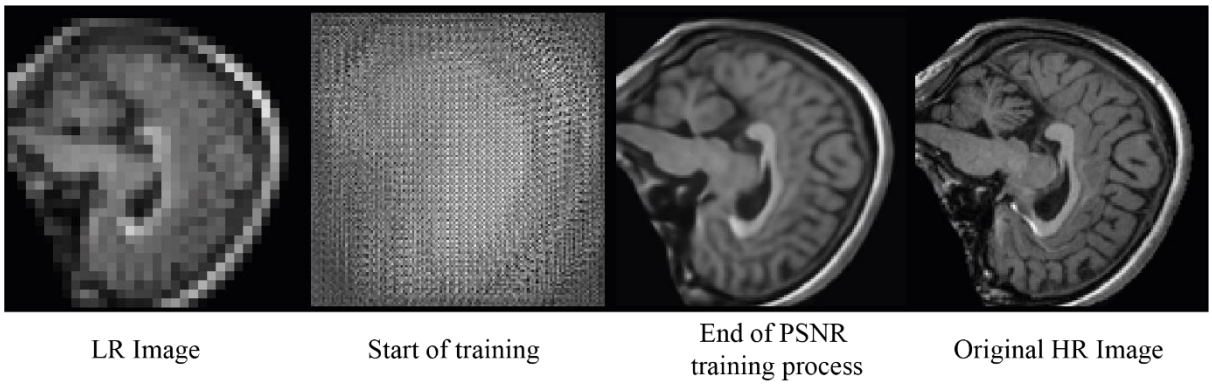

Figure 8. Comparison between the first and last PSNR model results with x4 scaling factor

The optimizer used was the Adam Optimizer [39] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ as proposed by Xintao et al. [22]. The model was developed using the PyTorch Python [40] library hosted by Google colab Pro, a Jupyter notebook environment that runs entirely in the cloud and offers the opportunity to use the powerful GPU machines necessary for the training of the model. The machines available for Pro users are the NVIDIA Tesla T4 and P100 GPUs with 16GB of RAM.

2.5 Performance metrics

Several image similarity methods were implemented in Matlab to perform a quantitative evaluation of the similarity between real and generated images, as well as to measure the perceptual quality of said images. The chosen metrics were peak signal-to-noise ratio (PSNR), structural similarity (SSIM), perceptual index (PI) and root-mean-square deviation (RMSE).

$$PSNR = 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE)$$

Equation 10. Peak Signal-to-Noise Ratio

The peak signal-to-noise ratio measures the quality of the generated image when compared to the original HR ground truth. Mean square error (MSE) is used for the calculation of PSNR as can be seen in Equation 10, here MAX_I stands for the maximum possible intensity value of the image [41].

$$SSIM(x, y) = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)}$$

Equation 11. Structural Similarity Index

The SSIM formula shown in Equation 11 [42] is a simplified version of the complete expression proposed by Wang et al. [43]. Here μ_A, μ_B are the mean values calculated over images A and B. σ_A^2 σ_B^2 represent the variances and σ_{AB} the covariance while C_1 and C_2 are constants. At its core, SSIM is a statistical measure, a product of three local dissimilarity factors (luminance, variance and correlation). The use of this similarity index is widespread, playing a major role in image quality assessment for almost two decades [42].

$$Perceptual\ Index = \frac{1}{2}\big((10 - Ma) + NIQUE\big)$$

Equation 12. Perceptual Index

As previously stated, the training of deep learning models using pixel-wise distance methods yields algorithms with outstanding performance when measured through PSNR or SSIM. High PSNR and SSIM scores are not always related to perceptual quality in images, they often result in blurry and unnatural-looking outputs. These observations ended up causing a shift towards perceptual quality maximization training methods [23]. The implementation for calculating the perceptual index metric is described in the 2018 PIRM challenge on perceptual super-resolution (SR) paper [23]. The proposed perceptual index combines the no-reference image quality measures of Ma et al. [44] and NIQUE [45] as seen in Equation 12.

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(I_n^{HR} - G(I_n^{LR})\right)^2}$$

Equation 13. Root-mean-square deviation

Following the work done in the 2018 PIRM challenge, RMSE is also calculated as a measure of reconstructed image accuracy. Lower RMSE scores are related to higher reconstruction accuracy and they often yield worse results in terms of perceptual quality [23].

2.6 Graphical user interface

Finally, a user-friendly interface was developed using Anvil Works [46], an online web app building platform that runs on python. This platform has a feature which allows to connect a built app with an existing backend code running in google Colab by creating a server and establishing an uplink with the google Colab notebook.

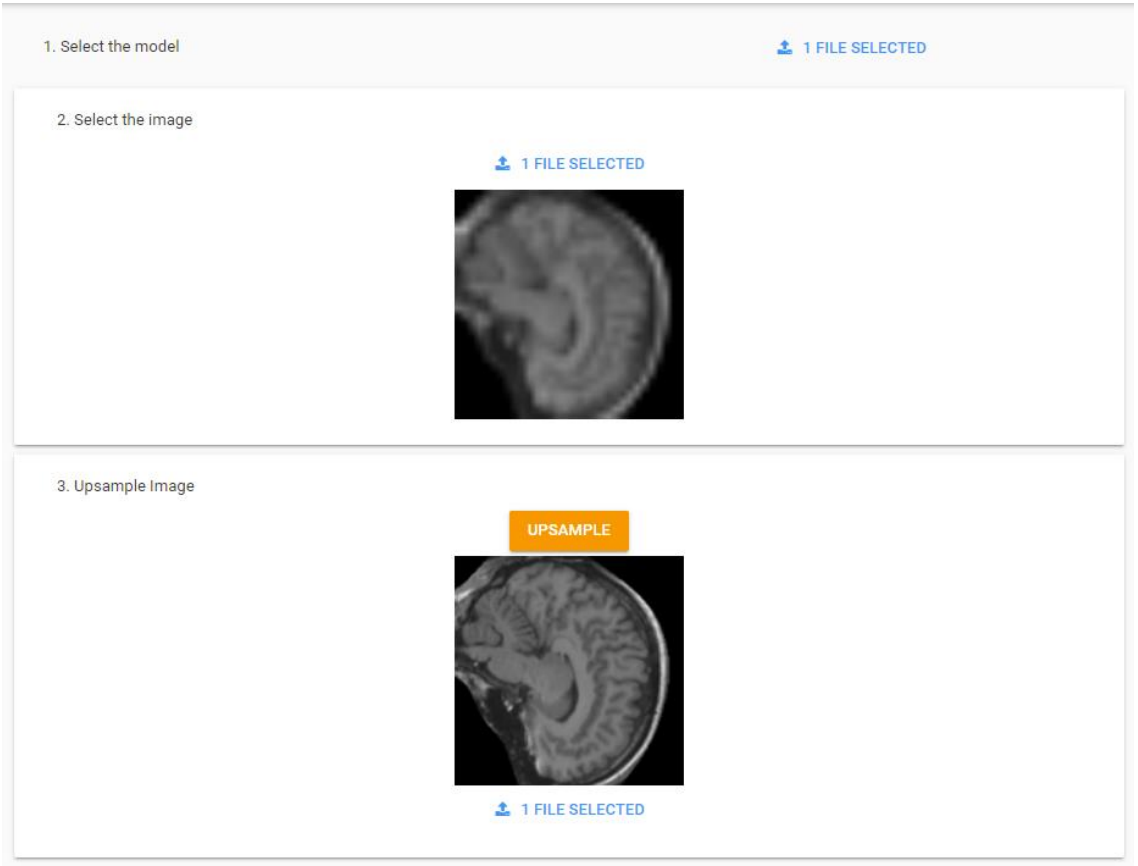

Figure 9. Graphical user interface developed in Anvil Works

The user is able to select a model (the network weights that were saved during the training), choose an image to up sample and display the super-resolution result without needing an in-depth understanding of how the computer code works. An example of the mentioned interface with a x4 scaling of a degraded brain MRI image can be seen in Figure 9.

