

Unsupervised Learning for Generalised Super-Resolution of 3D Anisotropic Medical Images via Domain Transfer

Michele Pascale

Primary Supervisor:

Dr. Jennifer Steeden

Secondary Supervisors:

Professor Vivek Muthurangu

Professor Owen Arthurs

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Master of Philosophy
of
University College London.

Department of Medical Physics and Biomedical Engineering
University College London

I, Michele Pascale, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Impact Statement

This thesis presents CLADE (Cycle Loss Augmented Degradation Enhancement), a completely unsupervised, deep learning approach for super-resolution reconstruction of anisotropic 3D medical images. By incorporating a cycle-consistent gradient mapping loss and weight demodulation process, CLADE enables resolution enhancement of MRI and CT images without requiring paired training data. This work contributes to ongoing efforts in medical imaging research with potential applications in clinical diagnostics and public health.

CLADE addresses certain limitations in existing super-resolution techniques for medical imaging. Specifically, traditional methods often rely on paired datasets that are difficult to obtain in clinical settings, or perhaps even physically impossible to acquire in settings such as abdominal or motile anatomical studies. Furthermore, existing approaches applied to natural images can produce high-resolution images, but at the expense of artifacts and inconsistencies with the risk of hallucination. This unsupervised framework that learns resolution mapping from disjoint patches offers an alternative approach to machine learning-based medical image enhancement. This methodology may be applicable to multimodal imaging studies and different imaging modalities, with opportunities to improve the generalizability of the model through more diverse datasets. The methodological aspects of this work may also contribute to applications in other areas of Deep Learning. The unsupervised learning approach could be adapted for other domains that require resolution enhancement, such as satellite imaging, microscopy, and industrial applications.

With respect to clinical applicability, CLADE shows potential for clinical applications in radiology and diagnostic imaging. By improving the resolution of

MRI and CT scans, it has the potential to enhance image-based diagnostics, in turn leading to better detection of pathologies such as tumors, lesions, or vascular abnormalities. This could be particularly useful in resource-limited healthcare settings, such as the NHS, where rapid high-resolution imaging equipment and protocols are less accessible. As a potential pre-processing tool in radiology workflows, CLADE could help improve the quality of lower-resolution scans, potentially reducing the need for repeated imaging and decreasing patient exposure to radiation in CT procedures. The resolution enhancement capabilities might also benefit image-guided interventions and surgical planning. In clinical trials and biomedical research, improved image quality could contribute to more reliable datasets.

The integration of CLADE into healthcare systems could offer economic benefits through the more efficient use of imaging infrastructure. Hospitals can improve their existing imaging capabilities without necessarily investing in new hardware and the contribution to further patient discomfort or burden. Additionally, if CLADE contributes to earlier disease detection, there could be long-term public health benefits through reduced costs associated with late-stage treatment. In the context of remote diagnostics, which are becoming increasingly more prevalent, CLADE could help improve the quality of images from portable or low-power devices, potentially supporting more accurate remote consultations in underserved regions. This research therefore contributes to the field of medical image reconstruction with a focus on developing a proof-of-concept methodology that can be extended to develop practical tools that have potential to improve healthcare imaging capabilities across different imaging centres and modalities.

Abstract

Three-dimensional (3D) imaging plays a crucial role in various medical applications, particularly in fields such as radiology and neurology, where detailed anatomical insights are crucial. However, in many clinical settings, anisotropic 3D volumes—characterized by uneven spatial resolution across different axes—are commonly acquired to balance scan quality and acquisition time. These anisotropic datasets often result from thick slices with low spatial resolution, which, while reducing scan time, can pose challenges for accurate analysis and interpretation. Deep learning (DL) offers a solution to recover high-resolution features through super-resolution reconstruction (SRR). Unfortunately, paired training data is unavailable in many 3D medical applications and therefore a novel unpaired approach is proposed; CLADE (Cycle Loss Augmented Degradation Enhancement). CLADE uses a modified CycleGAN-based architecture with a cycle-consistent gradient mapping loss and weight demodulation process. This approach is trained in an unsupervised fashion to learn SRR of the low-resolution dimension, from disjoint patches of the high-resolution plane within the anisotropic 3D volume data itself via resolution domain transfer. The feasibility of CLADE in abdominal Magnetic Resonance Imaging (MRI) and abdominal Computed Tomography (CT) imaging is demonstrated, with improvements in CLADE image quality over low-resolution volumes, conventional Cycle-GAN and state-of-the-art self-supervised SRR; SMORE (Synthetic Multi-Orientation Resolution Enhancement).

Acknowledgements

I would like to express my heartfelt gratitude to the following people, without whom this thesis would not have been possible. First and foremost, I am deeply grateful to my family for their unwavering support throughout the course of my MPhil. They have stood by me through all the highs and lows, and their encouragement has been a constant source of strength. Secondly, but most importantly, I wish to extend my deepest thanks to my primary supervisor, Jenny. There are no words to fully capture the extent of my gratitude - you are the reason I am here today with this thesis in hand. Your unrelenting support and the comfort you provided me through not only the challenges of this academic journey but also the personal difficulties I have faced has been profoundly impactful. I will forever be indebted to your kindness, guidance, and unwavering belief in me. To my secondary supervisor, Vivek, thank you for your wisdom and guidance, particularly during times when things didn't go as planned. Though we may have had our disagreements about the superiority of LaTeX formatting over the years, your insights and contribution to my scientific knowledge-base have proven invaluable. Both you and Jenny have shaped me into the scientist and researcher I am today, and the lessons I've learned from you will stay with me always. I would also like to thank Professor Alejandro Frangi and Dr. Nishant Ravikumar, who first introduced me to the world of machine learning and medical imaging during my master's degree. This journey has been an incredible one, and I will always look back on it fondly. Thank you all.

Publications & Abstracts

1. CLADE: Cycle Loss Augmented Degradation Enhancement for Unpaired Super-Resolution of Anisotropic Medical Images, arXiv pre-print, Pascale et al. [1]
2. Unpaired Super-Resolution for Anisotropic MRI using a Gradient Mapping Adversarial Loss, Digital Poster presented at ISMRM 2023 in Toronto, Pascale et al. [2]
3. CycleGAN-based Degradation Enhancement for Unpaired Super-Resolution of Anisotropic Medical Images, Digital Poster presented at ISMRM 2024 in Singapore, Pascale et al. [3]

Contents

1	Introduction	20
1.1	3D Medical Imaging	20
1.1.1	Magnetic Resonance Imaging	20
1.1.2	Computed Tomography	22
1.1.3	Anisotropic Medical Imaging	23
1.1.4	Clinical Needs & Requirements	24
1.2	Image Super-Resolution	25
1.2.1	Problem Formulation and Notation	26
1.3	Machine Learning Background	27
1.3.1	Neurons, Layers, and Activation Functions	28
1.3.2	Convolutional Neural Networks	28
1.3.3	Convolutional Kernels	29
1.3.4	Gradient Descent	30
1.3.5	Backpropagation	30
1.3.6	Supervised vs. Unsupervised Learning	31
1.3.7	Generative Adversarial Networks (GANs)	32
1.3.8	CycleGAN	32
1.4	Machine Learning for Super-Resolution	35
1.4.1	Supervised SRR Approaches	35
1.4.2	Unsupervised Approaches	40
1.5	Hypothesis	43

2	Methodology	45
2.1	Overview	45
2.2	CLADE Architecture	45
2.2.1	Generator Networks	46
2.2.2	Discriminator Networks	47
2.3	Gradient Mapping Cycle-Consistency Loss	48
2.4	Patched Training and Reconstruction	49
2.5	Imaging Datasets	50
2.5.1	MRI Dataset	50
2.5.2	CT Dataset	50
2.6	Experimental Network Setup	51
2.6.1	CLADE Network Optimisation	51
2.6.2	Assessing the Importance of the Weight Demodulation and Gradient Mapping Loss	52
2.6.3	Assessing Supervised vs. Unsupervised Training	52
2.6.4	Comparison with State-of-the-art SRR Method: SMORE	54
2.7	Image Quality Metrics	54
2.7.1	Ground Truth Image Pairs: PSNR, SSIM, MSE and MAE	54
2.7.2	No Ground Truth: PIQUE, NIQE and BRISQUE	56
2.8	Quantitative Image Quality Assessment	56
2.8.1	Statistical Analysis Methodology	57
3	Results	58
3.1	CLADE Network Optimisation	58
3.1.1	Loss Weighting Optimisation	58
3.1.2	Epoch Optimisation	63
3.1.3	Patch Reconstruction Stride Length	68
3.2	Importance of Weight Demodulation	70
3.2.1	MRI weight demodulation	70
3.2.2	CT weight demodulation	71
3.3	Supervised Training	71

<i>Contents</i>	<i>10</i>
3.3.1 MRI Supervised Networks	72
3.3.2 CT Supervised Networks	76
3.4 Unsupervised Training	81
3.4.1 MRI Unsupervised Networks	81
3.4.2 CT Unsupervised Networks	85
3.5 Quantitative Image Quality	88
4 Discussion	93
5 Conclusion	95
Bibliography	98

List of Figures

1.1	A 3D representation of an image of a skull, outlining the differences between pixels, voxels and slice thickness. Image used under fair use from [5].	21
1.2	A 2D slice reconstruction of the brain (b) as reconstructed using an Inverse Fourier Transform from its corresponding k-space acquisition (a). Image used under Creative Commons Licensing from [6]. .	22
1.3	An example of the CT acquisition strategy (image derived from [7] under fair use).	23
1.4	An example of an image from a 3D anisotropic MRI acquisition. In this example, two directions have high spatial resolution ($\approx 1\text{mm}$), while the third direction (shown in both the Sagittal and Transverse view) has lower spatial resolution ($\approx 6\text{mm}$).	24
1.5	A simple illustration of a 2D Convolutional Kernel showing the calculation of a feature map from the element-wise product of the input tensor with the kernel. Image used under fair use from [13].	30
1.6	Constituents of the CycleGAN architecture	34
1.7	3D-DCSRN Architecture (image derived under fair use from [18]). .	36
1.8	ResCNN Architecture (image derived under Creative Commons Licensing from [19]).	36
1.9	ArSRR Architecture (image derived under Creative Commons Licensing from [20]).	37
1.10	SRGAN Generator and Discriminator architecture (image derived under fair use from [21]).	38

1.11	G-GAINSR Architecture (image derived funder Creative Commons Licensing from [24]).	38
1.12	SRFeat Generator Architecture (image derived under fair use from [25]).	39
1.13	SRFeat Discriminator Architecture (image derived under fair use from [25]).	39
1.14	CinCGAN architecture, where G_1 , G_2 and G_3 are generators and SR is a pre-trained EDSR network. D_1 and D_2 are discriminators. The G_1 , G_2 and D_1 compose the first LR \rightarrow clean LR CycleGAN model, mapping the degraded LR images to clean LR images. The G_1 , SR, G_3 and D_2 compose the second LR \rightarrow HR CycleGAN model, mapping the LR images to HR images (image and caption derived under fair use from [26]).	40
1.15	The network architecture of UASSR, which consists of (A) disentangled resolution translation, (B) self reconstruction and (C) cross-cycle reconstruction. Specifically, the inputs l and h are unpaired LR and HR images. Different colours show different domains, i.e., yellow represents the LR domain, green represents the HR domain and white shows the shared content space (image and caption derived under fair use from [30]).	42
1.16	High-level outline of the SMORE super-resolution pipeline. The pipeline consists a PSF estimation framework (A), synthetic data generation and training of WDSR network (B) and inference pipeline (C) (image derived under fair use from [32]).	43
2.1	CLADE generator architecture shown for $G_X : X \rightarrow Y$. ModConv denotes the 2D modified convolution layers. Each modified convolutional layer performs the weight-demodulation process.	46
2.2	CLADE PatchGAN discriminator architecture.	48

- 3.1 Image quality from a MRI test subject, using CLADE (with \mathcal{L}_{gmap}) across the 10 models with the lowest PIQUE scores. Blue boxes show magnified regions; red box highlights the selected final loss weighting. 60
- 3.2 Image quality from a CT test subject, using CLADE (with \mathcal{L}_{gmap}) across the 10 models with the lowest PIQUE scores. Blue boxes show magnified regions; red box highlights the selected final loss weighting. 62
- 3.3 Example image quality from one subject in the MRI test dataset, at each epoch when training the final MRI CLADE (with \mathcal{L}_{gmap}) ($\lambda_{cyc} = 1, \lambda_{ident} = 1, \lambda_{cyc} = 5$). Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the MRI test dataset. The red box indicates the chosen epoch based on PIQUE scores (see Table 3.3) as well as visual image quality. 65
- 3.4 Example image quality from one subject in the CT test dataset, at each epoch when training the final CT CLADE (with \mathcal{L}_{gmap}) ($\lambda_{cyc} = 1, \lambda_{ident} = 1, \lambda_{cyc} = 10$). Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the CT test dataset. The red box indicates the chosen epoch based on PIQUE scores (see Table 3.4) as well as visual image quality. 67
- 3.5 Example image quality from one subject in the MRI test dataset, at each stride length, using the final MRI CLADE during inference. Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the MRI test dataset. The red box indicates the chosen patch size based on PIQUE score, inference time and visual image quality . . . 69

- 3.6 Example image quality from one subject in the CT test dataset, at each stride length, using the final CT CLADE (with \mathcal{L}_{gmap}) during inference. Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the CT test dataset. The red box indicates the chosen patch size based on PIQUE score, inference time and visual image quality. 69
- 3.7 An example low-resolution sagittal MRI slice from one subject in the test data, which shows normalisation errors when SRR is applied using a Conventional CycleGAN. These artifacts are removed when SRR is applied using CLADE (no \mathcal{L}_{gmap}). Magnified regions within the blue box are displayed beneath each image. 70
- 3.8 The conventional CycleGAN suffered from mode collapse when trained using the CT data. The top row shows the patched image reconstruction from the model at epochs 1-5, on one of the test CT volumes, with a stride of 6 (as used for CLADE). The bottom row shows that where there are no overlapping patches (i.e. stride length is equal to the patch size), then all patches remain identical, demonstrating mode collapse. 71
- 3.9 MRI paired networks, applied to paired data (coronal). Although there are some improvements to the resolution, the networks fail to faithfully reconstruct the original ground-truth high resolution images. It can clearly be seen that although resolution is marginally improved, the downsampling artifacts and degradations still remain present in the reconstructions. 73

- 3.10 MRI paired networks, applied to unpaired data (sagittal). In contrast to the paired data case, the networks fail to perform any reasonable form of super-resolution across all of the networks. This gives further evidence that the simple downsampling method used to generate paired data doesn't faithfully represent the complex degradations present in the true low-resolution data as pictured here. 75
- 3.11 CT Paired Network, applied to paired data (in transverse orientation). Here, we can see that the resolution is increased and the reconstructed images do begin to sharpen quite significantly. However, we can see that there are many more artifacts present, alongside some potential hallucinations from the network architectures, particularly in CLADE (no \mathcal{L}_{gmap}). 78
- 3.12 CT paired networks, applied to paired data (sagittal). Here, we see that marginal super-resolution is performed by the networks, however there is some clear hallucination in the case of CLADE (no \mathcal{L}_{gmap}), likely as a result of the inefficiency of the downsampling method when generating the synthetic data. 80
- 3.13 An example of sagittal image quality from SRR applied to a low-resolution MRI data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the MRI data. Magnified regions within the blue box are displayed beneath each image. 83
- 3.14 An example of transverse image quality from SRR applied to a low-resolution MRI data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the MRI data, where this figure shows the resulting volume reformatted in the (low-resolution) transverse plane. Magnified regions within the blue box are displayed beneath each image. 84

3.15	An example of sagittal image quality from SRR applied to a low-resolution CT data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the CT data. Magnified regions within the blue box are displayed beneath each image.	86
3.16	An example of coronal image quality from SRR applied to a low-resolution CT data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the CT data, where this figure shows the resulting volume reformatted in the (low-resolution) coronal plane. Magnified regions within the blue box are displayed beneath each image.	87
3.17	Box-plot for MRI Metrics, demonstrating the superiority of CLADE throughout all other comparisons.	89
3.18	Box-plot for CT Metrics, demonstrating the superiority of CLADE throughout all other comparisons.	90

List of Tables

3.1	PIQUE scores from MRI CLADE (with \mathcal{L}_{gmap}) hyper-parameter optimisation, shown in ascending PIQUE order.	59
3.2	PIQUE scores from CT CLADE (with \mathcal{L}_{gmap}) hyper-parameter optimisation, shown in ascending PIQUE order.	61
3.3	PIQUE scores ($\mu \pm \sigma$) as calculated across all 15 volumes in the MRI test dataset, assessed with a stride length of 6. The scores are compared across different epochs for the final MRI CLADE (with \mathcal{L}_{gmap}) model ($\lambda_{cyc} = 1, \lambda_{ident} = 1, \lambda_{cyc} = 5$). The chosen number of epochs is outlined in bold, based on PIQUE score and visual image quality (see Figure 3.3).	64
3.4	PIQUE scores ($\mu \pm \sigma$) as calculated across all 15 volumes in the CT test dataset, assessed with a stride length of 6. The scores are compared across different epochs for the final CT CLADE (with \mathcal{L}_{gmap}) model ($\lambda_{cyc} = 1, \lambda_{ident} = 1, \lambda_{cyc} = 10$). The chosen number of epochs is outlined in bold, based on PIQUE score and visual image quality (see Figure 3.4).	66
3.5	Impact of stride length on PIQUE scores and inference time across 15 MRI test volumes for the final MRI CLADE model ($\lambda_{cyc} = 1, \lambda_{ident} = 1, \lambda_{gmap} = 5, 9$ epochs)	68
3.6	Impact of stride length on PIQUE scores and inference time across 15 CT test volumes for the final CLADE model ($\lambda_{cyc} = 1, \lambda_{ident} = 1, \lambda_{gmap} = 10, 10$ epochs).	68

3.7	SSIM, PSNR, MSE and MAE scores ($\mu \pm \sigma$) for the MRI paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.	72
3.8	PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the MRI paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.	72
3.9	PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the MRI paired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes.	74
3.10	SSIM, PSNR, MSE and MAE scores ($\mu \pm \sigma$) for the CT paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.	76
3.11	PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the CT paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.	77
3.12	PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the CT paired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse. Furthermore, SMORE is not included due to its inability to be trained in a paired fashion.	79
3.13	PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the MRI unpaired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes.	82
3.14	PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the CT unpaired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse.	85

3.15	PIQUE scores ($\mu \pm \sigma$) as calculated across all slices in the CT and MRI test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse.	91
3.16	Quantitative edge sharpness and estimated SNR ($\mu \pm \sigma$) as calculated across all slices in the CT and MRI test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse.	91

Chapter 1

Introduction

1.1 3D Medical Imaging

Three-dimensional (3D) Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) have become integral components of medical diagnosis, facilitating a comprehensive evaluation of anatomy. These imaging modalities enable the visualisation of internal structures in great detail, providing critical insights that aid in the diagnosis, treatment planning, and monitoring of various medical conditions.

1.1.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) operates by utilising radio-frequency (RF) pulses and magnetic gradients to generate detailed images of internal body structures. RF pulses are applied to the body within a strong magnetic field, causing the alignment of hydrogen atoms in tissues to temporarily shift away from their alignment with the main magnetic field. As these atoms return to their original alignment, they emit signals that are detected by the MRI detector coils. The process of acquiring MRI images is inherently slow because each data point in so-called *k-space* is collected sequentially, often resulting in long acquisition times. Numerous advancements have emerged to expedite the imaging processes, including trajectory optimisation, under-sampling, and partial Fourier acquisitions [4]. These accelerated acquisition techniques necessitate sophisticated image reconstruction algorithms to faithfully recreate the acquired images. Despite the use of these methods for accelerating acquisition, simple 3D anisotropic methods remain advantageous due to

their efficiency in capturing relevant anatomical information, while reducing acquisition time and maintaining image quality in two of the spatial dimensions. These anisotropic methods are particularly valuable due to their applicability to simple, widely accessible “out-of-the-box” sequences, with fast and simple reconstructions.

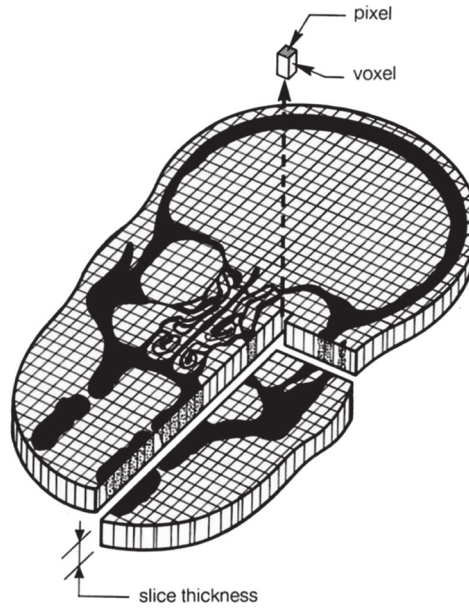


Figure 1.1: A 3D representation of an image of a skull, outlining the differences between pixels, voxels and slice thickness. Image used under fair use from [5].

K-space is a critical concept in MRI, referring to the data acquisition domain where spatial frequencies are collected. The centre of *k*-space contains information about the overall contrast and structure of the image, while the edges of *k*-space contain data related to the finer details and resolution. In 3D MRI, voxels (3D pixels) in *k*-space are sampled across the x , y , and z axes and a Fourier transform is then applied (along the k_x , k_y , and k_z axis) to convert the data into the 3D spatial image domain, revealing the anatomical structures of the scanned object (see Figure 1.2).

The sequential nature of the acquisition of MRI data can be time-consuming, particularly for high-resolution 3D imaging, leading to prolonged scan times and potential distortions due to physiological or voluntary motion, such as breathing.

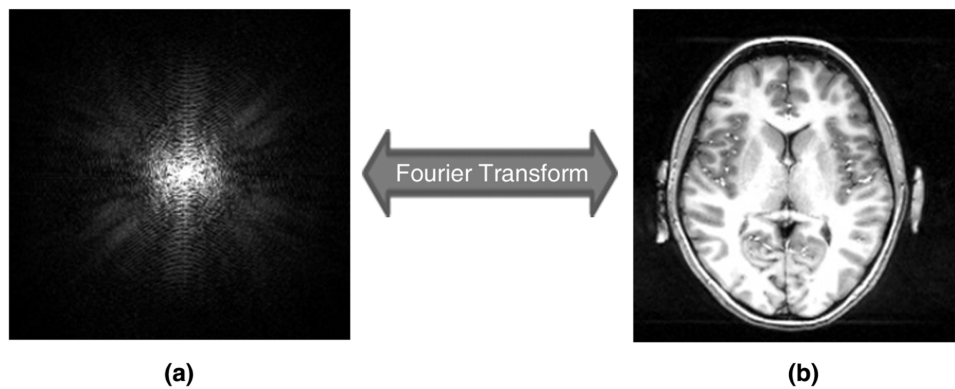


Figure 1.2: A 2D slice reconstruction of the brain (b) as reconstructed using an Inverse Fourier Transform from its corresponding k-space acquisition (a). Image used under Creative Commons Licensing from [6].

1.1.2 Computed Tomography

Computed Tomography (CT) imaging is another popular diagnostic tool within medical imaging, providing detailed cross-sectional images of the internal structures of the human body. The principle underlying CT imaging involves the attenuation of X-rays as they pass through the body, as different tissues within the body attenuate X-rays to varying degrees, depending on their density and composition. Dense structures such as bones attenuate more X-rays, appearing as bright areas on the resulting images, while less dense tissues like muscles and organs attenuate fewer X-rays, appearing darker. During a CT scan, an X-ray source rotates around the patient, emitting a narrow beam of X-rays. Detectors opposite the source measure the intensity of the X-rays after they have traversed through the body. By collecting a series of 2D image slices from multiple angles, these images are then processed to reconstruct a 3D representation of the imaged anatomy (see Figure 1.3).

