



MASTER THESIS IN ELECTRICAL ENGINEERING

Advanced machine learning for single-acquisition isotropic fetal MRI

Author:

Kay LÄCHLER

Section:

Electrical and Electronic Engineering

Supervisors:

Dr. Pol DEL AGUILA PLA (CIBM / BIG)

&

Dr. Meritxell BACH CUADRA (CIBM)

&

Prof. Michael UNSER (BIG)

Lausanne — 19.08.2022

Contents

1	Introduction	2
1.1	Anisotropic resolution in 3D imaging	2
1.2	3D fluorescence microscopy	3
1.2.1	Isotropic resolution in fluorescence microscopy	4
1.3	Fetal brain MRI	4
1.3.1	Simplified MRI model	5
1.3.2	Isotropic resolution in fetal brain MRI: State of the art	6
1.4	Aim of this Project	7
2	Restoring isotropic resolution in 3D imaging using a sample-specific CNN	8
2.1	Fluorescence microscopy	9
2.2	Fetal brain MRI	11
2.2.1	Generating the training data	11
2.2.2	Network architecture and training	12
2.2.3	Prediction	13
2.2.4	Using multiple volumes	14
2.3	Fetal brain MRI data	15
2.3.1	Simulated data	15
2.3.2	Clinical acquisitions	15
3	Results	17
3.1	Fluorescence microscopy	17
3.2	Fetal brain MRI	18
3.2.1	Simulated data	18
3.2.2	Clinical Acquisitions	21
3.2.3	Using multiple volumes	21
4	Discussion	24
4.1	Fluorescence microscopy	24
4.2	Fetal brain MRI	25
5	Appendix	30
5.1	Additional work	30
5.2	Supplementary figures	30

Chapter 1

Introduction

1.1 Anisotropic resolution in 3D imaging

In the scope of this project, 3D imaging refers to the acquisition of volumetric signal intensity data, such as fluorescence-light emission in the case of microscopy and proton-spin density in magnetic resonance imaging (MRI). This data is represented in the form of voxels¹, which describe a volume of width Δx , height Δy , and thickness Δz as depicted in Figure 1.1. Volumetric (3D) images often exhibit anisotropic resolution, meaning the resolution is different in the x -, y - and z -axes, which is a result of the acquisition process. In the applications concerning this project, volumetric data is acquired as a series of "2D" slices, which are stacked to create a 3D image. For simplicity, we will define the x - and y -axis such that the slices lie in the xy -plane, also denoted as *in-plane*, and the slices are stacked along the z -axis, also denoted as *through-plane*. Each slice is a collection of $N_x \times N_y$ voxels, which lets us define its width $w = N_x \Delta x$, height $h = N_y \Delta y$ and thickness Δz . The slice separation Δs is defined as the distance at which the slices are stacked along the z -axis. If $\Delta s = \Delta z$ the slices are stacked adjacent to each other, if $\Delta s < \Delta z$ the slices overlap, and if $\Delta s > \Delta z$ there are gaps between them. Given a series of N_z slices, the total thickness of the volumetric image is thus given by $d = N_z \Delta s$.

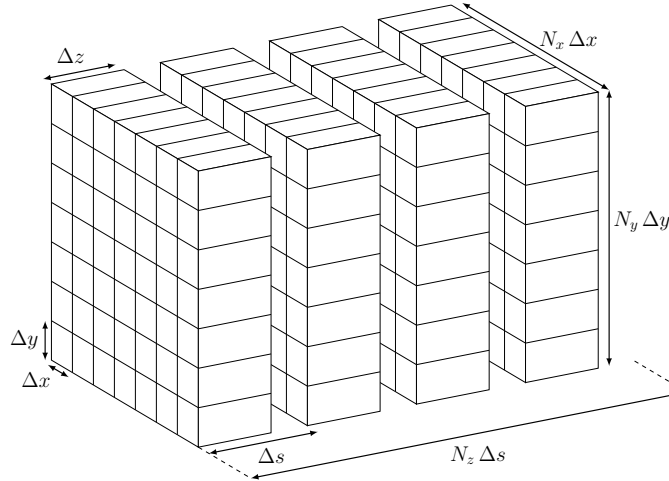


Figure 1.1: Diagram showing the definition of some 3D imaging parameters.

Assuming that $\Delta x = \Delta y$, which is true for most 3D imaging techniques, we can define the resolution ratio as $r = \frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y}$, which represents the resolution anisotropy present in a volumetric image. As the

¹voxel = volumetric pixel

slice thickness is proportional to the acquired signal strength [1], increasing Δz leads to a higher signal-to-noise ratio (SNR), while increasing Δs reduces the acquisition time. However, as Δz increases, r increases as well, and if $\Delta s > \Delta z$, the volume is undersampled, resulting in missing information. For that reason, a popular approach to both increase the SNR and decrease the acquisition time, while also optimizing the overall resolution, is to choose $\Delta x = \Delta y < \Delta z = \Delta s$, resulting in thicker and fewer slices, but still performing a complete sampling. The consequence of this choice is an anisotropic resolution, which is higher in-plane than through-plane ($r > 1$). This not only reduces the visual fidelity of the volume, when looking at the xz - or yz -slices, but also limits the use of automated numerical analysis tools. Restoring isotropic resolution in 3D images is thus an important step to facilitate both the visual and numerical analysis of volumetric data.

1.2 3D fluorescence microscopy

We started this project by trying to adapt the existing work on isotropic resolution reconstruction in 3D fluorescence microscopy to fetal brain MRI. For this reason, we will briefly summarize the fundamentals of 3D fluorescence microscopy here and present the paper that inspired this project. Two common techniques to produce 3D images in fluorescence microscopy are confocal microscopy [2] and light-sheet fluorescence microscopy (LSFM) [3].

