Anisotropic to Isotropic Reconstruction of Infant Brain MRI using Deep Learning

Alma Lennartsson

Master's thesis
2026:E7

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematics

Centrum Scientiarum Mathematicarum

LTH
FACULTY OF
ENGINEERING

Anisotropic to Isotropic Reconstruction of Infant Brain MRI using Deep Learning

---

Master's Thesis in Biomedical Engineering

Alma Lennartsson

January, 2026

Supervisors: Jacob Vogel, Linda Karlsson, Karl Åström
Examiner: Mikael Nilsson

DeMON Lab
Faculty of Engineering LTH
Centre for Mathematical Sciences

Abstract

Brain magnetic resonance imaging (MRI) is an important tool for diagnosing neurological diseases and is widely used in research to study brain development and aging. However, acquiring high-resolution (HR) isotropic MRI is time- and resource-consuming, which often leads to acquisition of low-resolution (LR) anisotropic MRI instead. In recent years, deep learning has shown promising advancements in synthesizing isotropic MRI from anisotropic scans. However, the models are often trained for certain tasks, and may need re-training or re-optimization to perform well in other scenarios. Consequently, this thesis aims to construct a deep learning pipeline specifically optimized for anisotropic to isotropic reconstruction of T2-weighted (T2w) infant brain MRI. The proposed approach is a supervised, 3D patch-based residual U-Net architecture, which has been carefully optimized in terms of network depth, patch size, input channels, residual units, and data augmentation. The final model outperforms a baseline interpolation-based framework, both with visual assessment and quantitative measures. However, the network presents limited generalization to unseen types of LR anisotropic input. To address this limitation, augmentation is introduced by adding more types of LR anisotropic MRI to the training data. This shows promising results and can, with further development, be utilized for research use, and eventually in clinical practice. In conclusion, this work successfully implements a deep learning pipeline for anisotropic to isotropic reconstruction in T2w infant brain MRI through systematic optimization of model hyperparameters.

Acknowledgements

I would like to express my deepest gratitude to my on-site supervisors Jacob Vogel and Linda Karlsson, for all the invaluable help and support during my project. It has been truly inspiring to work with you during this term and I have learned so much.

Secondly, I want to thank all the members of the DeMON Lab for the warm welcoming to the group. Thank you for always helping me out and for all the fun breaks, lunches and after works.

I also want to thank my supervisor at LTH, Kalle Åström, for guidance and support during the whole project, and my examiner Mikael Nilsson for valuable inputs, as well as Jakob Seidlitz who suggested this project and provided consistent assistance ever since.

Lastly, I want to thank my dear friends from the Biomedical Engineering program; Edvin Andersson, Hilma Bettinger, Emma Friberg, Ebba Knubbe and Vilma Sterner. It has been wonderful spending the past five years with you. And thank you to my family; Alice Lennartsson, Pernilla Ekberg Lennartsson and Ola Lennartsson, for always supporting me.

Sammanfattning

Användning av artificiell intelligens för att förbättra kvaliteten på MR-bilder av spädbarnshjärnan

Kliniska MR-bilder har ofta låg upplösning på grund av tids- och resursbrist på sjukhus. I det här projektet undersöks om artificiell intelligens kan användas för att förbättra upplösningen av MR-bilder - ett sätt att underlätta för både forskning och diagnostik.

Magnetkameran (MR) är ett viktigt verktyg för att studera hjärnan och diagnostisera neurologiska sjukdomar. Tyvärr finns det vissa svårigheter med MR. För det första är det resurskrävande, vilket innebär att sjukvården måste begränsa användningen, och för det andra krävs långa tider i MR-maskinen för att få bra bilder. Långa bildtagningstider resulterar i att patienten hinner röra på sig mer i maskinen, vilket introducerar rörelseartefakter i bilderna. Detta skapar en konstant balansgång inom sjukvården, där man strävar efter att minimera tiden i maskinen och samtidigt maximera bildkvaliteten.

Begränsningarna med MR är särskilt framträdande när det kommer till spädbarn, eftersom det inte finns något enkelt sätt att försäkra sig om att barnet är stilla under bildtagning. Det har lett till att man låter barnet vara kortare tid i MR-maskinen, vilket resulterar i lågupplösta bilder som är svåra att använda inom både forskning och diagnostik.

Därför utreder vi i det här projektet om det finns något sätt att förbättra MR-bilder av just spädbarnshjärnor. Idén är att använda artificiell intelligens för att rekonstruera den lågupplösta MR-bilden till en högupplöst bild. Detta görs genom att träna en djupinlärningsmodell på flera MR-bilder av spädbarnshjärnor. Under

träningen får modellen se både lågupplösta bilder och deras högupplösta motsvarigheter. Målet är att modellen till slut ska lära sig att självständigt skapa en korrekt högupplöst bild baserat enbart på den lågupplösta informationen.

Men fungerar det? För att utvärdera resultatet gav vi modellen olika typer av lågupplösta testbilder som den fick prova att rekonstruera. Det visade sig att den lyckades med sin uppgift om testbilden hade liknande grad av låg upplösning som bilderna den var tränad på. Men om testbilden, till exempel, hade ännu lägre upplösning än de bilder som användes under träningen, var det svårt för modellen att lyckas.

Slutsatsen från det här projektet är att det går att använda artificiell intelligens, alltså djupinlärningsmodeller, för att rekonstruera lågupplösta MR-bilder av spädbarnshjärnor till högupplösta. För att modellen ska fungera mer generellt krävs dock att den tränas på fler variationer av lågupplösta bilder.

På sikt kan vår modell få praktisk användning inom både forskning och sjukvård. Genom att rekonstruera kliniska lågupplösta MR-bilder till högupplösta kan forskare få bättre möjligheter att studera hur spädbarnshjärnan utvecklas. Det finns även potential för framtida tillämpningar inom sjukvården, där förbättrad bildkvalitet kan bidra till säkrare och mer effektiv diagnostik.

Acronyms and Abbreviations

MRI	Magnetic Resonance Imaging
HR	High-Resolution
LR	Low-Resolution
SR	Super-Resolution
T1w	T1-weighted
T2w	T2-weighted
GT	Ground Truth
CNN	Convolutional Neural Network
ReLU	Rectified Linear Unit
LReLU	Leaky Rectified Linear Unit
PReLU	Parametric Rectified Linear Unit
MSE	Mean Square Error
MAE	Mean Absolute Error
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
LPIPS	Learned Perceptual Image Patch Similarity
MONAI	Medical Open Network for AI
GAN	Generative Adversarial Network

Contents

Chapter 1

Introduction

1.1 Background

1.1.1 Brain MRI

Magnetic resonance imaging (MRI) is a non-invasive imaging technique that uses magnetic fields and radio frequency pulses to generate images of soft tissues in the human body. In clinical practice, brain MRI is one of the most important tools for diagnosing neurological diseases and detecting brain abnormalities. Furthermore, the information provided from the MRI helps to evaluate disease progression or plan appropriate treatment strategies, such as surgical interventions. In research settings, MRI is used to study brain development, aging, and neurodegenerative processes.

During MRI acquisition, the image contrast can be adjusted to enhance specific parts of the brain. The contrasts are known as image modalities, and can be acquired by altering physical parameters such as echo time or repetition time. The most used modalities for brain MRI are T1-weighted (T1w) and T2-weighted (T2w).

T1w MRI highlights fatty tissue and excels in visualizing anatomical structure. This makes T1w MRI good for detecting structural abnormalities such as tumors or atrophy, making it useful to diagnose Alzheimer's disease, multiple sclerosis and cancer [1]. Due to the anatomical sharpness, T1w MRI can also be used for several downstream tasks such as brain volumetry, tissue segmentation, cortical thickness estimation, and for guiding surgical planning [1].

T2w MRI, on the other hand, enhances signals from watery tissues. Cere-

brospinal fluid appears bright, making T2w images particularly useful for identifying edema, inflammation, and abnormalities involving increased fluid content in the brain [1].

1.1.2 MRI Resolution

MRI information is acquired from three orthogonal orientations: axial (going from the top of the head to the bottom), coronal (going from front of the head to the back), and sagittal (going from ear to ear). Together, these acquisitions form a three-dimensional (3D) representation of the brain, which can be visualized from all spatial directions. The spatial resolution of an MRI image is determined by the voxel size, typically measured in millimeters (mm). An MRI volume with equal resolution in all three spatial dimensions, i.e. composed of cubic voxels, is referred to as isotropic and provides consistent image quality along all axes. In clinical practice, MRI scans are often acquired with anisotropic resolution, where one spatial dimension has lower resolution than the others. The high-resolution (HR) dimensions are referred to as the in-plane resolution, while the low-resolution (LR) dimension is known as the through-plane resolution. This can be interpreted as imaging the brain in a series of slices, where each slice lies in the in-plane dimensions and the through-plane resolution corresponds to the slice thickness.

Due to differences in acquisition times, isotropic T1w MRI is more common than isotropic T2w MRI. T1w MRI is faster to acquire, since it has shorter echo- and repetition times. Hence, the clinical standard is, in many cases, to acquire LR anisotropic T2w MRI and HR isotropic T1w MRI [2].

1.1.3 Challenges

There are several constraints for brain MRI, especially when used in clinics. To receive HR isotropic images, very long scanning times are acquired. Besides being resource-consuming, long scanning times may also result in motion corruptions from the movement of the patient. Therefore, clinicians acquire anisotropic MRI, which reduces the scan duration. In a resource rich setting, the optimal choice would be a HR isotropic 3D volume available for assessment in all directions, which could enable better diagnosis and understanding of the brain structure. Furthermore, HR isotropic MRI enables several downstream tasks that are dependent on isotropic

3D volumes, such as brain volumetry and 3D segmentation [3], for good results. Furthermore, since isotropic data is preferred in research settings, increasing the quantity and availability of isotropic data would facilitate for researchers.

As a result, researchers have begun to explore the field of MRI synthesis using deep learning. This includes anisotropic to isotropic reconstruction, as well as other super-resolution (SR) approaches, i.e. generating HR MRI images from LR MRI images. Successful MRI synthesis could enable better diagnosis and understanding of the brain and facilitate for both clinicians and researchers.

1.2 Related Research

To navigate the field of MRI SR, the following differences are crucial to understand.

- **Overall approach:** This defines the general strategy used to enhance image resolution. Early SR methods relied on transformation or interpolation techniques to upsample LR MRI [4]. These approaches were later extended with model-based and machine learning methods, including sparse representation and dictionary learning. The current state-of-the-art methods are based on deep learning, which learn complex mappings from LR to HR MRI and consistently outperform earlier approaches.

- **Resolution direction:** The resolution direction describes which spatial dimensions of the MRI volume are enhanced:

  - *In-plane super-resolution:* Upsampling within the imaging plane (2D), i.e. , improving resolution within individual slices.

  - *Through-plane super-resolution:* Enhancing resolution along the slice thickness direction, commonly used for anisotropic to isotropic reconstruction.

  - *Super-resolution across field strengths:* Aims to reconstruct a LR MRI from a low magnetic field scanner, to a HR MRI mimicing the output of a high magnetic field scanner.

  - *Full isotropic super-resolution:* Upsampling a LR isotropic 3D volume to a HR isotropic 3D volume across all dimensions.

- *Slice-to-volume reconstruction (SVR):* Reconstructing a HR 3D volume by combining multiple stacks of slices acquired from different orientations.
- *Multi-modality super-resolution:* Using information from multiple MRI modalities (e.g. , T1w and T2w) to improve spatial resolution, or using one modality to synthesize or enhance another.

As for the interpolation and transformation SR methods, an early approach was performed by Mahmoudzadeh et al. in 2014 [5]. They used an SVR approach which requires multiple anisotropic volumes in three orthogonal planes (axial, coronal, sagittal) to reconstruct a single HR isotropic volume. All anisotropic volumes were first interpolated to the HR isotropic grid and thereafter combined together. To ensure correct fusion of the volumes, one fixed image (the axial LR MRI) acted as template for registering the moving images (the coronal and sagittal LR MRIs). This approach is dependent on three separate anisotropic volumes to function, which is time-consuming to generate.

Moving on to the machine learning SR methods, a common approach is to use sparse representations and train a dictionary between LR and HR images. Yang et al. [6] used this method and Jia et al. [7] used a similar approach specifically adjusted to anisotropic to isotropic MRI reconstruction.

As for now, deep learning based models are dominating the field of MRI SR. The main difference between these models is the choice of learning paradigm, where methods may be *supervised*, requiring paired LR–HR data; *unsupervised*, learning without available HR ground truth; or *self-supervised*, where outcome labels are derived from the data itself.

One of the earlier approaches of deep learning SR was presented by Dong et al. [8], which used a patch-based, supervised CNN to reconstruct 2D HR images, corresponding to in-plane SR. Pham et al. followed by applying a similar approach specifically to 3D brain MRI [9], extending to full isotropic SR. With available HR isotropic MRI, it is possible to synthesize a corresponding LR MRI from each HR MRI. The LR MRI is then interpolated to the size of the HR image to facilitate the training. Pham et al. extended the network with residual learning [10], which resulted in a better performing model [11]. Many have followed the implementation of residual learning in MRI SR [12] [13], one example being Du et al. [14] which

introduced a through-plane SR method with both long and short skip connections as well as self-similarity, which improved performance and reduced computation time.

Supervised methods require either real HR-LR pairs or real HR images that can be used to generate synthetic HR-LR pairs. This is not always available, since medical data availability is very restricted due to patient privacy and sensitive information within the data. Therefore, many SR methods for MRI adopt a self-supervised learning strategy in order to avoid the limited data problem. Jog et al. [15] introduced this by using only the available LR MRI as training data to perform through-plane SR. By downsampling to an even lower resolution, LR2, and learn the mapping from LR2-LR, it was possible to apply the same mapping for LR-HR.

