A Comparison of Single Image Super-Resolution Algorithms for Cardiac MRI Scans

Frank T. Usher supervised by Dr Lian Gan

Abstract*—Objective:*** The comparison of two Single Image Super-Resolution (SISR) deep learning algorithms, Very Deep Super Resolution (VDSR) and a Super-Resolution Generative Adversarial Network (SRGAN), to enhance spatial resolution of low-resolution (LR) cardiac magnetic resonance imaging (MRI) scan slices, and to assess their utility as input to bi-ventricular segmentation programs. *Approach:* both networks were modified to accept MRI input data, and two artificial low-resolution (LR) MRI datasets were created from a single high-resolution (HR) dataset. The two datasets were divided into 806 HR/LR image training pairs and 225 testing pairs. The networks were trained and compared subjectively and by comparison of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). *Results:* For both datasets, SRGAN produced the most realistic output, and was most effective in restoring cardiac structural demarcation. VDSR was most effective in restoring correct cardiac structural form during SR, outperforming SRGAN and Bicubic interpolation on measures of PSNR and SSIM for Dataset 1 (P<0.001) and outperformed bicubic (P<0.001) and SRGAN for both measures in Dataset 2, but its poor edge-restoration can cause blending of cardiac boundaries, complicating the segmentation task. Some forms restored by SRGAN had been visibly altered compared to the ground truth, which could affect ventricular volume and wall thickness determination. SRGAN was outperformed on measures of PSNR and SSIM by VDSR and bicubic interpolation for Dataset 1 (P<0.001) and appeared to be outperformed by VDSR for both measures in Dataset 2. *Significance:* SISR networks using MSE loss appear to be more suitable than both GAN-based networks and interpolation methods for generating HR MRI images suitable for the ventricular segmentation task. SRGAN appeared to be more effective than bicubic interpolation for SR of poorer quality LR input data.

***Index Terms*—Deep Learning, GAN, MRI, Single Image Super-resolution**

I. Introduction

One limitation of cardiac magnetic resonance imaging (MRI) is its long acquisition time, often requiring the patient to hold their breath for extended periods to reduce motion artifact (blurring) caused by bodily movements. Acquisition time can be reduced if smaller-matrix images are captured, but this comes at the cost of reduced spatial detail (a lower resolution scan of the same size) or reduced coverage (maintaining scan quality while reducing spatial volume) [1], as well as a lower signal-to-noise ratio [3]. This diminishes the ability of professionals and cardiac segmentation software to accurately identify and diagnose cardiac diseases. The use of deep learning, specifically single-image super-resolution (SISR) algorithms to posteriorly increase the spatial detail of low-resolution (LR) scans is proving itself to be a promising solution to this trade-off.

Additionally, there is a growing field of study concerned with automatic in-slice and 3D ventricular segmentation programs, which use deep learning to identify the left or right ventricle on an MRI scan. This aids diagnosis by helping to determine quantitative measures such as ventricular volume and mass fraction, and has historically been a manual task performed by an expert [24]. At its core, ventricular segmentation relies on high quality scans to deliver good results; one study on right ventricle segmentation used high-resolution (HR) MRI input scans of between 0.75 mm/pixel and 1.6 mm/pixel [30]. It is envisaged that a LR MRI scan could be obtained, passed through a SISR network to increase the resolution, before passing the super-resolved (higher resolution) scan to a segmentation program. This process could provide an expert with data relevant to diagnosis in a matter of seconds, where previously it would have taken a cardiologist up to 15 minutes [30]. As such, deep learning methods have the potential to revolutionise the speed and facility of cardiac diagnosis.

This project focuses on the first stage of the process, aiming to compare two well-known SISR algorithms, a Very Deep Super-Resolution (VDSR) network and a Super-Resolution Generative Adversarial network (SRGAN), for the purposes of improving the resolution of individual cardiac MRI image slices. A HR sagittal plane (parallel to the long and sagittal axes of the body) dataset was obtained from the HVSMR 2016 Challenge on whole-heart and great vessel segmentation in MRI [29]. This dataset was cropped and 4× downsampled (the factor by which image resolution is decreased in each dimension) to create two corresponding artificial LR datasets with in-plane resolution of 3.6 × 3.6 mm/pixel, which were subsequently super-resolved by the networks to restore the original resolution of 0.9 × 0.9 mm/pixel, acceptable for input to a ventricular segmentation program, as shown in Fig. 1.

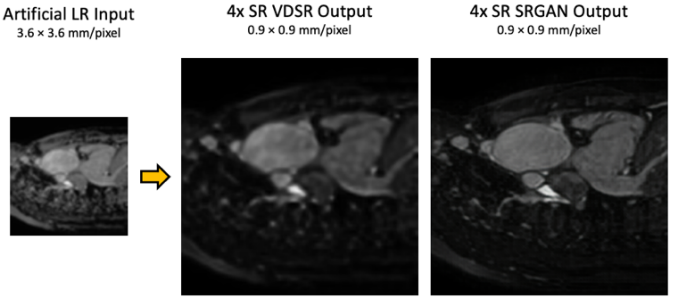

Figure 1 - Artificial LR input image from Dataset 2, and super-resolved output for both networks.

Each dataset was partitioned into 806 HR/LR training image pairs and 225 testing pairs. Performances were assessed by

inspection and by comparing peak signal-to noise-ratio (PSNR) and structural similarity (SSIM) scores (explained in *Theory. D*). Statistical significance was determined using one-tailed Wilcoxon signed-rank testing [33]. A 4× downsampling factor was chosen so the artificial LR images would have a similar resolution to real LR images. For example, the LR MRI reference dataset used in this study had in-plane resolution of 3 × 3 mm/pixel. 3× downsampling would have generated images which were closer to this resolution, but super-resolution (SR) effects are harder to analyse for smaller upsampling factors.

These two networks were chosen due to their significant differences in function and objective. VDSR optimises its image output to be as similar as possible to the ground truth (original HR image) by minimising the mean squared error (MSE), which measures the intensity difference between corresponding pixels in the two images. In contrast, SRGAN can create more realistic images by adding features and fine texture details which aren't present in the ground truth [7]. This ability to hallucinate details is harmless in more benign applications, but when applied to cardiac MRI scans it could reduce the output image's reliability as a diagnostic tool.

A. Objectives

With so much academic focus being placed on the application of generative adversarial networks (GANs) to the SR problem, other deep learning networks are receiving less attention, particularly those focused on MSE reduction. However, despite the many merits of GAN-based SISR algorithms, they are theoretically poorly suited to medical imaging applications. Input scans to bi-ventricular identification programs must accurately display cardiac features, and networks aiming to reduce the MSE loss, such as VDSR, could potentially be better suited to this task than GAN based networks due to their stricter preservation of image features.

This study investigates differences in performance between the two SISR networks of Cardiac MR scans and assesses the impact of hallucinated details created by SRGAN on upsampled MRI images, as even minor imagined details on scans could lead to inaccurate diagnoses when used as input to ventricular segmentation programs.

II. Theory

A. An Overview of Neural Networks

Neural networks are a method of deep learning, which itself is a type of machine learning. Any neural network consists of three types of layers: an input layer, an output layer and one or several hidden layers in between. Each layer is made up of nodes, and each connection between two nodes in adjacent layers is assigned a weight corresponding to the signal strength through that connection.