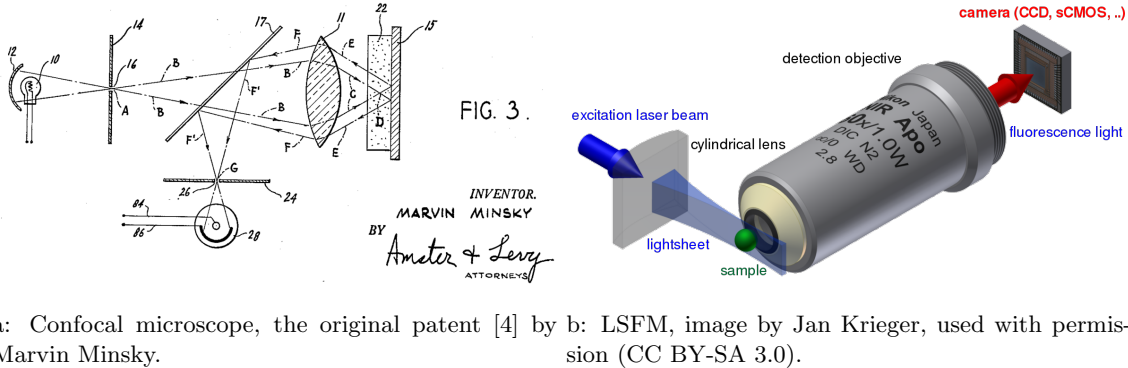


Figure 1.2: Simple diagrams, showing the two main techniques to produce 3D fluorescence microscopy images.

In confocal microscopy (Figure 1.2a), slices are acquired one pixel at a time, by focusing a light beam on one single point in the sample. The fluorescent emission is then filtered by a pinhole aperture before being acquired by the sensor, which reduces the influence of scattered out-of-focus emissions and gives rise to the name *confocal*. By sequentially moving the focal point in the xy -plane, an entire slice is acquired. Then, the focal point is moved along the z -axis and the process is repeated until the entire volume has been aggregated. The best in-plane resolution of confocal microscopes can reach around $\Delta x = \Delta y = 200$ nm, while the best through-plane resolution is $\Delta z = 790$ nm [5], resulting in a resolution ratio of $r = 3.95$.

In LSFM (Figure 1.2b), a laser together with a cylindrical lens produce a thin plane of light, which is used to excite thin slices of the sample. With this technique, the fluorescence emission of entire slices can be captured at once, massively increasing acquisition speed compared to confocal microscopy. However, generating a homogeneous light plane is difficult because the plane thickness is affected by divergence [6], which is one of the limiting factors concerning the through-plane resolution. Standard LSFM can reach an in-plane resolution of around $\Delta x = \Delta y = 290$ nm, with a through-plane resolution of $\Delta z = 820$ nm, corresponding to a resolution ratio of $r = 2.828$ [5]. In this project, we will only focus on LSFM, as this technique is superior in both resolution ratio and acquisition speed compared to confocal microscopy, at the cost of a comparably minimal loss of resolution.

1.2.1 Isotropic resolution in fluorescence microscopy

In their 2017 paper [7], Weigert et al. developed a pipeline to restore isotropic resolution of 3D fluorescence microscopy images by training a sample-specific neural network, which will be explained in detail in Section 2.1. In short, they use a simplified model of the fluorescence microscopy image formation process,

$$p = D_{x,r}(\tilde{h} * g), \quad (1.1)$$

where g is the original volumetric image, \tilde{h} is the rotated version of the PSF, $D_{x,r}$ is the downsampling operator in the x -axis by the resolution factor r , and p is the resulting volumetric image with a simulated bad resolution in the x -axis. The rotated PSF \tilde{h} is generated by rotating the original (or estimated) PSF h of the imaging system, such that the z -axis of h aligns with the x -axis of g , as demonstrated in Figure 1.3. The simulated low-resolution volume p is then upsampled with *bicubic* interpolation along the x -axis to create the upsampled volume $p_{\text{up}} = U_{x,r} p$. The xy -slices of p_{up} are then used as the training inputs, while the xy -slices of g are used as the training targets. After training, the network is then applied to the xz -slices of g to generate the predicted volume \tilde{g} with isotropic resolution. They tested their pipeline on synthetic, as well as real 3D fluorescence microscopy data and managed to produce impressive results.

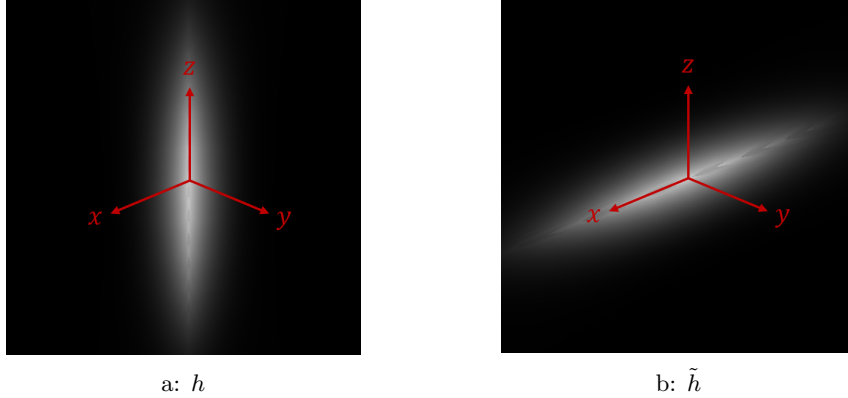


Figure 1.3: The original PSF h (a) and its rotated version \tilde{h} (b), used during the training data generation.

By incorporating the PSF of the acquisition system and the resolution ratio to generate the training data, this strategy can be used on all types of 3D microscopy images, and with some adjustments, it can be made to work on any 3D imaging data compatible with a convolutional model.

1.3 Fetal brain MRI

MR scans of the fetal brain have become a common addition to prenatal sonography if there is a known risk of neurodevelopmental disabilities or in cases where indications of an abnormality have been found during sonography. As such, it helps to improve pregnancy management and delivery preparations.

Acquiring MR images of the fetal brain [8] poses several challenges compared to the adult brain. The fetal brain volume is typically around $100 - 350 \text{ cm}^3$ [9] during the third trimester, which is up to 12 times smaller than the adult brain, which usually measures around 1200 cm^3 [10]. Furthermore, because the fetus is still inside the mother’s womb, the RF antenna of the MRI scanner is further away, reducing the signal strength and lowering the SNR. Furthermore, because it is impossible to restrict the movement of the fetus during acquisition, movement artefacts that deteriorate the image quality are a common issue in fetal brain MRI. In order to mitigate these, a common approach is to speed up the image acquisition, which is done by using very fast protocols, such as the HASTE, SS-FSE, and

SSH-TSE sequences [11, 12]. In particular, these sequences are able to acquire an entire T_2 -weighted image (T2WI) slice in less than 1 s [13]. An example of three orthogonal fetal brain MRI acquisition series of the same subject are shown in Figure 1.4, where we can see the effect of motion during the acquisition process, as well as the anisotropic resolution.