3. RESULTS

To assess the network's performance and generative capabilities a study consisting of different amplification factors was developed using the validation sets of both brain and abdominal MRI scans. The original images were degraded using a bicubic kernel by 2, 4 and 8-fold. The generator´s task was to create super-resolution images as close as possible to the original HR data.

To have a baseline and understand the quality of the obtained SR images, 4 other methods were used and evaluated. Nearest neighbor is an interpolation technique that determines the intensity of a new interpolated pixel by replicating the intensity of the nearest pixel (neighbor) available. Bilinear interpolation consists of applying a linear interpolation in two directions, it uses the 4 nearest neighbors and performs a weighted average to calculate a new pixel's intensity value. The last technique is bicubic interpolation which uses cubic splines or other polynomial techniques to estimate the new pixel values. The last 2 columns of the comparative figures are the proposed super-resolution model after the first training stage (PSNR-oriented model) and after the adversarial training (ESRGAN). The PSNR-oriented model behaves like a standard deep network, trained with pixel-wise metrics only, and will serve as a benchmark for PSNR and SSIM results as well as to compare the perceptual quality of the generated images.

The comparative tables show the numeric results of the proposed methods. The metrics used, addressed in the Methods section, are calculated in Matlab over the complete validation sets for both fetal and brain MRI images. For the PSNR and SSIM metrics the best results are related to the highest values (indicated by an upwards arrow next to the name) while the perceptual index as well as RMSE should be as low as possible (illustrated by a downwards arrow).

Lastly, a different type of super-resolution task was performed using the brain MRI model. The trained generative networks from the already mentioned reconstruction task were used to generate images of higher resolution than the already high-resolution HR scans. The resulting image's dimensions were 256x256, 512x512 and 1024x1024 whilst the HR images used as the starting point were 128x128 pixels in size.

X2 AMPLIFICATION (HR: 128x128 | LR: 64x64)

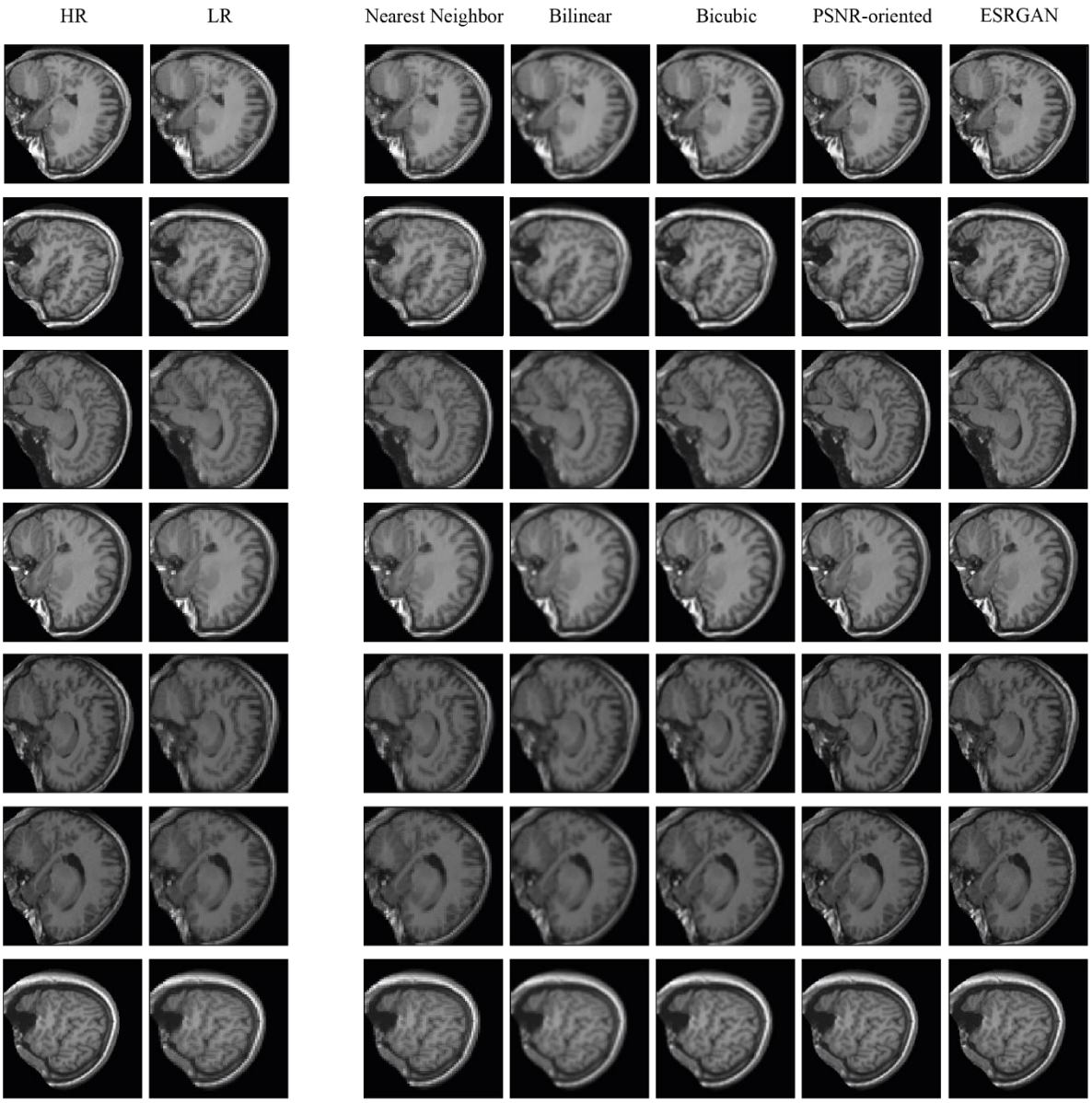

Figure 10. X2 Amplification on brain MRI validation set

X4 AMPLIFICATION (HR: 128x128 | LR: 32x32)