Another example is the self-supervised through-plane SR pipeline SMORE [16]. In SMORE, the same LR image is used for both training and testing, and synthetic LR–HR training pairs are generated directly from this image. Starting from the LR anisotropic MRI volume with low resolution in the through-plane (z) direction, 2D patches extracted from the HR in-plane (x, y) slices are used as ground truth (GT). The anisotropic image is then upsampled to an isotropic grid, followed by a blurring along one of the in-plane directions (e.g. , the x-axis). From this blurred image, new 2D patches that are LR in the in-plane (x, y) directions are extracted and paired with the corresponding GT in-plane patches to form synthetic LR–HR training pairs. These patch pairs are used to train a CNN. After training, the learned mapping is applied to the real through-plane (z) direction of the input image, producing HR 2D patches that can be reconstructed into an isotropic 3D volume. To ensure that the learned mapping generalizes across directions, the patch size is kept small so that the network learns local image features rather than global structural information. However, training and testing on the same sample, commonly referenced as zero-shot super resolution (ZSSR) [17] have a drawback. It prohibits too complex and deep network structures, since training has to be performed for each test sample.

Another self-supervised through-plane SR pipeline is the SIMPLE model, presented by Benisty et al. [18]. It is solely trained on real anisotropic data and uses the information in all directions to supervise each other. This is performed with a patch-based Generative Adversarial Network (GAN) U-Net with two important losses, in-plane reconstruction loss and cross-plane consistency loss. Where the in plane reconstruction is directly supervised by the HR in-plane slices (i.e. computed in 2D), and the cross-plane consistency loss ensures consistency in all directions

(3D).

Iwamoto et al. [19] suggests an unsupervised multi-modality through-plane SR approach, where SR for T2w MRI is guided by HR T1w MRI. As for the LR-HR pairs, the method produces a lower LR T2w MRI and lets the actual LR T2w MRI act as GT. The HR T1w MRI is inputted and concatenated in several layers of the network. Not using the ZSSR approach is an intentional choice to allow longer training times and deeper network structures.

Furthermore, in the area of SR between field strengths, Gicquel et al. [20] developed a supervised pipeline for SR of 3 Tesla (T) MRI to 7T MRI. Two models were used; a U-Net with residual units, and a U-Net integrated GAN with residual units.

In general, CNN-based methods, often extended to U-Nets or GANs, are well established in the field. The use of residual units within the network have also been widely adopted. To address data unavailability, many methods rely on unsupervised or/and patch-based approaches. Furthermore, most studies are trained specifically on adult brain MRI data. Consequently, SR of infant brain MRI remains underexplored, particularly within the context of supervised through-plane reconstruction.

1.3 Purpose and Objectives

1.3.1 Objectives

This project will focus on anisotropic to isotropic MRI reconstruction, also known as through-plane super-resolution, using deep learning. The aim is specifically adjusted to the needs of the University of Pennsylvania and the Children's Hospital of Pennsylvania (CHOP), whom have requested a deep learning pipeline to convert a T2w MRI infant brain dataset from anisotropic to isotropic. Hence, the setup is also conducted based on the data availability at CHOP, which follows the clinical standard of isotropic T1w MRI and anisotropic T2w MRI. Since the CHOP in some cases also acquired isotropic T2w MRI from patients, a supervised approach can be used. A model will be trained using LR-HR pairs where the isotropic T2w MRI serves as the ground truth (GT).

1.3.2 Purpose

The purpose of the project spans both the overall effect to the research within the field of anisotropic to isotropic MRI reconstruction, as well as the direct implementation opportunities for the University of Pennsylvania.

Global Purpose

From a global perspective, the results will add to the development of generating sharper MRI, which have effects both in clinics and in research. Being able to generate HR isotropic MRI will enable the following:

- Better diagnosis, treatment planning and brain monitoring for clinicians.
- Opportunities for downstream tasks which are dependent on the availability of isotropic 3D MRI, such as brain segmentation and myelin maturation tracking.
- Possibilities to generate more brain MRI data, facilitating for future research.

Direct Purpose

As for the University of Pennsylvania and CHOP, the deep learning pipeline can be directly applied to their purpose of generating isotropic MRI from anisotropic MRI from real clinical infant brain data. As for now, CHOP has a large collection of anisotropic T2w infant brain MRI. The application of an anisotropic to isotropic reconstruction model, would enable all existing data to be further used in specific research tasks. The primary focus will be to use the data for myelination quantification.

Myelination quantification refers to the process of myelin development in the brain. Myelin acts as a fatty insulation around the axons of the nerves. This insulation enables faster signaling between nerves and is hence dependent for brain function. During early infancy, myelination undergoes rapid development. The myelination process is still not well-understood and researchers aim to map it more thoroughly [21] [22]. The aim of the University of Pennsylvania and CHOP, and by extension this project, is to use the synthesized isotropic T2w MRI to create models of myelin development which could be used to benchmark patients with various conditions.

Additionally, infant brain MRI data is fairly complicated to acquire due to logistical issues during the acquisition. To avoid motion artifacts, the infant is required to be very still, which is typically achieved by letting the infant be asleep while in the scanner. However, that requires the scanner to be quiet, to avoid awakening the infant, which is not always possible. Another option is to sedate the baby, which clinicians also want to avoid. Consequently, a deep learning pipeline is a possible solution to these issues. Since CHOP already have large quantities of anisotropic MRI data, these could be directly reconstructed to isotropic images, instead of acquiring additional scans.

1.4 Limitations

1.4.1 Data Availability

Data availability is a major challenge in deep learning, as most models require large datasets to perform well. This issue is particularly prominent in the medical domain, where data often contain sensitive patient information. As a result, relatively few medical datasets are openly available.

In this project, the open-source dataset Baby Open Brain's Repository [23], containing isotropic MRI scans of infant brains is used. More detailed information about the dataset is provided in Section 3.2. To our knowledge, it is currently the only publicly available dataset that provides images with these specific characteristics. The dataset is comparatively small for training a deep learning application, consisting of only 71 subjects. It would have been desirable to include a larger dataset to enable better training and results.

1.4.2 Computing Resources

Training deep learning models requires substantial computational resources. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. All model training is performed on the High-Performance Computing (HPC) cluster BerzeLiUs. However, even with HPC clusters, training can take time, especially when creating deeper networks and extending the training data. In addition, job scheduling and

queue times on the cluster can further affect the overall training timeline.

Chapter 2

Theory

2.1 Neural Networks

Neural networks form the base of deep learning applications and are inspired by the structure and function of the human brain. A neural network consists of interconnected units which transmit information through weighted connections. By iteratively updating the weight on each connection, based on a computed loss between the output and the ground truth (GT), the neural network optimizes the model to reach the desired result. The process of updating the weights is referenced as backpropagation and aims to find the minima of the loss function.

2.1.1 Perceptron

Figure 2.1 illustrates the perceptron, which is the building block of all neural networks. It was first introduced by Rosenblatt [24] in 1957, and was the first model to present machine learning by adapting its parameters based on the input data to improve its performance. Figure 2.1 illustrates the perceptron which is based on a node which receives a set of inputs $x_1, x_2, \ldots, x_n$, each of which is multiplied by a corresponding weight $w_1, w_2, \ldots, w_n$. The weighted inputs are summed and a node-specific bias b is added. Then the result is passed through an activation function. By stacking and combining several perceptrons, a neural network is formed, which manage to learn complex patterns in the data.

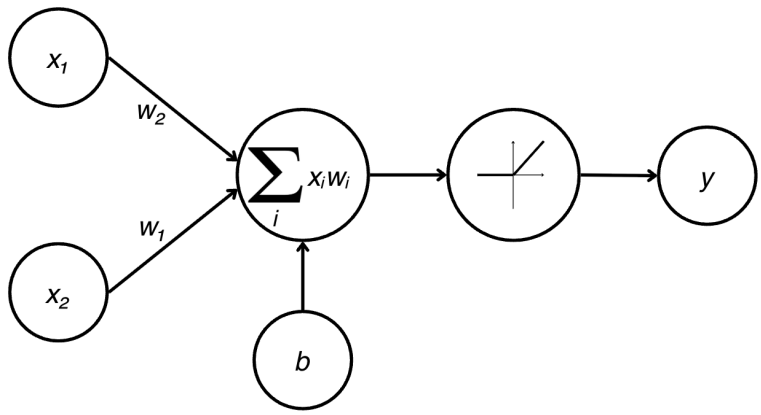


Figure 2.1: A perceptron with inputs x_1, x_2, and corresponding weights w_1, w_2, which form a weighted sum. The weighted sum is passed through an activation function to produce the output y.

2.1.2 Activation Functions

The activation function, which is applied after the weighted sum, is the last step before producing the output of a neuron. It is a crucial component of neural networks, as it introduces non-linearity into the model. Without activation functions, a neural network would simply compute linear combinations of its inputs.

Several activation functions are commonly used in deep learning. The Rectified Linear Unit (ReLU) is one of the most widely adopted functions and outputs zero for negative input values while leaving positive values unchanged. Since all negative values outputs zero, negative valued neurons become inactive and may harm the learning process. To address this problem, the Leaky Rectified Linear Unit (LReLU) was introduced [25]. LReLU multiplies all negative outputs with a small value to restore its impact, preventing inactive neurons. Notably, LReLU have showed negligible impact on accuracy compared to ReLU. A further extension of ReLU and LReLU is the Parametric Rectified Linear Unit (PReLU)[26]. The PReLU differs from the LReLU in presenting the negative multiplier value as learnable. The different ReLU activation functions are visualized in Figure 2.2 and can be formally defined as in Eq. (1)

$$f(y_i) = \begin{cases} y_i, & \text{if } y_i > 0, \\ a_i y_i, & \text{if } y_i \leq 0. \end{cases} \tag{2.1}$$

Where, y_i is the input of the activation function f on the i-th channel, and a_i is a coefficient controlling the slope of the negative part. When $a_i = 0$, the activation function reduces to ReLU. If a_i is a fixed, non-zero value, the function corresponds to LReLU. When a_i is a learnable parameter, the activation function is referred to as PReLU.

Other commonly used activation functions include the sigmoid function, which maps inputs to the range $(0, 1)$, commonly used for binary classification problems, and the hyperbolic tangent (tanh), which maps inputs to the range $(-1, 1)$, enabling negative data to restore negative.

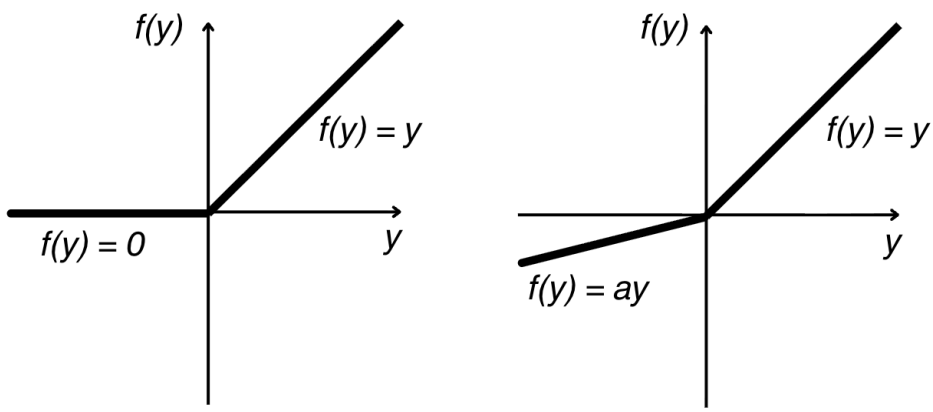


Figure 2.2: The activation function is ReLU, when $f(y) = 0$ for all negative values, and PReLU when $f(y) = ay$ for all negative values, where a is a learnable parameter.

2.1.3 Optimization method

During network training, the model weights are iteratively updated based on the computed loss function. The aim is to minimize the loss function by finding the optimal weight setup.

Gradient Descent

Gradient descent is one of the simplest optimization algorithms used in deep learning. It relies on computing the gradient of the loss function with respect to each individual weight. Each weight is then updated by taking a step in the direction of the steepest descent. The step size is determined by the learning rate, which must be

carefully chosen to ensure convergence. The goal is to minimize the loss function, however, with a poorly chosen learning rate, the optimization process may converge slowly or become trapped in a local minimum.

Stochastic Gradient Descent (SGD)

Computing gradients over the entire training dataset can be computationally expensive and may also increase the risk of being trapped in a local minima. To address this, Stochastic Gradient Descent (SGD) estimates the gradient using smaller mini-batches of the data, and updates the weights based on this. This procedure is less computationally demanding and also helps escaping local minima and plateaus.

Adaptive Moment Estimation (Adam)

The Adaptive Moment Estimation (Adam) algorithm differs from the above by using an adaptive learning rate. The learning rate is scaled and adapted for each weight, based on the first and second moments of the gradients. Adam aims to improve convergence and stability.

2.2 Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a type of neural network commonly used in image analysis due to its ability to model spatially structured data. CNNs are composed of convolutional blocks that rely on the mathematical definition of convolution:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(\tau)\, g(x-\tau)\, d\tau. \tag{2.2}$$

Where $(f * g)(x)$ is the resulting convolved signal of functions f and g, at position x, integrated over τ. When performing a convolution in a CNN, it is discrete instead of continuous. The input function f represents the image, while the function g is referred to as the kernel. The kernel slides over the input image and computes the discrete convolution, producing a convolved output known as a feature map. An activation function is applied to the feature map to introduce non-linearity.

2.2.1 Pooling

Pooling layers are used in CNNs to downsample the input along the spatial dimensions. This reduces the spatial size of the feature maps and hence, the number of computations, resulting in a more efficient network. Additionally, the pooling layers introduce a translational invariance, which allows variation between images. For instance, if an edge is slightly moved in one image, it will not affect the overall result. Lastly, the pooling allows hierarchical learning, where the higher layers learn more global features, compared to the lower levels focusing on edges and finer details.

The most commonly used pooling operation is max pooling, which selects the maximum value within a local neighborhood of the feature map. This process compresses the dimensions while still keeping the most important information. Pooling can also be implemented using a convolutional layer with appropriate stride and zero-padding which results in a similar downsampling effect.

2.3 Residual Unit

Conceptually, deeper neural networks are capable of learning more complex patterns, which can lead to improved performance. However, increasing network depth also makes training more difficult. Two well-known challenges associated with deep networks are the vanishing and exploding gradient problem, as well as the degradation problem.