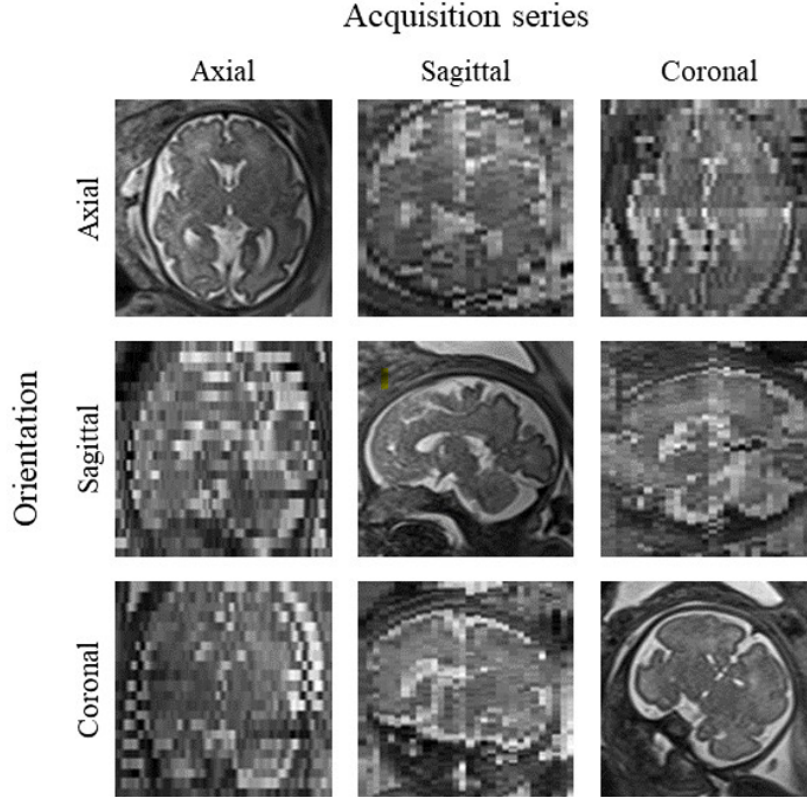


Figure 1.4: Example of a fetal brain MRI, showing three orthogonal acquisition series of the same subject (30 weeks gestational age), made using the HASTE sequence.

1.3.1 Simplified MRI model

In order to simulate the through-plane resolution in the in-plane slices, we need to define a simple mathematical model to represent the MRI acquisition process. Real MRI acquisitions are performed in Fourier space, and the approach is completely different from 3D fluorescence microscopy. However, if we neglect the presence of patient motion and bias field, the simplest way of modelling both processes is similar, using only the convolution with the system PSF followed by a downsampling step. The only change compared to (1.1) is that we additionally add white Gaussian noise N_G after the downsampling step, which gives us the MRI model

$$p = D_{x,r}(\tilde{h} * g) + N_G. \quad (1.2)$$

Here, $g = U_{z,r} f$ is the original volume f after *bicubic* upsampling in the z -direction by the resolution ratio r . This means g has isotropic voxel sizes, but is still blurred in the through-plane direction z . We chose to use the upsampled volume here because having isotropic voxel sizes simplifies the convolution operation ($\tilde{h} * g$) and also provides more training data, as there are more xy -slices to train on. The rotated PSF \tilde{h} corresponds to the original PSF h after aligning the z -axis of h with the

x -axis of g . The three-dimensional PSF h depends on the acquisition parameters, but can generally be approximated by a $\text{sinc}(\cdot)$ function with full width at half maximum (FWHM) of its central peak $\text{FWHM} = 1.2\Delta x = 1.2\Delta y$ in the in-plane slices and a Gaussian with $\text{FWHM} = \Delta z$ in the through-plane direction [14]. We can thus define the PSF of the MRI acquisition system as

$$h(x, y, z) = \text{sinc}(a\sqrt{x^2 + y^2}) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) \quad (1.3)$$

with $a = \frac{1.0055}{\Delta x}$ and $\sigma = \frac{\Delta z}{2\sqrt{2\ln(2)}}$, corresponding to the above-mentioned FWHM². As can be seen from (1.3), we chose the $\text{sinc}(\cdot)$ to act on the orthogonal distance from the z -axis. An alternative option would have been to define this in-plane $\text{sinc}(\cdot)$ function as $\text{sinc}(x)\text{sinc}(y)$, acting on the x - and y -axes independently. The difference between the two can be easily seen in Fourier space, where the former describes a disc in the xy -plane and the latter corresponds to a square. Since we couldn't find a clear opinion on this choice in the literature, we chose the disc option because it is angularly isotropic, displaying no orientation preferences. This is unlike the square, which favors orientations around $\pm\frac{\pi}{4}$. The resulting radial fringe patterns also correspond well to the ones that can be observed in real images, validating this choice.

1.3.2 Isotropic resolution in fetal brain MRI: State of the art

Multiple iterative optimization algorithms have been developed to restore isotropic resolution in MRI [15, 16, 17, 18, 19, 20, 21, 22]. Such algorithms use multiple (approximately) orthogonal image stacks of the same subject to reconstruct the complete volume with isotropic resolution. They are usually comprised of a registration and motion estimation step, and a restoration step. In the registration step, all the volumes are aligned to a reference, using volume-to-volume rigid body registration and then each volume is corrected for motion using slice-to-volume registration. This step is very important since motion can heavily deteriorate the reconstruction results. Once the volumes are registered and corrected for motion, there are several possibilities to reconstruct the high-resolution volume, using either interpolation, Bayesian techniques or variational optimization algorithms. The common principle of all these algorithms is to either minimize the system energy or to maximize a probability density function through an iterative optimization algorithm. In this project, we will use the reconstructions generated by the MIALSRTK [22, 23, 24] pipeline, provided by our collaborator Dr. Hélène Lajous, as a reference for the performance of this type of approaches.

Besides iterative optimization methods, several self-supervised neural network-based approaches have been proposed to restore isotropic resolution, like SMORE [25], STRESS [26] and the work of Kebiri et al. [27]. In this section, we will briefly mention the main features of each of those approaches and how they relate to this project.

The approach of SMORE is very similar to what we're trying to achieve in this project. Zhao et al. developed a self-supervised convolutional neural network (CNN) to improve the through-plane resolution of MR images, as well as reduce aliasing. They did this by training the network on low-resolution (LR) images created by filtering in the in-plane direction, which is then applied to the through-plane direction. SMORE uses a ResNet-based network with a small patch size, since their goal is to focus on local structures. Their model was tested on a broad collection of MR data, which included real MR images as well as filtered and downsampled ones, which provided easier numerical evaluation. The main difference between SMORE and this project, besides the different network architecture, lies in the fact that SMORE was developed for adult brain MR images, which differ greatly from fetal brain MR images in terms of size and complexity of the structures, as well as motion artefacts and SNR.