Despite the popularity of CT imaging due to its simplicity and relative speed, it results in substantially higher radiation exposure than conventional X-ray procedures. Repeated exposure to such high doses of ionising radiation has been linked to an increased risk of cancer. Therefore, anisotropic 3D imaging has emerged as a widely employed approach, offering expedited acquisition times and diminished

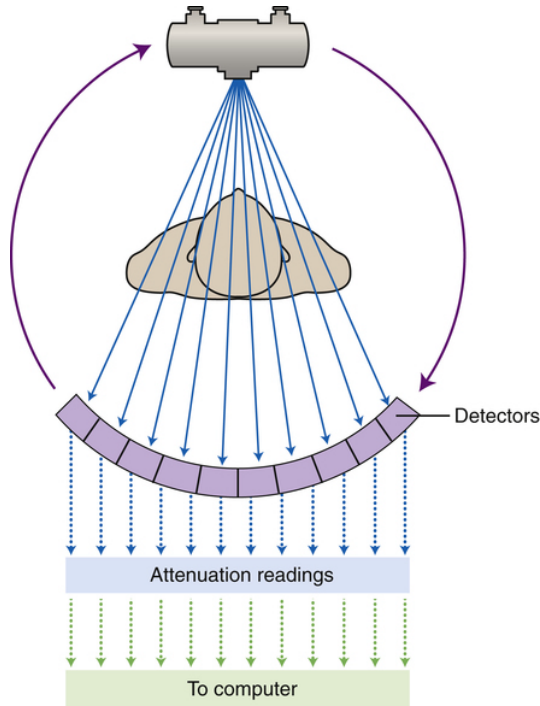


Figure 1.3: An example of the CT acquisition strategy (image derived from [7] under fair use).

radiation dosage in the context of CT imaging.

1.1.3 Anisotropic Medical Imaging

In isotropic imaging the spatial resolution is uniform in all directions, however anisotropic imaging (Figure 1.4) allows for differential resolution along distinct spatial dimensions. This variability in resolution can be influenced by factors such as imaging system parameters and acquisition geometry.

In the context of 3D MRI, anisotropic imaging is normally achieved by maintaining the spacing between the k -space lines but collecting fewer phase-encoding steps in the k_z -direction. This means that we do not collect as much high-frequency/resolution information in the k_z/z -direction, resulting in image voxels which are lower resolution in one of the 3D image directions (z). The high-resolution x - y plane is commonly referred to as in-plane, whereas the lower-resolution the z plane is referred to as through-plane. Although collecting fewer k_z lines reduces the scan time, the resulting discrepancy in resolution can present challenges in accurately representing anatomical structures, potentially compromis-

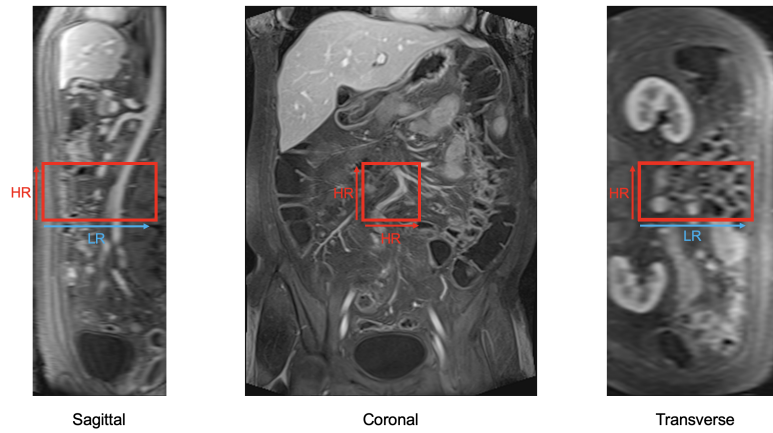


Figure 1.4: An example of an image from a 3D anisotropic MRI acquisition. In this example, two directions have high spatial resolution ($\approx 1\text{mm}$), while the third direction (shown in both the Sagittal and Transverse view) has lower spatial resolution ($\approx 6\text{mm}$).

ing diagnostic accuracy and subsequent image analysis procedures. 2D multi-slice MRI can sometimes be used as an alternative to 3D MRI, whereby 2D multi-slice MRI acquires individual slices separately, often with gaps. This leads to potential misalignment and lower through-plane resolution. In contrast, 3D MRI captures an entire volume at once, providing higher isotropic resolution and allowing for improved signal-to-noise ratio across the entire volume compared to the separate 2D slice acquisitions.

In comparison to MRI, CT achieves higher spatial resolution due to its use of X-rays, which have short wavelengths and can capture fine structural details, while MRI excels in tissue contrast by exploiting differences in proton density and relaxation times. The primary reasons for the lack of isotropy in CT imaging can be attributed to the characteristics of the X-ray beam and the reconstruction algorithms employed. X-ray beams in CT scanners are typically collimated into a thin fan or cone shape, resulting in higher resolution in the plane perpendicular to the beam's direction.

1.1.4 Clinical Needs & Requirements

For super-resolution reconstruction (SRR) to be clinically viable, it must satisfy several key requirements that address the specific needs of radiological diagnosis.

In MRI applications, a primary objective is to enhance through-plane resolution to match in-plane resolution, typically achieving isotropic voxels of approximately 1mm^3 or better. Similarly, for CT imaging, resolution enhancement must improve soft tissue differentiation while maintaining anatomical accuracy. This level of detail is critical for detecting subtle pathological features that might otherwise remain hidden in standard-resolution images, including early-stage tumors, fine vascular networks, and microstructural tissue abnormalities.

Clinical utility demands that the super-resolution method preserves essential anatomical details without introducing artifacts that could lead to misdiagnosis. Particularly important is the accurate representation of small but clinically significant structures, such as tumors, vessel wall characteristics, and nuanced textural variations in tissue. The preservation of these details directly impacts diagnostic confidence and accuracy.

Beyond image quality considerations, practical implementation requirements are equally important. The SRR method must integrate seamlessly into existing radiological workflows, maintaining compatibility with DICOM standards without necessitating significant procedural modifications. Processing speed is another critical factor, with real-time or near-real-time performance necessary to avoid introducing delays in the diagnostic process. Furthermore, an ideal super-resolution approach should demonstrate robust generalizability across diverse imaging modalities and anatomical regions. This versatility ensures broader clinical applicability without requiring extensive retraining for each new application context.

When these specifications are met, super-resolution technology can substantially enhance medical imaging capabilities, potentially improving diagnostic precision, facilitating earlier disease detection, and optimizing treatment planning—all without increasing scan duration, radiation exposure, or equipment costs.

1.2 Image Super-Resolution

There are many types of interpolation techniques which are frequently employed to mitigate the resolution disparities caused by anisotropic imaging, however often

they do not completely compensate for the loss of detail in the through-plane direction. Thus, careful consideration and optimisation of acquisition parameters are essential to minimise these concerns and ensure attainment of high-quality imaging outcomes.

To address the issues that arise with interpolation-based techniques, Super-resolution (SR) [8] can be employed. In the context of imaging, SR techniques pertain to the procedure of augmenting the resolution of an image beyond its original resolution by utilising one (or more) observations of the same scene. SR methods can be broadly described by two main approaches: Single Image Super-Resolution (SISR) and Multiple Image Super-Resolution (MISR). As indicated by its name, SISR enables the enhancement of the target image resolution, without necessitating the acquisition of any additional data. In the context of this thesis, we will focus on SISR approaches exclusively.

1.2.1 Problem Formulation and Notation

Given a set of low-resolution images $Y \subset \mathbb{R}^{m \times n}$, where each image $\mathbf{y} \in Y$, the goal of SISR is to estimate a high-resolution image $\mathbf{x} \in X$, where $X \subset \mathbb{R}^{M \times N}$; with $M > m \in \mathbb{Z}$ and $N > n \in \mathbb{Z}$.

Let $H : X \rightarrow Y$ denote the downsampling operator that maps a high-resolution image to a low-resolution image. The relationship between the low-resolution image \mathbf{y} and the high-resolution image \mathbf{x} can be modelled as:

$$\mathbf{y} = H(\mathbf{x}) + \epsilon \quad (1.1)$$

where ϵ represents noise and other distortions in the imaging process. The goal of SISR is to find an estimate $\hat{\mathbf{x}}$ of the high-resolution image \mathbf{x} given the observed low-resolution image \mathbf{y} ; this can be formulated as a typical optimisation problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{y}) \quad (1.2)$$

where $\mathcal{L}(\mathbf{x}, \mathbf{y})$ is an arbitrary loss function that penalises the discrepancy be-

tween the observed low-resolution image y and the reconstructed high-resolution image \hat{x} . The main challenge in solving the SISR problem, lies in the ill-posed nature of the inverse problem, where multiple high-resolution images can lead to the same low-resolution image.

Traditional super-resolution methods rely on interpolation techniques, such as nearest neighbour, bilinear, bi-cubic, B-spline, or Lanczos resampling. However, these interpolant-based approaches often result in “blocky” edges or blurring, due to the extrapolation of missing high-frequency information from neighbouring pixels [9]. Alternatively, traditional reconstruction methods have sought to address the ill-posed inverse problem directly, by incorporating prior knowledge or constraints into the reconstruction to regularise the solution. These traditional techniques commonly fall into two categories: Constrained Least Squares (CLS) minimisation methods, which rely on assumptions such as the smoothness of the image, and Bayesian-based maximum a posteriori (MAP) stochastic approaches, which necessitate knowledge of the posterior probability density function (PDF) of the original image [8]. Traditional approaches are grounded in well-established mathematical principles, ensuring consistent and interpretable results. However, they must be tailored on a case-by-case basis, resulting in a lack of adaptability and generalisability. Additionally, the mathematical assumptions may not hold for diverse datasets, potentially leading to reduced image quality and robustness in reconstruction. Recent advancements in Machine Learning (ML) have significantly enhanced and, in many cases, replaced traditional Super-Resolution Reconstruction (SRR) methodologies by overcoming these limitations. Nonetheless, these ML approaches have their own limitations, including the need for large datasets, challenges with network interpretability, and potential issues with robustness due to network hallucination [10].

1.3 Machine Learning Background

A neural network is a machine learning computational model, inspired by the biological neurons found in the human brain. This section offers a formal, mathematical exposition of the foundational components that constitute neural networks.

1.3.1 Neurons, Layers, and Activation Functions

A neural network is organised into layers, denoted as l , with each layer consisting of multiple neurons. Let $\mathbf{a}^{(l)}$ represent the activations of layer l , and $\mathbf{z}^{(l)}$ denote the pre-activation outputs (which form the inputs to the activation function) of the neurons in layer l . Each neuron in l has associated weights $\mathbf{W}^{(l)}$ and biases $\mathbf{b}^{(l)}$, where the pre-activation outputs $\mathbf{z}^{(l)}$ are computed as:

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)} \quad (1.3)$$

where $\mathbf{a}^{(l-1)}$ is the activation vector from the previous layer.

Subsequently, the activations $\mathbf{a}^{(l)}$ of layer l are obtained by applying an activation function f to the *pre-activation outputs* $\mathbf{z}^{(l)}$:

$$\mathbf{a}^{(l)} = f(\mathbf{z}^{(l)}) \quad (1.4)$$

Each neuron's output $\mathbf{z}^{(l)}$ is then passed through an activation function $\sigma^{(l)}(\cdot)$ to introduce non-linearity into the network:

$$\mathbf{a}^{(l)} = \sigma^{(l)}(\mathbf{z}^{(l)}) \quad (1.5)$$

where $\sigma^{(l)}(\cdot)$ is typically a non-linear function such as ReLU [11], sigmoid, or tanh. These activation functions allow the neural network to learn complex patterns in the data by transforming the weighted sum of inputs into the neuron's output.

In summary, the computation of layer l of a neural network can be expressed as:

$$\mathbf{a}^{(l)} = \sigma^{(l)}(\mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}) \quad (1.6)$$

1.3.2 Convolutional Neural Networks

The majority of ML-driven methodologies, particularly those pertaining to image-related tasks, heavily rely on the utilisation of Deep Learning (DL) techniques; specifically, Convolutional Neural Networks (CNNs) [12]. CNNs are a specialised

class of deep neural networks, designed specifically for the processing of visuospatial data, drawing inspiration from an abstracted variant of the human visual system. Fundamentally, these networks aim to simulate the capability of the human brain to identify shapes and objects through the systematic dissection of images into smaller (potentially related) entities known as features.

1.3.3 Convolutional Kernels

Convolutional kernels play a crucial role in the extraction of features from the input data in CNNs. These kernels are small matrices that are applied to the input data through convolution operations; by element-wise multiplication of the kernel values with the corresponding input values at each spatial location, followed by the summation of the results.

Formally, in the case of a 2D image, let I denote the 2D input data matrix of size $N \times M$, and C represent the convolutional kernel matrix of size $K \times L$. The resulting feature map F is computed as follows:

$$F(i, j) = \sum_{m=1}^K \sum_{n=1}^L I(i - m, j - n) \cdot C(m, n), \quad (1.7)$$

where i and j iterate over the spatial dimensions of the feature map, m and n iterate over the spatial dimensions of the convolutional kernel (see Figure 1.5).

These operations generate feature maps that effectively highlight specific patterns or features present in the input data. In a CNN, multiple kernels are simultaneously applied to the input data, enabling the extraction of diverse features. The optimised weights (and biases) for each kernel are learnt through the iterative process of training (using backpropagation 1.3.5 and gradient descent 1.3.4), based on the discrepancy between the predicted and actual outputs. As a result, CNNs are capable of autonomously acquiring hierarchical representations of features from intricate data, rendering them highly effective tools for tasks including image recognition, object detection, and classification and SRR.

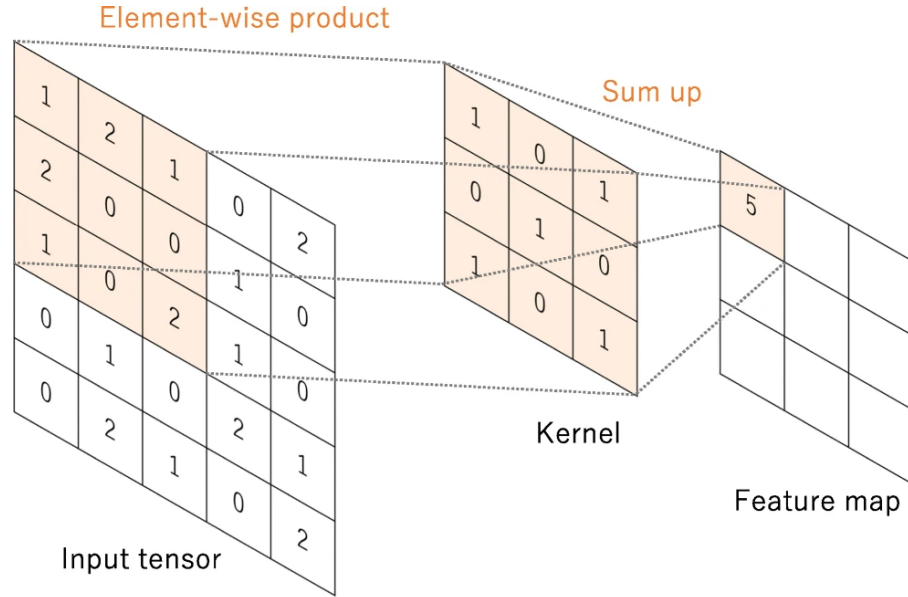


Figure 1.5: A simple illustration of a 2D Convolutional Kernel showing the calculation of a feature map from the element-wise product of the input tensor with the kernel. Image used under fair use from [13].

1.3.4 Gradient Descent

Gradient descent is an optimisation algorithm used to minimise an arbitrary loss function, $L(\theta)$, where θ represents the parameters of the model (weights and biases). Intuitively, gradient descent works by iteratively adjusting the parameters in the direction that reduces the loss the most.

Mathematically, the update rule for gradient descent is given by:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} L(\theta), \quad (1.8)$$

where η is the learning rate, a small positive number that controls the step size, and $\nabla_{\theta} L(\theta)$ is the gradient of the loss function with respect to the parameters of the model. This gradient points in the direction of the steepest ascent, so subtracting it moves the parameters in the direction of the steepest descent.

1.3.5 Backpropagation

Backpropagation is an algorithm used to efficiently compute the gradients of the loss function with respect to the parameters of the model, which is essential for gradient descent. Backpropagation utilises the chain rule of calculus to propagate

the error backward through the network layers, where the error term for each layer $\delta^{(l)}$, is calculated as:

$$\delta^{(L)} = \nabla_{\mathbf{a}^{(L)}} L \odot \sigma'(\mathbf{z}^{(L)}), \quad (1.9)$$

$$\delta^{(l)} = (\mathbf{W}^{(l+1)})^T \delta^{(l+1)} \odot \sigma'(\mathbf{z}^{(l)}), \quad (1.10)$$

where \odot denotes the element-wise multiplication, $\delta^{(L)}$ is the error term for the output layer, and σ' is the derivative of the activation function.

Finally, the gradients of the loss function, L , with respect to the weights and biases are given by:

$$\nabla_{\mathbf{W}^{(l)}} L = \delta^{(l)} (\mathbf{a}^{(l-1)})^T, \quad (1.11)$$

$$\nabla_{\mathbf{b}^{(l)}} L = \delta^{(l)}. \quad (1.12)$$

By combining these gradients with the gradient descent update rule, we can iteratively adjust the network's parameters to minimise the loss function, thereby training the neural network.

1.3.6 Supervised vs. Unsupervised Learning

When training neural networks, two primary paradigms are typically considered: supervised learning and unsupervised learning. Supervised Learning addresses the task of learning a mapping function from input-output pairs provided in a labelled dataset. Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ represent a labelled dataset, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the input features and $y_i \in \mathcal{Y}$ denotes the corresponding labels. The goal in supervised learning is to infer a function $f : \mathbb{R}^d \rightarrow \mathcal{Y}$ that maps inputs to outputs. This function is learned by minimising a loss function $\mathcal{L}(f(\mathbf{x}), y)$ over the training dataset \mathcal{D} , where \mathcal{L} quantifies the discrepancy between predicted outputs and true labels.

In contrast, Unsupervised Learning operates on datasets lacking explicit output labels, instead aiming to discover hidden patterns or structures inherent in the data. Given an unlabelled dataset $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^d$, the objective in unsupervised learning is to learn meaningful representations, clusters, or distributions from the data. Clustering (partitioning data into groups based on similarity mea-

tures), dimensionality reduction (extracting a compact representation of the data by capturing its essential characteristics in a lower-dimensional space), and density estimation (modelling the underlying distribution of the data points) are common tasks in unsupervised learning.

1.3.7 Generative Adversarial Networks (GANs)

While supervised and unsupervised learning form the foundational pillars of machine learning, recent advancements have introduced innovative techniques that transcend traditional paradigms. One such paradigm-shifting methodology is Generative Adversarial Networks (GANs), proposed by Goodfellow et al. [14]. GANs represent a framework for training generative models by pitting two neural networks against each other in a competitive setting. The core idea behind GANs involves training two neural networks concurrently: a generator (G) and a discriminator (D). The generator aims to produce synthetic data samples (\hat{y}) that closely resemble the true data distribution (Y , with elements $y \in Y$), while the discriminator endeavours to distinguish between genuine samples (y) and synthetic samples (\hat{y}). The generator and discriminator models are trained in a zero-sum, two-player minimax game with an adversarial loss function, \mathcal{L}_{adv} (Equation 1.13):

$$\min_G \max_D \mathcal{L}_{adv}(D, G) = \mathbb{E}_{y \sim p_{data}(y)} [\log D(y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1.13)$$

The networks iteratively improve their performance until an equilibrium, known as the Nash equilibrium, is reached, where the generator generates realistic samples indistinguishable from true data according to the discriminator.

1.3.8 CycleGAN

The CycleGAN [15] expands the GAN concept to enable image-to-image style transfer between two disjoint image collections X and Y (with images defined as $x \in X$ and $y \in Y$). The CycleGAN attempts to translate features from domain X to domain Y (and vice-versa) using two generators to simultaneously learn the forward and backward generator mapping functions, $G_X : X \rightarrow Y$ and $G_Y : Y \rightarrow X$,

respectively. Each generator is paired with a corresponding discriminator, D_X and D_Y , respectively, which are trained to distinguish whether an image from the opposite domain is real or generated (see Figure 1.6(a)).

The adversarial loss is calculated for the forward and backward generator mappings, as defined in Equation 1.14 and Equation 1.15, respectively.

$$\mathcal{L}_{adv}(D_Y, G_X) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(G_X(x)))] \quad (1.14)$$

$$\mathcal{L}_{adv}(D_X, G_Y) = \mathbb{E}_{x \sim p_{data}(x)}[\log D_X(x)] + \mathbb{E}_{y \sim p_{data}(y)}[\log(1 - D_X(G_Y(y)))] \quad (1.15)$$

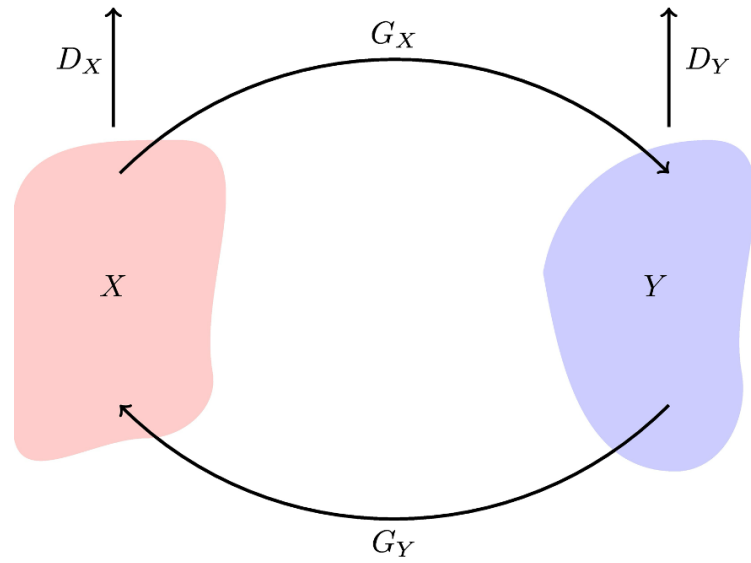
Furthermore, to constrain the space of possible mapping functions, the CycleGAN imposes cycle-consistency, which means that the mappings must be approximately bijective and therefore mapping between the two domains should be invertible. Cycle-consistency is achieved through forward and backward cycle-consistency functions, which are defined as $G_X(G_Y(y)) \approx y$ and $G_Y(G_X(x)) \approx x$, respectively (see Figure 1.6(b) and Figure 1.6(c)), where the cycle-consistency loss is defined in Equation 1.16.

$$\begin{aligned} \mathcal{L}_{cyc}(G_X, G_Y) = & \mathbb{E}_{x \sim p_{data}(x)} [\|G_Y(G_X(x)) - x\|_1] + \\ & \mathbb{E}_{y \sim p_{data}(y)} [\|G_X(G_Y(y)) - y\|_1], \end{aligned} \quad (1.16)$$

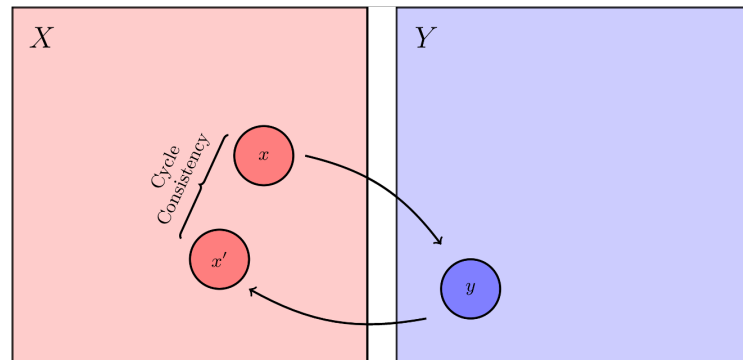
where $\|\cdot\|_1$ denotes an L_1 norm. Therefore, the overall loss function for a conventional CycleGAN, \mathcal{L}_{GAN} , consists of the weighted sum of the adversarial losses and the cycle-consistency loss, as defined in Equation 1.17.

$$\begin{aligned} \mathcal{L}_{GAN}(D_X, D_Y, G_X, G_Y) = & \mathcal{L}_{adv}(D_Y, G_X) + \\ & \mathcal{L}_{adv}(D_X, G_Y) + \\ & \lambda_{cyc} \mathcal{L}_{cyc}(G_X, G_Y), \end{aligned} \quad (1.17)$$

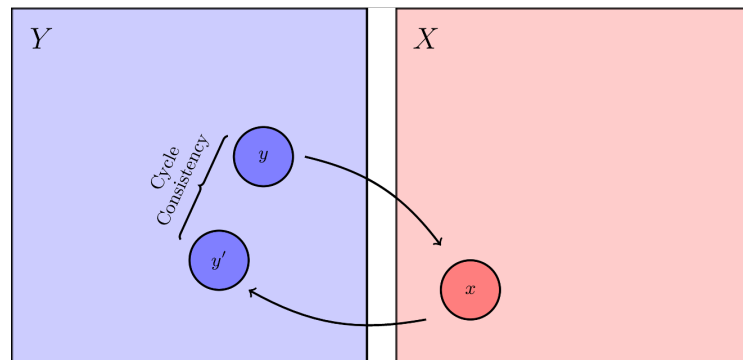
where λ_{cyc} is the weighting for the cycle-consistency loss term.