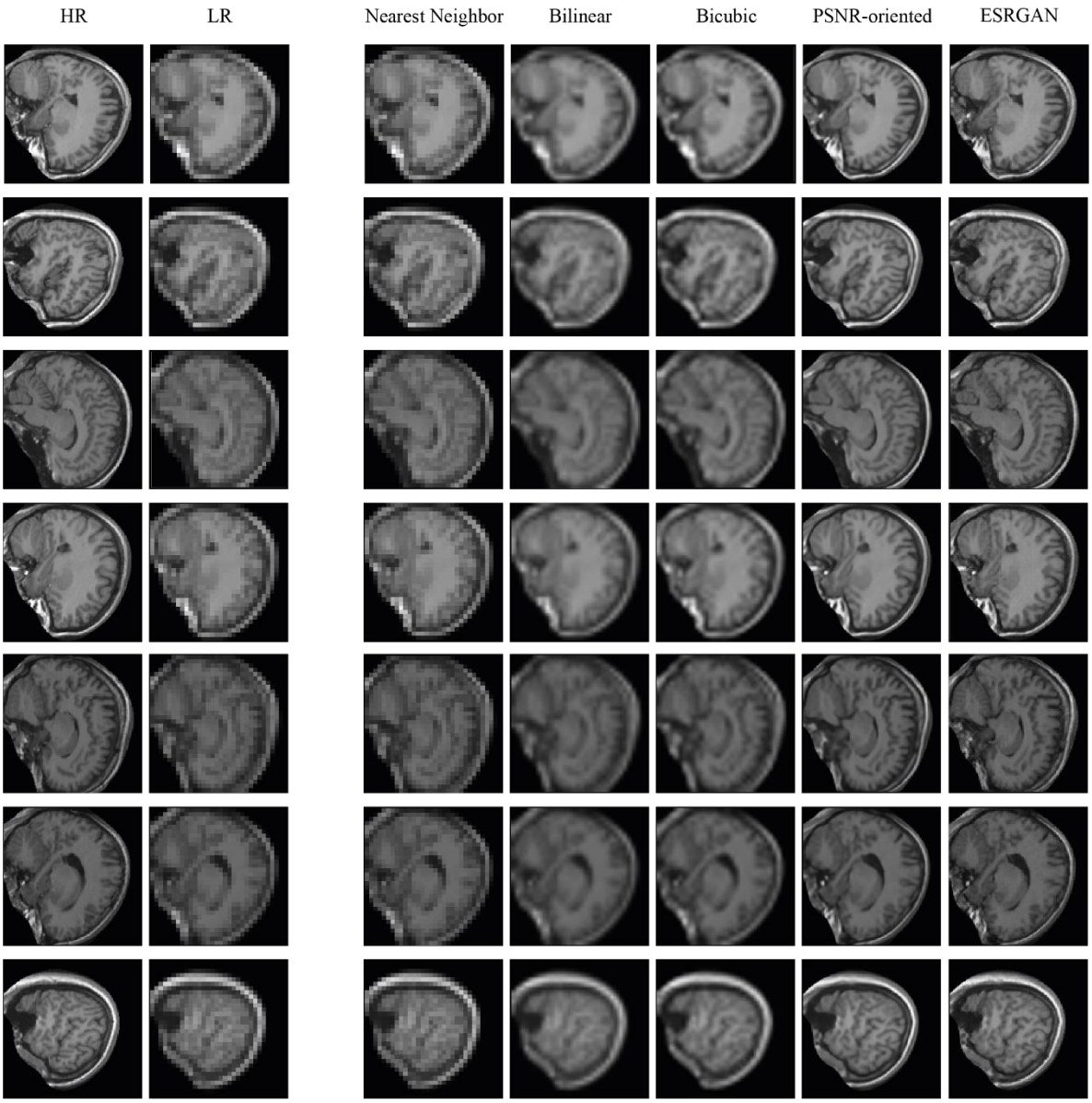

Figure 11. X4 Amplification on brain MRI validation set

X8 AMPLIFICATION (HR: 128x128 | LR: 16x16)