The weights in the network must be trained to minimise the error between its output and its desired output. This is done using something called the *back propagation algorithm;* the breakthrough that led to the current explosion in performance being witnessed in the field [21]. Training a neural network involves several steps:

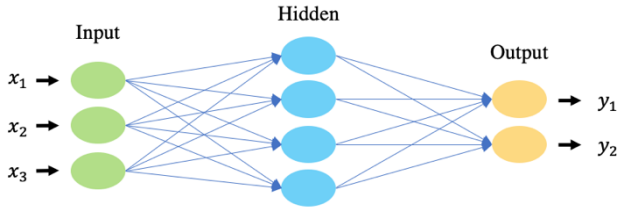

Figure 2 - Structure of a neural network, showing inputs and outputs, layers, nodes and nodal connections.

1. The network input is taken from the training data in the form of *{input, desired output}*. In the case of SISR, *input* is an artificial LR image, and *desired output* is its corresponding HR ground truth image.

2. Once an output (SR image) has been obtained, the error between *output* and *desired output* is calculated. Different loss (error) functions are used depending on the type of network and its goals.

3. The error between *output* and *desired output* is 'propagated' back through each layer of the network using the back propagation algorithm, starting from the output layer and arriving back at the input layer. The error at each node is calculated and its weight is updated to minimise this error.

The process from step 1 to 3 is called an *epoch* [22]. If a network has been trained to 100 epochs, it means these steps were performed 100 times, and each time the error between *output* and *desired output* is reduced. Another important feature is the *learning rate*, which governs how much the weights are changed per iteration. If this value is too high, the output won't converge. If it is too low, the output will converge too slowly [21].

B. Image Super-Resolution Methods

SISR aims to recover a single high-resolution (HR) image from a LR version of the same image. The problem is ill-posed since many different HR images can be created from the same LR input image, and this becomes increasingly pronounced for greater upscaling factors [7]. Essentially, the goal is to find the inverse of an unknown degradation function applied to the ground truth [6]. Deep learning models do this using a Convolutional Neural Network (CNN) [14], a deep neural network arranged into specialised layers designed to independently identify and extract image features, such as edges, and increase resolution by filling in unknown details.

Initially, simple mathematical interpolation techniques were used for upsampling LR images [9]. Since then, researchers have developed a variety of SR models based on deep learning. Most of these models use supervised learning, which means HR ground truth images and their corresponding artificial LR images are used to train the network. All deep learning SR models involve a combination of four things: A basic framework, an upsampling method, a network design, and a learning strategy. Different combinations of these components can be used to create SR models with varying purposes [4]. The methods in this section have been grouped by their framework.

i. Interpolation Techniques

Interpolation (nearest neighbour, bilinear, bicubic) estimates the value of an unknown pixel from the known values of neighbor pixels [9]. Bicubic interpolation is the most commonly used interpolation technique in deep learning and considers the closest 4×4 grid of known pixels, with the influence of each being weighted by its distance from the unknown pixel.

ii. Pre-Upsampling Methods: SRCNN, VDSR

It's difficult to directly map between a LR image and a ground truth image. In the pre-upsampling framework, a traditional interpolation method is first used to increase the dimensions of the LR image to those of the ground truth. The interpolated image can then be mapped to the ground truth and refined using a deep neural network [4]. The first pre-upsampling SR method was called Super Resolution Convolutional Neural Network (SRCNN). It first upsampled the LR images using bicubic interpolation, before applying a three-layer CNN to recreate finer details, where the CNN is trained to reduce the MSE loss [14].

The Very Deep Super-resolution (VDSR) network works like SRCNN but improves performance by using 20 convolutional layers in the CNN instead of three. It does this by using a residual loss function, shown in (1).

$$e_r = 1/2||r - f(x)||^2 \quad (1)$$

where r is the residual image, and $f(x)$ is the network prediction. The residual image shows the pixel intensity differences between the SR and ground truth images. Training efficiency is improved by predicting only this residual image, rather than the HR output itself, as is the case in SRCNN. During training, the weights of the network are retrained using the back propagation algorithm to minimise the residual error. The residual image is much less informationally dense than the SR image itself, shown in Fig. 3. This means that substantially deeper networks can be trained, and converge, in a shorter time than was previously possible [12]. VDSR also employs multi-scale training, meaning it can use 2×, 3× and 4× downsampled images during training, and reduce the MSE loss for all three rather than having to separately retrain the network for each scale factor.

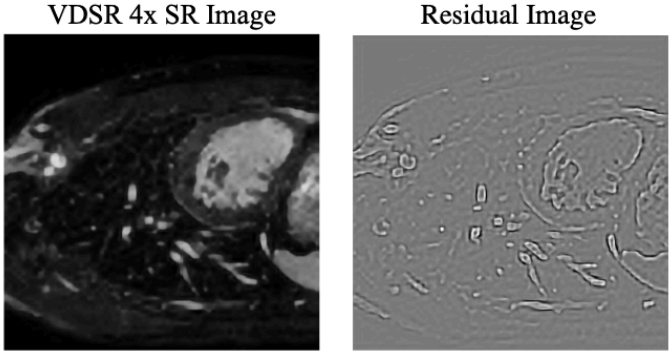

Figure 3 - VDSR SR testing image, and the corresponding residual image between the SR and ground truth images.

iii. Post-Upsampling Methods

Post-upsampling methods work in the opposite way to pre-upsampling methods. They first feed LR input images into deep CNNs, and then carry out the interpolation upsampling step afterwards. This improves computational efficiency as LR images are less informationally dense [11].

iv. Residual Networks (ResNets): CARN

Residual Networks (ResNets) are categorised by their use of a residual loss function (1) [12] VDSR is also a ResNet, but since it has a pre-upsampling structure it was included in section *B. ii.* Another network in this category is the Cascading Residual Network (CARN) [13]. It reduces the number of computational operations by giving the model access to various levels of features, and therefore more information. CARN is suitable for practical applications as it's fast and lightweight [13], but with slightly reduced performance.

v. Generative Adversarial Networks: SRGAN, ESRGAN

The Super-Resolution GAN (SRGAN) was the first SR method to use something called a generative adversarial network (GAN). Previous SR methods lacked fine texture detail in the SR images at high upscaling factors. This is largely due to the choice of loss function. Most SR upscaling methods before SRGAN focused on reducing the MSE loss. However, this doesn't necessarily lead to a perceptually better result, as it only assesses pixel-level differences in an image, failing to capture relevant texture details. As a result, the SR image has soft edges and lacks fine detail, appearing unrealistic and perceptually unsatisfying [7].

SRGAN attempts to solve this issue by inferring more natural looking images, particularly for greater scaling factors (4× and higher). It does this by implementing a new loss function, called perceptual loss, which assesses a solution based on perceptually relevant image characteristics. It comprises an adversarial loss and a content loss element. Simply, SRGAN consists of two separate networks "competing" against one another. The first network, called the generator, generates an SR image using a ResNet with content loss, similar to VDSR. The second network, called the discriminator, takes either an SR image or a HR ground truth image as input and attempts to determine if the image is a 'real' HR image, or an artificially generated one. The output of the discriminator is binary; 1 if it believes the image is 'real', and 0 if it thinks the image has been generated. Both networks are trained simultaneously. The generator is trained to 'fool' the discriminator by generating ever more realistic images, and the discriminator is trained to better discriminate between 'real' and generated images [4].

Enhanced SRGAN (ESRGAN) [8] improved upon this idea by creating a discriminator that predicted relative realness by outputting a value between 0 and 1 rather than the binary output used in SRGAN. The perceptual loss was also modified to improve texture recovery and brightness consistency, improving image realness and reducing the unwanted noise added during upsampling.