(a) Abstracted CycleGAN model, showing the bijective generator mappings and corresponding discriminators.



(b) Forward cycle-consistency loss, showing the discrepancy between the original element x and the “cycled” element x' in the direction $X \rightarrow Y \rightarrow X$.



(c) Backward cycle-consistency loss, showing the discrepancy between the original element y and the “cycled” element y' in the direction $Y \rightarrow X \rightarrow Y$.

Figure 1.6: Constituents of the CycleGAN architecture

1.4 Machine Learning for Super-Resolution

Supervised deep learning super-resolution methods leverage CNNs to learn to effectively capture and exploit intricate image features, enabling them to generate high-quality, realistic-looking images from large datasets of low-resolution and high-resolution image pairs. Compared to traditional methods, these deep learning approaches have demonstrated superior SRR performance in terms of image quality, preservation of details, and reduction of artifacts. However, often in the context of medical imaging, high-quality datasets are either unavailable or impossible to acquire (due to scan time, or physiological motion). As a result, fully supervised training methods prove difficult, and so unsupervised super-resolution techniques are often required to mitigate these limitations.

1.4.1 Supervised SRR Approaches

One of the simplest architectures to perform supervised super-resolution involves the use of a 3D Convolutional Neural Network with paired training data samples. One such approach (SR-CNN3D) [16], was originally developed to enhance the resolution of 3D MRI brain images, where paired low-resolution images were synthetically created through Gaussian blurring and downsampling of high-resolution images (by an isotropic scale factor of 2). Subsequently, 3D patches, sized $25 \times 25 \times 25$, were utilised to train the SR-CNN3D network for the low-resolution to high-resolution mapping.

Building on the ideas behind SR-CNN3D, and drawing inspiration from DenseNet [17], a widely recognised architecture for object recognition tasks, 3D Densely Connected Convolutional Neural Networks (3D-DCSRN) [18] were introduced. This architecture stands out due to its accelerated training process, lightweight nature facilitated by weight sharing, and reduced over-fitting owing to parameter reduction and feature reuse. Here, a novel architecture was introduced specifically tailored for further improving the resolution of 3D structural brain MRI images, demonstrating 4x resolution enhancement, via the use of patch-based training scheme which uses the densely-connected blocks as described in the original DenseNet architecture.

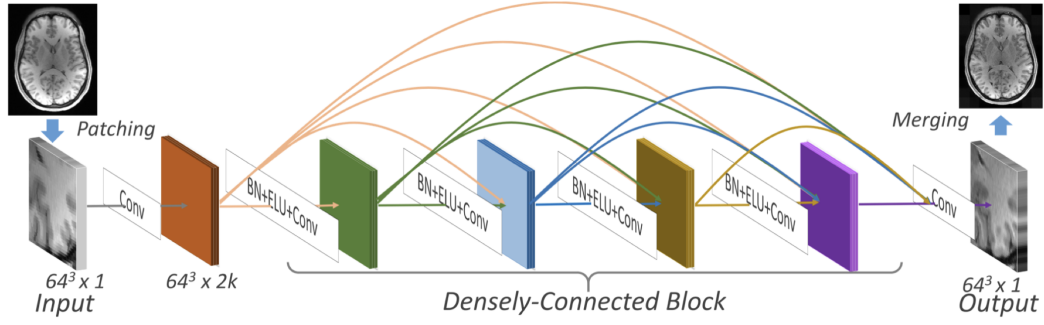


Figure 1.7: 3D-DCSRN Architecture (image derived under fair use from [18]).

In contrast to these methods, ResCNN [19] specifically focuses on super-resolving anisotropic 3D MRI images. In this work, the authors introduce a technique based around residual learning, incorporating both long and short skip connections into the network. This dual strategy not only mitigates the vanishing gradient issue, but also effectively restores high-frequency details inherent in the target MRI images.

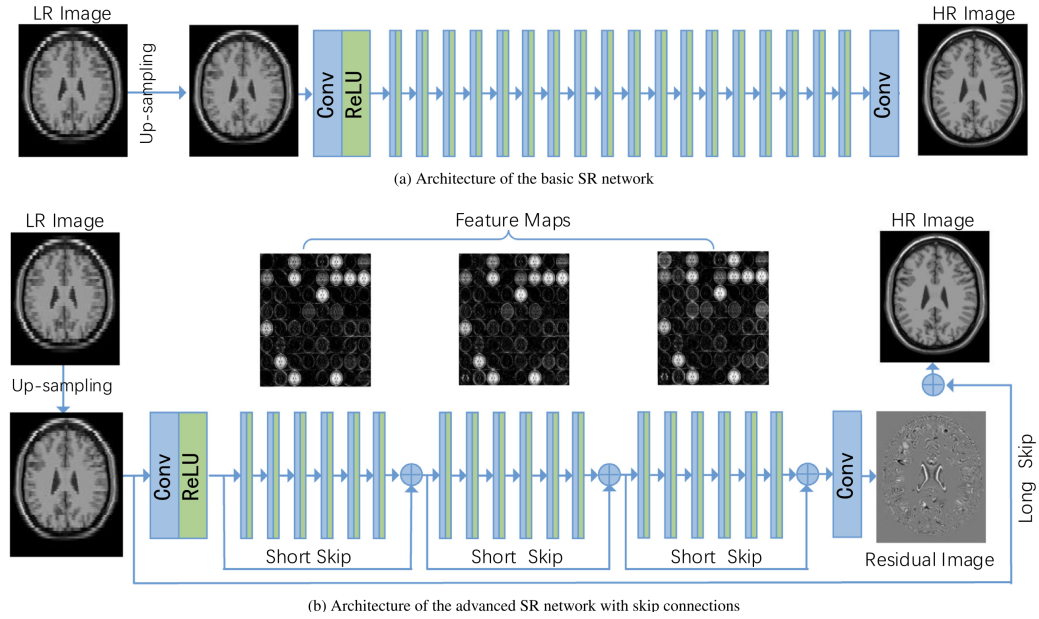


Figure 1.8: ResCNN Architecture (image derived under Creative Commons Licensing from [19]).

Furthermore, the authors optimise memory usage and computational complexity by exploiting the cross-plane self-similarity within the MRI images. Although ResCNN requires synthetic generation of low-resolution images to create paired

training data, the authors utilise 2D patches to reconstruct the 3D volumes, exploiting the inherent self-similarity within the images, which represents a departure from traditional 3D patches.

Other studies have tried to eliminate the need for training multiple networks tailored to specific resolution scales. Arbitrary Scale Super-Resolution (ArSRR) [20] adopts a continuous implicit neural function to model both low-resolution and high-resolution images, accommodating varying sampling rates. Leveraging the seamless continuity of this learned implicit function, the network theoretically allows for arbitrary super-resolution scaling from any low-resolution input image.

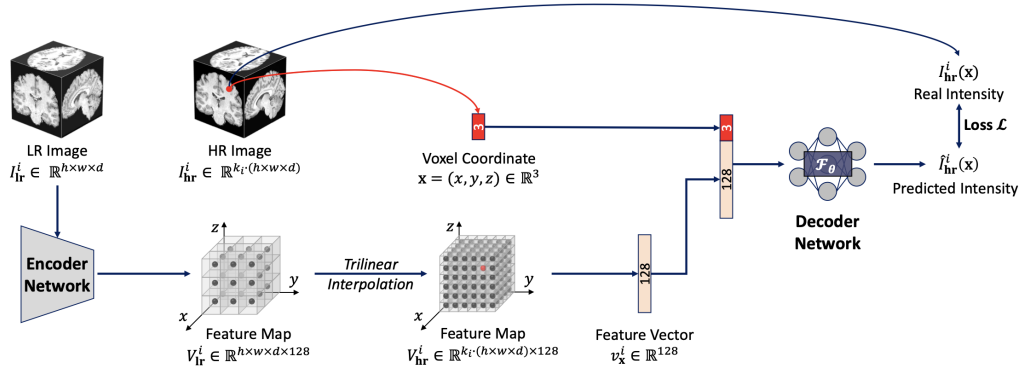


Figure 1.9: ArSRR Architecture (image derived under Creative Commons Licensing from [20]).

Despite this advancement, the ArSRR network still relies on synthetically generated paired 3D low-resolution and high-resolution image patches, created using cubic interpolation with a scale factor, k , randomly sampled from a uniform distribution $\mathcal{U}(2, 4)$.

Building on the foundation of traditional super-resolution techniques, generative adversarial networks (GANs) emerged as a powerful alternative, offering innovative approaches to enhance image resolution. Among the earliest GAN-based approaches for super-resolution, SRGAN [21] introduced a novel perceptual loss function, alongside a state-of-the-art super-resolution architecture.

The authors used a ResNet [22] architecture (specifically an SRResNet), where the content-loss is acquired via sampling of feature maps from a pre-trained VGG network [23] and by calculation of the Euclidean distance between the feature rep-

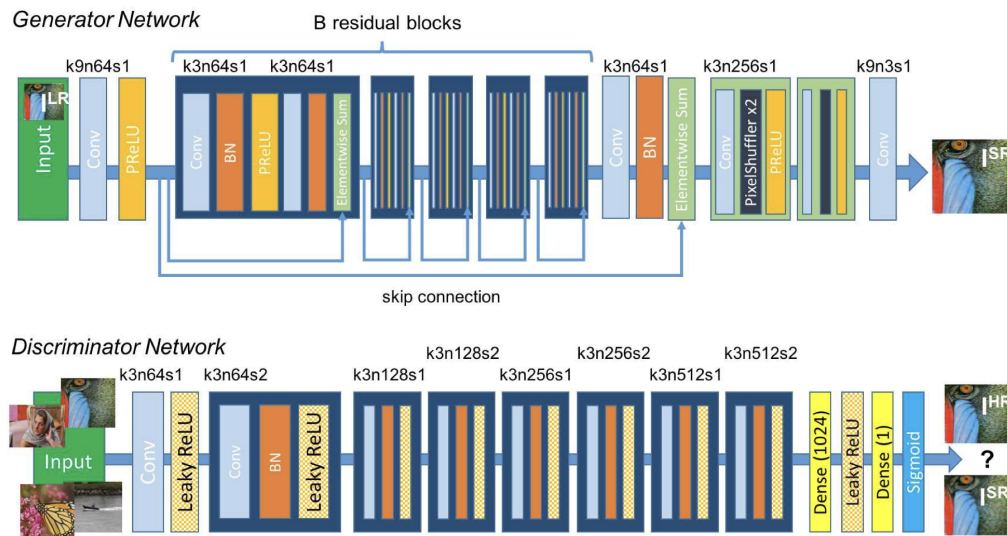


Figure 1.10: SRGAN Generator and Discriminator architecture (image derived under fair use from [21]).

representations of their reconstructed and reference images.

Another GAN-based method, G-GANISR [24], utilises a generator and discriminator trained adversarially to achieve superior super-resolution results. Its architecture combines dense and skip connections, enhancing its ability to learn complex image features.

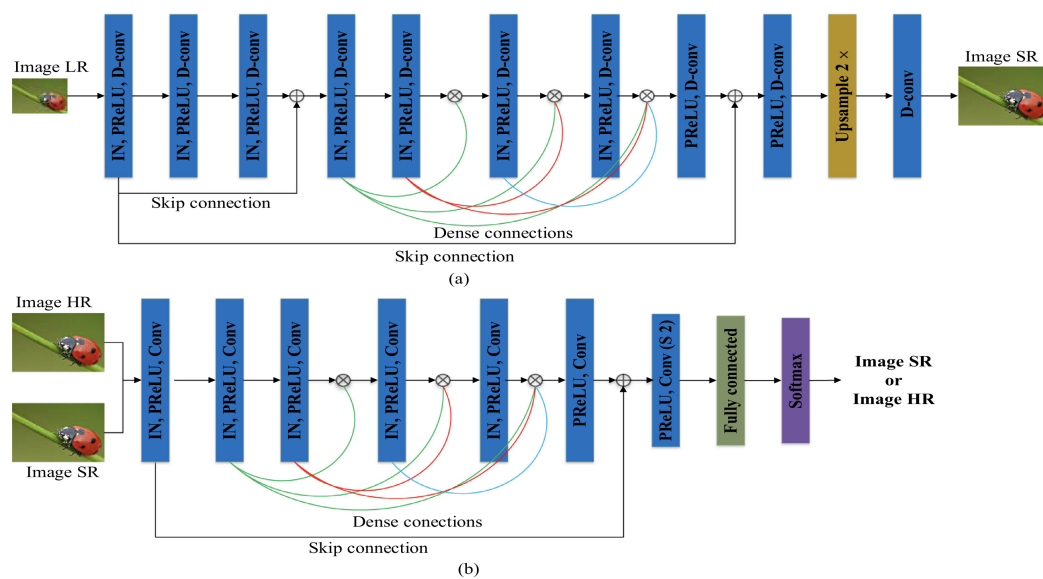


Figure 1.11: G-GAINSR Architecture (image derived funder Creative Commons Licensing from [24]).

The effectiveness of G-GANISR is showcased through its ability to achieve up to 8x super-resolution on synthetically degraded natural images (created through Gaussian filtering, followed by downsampling). Unlike conventional methods that rely on patch-based approaches, G-GANISR operates directly on full-size images, offering a streamlined approach that doesn't necessitate further pre/post-processing.

An alternative GAN-based method is Single Image Super-Resolution with Feature Discrimination (SRFeat) [25] which aims to increase the fidelity of the resulting output images, by addressing a notable drawback of GANs: their tendency to neglect high-frequency noise within input images. SRFeat integrates an extra discriminator model, which operates directly on the feature maps of the images (rather than image pixels), guiding the generator to emphasise structural high-frequency image components while suppressing unwanted noisy artifacts.

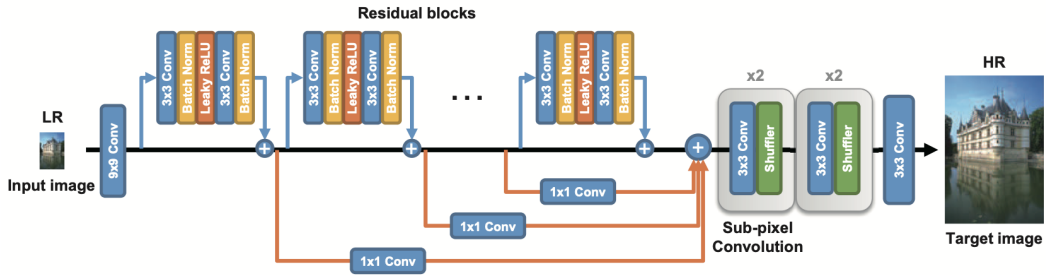


Figure 1.12: SRFeat Generator Architecture (image derived under fair use from [25]).

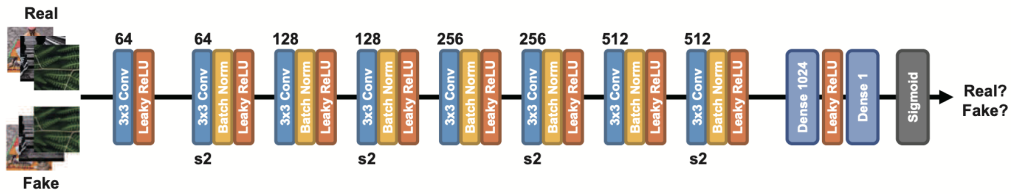


Figure 1.13: SRFeat Discriminator Architecture (image derived under fair use from [25]).

In addition, SRFeat introduced a tailored feature loss function that operates in tandem with the feature discriminator, facilitating further guidance for the generator during training. Moreover, the generator architecture is enhanced through the incorporation of both long and short skip connections, aimed at augmenting the model's proficiency in capturing and reconstructing high-frequency image features.

1.4.2 Unsupervised Approaches

Inspired by the CycleGAN [15], Cycle-in-Cycle GANs (CinCGAN) [26] were proposed for SISR on unpaired datasets of natural images. It should be noted that although this is an unsupervised approach, it relies on a pre-trained EDSR [27] network for super-resolution.

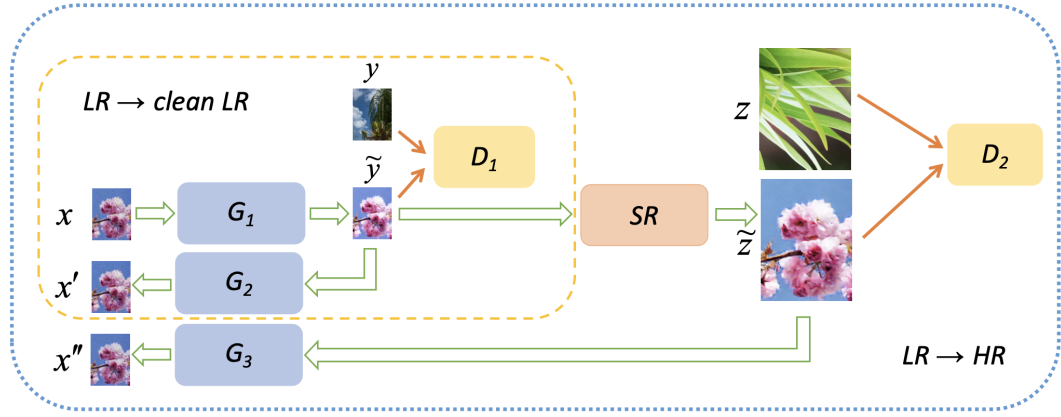


Figure 1.14: CinCGAN architecture, where G_1 , G_2 and G_3 are generators and SR is a pre-trained EDSR network. D_1 and D_2 are discriminators. The G_1 , G_2 and D_1 compose the first $LR \rightarrow \text{clean LR}$ CycleGAN model, mapping the degraded LR images to clean LR images. The G_1 , SR, G_3 and D_2 compose the second $LR \rightarrow HR$ CycleGAN model, mapping the LR images to HR images (image and caption derived under fair use from [26]).

The CinCGAN acts to super-resolve images between two disjoint sets of LR and HR images respectively. It achieves this task by taking the noisy input image and maps it to a noise-free low-resolution latent space. Then, the intermediate image is up-sampled using a pre-trained EDSR network, followed by fine-tuning during training to achieve a high-resolution output image (see Figure 1.14). It is important to note here that the authors make a distinction between “clean” LR images and standard LR images. Specifically, in the first stage of the architecture, a mapping is learned from a real-world LR image $x \in X$ to a “clean” LR image $y \in Y$. These “clean” images are derived via bi-cubic downsampling of HR images $z \in Z$, where it can be assumed that each z is noise-free. The use of a Cycle Consistency Loss (as described in Equation 1.16), permits the unsupervised fine-tuning step within the network. Furthermore, the authors include a total variation (TV) [28] loss to impose spatial smoothness during the training process.

Unsupervised arbitrary scale super-resolution reconstruction (UASSR) [29], is a truly unsupervised CycleGAN method based on disentangled representation learning. It was developed for arbitrary SRR of anisotropic 3D medical images. UASSR can be deconstructed into three main components which are jointly optimised to perform super-resolution on the disjoint sets of images (the architectural configuration is illustrated in Figure 1.15):

1. **Resolution Translation:** acts by encoding an input image into a domain-invariant content space \mathcal{C} and a domain-specific resolution space \mathcal{A} .
2. **Disentangled Representation Learning:** by training for self-reconstruction and cycle-consistent reconstruction (as described in Section 1.3.8). This uses both the content and resolution content of each image to translate between the two (as well as between itself for self-reconstruction).
3. **Structure Invariant Learning:** by preserving image content and spatial structure within images during the cross-domain translation process, via the use of a perceptual loss (which uses a pre-trained VGG16 network [23] as a feature extractor).

Another popular self-supervised super-resolution technique is Synthetic Multi-Orientation Resolution Enhancement (SMORE) [31]. SMORE utilises true anisotropic image volumes to generate synthetic (LR, HR) image pairs, using two EDSR [27] networks to remove aliasing and perform super-resolution, respectively. Building on this foundation, a newer variant of SMORE [32] has been proposed, which eliminates the need for two EDSR networks by accurately simulating LR images from true anisotropic data. This is achieved using the ESPRESO (Estimate the Slice Profile for Resolution Enhancement from a Single Image Only) [33] algorithm to estimate the Point Spread Function (PSF), which is then used to simulate the LR degradation kernel.

SMORE first estimate the PSF of the anisotropic input volume using ESPRESO (see Figure 1.16 (A)). This is used during training, to degrade 2D patches which are extracted from high-resolution data, creating paired training data

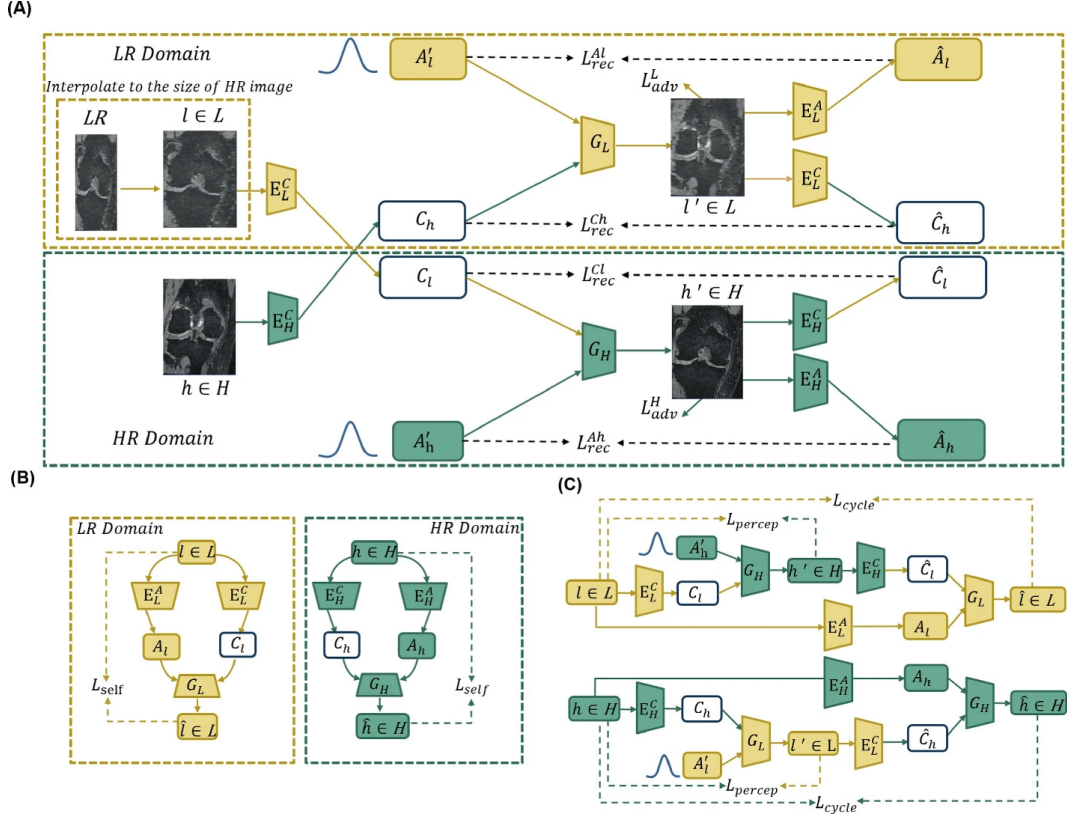


Figure 1.15: The network architecture of UASSR, which consists of (A) disentangled resolution translation, (B) self reconstruction and (C) cross-cycle reconstruction. Specifically, the inputs l and h are unpaired LR and HR images. Different colours show different domains, i.e., yellow represents the LR domain, green represents the HR domain and white shows the shared content space (image and caption derived under fair use from [30]).

for the Wide Activation Super-Resolution (WDSR) [34] network (see Figure 1.16 (B)). At inference, the low-resolution through-plane 2D slices are super-resolved using the trained WDSR, and then stacked and averaged to produce an isotropic output 3D volume (see Figure 1.16 (C)). For simplicity, any further reference to SMORE in this thesis will pertain to this newer variant.