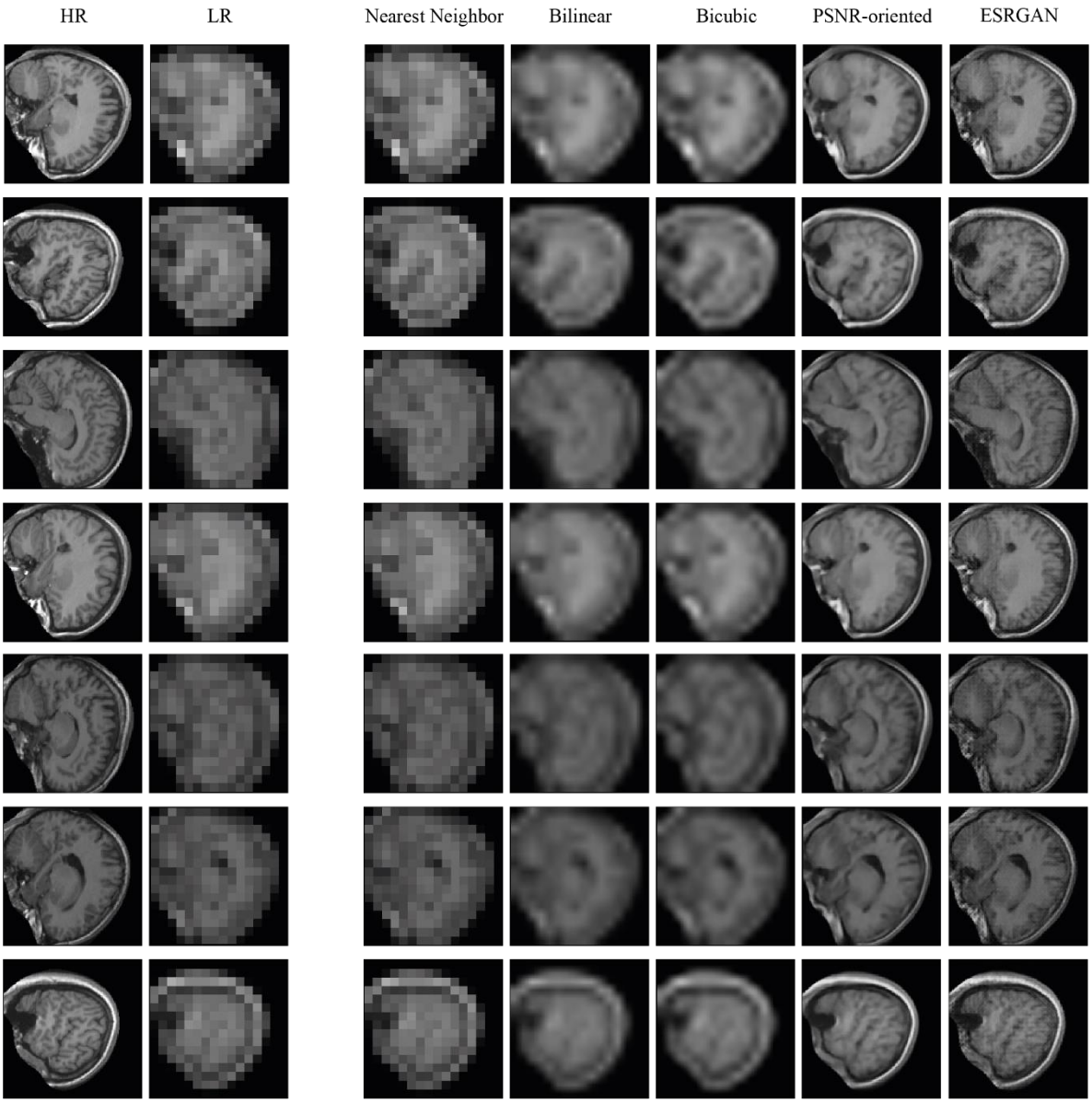

Figure 12. X8 Amplification on brain MRI validation set

Brain Images	*PSNR*↑	*SSIM*↑	*P.I.*↓	*RMSE*↓
X2				
ESRGAN	32.78	0.91	22.12	5.98
PSNR-Oriented	36.79	0.97	22.16	3.61
BICUBIC	34.18	0.96	22.85	5.32
BILINEAR	31.97	0.93	23.17	6.96
NEAREST	30.55	0.92	22.6	8.191
X4				
ESRGAN	31.43	0.91	22.28	7.08
PSNR-Oriented	34.40	0.95	22.45	4.98
BICUBIC	27.84	0.82	23.42	11.28
BILINEAR	26.74	0.77	23.66	12.85
NEAREST	25.67	0.75	22.64	14.37
X8				
ESRGAN	25.75	0.74	22.31	14.23
PSNR-Oriented	26.67	0.77	22.70	12.85
BICUBIC	23.74	0.62	23.94	18.16
BILINEAR	23.1	0.58	23.99	19.41
NEAREST	22.47	0.57	23.48	20.48

Table 1. Brain MRI results

As seen in figures 10, 11, and 12 the ESRGAN model is capable of reconstructing the original image by amplifying the degraded input data. In the x2 amplification we can see that the best results in terms of pixel-wise metrics (as seen in Table 1) are the ones generated by the PSNR-oriented model. Even though the ESRGAN model was outperformed in the mentioned similarity metrics, the best perceptual quality results (measured by P.I.) are generated by ESRGAN. The difference between the models is accentuated in the x4 and x8 tests. We can see that in the x2 amplification, the results obtained by ESRGAN in terms of PSNR, SSIM and RMSE are worse than the ones belonging to simpler interpolation methods, but when the up sampling task is more complex, the deep models show the top results.

X2 AMPLIFICATION (HR: 128x128 | LR: 64x64)

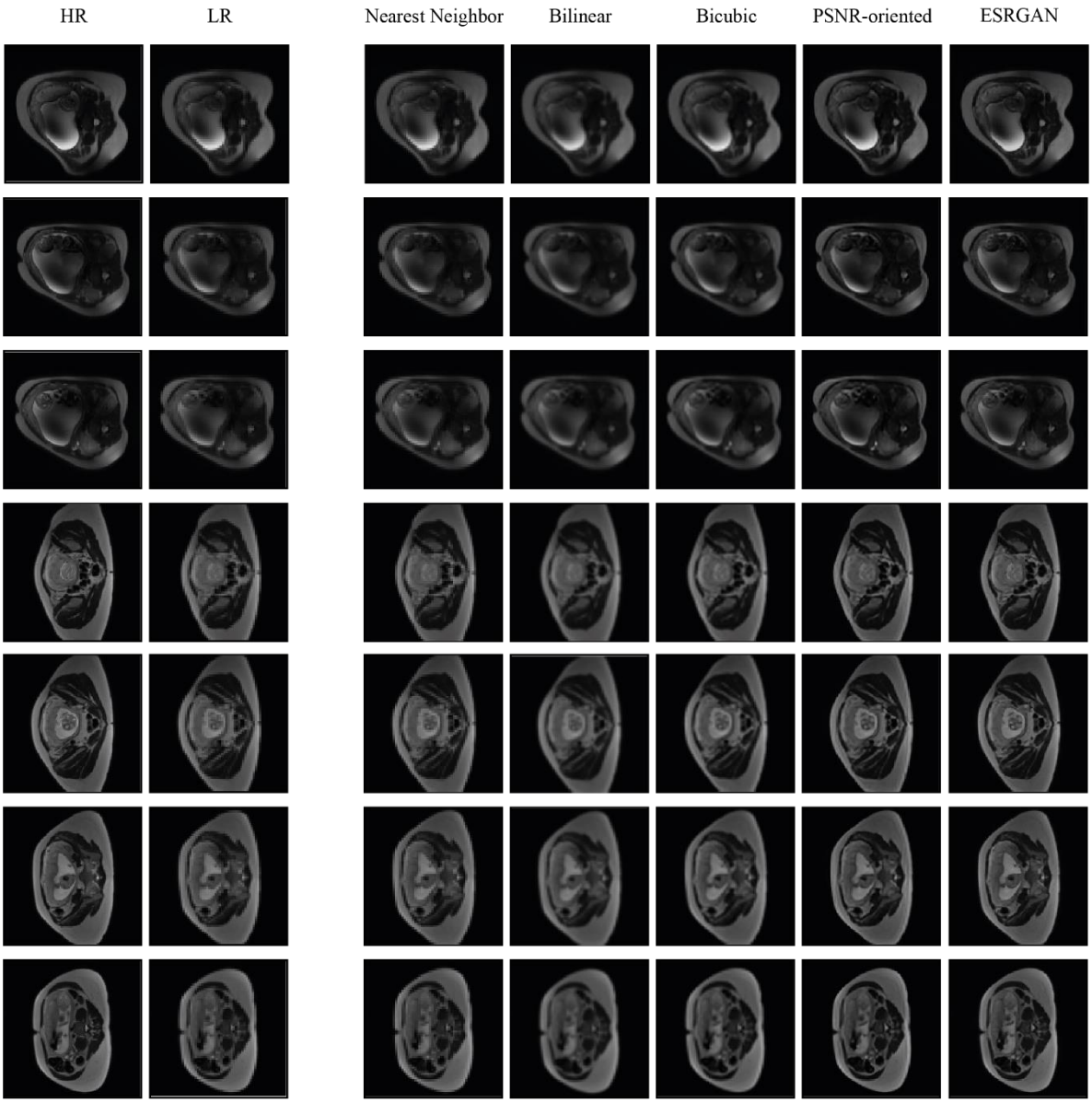

Figure 13. X2 Amplification on fetal MRI validation set

X4 AMPLIFICATION (HR: 128x128 | LR: 32x32)