C. Previous Applications of SISR methods to Cardiac MRI

Several studies have already investigated the use of SISR methods to improve MRI quality. J. Liu et al (2019) [31] proposed an edge enhanced SRGAN (EE-SRGAN) for MRI

SISR in the slice-select direction, to reduce information loss and artifacts which could affect disease diagnosis and treatment. The use of a GAN improved the visual reality of reconstructed images, and an edge enhanced loss function in the generator improved texture detail. The study suggested that the MRI-specialised EE-SRGAN provided a better visual effect beneficial to the segmentation task compared with other state-of-the-art SR networks, but the method was not assessed on cardiac MRI images.

E. Masutani et al (2020) [1] trained CNNs on short-axis cardiac MRI images to perform image SR using artificially generated LR data. CNNs were compared against bicubic interpolation and Fourier-based zero padding using SSIM between HR ground truth and each upscaling method, and clinical performance was evaluated by measuring left ventricular volumes. All CNNs outperformed the interpolation methods at upsampling factors ranging from two to 64 (P<0.001). SR images yielded left ventricular volumes comparable to those from HR images. Super-resolving HR images seemed to further enhance anatomic detail, and the method also appeared to improve image quality in other planes. However, the study compared two CNNs which both used MSE loss, and a GAN-based network was not investigated.

D. Methods of Performance Evaluation

A key challenge in SR is how to measure effectiveness. Distinguishing perceptually better images is a simple task for humans but challenging for an algorithm [2]. Image Quality Assessment (IQA) methods have helped drive considerable progress in the field. While by objective measures of quality, SR networks continue to improve, the generated images don't necessarily look more convincing. In recent years with the introduction of GANs, more subjective IQA methods have been introduced, with a greater focus on perceptual quality [6].

Subjective evaluation methods rely on human judgement [10] and are rather impractical and time consuming [6]. One example is the Mean Opinion Score (MOS), which measures the mean of subjective human image quality ratings from 1 to 5. Conversely, objective image evaluation methods are fast and use comparisons based on numerical criteria [10]. In this study, two objective measures were used: PSNR and SSIM [4].

Peak signal-to-noise ratio (PSNR) can be defined as [6]: *"A mathematical measure of image quality based on the pixel-level differences between two images."* In this case, the two images are the SR generated image and the HR ground truth image. It can be described by (2).

$$PSNR = 10\log \frac{L^2}{1/N \sum_{i=1}^{N} (I(i) - \hat{I}(i))^2} \qquad (2)$$

where L is the maximum pixel value, 255 for 8-bit images, and N is the number of pixels in each image. The denominator calculates *MSE*, defined as the mean square of the greyscale intensity difference between each corresponding pixel $I(i)$ and $\hat{I}(i)$ in the two images (a value between 0 and 255, 0 = black, 255 = white) [7]. Since a zero MSE is ideal, a greater PSNR indicates the images are more similar [10]. PSNR is simple to calculate and has a clear physical meaning. However, it is a poor indicator of perceived visual quality [7].

The Structural Similarity Index (SSIM) measures the similarity in structure between the SR image and the HR ground truth image, rather than the pixel-level difference. It does this by comparing the similarity of the mean luminance, the contrast and the correlation between structures (such as edges) in the two images, assigning a relative importance to each of these three criteria [4]. This is represented by (3).

$$SSIM(I, \hat{I}) = [C_l(I, \hat{I})]^{\alpha} [C_c(I, \hat{I})]^{\beta} [C_s(I, \hat{I})]^{\gamma} \qquad (3)$$

where α, β and γ are control variables adjusting the relative importance of each term, and $C_l(I, \hat{I})$, $C_c(I, \hat{I})$ and $C_s(I, \hat{I})$ are luminance, contrast and structure comparisons, respectively [4]. One drawback of objective methods is that they need a HR ground truth image with which to compare the SR image. This means they can only be used on artificially downsampled datasets, as real LR images have no ground truth counterpart.

VDSR and SRGAN were chosen for this study as they are two pure examples of an MSE loss and adversarial network that lent themselves strongly to comparison. Other networks were also considered, such as SRCNN, CARN and ESRGAN. However, VDSR is essentially a deeper version of SRCNN, CARN is lightweight at the expense of output quality, and there is limited publicly available code for ESRGAN.

III. METHODOLOGY

First, existing VDSR and SRGAN programs were modified to accept MRI input data, and two artificial LR MRI datasets were created from the HR ground truth set. Dataset 1 was created by simple 4× bicubic downsampling of the HR scans. Dataset 2 was created by altering the downsampled images to resemble the LR reference dataset as closely as possible through histogram matching and smoothing. Both networks were then trained and tested on these datasets, paired with their corresponding ground truth images. Results were compared subjectively and by comparison of PSNR and SSIM scores.

A. Creating the Datasets

In this study, a dataset of HR cardiac scans was downsampled to generate two artificial LR datasets. Each of these two datasets was then split into 806 training pairs and 225 testing pairs. A pair consisted of a HR ground truth image and its corresponding downsampled LR image. A separate reference dataset of real LR Cardiac MRI scans was used to provide a point of comparison when creating artificial LR images.

The LR MRI reference dataset was taken from scans acquired during a previous study [27] containing 18 healthy participants acquired from Newcastle Hospital's NHS Foundation Trust (structurally normal hearts, age range 21–50). A Phillips Achieva at 3.0 T was used, with a cardiac phased array coil. Multi-slice, multi-phase fast-field echo cine scans in three orthogonal planes were acquired during free breathing. Field of View: 240 mm (antero-posterior) × 240 mm (inferior-superior)

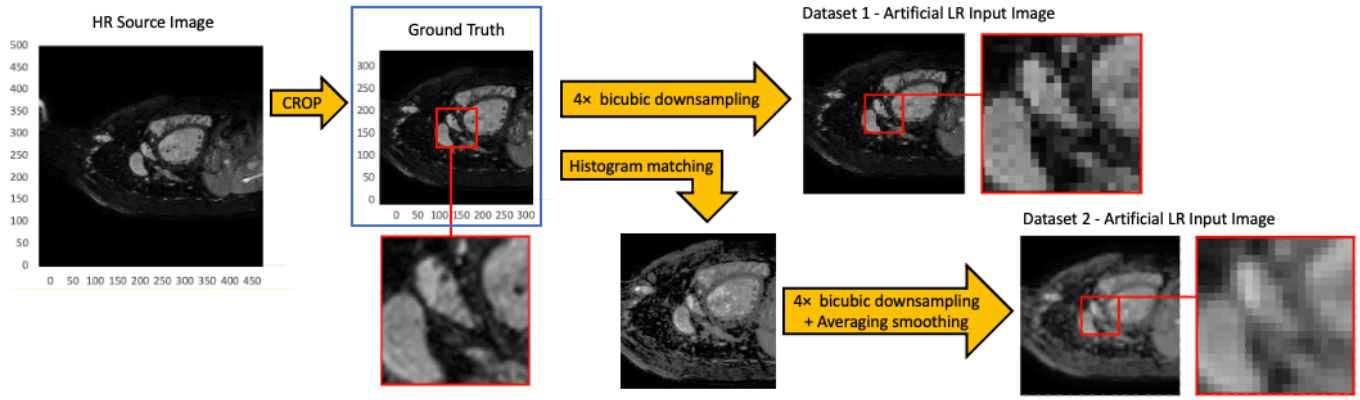

Figure 4 – The method used to create both datasets from original uncropped HR source image slices. To create the HR ground truth dataset, each image was cropped to a scale of 0.7 from the centre. From the ground truth images, two artificial LR datasets were created. In Dataset 1 this was done by 4× bicubically downsampling each image. In Dataset 2, each image was histogram matched with a real LR reference image, before applying 4× bicubic downsampling with 8 × 8 kernel averaging smoothing. Both datasets comprised 806 ground truth/LR slice pairs in the training dataset, and 225 pairs in the testing dataset.