²The maximum of the $\text{sinc}(\cdot)$ function is 1. Given $\text{sinc}(x_0) = \frac{\sin(\pi x_0)}{\pi x_0} = 0.5$, with $x_0 \approx 0.6033$, the initial FWHM thus corresponds to $\text{FWHM}' = 2x_0$. The multiplicative factor a is the ratio of the initial FWHM' and the desired $\text{FWHM} = 1.2\Delta x$, which results in $a = \frac{\text{FWHM}'}{\text{FWHM}} = \frac{2x_0}{1.2\Delta x} = \frac{1.0055}{\Delta x}$. Also remember that $\Delta x = \Delta y$.

The concept of STRESS was to train a super-resolution network by exploiting both spatial and temporal information of dynamic fetal MRI time series to improve the through-plane resolution. Xu et al. did this by simulating interleaved slice acquisitions along the high-resolution axis and using the temporal cross-information between the interleaved series. This pipeline was again evaluated on both real and simulated data. Here, the difference compared to this project is that STRESS works with dynamic fetal MRI data, which is used to track fetal motion and the dynamics of fetal function. This application differs from the static brain images that we worked with, not only by focusing on the movement, but also by considering the entire fetal body instead of only the brain. Additionally, STRESS also uses a different network architecture.

Kebiri et al. took a slightly different approach than the previously mentioned works. First, they work with diffusion-weighted fetal brain MR images, which have a vastly different sharpness and contrast than both T_1 - and T_2 -weighted MR images. Second, they use a different form of prediction to increase the through-plane resolution. Instead of upsampling the volume and using the network to correct the upsampled volume, they use two isolated in-plane slices as the input to their model, linearly combine their latent space encoding, and run the decoder step on the combined slice to get the predicted in-between slice. They evaluated their pipeline on fetal MRI data from the developing Human Connectome Project (dHCP) and showed that they can indeed restore the through-plane resolution. The technique of predicting the in-between slices is different from our approach, but still interesting to consider.

1.4 Aim of this Project

In this Master Project, the idea is to get inspired by the work of Weigert et al. [7] on sample specific neural networks for isotropic resolution restoration in 3D fluorescence microscopy, in order to adapt the same principle to fetal brain MRI data. More specifically, the goal is to create a pipeline for restoring the 3D isotropic resolution in fetal brain MR images from a single acquisition, using a sample-specific neural network, to facilitate manual and automatic analysis of the developing brain. This pipeline should be capable of improving the through-plane resolution of a given volumetric image by only training on that same volume. To this end, a simplified mathematical model of the acquisition system is used to simulate the through-plane acquisition in the in-plane slices, which is then used together with the original in-plane slices to teach the network how to restore the through-plane resolution.

Chapter 2

Restoring isotropic resolution in 3D imaging using a sample-specific CNN

Generally, neural networks are trained once on a large dataset, to then be applied on unseen data and generate fast predictions. The idea behind this approach is that training the network is time-consuming, and thus it should be done during development rather than during application. In order for the network to be general, so that it can correctly predict results on unseen data, the training data needs to contain sufficiently varying samples. Although neural networks have been successfully used in many domains, their performance completely depends on the quality and quantity of the training data. More generally, a neural network is only as good as the data it has been trained on. Sometimes this can lead to cases where a neural network introduces erroneous, but realistic features it has learned from the training data. This phenomenon is called AI hallucination [28], and has a detrimental effect on the network performance. Furthermore, the uncertainty of knowing whether a specific shape is real or artificial is dangerous when working with medical imaging data, and should thus be minimized as much as possible.

In the case of restoring isotropic resolution in 3D imaging, another approach for training a neural network presents itself. Since it is often impossible to acquire 3D images with an isotropic resolution, good SNR and no artefacts, especially in the case of fetal brain MRI, constructing a large ground-truth dataset is unfeasible. However, the particular property of these images offers another possibility in generating a training dataset; to use the higher-resolution in-plane slices of the volume as the ground-truth. If we can approximate the mathematical model of the acquisition system, we can simulate the image acquisition on these in-plane slices, and teach the network to learn the inverse operation. We can then apply the same network on the through-plane slices to restore isotropic resolution. This approach seems particularly interesting, considering that it allows to train a network on a single volumetric image, on which it can ultimately restore isotropic resolution. Previous projects [7, 25, 26, 27] have already used this idea and demonstrated that it can work for several 3D imaging applications.

Using a sample-specific neural network for 3D imaging tasks inherently comes with some advantages. First, assuming that the shapes and structures in the image are mostly orientation-invariant, the in-plane slices will be similar to the through-plane slices, reducing the risk of AI hallucination. Second, the sample-specific network is trained anew for every image acquisition. This enables it to easily handle changes in acquisition parameters and other experimental conditions that lead to shifts in the image distribution.

2.1 Fluorescence microscopy

In order to develop a better understanding of the process described by Weigert et al., we attempted to reproduce their results, using the same synthetic dataset of cell membranes and nuclei, which was kindly provided by Dr. Martin Weigert. During this attempt, several questions concerning the exact methodology arose. Some of these were due to inconsistent reporting in the original article [7], for example how to generate the isotropic average of the PSF or if they upsample the volume before applying the acquisition model. Others were due to discrepancies between the explanations and the provided data: for example, the exact sequence of operations described in [7] and the addition of artificial noise during the training data generation do not seem to be consistent with the provided data. While these unanswered questions complicated the reproduction effort, they also encouraged us to delve deeper into the subject and propose our own solutions. In the end, we didn't manage to reproduce the original results completely. However, the process contributed to the improvement of some aspects for use in fetal brain MRI data.

Figure 2.1 shows our interpretation of the workflow in [7], containing the training data generation, neural network training and prediction. Since we were mainly working with synthetic 3D fluorescence microscopy data, the diagram includes the initial acquisition simulation from the synthetic ground-truth volume.