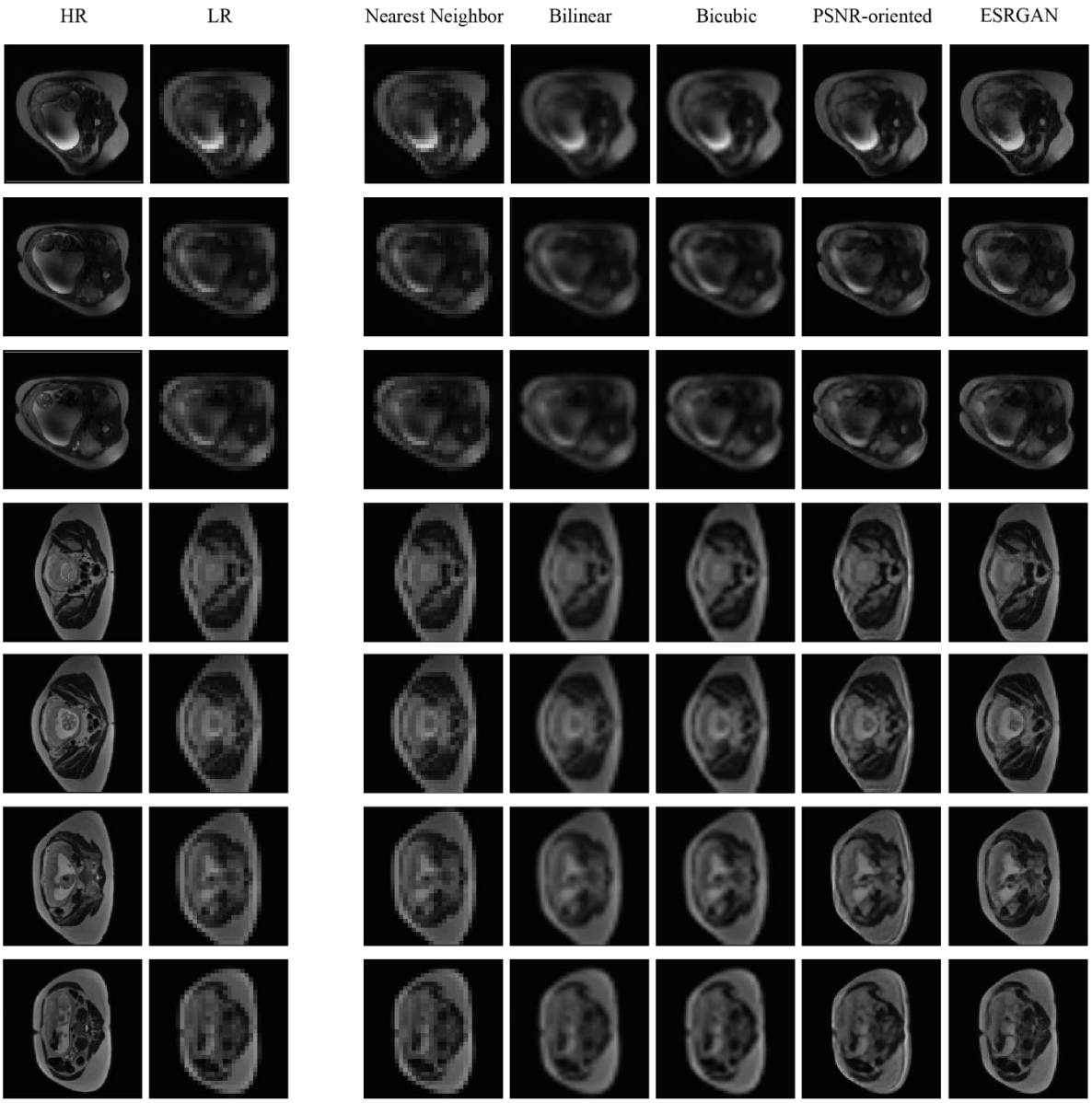

Figure 14. X4 Amplification on fetal MRI validation set

X8 AMPLIFICATION (HR: 128x128 | LR: 16x16)