× 142 mm (left-right), resolution: 3 × 3 × 3 mm, temporal resolution: 50–55 ms, 20 phases (Ph) in a cardiac cycle.

The HR MRI data was obtained from an open dataset of scans acquired during clinical practice at Boston Children's Hospital, MA. It included a variety of congenital heart defects, with some subjects having undergone interventions. Imaging was performed in the axial view, perpendicular to the long axis of the body, on a Phillips Achieva at 1.5 T without contrast agent, during free breathing. ECG and respiratory-navigator gating were used to remove artifact. Image dimension and resolution varied across subjects, with an average of 390 × 390 × 165 mm and 0.9 × 0.9 × 0.85 mm, respectively [28].

The HR data was chosen as it contained scans of subjects suffering from several congenital defects, enabling variation in the training data. The data was provided in the NIfTI format [16], a common medical imaging format capable of storing both 3D spatial and directional blood-flow velocity data. VDSR and SRGAN are both SISR algorithms and required single image inputs. Therefore, each NIfTI file was converted into a set of PNG images, each one storing a single sagittal plane greyscale slice of the scan. To reduce capture time, 3D MRI scans often have a higher resolution in one of the planes, therefore scans in other planes were not included in the study.

Scans in the HR dataset contained a large field of view in the sagittal plane, which included the neck and lower torso. These regions were not pertinent to the study, and their inclusion would have led to longer training times and could have detracted from the upsampling accuracy in the cardiac region. Therefore, each HR image was cropped to a scale of 0.7 of its original width and height from the centre, as seen in Fig. 4.

To create the first artificial LR dataset, the cropped ground truth images were each 4× bicubically downsampled to create corresponding LR images, with no smoothing applied. This is a common way to create artificial LR data for deep learning. However, the downsampled images aren't representative of how a LR MRI scan would appear in real-life, as fine details like blood vessels, appearing as white spots on a scan, aren't always removed. Instead, they are simply reduced to grainy dots in the LR image that an SR network can learn to restore.

For the second dataset, the goal was to create artificial LR images which resembled real LR scans. To achieve this, the first step was to histogram match the greyscale intensity values and frequencies of the HR image slices with those of a sample image from the real LR reference dataset. Histogram matching transforms a given image so that its pixel intensity histogram, which is a representation of the total number of pixels of each greyscale intensity between 0 (black) and 255 (white) [23], more closely resembles the histogram of a reference image [17]. It essentially aims to change the grey values of an image to match those of the reference image. In MRI, darker regions represent soft tissue, and lighter regions represent calcification.

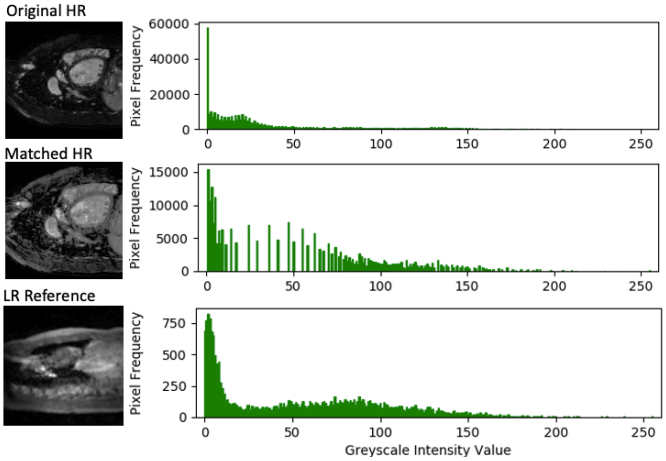

Figure 5 - Images and their corresponding histogram distributions on the right. Top left, an example of a cropped ground truth HR image from the training dataset. Middle left, the same HR image after histogram matching with the reference image. Bottom left, a real LR sagittal image slice used as the reference image.

It can be seen in Fig. 5 that nearly all pixels in the original HR image have an intensity below 50, indicating a relatively dark image. There are some brighter regions, but very few pixels have an intensity above 150. The real LR reference image at the bottom of the figure has a greater distribution of pixel intensities. The dark background means there are still many low

intensity pixels, but there is a more distributed spread between 0 and 200, with some very bright outlying pixels. The central image shows the HR image after having been histogram matched with the reference image. It has a much more even pixel intensity spread between 0 and 200, with some very bright regions, similar to the reference image. Differences in resolution and image composition mean the pixel frequency axes are different, but what's important is the relative number of each grey value within an image.

After matching, the HR images contained grainy noise which had to be removed during downsampling. In existing SR studies, a blurring filter is often applied to remove fine details present in the ground truth, as well as any noise created during downsampling that would not be present in real LR images. [7]. Various smoothing filters were tested, including Gaussian, used in the original SRGAN paper [7] and bicubic, and using varying kernel sizes (a small matrix which according to its size and arrangement creates a particular smoothing effect when convoluted with an image [19]). It was found that single-pass (only applied once) averaging smoothing using an 8x8 kernel size, followed by 4× bicubic downsampling created an artificial dataset that most closely resembled the reference dataset, as shown in Fig. 6. An example of a ground truth image and LR images from both datasets can be seen in Fig. 4.

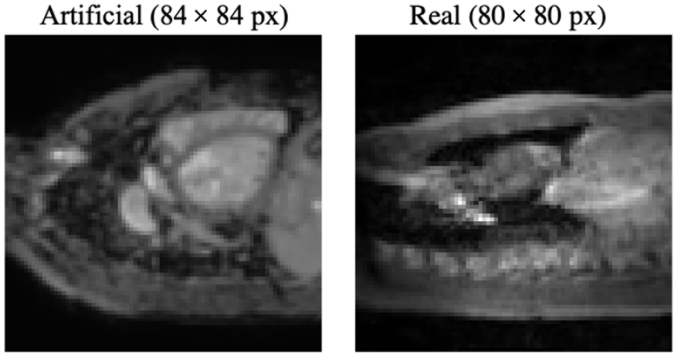

Figure 6 - Left, example of an artificial LR image from Dataset 2, after being 4× downsampled, histogram matched and 8x8 kernel averaging smoothed. Right, real LR MRI reference slice.

B. Network Configuration and Training

The VDSR network was obtained from a script [20] on the MathWorks website, and the SRGAN was based on a TensorFlow implementation [25] of the original paper [7]. Each network was subsequently modified to be trained on greyscale MRI slices, rather than the RGB colour images they were originally designed for. Information on these modifications is presented in appendix A.

The VDSR network was trained for a total of 100 epochs using stochastic gradient descent with momentum optimisation. The initial learning rate was 0.1 and decreased by a factor of 10 every 10 epochs. Since the network has 20 layers, training was time-consuming. Choosing a high learning rate accelerates this process, and it has been demonstrated [12] that an initial learning rate of 0.1 converges most quickly. However, large learning rates can lead to exponential gradient divergence, preventing the network from training successfully. This is solved by specifying a maximum gradient threshold of 0.01, which clips gradients that have become too large. The network was trained using the minibatch method, with a large batch size of 64. This means that the average error for 64 images is taken, and the weights are updated once based on this error. A large batch size reduces how often the weights are updated and speeds up training [22].