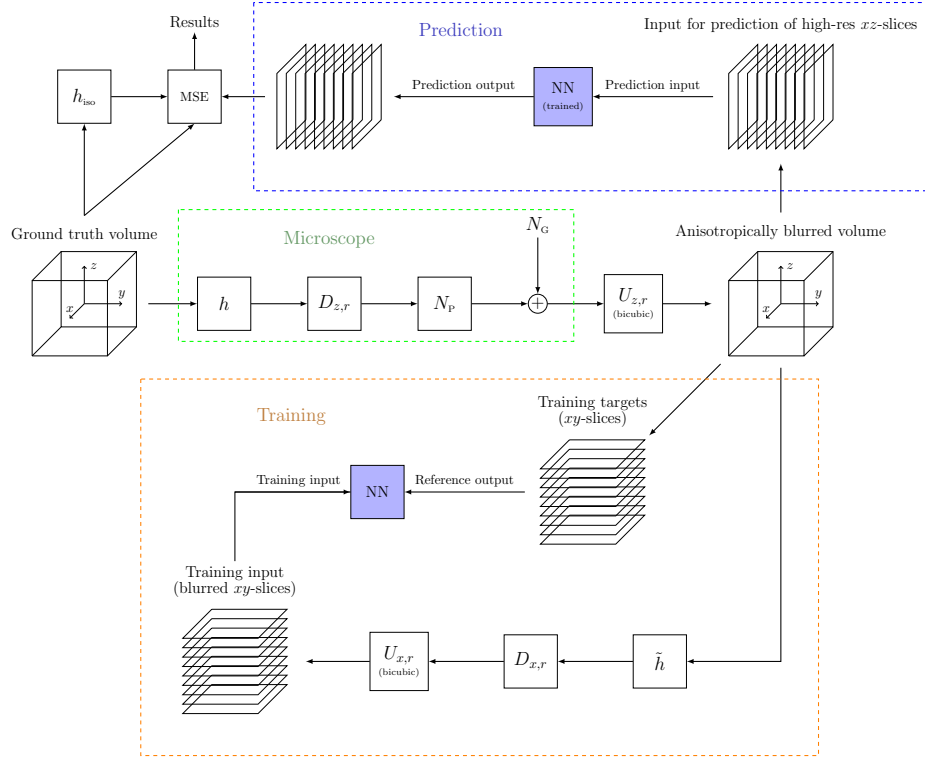


Figure 2.1: Flowchart depicting the synthetic image acquisition process (green), the training data generation (orange) and the prediction process (blue) implemented for isotropic resolution reconstruction in 3D fluorescence microscopy.

Starting from the synthetic ground-truth volume f , which has isotropic resolution and no blurring or noise, we apply the complete acquisition model

$$g = N_P[D_{z,r}(h * f)] + N_G \quad (2.1)$$

by first convolving it with the PSF h and then downsampling it in the z -direction with a downsampling

factor equal to the resolution ratio r using nearest-neighbor interpolation. In contrast to (1.1), we also add Poisson noise N_P and Gaussian noise N_G here. The resulting blurred and downsampled volume corresponds to what we would expect to obtain from a real 3D fluorescence microscopy acquisition. We then upsample this blurred and downsampled volume in the z -direction with bicubic interpolation, using an upsampling factor equal to the resolution ratio r , to obtain an anisotropically blurred volume that has the same size as the ground-truth volume. The xy -slices of this volume are going to be used as the targets the network is trained to reconstruct. To generate the training inputs, we apply the simplified acquisition model (1.1) to the blurred and upsampled volume. This results in a simulated bad resolution in the x -axis, meaning we can use the xy -slices of this volume as the training inputs.

Once the neural network has been trained, we can use the xz -slices from the anisotropically blurred volume as the input to the network and predict the target volume with isotropic resolution. Since we have the ground-truth volume in this case, we can compare the prediction directly to that. We also compare the prediction to the isotropically blurred volume, which is the ground-truth volume convolved with the isotropic PSF h_{iso} , corresponding to a Gaussian with the same standard deviation in all axes. This isotropically blurred volume is what we would expect to get from an ideal imaging process.

The neural network used in [7] is a classic UNet [29] with an additional top-level skip connection, as shown in Figure 2.2. This additional skip connection converts the UNet into a residual network, learning the difference between the input and output instead of the direct input-output operation. This simple modification has shown to considerably improve performance in image processing scenarios [30, 31].

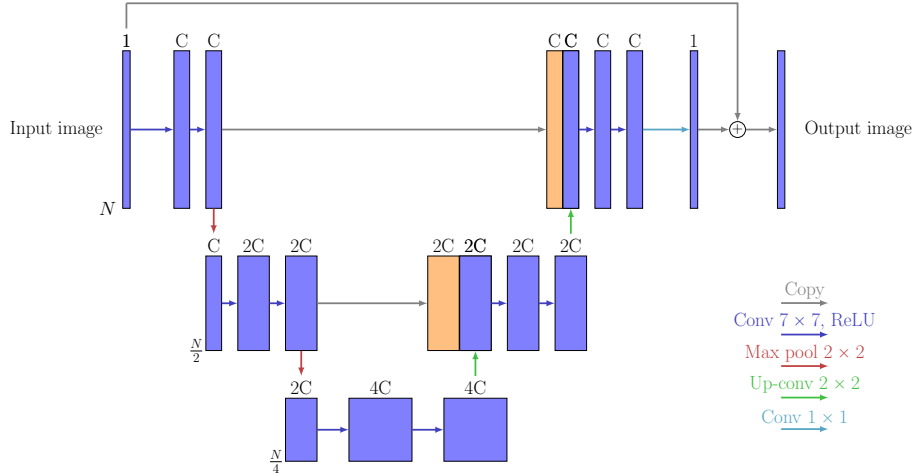


Figure 2.2: UNet architecture used in [7] with $C = 16$.

The UNet proposed in [7] has 2 encoder and 2 decoder steps, resulting in a rather shallow network, which is preferred for sample-specific applications, since they are naturally prone to overfit. The initial number of channels was set as $C = 16$ and the convolutional kernel was chosen to be of size $k = 7$ for all convolutional layers except the last one, which has $k = 1$. The PSNR loss

$$\text{PSNR}(g, \tilde{g}) = -10 \log_{10} \left(\frac{\max(g)^2}{\text{MSE}(g, \tilde{g})} \right) \quad (2.2)$$

with

$$\text{MSE}(g, \tilde{g}) = \frac{1}{WH} \sum_i^W \sum_j^H (g[i, j] - \tilde{g}[i, j])^2 \quad (2.3)$$

was minimized using the Adam optimizer [32] and the UNet was trained for 100 epochs with a learning rate of 0.005.

To reproduce their work, we implemented the pipeline as shown in Figure 2.1 in Python using NumPy [33] and the UNet presented in Figure 2.2 using PyTorch [34]. The training was conducted on an NVIDIA GeForce RTX 3090 with 24 GB of RAM. Since the original paper did not mention specific numbers for the noise and patch size, we experimented with both of them, but we did not manage to find any choice that could reproduce their results. We also tried to train the network with and without normalizing the input data p_{up} , but could still not produce matching results.