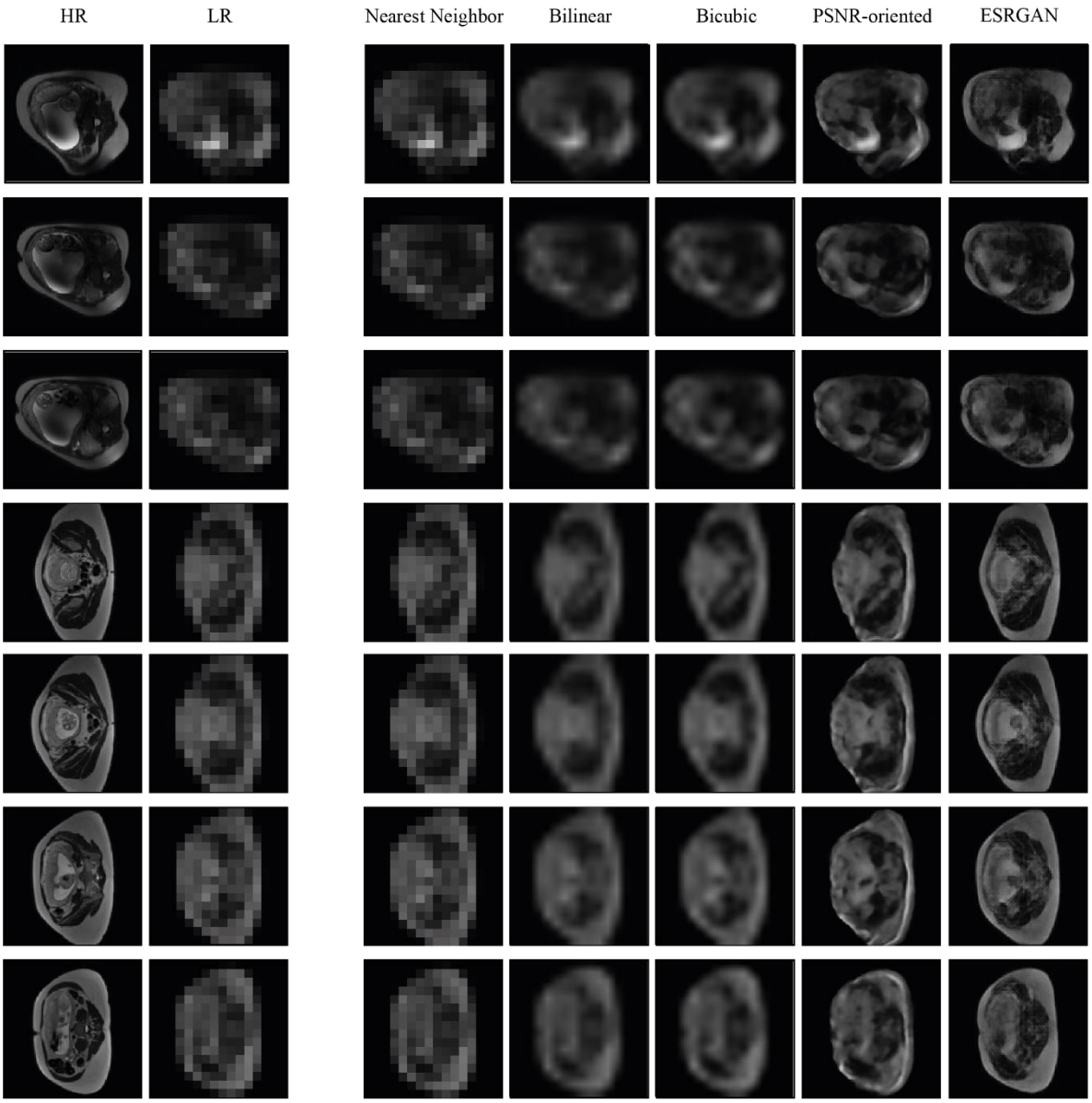

Figure 15. X8 Amplification on fetal MRI validation set

Fetal Images	*PSNR*↑	*SSIM*↑	*P.I.*↓	*RMSE*↓
X2				
ESRGAN	35.37	0.96	22.44	4.01
PSNR-Oriented	37.87	0.97	22.49	3.08
BICUBIC	34.71	0.95	23.09	4.51
BILINEAR	32.79	0.93	23.29	5.62
NEAREST	31.40	0.93	22.91	6.56
X4				
ESRGAN	29.26	0.86	22.67	8.51
PSNR-Oriented	31.29	0.89	23.10	6.51
BICUBIC	28.74	0.83	23.63	8.96
BILINEAR	27.65	0.80	23.73	10.13
NEAREST	26.51	0.78	23.12	11.52
X8				
ESRGAN	24.94	0.71	22.71	13.74
PSNR-Oriented	26.42	0.76	23.41	11.37
BICUBIC	24.57	0.65	24.02	14.41
BILINEAR	23.90	0.61	24.06	15.50
NEAREST	23.21	0.60	23.53	16.78

Table 2. Fetal MRI results

The results shown in Table 2 follow a similar distribution to the Brain MRI images, the ESRGAN model can achieve the highest perceptual quality while the PSNR-oriented model outperforms every other method in pixel-wise metrics. The results obtained in the fetal MRI images are worse in terms of the presented metrics than the brain MRI and, visually, both deep models fail to reconstruct the fine details of the original HR images in both x4 and x8 amplification, all of which is addressed in detail during the discussion of the results.

The mentioned results showcase a super-resolution performed on LR degraded images to reconstruct the original HR data, but once the networks have been trained, they can also be used to perform super-resolution on the already HR images, as seen in Figure 16 and 17. Here the input image is the HR scan, super-resolution of 2,4 and 8-fold was performed, the textures and details are compared and can be seen in the zoomed-in areas of the figures. The resulting images were of size 256x256, 512x512 and 1024x1024 respectively, obtained from the original full-size HR scans of 128x128 pixels.

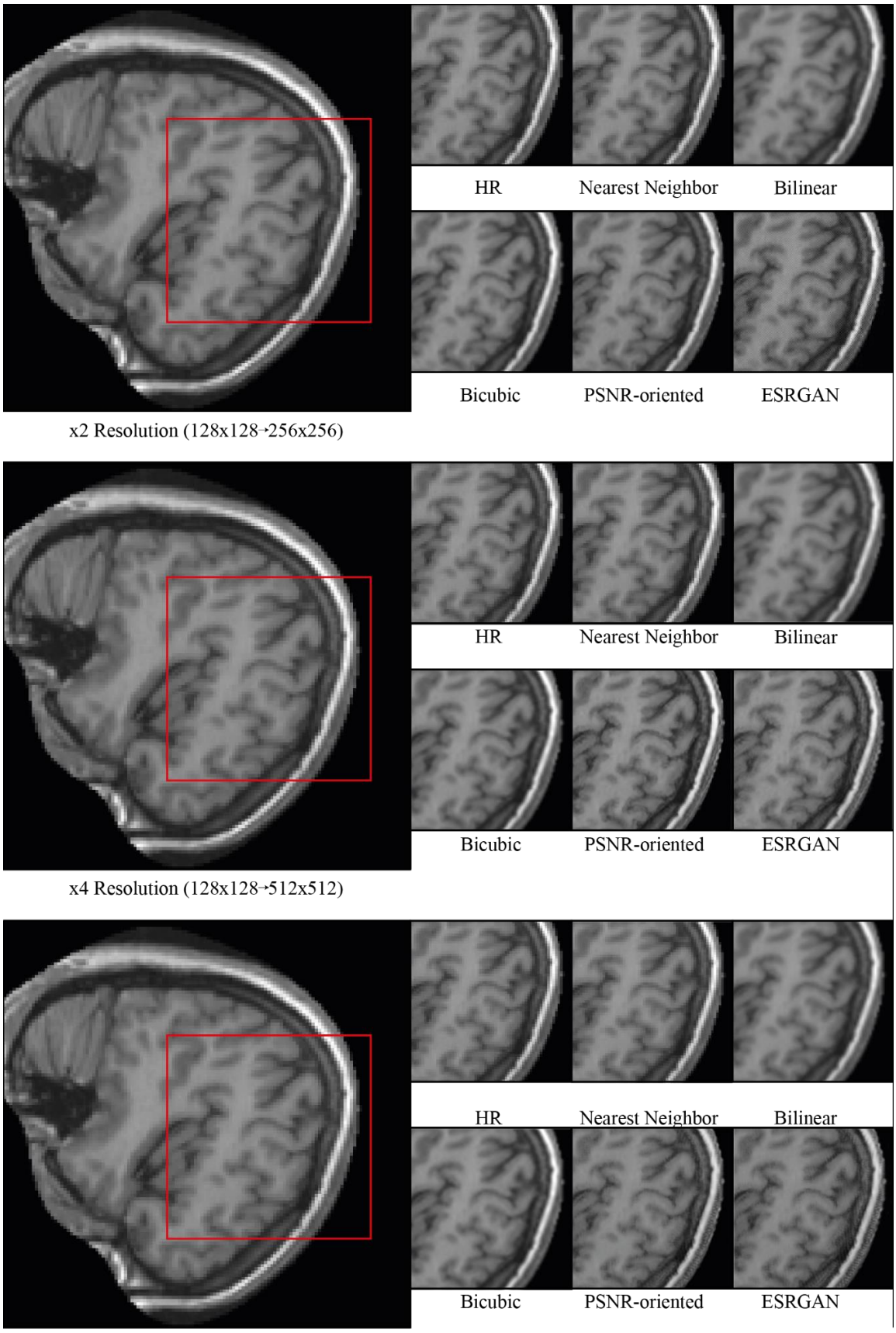

Figure 16. Comparison of brain MRI super-resolution on HR images.