The vanishing and exploding gradient problem happens during backpropagation, the process used to update network weights. Backpropagation relies on the gradient of the loss function, which is computed using the chain rule across all layers of the network. As gradients are repeatedly multiplied when propagating through deep architectures, they may either shrink towards zero or grow uncontrollably large, both which hinder learning.

The vanishing and exploding gradient problem can be partially solved by using ReLU activation functions instead of sigmoid or hyperbolic tangent functions, or with normalization techniques, such as batch normalization, which help stabilize gradient propagation [27].

Another challenge specific to very deep networks is the degradation problem which refers to the phenomenon where adding more layers to a deep network leads

to higher training error.

Residual learning has proven to be an effective solution to the problems of training too deep networks [10]. The key idea of a residual unit, illustrated in Figure 2.3, is the introduction of shortcut connections that allow the input x to bypass one or more weight layers and be added directly to the output. The output y of a residual unit is therefore defined as

$$y = F(x) + x. \tag{2.3}$$

where $F(x)$ represents the mapping learned by the stacked layers.

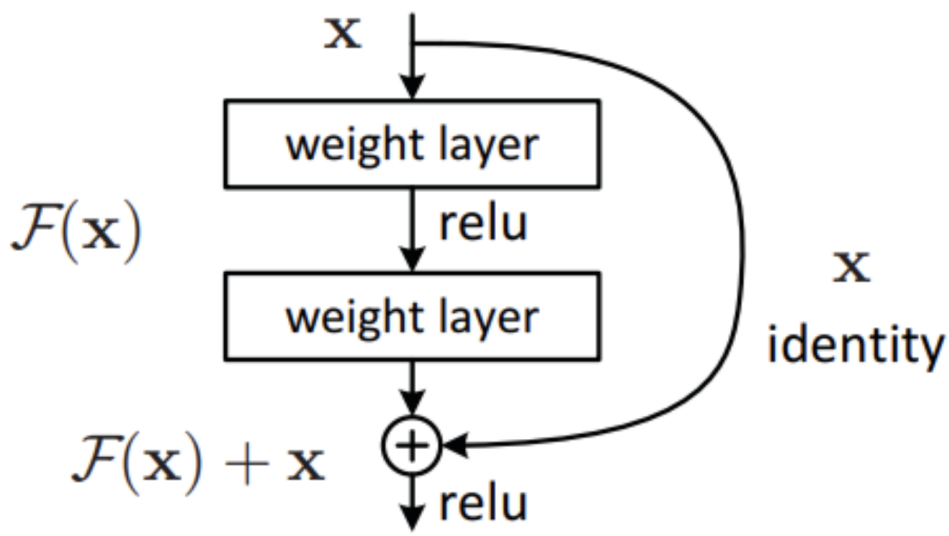


Figure 2.3: The residual unit presented by He et al. [10].

2.4 U-Net

The U-Net is a widely used neural network architecture in medical image analysis and was first introduced by Ronneberger et al. [28] in 2015. Figure 2.4 illustrates the classical U-Net architecture.

The U-Net consists of three main components: an encoding path, a bottleneck, and a decoding path. The encoding path performs a series of convolutional operations followed by activation functions. After each encoding stage, the spatial resolution of the feature maps is reduced, with some kind of pooling, while the number of feature channels is increased. This process enables the network to capture high-level

contextual information.

At the bottleneck, the feature maps reach their smallest spatial resolution while the number of feature channels is maximized, representing the most compressed representation of the input image.

The decoding path progressively reconstructs the spatial resolution through up-convolutions, while simultaneously reducing the number of feature channels. At each decoding stage, the upsampled feature maps are concatenated with corresponding feature maps from the encoding path via skip connections. These skip connections directly transfer feature information from earlier encoding layers to the decoding layers.

Skip connections restore high spatial resolution information and enable more precise localization of fine details. From the deeper layers, on the other hand, more global contextual information is provided. The combination of these two components enables U-Net to achieve strong performance, particularly in segmentation tasks, but also in other applications [28].

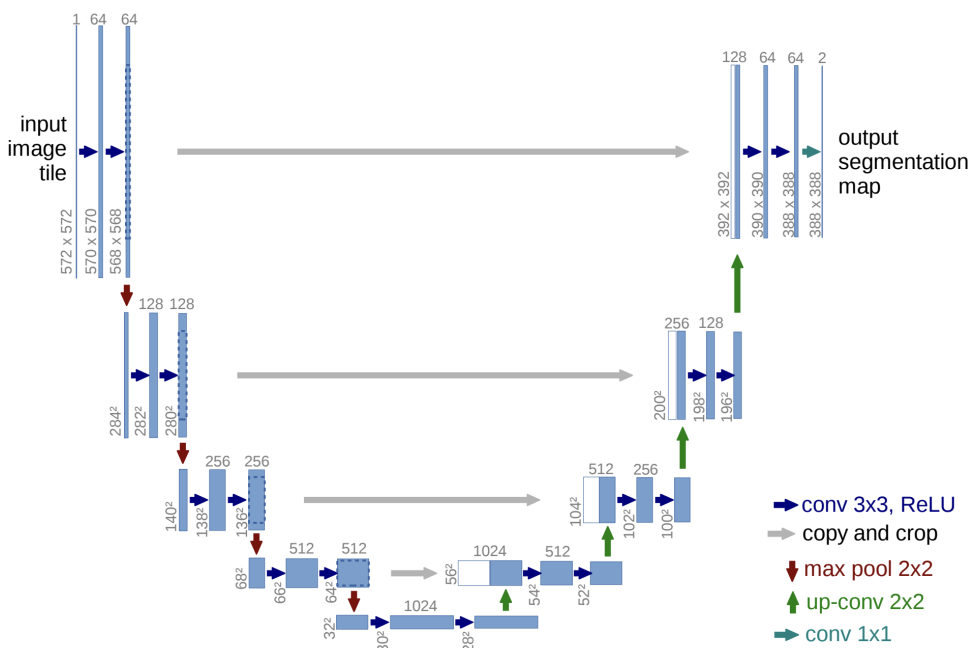


Figure 2.4: The U-Net architecture presented by Ronneberger et al. [28].

2.4.1 Residual U-Net

A residual U-Net will simply implement the residual units, with its shortcut connections, into chosen layers of the U-Net. Combined with the skip connections from the U-Net, it creates a network which can benefit from both architectures, using the residual units to hinder degradation and the skip connections to preserve spatial localization.

Similar setups have previously been used for MRI SR tasks [20] [29]. However, the ResU-Net is even more popular for MRI segmentation tasks [30] [31]. In this paper, the effect of the residual units within a U-Net made for anisotropic to isotropic reconstruction will be further investigated.

2.5 Loss Function

The loss function defines how to compare the output with the GT. The function is computed in every iteration and serves as the basis for the backpropagation. Accordingly, there are several different loss functions, all presenting slightly different properties [32].

2.5.1 Mean Square Error (MSE)

The MSE computes the average square distance between the predicted and the actual value. Due to squared values, outliers have a stronger effect on the loss. It is defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(I_1^{(i)} - I_2^{(i)}\right)^2. \tag{2.4}$$

where n is the total number of pixels, and $I_1^{(i)}$ and $I_2^{(i)}$ represent the intensity values of the i-th pixel in the two compared images.

2.5.2 Mean Absolute Error (MAE)

MAE computes the average absolute value, and is less sensitive to outliers due to using absolute value instead of squared value. MAE is defined as

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|I_1^{(i)} - I_2^{(i)}\right|. \tag{2.5}$$

where n is the total number of pixels, and $I_1^{(i)}$ and $I_2^{(i)}$ represent the intensity values of the i-th pixel in the two compared images.

2.5.3 Perceptual Loss

Perceptual losses differ from pixel-wise losses (as MSE and MAE) since it uses feature extraction to compare differences between images. This results in a stronger alignment with actual human perception, providing a more accurate judgment [33]. It is specifically useful for different types of image generation tasks, SR included [34]. It operates by using a pretrained network to extract features from output and GT, a common choice is the VGG network [35].

2.6 Evaluation Metrics

To evaluate and quantify the result of a model, evaluation metrics can be used. There exists several metrics, adapted for different tasks [36].

2.6.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR is a widely used metric which relies on the MSE between two images. It is defined as

$$\text{PSNR} = 10\log_{10}\left(\frac{I_{\text{max}}^2}{\text{MSE}}\right). \tag{2.6}$$

where I_{max}^2 represents the square of the maximum possible voxel intensity value in the image, and MSE denotes the mean squared error, defined as the average of the squared differences between corresponding voxel intensities.

2.6.2 Structural Similarity Index Measure (SSIM)

SSIM was introduced by Wang et al. to further extend the performance of evaluation metrics [37]. It aims to align closer with the human perception, by evaluating more factors than just voxel intensities. It is built of three components, which measures luminance, contrast and structure in the images. The overall definition is as follows

$$\text{SSIM}(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}. \tag{2.7}$$

where $l(x,y)$ measures the difference in luminance, $c(x,y)$, the difference in contrast, and $s(x,y)$, the difference in structure between the two images. The parameters α, β, and γ are weighting constants that control the relative importance of the luminance, contrast, and structure components. For simplicity, these parameters are commonly set to $\alpha = \beta = \gamma = 1$ [38].

2.6.3 Perceptual Metric

A perceptual metric operates similar to a perceptual loss, with the use of a pretrained network for feature extraction and comparison. The difference lies in the intended use, where the metric aims to present the result instead of computing the gradient of the loss function.

2.7 Augmentations

Augmentations is a widely used method to enhance deep learning performance. The main idea is to slightly shift or adjust the original data to create more samples. Common augmentations include rotation, translation, blurring etc. It is specifically useful when aiming to generalize data with site-specific variations, which is often the case in medical imaging. Models trained on augmented data usually present a more generalized and robust learning process, which can reduce overfitting [39].

Augmentation has been adopted in several MRI SR tasks, one example being the SMORE [18] model, where rotation and flipping of the patches was performed. Furthermore, Ronneberger et al. describes the importance of augmentations when presenting the U-Net architecture [28].

Chapter 3

Method

3.1 Aim

The aim of the project is to implement a 3D residual U-Net which can successfully output an isotropic T2w MRI. Since both isotropic T1w MRI and anisotropic T2w MRI are available, the model will use two input channels to take advantage of all available information. The model will be patch-based, processing one patch at a time. By reconstructing the output patches, a full 3D isotropic volume will be acquired. The approach will be supervised, using real T2w isotropic MRI patches as GT. The setup is visualized in Figure 3.1. The network is trained with data from the open-source dataset Baby Open Brain's Repository [23], see Section 3.2, consisting of 71 isotropic T1w and T2w infant brain MRIs.

This chapter presents the complete process of developing the proposed model, from data preprocessing, to the adjustments and evaluations performed to optimize model performance.

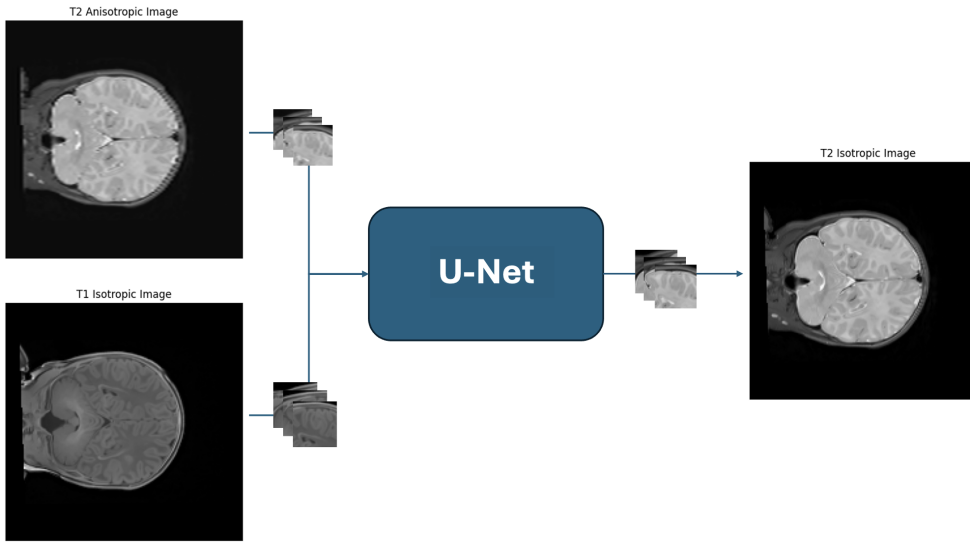


Figure 3.1: The proposed model setup which inputs 3D patches in two channels, T2w anisotropic and T1w isotropic, and outputs 3D patches which are reconstructed to a T2w isotropic MRI. The network used is a supervised residual U-Net.

3.2 Data

3.2.1 Baby Open Brain's Repository

The Baby Open Brain's (BOB's) Repository [23] is an open-source dataset collected at University of Minnesota. It consists of 71 3D MRI images of infant brains, aged 1 to 9 months, from a total of 51 participants. For each of the 71 images, both T1w and T2w MRI are collected. All images are HR isotropic with a resolution of $0.8 \times 0.8 \times 0.8$ mm^3. Additional details are presented in Table 3.1. Figure 3.2 shows an example scan from the dataset.