2.2 Fetal brain MRI

The main part of this project was to develop a pipeline for restoring isotropic resolution in fetal brain MR images using only one single volume, which we will denote as single-acquisition isotropic resolution (SAIR). To this end, we simulate the through-plane MRI acquisition on the in-plane slices using the simplified MRI acquisition model (1.2). With these slices, we can then train a sample-specific CNN to restore the through-plane resolution.

2.2.1 Generating the training data

The fetal brain MRI training data generation, depicted in Figure 2.3, is a direct adaptation of the previously implemented fluorescence microscopy pipeline (Figure 2.1). We start with the MRI acquisition volume, which intrinsically has an anisotropic resolution. In the first step, we upsample the original volume o in the z -direction by the resolution ratio r , using bicubic interpolation, to get a volume

$$o_{up} = U_{z,r}o$$

with isotropic voxel sizes, which is still anisotropically blurred. As previously mentioned, this is done to increase the amount of data we can use for training, and also to simplify the upcoming acquisition simulation. As the next step, we incrementally rotate the volume o_{up} by an angle $\theta = \frac{\pi}{n+1}$ around the z -axis, where n is the number of total rotated volumes, and concatenate all the resulting volumes together along the z -axis, resulting in the augmented volume

$$g = (o_{up}, R_{z,\theta}o_{up}, R_{z,2\theta}o_{up}, \dots, R_{z,n\theta}o_{up})^T.$$

The xy -slices of g represent the training targets for our network. The training inputs are created by applying the simplified MRI acquisition model (1.2) to each of the rotated volumes and concatenating them, resulting in the augmented, double-anisotropically blurred volume

$$p = (D_{x,r}(\tilde{h} * o_{up}) + N_G, D_{x,r}(\tilde{h} * R_{z,\theta}o_{up}) + N_G, \dots, D_{x,r}(\tilde{h} * R_{z,n\theta}o_{up}) + N_G)^T.$$

The xy -slices of p represent the training inputs for our neural network.

Notably, the main difference compared to the data generation in [7] is the use of multiple rotated volumes. By incrementally rotating o_{up} around the z -axis, we can augment the training data by a factor n , which we set to $n = 10$ for all results in this report. This only works because we assume that the original PSF h of the imaging system is rotationally invariant with respect to the z -axis. Other considerations for this operation are the boundary conditions and interpolation used for the rotation. Concerning the boundary conditions we use zero padding, since we usually work with pre-extracted brain volumes, where only voxels in a certain radius around the center are non-zero. Furthermore, we use bicubic interpolation for all the rotations to lose as little information as possible. In addition to providing more training data, this approach also improves the rotational invariance of the network, because the network learns how to reconstruct all the shapes and structures at multiple orientations. Finally, providing more diverse and a larger quantity of training data reduces the risk of overfitting, which is one of the main problems of sample-specific networks.

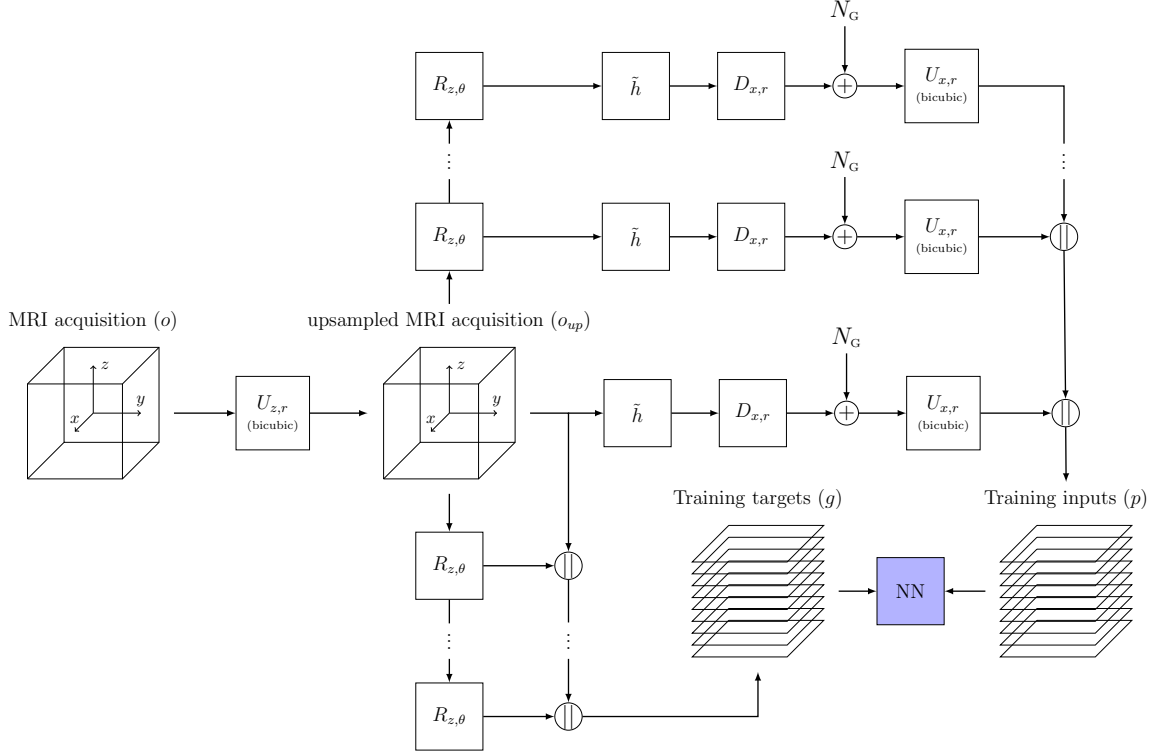


Figure 2.3: Flowchart depicting the training data generation from fetal MR images. $U_{z,r}$ and $U_{x,r}$ represent the bicubic upsampling operators by a factor r in the z - and x -direction respectively. $R_{z,\theta}$ is the rotation operator, rotating a volume by an angle θ around the z -axis. The concatenation operator \parallel concatenates two volumes along the x -axis. \tilde{h} is the rotated PSF, aligned along the z -axis as shown in Figure 1.3. $D_{x,r}$ is the downsampling operator by a factor r in the x -direction, using nearest-neighbor interpolation. As before, N_G represents additive white Gaussian noise.