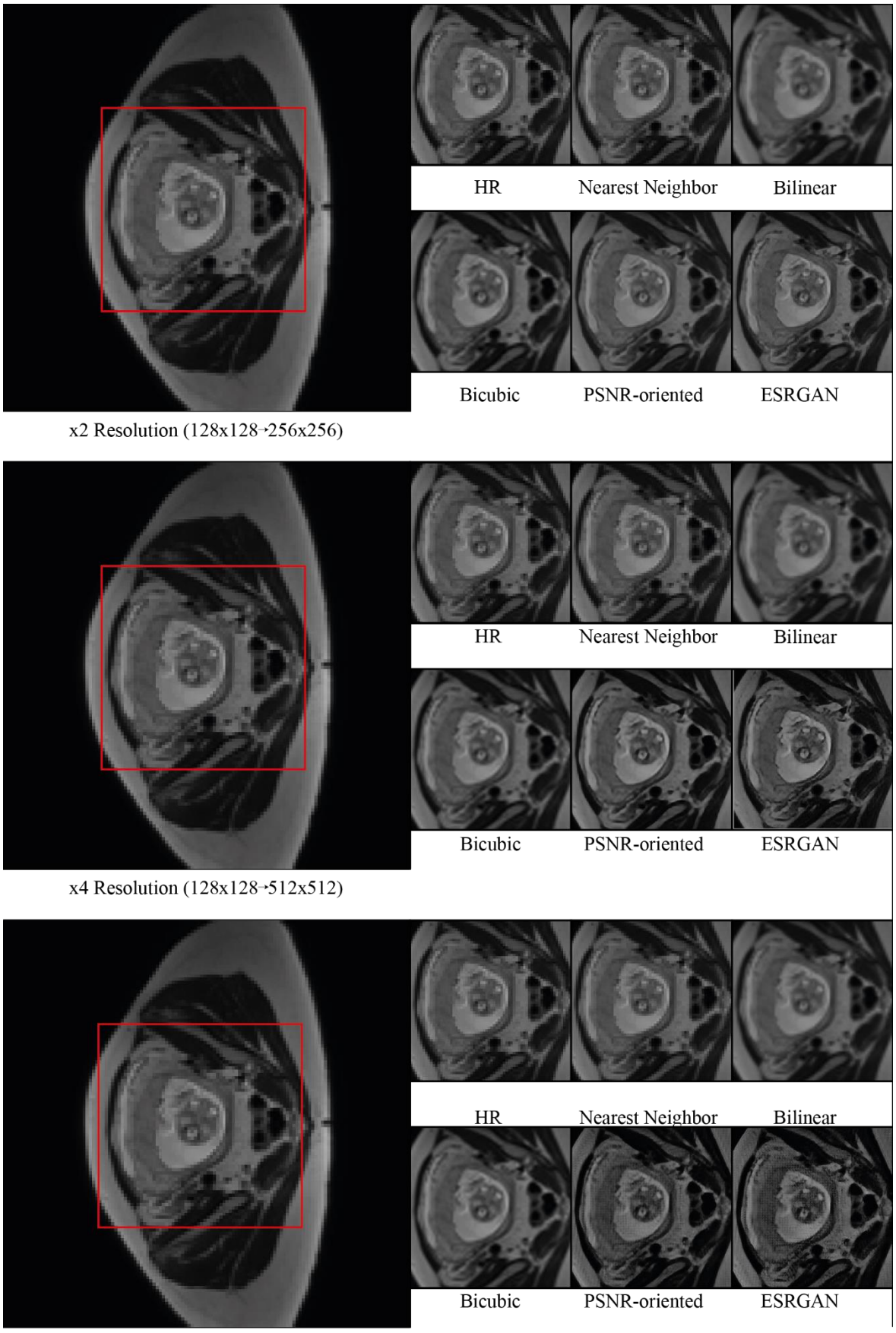

Figure 17. Comparison of fetal MRI super-resolution on HR images.

4. DISCUSSION

The obtained results showcase the ability of the trained models to successfully fulfill the super-resolution tasks. Nevertheless, the performance and quality of the results are not constant across all tests. Comparatively, the results of the brain MRI model are numerically and visually superior to fetal MRI. The main limitation when working with fetal MRI is the quality of the HR scans, which is inferior to that of the adult brain dataset. Also, the dataset size is 3 times smaller, affecting the overall quality of the training process. Even with these limitations the results obtained in x2 up-sampling were satisfactory and demonstrate the potential application of these networks to the brain and fetal imaging once properly trained.

By closely analyzing the super-resolution images in the x2 amplification we can detect certain image artifacts product of a suboptimal training process. During training, a batch of images is loaded and fed to the network to avoid using the whole dataset at all times. These images are handled by the GPUs provided by Google Colab which have 16GB of RAM. The optimal batch size following the recommendations of Xintao et al. [22] was 16, a number of images that was not possible to handle during the training of the x2 networks. Considering the fact that the HR images are of size 128x128 pixels, the x2 degraded LR images were of 64x64 pixels, loading a batch of size 16 containing these images caused GPU memory problems (notably there is also a large portion of memory allocated to the PyTorch library and models). The training was performed with batches of size 4, which caused the generator to undergo a worse training process compared to the x4 and x8 amplifications.

The restored images after an x8 amplification also suffer from artifacts and lack of fine detail. In this case, the main limitation is the LR image size and information. When HR images are degraded by 8-fold we obtain LR images of size 16x16 pixels, losing all information on structures and details. If the training of the model were to be performed with images of bigger sizes (Xintao et al. used HR images of 2K resolution [47]), the x8 amplification would start from an LR image that still contains enough information to reconstruct all the details.

Lastly, analyzing the results of performing super-resolution on HR images it can be seen that both deep models can obtain satisfactory results compared to the blurrier images product of simpler interpolation techniques. The ESRGAN images have richer textures and details, following the tendency seen in previous experiments where it outperformed the PSNR-oriented model in terms of perceptual quality. We can see how the same problems mentioned before are present in these images, the x2 amplification features checker-board-like artifacts and the x8 produces unrealistic textures. As seen during the previous tests, the x4 amplification model is the most balanced of the three, a product of richer LR images and satisfactory training with optimal hyper parameters.

The proposed models show promising results in image super-resolution tasks with potential applications in the reconstruction of HR scans and the generation of images with superior resolution. These models, if properly trained, could be used in the reconstruction of fast or open MRI acquisitions without sacrificing image quality and optimizing acquisition times and costs. They could also be used on normal-time MRI acquisitions to obtain super-resolution images of the patient without the need of using a research-grade

high magnetic flux scanner, enhancing details and textures and improving the outcome of diagnosis and detection tasks.

Potential future steps for improving the model include making the up-sampling process progressive; instead of performing the x8 super-resolution in a single step, we could obtain it through 3 consecutive x2 operations. Introducing this new way of up-sampling also opens up the opportunity to implement the 'Triplet Loss' developed by Mahapatra et al. [29] which yielded outstanding results. Finally, a noise elimination step previous to the amplification could be introduced to cancel out potential artifacts that real LR MRI images tend to have. In the present model, there is no training for removing image noise which could cause a decrease in performance. In a real clinical scenario, MRI images obtained with shorter or open scans tend to have artifacts and image noise which should be removed before the up-sampling. Following the work done by Yuan Y. et al. [31] a cycle-GAN could be implemented to perform domain transfer between the noisy LR data and the clean LR images which would be fed to the amplification model.