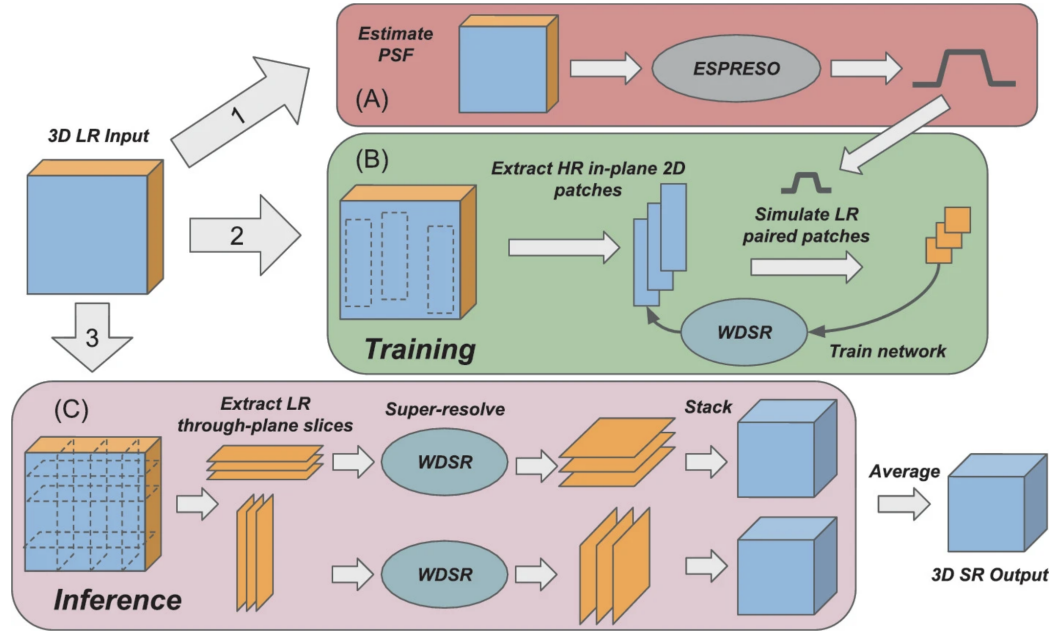


Figure 1.16: High-level outline of the SMORE super-resolution pipeline. The pipeline consists a PSF estimation framework (A), synthetic data generation and training of WDSR network (B) and inference pipeline (C) (image derived under fair use from [32]).

1.5 Hypothesis

Many of the methods described in Section 1.4 (specifically, SRGAN [21], G-GANISR [24] and SRFeat [25]) are primarily designed for the super-resolution of natural images. While the underlying network architectures and techniques are similar, their application in medical image super-resolution presents distinct challenges. Specifically, the degradation kernels inherent to various medical imaging modalities are significantly more complex and challenging to simulate, compared to those in natural images, which typically exhibit simpler and more predictable noise characteristics, allowing for straightforward assumptions and simulations. In scenarios where the degradation kernels are unknown or difficult to accurately simulate, many supervised methods do not perform optimally due to incorrect assumptions or inadequate simulation of these kernels. Simpler degradation techniques, such as Gaussian blurring or bi-cubic downsampling, as seen in Ar-SRR [20], may be insufficient for accurately representing the complex and modality-specific degradation processes present in many medical imaging modalities. Another important consideration is

that many medical imaging modalities are inherently anisotropic, meaning that they cannot physically facilitate the acquisition of truly “paired” imaging datasets and although these can be simulated, there are a lack of approaches that permit truly unsupervised super-resolution of this kind.

Our proposed methodology focuses on addressing the issue of unsupervised deep learning-based single image super-resolution reconstruction (DL-SRR) within the framework of 3D anisotropic medical images. Our approach does not rely on any predefined assumptions regarding degradation kernels or domain feature distributions. Our goal is to “implicitly” learn the degradation kernel and corresponding inverse mapping in order to produce high-resolution images from low-resolution image patches originating from separate domains. In order to achieve this, we leverage the theory of unpaired image-to-image translation via CycleGAN-based architectures by redefining our SRR problem as an image-to-image domain translation task involving separate patches within our 3D images across the LR and HR imaging planes respectively. More concretely, we propose CLADE (Cycle Loss Augmented Degradation Enhancement), an entirely unsupervised DL-SRR approach for anisotropic 3D medical images that learns directly from the anisotropic volumes themselves. This avoids the need to learn an explicit degradation kernel, simplifying the overall DL-SRR scheme. The framework relies on the observation that small two-dimensional (2D) patches extracted from a 3D volume contain similar visual features, irrespective of their orientation. This is coupled with a 2D patch-based approach that learns HR features from patches in the HR plane of the anisotropic volume, to guide DL-SRR in the LR direction in an unpaired fashion.

Chapter 2

Methodology

2.1 Overview

In this chapter we will cover:

- Proposed CLADE Network Architecture
- Proposed Gradient Mapping Cycle-Consistency Loss
- Proposed patch-based reconstruction
- Datasets
- Optimisation of the CLADE framework;
 - Loss Weightings (λ_{cyc} , λ_{ident} , λ_{gmap})
 - Stride Length
 - Number of Epochs
- Comparison of supervised and unsupervised training methods
- Comparison of CLADE with other state-of-the-art SRR methods

2.2 CLADE Architecture

The architecture of the proposed network, CLADE, closely resembles the conventional CycleGAN framework by Zhu et al. [15], but with specific adaptations for our application. These modifications include:

- a weight demodulation process within the generators
- integration of an image-based gradient-mapping cycle-consistency loss
- operating on disjoint image patches (rather than complete images) during both training and inference
- a patch-reconstruction algorithm employed at inference, facilitating the assembly of high-resolution full-sized images from the processed patches

2.2.1 Generator Networks

CLADE contains two generators (G_X and G_Y), each of which contain an encoder block, six residual blocks, and a decoder block (Figure 2.1), similar to the conventional CycleGAN.

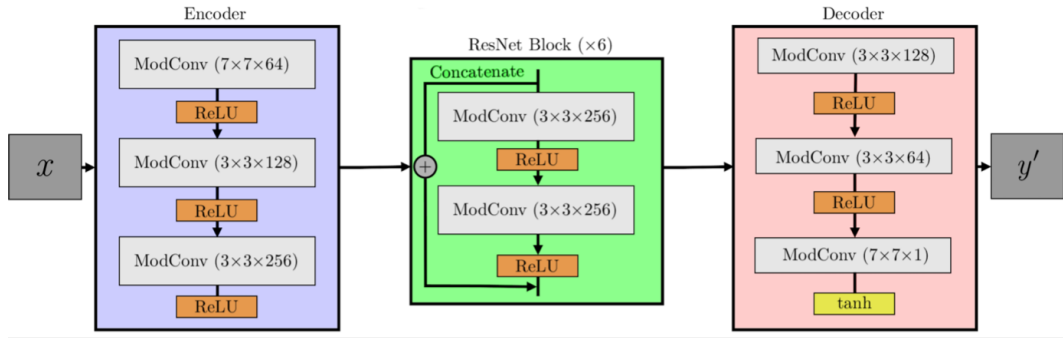


Figure 2.1: CLADE generator architecture shown for $G_X : X \rightarrow Y$. ModConv denotes the 2D modified convolution layers. Each modified convolutional layer performs the weight-demodulation process.

However, unlike the conventional CycleGAN, we replaced instance normalisation layers with a weight demodulation process applied directly to the convolutional weights; this aims to reduce the presence of droplet or block-like noise artifacts which are commonly seen in CycleGANs [35]. This approach is similar to that used in StyleGAN2 [36], where they substitute instance normalisation with a combination of a weight modulation and demodulation. In StyleGAN2, this eliminates the stylistic influence vector, s (intrinsic to the StyleGAN architecture), from the output convolutional feature maps, w , by scaling the weights of the convolutional filters such that $w' = s \cdot w$. As CLADE does not use style vectors, we can assume

that $s = 1 = w' = w$, therefore removing the need for the modulation step entirely. However, we retain the demodulation step, prior to the convolution operation, by normalising the convolutional weights themselves by the reciprocal square root of the sum of squared weights (Equation 2.1):

$$w''_{ijkl} = \frac{w'_{ijkl}}{\sqrt{\sum_{ijkl} (w'_{ijkl})^2 + \epsilon}}, \quad (2.1)$$

where i denotes the index of the output channel, j denotes the horizontal position in the kernel, k denotes the vertical position in the kernel, and l denotes the input channel index; furthermore, we chose $\epsilon = 1 \times 10^{-8}$ as our numerical stability constant to avoid numerical division issues [36].

This technique is directly inspired by weight normalisation [37], a technique used to improve training stability and convergence speed by directly normalising the layer weights themselves. It also acts as a form of regularisation, but unlike techniques such as batch normalisation which introduce noise into the gradients, the dependence on the normalisation process is removed. Although analogous, there are some distinct differences between traditional weight normalisation and weight demodulation. Specifically, weight normalisation is a technique used in neural networks to rescale the weights before they are used in the model, ensuring that each weight vector has a consistent magnitude, which helps to stabilise and accelerate the training process by promoting more consistent updates during learning. Instead, in weight demodulation, the normalisation occurs after the weights have been used in each iteration, primarily to stabilise the output.

2.2.2 Discriminator Networks

The discriminator networks in CLADE, D_X and D_Y , are identical to those in the conventional CycleGAN [15], each containing a PatchGAN [38] (Figure 2.2).

Unlike the generator networks, these discriminators do have instance normalisation layers present, as block-like noise artifacts only occur during image generation and not during the discriminative phase of training.

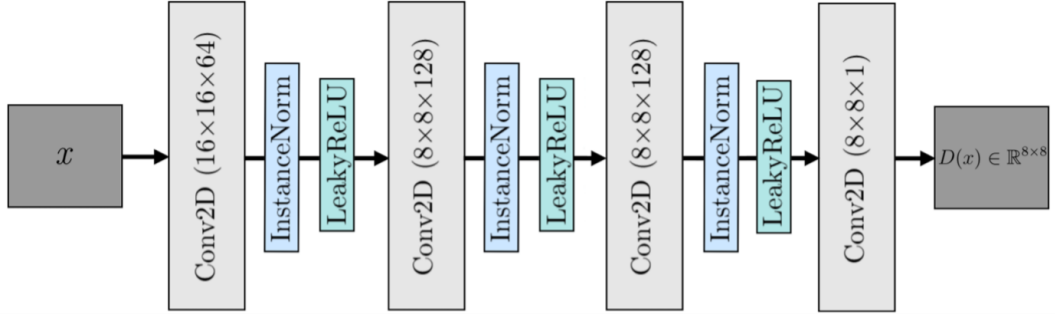


Figure 2.2: CLADE PatchGAN discriminator architecture.

2.3 Gradient Mapping Cycle-Consistency Loss

The conventional CycleGAN architecture has been noted to suffer from deformation errors at the boundaries of soft tissues when applied to anatomical data [39]. Our proposed CLADE architecture acts on local 2D patches of the image volume, and therefore the generator has no contextual knowledge of the global image features. Taking advantage of this fact, we opted to investigate the addition of an absolute gradient mapping between the local patches of the cycled high-resolution images to reduce deformation errors, analogous to a Mixed Gradient Loss [40]. This gradient mapping loss, \mathcal{L}_{gmap} , aims to preserve local edge sharpness between gradient maps of input images and “cycled” images:

$$\mathcal{L}_{gmap}(G_X, G_Y) = \mathbb{E}_{y \sim p_{data}(y)} [\|(\mathbf{S}_x(G_Y(G_X(y))) - \mathbf{S}_x(y)) + (\mathbf{S}_y(G_X(G_Y(y))) - \mathbf{S}_y(y))\|_1], \quad (2.2)$$

where \mathbf{S}_x and \mathbf{S}_y denote the image gradient acquired using a Sobel operator in the x and y directions, respectively (see Equations 2.3 and 2.4 respectively).

$$\mathbf{S}_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad (2.3)$$

$$\mathbf{S}_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (2.4)$$

The Sobel operator was chosen as it is easy to implement, fast to calculate, and does not require any manual tuning. We combine this gradient mapping loss with the conventional CycleGAN loss, \mathcal{L}_{GAN} (Equation 1.17), to define the CLADE total loss, \mathcal{L}_{CLADE} , as follows:

$$\mathcal{L}_{CLADE}(D_X, D_Y, G_X, G_Y) = \mathcal{L}_{GAN}(D_X, D_Y, G_X, G_Y) + \lambda_{gmap} \mathcal{L}_{gmap}, \quad (2.5)$$

where λ_{gmap} is the scalar loss weighting for \mathcal{L}_{gmap} .

2.4 Patched Training and Reconstruction

The training datasets were created by first pre-processing the anisotropic volumes using a cubic-spline interpolation, to create isotropic pixels and ensure consistency between the low-resolution and high-resolution data. Multiple patches of size (32×32) were then extracted from these training volumes (see Section 2.6.3.2 for further information). This patch size was chosen to be (32×32) , as this enforces the network to extract local structural information from within the patches and places a restriction on the effective receptive field [41]. This, in turn, reduces the likelihood of learning specific anatomical features and the potential risk for network hallucination.

At inference CLADE can be applied to perform SRR on a full-size 3D image volume, which is not of fixed size. To achieve this, the 3D image volume must first be pre-processed using cubic-spline interpolation, to create isotropic pixels (in the same way as is performed for training). The interpolated volume is then split into 2D slices, $x \in X$, of shape $m \times n$, where one of the dimensions of these 2D images is the original low-resolution (through-plane) dimension. Each of these 2D images is then deconstructed into overlapping (32×32) pixel patches, where the amount of overlap between consecutive patches is determined by a given stride length. These patches are then independently passed through the SRR network to super-resolve each patch, producing an output patch $y \in Y$. The final images are created using a sliding-window patch reconstruction algorithm that stitches together the resultant super-resolved patches and then normalises the images by dividing by the number

of times each pixel was subsequently sampled (using the Hadamard product).

2.5 Imaging Datasets

Our proposed CLADE approach can be applied to any anisotropic 3D dataset, however we chose to demonstrate its utility on both 3D MRI and 3D CT anisotropic images. All SRR networks were trained independently for the MRI and CT datasets due to differences in resolution and visual contrast of the datasets.

2.5.1 MRI Dataset

Volumetric Interpolated Breath-hold Examination (VIBE) is a popular MRI technique that enables full abdominal 3D coverage with high in-plane resolution and lower through-plane resolution, within an achievable breath-hold time of ≈ 20 seconds [42]. VIBE produces T1-weighted anisotropic 3D images and permits simultaneous evaluation of soft tissue and vasculature. In this study, data were retrospectively collected from 60 patients who undertook abdominal 3D VIBE imaging after gadolinium-based contrast agent administration as part of a clinical protocol conducted at University College Hospital, London, UK. The local research ethics committee approved the study (Ref: 10/H0/720/91). Whole abdominal 3D coverage was achieved in a breath-hold of 21s, with the following imaging parameters: coronal in-plane orientation, matrix size $\approx 185 \times 330$ (Right-Left, Head-Foot, RL-HF), ≈ 23 slices (in Anterior-Posterior, AP direction), field-of-view: $\approx 225 \times 400 \times 150$ mm, in-plane resolution $\approx 1.2 \times 1.2$ mm (range: 0.8-1.6 mm) and through-plane resolution ≈ 6.6 mm (range: 5.2-7.6 mm). This resulted in a super-resolution factor of ≈ 5.5 , and hence pre-processing using cubic-spline interpolation, resulted in ≈ 128 slices in the AP direction per 3D volume (range: 90-215). Of these 60 3D MRI VIBE datasets, 45 volumes were used for training (resulting in a total of 23,068 patches, see Section 2.6.3), and 15 volumes were reserved for testing of the SRR networks .

2.5.2 CT Dataset

The open-source DeepLesion CT dataset [43] comprises of a diverse collection of 32,120 CT slices. In this study, we chose a subset of 60 anisotropic 3D CT volumes

of the abdomen, with a comparable anisotropy ratio to the MRI data (i.e. a super-resolution factor of ≈ 5.5). The data were collected with the following imaging parameters: transverse orientation, matrix size 512×512 (RL-AP), ≈ 22 slices (in HF direction, range: 13-83), in-plane resolution $\approx 0.95 \times 0.95$ mm (range: 0.90-0.98 mm) and through-plane resolution of 5.0 mm. This resulted in a super-resolution factor of ≈ 5.3 , and hence pre-processing using cubic-spline interpolation resulted in ≈ 118 slices in the HF direction per 3D volume (range: 67-443). The datasets were all cropped to 320×320 matrix in-plane, to remove air around the body. Of these 60 3D CT datasets, 45 volumes were used for training (resulting in a total of 22,520 patches, see Section 2.6.3), and 15 volumes were reserved for testing of the SRR networks.

2.6 Experimental Network Setup

We performed a number of experiments, specifically to:

- Optimise the CLADE Network Architecture
- Assess the impact of Weight Demodulation and the Gradient Mapping Loss
- Assess the impact of Supervised vs. Unsupervised training
- Compare the image quality from CLADE to other SRR methods

All networks were trained in Python3 using Tensorflow 2.0 on an NVIDIA RTX A6000 GPU with 48GB of available memory. Adam [44] was used as the optimiser, with a fixed learning rate of 2×10^{-4} and the default momentum coefficient $\beta = 0.5$.

2.6.1 CLADE Network Optimisation

We performed hyper-parameter optimisation of CLADE using all of the (unpaired) training data over one epoch, to optimise:

- The weightings for CLADE cycle-consistency ($\lambda_{cyc} \in \{0.1, 1, 10\}$), identity ($\lambda_{ident} \in \{0, 1, 10\}$), and gradient-mapping ($\lambda_{gmap} \in \{1, 5, 10\}$) losses to produce the best SRR image quality

- The number of epochs in training $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
- The patch reconstruction stride length $\{6, 8, 12, 16, 20\}$ as we observed a trade-off between image quality and reconstruction time

The no-reference image quality metric, PIQUE (Perception-based Image Quality Evaluator) [45], subjective image quality and reconstruction time over the test dataset were assessed for these hyper-parameters. As PIQUE is a 2D metric, we calculated PIQUE on all 2D slices in the LR orientation, and the mean was used for comparison. Due to extensive training times, for all models across the entire dataset, only one epoch was used for all models as a comparison.

2.6.2 Assessing the Importance of the Weight Demodulation and Gradient Mapping Loss

We wanted to assess the importance of the proposed weight demodulation and gradient mapping loss (with the optimised hyper-parameters from above), therefore we tested the following networks:

- Conventional CycleGAN
- CLADE without gradient mapping loss (CLADE no \mathcal{L}_{gmap}) i.e. $\mathcal{L}_{gmap} = 0$
- CLADE with gradient mapping loss (CLADE with \mathcal{L}_{gmap})

2.6.3 Assessing Supervised vs. Unsupervised Training

As described in the literature section (see Section 1.4), many SRR models are trained using paired training data, where the low-resolution pairs are synthetically created from the high-resolution data. It is possible to do this using our in-plane data. Therefore, we assessed the effectiveness of CLADE when it is trained:

- In a supervised fashion, from synthetic paired training data created from the high-resolution plane of the anisotropic 3D data
- In a truly unpaired fashion, from independent patches taken from the low-resolution and high-resolution slice planes of the anisotropic 3D data

The supervised CLADE networks were assessed on both synthetic low-resolution data and on the prospective low-resolution data (taken from the sagittal plane for both the MRI and CT data) to assess whether the degradation process implicitly learned by the unpaired networks aligns with the simple degradation model applied to the synthetic data. This approach helps determine if the degradation mechanisms in the prospective datasets are more complex and thus differ from the uniform degradation assumed in the synthetic data. It allows us to evaluate whether the unpaired networks can adapt to and accurately handle the complexities of real-world degradations, highlighting if a standardised physically-based degradation model is inadequate.

2.6.3.1 Creation of Paired Synthetic Training Data

For training of the supervised networks, image pairs were generating by creating synthetic low-resolution images from the high-resolution plane of the anisotropic volume (see Algorithm 1):

Algorithm 1 Sinc Interpolation Downsampling

- 1: **Input:** High-resolution image $I_{high}(x, y)$, Anisotropy factor α , Number of pixels in x -direction N_x
 - 2: **Output:** Low-resolution image $I_{low}(x, y)$
 - 3: Compute Fourier transform:
 - 4: $\tilde{I}_{high}(k_x, k_y) \leftarrow \mathcal{F}[I_{high}(x, y)]$
 - 5: Zero-fill in Fourier domain:
 - 6: $\tilde{I}_{low}(k_x, k_y) \leftarrow \begin{cases} \tilde{I}_{high}(k_x, k_y) & \text{if } |k_x| \leq \alpha \cdot \frac{N_x}{2}, \\ 0 & \text{otherwise} \end{cases}$
 - 7: Perform inverse Fourier transform:
 - 8: $I_{low}(x, y) \leftarrow \mathcal{F}^{-1}[\tilde{I}_{low}(k_x, k_y)]$
 - 9: **return** $I_{low}(x, y)$
-

For training four random 2D patches of size (32×32) are extracted from each high-resolution slice in the volume (the high-resolution slices are taken from the coronal plane for MRI data and the transverse plane for CT data), and the paired patch taken from the same location in the synthetically generated low-resolution image.

2.6.3.2 Creation of Unpaired Training Data

The interpolated 3D volumes are separated into 2D slices in both the low-resolution and high-resolution planes. For the MRI data, the coronal plane is high-resolution in both spatial directions, and the sagittal plane was chosen as the low-resolution plane (although the transverse plane could have been chosen). For the CT data, the transverse plane is high-resolution in both spatial directions, and the sagittal plane was chosen as the low-resolution plane (similarly, the coronal plane could have been chosen). To create the training datasets, we extract four 2D patches of size (32×32) pixels were extracted from each slice in the training volumes.

2.6.4 Comparison with State-of-the-art SRR Method: SMORE

We will compare the resulting image quality from supervised and unsupervised CLADE (with \mathcal{L}_{gmap} and no \mathcal{L}_{gmap}) networks to SMORE, a state-of-the-art self-supervised SRR approach (as described in Section 1.4.2¹ [31, 32]).

2.7 Image Quality Metrics

To evaluate the image quality of the various network architectures, a selection of image quality metrics was made based on their ability to accommodate both paired and unpaired variants. These metrics were chosen to meet the requirement of calculating quality with and without ground truth image pairs.

2.7.1 Ground Truth Image Pairs: PSNR, SSIM, MSE and MAE

Where networks are trained in a supervised manner, using paired synthetic low-resolution and ground-truth image pairs exist, it was possible to calculate the following imaging metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Mean Squared Error (MSE) and Mean Absolute Error (MAE).

PSNR is a metric used heavily within image processing and compression algorithms to quantitatively assess the fidelity of reconstructed images compared to their originals. It measures the ratio of the maximum possible signal power to the power of corrupting noise, usually expressed in decibels (dB). The Peak Signal-to-Noise

¹The SMORE implementation that was used is publicly accessible here: <https://gitlab.com/iacl/smores/-/tree/main>

Ratio (PSNR) is calculated as:

$$\text{PSNR} = \begin{cases} \infty & \text{if } MSE = 0 \\ 10 \cdot \log_{10} \left(\frac{P_{\max}^2}{MSE} \right) & \text{if } MSE > 0 \end{cases} \quad (2.6)$$

where P_{\max} represents the maximum possible pixel value (e.g., 255 for 8-bit images) and MSE denotes the Mean Squared Error, defined as the average squared difference between the original and reconstructed images.

Furthermore, SSIM [46, 47] is a metric used to measure the similarity between two images, x and y , by comparing their luminance, contrast, and structure, providing a score that reflects perceived visual quality; it is computed as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (2.7)$$

where x and y represent the compared images, μ_x and μ_y denote their respective means, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance between x and y . The scalar constants c_1 and c_2 are included to prevent division by zero errors.

The MSE is a metric that measures the average squared difference between the corresponding pixel values of two images, x and y . It quantifies the cumulative squared error and is defined as:

$$\text{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2, \quad (2.8)$$

where N is the total number of pixels, and x_i and y_i represent the pixel values of the two images at the i -th position. A lower MSE value indicates a higher similarity between the two images.

Similarly, the MAE is a metric that measures the average absolute difference between the pixel values of two images, x and y , giving a more interpretable error

than MSE by focusing on the absolute differences:

$$\text{MAE}(x, y) = \frac{1}{N} \sum_{i=1}^N |x_i - y_i|, \quad (2.9)$$

where N is the number of pixels, and x_i and y_i are the pixel values at the i -th position. MAE provides a direct understanding of the average error between the images.