Table 3.1: MRI acquisition details for the BOB's repository dataset

Item	**Details**
Scanner	Siemens 3T Prisma
Head coil	32-channel
T1w structural (MPRAGE)	TR 2400 ms; TE 2.24 ms; TI 1600 ms; flip angle 8°; resolution $0.8 \times 0.8 \times 0.8$ mm^3
T2w structural (turbo spin-echo)	Turbo factor 314; echo train length 1166 ms; TR 3200 ms; TE 564 ms; resolution $0.8 \times 0.8 \times 0.8$ mm^3; variable flip angle

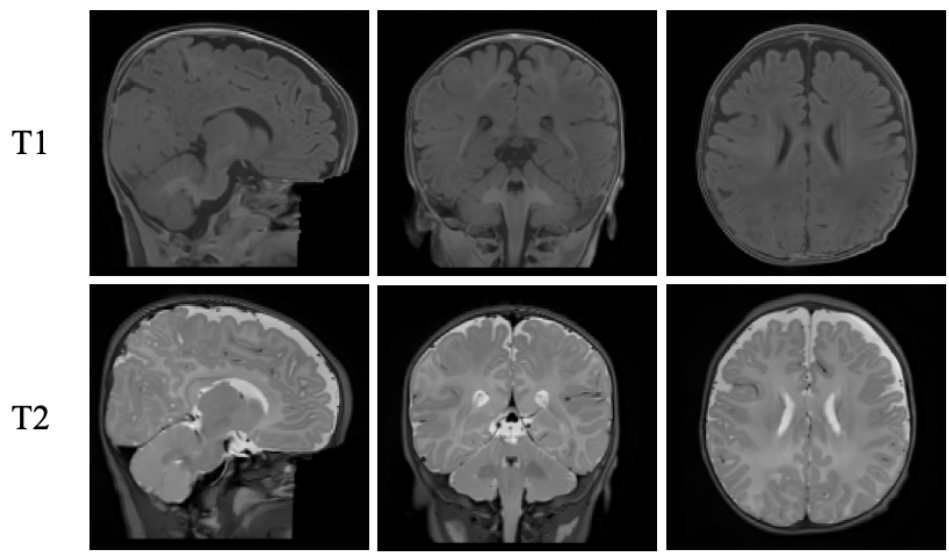


Figure 3.2: MRI image of a brain from a 3 month old infant in BOB's repository dataset.

Prior Preprocessing

The BOB's repository dataset was preprocessed prior use, to a spatial resolution of $182 \times 218 \times 182$ voxels with a voxel size of $1 \times 1 \times 1$ mm^3. In addition, the images were defaced and rigidly registered to a common reference space [23].

3.3 Preprocessing

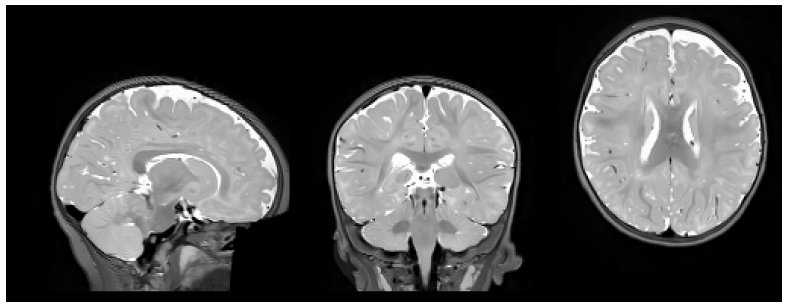
3.3.1 Synthesize Anisotropy

Since the proposed deep learning approach relies on supervised training, LR-HR pairs, i.e. anisotropic and isotropic versions of the MRI, are required. The BOB's dataset only contains HR isotropic MRI, therefore the LR anisotropic T2w MRI is synthesized from the already existing HR isotropic images. The proposed method is well established and used in several previous studies [3][16][18][19][40]. The first step is to downsample the isotropic T2w image to anisotropic size. Secondly, interpolation-based upsampling is performed to ensure size alignment between images.

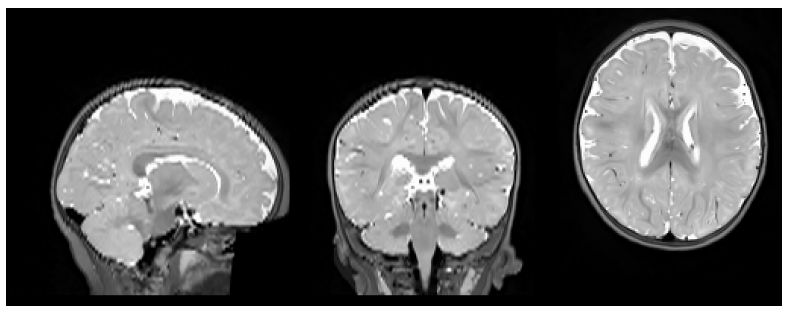
- **Downsampling:** The downsampling is performed along one direction d by removing slices according to a specified downsampling factor r. Given an original MRI volume of size $x \times y \times z$, the downsampled volume is reduced by a factor of r along direction d, while the other two dimensions are preserved. This results in a volume of size $\frac{x}{r} \times y \times z$. The model in this project is trained on images downsampled in x-direction, i.e. anisotropic in the axial dimension

- **Upsampling:** The downsampled volumes are upsampled back to the original size $x \times y \times z$. Conceptually, this step reconstructs the slices removed during the downsampling process. Upsampling is performed using cubic spline interpolation, which estimates intermediate voxel values based on neighboring samples and provides a smooth reconstruction. The purpose of the upsampling process is to match the shape of the isotropic and anisotropic image. It also enables the use of different downsampling factors without having to adjust the input dimensions of the network, which results in a more generalized model, since it can input any type of anisotropic T2w MRI. Additionally the interpolated LR anisotropic image serves as a baseline/benchmark which can be used for evaluating results.

Figure 3.3 shows a HR isotropic T2w MRI and its corresponding synthesized LR anisotropic MRI, downsampled with a factor 2 in axial direction.



(a) HR isotropic T2w MRI



(b) Synthesized LR anisotropic MRI.

Figure 3.3: Real HR image and its corresponding synthesized LR image. The synthesized image was downsampled with a factor 2 in axial direction

3.3.2 Normalization

The preprocessed images from the BOB's Repository dataset contain voxel intensity values in the range $[0, 136.2705]$. Such a wide intensity range can negatively affect the network training and the evaluation of the model performance, as many loss functions and evaluation metrics are based on voxel-wise intensity differences.

To address these issues, intensity normalization is applied. Several normalization strategies exist, and the choice of method depends on the characteristics of the data and the imaging modality. For the BOB's Repository images, a masked and clipped percentile-based min-max normalization was employed.

The normalization procedure consists of the following steps:

- **Masking:** A mask is applied to exclude non-informative background regions from the normalization process. The chosen mask excludes all zero-values from the normalization process.

- **Percentile clipping:** To reduce the influence of outlier intensities, voxel val-

ues are clipped to the 1^{st} and 99^{th} percentiles of the masked intensity distribution.

- **Min-max normalization:** The clipped intensities are linearly rescaled to the range $[0, 1]$ using min-max normalization.

3.3.3 Patch Extraction

A patch-based learning strategy was adopted for two main reasons. First, since the dataset is small, consisting of only 71 images, extracting patches results in a larger training set. Second, learning from smaller patches is less computationally demanding than learning directly from full 3D volumes [18].

The patch size was inspired by recent works[18] [16], where patches of size $64 \times 64 \times 64$ mm^3 and 2D patches of size 32×32 mm^2 where successfully implemented. In this project 3D patches of size $64 \times 64 \times 64$ mm^3 and $32 \times 32 \times 32$ mm^3 were tested and evaluated.

To ensure compatibility between patch dimensions and image size, the volumes were initially zero-padded from shape $182 \times 218 \times 182$ mm^3 to shape $192 \times 224 \times 192$ mm^3. A stride corresponding to a 50% overlap between neighboring patches was used, resulting in each voxel being included in multiple patches.

Patch extraction was performed in a consistent order across all samples. During training, the loss was computed on a patch-wise basis between the predicted patch and the corresponding GT patch. After inference, the individual output patches were recombined to reconstruct the final output volume.

3.3.4 Augmentation

Since the BOB's dataset is relatively small, augmentations could be of important value, since it could make the model more generalizable. In this case, augmentations were produced by introducing more types of anisotropy, by adjusting the downsampling rate.

The first model was trained without augmentations with a downsample rate of 2 (D2), i.e. removing every other slice, and an axial downsampling direction. For the augmented models, the first step included adding one augmentation, i.e. changing the downsampling rate to 4 (D4). Figure 4.1b shows the two types of training

data, where the use of both anisotropic versions, i.e. D2+D4, corresponds to the augmented model.

In further implementations, more downsampling factors, as well as downsampling in different directions, will be conducted to simulate variations in the dataset.

3.4 Model Architecture

The model chosen for the task is a residual U-Net from the Medical Open Network for Artificial Intelligence (MONAI) [41].

Figure 3.4 visualizes the overall structure of the MONAI U-Net, with example hyperparameters defined as in Table 3.2. The U-Net contains three types of units, the encoder unit (pink), the bottleneck unit (blue) and the decoder unit (peach). The encoding residual unit downsamples the spatial dimension with a strided convolution (denoted as the Conv2D block in the figure), and is built out of several convolution blocks followed by PReLU activation functions. The number of blocks within the unit is defined by the hyperparameter residual size and can be adjusted. The bottleneck unit is identical to the encoder unit with exception for the absence of a downsampling block. It is only used in the bottleneck of the U-Net, hence keeping the spatial dimension constant between the last two layers. The decoder unit performs an upsampling in the spatial dimension with a transposed convolution. Thereafter, it consists of solely one convolutional block and PReLU activation within the residual connection. It is never adjusted, hence the hyperparameter residual size only affects the encoding path. The skip connections add information from the encoding path to the decoding path with concatenation. The resulting output is consistently a 1-channel 3D patch. The depth, defined by the hyperparameter channels per level, is changed and evaluated during the project. The example U-Net architecture is therefore solely an example of a possible depth choice.

Table 3.2: Hyperparameters in Example U-Net

Hyperparameter	**Example U-Net**
Spatial dimensions	3
Input channels	2
Output channels	1
Channels per level	[16, 32, 64, 128]
Normalization	None
Residual size	*not defined*
Activation function	PReLU

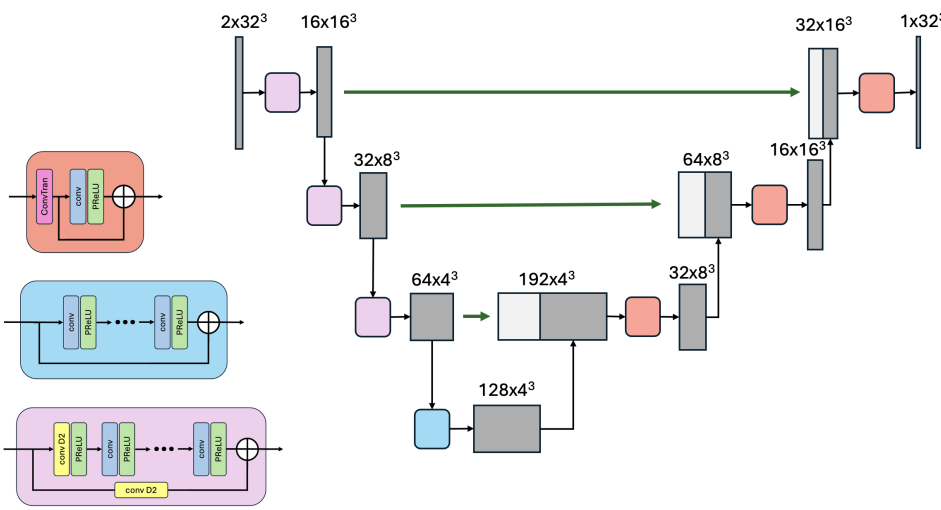


Figure 3.4: Example U-Net architecture. The hyperparameters for this specific architecture are defined in Table 3.2. The colored units are residual blocks. The green arrows represent the skip connections.

3.5 Implementation

Table 3.3 presents the implementation details of the proposed approach. The dataset, described in section Data 3.2, was split into training, validation and test data, with 49 training images (70%), 11 validation images (15%) and 11 test images (15%). Further input specifications are described in section Preprocessing 3.3.

The training loop uses an Adam optimizer with a learning rate of 10^{-4}. Mean squared error (MSE) is used as the loss function and is computed on a patch-wise basis. The batch size is set to 2, and the number of training epochs is determined using an early stopping strategy, which saves the model achieving the best perfor-

mance on the validation set, and stops training with a patience of 5.

Several network hyperparameters are variable and are systematically evaluated throughout the model development process, these are denoted as *variable* in Table 3.3. However, the spatial dimensions, number of output channels, normalization strategy, and activation function are fixed across all evaluated architectures.

Table 3.3: Implementation details of the proposed U-Net model.

Parameter	**Value**
Dataset and input	
Dataset	BOB's repository
Dataset size	71 images
Train / validation / test split	49 / 11 / 11 images
Normalization	Min-max w. percentile clip
Image size	$192 \times 224 \times 192$ mm^3
Patch size	*variable*
Training loop	
Optimizer	Adam
Learning rate	10^{-4}
Loss function	MSE
Batch size	2
Number of epochs	early stopping based on val. result
Network	
Spatial dimensions	3
Input channels	*variable*
Output channels	1
Channels per level	*variable*
Normalization	None
Residual size	*variable*
Activation function	PReLU
Training data	*variable*

3.6 Optimizing Model

In the following section the different architecture adjustments, which have all been tested and evaluated, are summarized. Table 3.4 presents all architectures and their specific configurations, i.e. the *variable* parameters from Table 3.3. The Train. data column represents the type of input data which was used to train the model, where the options are; anisotropic downsampled with a step size of 2 (D2) and anisotropic

downsampled with a step size of 4 (D4).

Table 3.4: Configurations of the evaluated U-Net architectures.

Configuration	Depth	Patch size	Input	Res. size	Train. data
Shallow U-Net	(16, 32, 64, 128, 256)	32x32x32	T1w+T2w	2	D2
Deep U-Net	(32, 64, 128, 256, 512, 1024)	32x32x32	T1w+T2w	10	D2
64mm patch	(32, 64, 128, 256, 512, 1024)	64x64x64	T1w+T2w	10	D2
T1w input	(32, 64, 128, 256, 512, 1024)	32x32x32	T1w	10	D2
T2w input	(32, 64, 128, 256, 512, 1024)	32x32x32	T2w	10	D2
No residual	(32, 64, 128, 256, 512, 1024)	32x32x32	T1w+T2w	0	D2
2 residual	(32, 64, 128, 256, 512, 1024)	32x32x32	T1w+T2w	2	D2
18 residual	(32, 64, 128, 256, 512, 1024)	32x32x32	T1w+T2w	18	D2
D2+D4 trained	(32, 64, 128, 256, 512, 1024)	32x32x32	T1w+T2w	10	D2+D4

3.6.1 Depth

To compare the effect of the network depth, two versions were tested, which are denoted in Table 3.4 as *Shallow U-Net* respectively *Deep U-Net*. The *Shallow U-Net* has 5 layers which results in 4 skip connections. The feature dimensions starts at 16 features in the first layer, and ends with 256 features in the bottleneck. The residual size is set to 2.