5. CONCLUSIONS

The results in super-resolution tasks demonstrate the possible applications of GANs in HR scan reconstruction and medical image synthesis. The trained networks were able to improve MRI scan resolution while maintaining realistic textures and details, outperforming typical image interpolation or PSNR-oriented models. They showed potential to also be utilized on regular time MRI acquisitions to obtain super-resolution images of the patient without the need for longer scans or more powerful machinery. Overall the use of the proposed models could improve the clinical landscape by optimizing costs, reducing wait times and improving the accuracy of the detection and diagnosis of lesions, diseases or abnormalities.

BIBLIOGRAPHY

[1] A. Berger, "Magnetic resonance imaging.," *BMJ*, vol. 324, no. 7328, p. 35, Jan. 2002, doi: 10.1136/bmj.324.7328.35.

[2] Y. Chen, A. G. Christodoulou, Z. Zhou, F. Shi, Y. Xie, and D. Li, "Mri super-resolution with gan and 3d multi-level densenet: Smaller, faster, and better," *arXiv Prepr. arXiv2003.01217*, 2020.

[3] R. F. Kilcoyne, M. L. Richardson, B. A. Porter, D. O. Olson, T. K. Greenlee, and W. Lanzer, "Magnetic resonance imaging of soft tissue masses.," *Clin. Orthop. Relat. Res.*, no. 228, pp. 13–19, Mar. 1988.

[4] K. Zhang *et al.*, "SOUP-GAN: Super-resolution MRI using generative adversarial networks," *Tomography*, vol. 8, no. 2, pp. 905–919, 2022.

[5] E. Van Reeth, I. W. K. Tham, C. H. Tan, and C. L. Poh, "Super-resolution in magnetic resonance imaging: A review," *Concepts Magn. Reson. Part A*, vol. 40A, no. 6, pp. 306–325, 2012, doi: https://doi.org/10.1002/cmr.a.21249.

[6] H. Chen *et al.*, "Real-world single image super-resolution: A brief review," *Inf. Fusion*, vol. 79, pp. 124–145, 2022.

[7] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep Learning for Image Super-Resolution: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, 2021, doi: 10.1109/TPAMI.2020.2982166.

[8] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[10] H.-I. Suk, S.-W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis.," *Brain Struct. Funct.*, vol. 220, no. 2, pp. 841–859, Mar. 2015, doi: 10.1007/s00429-013-0687-3.

[11] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation.," *Neuroimage*, vol. 108, pp. 214–224, Mar. 2015, doi: 10.1016/j.neuroimage.2014.12.061.

[12] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1930–1943, Aug. 2013, doi: 10.1109/TPAMI.2012.277.

[13] H. R. Roth *et al.*, "Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation.," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1170–1181, May 2016, doi: 10.1109/TMI.2015.2482920.

[14] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[15] K. Fukushima, "Neocognitron: a self organizing neural network model for a

mechanism of pattern recognition unaffected by shift in position.," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980, doi: 10.1007/BF00344251.

[16] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex.," *J. Physiol.*, vol. 195, no. 1, pp. 215–243, Mar. 1968, doi: 10.1113/jphysiol.1968.sp008455.

[17] I. S. Mohamed, "Detection and Tracking of Pallets using a Laser Rangefinder and Machine Learning Techniques," 2017. doi: 10.13140/RG.2.2.30795.69926.

[18] I. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, 2017, pp. 214–223.

[20] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv Prepr. arXiv1701.04862*, 2017.

[21] K. Roth, A. Lucchi, S. Nowozin, and T. Hofmann, "Stabilizing training of generative adversarial networks through regularization," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[22] X. Wang *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, p. 0.

[23] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 pirm challenge on perceptual image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, p. 0.

[24] A. Hindupur, "GAN Zoo," *GitHub repository*. GitHub, 2018.

[25] * Hanxiao Zhang Xiaoliang Wang Frank Jiang Kui Fu Jiansheng Peng, "Image Super-Resolution Based on Generative Adversarial Networks: A Brief Review," *Comput. Mater. \& Contin.*, vol. 64, no. 3, pp. 1977–1997, 2020, doi: 10.32604/cmc.2020.09882.

[26] R. R. Schaller, "Moore's law: past, present and future," *IEEE Spectr.*, vol. 34, no. 6, pp. 52–59, 1997, doi: 10.1109/6.591665.

[27] R. P. França, A. C. Borges Monteiro, R. Arthur, and Y. Iano, "Chapter 3 - An overview of deep learning in big data, image, and signal processing in the modern digital age," in *Trends in Deep Learning Methodologies*, V. Piuri, S. Raj, A. Genovese, and R. Srivastava, Eds. Academic Press, 2021, pp. 63–87. doi: https://doi.org/10.1016/B978-0-12-822226-3.00003-9.

[28] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[29] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, "Image super-resolution using progressive generative adversarial networks for medical image analysis.," *Comput. Med. imaging Graph. Off. J. Comput. Med. Imaging Soc.*, vol. 71, pp. 30–39, Jan. 2019, doi: 10.1016/j.compmedimag.2018.10.005.

[30] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image

translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[31] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 701–710.

[32] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2472–2481.

[33] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[34] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[35] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN," *arXiv Prepr. arXiv1807.00734*, 2018.

[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prepr. arXiv1409.1556*, 2014.

[37] C. Sammut and G. I. Webb, Eds., "Mean Absolute Error," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2010, p. 652. doi: 10.1007/978-0-387-30164-8_525.

[38] I. C. L. & K. C. London, Ed., "Hammers adult brain dataset," 2008. [Online]. Available: https://brain-development.org/brain-atlases/

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, 2014.

[40] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems*, 2019, vol. 32. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

[41] M. Nadipally, "Chapter 2 - Optimization of Methods for Image-Texture Segmentation Using Ant Colony Optimization," in *Intelligent Data Analysis for Biomedical Applications*, D. J. Hemanth, D. Gupta, and V. Emilia Balas, Eds. Academic Press, 2019, pp. 21–47. doi: https://doi.org/10.1016/B978-0-12-815553-0.00002-1.

[42] J. Nilsson and T. Akenine-Möller, "Understanding ssim," *arXiv Prepr. arXiv2006.13846*, 2020.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004, doi: 10.1109/TIP.2003.819861.

[44] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, “Learning a no-reference quality metric for single-image super-resolution,” *Comput. Vis. Image Underst.*, vol. 158, pp. 1–16, 2017.

[45] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a ‘Completely Blind’ Image Quality Analyzer,” *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013, doi: 10.1109/LSP.2012.2227726.

[46] A. Works, Ed., “Anvil Works Web Apps,” [Online]. Available: https://anvil.works/

[47] R. Timofte *et al.*, “NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1110–1121. doi: 10.1109/CVPRW.2017.149.