The SRGAN was trained for 200 epochs, comprising a 100-epoch generator initialisation phase optimising only MSE loss, followed by a further 100 epochs once the discriminator had been introduced, optimising for image realness. The initialisation stage is crucial, as the MSE between the SR and ground truth images must already be small for the discriminator to work effectively. An initial learning rate of 0.0001 was chosen, which was smaller than that of VDSR, but decreased less frequently, by a factor of 10 every 50 epochs. SRGAN also used minibatch training, with a smaller batch size of 16 since the 16-layer ResNet in the generator is easier to train than the 20-layer VDSR. Both networks were trained on identical versions of the two datasets discussed in *Methodology*. *B*. The extra 100 training epochs of SRGAN enabled the discriminator to have a noticeable effect on the output. As shown in Fig. 7, VDSR trains quickly, approaching convergence within the first 10 training epochs, so an extra 100 epochs would not have led to a significant improvement in its output.

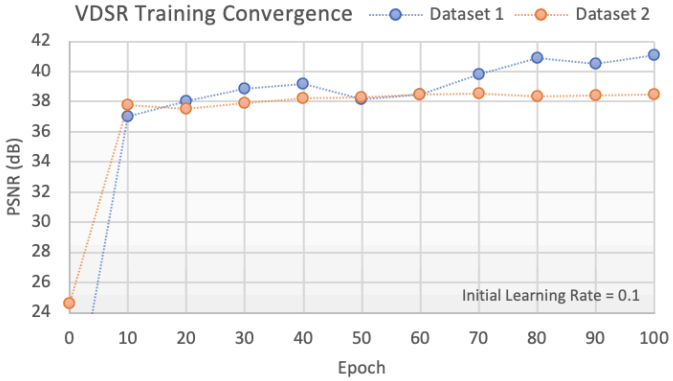

Figure 7 - VDSR PSNR performance for both datasets. After 10 epochs the network is already approaching convergence.

SISR training is usually performed using a specialized CUDA-capable NVIDIA™ Graphics Processing Unit (GPU). Using a GPU speeds up training significantly, and CUDA is software developed by NVIDIA™ to further accelerate GPU neural network training. Due to lack of access to such a machine, the training of both networks was performed using a Central Processing Unit (CPU), on an Intel i7-8700, 64GB machine. CPUs are located directly on the circuit board and are normally less powerful than GPUs.

Once trained, the two networks were evaluated on a testing set of 225 ground truth/LR slice pairs, comprising all slices of scan 17. These were kept separate from the training dataset, as testing directly on the training data yields superior results to what can be achieved on real-world input data. This is why PSNR scores in Fig. 7 are greater than those achieved in testing. Results for two of the testing slices can be seen in Figs. 9 and 10. These slices were displayed as they include a diverse range of cardiac structures, as labelled in Fig. 8.

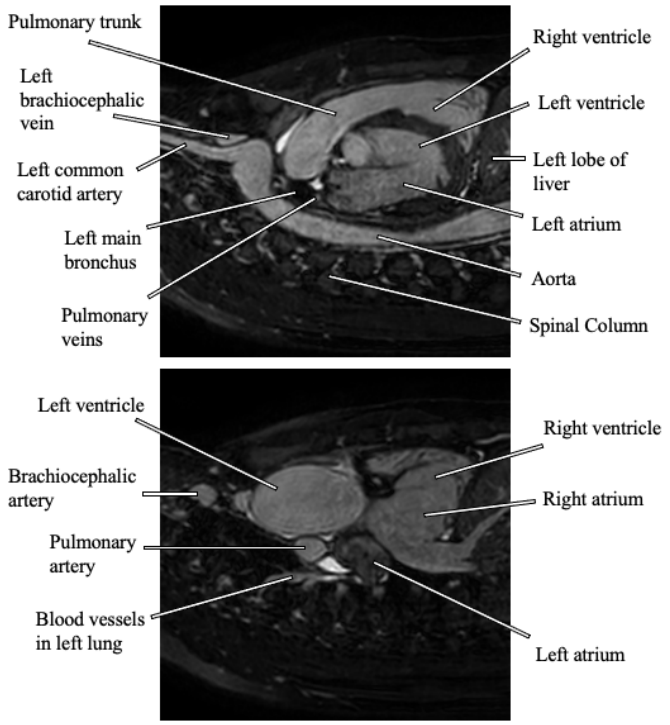

Figure 8 – Labelled versions of the sagittal MRI testing slices shown in Figs. 9 and 10.

Results were compared separately for both datasets using side-by-side comparison, and through calculation of the mean and standard deviation of PSNR and SSIM scores across the testing dataset for VDSR, SRGAN and bicubically interpolated SR images, allowing each method to be analysed regarding its utility as input to a ventricular segmentation program. Statistical significance of results was determined using the Wilcoxon signed-rank test [33], with a type I critical error threshold of 0.05. This test was selected as samples were paired (performed on the same testing dataset), and normal distribution of scores could not be assumed.

IV. Results and Discussion

Mean PSNR and SSIM scores achieved by each SR method for both datasets ± standard deviations are displayed in Table. 1. Results of Wilcoxon signed-rank significance tests are displayed in Table 2. Directional differences were compared between sample means, μ, of PSNR and SSIM scores for both datasets. For all tests, sample size, N = number of testing slices since all value pairs had non-zero difference. The null hypothesis (equal population means) was rejected in all cases for a type I error threshold of 0.05.

	Dataset 1 (No smoothing)	Dataset 2 (Matching + smoothing)
Mean PSNR		
Bicubic	30.48 ± 1.44	18.11 ± 1.90
VDSR	31.17 ± 1.41	25.72 ± 2.69
SRGAN	28.94 ± 2.04	N/A
Mean SSIM		
Bicubic	0.835 ± 0.002	0.592 ± 0.058
VDSR	0.852 ± 0.020	0.696 ± 0.053
SRGAN	0.831 ± 0.024	N/A

Table 1 – PSNR and SSIM testing scores, displayed as sample means ± standard deviations for the 225 testing slices.

1-tailed Wilcoxon signed-rank test *P* values				
	N	Alternative Hypothesis	PSNR	SSIM
Dataset 1	225	$\mu_{VDSR} > \mu_{Bicubic}$	$P<0.001$	$P<0.001$
	225	$\mu_{VDSR} > \mu_{SRGAN}$	$P<0.001$	$P<0.001$
	225	$\mu_{Bicubic} > \mu_{SRGAN}$	$P<0.001$	$P<0.001$
Dataset 2	225	$\mu_{VDSR} > \mu_{Bicubic}$	$P<0.001$	$P<0.001$

Table 2 – Results of Wilcoxon signed-rank significance tests. N = sample size = number of testing slices. μ = sample means.