2.7.2 No Ground Truth: PIQUE, NIQE and BRISQUE

Alternatively, no-reference image quality assessment metrics exist, which evaluate image quality without needing a reference image or ground truth for comparison. These were used to assess image quality from both the supervised networks and unsupervised networks. The no-reference image quality assessment metrics utilised in this study are PIQUE (Perception-based Image Quality Evaluator) [48], NIQE (Natural Image Quality Evaluator) [49], and BRISQUE (Blind/Reference-less Image Spatial Quality Evaluator) [50]. PIQUE measures image quality by detecting perceptually significant artifacts, with lower scores indicating better quality. NIQE assesses image quality based on statistical features extracted from natural scenes, where deviations from these natural image statistics are quantified, and lower scores imply higher quality. BRISQUE evaluates image quality using spatial domain features, specifically by analysing scene statistics of locally normalised luminance coefficients, where lower scores similarly correspond to better image quality.

2.8 Quantitative Image Quality Assessment

For the final unsupervised networks, quantitative edge sharpness (ES) and estimated Signal-to-Noise Ratio (SNR) image quality metrics were assessed in the test datasets for MRI and CT images from original low-resolution data, SMORE, the conventional CycleGAN, CLADE (no \mathcal{L}_{gmap}) and CLADE (with \mathcal{L}_{gmap}). The quantitative edge sharpness and SNR measurements were made using in-house plug-ins for the OsiriX DICOM viewing platform [51].

Edge sharpness was quantified on 3D MPR images, as described previously

[52], by measuring the maximum gradient of the pixel intensities perpendicularly across the border of interest. Pixel intensities were filtered using a Savitzky–Golay filter to remove the effect of noise (window width = 5 pixels, third-order polynomial) before differentiation. Edge sharpness was taken as the maximum gradient of the filtered pixel intensities. In the MRI data, Edge sharpness was measured across four distinct anatomical regions: the abdominal aorta, liver, lower pole of the kidney, and the spleen, and an average edge sharpness value was taken for comparison. In the CT data, edge sharpness was measured in one region: across a bony structure.

Estimated SNR was calculated by dividing the mean signal intensity in a region of interest (ROI) (drawn in the kidney - an area of contrast uptake - in MRI, and in a bony structure in CT) by the standard deviation of the pixel values within a ROI drawn in an area of no signal (air in the stomach in MRI and air outside the body in CT).

2.8.1 Statistical Analysis Methodology

Statistical analyses were performed using Python (`statsmodels` 0.13.5). Quantitative metrics were compared using one-way repeated measures analysis of variance (ANOVA) with post-hoc testing using Tukey Honest Significant Difference (HSD) multiple comparisons of means to determine group-wise relationships and significant results.

Chapter 3

Results

3.1 CLADE Network Optimisation

3.1.1 Loss Weighting Optimisation

CLADE (with \mathcal{L}_{gmap}) models were trained for one epoch (with a stride of 6) with varying loss weightings: $\lambda_{cyc} \in \{0.1, 1, 10\}$, $\lambda_{ident} \in \{0, 1, 10\}$, and $\lambda_{gmap} \in \{1, 5, 10\}$. Each resulting network was applied to all 15 volumes in the test set, and the quantitative PIQUE scores calculated for comparison.

3.1.1.1 MRI Loss Optimisation

The quantitative results for the MRI CLADE networks can be found in Table 3.1 and representative visual image quality seen in Figure 3.1. The final chosen loss weighting for CLADE network for MRI data was $\lambda_{cyc} = 1$, $\lambda_{ident} = 1$, and $\lambda_{gmap} = 5$, based on PIQUE score and visual image quality. Specifically, we are visually evaluating network hallucinations, introduction / worsening of noise and visual image sharpness.

3.1.1.2 CT Loss Optimisation

The quantitative results for the CT CLADE networks can be found in Table 3.2 and representative visual image quality seen in Figure 3.2. The final chosen loss weighting for CLADE network for CT data was $\lambda_{cyc} = 1$, $\lambda_{ident} = 1$, and $\lambda_{gmap} = 10$, based on PIQUE score and visual image quality. This is evaluated visually in the same way as in Section 3.1.1.1.

Table 3.1: PIQUE scores from MRI CLADE (with \mathcal{L}_{gmap}) hyper-parameter optimisation, shown in ascending PIQUE order.

λ_{cyc}	λ_{ident}	λ_{gmap}	PIQUE ($\mu \pm \sigma$)
0.1	0	1	20.20 ± 2.98
0.1	0	5	22.51 ± 4.28
0.1	1	1	22.74 ± 4.76
0.1	0	10	27.48 ± 3.94
0.1	1	10	30.98 ± 5.34
0.1	10	1	31.25 ± 4.66
0.1	10	5	31.66 ± 4.28
0.1	1	5	32.15 ± 4.17
0.1	10	10	35.25 ± 4.90
1	0	1	31.34 ± 4.37
1	0	5	30.57 ± 5.10
1	0	10	30.28 ± 4.94
1	1	1	37.78 ± 4.51
1	1	5	25.55 ± 4.71
1	1	10	31.26 ± 4.66
1	10	1	27.54 ± 4.88
1	10	5	26.22 ± 4.62
1	10	10	32.05 ± 4.89
10	0	1	44.94 ± 6.23
10	0	5	29.40 ± 5.13
10	0	10	38.62 ± 5.70
10	1	1	36.23 ± 5.55
10	1	5	34.67 ± 5.13
10	1	10	38.38 ± 5.83
10	10	1	36.31 ± 5.18
10	10	5	36.71 ± 5.81
10	10	10	36.72 ± 5.45

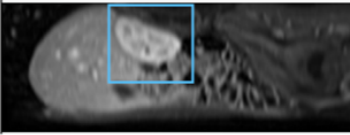

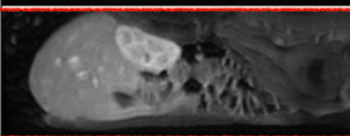


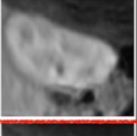
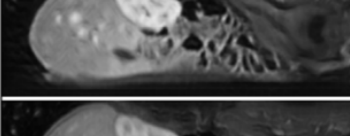
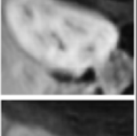
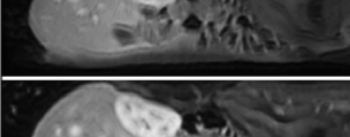
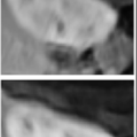
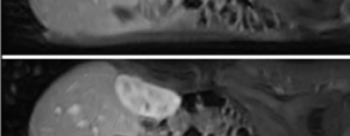
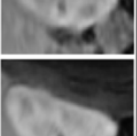
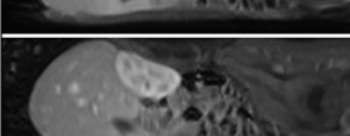
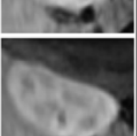
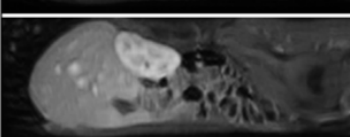
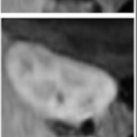
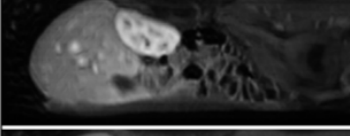

λ_{cyc}	λ_{ident}	λ_{gmap}	Low-resolution	0.1	0.1	0.1	0.1	1	0.1	1	10	1	10	1	1	PIQUE
				0	0	0	1	1	0	10	0	10	0	0	0	
				1	5	1	1	5	10	5	5	1	5	10	5	
			Sagittal Slice		Zoom											
																
																
																
																
																
																
																
																
																

Figure 3.1: Image quality from a MRI test subject, using CLADE (with \mathcal{L}_{gmap}) across the 10 models with the lowest PIQUE scores. Blue boxes show magnified regions; red box highlights the selected final loss weighting.

Table 3.2: PIQUE scores from CT CLADE (with \mathcal{L}_{gmap}) hyper-parameter optimisation, shown in ascending PIQUE order.

λ_{cyc}	λ_{ident}	λ_{gmap}	PIQUE ($\mu \pm \sigma$)
0.1	0	1	27.65 ± 4.45
0.1	0	5	32.00 ± 3.23
0.1	0	10	34.78 ± 3.66
0.1	1	1	35.15 ± 5.08
0.1	1	5	31.23 ± 3.06
0.1	1	10	24.47 ± 3.44
0.1	10	1	35.72 ± 4.20
0.1	10	5	33.08 ± 4.23
0.1	10	10	30.73 ± 3.80
1	0	1	36.29 ± 3.04
1	0	5	28.26 ± 4.92
1	0	10	34.34 ± 3.75
1	1	1	33.67 ± 5.86
1	1	5	28.97 ± 3.89
1	1	10	31.24 ± 3.93
1	10	1	31.77 ± 3.80
1	10	5	33.51 ± 4.42
1	10	10	38.91 ± 4.43
10	0	1	30.86 ± 4.18
10	0	5	32.14 ± 5.20
10	0	10	36.45 ± 5.27
10	1	1	33.79 ± 3.92
10	1	5	36.91 ± 5.27
10	1	10	38.68 ± 4.88
10	10	1	37.91 ± 4.87
10	10	5	38.23 ± 5.54
10	10	10	42.68 ± 5.46

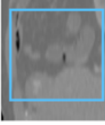
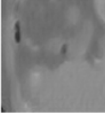
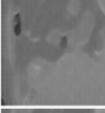
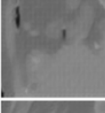
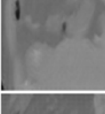
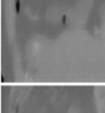
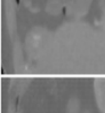
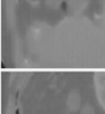
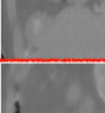
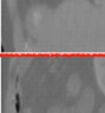
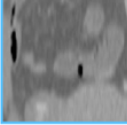
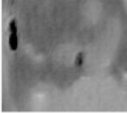
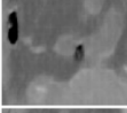
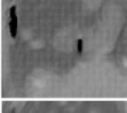
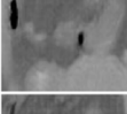
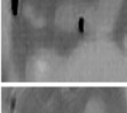
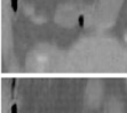
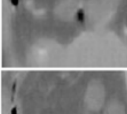
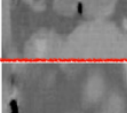
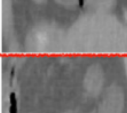
λ_{cyc}	0.1	0.1	1	1	0.1	10	0.1	1	1	0.1
λ_{ident}	1	0	0	1	10	0	1	1	10	0
λ_{gmap}	10	1	5	5	10	1	5	10	1	5
Sagittal Slice										
										
PIQUE	75.89	24.47	27.65	28.26	28.97	30.73	30.86	31.23	31.24	32.00

Figure 3.2: Image quality from a CT test subject, using CLADE (with \mathcal{L}_{gmap}) across the 10 models with the lowest PIQUE scores. Blue boxes show magnified regions; red box highlights the selected final loss weighting.

3.1.2 Epoch Optimisation

Training for the CLADE (with \mathcal{L}_{gmap}) models with the optimised loss functions took ≈ 36 hours for 10 epochs. For each epoch, the network was applied to all 15 volumes in the test set, and the quantitative PIQUE scores calculated for comparison.

3.1.2.1 MRI Epoch Optimisation

The quantitative PIQUE scores at each epoch for the MRI data can be found in Table 3.3 and representative image quality at each epoch seen in Figure 3.3. The final CLADE (with \mathcal{L}_{gmap}) model was chosen from the ninth epoch for MRI data, based on PIQUE score and visual image quality.

3.1.2.2 CT Epoch Optimisation

The quantitative PIQUE scores at each epoch for the CT data can be found in Table 3.4 and representative image quality at each epoch seen in Figure 3.4. The final CLADE (with \mathcal{L}_{gmap}) model was chosen from the tenth epoch for CT data, based on PIQUE score and visual image quality.

Table 3.3: PIQUE scores ($\mu \pm \sigma$) as calculated across all 15 volumes in the MRI test dataset, assessed with a stride length of 6. The scores are compared across different epochs for the final MRI CLADE (with \mathcal{L}_{gmap}) model ($\lambda_{cyc} = 1$, $\lambda_{ident} = 1$, $\lambda_{cyc} = 5$). The chosen number of epochs is outlined in bold, based on PIQUE score and visual image quality (see Figure 3.3).

Epoch	PIQUE ($\mu \pm \sigma$)
1	33.73 ± 3.57
2	44.12 ± 0.71
3	23.34 ± 4.75
4	35.53 ± 5.93
5	26.52 ± 4.68
6	41.62 ± 5.18
7	39.08 ± 5.05
8	31.46 ± 5.43
9	27.42 ± 4.19
10	30.44 ± 4.70



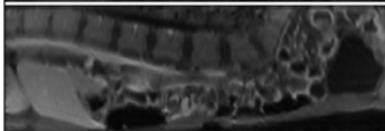
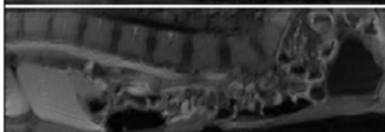

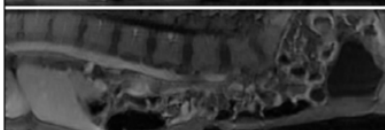
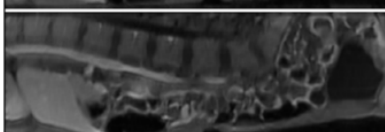


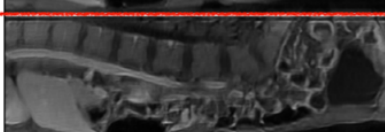
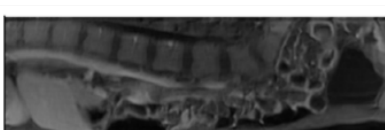


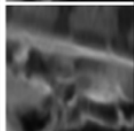



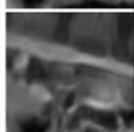



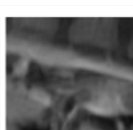
Epoch	Low-resolution	1	2	3	4	5	6	7	8	9	10	PIQUE
												
												

Figure 3.3: Example image quality from one subject in the MRI test dataset, at each epoch when training the final MRI CLADE (with \mathcal{L}_{gmap}) ($\lambda_{cyc} = 1, \lambda_{ident} = 1, \lambda_{cyc} = 5$). Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the MRI test dataset. The red box indicates the chosen epoch based on PIQUE scores (see Table 3.3) as well as visual image quality.

Table 3.4: PIQUE scores ($\mu \pm \sigma$) as calculated across all 15 volumes in the CT test dataset, assessed with a stride length of 6. The scores are compared across different epochs for the final CT CLADE (with \mathcal{L}_{gmap}) model ($\lambda_{cyc} = 1$, $\lambda_{ident} = 1$, $\lambda_{cyc} = 10$). The chosen number of epochs is outlined in bold, based on PIQUE score and visual image quality (see Figure 3.4).

Epoch	PIQUE ($\mu \pm \sigma$)
1	30.06 ± 4.68
2	36.28 ± 4.33
3	33.26 ± 5.87
4	37.76 ± 3.90
5	33.00 ± 5.20
6	34.23 ± 5.13
7	29.73 ± 4.83
8	37.67 ± 3.99
9	29.08 ± 4.25
10	28.24 ± 4.45


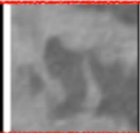
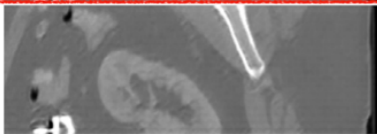
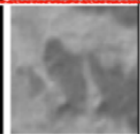

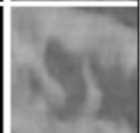

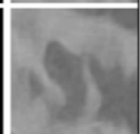

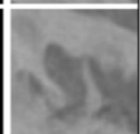
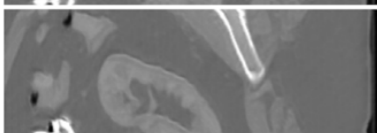
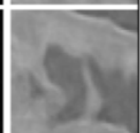

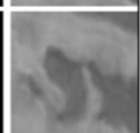

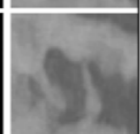

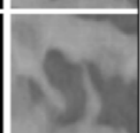
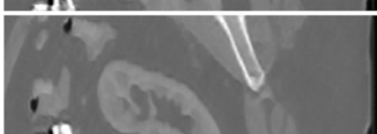
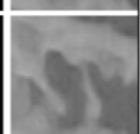
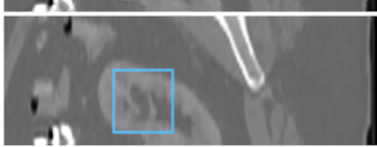
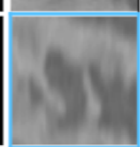
Epoch	Low-resolution	Sagittal Slice	Zoom	PIQUE
10				28.24
9				29.08
8				37.67
7				29.73
6				34.23
5				33.00
4				37.76
3				33.26
2				36.28
1				30.06
	Low-resolution			75.89

Figure 3.4: Example image quality from one subject in the CT test dataset, at each epoch when training the final CT CLADE (with \mathcal{L}_{gmap}) ($\lambda_{cyc} = 1$, $\lambda_{ident} = 1$, $\lambda_{cyc} = 10$). Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the CT test dataset. The red box indicates the chosen epoch based on PIQUE scores (see Table 3.4) as well as visual image quality.

3.1.3 Patch Reconstruction Stride Length

Stride length in patch reconstruction affects both image quality and reconstruction time. We tested stride lengths of 6, 8, 12, 16, and 20, calculating PIQUE scores and reconstruction times across the 3D test volumes using optimised CLADE models (with \mathcal{L}_{gmap}). Results for the 15 MRI test volumes are shown in Table 3.5, with representative images in Figure 3.5, while Table 3.6 and Figure 3.6 present the results for the 15 CT volumes.

Table 3.5: Impact of stride length on PIQUE scores and inference time across 15 MRI test volumes for the final MRI CLADE model ($\lambda_{cyc} = 1$, $\lambda_{ident} = 1$, $\lambda_{gmap} = 5$, 9 epochs)

Stride Length	PIQUE ($\mu \pm \sigma$)	Inference Time per 3D Volume (s) ($\mu \pm \sigma$)
6	27.42 ± 4.19	317 ± 132
8	29.17 ± 4.09	195 ± 80
12	26.61 ± 3.78	103 ± 38
16	31.25 ± 5.32	72 ± 25
20	25.41 ± 3.88	53 ± 16

Table 3.6: Impact of stride length on PIQUE scores and inference time across 15 CT test volumes for the final CLADE model ($\lambda_{cyc} = 1$, $\lambda_{ident} = 1$, $\lambda_{gmap} = 10$, 10 epochs).

Stride Length	PIQUE ($\mu \pm \sigma$)	Inference Time per 3D Volume (s) ($\mu \pm \sigma$)
6	28.31 ± 4.42	292 ± 133
8	28.20 ± 4.53	183 ± 74
12	24.69 ± 4.53	100 ± 34
16	28.59 ± 4.81	73 ± 19
20	21.49 ± 4.84	54 ± 16

The optimal balance between image quality and reconstruction time for MRI CLADE (with \mathcal{L}_{gmap}) was found with a stride length of 12 pixels. This resulted in ≈ 204 patches per slice (range: 132-300). The optimal balance between image quality and reconstruction time for CT CLADE (with \mathcal{L}_{gmap}) was found with a stride length of 6 pixels. This resulted in ≈ 581 patches per slice (range: 294-1078).

Stride	Low-resolution	6	8	12	16	20
Sagittal Slice						
Zoom						
PIQUE	62.44	27.42	29.17	26.61	31.25	25.41

Figure 3.5: Example image quality from one subject in the MRI test dataset, at each stride length, using the final MRI CLADE during inference. Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the MRI test dataset. The red box indicates the chosen patch size based on PIQUE score, inference time and visual image quality .

Stride	Low-resolution	6	8	12	16	20
Sagittal Slice						
Zoom						
PIQUE	75.89	28.31	28.20	24.69	28.59	21.49

Figure 3.6: Example image quality from one subject in the CT test dataset, at each stride length, using the final CT CLADE (with \mathcal{L}_{gmap}) during inference. Magnified regions within the blue box are displayed beneath each image. PIQUE score shows the mean value across all 15 volumes in the CT test dataset. The red box indicates the chosen patch size based on PIQUE score, inference time and visual image quality.

3.2 Importance of Weight Demodulation

To assess the impact of the proposed weight demodulation we compared the conventional CycleGAN to CLADE (no \mathcal{L}_{gmap}). These networks were trained using the same number of epochs, stride length, and the same optimised loss weightings for λ_{cyc} and λ_{ident} , but with λ_{gmap} set to 0.

3.2.1 MRI weight demodulation

When trained and tested using the MRI data, the conventional CycleGAN architecture (with instance normalisation) resulted in “blocky” grid artifacts with signal voids in some MRI super-resolved volumes. However, these errors were successfully removed when using CLADE (no \mathcal{L}_{gmap}) as seen in Figure 3.7. This demonstrating the utility of the weight demodulation process in improving MRI image quality.

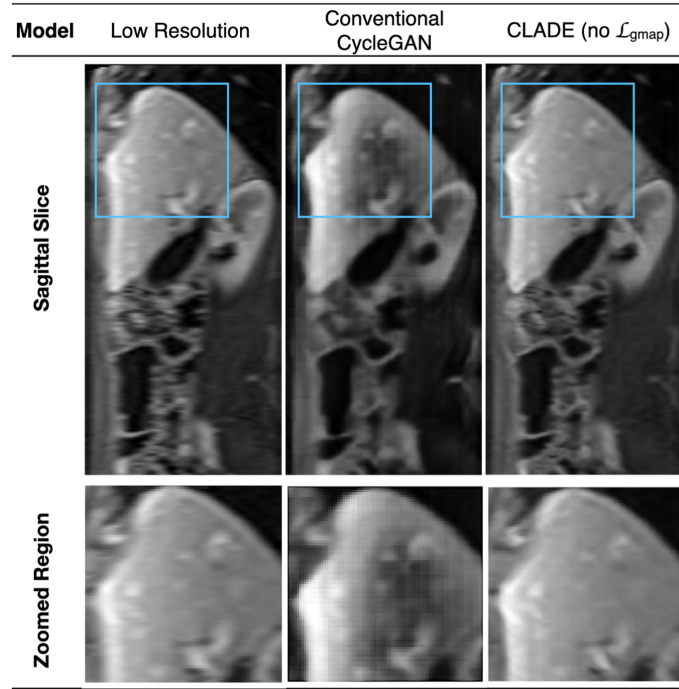


Figure 3.7: An example low-resolution sagittal MRI slice from one subject in the test data, which shows normalisation errors when SRR is applied using a Conventional CycleGAN. These artifacts are removed when SRR is applied using CLADE (no \mathcal{L}_{gmap}). Magnified regions within the blue box are displayed beneath each image.

3.2.2 CT weight demodulation

When trained and tested using the CT data, the conventional CycleGAN (with instance normalisation) suffered from mode collapse (see Figure 3.8). This may be due to the lack of visual contrast in the CT datasets, compared to the MRI data, which will likely have made learning these domain mappings a more challenging task.