The *Deep U-Net* has 6 layers which results in 5 skip connections. It starts at 32 features, and ends with 1024 features in the bottleneck. It has larger residual units of size 10.

3.6.2 Patch Size

The selected patch size was inspired by recent literature [18] [16]. Consequently, patch sizes of $64 \times 64 \times 64$ mm^3 and $32 \times 32 \times 32$ mm^3 were evaluated. In Table 3.4, the *Deep U-Net* conduct the smaller patch size and the *64mm patch* configuration represents the larger patch size.

3.6.3 Input Channels

The primary choice of input channels was motivated by the availability of clinical MRI data. As previously mentioned, it is common in clinical practice to acquire anisotropic T2w MRI together with isotropic T1w MRI [2]. Since these two modalities provide complementary anatomical information, it was therefore natural to use both as input in order to develop a well-performing model.

To further investigate the influence of input modality, configurations using a single input channel were also evaluated. Two cases were considered: one using only HR T1w MRI as input and one using only LR T2w MRI. All comparisons were performed using the deeper U-Net architecture, with the only differences being the number of input channels and the selected input modality. In Table 3.4 the input versions are denoted as *Deep U-Net*, for double input, and *T1w input* and *T2w input*, for the single inputs.

3.6.4 Residual Unit

To investigate the effect of residual units, several variants of the deeper U-Net architecture were evaluated, differing only in the configuration of the residual units. The following residual sizes were tested: 2, 10, and 18. In addition, a U-Net architecture without residual connections was also evaluated. All versions are summarized in Table 3.4 where *Deep U-Net*, has residual units of size 10, and the rest are denoted as *No residual*, *2 residual* and *18 residual*.

3.6.5 Normalization

Normalization layers can be incorporated into the network architecture to improve training stability [27]. Common normalization choices include batch normalization, instance normalization, group normalization, and layer normalization. All mentioned normalizations were tested, but none led to improved performance. Most resulted in stitching artifacts when later reconstructing from patches. For this reason, the model was solely trained without normalization layers.

3.7 Evaluation

For evaluation of the proposed model, visual assessment in combination with quantitative evaluation metrics was used. The evaluation approach directly depends on the downstream tasks of the application. If the generated images are to be used for visual tasks, such as diagnosis, visual assessment serves as the most important evaluation. However, if it is to be used for computer-aided calculations, such as segmentation, quantitative metrics should be considered.

3.7.1 Visual Assessment

The visual assessment was performed by visually comparing the reconstructed model output MRI with the corresponding GT MRI. This evaluation was conducted by the author and the project supervisors. The comparison involves analyzing blurriness, edges and overall alignment between the output image and the corresponding GT. The results are presented visually throughout this report, allowing the reader to perform an independent visual assessment.

3.7.2 Evaluation Metrics

The use of evaluation metrics allows for a quantitative analysis of the results. To evaluate the models, three main evaluation metrics were computed; PSNR, SSIM and a perceptual metric. The perceptual metric is implemented by MONAI, and uses the network pretrained by Chen, et al. [42]. The metrics were computed between the reconstructed output and its corresponding GT T2w MRI. The mean was computed and is presented in the results. The synthesized T2w LR anisotropic MRI, i.e. one of the input images, is also evaluated and used as a benchmark.

Chapter 4

Results

The following section presents the result of the best-performing model and compares it with a benchmark method. In addition, the results from the model optimization process is described to further motivate the design choices of the final model.

4.1 Benchmark

To evaluate the model performance, a benchmark was chosen for comparison. In this case, the interpolated anisotropic T2w MRI serves as the benchmark.

Two different LR anisotropic MRI was used, D2 and D4, both visualized in Figure 4.1b. The benchmarks are further on referenced as *Interpolated D2* and *Interpolated D4*

4.2 Best-performing Model

Figure 4.2 presents the best-performing model, supported by both visual assessment and evaluation metrics, see Table 4.2 where it is denoted as *Deep U-Net*. For comparison, the evaluation metrics for the benchmark model are consistently worse.

The specific model configurations are detailed in Table 4.1 combined with the overall implementation details presented in Table 3.3. The results indicate that a smaller patch size of $32 \times 32 \times 32$ mm^3 is preferable, along with a deeper U-Net architecture with an increased number of channels per level and larger residual units in

the encoding path. Furthermore, the use of both isotropic T1w MRI and anisotropic T2w MRI as model input led to improved performance.

This model was trained and evaluated exclusively on anisotropic MRI data with a slice thickness of 2 mm. The impact of varying slice thicknesses is further investigated in Section 4.3.5.

Table 4.1: Configuration of the best-performing model

Hyperparameter	**Best-performing model**
Input channels	2
Channels per level	[32, 64, 128, 256, 512, 1024]
Number of residual units	10
Patch size	$32 \times 32 \times 32\,\text{mm}^3$
Training data	D2

Table 4.2: Quantitative evaluation metrics from all tested U-Net configurations. Each configuration is based on the best-performing model configuration, but with one or more adjusted variables, specified in Table 3.4. The metrics are a mean from all images in the test set and the highlighted metrics show the best result.

Configuration	**PSNR**	**SSIM**	**Perceptual**	**Note**
Interpolated D2	29.85	0.975	3.67×10^{-5}	Benchmark
Interpolated D4	27.12	0.946	9.28×10^{-5}	Benchmark
Shallow U-Net	36.51	0.982	5.44×10^{-6}	
Deep U-Net	**38.27**	**0.993**	$\mathbf{2.96 \times 10^{-6}}$	
64 mm patch	37.54	0.990	3.82×10^{-6}	
T2w input	37.15	0.991	6.13×10^{-6}	
T1w input	18.85	0.750	2.74×10^{-3}	
No residual	36.07	0.988	5.27×10^{-6}	
2 residual	37.11	0.991	3.51×10^{-6}	
18 residual	37.21	0.991	3.42×10^{-6}	
D2 + D4 trained	37.79	0.992	4.03×10^{-6}	D2 tested
D2 + D4 trained	34.10	0.981	1.54×10^{-5}	D4 tested

4.3 Alternative Models

The following sections presents the individual results from the process of achieving the best-performing model.

4.3.1 Depth Comparison

The depth of the U-Net architecture and the number of feature channels per layer were found to have a notable impact on model performance. Figure 4.3 llustrates synthetic images generated from two different U-Net architectures, referenced as *Shallow U-Net* and *Deep U-Net*. The corresponding model specifications are summarized in Table 3.4, where the *Deep U-Net* is more complex due to its increased depth and higher number of feature channels per level.

For further comparison, the evaluation metrics are presented in Table 4.2. Both the metric values and visual assessment indicate performance differences between the *Shallow U-Net* and *Deep U-Net*, with the deeper architecture presenting superior results.

4.3.2 Patch Size Comparison

Two patch sizes, $32 \times 32 \times 32$ mm^3 and $64 \times 64 \times 64$ mm^3 , were evaluated. Figure 4.4 shows the results obtained while varying the patch size. Evaluation metrics for both patch sizes are reported in Table 4.2, with configuration details stated in Table 3.4. Both visual assessment and quantitative metrics indicate that the smaller patch size of $32 \times 32 \times 32$ mm^3 is preferable.

4.3.3 Input Comparison

To motivate the choice of double input channels, all possible input combinations were tested. This includes double input of T1w isotropic MRI and T2w anisotropic MRI, single input of T2w anisotropic MRI, as well as single input of T1w isotropic MRI, see Table 3.4. Figure 4.5 presents the output of the different input variants. It is evident that the single input with T1w isotropic MRI performs substantially worse than the other alternatives, with visible stitching artifacts from patch reconstruction. The remaining variants show comparable results when visually assessing. Although, quantitative evaluation metrics present further information. As reported in Table 4.2, the double input model, denoted as *Deep U-Net*, achieves a higher PSNR compared to the single-input variants, with a PSNR of 38.27 dB for the combined T1w+T2w input, compared to 37.15 dB for the single-input configuration. Overall, these results indicate that including the anisotropic T2w MRI as part of the input is crucial for achieving acceptable results with the proposed model.

4.3.4 Residual Unit

The size of the residual units in the encoder path was evaluated and compared. The encoder (pink) and bottleneck (blue) block in Figure 3.4 represent example residual units. Increasing the size corresponds to adding more convolutional blocks within the shortcut connection. The sizes tested were 0, 2, 10 and 18, see Table 3.4. Figure 4.6 visualizes the model output from the different residual unit choices and their corresponding evaluation metrics are presented in Table 4.2.

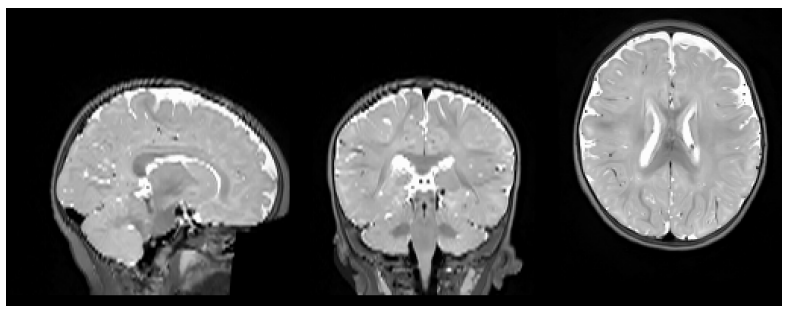
Visual assessment supports the residual unit of size 10 and proves that not using any residual unit performs worse. The quantitative metrics conform with the visual evaluation.

4.3.5 Augmentations

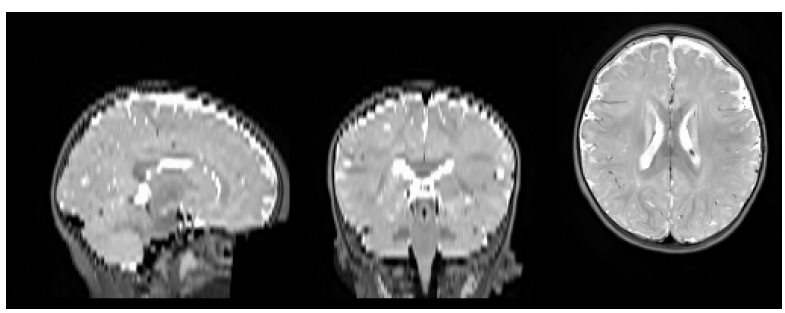
This section compares two model configurations, one which is considered the augmented model. The augmented model differs from the others due to its addition of more training data. It is trained on D2+D4 inputs, resulting in double the data size. Note that a properly augmented training setup would require substantially more augmentations.

The result is presented in Figure 4.7. Two models are presented here, one trained on solely D2 input, i.e. the model denoted as *Deep U-Net*, and one trained on D2+D4 input, i.e. the augmented version. Furthermore, each model is tested with two types of LR anisotropic MRI data, D2 or D4. Both model performs well when testing with D2 MRI, although the *Deep U-Net* slightly outperforms the augmented when comparing the quantitative metrics in Table 4.2. When testing with D4 MRI, the augmented model performs substantially better than the non-augmented model.

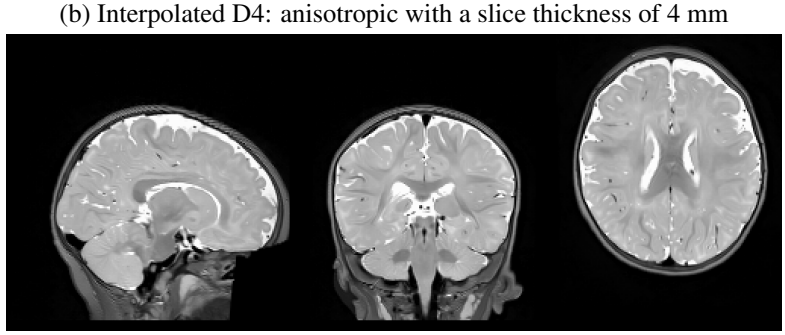
In conclusion, the augmented model proves ability to reconstruct from two kinds of anisotropy, compared two the non-augmented model which can only reconstruct one kind of anisotropy. The best result, both visually and quantitatively, is still received from the *Deep U-Net*, however, the most generalizable result is obtained from the augmented model.