2.2.2 Network architecture and training

Similar to the fluorescence microscopy case, we also used the improved UNet [29] architecture (Figure 2.2) for the application in fetal brain MRI. However, after implementing a modular version of this network and performing some grid search tests on the network depth and initial channels (see supplementary Figure 5.6), we noticed that a network with only one single encoder / decoder step and 32 initial channels already produces sufficient results, while also being more resistant to overfitting and much faster to train. As we know, the network will need to be retrained for each new acquisition, meaning that time is of the essence. For that reason, we chose to use this very shallow network, depicted in Figure 2.4, that can be trained in little time and still provides good results.

We also changed the loss function from the PSNR (2.2) to the MSE (2.3), which is more sensitive to small errors and differentiable. This change in loss function improved the training process and resulted in faster convergence. Unlike in fluorescence microscopy, we did not separate the input images into smaller patches, because the fetal brain MR images are already very small. The network was trained for 100 epochs using the ADAM optimizer [32] with a learning rate of 0.0005 and a batch size of 20. The training was done on an NVIDIA GeForce RTX 3090 with 24 GB of RAM, on which training the network for a single volume took less than five minutes when using $n = 10$ rotated volumes.

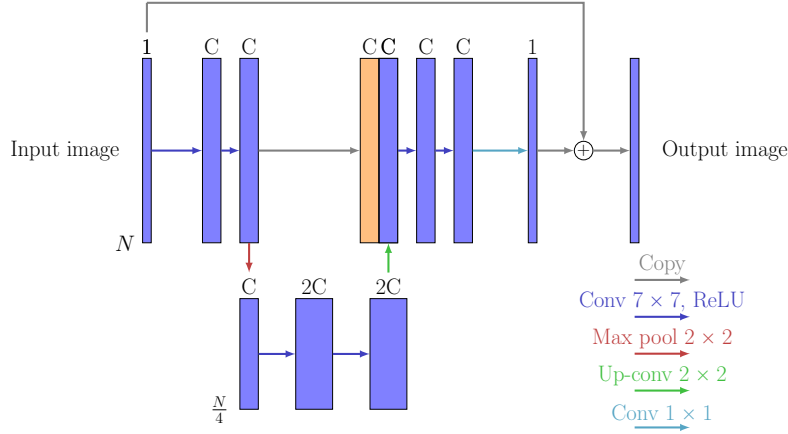


Figure 2.4: UNet architecture used for fetal brain MRI with $C = 32$.

2.2.3 Prediction

Once the network has been trained, we can use it to restore isotropic resolution in the original volume. Figure 2.5 shows the workflow diagram. The first step, upsampling the volume, is the same as in the training data generation, meaning we can use the upsampled volume that was previously generated. This step is thus only included for completeness. Next, since we assume that the original system PSF h is rotationally invariant with respect to the z -axis, we can again take advantage of this fact by applying the trained network to the xz -slices of multiple rotated versions of the upsampled volume, similar to what was done during training. The results are multiple volumes v_k with slightly differently predicted isotropic resolution, that are rotated by some angle $k\theta = k\frac{\pi}{M+1}$ around the z -axis, with $k \in \{0, 1, \dots, M\}$. We used $M = 15$ for all results in this report. Finally, we can then combine all these different predictions using Fourier burst accumulation (FBA) [35], which was originally proposed to create sharp images from multiple differently blurred images of the same scene. We adapted this idea to three dimensions, where the different volumes are first rotated back by the corresponding angle $-k\theta$, before performing the weighted averaging in the Fourier domain, given by

$$u_p(\mathbf{x}) = \mathcal{F}^{-1} \left(\sum_{i=1}^M w_i(\omega) \cdot \hat{v}_i(\omega) \right) (\mathbf{x}), \quad (2.4)$$

where \hat{v}_i is the Fourier Transform of volume v_i and the weights are given by

$$w_i(\omega) = \frac{|\hat{v}_i(\omega)|^p}{\sum_{j=1}^M |\hat{v}_j(\omega)|^p}, \quad (2.5)$$

where $w_i(\omega)$ controls the contribution of the frequency ω of volume v_i to the final reconstruction u_p . From this definition, we can see that the contribution of each frequency for a given volume is given by its Fourier magnitude. Since a higher Fourier magnitude means that this frequency is well-represented in that image, this means that FBA combines images based on their directional sharpness. A higher value of p results in a larger difference between the contributions. For $p = 0$, each image contributes the same amount, corresponding to a simple arithmetic mean, and as $p \rightarrow \infty$, for each given frequency ω only the volume with the largest Fourier magnitude at that frequency will contribute to the final result. After visual tests, we chose to use $p = 2$, as this value favors the prominent frequencies, but also includes enough other contributions.

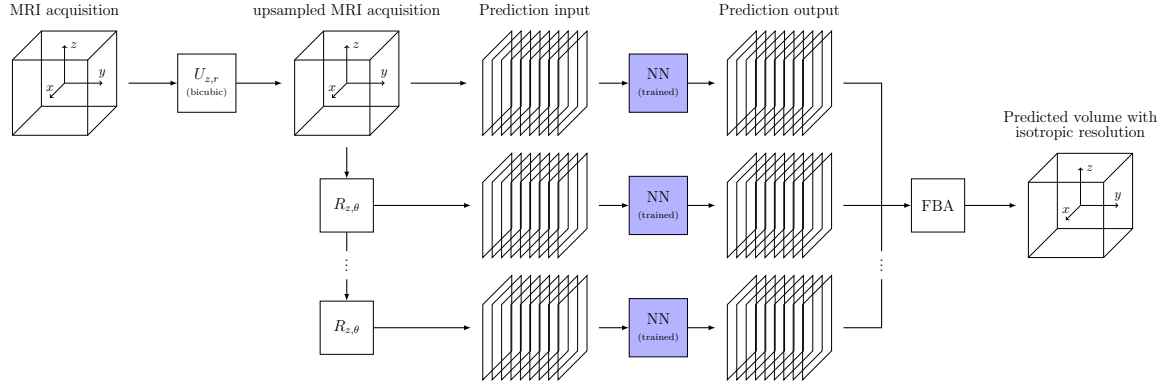


Figure 2.5: Flowchart depicting the prediction process for fetal MR images.

2.2.4 Using multiple volumes

Although the previously presented SAIR pipeline works well when using only one single acquisition (volume), it is straightforward to extend the idea to multiple series of the same subject. Most current MRI acquisition sequences will acquire at least three orthogonal series of the subject, so we can make use of them without adding any additional requirements to the scans. We can make use of multiple volumes either only for training, only for prediction, or for both.