A. Dataset 1 - 4× Bicubic Downsampling

Representative SR output testing images from Dataset 1 are shown in Fig. 9. Both networks generated noticeably sharper SR images than bicubic interpolation. On inspection, SRGAN was more effective in restoring fine details and sharp edges, which is particularly visible around the left brachiocephalic vein highlighted in green. All three PSNR scores were similar for this slice, with the highest being the 30.07 dB achieved by VDSR. A higher PSNR score is desirable, as indicated in (2). Across the whole testing dataset, mean PSNR scores were 31.17 dB for VDSR, 28.94 dB for SRGAN and 30.48 dB for bicubic interpolation. VDSR outperformed the other two methods for PSNR scores (P<0.001), beating bicubic interpolation on all slices, and SRGAN on 98.7% of slices (222/225). Despite blurry edge restoration and lack of detail, bicubic interpolation achieved a similar PSNR score to the two deep learning networks, and even outperformed SRGAN (P<0.001), performing better on 85.3% of slices (192/225).

This highlights the primary weakness of optimising for PSNR, which is slightly biased towards oversmoothed results. This is especially true when the LR and ground truth are similar, as is the case for the 4× bicubically downsampled images where pixel intensities in the LR images are taken directly from the HR image. Large areas of similar pixel intensity in MRI, such as the background, the aorta and the pulmonary trunk intensify this bias, as there are few high frequency details in these regions to be lost during downsampling. It should be noted that these scores are competitive with other studies [7] despite the limited number of training epochs performed. This is in part because MRI image slices are often lower resolution and contain less detail than the image datasets these networks are designed to be run on, such the benchmark DIV2K dataset which contains 900 diverse HR colour images [29].

Conversely, to obtain a high SSIM score, high-frequency (less blurry) edges and details must be accurately restored. For the slice in Fig. 9, VDSR and SRGAN obtained scores of 0.843 and 0.841, respectively, greater than the 0.822 achieved with bicubic interpolation. Visually, both deep learning networks demonstrated sharper edge restoration than bicubic interpolation, this can be seen in the restoration of the left common carotid artery and the left atrium, highlighted in green and yellow, respectively. Across the whole testing dataset, mean SSIM scores were 0.852 for VDSR, 0.831 for SRGAN and 0.835 for bicubic interpolation. VDSR achieved a greater SSIM score than both SRGAN and bicubic interpolation for all 225 slices (P<0.001). Bicubic interpolation narrowly outperformed SRGAN, achieving a greater SSIM score on 55% of slices (125/225). SRGAN generated the most visually

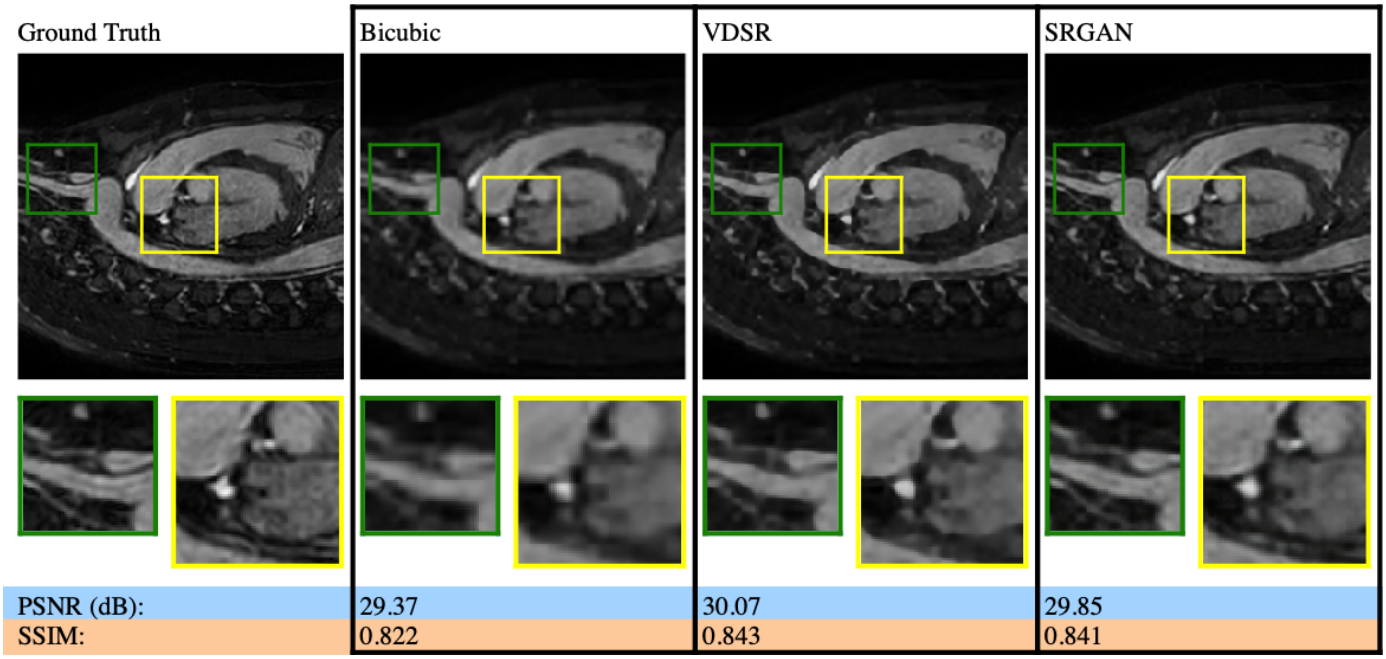

Figure 9 – Testing results for Dataset 1 (4× Bicubically Downsampled). A HR ground truth testing slice, alongside super-resolved outputs of the slice, for bicubic interpolation, VDSR and SRGAN networks. Enlarged sections of each image can be seen in the box of the corresponding colour shown beneath, and the respective PSNR and SSIM scores are displayed at the bottom of the figure for comparison.

satisfying output, recreating realistic high-frequency details (visible in the area below the left atrium highlighted in yellow). However, its output in fact differed the most from the ground truth. Other studies confirm these findings. SR networks optimising MSE almost always generate images with higher PSNR and SSIM scores than SRGANs [10].

Neither network recreated structures varying significantly from the ground truth in shape or area. However, this could be partly due to the limited number of training epochs performed for SRGAN, as longer training often intensifies detail hallucination. The superior PSNR and SSIM scores achieved by VDSR suggest that outputs generated by this network were most similar both pixel-wise and structurally to the ground truth. This indicates that VDSR most effectively restored the original forms of the ground truth, and it is likely that scans generated by this network would be the most useful for accurate determination of ventricular wall thickness and volume.

B. Dataset 2 - 4× Bicubic Downsampling with Histogram Matching and 8x8 Kernel Averaging Smoothing

Fig. 10 shows SR output slices for Dataset 2. Here, the merits of a deep learning approach are more evident. It is important to restore clear demarcations between cardiac boundaries to facilitate the ventricular segmentation task. In the second dataset, smoothing and histogram matching made these boundaries difficult to discern in the artificial LR images, reducing the greyscale gradient between cardiac features and the darker regions which separated them in the ground truth. Bicubic interpolation was incapable of restoring the grey levels and detail of the ground truth, as it only had access to the information given in the heavily modified and smoothed LR image. VDSR and SRGAN were much more successful in this task, as they had already encountered similar ground truth images during training.

Fig. 10 shows that SRGAN was more effective than VDSR in restoring the sharp region boundaries and grey levels of the ground truth and again generated the most visually convincing output of the three methods. VDSR restored a large portion of the original grey levels, but its tendency to produce soft and unrealistic looking edges is more pronounced here than in the previous dataset. Many subtle features were not restored or were difficult to identify in the resolved image.