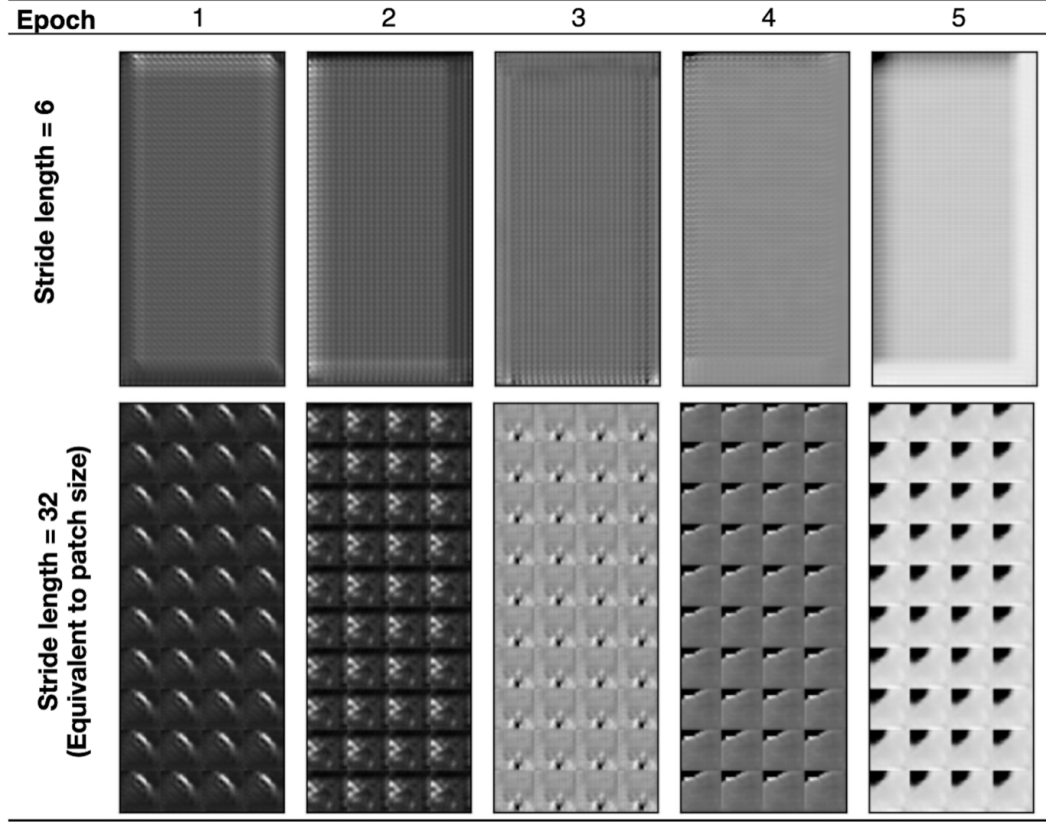


Figure 3.8: The conventional CycleGAN suffered from mode collapse when trained using the CT data. The top row shows the patched image reconstruction from the model at epochs 1-5, on one of the test CT volumes, with a stride of 6 (as used for CLADE). The bottom row shows that where there are no overlapping patches (i.e. stride length is equal to the patch size), then all patches remain identical, demonstrating mode collapse.

3.3 Supervised Training

The Conventional CycleGAN, CLADE (no \mathcal{L}_{gmap}) and CLADE (with \mathcal{L}_{gmap}) networks were trained using synthetic paired data, and subsequently tested on synthetic low-resolution data, as well as the prospective low-resolution data. For the synthetic

test data, we compared paired-image metrics SSIM, PSNR, MSE and MAE, as well as no-reference image metrics PIQUE, NIQE and BRISQUE, as calculated across all 15 test volumes.

3.3.1 MRI Supervised Networks

The supervised networks are trained from synthetic low-resolution MRI patches, taken from the coronal plane. The paired-image metrics from the synthetic test data (created from the coronal plane) are shown in Table 3.7 and the no-reference metrics for this synthetic MRI test data are shown in Table 3.8. Furthermore, images for the paired networks, applied to both paired and prospective datasets are displayed in Figure 3.9 and Figure 3.10 respectively.

Table 3.7: SSIM, PSNR, MSE and MAE scores ($\mu \pm \sigma$) for the MRI paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.

Model	SSIM: Coronal Orientation	PSNR: Coronal Orientation	MSE: Coronal Orientation	MAE: Coronal Orientation
Conventional CycleGAN	0.583 ± 0.1	18.8 ± 2.8	0.01 ± 0.01	0.1 ± 0.04
CLADE (no \mathcal{L}_{gmap})	0.738 ± 0.08	27.0 ± 2.6	0.003 ± 0.006	0.03 ± 0.02
CLADE (with \mathcal{L}_{gmap})	0.756 ± 0.08	27.3 ± 2.7	0.002 ± 0.005	0.03 ± 0.02

Table 3.8: PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the MRI paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.

Model	PIQUE: Coronal Orientation	NIQE: Coronal Orientation	BRISUQE: Coronal Orientation
Low-resolution	73.1 ± 5.50	8.38 ± 1.68	41.5 ± 5.40
Conventional CycleGAN	31.5 ± 11.0	6.68 ± 1.90	29.7 ± 14.3
CLADE (no \mathcal{L}_{gmap})	39.8 ± 4.39	6.41 ± 1.53	36.4 ± 4.11
CLADE (with \mathcal{L}_{gmap})	23.7 ± 4.83	6.21 ± 1.68	23.5 ± 7.60

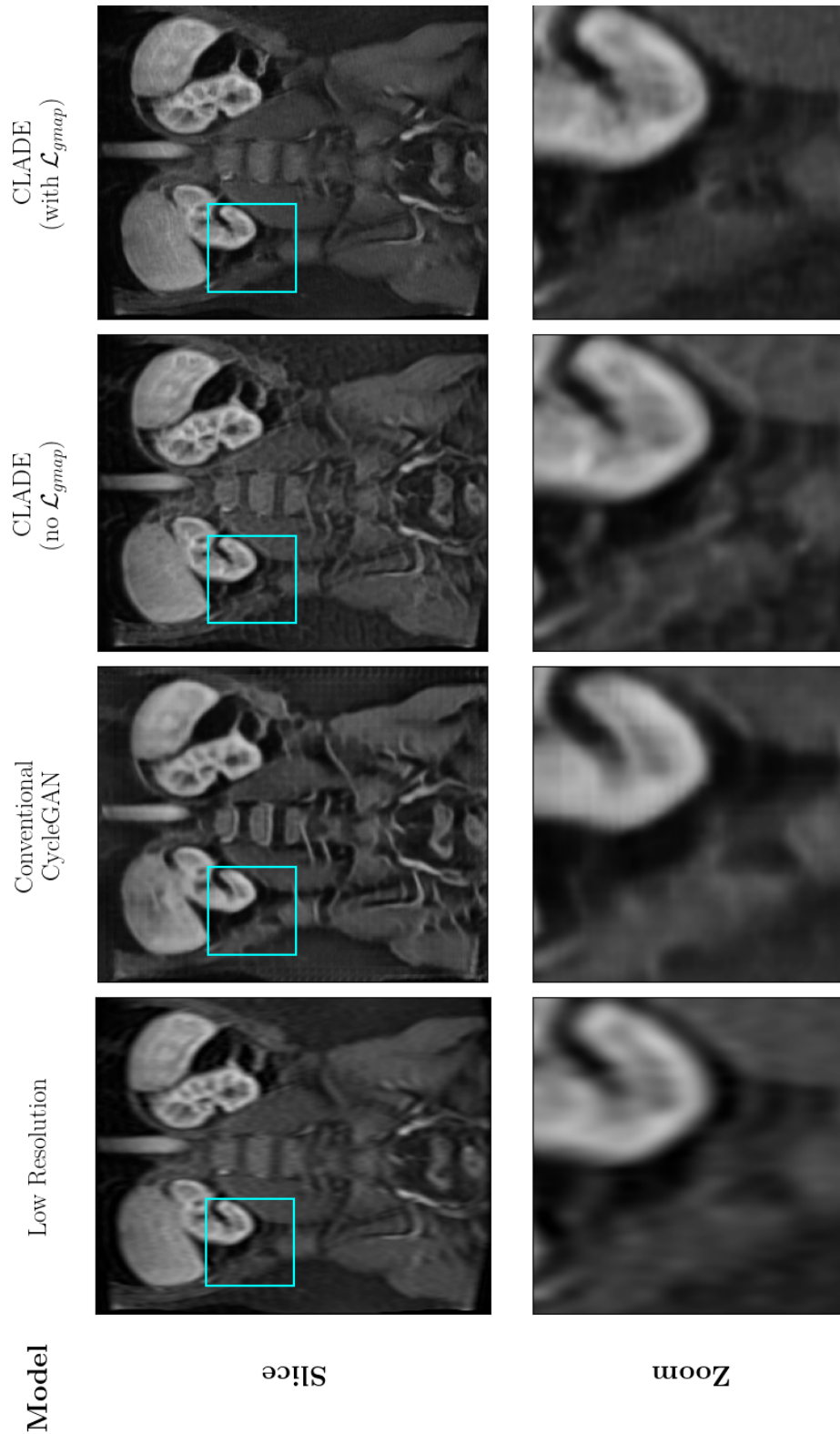


Figure 3.9: MRI paired networks, applied to paired data (coronal). Although there are some improvements to the resolution, the networks fail to faithfully reconstruct the original ground-truth high resolution images. It can clearly be seen that although resolution is marginally improved, the downsampling artifacts and degradations still remain present in the reconstructions.

These results show that the networks achieve good image quality when trained using paired synthetic data. On synthetic test data, the CLADE networks outperform the Conventional CycleGAN in terms of SSIM, PSNR, MSE, MAE. In the case of the no-reference metrics, CLADE with \mathcal{L}_{gmap} outperforms the Conventional CycleGAN, but the supervised CLADE network with no \mathcal{L}_{gmap} does not outperform the Conventional CycleGAN. Furthermore, the supervised CLADE network with \mathcal{L}_{gmap} , consistently outperforms the supervised CLADE network with no \mathcal{L}_{gmap} on synthetic data. These networks were also applied to the prospective low-resolution data, in both the transverse and sagittal planes. The no-reference metrics as calculated across the 15 test data sets are shown in Table 3.9 and images from one of the orientations (sagittal) is displayed in Figure 3.10.

Table 3.9: PIQUE, NIQUE and BRISQUE scores ($\mu \pm \sigma$) for the MRI paired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes.

MRI Orientation		PIQUE	NIQUE	BRISQUE
Transverse				
	Low-resolution	65.0 ± 5.48	8.48 ± 1.40	44.3 ± 5.86
	Conventional CycleGAN	52.0 ± 6.91	7.68 ± 1.86	46.8 ± 5.99
	CLADE (no \mathcal{L}_{gmap})	49.2 ± 6.39	9.78 ± 12.41	45.9 ± 6.87
	CLADE (with \mathcal{L}_{gmap})	36.8 ± 7.28	8.00 ± 1.73	34.7 ± 8.38
Sagittal				
	Low-resolution	63.3 ± 6.04	9.56 ± 2.37	32.3 ± 6.79
	Conventional CycleGAN	48.7 ± 11.4	8.30 ± 1.62	28.2 ± 6.24
	CLADE (no \mathcal{L}_{gmap})	48.9 ± 6.97	9.24 ± 2.00	26.4 ± 5.44
	CLADE (with \mathcal{L}_{gmap})	41.6 ± 9.39	8.86 ± 2.26	23.4 ± 5.20

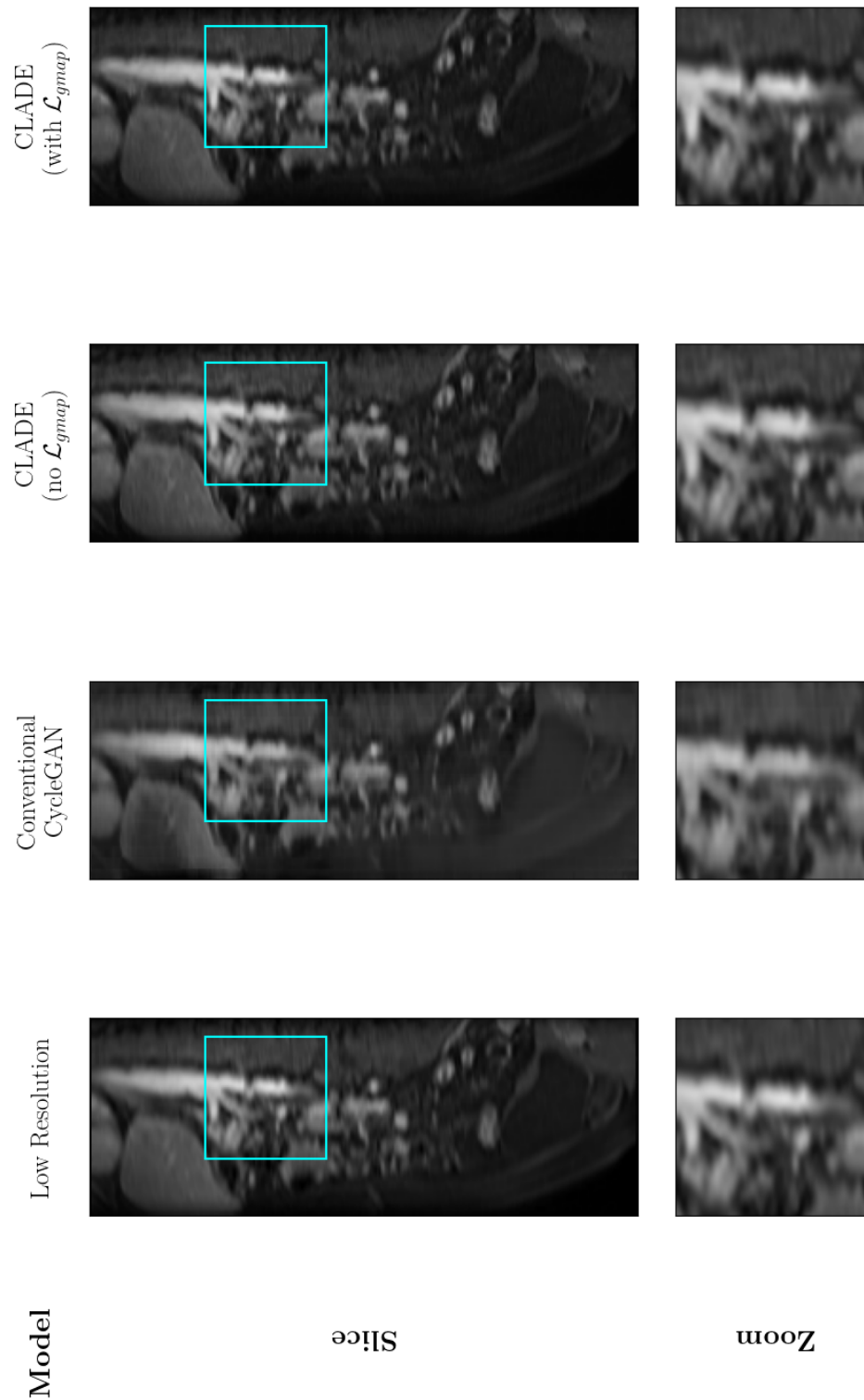


Figure 3.10: MRI paired networks, applied to unpaired data (sagittal). In contrast to the paired data case, the networks fail to perform any reasonable form of super-resolution across all of the networks. This gives further evidence that the simple downsampling method used to generate paired data doesn't faithfully represent the complex degradations present in the true low-resolution data as pictured here.

Comparing the no-reference metrics from the synthetic test data and the prospective test data, it can be seen that the image quality from the prospective test data is significantly lower, indicating that the degradation process is not accurately captured when generating the synthetic datasets. Furthermore, although the respective PIQUE scores indicates that the Conventional CycleGAN outperforms CLADE (no \mathcal{L}_{gmap}), visually there is the presence of more patching artefacts and blur (as can be seen in Figure 3.9).

3.3.2 CT Supervised Networks

The supervised CT networks were trained from synthetic low-resolution patches, taken from the transverse plane. The conventional CycleGAN suffered from model collapse (as displayed in Figure 3.8), and therefore these results are not included. The paired-image metrics from the synthetic test data (created from the transverse plane) are shown in Table 3.10 and the no-reference metrics for this synthetic CT test data are shown in Table 3.11. Furthermore, images for the paired networks, applied to both paired and prospective datasets are displayed in Figure 3.11 and Figure 3.12 respectively.

Table 3.10: SSIM, PSNR, MSE and MAE scores ($\mu \pm \sigma$) for the CT paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.

Model	SSIM: Transverse Orientation	PSNR: Transverse Orientation	MSE: Transverse Orientation	MAE: Transverse Orientation
CLADE (no \mathcal{L}_{gmap})	0.894 ± 0.02	28.1 ± 2.95	0.002 ± 0.001	0.03 ± 0.01
CLADE (with \mathcal{L}_{gmap})	0.914 ± 0.01	29.7 ± 3.34	0.001 ± 0.002	0.03 ± 0.02

Table 3.11: PIQUE, NIQE and BRISQUE scores ($\mu \pm \sigma$) for the CT paired networks, applied to synthetically generated paired test datasets. These values are calculated across all slices in the test volumes.

Model	PIQUE: Transverse Orientation	NIQE: Transverse Orientation	BRISUQE: Transverse Orientation
Low-resolution	64.4 ± 6.52	8.31 ± 0.943	46.9 ± 6.24
CLADE (no \mathcal{L}_{gmap})	39.9 ± 5.57	6.51 ± 0.910	31.9 ± 9.11
CLADE (with \mathcal{L}_{gmap})	39.5 ± 7.33	7.05 ± 0.774	33.8 ± 6.65

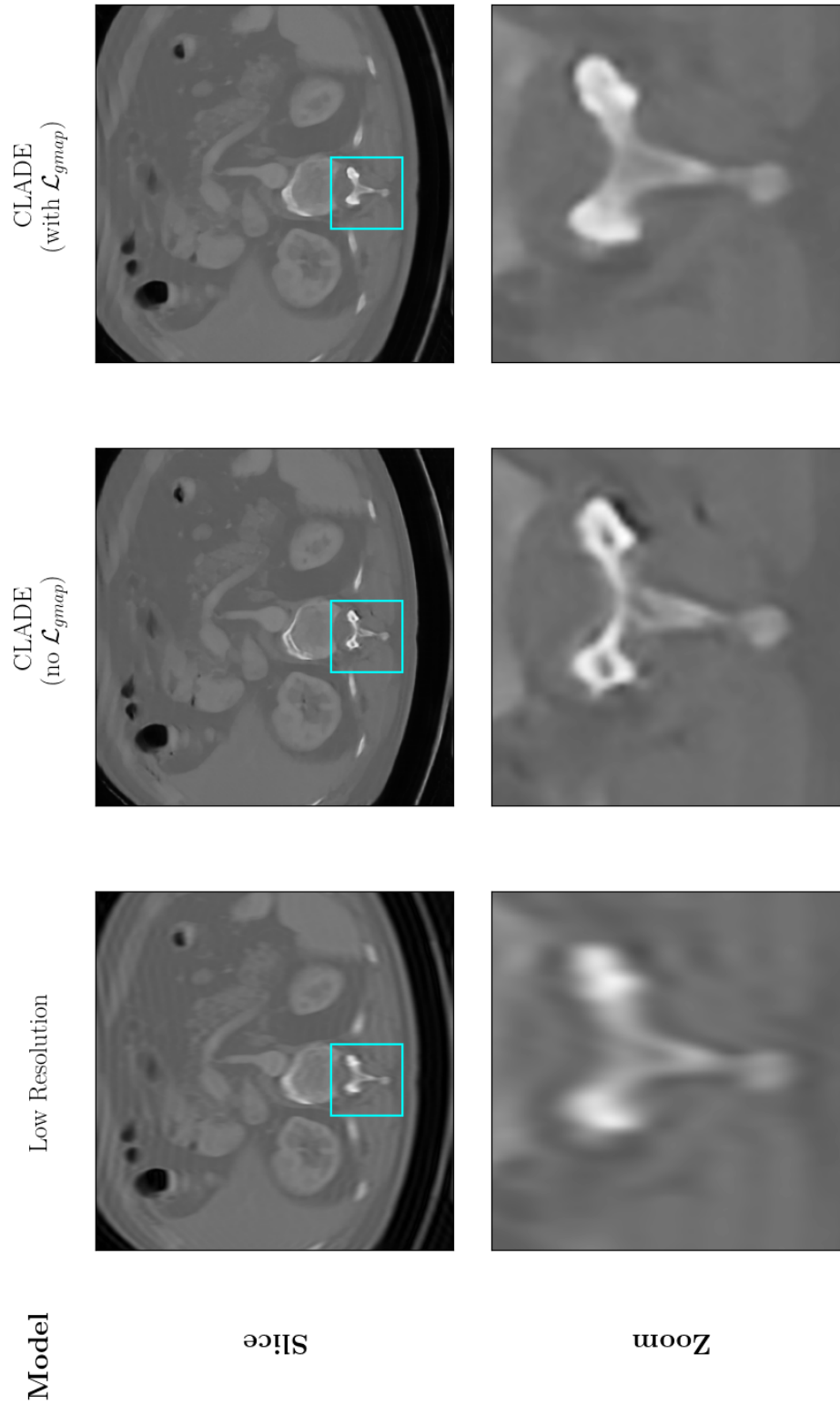


Figure 3.11: CT Paired Network, applied to paired data (in transverse orientation). Here, we can see that the resolution is increased and the reconstructed images do begin to sharpen quite significantly. However, we can see that there are many more artifacts present, alongside some potential hallucinations from the network architectures, particularly in CLADE (no \mathcal{L}_{gmap}).

These results show that the networks achieve good image quality when trained using paired synthetic data. On synthetic test data, the CLADE network with \mathcal{L}_{gmap} , consistently outperforms the supervised CLADE network no \mathcal{L}_{gmap} on synthetic data in terms of SSIM, PSNR, MSE, MAE (and no-reference metrics). These networks were also applied to the prospective low-resolution data, in both the coronal and sagittal planes. The no-reference metrics as calculated across the 15 test data sets are shown in Table 3.12 and images from one of the orientations (sagittal) is displayed in Figure 3.12.

Table 3.12: PIQUE, NIQUE and BRISQUE scores ($\mu \pm \sigma$) for the CT paired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse. Furthermore, SMORE is not included due to its inability to be trained in a paired fashion.

CT Orientation		PIQUE	NIQUE	BRISQUE
Coronal				
	Low-resolution	77.0 ± 4.12	11.9 ± 2.49	33.8 ± 7.16
	CLADE (no \mathcal{L}_{gmap})	47.6 ± 8.00	10.7 ± 2.11	33.8 ± 8.27
	CLADE (with \mathcal{L}_{gmap})	51.4 ± 7.40	11.2 ± 2.20	31.5 ± 6.52
Sagittal				
	Low-resolution	73.9 ± 4.98	11.4 ± 3.42	48.5 ± 7.97
	CLADE (no \mathcal{L}_{gmap})	36.3 ± 9.86	11.6 ± 2.37	40.9 ± 8.98
	CLADE (with \mathcal{L}_{gmap})	39.4 ± 9.35	11.4 ± 3.54	41.5 ± 5.94

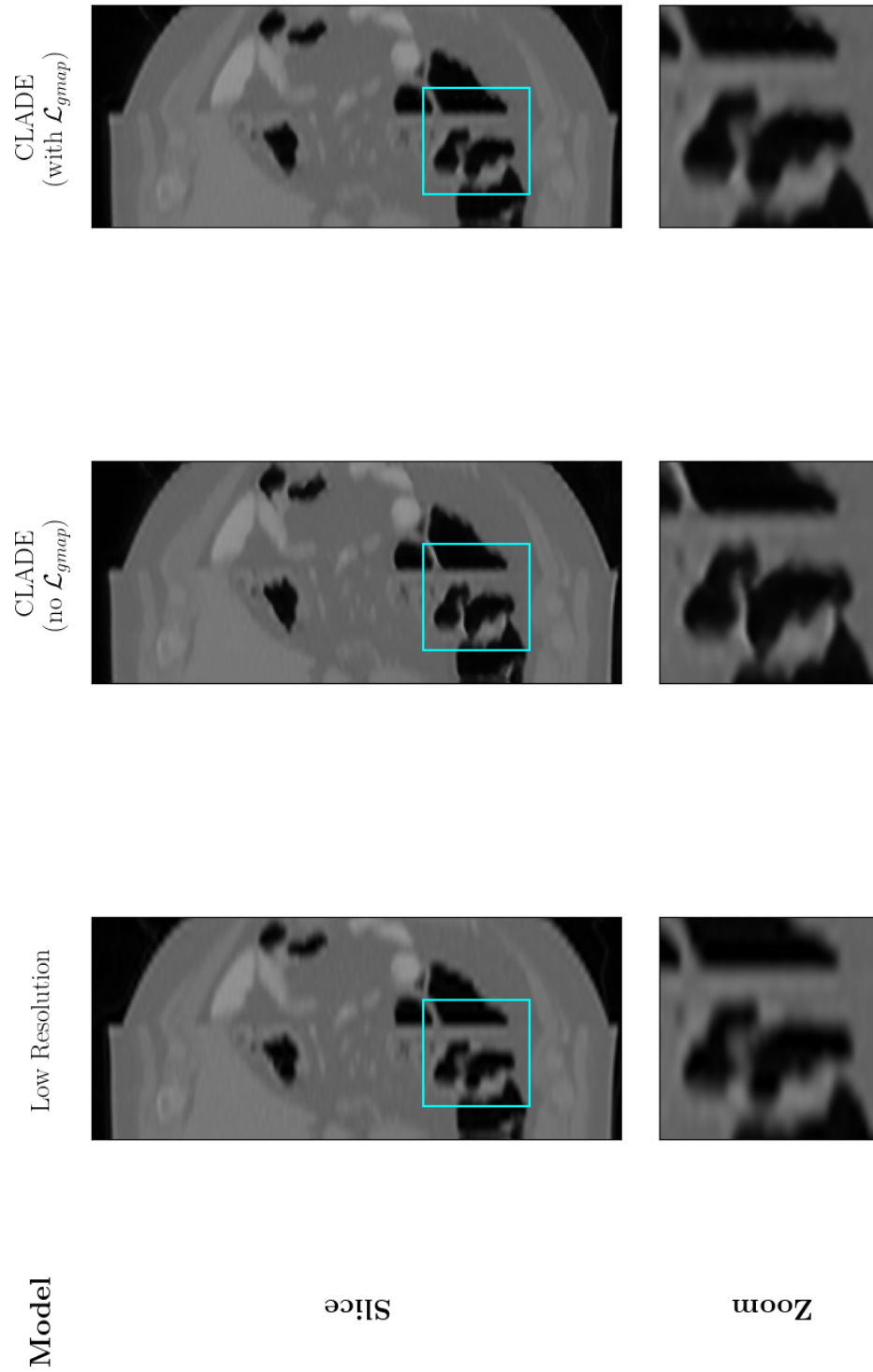


Figure 3.12: CT paired networks, applied to paired data (sagittal). Here, we see that marginal super-resolution is performed by the networks, however there is some clear hallucination in the case of CLADE (no \mathcal{L}_{gmap}), likely as a result of the inefficiency of the downsampling method when generating the synthetic data.