(a) Interpolated D2: anisotropic with a slice thickness of 2 mm

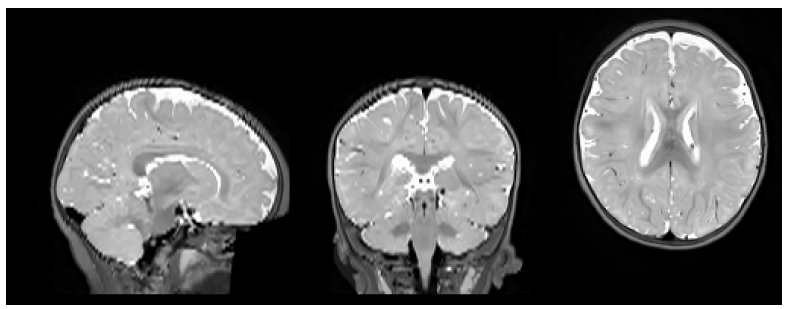


(b) Interpolated D4: anisotropic with a slice thickness of 4 mm

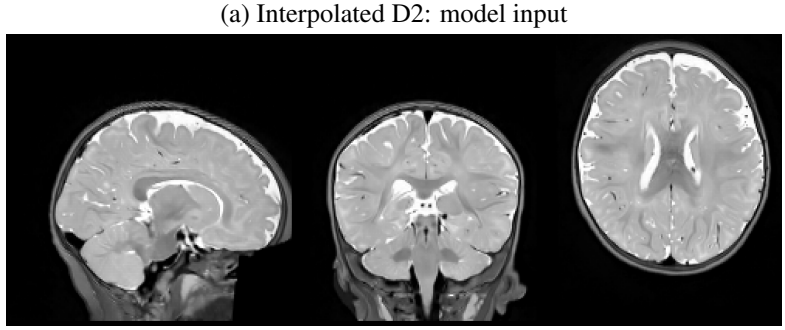


(c) Isotropic GT

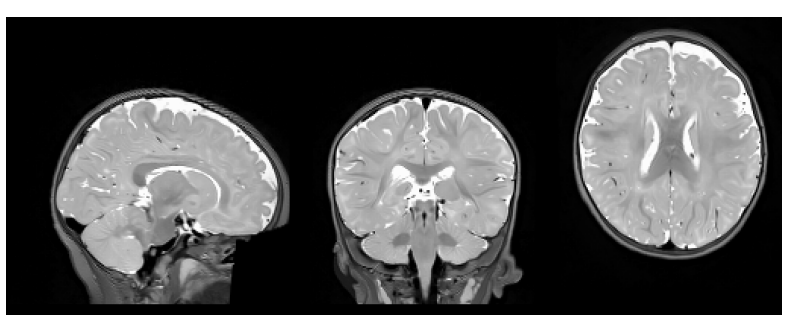
Figure 4.1: The anisotropic inputs used as a benchmark for evaluating the model performance. Two different grades of anisotropy is presented in (a) and (b). Both images are interpolated with cubic spline interpolation to restore the shape. (c) shows the T2w isotropic GT



(a) Interpolated D2: model input

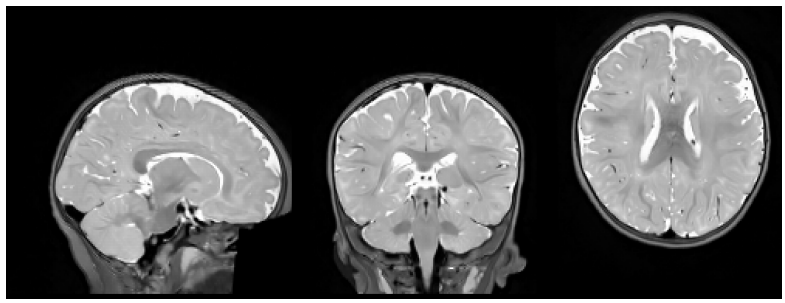


(b) Model output

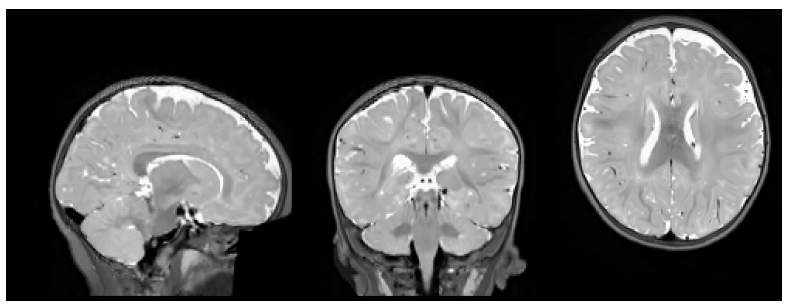


(c) Isotropic GT

Figure 4.2: Results from the best-performing model, the configuration is summarized in Table 4.1. Figure (a) presents the T2w anisotropic input, (b) the model output, and (c) the T2w isotropic GT.



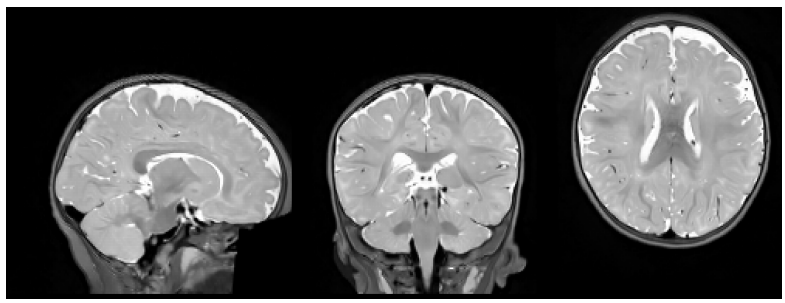
(a) Deep U-Net



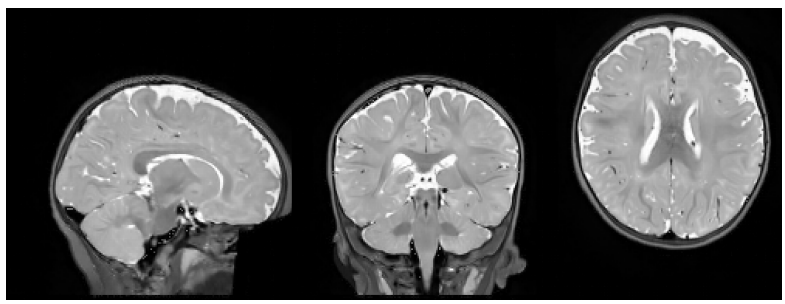
(b) Shallow U-Net

(c) Isotropic GT

Figure 4.3: Output from two models differing in network depth and feature channels, (a) presenting the deeper network compared to (b) which have fewer layers and feature channels. Network configurations are specified in Table 3.4.



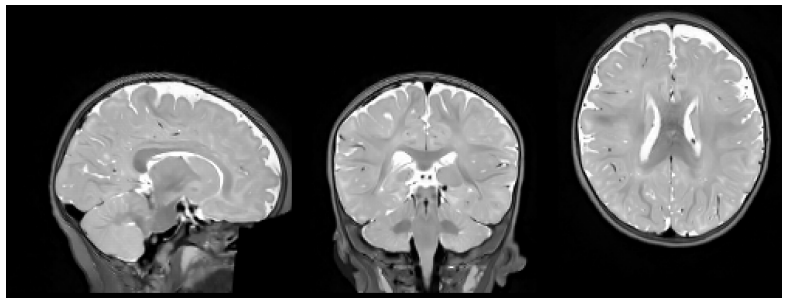
(a) Patch size: $32 \times 32 \times 32$ mm^3



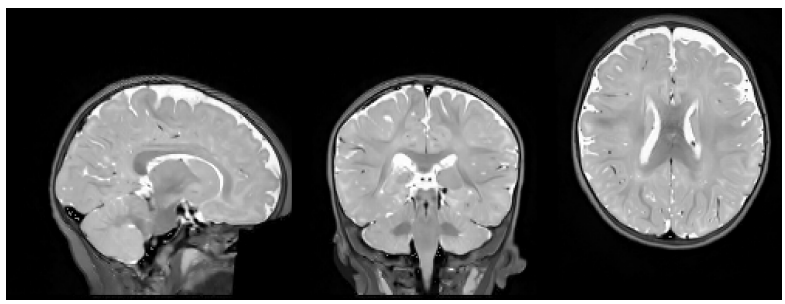
(b) Patch size: $64 \times 64 \times 64$ mm^3

(c) Isotropic GT

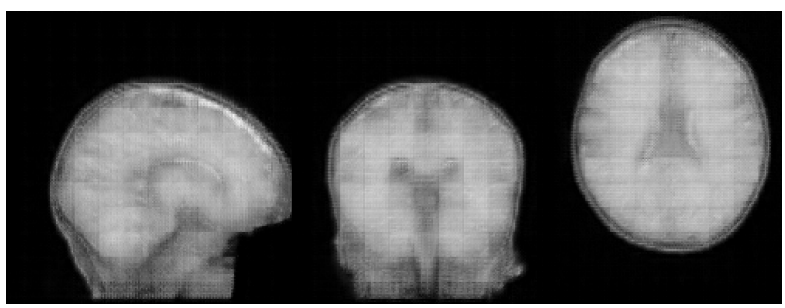
Figure 4.4: Output from the deep U-Net, with altering patch size of (a), $32 \times 32 \times 32$ mm^3, and (b), $64 \times 64 \times 64$ mm^3.



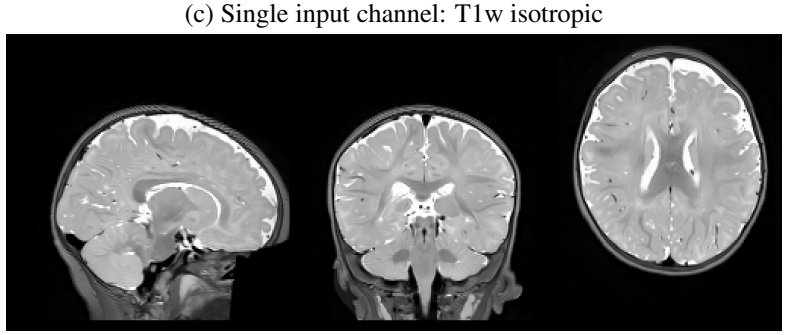
(a) Double input channels: T1w isotropic and T2w anisotropic



(b) Single input channel: T2w anisotropic

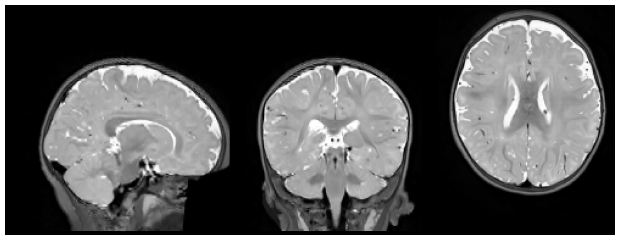


(c) Single input channel: T1w isotropic

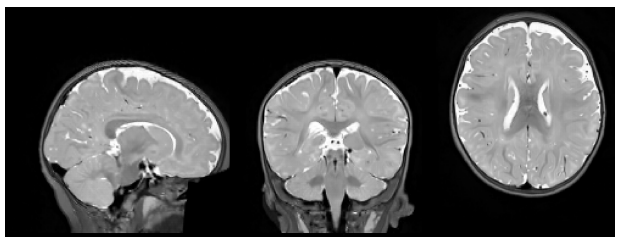


(d) Isotropic GT

Figure 4.5: Output from three models differing in the input data and the number of input channels, with (a), trained on T1w isotropic and T2w isotropic input, (b), trained on only T2w anisotropic input, and (c), trained on only T1w isotropic input.



(a) No residual unit



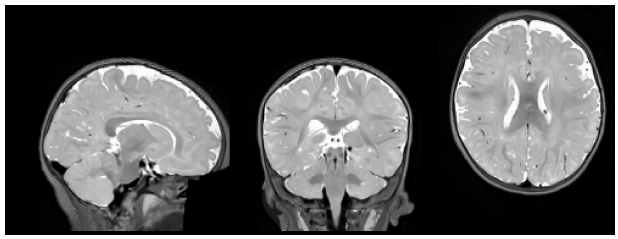
(b) Residual unit of size 2

(c) Residual unit of size 10

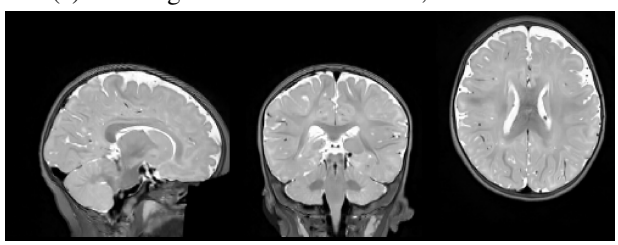
(d) Residual unit of size 18

(e) Isotropic GT

Figure 4.6: Residual unit comparison with residual units of size 0, 2, 10, and 18.

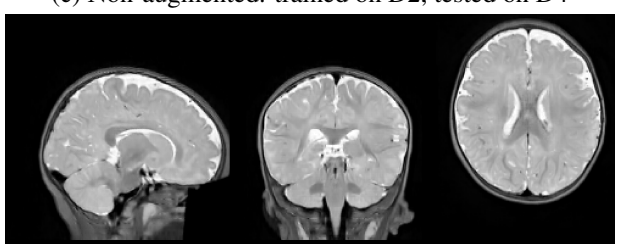


(a) Non-augmented: trained on D2, tested on D2

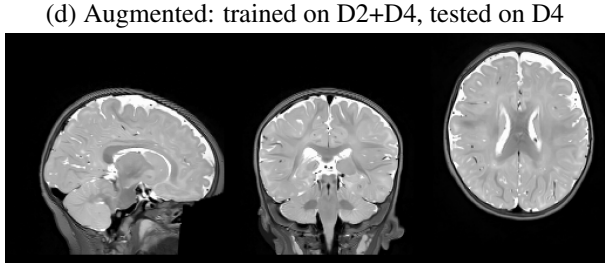


(b) Augmented: trained on D2+D4, tested on D2

(c) Non-augmented: trained on D2, tested on D4



(d) Augmented: trained on D2+D4, tested on D4



(e) Isotropic GT

Figure 4.7: Augmentation comparison. (a) and (c) are trained solely on D2 MRI, and (b) and (d) are trained on D2+D4 MRI. Furthermore, (a) and (b) are tested on D2 MRI, and (c) and (d) are tested on D4 MRI.

Chapter 5

Discussion

The primary goal of this project was to implement a deep learning pipeline capable of reconstructing isotropic T2w MRI from anisotropic T2w MRI. The pipeline is considered successful if the model outputs an isotropic T2w MRI which outperforms the benchmark, i.e. the interpolated anisotropic T2w MRI. Preferably, the output MRI also present quality close to the GT isotropic MRI.

According to these criteria, the proposed deep learning pipeline can be considered successful and is capable of performing anisotropic to isotropic MRI reconstruction. However, several additional factors contribute to the overall performance of the model, including generalizability and implementation feasibility. These factors, along with the overall performance will be further discussed in this chapter.

5.1 Overall Performance

Depth

Figure 4.2 presents the output of the best-performing model, together with the input (i.e. , the benchmark) and the GT MRI. The results indicate that a deeper model, with an increased number of layers as well as a higher number of feature channels per layer, outperforms the more shallow configuration, as shown in Figure 4.3. This behavior is expected, as deeper networks are generally able to model more complex representations and capture higher-level features, though often requiring more data and greater compute times. Consequently, the network could eventually reach a

limit, where an even deeper architecture would not enhance the results.