One risk of VDSR is that poorly defined edges could lead to adjacent cardiac regions being merged together by an identification program. In the yellow box in Fig. 10, the small vessel below the left ventricle appears in the ground truth as a bright form clearly separated from other structures. However, in the VDSR SR image this distinction in brightness remained unrestored, with no clear separation between this vessel and the left ventricular wall, contrary to the SRGAN image.

Across the whole testing dataset, mean PSNR and SSIM scores were 25.72 dB and 0.696, respectively for VDSR and 18.11 dB and 0.592, respectively for bicubic interpolation. VDSR outperformed bicubic interpolation for both measures (P<0.001), achieving a superior PSNR score on 91.1% of slices (205/225), and a superior SSIM score on 88.9% of slices (200/225). In this dataset, SRGAN also appeared to largely outperform bicubic interpolation for both measures but achieved lower scores for both measures compared to VDSR.

The difference in scores between the two networks appeared to be more pronounced in Dataset 2. This highlights the problem with optimising for image realness, as when more features must be restored, SRGAN's output can diverge further from reality. SRGAN has noticeably changed the shape and area of the brachiocephalic artery cross-section (highlighted in green), while VDSR has been more effective in maintaining the

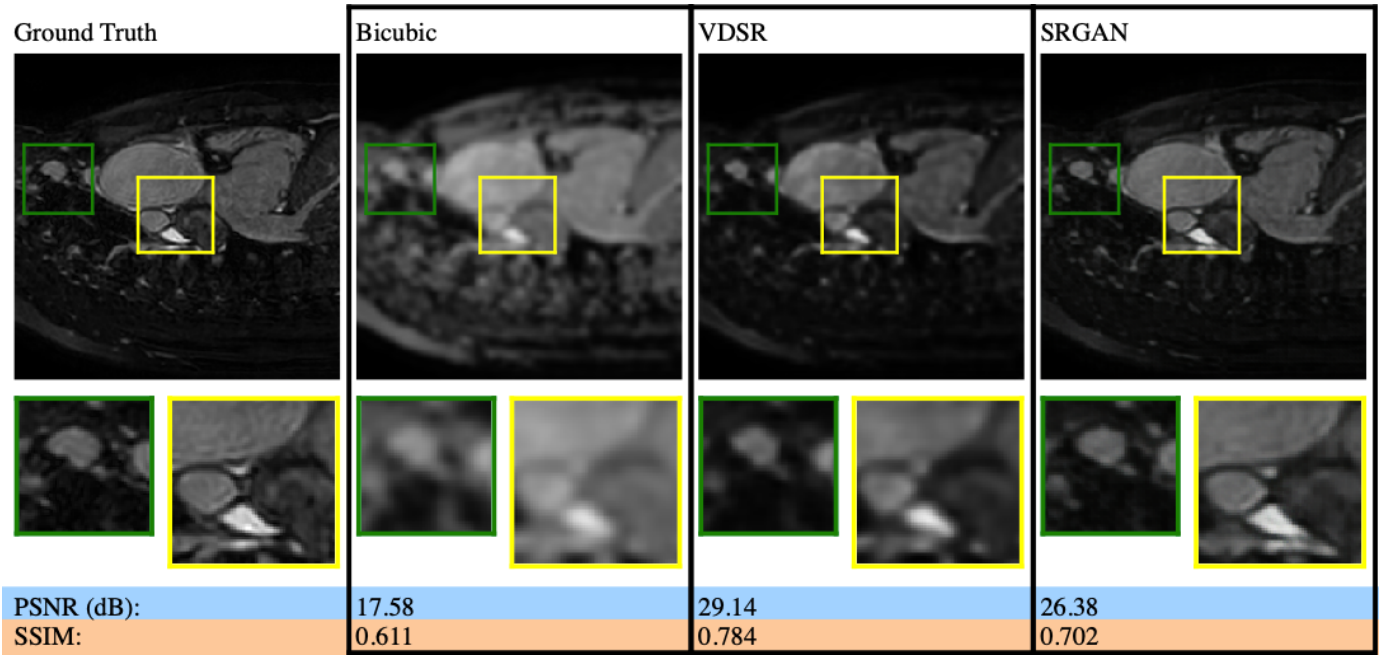

Figure 10 - Testing results for Dataset 2 (4× Bicubically Downsampled Dataset with histogram matching and 8x8 kernel averaging smoothing). A HR ground truth testing slice, alongside super-resolved outputs of the slice, for bicubic interpolation, VDSR and SRGAN networks. Enlarged sections of each image can be seen in the box of the corresponding colour shown beneath, and the respective PSNR and SSIM scores are displayed at the bottom of the figure for comparison.

original form, despite poorer detail restoration. To provide good input for a ventricular segmentation program, conserving tissue shape is more important than adding fine detail. For example, in left ventricular hypertrophy the myocardium thickens due to the increased muscle workload in the ventricle wall [26], and it is essential that this thickness is not altered during SR.

This study had some limitations. Network performance was assessed using artificially downsampled images, which can never be identical to real LR scans. Also, the training dataset had restricted breadth, as testing images were sourced from a single scan.

Note: Due to time constraints and issues loading pre-trained weights in SRGAN, it was not possible to obtain SRGAN resolved images nor mean scores for Dataset 2 through the testing script. Instead, an output was achieved by placing the slice used in Fig. 10 into the training dataset. Since SRGAN had already encountered this image during training, results and detail restoration will be slightly better than would be expected in reality.

V. Next Steps

To further validate the results of this study, SRGAN should be trained to several thousand epochs against the discriminator before in order to reach its full potential. While this could take weeks on a CPU, it could be managed in several hours were the experiment to be repeated on a CUDA-capable NVIDIA™ GPU machine. The subjects in this study were children with congenital heart defects, but the networks could equally be trained and tested on subjects with other hereditary or acquired defects. The experiment could also be repeated in the axial plane, which is commonly used in cardiac disease diagnosis as it gives a clear view of the structures within the heart. A MOS study could be conducted to determine the perceived quality of VDSR and SRGAN-resolved MRI images, as well as a study to discover the effectiveness of ventricular identification and volume estimation programs when provided with super-resolved VDSR and SRGAN cardiac images versus real HR MRI images.

Identification programs often make use of the velocity data provided in modern 4D-flow MRI scans to identify the positions and volumes of arteries and chambers within the heart, but accurate ventricle segmentation from blood flow data is made difficult by low contrast between blood and the myocardium in LR scans [32]. Therefore, it would be useful to apply these SISR methods to the velocity component of these scans, which can be upsampled in a similar way to the spatial data, helping remove noise and add finer detail.

VI. Conclusion

This study has demonstrated the feasibility of deep learning based super-resolution methods to improve the quality of low-resolution Cardiac MRI scans. SRGAN was more effective than VDSR and bicubic interpolation in restoring clear cardiac structural demarcation, which facilitates the ventricular segmentation task. SRGAN also generated the most visually convincing output for both datasets.

For Dataset 1 (4× bicubically downsampled), VDSR outperformed SRGAN and bicubic interpolation across the whole testing dataset for measures of PSNR and SSIM (P<0.001). For Dataset 2 (4× bicubically downsampled, matched and smoothed), VDSR outperformed bicubic interpolation for measures of PSNR and SSIM (P<0.001) and equally appeared to outperform SRGAN for both measures. This suggests that for both datasets, the VDSR-generated output

was most similar to the ground truth both structurally and pixel-wise. This indicates that VDSR was most effective in restoring the original shapes and forms of the ground truth, and consequently SR output scans would have the most utility as a diagnostic tool, and for determining ventricular wall thickness and volume. However, there is a risk VDSR's poor edge restoration could lead to separate regions being grouped together by a segmentation program.