Comparing the no-reference metrics from the synthetic test data and the prospective test data, similarly to the MRI data, it can be seen that the CT image quality from the prospective test data is much lower, which also indicates that the degradation process was not accurately captured by the network from the synthetic datasets. Furthermore, the PIQUE scores in the sagittal view of the prospective dataset look markedly similar to those in the transverse view from the synthetically generated dataset. In contrast however, there is not a drastic visual difference between the synthetic and prospective image quality (see Figure 3.11 and Figure 3.12 respectively). This is likely due to the heuristic nature of the no-reference metrics, which may prioritise different image features more highly than others.

3.4 Unsupervised Training

The Conventional CycleGAN, CLADE (no \mathcal{L}_{gmap}) and CLADE (with \mathcal{L}_{gmap}) networks were trained using unpaired data and compared to SMORE. These networks were subsequently tested on the prospective low-resolution test data. We compared no-reference image metrics PIQUE, NIQE and BRISQUE, as calculated across all 15 test volumes.

3.4.1 MRI Unsupervised Networks

The unsupervised MRI networks are trained from disjoint high-resolution patches taken from the coronal plane, and low-resolution patches taken from the sagittal plane. The no-reference image metrics from the prospective test data are shown in Table 3.13. Comparing these results to the no-reference metrics calculated from the prospective test datasets reconstructed from the supervised networks, it can be seen that the unsupervised networks significantly outperform the supervised networks in terms of PIQUE, NIQE and BRISQUE (with the only exception of the NIQE score for the Conventional CycleGAN). It should also be noted that CLADE significantly outperforms the state-of-the-art SMORE SRR technique, in terms of PIQUE, NIQE and BRISQUE, with CLADE (with \mathcal{L}_{gmap}) consistently outperforming CLADE (no \mathcal{L}_{gmap}) for the MRI test data. The resultant image quality from the final unsupervised networks in both the sagittal and transverse orientations

for the MRI test dataset are displayed in Figure 3.13 and Figure 3.14 respectively.

Table 3.13: PIQUE, NIQUE and BRISQUE scores ($\mu \pm \sigma$) for the MRI unpaired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes.

MRI Orientation		PIQUE	NIQUE	BRISQUE
Transverse				
	Low-resolution	65.0 ± 5.5	8.5 ± 1.4	44.3 ± 5.9
	SMORE	42.1 ± 7.1	6.7 ± 1.5	39.7 ± 5.8
	Conventional CycleGAN	43.5 ± 7.3	12.4 ± 6.5	23.5 ± 15.1
	CLADE (no \mathcal{L}_{gmap})	34.8 ± 8.2	6.5 ± 1.0	31.6 ± 6.9
	CLADE (with \mathcal{L}_{gmap})	29.8 ± 6.3	6.3 ± 1.6	17.9 ± 9.7
Sagittal				
	Low-resolution	63.3 ± 6.0	9.5 ± 2.1	32.0 ± 5.3
	SMORE	40.4 ± 7.6	7.6 ± 2.4	26.8 ± 7.1
	Conventional CycleGAN	34.2 ± 7.5	7.5 ± 2.4	23.6 ± 7.1
	CLADE (no \mathcal{L}_{gmap})	34.2 ± 7.5	7.4 ± 2.3	23.4 ± 6.7
	CLADE (with \mathcal{L}_{gmap})	28.0 ± 5.9	5.6 ± 3.0	19.8 ± 9.6

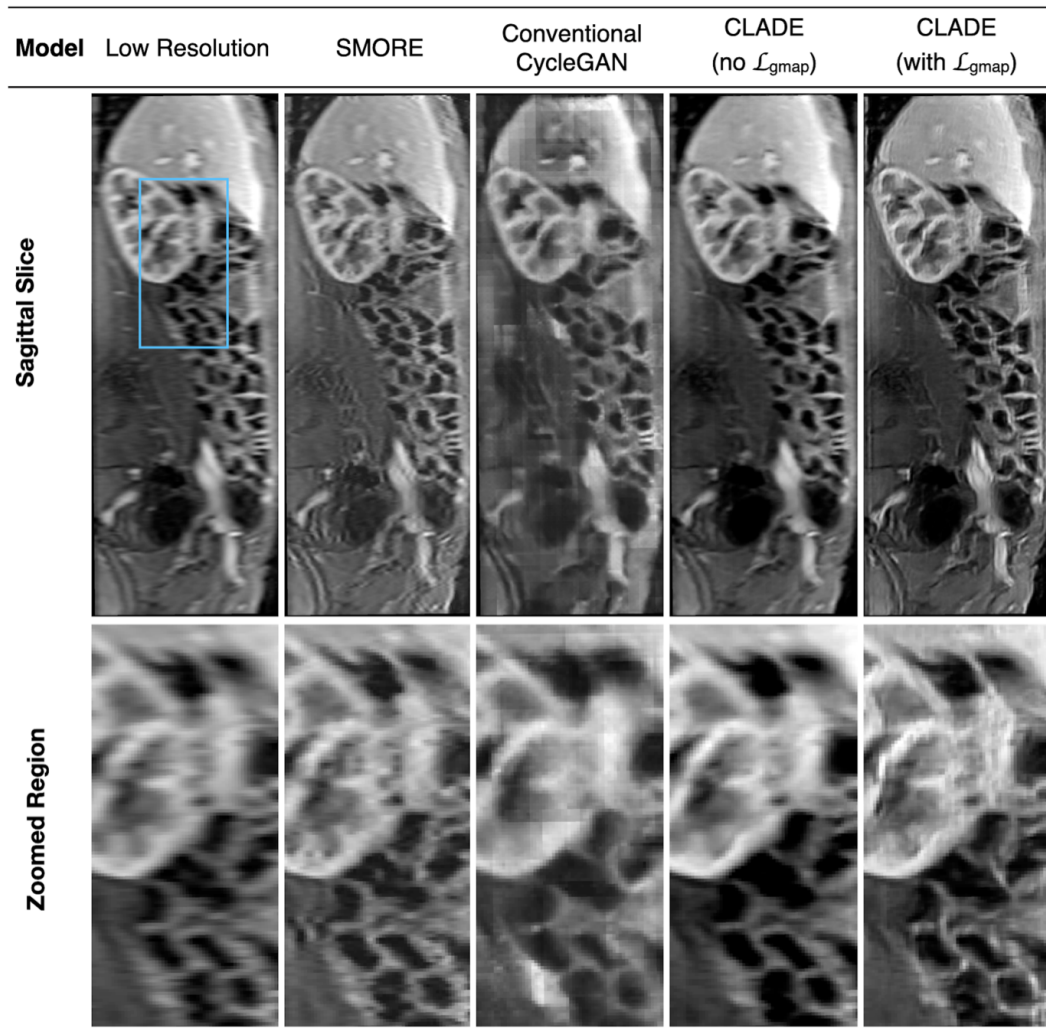


Figure 3.13: An example of sagittal image quality from SRR applied to a low-resolution MRI data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the MRI data. Magnified regions within the blue box are displayed beneath each image.

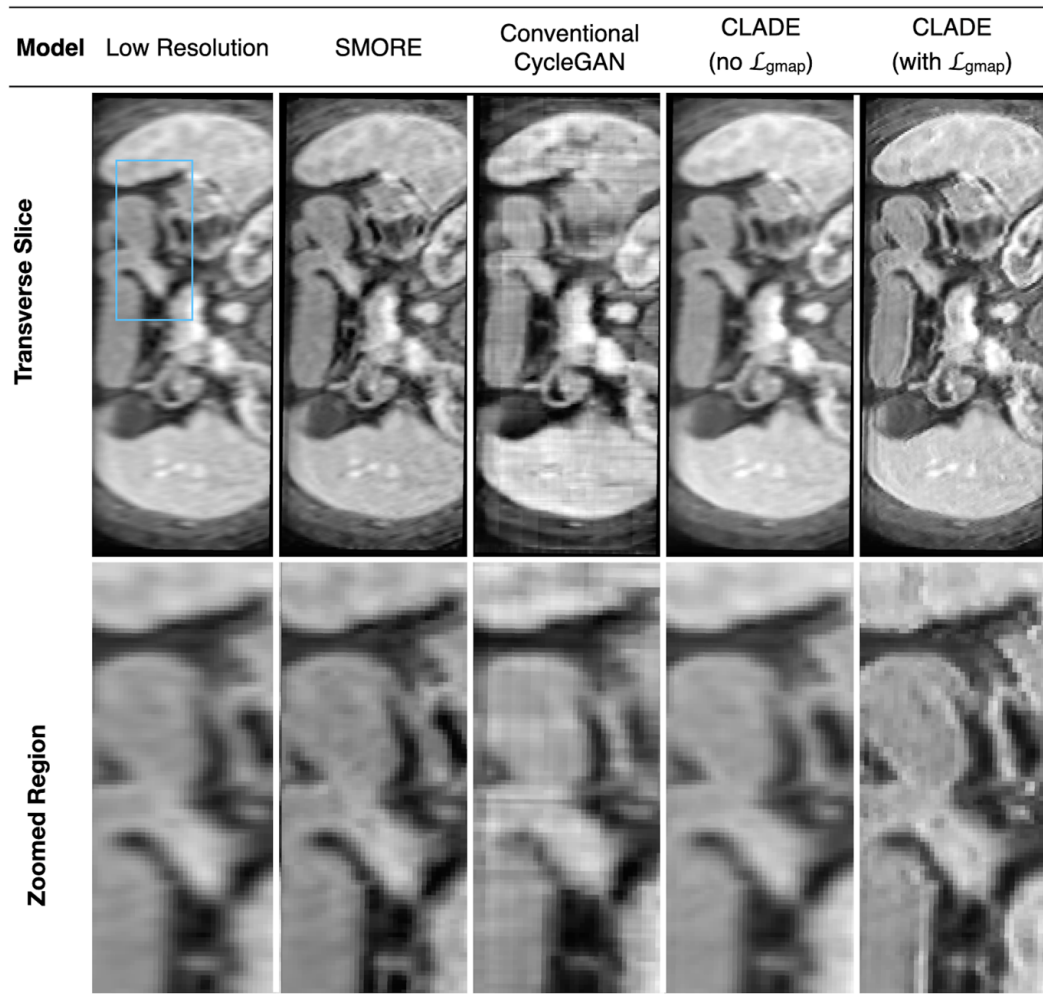


Figure 3.14: An example of transverse image quality from SRR applied to a low-resolution MRI data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the MRI data, where this figure shows the resulting volume reformatted in the (low-resolution) transverse plane. Magnified regions within the blue box are displayed beneath each image.

3.4.2 CT Unsupervised Networks

The unsupervised CT networks are trained from disjoint high-resolution patches taken from the transverse plane, and low-resolution patches taken from the sagittal plane. The no-reference image metrics from the prospective CT test data are shown in Table 3.14.

Table 3.14: PIQUE, NIQUE and BRISQUE scores ($\mu \pm \sigma$) for the CT unpaired networks, applied to the prospective test datasets. These values are calculated across all slices in the test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse.

CT Orientation		PIQUE	NIQUE	BRISQUE
Coronal				
	Low-resolution	77.0 ± 4.1	11.9 ± 2.5	33.8 ± 7.2
	SMORE	56.7 ± 7.4	9.9 ± 1.9	40.8 ± 10.0
	CLADE (no \mathcal{L}_{gmap})	36.0 ± 7.5	8.7 ± 2.0	30.3 ± 7.7
	CLADE (with \mathcal{L}_{gmap})	30.3 ± 7.4	7.5 ± 1.4	27.8 ± 7.5
Sagittal				
	Low-resolution	73.9 ± 5.0	11.4 ± 3.4	48.5 ± 8.0
	SMORE	53.9 ± 9.2	11.1 ± 4.3	45.4 ± 12.4
	CLADE (no \mathcal{L}_{gmap})	26.7 ± 8.4	10.3 ± 4.7	33.7 ± 14.1
	CLADE (with \mathcal{L}_{gmap})	22.3 ± 7.8	8.4 ± 3.2	31.6 ± 9.1

The resultant image quality from the final unsupervised networks in both the sagittal and coronal orientations for the CT test dataset are displayed in Figure 3.15 and Figure 3.16 respectively.

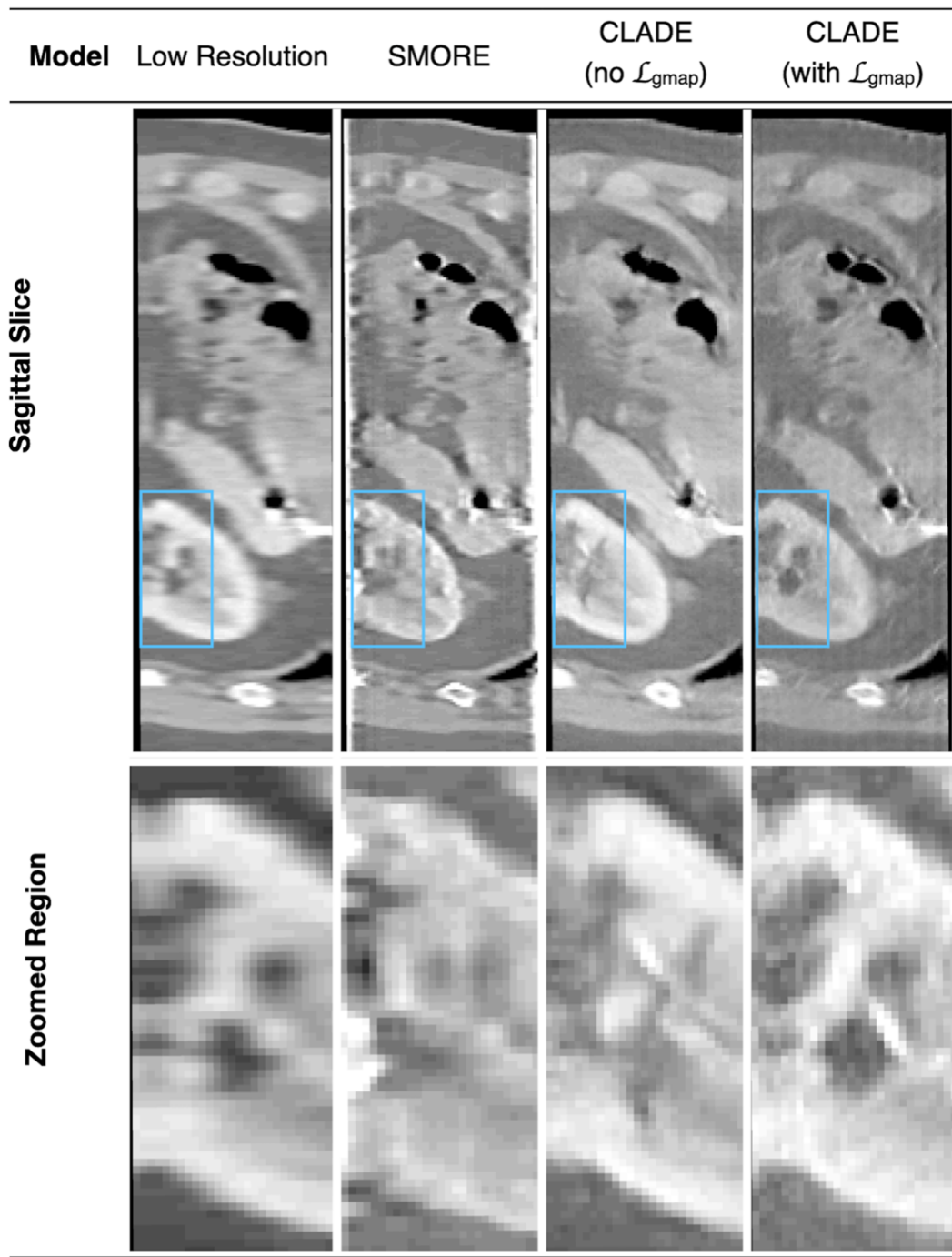


Figure 3.15: An example of sagittal image quality from SRR applied to a low-resolution CT data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the CT data. Magnified regions within the blue box are displayed beneath each image.

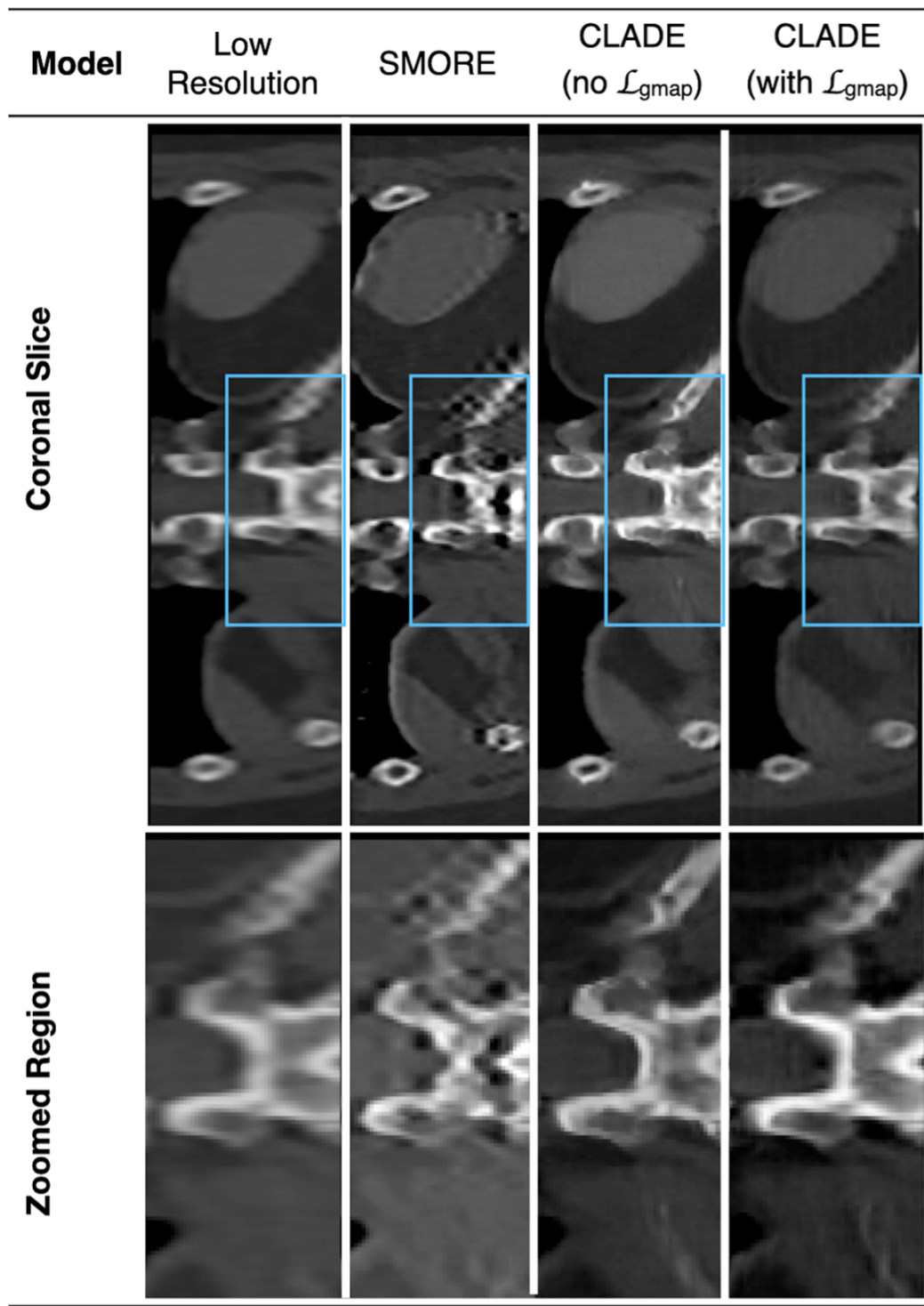


Figure 3.16: An example of coronal image quality from SRR applied to a low-resolution CT data from one subject in the test dataset. Note that the SRR models are applied in the low-resolution sagittal plane for the CT data, where this figure shows the resulting volume reformatted in the (low-resolution) coronal plane. Magnified regions within the blue box are displayed beneath each image.

Similar to the MRI data, it can be seen that compared to the no-reference metrics calculated from the supervised networks, the unsupervised networks significantly outperform the supervised networks in terms of PIQUE, NIQUE and BRISQUE. It should be noted that the Conventional CycleGAN suffered from mode collapse and so these results are not shown here. It should also be noted that CLADE significantly outperforms the state-of-the-art SMORE SRR technique, in terms of PIQUE, NIQUE and BRISQUE, with CLADE (with \mathcal{L}_{gmap}) consistently outperforming CLADE (no \mathcal{L}_{gmap}) for the CT test data.

3.5 Quantitative Image Quality

Quantitative scores comparing image quality of the original low-resolution volumes with DL-SRR using SMORE, conventional CycleGAN, CLADE (no \mathcal{L}_{gmap}) and CLADE (\mathcal{L}_{gmap}) can be found in Table 3.15, and Figure 3.17 and Figure 3.18 for MRI and CT respectively. Quantitative edge sharpness and SNR measurements from all DL-SRR networks are shown in Table 3.16, and Figure 3.17 and Figure 3.18, for MRI and CT respectively. The PIQUE scores for all DL-SRR networks were significantly better than the low-resolution volumes for MRI and CT ($p < 0.05$), with the exception of the conventional CycleGAN for the CT data that did not train. For both MRI and CT, CLADE (\mathcal{L}_{gmap}) produced images with the significantly better PIQUE scores ($p < 0.05$) compared to the other DL-SRR methods including SMORE and CLADE ($\mathcal{L}_{gmap} = 0$). Importantly, these trends were the same for PIQUE scores calculated from both low-resolution planes of the MRI and CT volumes.