For example, repeated downsampling in the spatial dimensions can lead to a loss of spatial information, such as edges and small anatomical structures. The skip connections in the U-net architecture are designed to compensate for this issue by preserving spatial information across the network. Nevertheless, even with skip connections, the lowest levels of the network may still have a limited contribution. In this model, i.e. the 6-layer U-net, the deepest levels operate on a spatial resolution of $1 \times 1 \times 1$ mm^3. It is therefore possible that these compressed representations do not significantly contribute to the final reconstruction. An interesting direction for future work would be to evaluate more network depths or to perform an explainability analysis in order to better understand the contribution of different layers.

Patch Size

The smaller patches of size $32 \times 32 \times 32$ mm^3 outperformed the larger patches of size $64 \times 64 \times 64$ mm^3. One definitive explanation is that smaller patches increase the amount of training data, as more samples can be extracted from each volume. However, this does not imply that a smaller patch size is always preferable. In deeper networks, where the spatial dimensions are progressively downsampled at each layer, the patch size cannot be too small. For the proposed model, changing the patch size to $16 \times 16 \times 16$ mm^3 would result in very small spatial dimensions in the deeper levels, potentially complicating the learning process and the networks ability to capture contextual information. Consequently, the choice of patch size represents a trade-off between maximizing the number of training samples and still preserving spatial context.

Patch size also influences the type of contextual information extracted in the model. Larger patches provide more global anatomical context, and using the full image would result in learning even more global features. In contrast, smaller patches encourage the model to focus on local structures, such as edges and tissue boundaries. As a result, models trained on smaller patches may rely on more generalizable local features rather than subject specific anatomy, a hypothesis which has been confirmed in previous work [16]. Based on this, it is possible that a model could use adult brain MRI data to improve performance on infant brain reconstruction, at least if training is based on small local patches which operate on generaliz-

able features. While infant and adult brain MRI differ in size, contrast and signal intensities [43], low-level image features, such as edges and brain structures, remain similar. Consequently, a model may learn features from adult brain data that are applicable to infant brain reconstruction. Combining adult and infant brain MRI data for training would enable access to larger datasets and potentially result in a more robust model. A possible strategy would be to pretrain the network on adult brain MRI data to learn generic image features, followed by fine-tuning on infant brain data.

Input Channels

The best-performing model has two input channels to make use of as much information as possible. The result of this is shown in Figure 4.5. Using a single input consisting only of anisotropic T2w MRI yields slightly inferior results, but remains a good alternative. In contrast, using only isotropic T1w MRI as input results in outputs that are substantially worse than the benchmark MRI.

The network is primarily optimized to enhance anisotropic T2w MRI by filling in missing through-plane information, rather than performing cross-modality synthesis, i.e. synthesizing T2w MRI from T1w MRI. As a result, the model appears to learn a task closer to structured de-blurring within the T2w domain, rather than generating T2w images from scratch based on T1w information. Cross-modality synthesis is generally a more challenging problem, since T1w and T2w MRI present actual differences [2]. Abnormalities and structures appearing distinct in one modality, can appear smoothed or not visible in another.

As for now, methods for cross-modality synthesis are usually more complex than the U-Net presented in this project. CycleGANs [44] and pix2pix [45] are common choices for similar tasks. Furthermore, when working in the field of multi contrast super-resolution, i.e. SR between modalities, transformer- and attention-based methods reach best performance. Consequently, the observed performance gap when T2w input is removed is expected, given the complexity of cross-modality tasks. Additionally, the limited size of the training dataset may as well negatively affect the generative learning opportunities.

Nevertheless, the double input outperforms the single input T2w version, as shown in Table 4.2. This highlights that the T1w MRI succeeds in providing addi-

tional information which enhance learning, as long as it is accompanied by the T2w MRI. A similar result is achieved in recent literature, where the isotropic T1w MRI acted solely as guidance to the SR process [19].

Residual Units

The network was constructed with residual units, where the configuration with residual units of size 10 achieved the best performance, as shown in Figure 4.6. It is also very clear that the network without any residual units performs substantially worse than the others. From this it can be concluded that incorporating residual units in the U-Net is beneficial, which was expected since several previous studies on SR in MRI has successfully incorporated residual units [20] [11] [14].

Interestingly, increasing the number of residual units beyond 10 did not lead to further improvements. In particular, the configuration with 18 residual units resulted in degraded performance. One possible explanation is that deep residual blocks increase model complexity and compression, which may lead to overfitting.

Augmentations

An interesting and promising result is the effect of the augmentations. The results clearly visualizes the difference when inputting a D4 anisotropic MRI in the augmented versus the non-augmented model. The augmented model outputs a high-quality MRI, proving that additional training data does affect performance. However, the model is still only capable of reconstructing the specific anisotropy it was trained on. Hence, by introducing additional data augmentations, the model could potentially learn to reconstruct those anisotropy patterns as well. At some point, with enough variety in the training data, a more generalizable model might be reached.

However, augmented data may not be able to fully replace new, unseen data. It is possible that, even with an increasing number of augmentations, the performance is limited, and the model would require new information for improvement, i.e. an expanded dataset. This has to be further tested to know for sure.

Loss Function

Some experimenting with loss functions was performed, since several previous studies highlighted the importance of loss functions, particularly the involvement of a perceptual loss to boost performance. One study, with a similar setup as in this project, evaluated the effect of the loss function within a U-Net structure for MRI SR [46]. It proved that the use of both perceptual loss, a VGG-19 network in this case, and pixel-wise loss, significantly enhanced the performance. In contrast, despite many attempts with perceptual loss implementation in this project, there was no real success. Once again, this could be connected with the way the model learns. If it learns very specific de-blurring patterns, perceptual loss within the training might not help to improve. Other loss functions which were tested did also not affect the result. This suggests that more training data, is needed for better learning.

Normalization

Incorporating normalization layers within the network led to impaired performance of the model. Usually, normalization helps stabilize and improve performance, and the use of batch normalization is common within deep learning [27]. However, recent studies have reported that batch normalizaton, specifically within SR, seems to negatively affect the model performance. The normalization affects the finer textures and may also introduce artifacts. This has led to a trend of omitting batch normalization within SR tasks [47]. Hence, the effect of normalization layers in this project aligns well with the literature, supporting the decision to not include any normalization layers in the U-Net.

Additionally, another explanation to the negative performance when including normalization could be the patch-wise approach. When including normalization, the output consistently showed images with stitching artifacts. This suggests that errors arise when patches are normalized independently, leading to inconsistencies between neighboring patches. An interesting direction for future work would be to incorporate normalization into a similar network trained on whole images rather than on patches.

5.2 Limitations

The primary limitation of the model is its robustness, i.e. its ability to generalize to data that differs slightly from the training distribution. A clear example of this behavior is shown in Figure 4.7, where a D4 anisotropic MRI input produces poor reconstruction quality when inferred using a model trained exclusively on D2 anisotropic data. This indicates that the model becomes overfitted, learning highly specific patterns rather than capturing more general anatomical structure.

Data

Several observations point to the size of the available dataset as the limiting factor. When the amount of training data is too small, the model struggle to learn all complex patterns, resulting in poor generalization. In such cases, performance may reach an upper limit beyond which further training does not lead to meaningful improvements.

To address this limitation, data augmentation was explored, resulting in a model performing well on both D2 and D4 anisotropy. Based on this observation, extended augmentation strategies have been considered, including the use of additional downsampling factors as well as downsampling along different spatial directions. The aim of the augmentation approach is to reach a more generalizable model by increasing both the size and diversity of the training dataset.

The augmentation process is still ongoing and the results obtained so far are promising and suggest that it could help for generalization. Beyond augmentation, another potential approach to increasing the size of the training data would be to, as already suggested, incorporate adult brain MRI data. For example, the model could be pretrained on adult brain images to learn generic image features, and then fine-tuned on infant brain data. Alternatively, an existing foundation model trained on adult brain MRI could be adapted to the task through fine-tuning.

Another way to address the limited data problem is to implement an unsupervised or self-supervised learning strategy. Several studies have shown promising results in learning only from the LR anisotropic data [15] [16] [18] [19]. The studies use different strategies, such as creating lower LR MRI and pairing with the LR MRI, using the in-plane slices as GT, or implementing a ZSSR method. With

limited supervised data available, it could also be an option to address a combined approach, where the real LR-HR pairs could be used as evaluation for the unsupervised method.

Model

The model architecture itself may impose limitations on performance. While the U-Net is widely used and effective for SR tasks, it is also sometimes outperformed by more complex architectures. The use of attention- and transformer-based methods, as well as more generative approaches such as Generative Adversarial Network (GAN) and diffusion models (DDPMs) [48] pose as the state-of-art methods, often outperforming the U-Net architecture [27]. In this project, the issue of pixel-wise and specialized learning without perceptual understanding, could be explained by the simpleness of the U-Net model. A GAN focuses on the perceptual quality by introducing a adversarial loss. The loss is computed between the generator and the discriminator, enabling a optimization competition between the two. Simultaneously, the generator uses a pixel-based loss similar to the one in this project. GAN-based SR has already been successfully implemented in several studies [49] [50] [51]. A possible improvement of this project would be to extend the model into a GAN, using the U-Net as generator and adding an additional discriminator. However, GANs come with certain drawbacks, such as training instability and mode collapse [27]. Furthermore, it is not certain that an extended network complexity would enhance the result, since the issue could still be the data limitation. With very limited data, model adjustments may not solve the issue. Due to this, a future approach would be to first extend the dataset with augmentations, and then implement more complex network structures.

Another option, instead of using a different architecture, is to introduce additional conditioning within the U-Net. The use of isotropic T1w input can already be seen as a way of conditioning. Inputting the T1w MRI in deeper layers as well, inspired by the work of Iwamoto [19], may affect the network in interesting ways. Further conditioning could be to incorporate the anisotropic slice thickness, which is already known, at some level of the network. Possibly, this could facilitate the learning for the model since it provides extra information.

5.3 Applications

The global purpose of the project was to contribute to the overall research within the field, eventually contributing to SR applications in clinical settings. The direct purpose of the project was to construct a deep learning pipeline which can be directly implemented as an application at University of Pennsylvania.

Global

The proposed model is capable of converting anisotropic T2w MRI into isotropic reconstructions. For a potential implementation in a clinical setting, several aspects must be considered. First, the model must be adapted to and evaluated on the specific data it is intended to operate on. This would involve training or fine-tuning the network on site-specific data, followed by validation. Since MRI data can vary in terms of anisotropy, slice thickness, and acquisition direction, the model may need to be adjusted accordingly. This could either be done by very precise training for the task or, which is preferable, by creating a more robust model. As already mentioned, such robustness may be achieved by extended data augmentation.

Another critical aspect is the safety and reliability of the model. Since the model generates new information, it is essential that clinically relevant details are neither omitted nor falsely introduced. For example, if a pathological structure such as a tumor is on the order of the slice thickness, it may not be visible in the anisotropic input image. During reconstruction to the isotropic volume, there is a risk that such a structure remains undetected, which could mislead clinicians. The model could possibly also hallucinate pathological findings, resulting in false-positive findings. In an ideal scenario, the model would learn contextual patterns, such as changes in surrounding tissue, that allow it to understand pathology even when it is not clearly visible in the input. Achieving this would require training on datasets that include a wide range of pathological cases, such as tumors or atrophy. It also requires a high-performance model with a perceptual understanding.

However, the performance can always be weighted against the intended use of the application. If clinicians are clearly informed that the reconstructed MRI represents an enhanced visualization of the original anisotropic input rather than new diagnostic information, the acceptable performance requirements can be adjusted

accordingly. In such a scenario, the reconstructed images offer improved visual interpretability while still potentially missing certain details. Importantly, in situations where no HR T2w MRI can be acquired, such reconstructions may still represent the best available option, if clinicians are aware of the applications limitations. This serves for diagnosing different brain diseases, as well as performing downstream tasks on the isotropic MRI. Some downstream task are not possible without a proper isotropic MRI, making the model useful in such cases.

Something else which is important to take into account, if implementing in a clinical setting, is the operational complexity. First of all, an application has to present high usability, with easy instructions and intuitive functionality. Secondly, the installation and maintenence has to be either performed by the developer or adapted for wider understanding. Lastly, the preprocessing steps should be minimized. If the application requires a highly preprocessed MRI, all these steps has to be performed in advance, taking time from the real use. With this in mind, preprocessing in this project was minimized, and methods like, for example, skull stripping was avoided due to this. The present preprocessing pipeline could also benefit from being integrated in the final application to ensure correct coregistration and normalization. Another thing to consider is to limit the input channels to only use T2w anisotropic input, since T1w isotropic input may sometimes not be available.

Beyond clinical use, the model also has potential for generating extended quantities of isotropic MRI data for research purposes. In this context, the primary concern is whether the generated images accurately reflect the statistical and anatomical properties of real isotropic brain MRI. If the model were to produce unrealistic or biased reconstructions, this could negatively impact downstream research. However, if the generated data can be shown to be representative of real brain anatomy, the model could serve as a valuable tool for research applications. In this setting, exact voxel-level accuracy is less critical than in clinical diagnosis, as the primary goal is to support the development and evaluation of new methods that may ultimately benefit patient care.

Direct

As for the direct implementation at the University of Pennsylvania, the process has already started. The idea is to re-train the model with a large amount of augmented

data from the current dataset, in combination with all available supervised LR-HR pairs at the University of Pennsylvania. This process require evaluations with the University of Pennsylvania data to ensure successful results.

If successful, the aim is to reconstruct all existing anisotropic T2w MRI to isotropic. With the isotropic T2w MRI available, myelination analysis and further research can be performed, as explained in Section 1.3.2.

5.4 Future Work

As already discussed, there are a wide range of possibilities for future work. The following list summarizes the most promising directions.