The results of this study suggest that SISR networks using MSE loss are more suitable than both GAN-based networks and interpolation methods for generating HR MRI images suitable for the ventricular segmentation task. SRGAN had impressive output realness. However, even with limited training time it was shown to noticeably alter the size and shape of cardiac forms within an MRI image, but still appears to be more effective than bicubic interpolation for SR of poorer quality LR input data.

APPENDIX

A. Preparing the Deep Learning Programs

The VDSR script [20] was originally in the form of a single '.mlx' (MathWorks Live Script) file [18] which needed to be turned into a fully-functioning MATLAB program. The helper, layers, testing and training functions were placed into separate '.m' files, and a main.m file was created to define global variables and initiate the program. VDSR was designed to take 3-layer RGB images as input, then separate them into individual R, G and B layers and extract only the intensity, since SR algorithms are only interested in pixel intensity, and not colour. These colour layers were then added back to the SR image after testing. However, MRI scans are already greyscale, so each slice was turned into a '.png' image with three identical greyscale layers so it could be accepted as input. The original program only bicubically downsampled testing and training images, so the more elaborate downsampling process used for dataset 2 was added. There was little instruction on how to make SRGAN work, or the location of certain datasets and files necessary for it to run correctly. As such, a lot of time went into getting it to train on the standard DIV2K dataset. Once this was done, the program was again modified to accept greyscale images, and the MRI training and testing datasets for this study were placed into the relevant folders.

ACKNOWLEDGMENT

F. U. Author thanks Dr Lian Gan for his help and guidance throughout this project.

REFERENCES

[1] Masutani, E., Bahrami, N., Hsiao, A., 2020. Deep learning single-frame and multiframe super-resolution for cardiac MRI. *Radiology*, 295(3), pp.552-561.

[2] Jinjin, G., Haoming, C., Haoyu, C., Xiaoxing, Y., Ren, J.,Chao, D., 2020. PIPAL: A large-scale image quality assessment dataset for perceptual image restoration. 10.1007/978-3-030-58621-8_37.

[3] Chen, Y., Shi, F., Christodoulou, A., Zhou, Z., Xie, Y., Li, D., 2018. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. 10.1007/978-3-030-00928-1_11.

[4] Wang, Z., Chen, J., Hoi, S., 2020. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1-1.

[5] Fadnavis, S., 2014. Image interpolation techniques in digital image processing: An overview. International Journal of Engineering Research and Application, 4, pp.2248-962270.

[6] Dong, C., Loy, C., He, K., Tang, X., 2014. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 38. 10.1109/TPAMI.2015.2439281.

[7] Ledig, C. et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. pp.105-114. 10.1109/CVPR.2017.19.

[8] Wang, X. et al. 2018. ESRGAN: Enhanced super-resolution generative adversarial networks.

[9] Riddhi, R., Hardik, V., Sapna, K., 2014. Analysis of single frame super resolution methods, International Journal of Engineering Development and Research, 2(1), pp.152-155.

[10] Kim, J., Lee, J., Lee, K., 2016. Accurate image super-resolution using very deep convolutional networks, pp.1646-1654, 10.1109/CVPR.2016.182.

[11] Shi, W. et al. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.

[12] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, pp.770-778. 10.1109/CVPR.2016.90.

[13] Ahn, N., Kang, B., Sohn, K., 2018. Fast, accurate, and lightweight super-resolution with cascading residual network, 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part X. 10.1007/978-3-030-01249-6_16.

[14] Jianxin, W., 2017. Introduction to convolutional neural networks.

[15] Rinck, P., 2018. Magnetic resonance in medicine: A critical introduction. 12th ed. BoD.

[16] Cox, R., 2021. Official definition of the nifti1 header. [online] Available at: <https://nifti.nimh.nih.gov/pub/dist/src/niftilib/nifti1.h> [Accessed 15 March 2021].

[17] Gonzalez, R., Woods, R., 2008. *Digital Image Processing (3rd ed.)*. Prentice Hall. p. 128. ISBN 9780131687288.

[18] Webb C.R., Domijan M., 2019. Writing script files: An introduction to MATLAB® for biologists. Learning Materials in Biosciences. Springer, Cham. 10.1007/978-3-030-21337-4_4

[19] Ludwig, J., 2021. Image convolution. [pdf] Portland State University. Available at: <http://web.pdx.edu/~jduh/courses/Archive/geog481w07/Students/Ludwig_ImageConvolution.pdf> [Accessed 12 March 2021].

[20] mathworks.com. 2021. Single image super-resolution using deep learning- MATLAB & simulink example. [online] Available at: <https://uk.mathworks.com/help/images/single-image-super-resolution-using-deep-learning.html> [Accessed 4 February 2021].

[21] Kishore, R., & Kaur, T., 2012. Backpropagation algorithm: an artificial neural network approach for pattern recognition. International Journal of Scientific & Engineering Research, 3(6), 6-9.

[22] Kim, P., 2017. *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*. Berkeley, CA: Apress.

[23] Jia, W., Zhang, H., He, X., & Wu, Q., 2006. A comparison on histogram-based image matching methods. 2006 IEEE International Conference on Video and Signal Based Surveillance, pp. 97-97. IEEE.

[24] Duan, J. et al. 2019. Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. IEEE Trans Med Imaging, pp.2151-2164. 10.1109/TMI.2019.2894322.

[25] Dong, H. et al. 2017. Tensorlayer: a versatile library for efficient deep learning development. In Proceedings of the 25th ACM international conference on Multimedia, pp. 1201-1204.

[26] Lorell, BH., Carabello, BA., 2000. Left ventricular hypertrophy: pathogenesis, detection, and prognosis. Circulation. 102(4):470-9, 10.1161/01.cir.102.4.470. PMID: 10908222.

[27] Gbinigie, H., Coats, L., Parikh, J., Hollingsworth, K., Gan, L., 2021. A 4D-flow cardiovascular magnetic resonance study of flow asymmetry and haemodynamic quantity correlations in the pulmonary artery. Physiological Measurement. 42. 10.1088/1361-6579/abdf3b.

[28] Pace, D., Dalca, A., Geva, T., Powell, A., Moghari, M., Golland, P., 2015. Interactive whole-heart segmentation in congenital heart disease, Medical Image Computing and Computer Assisted Interventions, Lecture Notes in Computer Science; 9351:80-88.

[29] Agustsson, E., Timofte, R., 2017. NTIRE 2017 challenge on single image super-resolution: Dataset and study. IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1122-1131, 10.1109/CVPRW.2017.150.

[30] Petitjean, C. et al. 2014. Right ventricle segmentation from cardiac MRI: A collation study. Medical Image Analysis. 10.1016/j.media.2014.10.004.

[31] Liu, J., Chen, F., Wang, X., Liao, H., 2019. An edge enhanced SRGAN for MRI super resolution in slice-selection direction. 10.1007/978-3-030-33226-6_2.

[32] Gupta, V., Bustamante, M., Fredriksson, A., Carlhäll, C., Ebbers, T., 2017. Improving left ventricular segmentation in four-dimensional flow MRI using intramodality image registration for cardiac blood flow analysis. Magnetic Resonance in Medicine. 79. 10.1002/mrm.26674.

[33] WILCOXON, F.. (1946. Individual comparisons of grouped data by ranking methods. Journal of economic entomology. 39. 269. 10.1093/jee/39.2.269.