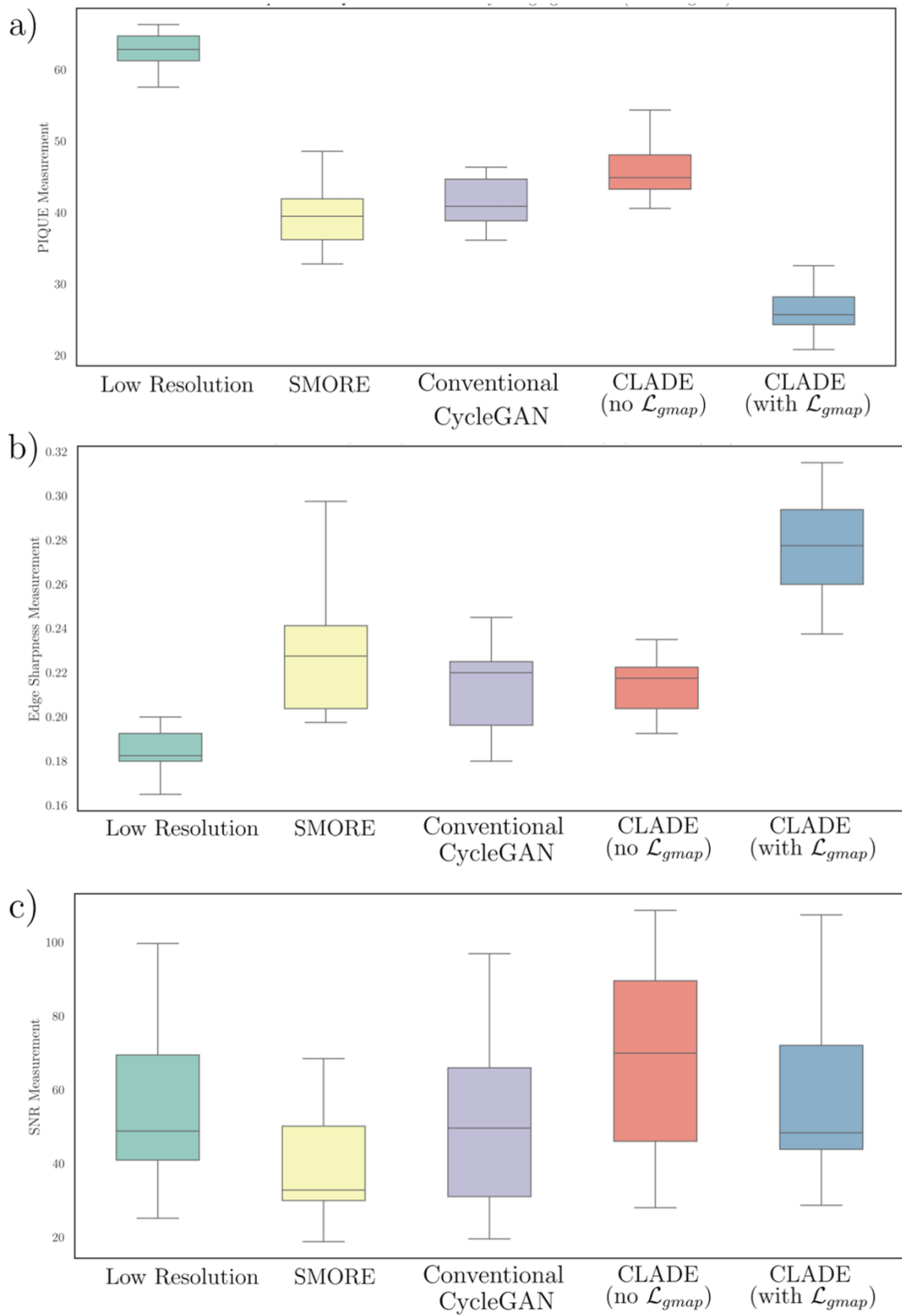


Figure 3.17: Box-plot for MRI Metrics, demonstrating the superiority of CLADE throughout all other comparisons.

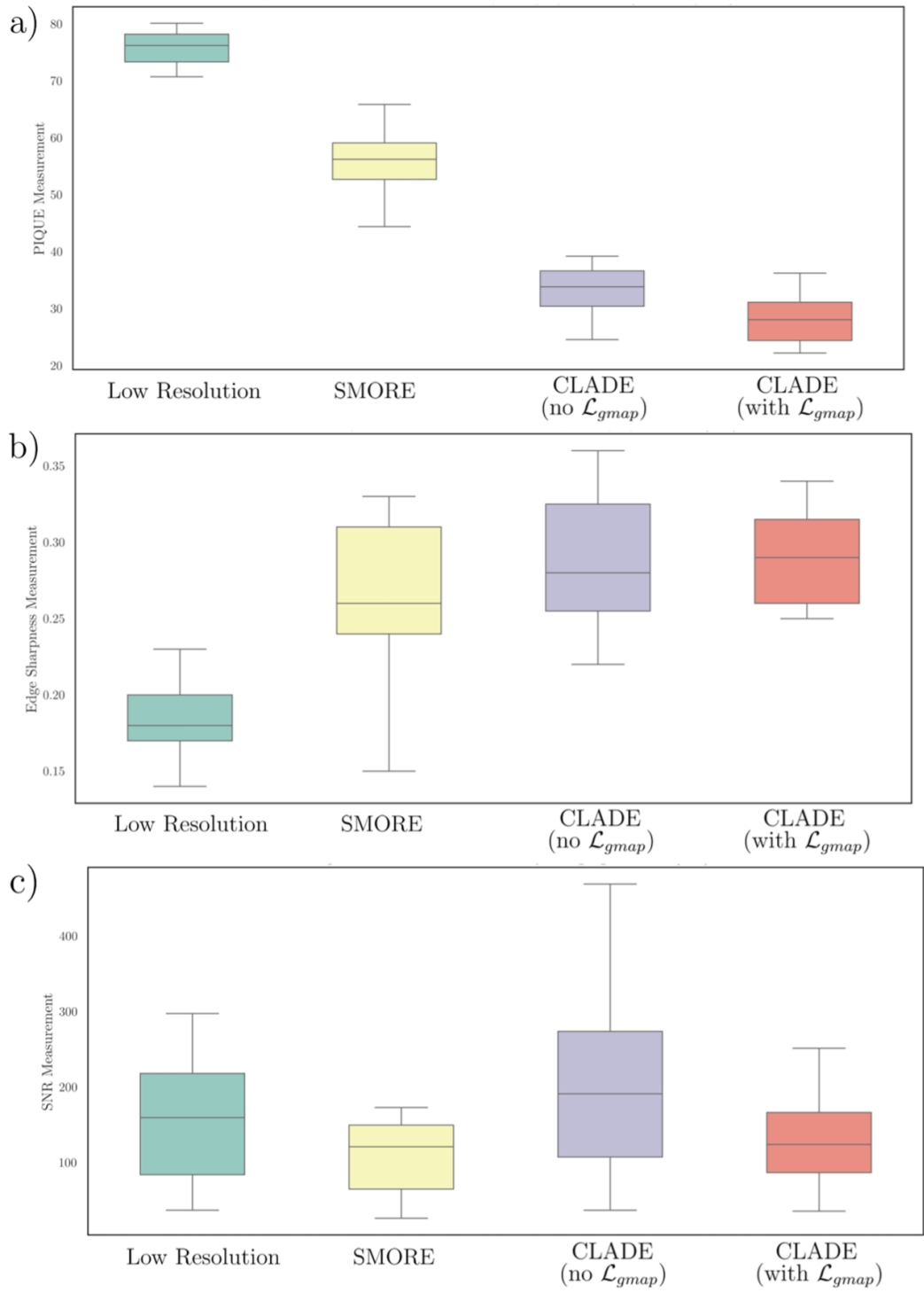


Figure 3.18: Box-plot for CT Metrics, demonstrating the superiority of CLADE throughout all other comparisons.

Table 3.15: PIQUE scores ($\mu \pm \sigma$) as calculated across all slices in the CT and MRI test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse.

Model	MRI PIQUE: Sagittal Orientation**	MRI PIQUE: Transverse Orientation	CT PIQUE: Sagittal Orientation**	CT PIQUE: Coronal Orientation
Low-resolution	62.4 ± 2.8	64.9 ± 3.3	74.0 ± 2.5	75.9 ± 3.0
SMORE	$39.5 \pm 4.5^*$	$41.0 \pm 4.7^*$	$52.2 \pm 5.7^*$	$55.7 \pm 5.5^*$
Conventional CycleGAN	$41.6 \pm 3.3^*$	$46.6 \pm 4.4^{*\#}$	-	-
CLADE (no \mathcal{L}_{gmap})	$45.8 \pm 3.7^{*\#\dagger}$	$51.1 \pm 3.4^{*\#\dagger}$	$29.5 \pm 5.5^{*\#}$	$33.4 \pm 4.3^{*\#}$
CLADE (with \mathcal{L}_{gmap})	$26.6 \pm 3.8^{*\#\dagger\ddagger}$	$28.4 \pm 4.0^{*\#\dagger\ddagger}$	$24.0 \pm 4.9^{*\#\dagger\ddagger}$	$28.2 \pm 4.5^{*\#\dagger\ddagger}$

** Denotes the orientation that the SRR was applied to.

* Denotes statistical differences from low-resolution data ($p < 0.05$).

Denotes statistical differences from SMORE ($p < 0.05$).

† Denotes statistical differences from conventional CycleGAN ($p < 0.05$).

‡ Denotes statistical differences from CLADE (no \mathcal{L}_{gmap}) ($p < 0.05$).

Table 3.16: Quantitative edge sharpness and estimated SNR ($\mu \pm \sigma$) as calculated across all slices in the CT and MRI test volumes. Note that the conventional CycleGAN is not included in the CT results due to model collapse.

Model	Edge Sharpness (mm^{-1})	Signal	Noise	SNR
MRI				
Low-resolution	0.19 ± 0.04	154.5 ± 34.2	3.1 ± 1.2	55.8 ± 21.5
SMORE	$0.23 \pm 0.05^*$	132.5 ± 46.0	3.6 ± 1.5	39.5 ± 15.4
Conventional CycleGAN	$0.21 \pm 0.05^*$	147.6 ± 46.4	3.3 ± 1.7	56.6 ± 32.7
CLADE (no \mathcal{L}_{gmap})	$0.22 \pm 0.04^*$	163.8 ± 39.8	2.6 ± 1.2	$73.9 \pm 35.3^\#$
CLADE (with \mathcal{L}_{gmap})	$0.28 \pm 0.04^{*\#\dagger\ddagger}$	162.5 ± 36.6	3.2 ± 1.3	56.7 ± 21.9
CT				
Low-resolution	0.19 ± 0.03	167.0 ± 53.5	1.5 ± 1.2	153.6 ± 81.5
SMORE	$0.26 \pm 0.05^*$	$101.4 \pm 41.5^*$	1.1 ± 0.9	118.2 ± 64.4
CLADE (no \mathcal{L}_{gmap})	$0.29 \pm 0.04^*$	$167.1 \pm 50.4^\#$	1.7 ± 1.2	196.4 ± 124.3
CLADE (with \mathcal{L}_{gmap})	$0.29 \pm 0.03^*$	$167.3 \pm 53.8^\#$	1.4 ± 1.4	130.6 ± 68.4

* Denotes statistical differences from low-resolution data ($p < 0.05$).

Denotes statistical differences from SMORE ($p < 0.05$).

† Denotes statistical differences from conventional CycleGAN ($p < 0.05$).

‡ Denotes statistical differences from CLADE (no \mathcal{L}_{gmap}) ($p < 0.05$).

The edge sharpness was significantly higher for all DL-SRR networks compared to the low-resolution volumes ($p < 0.05$) for both MRI and CT data. In MRI, CLADE (\mathcal{L}_{gmap}) was found to have a significantly higher edge sharpness than SMORE, the conventional CycleGAN and CLADE (no \mathcal{L}_{gmap}). No significant differences were found in the estimated SNR between the low-resolution data and any of the DL-SRR networks in either the MRI or CT data. All networks had similar

estimated SNR, with the only significant difference found between MRI CLADE (no \mathcal{L}_{gmap}) and MRI SMORE (due to higher signal and lower noise measurements in CLADE (no \mathcal{L}_{gmap})).

Chapter 4

Discussion

A critical aspect in evaluating CLADE’s clinical applicability involves validation against established gold standards. Ideally, super-resolution reconstruction methods should be benchmarked against high-resolution ground-truth images acquired using high-fidelity imaging protocols. However, obtaining true high-resolution isotropic datasets in medical imaging often proves challenging due to scan time constraints and patient motion. Despite these limitations, several validation approaches could strengthen the robustness of our proposed method.

As included in this thesis, the most direct validation method would involve acquiring isotropic high-resolution 3D MRI or CT scans and artificially downsampling them to simulate the low-resolution anisotropic data used for CLADE training. The reconstructed images could then be compared to the original high-resolution scans using quantitative metrics such as Peak Signal-to-Noise Ratio, Structural Similarity Index, and Mean Absolute Error. This approach has been successfully implemented in brain imaging studies, where ultra-high-field MRI serves as a reference standard compared to standard field strength imaging. Similarly, for CT, reference data could be derived from high-resolution acquisitions from cadaveric imaging datasets, phantom-based images or dedicated high-dose scans.

Building on these quantitative assessments, another robust validation approach involves utilizing longitudinal patient imaging data where high-resolution follow-up scans exist. In cases where lesions are detected at later stages on high-resolution scans, their earlier presence could be assessed retrospectively in CLADE-enhanced

images from previous low-resolution scans. This would demonstrate whether CLADE enables earlier and more accurate lesion detection in clinical practice.

The main limitation to these approaches, however, is the lack of support for evaluation of images that contain tissue motility, which would render use in anatomical sites such as the abdomen or heart less robust. To address this challenge, in the absence of paired high-resolution scans, expert radiologist assessments could offer an alternative gold standard. Radiologists could review CLADE-enhanced images alongside original low-resolution scans, evaluating them based on diagnostic confidence, anatomical clarity, and lesion detectability. Blind studies comparing the clinical usability of CLADE-enhanced images to standard scans could provide real-world validation of its effectiveness in clinical settings.

While initially demonstrated on abdominal MRI and CT datasets, CLADE's principles can extend to other anatomical regions given appropriate training data. Different anatomical regions present unique challenges and require specific validation standards. Brain imaging validation could involve ultra-high-field MRI or histological validation using post-mortem imaging studies, along with comparisons to diffusion-weighted imaging data to assess white matter microstructure preservation. For cardiac imaging, high-resolution cine MRI or 4D flow imaging could serve as gold standards, with additional comparisons to contrast-enhanced CT or late gadolinium enhancement MRI for scar and ischemia detection. Musculoskeletal imaging validation might utilize micro-CT or ultra-high-field MRI of joints and cartilage to assess fine structural preservation, while thoracic imaging could be validated against high-resolution lung CT from dedicated thin-slice protocols, with additional comparisons to functional imaging modalities to validate anatomical feature preservation. These expanded validation strategies would not only strengthen CLADE's clinical applicability but also pave the way for its potential integration into routine clinical workflows across multiple medical domains and imaging modalities.

Chapter 5

Conclusion

In conclusion, this thesis presents CLADE, a novel unsupervised patch-based method for super-resolving anisotropic medical images, evaluated on true anisotropic datasets from both MRI and CT scans. One of the key strengths of CLADE is its versatility across different imaging modalities, including MRI and CT, without requiring any supervised guidance. This is of significant benefit in medical imaging, where obtaining paired high-resolution and low-resolution datasets is often infeasible due to limitations in data acquisition. By operating in a completely unsupervised fashion, CLADE eliminates the need for paired datasets, making it particularly valuable in clinical contexts where such data is scarce. The model's ability to generate high-resolution images from low-resolution inputs without paired examples opens new possibilities for improving diagnostic imaging.

A distinctive feature of CLADE is its approach to learning a resolution mapping between disjoint patches from the low-resolution and high-resolution planes, enabling effective super-resolution of the entire image. Unlike traditional methods that require supervision through paired patches, CLADE operates without this constraint, allowing it to independently super-resolve patches and reconstruct the full image without explicit pairing information. This disjoint patch-based approach enhances the model's flexibility, consistently outperforming existing methods such as SMORE and CycleGAN in terms of image quality, as demonstrated in our experiments.

The incorporation of a weight demodulation process further strengthens

CLADE by stabilising training and reducing visual artifacts and normalisation errors that typically arise during patch reconstruction. These issues, which may be prominent in conventional patch-based architectures like the CycleGAN, are effectively addressed via the weight demodulation process, resulting in more accurate and visually coherent image reconstructions.

Additionally, the introduction of a gradient-mapping loss played a critical role in enhancing both the sharpness and perceptual quality of the super-resolved images. By aligning edge gradients between the high-resolution input and the super-resolved low-resolution output, this loss contributed to improved PIQUE scores and sharper visual results. However, while effective, this loss is not without limitations; it acts solely on image gradients, without accounting for underlying motion or noise, which in some cases could lead to the sharpening of unwanted features or artifacts.

Despite the promising results, there are several limitations to this study. A key challenge was the difficulty in quantitatively assessing image quality due to the lack of paired training data in anisotropic medical scans, which inherently do not have ground-truth comparisons. While no-reference image quality metrics such as PIQUE, NIQUE, and BRISQUE were used, these metrics remain somewhat heuristic. A more robust evaluation would require expert clinician scoring, which was logistically unfeasible for this study given its broad experimental scope. Moreover, while CLADE demonstrated superior performance over existing models, this study did not compare it against some newer architectures due to time constraints and the unavailability of source code. Future work should address this by incorporating comparisons to more recently developed models. Another limitation was the absence of task-specific evaluations, such as image registration, detection, or segmentation, which would require high-resolution isotropic data for ground-truth comparison. Moving forward, it will be important to assess CLADE’s performance on such tasks with access to high-resolution isotropic 3D data.

Additionally, CLADE does not explicitly handle motion artifacts, such as ghosting in MRI scans, which may be exacerbated during the super-resolution process. Future work should investigate techniques for concurrent motion artifact

suppression to further enhance image quality. A broader evaluation across larger datasets, sourced from multiple sites, vendors, and imaging contrasts, will also be critical to validate the method’s generalisability under domain shift, although the successful application across both MRI and CT data suggests strong potential for wide applicability. Finally, the current approach trained CLADE using patches from a single low-resolution dimension. Future research should explore training with disjoint patches from both low-resolution dimensions simultaneously to assess whether this leads to further improvements.

In summary, while CLADE introduces an effective and flexible method for the super-resolution of anisotropic medical images, there remain several avenues for future exploration, including task-specific evaluations, motion artifact suppression, and testing on larger, more diverse datasets. Addressing these will be key to refining and expanding CLADE’s utility across various medical imaging applications.

Bibliography

- [1] Michele Pascale, Vivek Muthurangu, Javier Montalt Tordera, Heather E Fitzke, Gauraang Bhatnagar, Stuart Taylor, and Jennifer Steeden. CLADE: Cycle Loss Augmented Degradation Enhancement for Unpaired Super-Resolution of Anisotropic Medical Images. *arXiv preprint arXiv:2303.11831*, 2023.
- [2] Michele Pascale, Vivek Muthurangu, and Steeden Jennifer. Unpaired Super-Resolution for Anisotropic MRI using a Gradient Mapping Adversarial Loss. In *ISMRM Proceedings 2023, Digital Poster, Toronto*, 2023.
- [3] Michele Pascale, Vivek Muthurangu, and Steeden Jennifer. CycleGAN-based Degradation Enhancement for Unpaired Super-Resolution of Anisotropic Medical Images. In *ISMRM Proceedings 2024, Digital Poster, Singapore*, 2024.
- [4] Robert W. Brown, Yu-Chung N. Cheng, E. Mark Haacke, Michael R. Thompson, and Ramesh Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley, April 2014.
- [5] Noortje Regensburg. Volumetric Measurements in Graves’ Orbitopathy. *Thesis, Universiteit van Amsterdam*, 2010.
- [6] Mahmood Qureshi, Muhammad Kaleem, and Hammad Omer. Journey Through K-space: an Interactive Educational Tool. *Biomedical Research*, 28(4):7, 2017.

- [7] Radiology Key. Chapter 16, Computed Tomography, 2016. [Online; accessed April 17, 2024].
- [8] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution Image Reconstruction: a Technical Overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003.
- [9] Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. Survey: Interpolation Methods in Medical Image Processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–1075, 1999.
- [10] Alexander Selvikvåg Lundervold and Arvid Lundervold. An Overview of Deep Learning in Medical Imaging Focusing on MRI. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.
- [11] Abien Fred Agarap. Deep Learning Using Rectified Linear Units (ReLU). *arXiv preprint arXiv:1803.08375*, 2018.
- [12] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.
- [13] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional Neural Networks: an Overview and Application in Radiology. *Insights into Imaging*, 9:611–629, 2018.
- [14] Ian Goodfellow et al. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27, 2014.
- [15] Jun-Yan Zhu et al. Unpaired Image-to-image Translation Using Cycle-consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
- [16] Cuong-Hai Pham et al. Brain MRI Super-resolution Using Deep 3D Convolutional Networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017.

- [17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely Connected Convolutional Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.
- [18] Yanchun Chen et al. Brain MRI Super Resolution Using 3D Deep Densely Connected Neural Networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018.
- [19] Jiayi Du et al. Super-resolution Reconstruction of Single Anisotropic 3D MR Images Using Residual Convolutional Neural Network. *Neurocomputing*, 392:209–220, 2020.
- [20] Qian Wu et al. An Arbitrary Scale Super-resolution Approach for 3D MR Images Via Implicit Neural Representation. *IEEE Journal of Biomedical and Health Informatics*, 27(2):1004–1015, 2022.
- [21] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, 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*, pages 4681–4690, 2017.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [23] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [24] Pourya Shamsolmoali et al. G-GANISR: Gradual Generative Adversarial Network for Image Super Resolution. *Neurocomputing*, 366:140–153, 2019.

- [25] Seung-Jun Park et al. SR-FEAT: Single Image Super-resolution with Feature Discrimination. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [26] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang 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*, pages 701–710, 2018.
- [27] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced Deep Residual Networks for Single Image Super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017.
- [28] Antonin Chambolle. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.
- [29] Ozan Oktay et al. Multi-input Cardiac Image Super-resolution Using Convolutional Neural Networks. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III*. Springer, 2016.
- [30] Jiale Wang, Runze Wang, Rong Tao, and Guoyan Zheng. UASSR: Unsupervised Arbitrary Scale Super-resolution Reconstruction of Single Anisotropic 3D Images Via Disentangled Representation Learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 453–462. Springer, 2022.
- [31] Chao Zhao et al. SMORE: a Self-supervised Anti-aliasing and Super-resolution Algorithm for MRI Using Deep Learning. *IEEE Transactions on Medical Imaging*, 40(3):805–817, 2021.
- [32] Sean W Remedios et al. *Self-supervised Super-resolution for Anisotropic MR Images with and Without Slice Gap*. Springer Nature, 2023.

- [33] Shuo Han, Samuel Remedios, Aaron Carass, Michael Schär, and Jerry L Prince. MR Slice Profile Estimation by Learning to Match Internal Patch Distributions. In *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27*, pages 108–119. Springer, 2021.
- [34] Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xinchao Wang, and Thomas Huang. Wide Activation for Efficient and Accurate Image Super-resolution. *arXiv preprint arXiv:1808.08718*, 2018.
- [35] Mikael Gladh and Daniel Sahlin. Image Synthesis Using CycleGAN to Augment Imbalanced Data for Multi-class Weather Classification. *Thesis, Linköping University*, 2021.
- [36] Tero Karras et al. Analyzing and Improving the Image Quality of StyleGAN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [37] Tim Salimans and Durk P Kingma. Weight Normalization: a Simple Reparameterization to Accelerate Training of Deep Neural Networks. *Advances in Neural Information Processing Systems*, 29, 2016.
- [38] Phillip Isola et al. Image-to-image Translation with Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [39] Jie Liu et al. Unsupervised Representation Learning for 3D MRI Super Resolution with Degradation Adaptation. *arXiv preprint arXiv:2205.06891*, 2022.
- [40] Zhi Lu and Ying Chen. Single Image Super-resolution Based on a Modified U-Net with Mixed Gradient Loss. *Signal, Image and Video Processing*, pages 1–9, 2022.

- [41] Wen Luo et al. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016.
- [42] Neil M Rofsky et al. Abdominal MR Imaging with a Volumetric Interpolated Breath-hold Examination. *Radiology*, 212(3):876–884, 1999.
- [43] Ke Yan et al. DeepLesion: Automated Mining of Large-scale Lesion Annotations and Universal Lesion Detection with Deep Learning. *Journal of Medical Imaging*, 5(3):036501–036501, 2018.
- [44] Diederik P Kingma and Jimmy Ba. ADAM: a Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [45] Narasimhan Venkatanath et al. Blind Image Quality Evaluation Using Perception Based Features. In *2015 Twenty First National Conference on Communications (NCC)*. IEEE, 2015.
- [46] Zhou Wang and Alan C Bovik. A Universal Image Quality Index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.
- [47] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [48] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind Image Quality Evaluation Using Perception Based Features. In *2015 Twenty First National Conference on Communications (NCC)*, pages 1–6. IEEE, 2015.
- [49] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely Blind” Image Quality Analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012.

- [50] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.
- [51] Antoine Rosset, Luca Spadola, and Osman Ratib. Osirix: an Open-source Software for Navigating in Multidimensional Dicom Images. *Journal of Digital Imaging*, 17:205–216, 2004.
- [52] Jennifer A Steeden et al. Rapid Whole-heart CMR with Single Volume Super-resolution. *Journal of Cardiovascular Magnetic Resonance*, 22(1):1–13, 2020.