- **Data augmentation:** An investigation whether extended data augmentation can improve model generalization serves as the next step in the process. The aim is to train the model on a wider range of anisotropic data, varying in slice thickness and acquisition direction.
- **Evaluation on other datasets:** To further evaluate the model, the existing pretrained model can be tested directly on other datasets, such as the CHOP dataset.
- **Dataset extension:** The ideal solution would be to incorporate additional infant brain MRI data in the training dataset. However, since infant brain MRI datasets are limited, an alternative approach is to investigate the effect of adding adult brain MRI data in the training dataset. Furthermore, it would also be of interest to train and evaluate the model exclusively on adult brain MRI data and compare the results with recent related studies. Lastly, addressing an unsupervised or self-supervised approach could help work around the data limitation problem.
- **Model adjustments:** Possible adjustments would be to try new architectures or extensions of the model, such as GAN or diffusion models, or to test an already existing foundation model and fine-tune it.

Chapter 6

Conclusion

The primary aim of this thesis was to develop a deep learning pipeline capable of anisotropic to isotropic reconstruction of T2w infant brain MRI. This resulted in a patch-based supervised residual U-Net which outperformed the interpolation-based benchmark. The best-performing network presents a PSNR of 38.27 dB, compared to the benchmark PSNR of 29.85 dB.

However, the main model was trained to reconstruct only one type of anisotropy, corresponding to a slice thickness of 2 mm. As a consequence, it performed poorly on anisotropic MRI with a different slice thickness. To address this limitation, augmentations were introduced. A second model, trained on two types of anisotropy, with slice thicknesses of 2 mm and 4 mm, managed to successfully reconstruct both types during testing. This result is promising for future work, where an extension of the augmentations could lead to a more generalized model.

While augmentations may improve performance and generalization, extending the dataset is likely the most effective way to achieve high contextual learning. Another direction for future work is to explore different model architectures, as well as the use of foundation models.

Since this work is specifically adjusted to infant brain MRI, potential applications would be focused on infant neuroimaging. One possible downstream task, enabled by successfully reconstructed isotropic T2w MRI, is myelination quantification. Furthermore, the model could possibly be re-trained on adult brain MRI data, resulting in different applications and downstream tasks.

Bibliography

[1] Ham B. J. Pyun S. B. Kim B. J. Tae, W. S. Current clinical applications of structural mri in neurological disorders. *Journal of clinical neurology (Seoul, Korea), 21(4), 277–293.*, 2025.

[2] Yidan Feng, Sen Deng, Jun Lyu, Jing Cai, Mingqiang Wei, and Jing Qin. Bridging mri cross-modality synthesis and multi-contrast super-resolution by fine-grained difference learning. *IEEE Transactions on Medical Imaging*, 44(1):373–383, 2025.

[3] Jinsha Tian, Canjun Xiao, and Hongjin Zhu. Isotropic brain mri reconstruction from orthogonal scans using 3d convolutional neural network. *Sensors*, 24(20), 2024.

[4] Abdulhamid Muhammad, Supavadee Aramvith, Khanita Duangchaemkarn, and Ming-Ting Sun. Brain mri image super-resolution reconstruction: A systematic review. *IEEE Access*, 12:156347–156362, 2024.

[5] Amir Pasha Mahmoudzadeh and Nasser H. Kashou. Interpolation-based super-resolution reconstruction: effects of slice thickness. *Journal of Medical Imaging*, 1(3):034007, 2014.

[6] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010.

[7] Yuanyuan Jia, Ali Gholipour, Zhongshi He, and Simon K. Warfield. A new sparse representation framework for reconstruction of an isotropic high spatial resolution mr volume from orthogonal anisotropic resolution scans. *IEEE Transactions on Medical Imaging*, 36(5):1182–1193, 2017.

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 184–199, Cham, 2014. Springer International Publishing.

[9] Chi-Hieu Pham, Aurélien Ducournau, Ronan Fablet, and François Rousseau. Brain mri super-resolution using deep 3d convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 197–200, 2017.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[11] Chi-Hieu Pham, Carlos Tor-Díez, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Passat, and François Rousseau. Multiscale brain mri super-resolution using deep 3d convolutional networks. *Computerized Medical Imaging and Graphics*, 77:101647, 2019.

[12] Akshay S Chaudhari, Zhongnan Fang, Feliks Kogan, Jeff Wood, Kathryn J Stevens, Eric K Gibbons, Jin Hyung Lee, Garry E Gold, and Brian A Hargreaves. Super-resolution musculoskeletal mri using deep learning. *Magnetic resonance in medicine*, 80(5):2139–2154, 2018.

[13] Ozan Oktay, Wenjia Bai, Matthew Lee, Ricardo Guerrero, Konstantinos Kamnitsas, Jose Caballero, Antonio de Marvao, Stuart Cook, Declan O'Regan, and Daniel Rueckert. Multi-input cardiac image super-resolution using convolutional neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 246–254. Springer, 2016.

[14] Jinglong Du, Zhongshi He, Lulu Wang, Ali Gholipour, Zexun Zhou, Dingding Chen, and Yuanyuan Jia. Super-resolution reconstruction of single anisotropic 3d mr images using residual convolutional neural network. *Neurocomputing*, 392:209–220, 2020.

[15] Amod Jog, Aaron Carass, and Jerry L. Prince. Self super-resolution for magnetic resonance images. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing*

and Computer-Assisted Intervention - MICCAI 2016, pages 553–560, Cham, 2016. Springer International Publishing.

[16] Can Zhao, Blake E. Dewey, Dzung L. Pham, Peter A. Calabresi, Daniel S. Reich, and Jerry L. Prince. 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.

[17] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. *CoRR*, abs/1712.06087, 2017.

[18] Rotem Benisty, Yevgenia Shteynman, Moshe Porat, Anat Ilivitzki, and Moti Freiman. Simple: Simultaneous multi-plane self-supervised learning for isotropic mri restoration from anisotropic data, 2025.

[19] Yutaro Iwamoto, Kyohei Takeda, Yinhao Li, Akihiko Shiino, and Yen-Wei Chen. Unsupervised mri super resolution using deep external learning and guided residual dense network with multimodal image priors. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(2):426–435, 2023.

[20] Malo Gicquel, Ruoyi Zhao, Anika Wuestefeld, Nicola Spotorno, Olof Strandberg, Kalle Åström, Yu Xiao, Laura EM Wisse, Danielle van Westen, Rik Ossenkoppele, Niklas Mattsson-Carlgren, David Berron, Oskar Hansson, Gabrielle Flood, and Jacob Vogel. Converting t1-weighted mri from 3t to 7t quality using deep learning, 2025.

[21] Mareike Grotheer, Mona Rosenke, Hua Wu, Holly Kular, Francesca R Querdasi, Vaidehi S Natu, Jason D Yeatman, and Kalanit Grill-Spector. White matter myelination during early infancy is linked to spatial gradients and myelin content at birth. *Nature communications*, 13(1):997, February 2022.

[22] Tugba Akinci D'Antonoli, Ramona-Alexandra Todea, Nora Leu, Alexandre N. Datta, Bram Stieltjes, Friederike Pruefer, and Jakob Wasserthal. Development and evaluation of deep learning models for automated estimation of myelin maturation using pediatric brain mri scans. *Radiology: Artificial Intelligence*, 5(5):e220292, 2023.

[23] Eric Feczko, Sally M Stoyell, Lucille A. Moore, Dimitrios Alexopoulos, Maria Bagonis, Kenneth Barrett, Brad Bower, Addison Cavender, Taylor A. Chamberlain, Greg Conan, Trevor KM Day, Dhruman Goradia, Alice Graham, Lucas Heisler-Roman, Timothy J. Hendrickson, Audrey Houghton, Omid Kardan, Elizabeth A Kiffmeyer, Erik G Lee, Jacob T. Lundquist, Carina Lucena, Tabitha Martin, Anurima Mummaneni, Mollie Myricks, Pranav Narnur, Anders J. Perrone, Paul Reiners, Amanda R. Rueter, Hteemoo Saw, Martin Styner, Sooyeon Sung, Barry Tiklasky, Jessica L Wisnowski, Essa Yacoub, Brett Zimmermann, Christopher D. Smyser, Monica D. Rosenberg, Damien A. Fair, and Jed T. Elison. Baby open brains: An open-source repository of infant brain segmentations. *bioRxiv*, 2024.

[24] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[25] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Atlanta, GA, 2013.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852, 2015.

[27] Mohammad Khateri, Serge Vasylechko, Morteza Ghahremani, Liam Timms, Deniz Kocanaogullari, Simon K. Warfield, Camilo Jaimes, Davood Karimi, Alejandra Sierra, Jussi Tohka, Sila Kurugol, and Onur Afacan. Mri super-resolution with deep learning: A comprehensive survey, 2025.

[28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[29] Xiaodan Hu, Mohamed A. Naiel, Alexander Wong, Mark Lamm, and Paul Fieguth. Runet: A robust unet architecture for image super-resolution. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 505–507, 2019.

[30] Agnesh Chandra Yadav, Maheshkumar H. Kolekar, and Mukesh Kumar Zope. Modified recurrent residual attention u-net model for mri-based brain tumor segmentation. *Biomedical Signal Processing and Control*, 102:107220, 2025.

[31] Rahman Maqsood, Fazeel Abid, Jawad Rasheed, Onur Osman, and Shtwai Alsubai. Optimal res-unet architecture with deep supervision for tumor segmentation. *Frontiers in Medicine*, Volume 12 - 2025, 2025.

[32] Omar Elharrouss, Yasir Mahmood, Yassine Bechqito, Mohamed Adel Serhani, Elarbi Badidi, Jamal Riffi, and Hamid Tairi. Task-based loss functions in computer vision: A comprehensive review, 2025.

[33] Pontus Andersson, Jim Nilsson, Tomas Akenine-Möller, Magnus Oskarsson, Kalle Åström, and Mark D. Fairchild. Flip: A difference evaluator for alternating images. *Proc. ACM Comput. Graph. Interact. Tech.*, 3(2), August 2020.

[34] Gustav Grund Pihlgren, Konstantina Nikolaidou, Prakash Chandra Chhipa, Nosheen Abid, Rajkumar Saini, Fredrik Sandin, and Marcus Liwicki. A systematic performance analysis of deep perceptual loss networks: Breaking transfer learning conventions, 2024.

[35] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016.

[36] Mukhriddin Arabboev, Shohruh Begmatov, Mokhirjon Rikhsivoev, Khabibullo Nosirov, and Saidakmal Saydiakbarov. comprehensive review of image super-resolution metrics: classical and ai-based approaches. *Acta IMEKO*, 13:1–8, 03 2024.

[37] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[38] Jim Nilsson and Tomas Akenine-Möller. *Understanding SSIM*. arXiv.org, June 2020.

[39] Tauhidul Islam, Md. Sadman Hafiz, Jamin Rahman Jim, Md. Mohsin Kabir, and M.F. Mridha. A systematic review of deep learning data augmentation in

medical imaging: Recent advances and future research directions. *Healthcare Analytics*, 5:100340, 2024.

[40] Aryan Kalluvila. Super-resolution of brain mri via u-net architecture. *International Journal of Advanced Computer Science and Applications*, 14, 01 2023.

[41] M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, Vishwesh Nath, Yufan He, Ziyue Xu, Ali Hatamizadeh, Andriy Myronenko, Wentao Zhu, Yun Liu, Mingxin Zheng, Yucheng Tang, Isaac Yang, Michael Zephyr, Behrooz Hashemian, Sachidanand Alle, Mohammad Zalbagi Darestani, Charlie Budd, Marc Modat, Tom Vercauteren, Guotai Wang, Yiwen Li, Yipeng Hu, Yunguan Fu, Benjamin Gorman, Hans Johnson, Brad Genereaux, Barbaros S. Erdal, Vikash Gupta, Andres Diaz-Pinto, Andre Dourson, Lena Maier-Hein, Paul F. Jaeger, Michael Baumgartner, Jayashree Kalpathy-Cramer, Mona Flores, Justin Kirby, Lee A. D. Cooper, Holger R. Roth, Daguang Xu, David Bericat, Ralf Floca, S. Kevin Zhou, Haris Shuaib, Keyvan Farahani, Klaus H. Maier-Hein, Stephen Aylward, Prerna Dogra, Sebastien Ourselin, and Andrew Feng. Monai: An open-source framework for deep learning in healthcare, 2022.

[42] Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *CoRR*, abs/1904.00625, 2019.

[43] Jessica Dubois, Marianne Alison, Serena J. Counsell, Lucie Hertz-Pannier, Petra S. Hüppi, and Manon J.N.L. Benders. Mri of the neonatal brain: A review of methodological challenges and neuroscientific advances. *Journal of Magnetic Resonance Imaging*, 53(5):1318–1343, 2021.

[44] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017.

[45] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[46] Mohammad Javadi, Rishabh Sharma, Panagiotis Tsiamyrtzis, Shishir Shah, Ernst L. Leiss, and Nikolaos V. Tsekos. From perception to precision: Navigating perceptual loss in mri super-resolution. In *2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 57–61, 2023.

[47] Dawa Chyophel Lepcha, Bhawna Goyal, Ayush Dogra, and Vishal Goyal. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*, 91:230–260, 2023.

[48] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *CoRR*, abs/2006.11239, 2020.

[49] Jiancong Wang, Yuhua Chen, Yifan Wu, Jianbo Shi, and James Gee. Enhanced generative adversarial network for 3d brain mri super-resolution. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3616–3625, 2020.

[50] Kuan Zhang, Haoji Hu, Kenneth Philbrick, Gian Marco Conte, Joseph D. Sobek, Pouria Rouzrokh, and Bradley J. Erickson. Soup-gan: Super-resolution mri using generative adversarial networks. *Tomography*, 8(2):905–919, 2022.

[51] Qi Wang, Lucas Mahler, Julius Steiglechner, Florian Birk, Klaus Scheffler, and Gabriele Lohmann. Disgan: Wavelet-informed discriminator guides gan to mri super-resolution with noise cleaning, 2023.

Master's Theses in Mathematical Sciences 2026:E7
ISSN 1404-6342

LUTFMA-3607-2026

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/