Using multiple volumes for training simply consists of generating the training inputs and targets for each volume according to the diagram in Figure 2.3, combining the datasets by shuffling, and then training the network on the combined dataset. We will denote this version of the pipeline as multi-training SAIR (MTSAIR). Training the network with multiple volumes is not only beneficial because it provides more training data, but the data is much more diverse, coming from different acquisitions. For each low-resolution axis in one volume, we have the same axis with a high resolution in another volume, providing more information on the shape of the structures.

To use multiple volumes only during prediction, we simply perform the training data generation (Figure 2.3) and prediction (Figure 2.5) on each of the volumes separately and then combine the resulting predicted volumes, as shown in Figure 2.6. In the combination process, we first align all the volumes to each other, using one of them as the reference for volume-to-volume rigid-body registration, and then use FBA to generate the final weighted average output. We will denote this version of the pipeline as multi-prediction SAIR (MPSAIR). As the input volumes all have a different low-resolution axis, combining them will increase the amount of information available for the final FBA prediction. In other words, the resulting volume combines the best directional sharpness of all the volumes, providing a more detailed isotropic resolution prediction.

Finally, we can combine both training and prediction on multiple volumes by training a single network on all volumes, applying that network on all the volumes for prediction, and then combining those prediction outputs using FBA. We will denote this version of the pipeline as multi-volume SAIR (MVS AIR). Table 2.1 provides a summary of the pipelines and their properties.

Pipeline	Multi-volume training	Multi-volume prediction
SAIR		
MTSAIR	X	
MPSAIR		X
MVS AIR	X	X

Table 2.1: Summary of the pipeline properties.

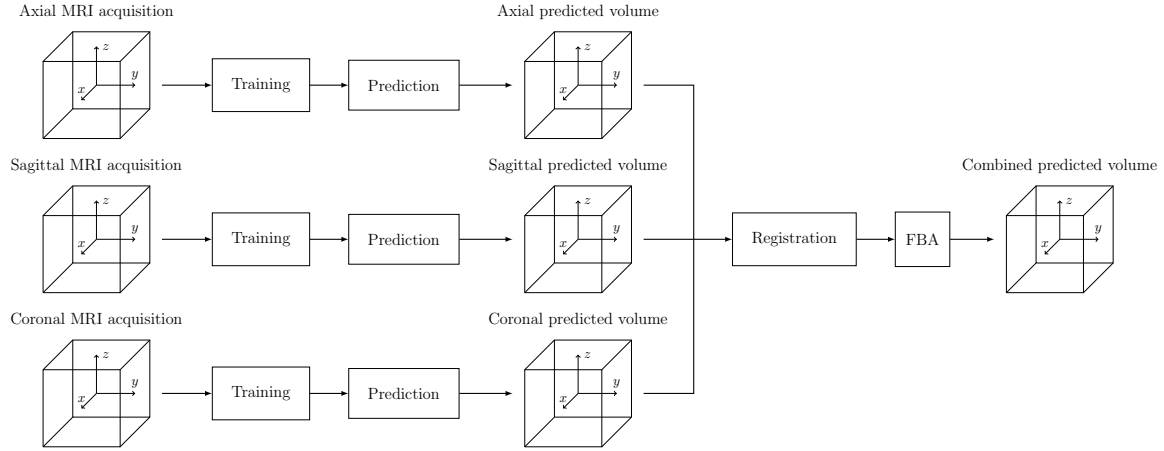


Figure 2.6: Flowchart showing the training and prediction process when using three orthogonal volumes and combining the results during prediction, as it is done in the MPSAIR pipeline.

2.3 Fetal brain MRI data

On one hand, evaluation and testing of the proposed workflow was done on simulated data provided by Dr. Hélène Lajous with FaBiAN [36, 37, 38], which offers a controlled environment and the ability to validate the results by comparing to a ground-truth (GT). On the other hand, we also used clinical data obtained within the FNS research project 182602 (Dr. Meritxell Bach Cuadra), which offers a proof of concept in real conditions and more practical value for the developments.

2.3.1 Simulated data

We used simulated data of four subjects of 26, 30 and 33 weeks of gestational age, containing 3 to 8 acquisitions each. Three of the subjects contained simulated motion artefacts and one didn't. All the simulated volumes had an in-plane resolution of $\Delta x = \Delta y = 1.1$ mm and a through-plane resolution of $\Delta z = 3.3$ mm, resulting in a resolution ratio of $r = 3$. Additionally, we obtained three different ground-truth volumes for each subject, namely one with both an in-plane B_1^+ bias field¹ and a through-plane slice profile², one with only the through-plane slice profile and the last one without any of the two. All the ground-truth volumes had an isotropic resolution of $\Delta x = \Delta y = \Delta z = 1.1$ mm.

2.3.2 Clinical acquisitions

The first real MRI data that we used to validate the initial developments were MRI acquisitions of the NIST/ISMRM system phantom that used the same acquisition parameters as real fetal brain MRI acquisitions. This data was great for early evaluations, as it didn't contain any movement artefacts and only simple shapes, allowing us to prove that the concept works on simplified data. Once this proof of concept had been achieved, we moved on to more realistic data.

In terms of real fetal brain MRI data, we worked with five different subjects, for each of which we had between 3 and 10 acquisitions from different orientations. Four subjects were scanned at $B_0 = 1.5$ T and had gestational ages of 26, 30 and 33 weeks. One of those four was specifically chosen for its lack of motion artefacts, the others displayed various amounts of motion. The fifth subject was acquired using $B_0 = 3$ T and also contained motion artefacts. All acquisitions were made

¹The B_1^+ bias field stands for inhomogeneities in the RF excitation field, which can be observed in real MRI acquisitions.

²The through-plane slice profile is an artefact observed in real MRI acquisitions that stems from the process of acquiring a 3D image using 2D slices.

using the HASTE sequence [11, 12], resulting in an in-plane resolution of $\Delta x = \Delta y = 1.125$ mm and a through-plane resolution of $\Delta z = 3.3$ mm, corresponding to a resolution ratio of $r = 2.933$. For each acquisition, we also received the corresponding manually generated brain masks and the super-resolution reconstruction with an isotropic resolution of $\Delta x = \Delta y = \Delta z = 1.125$ mm, generated with the MIALSRTK pipeline [22, 23], using all the available acquisitions, which could be used as a reference volume for visual comparison. Figure 2.7 shows an example of the real fetal brain MRI data, specifically it shows three orthogonal acquisitions of a subject of 30 weeks gestational age and the corresponding MIALSRTK super-resolution reconstruction.

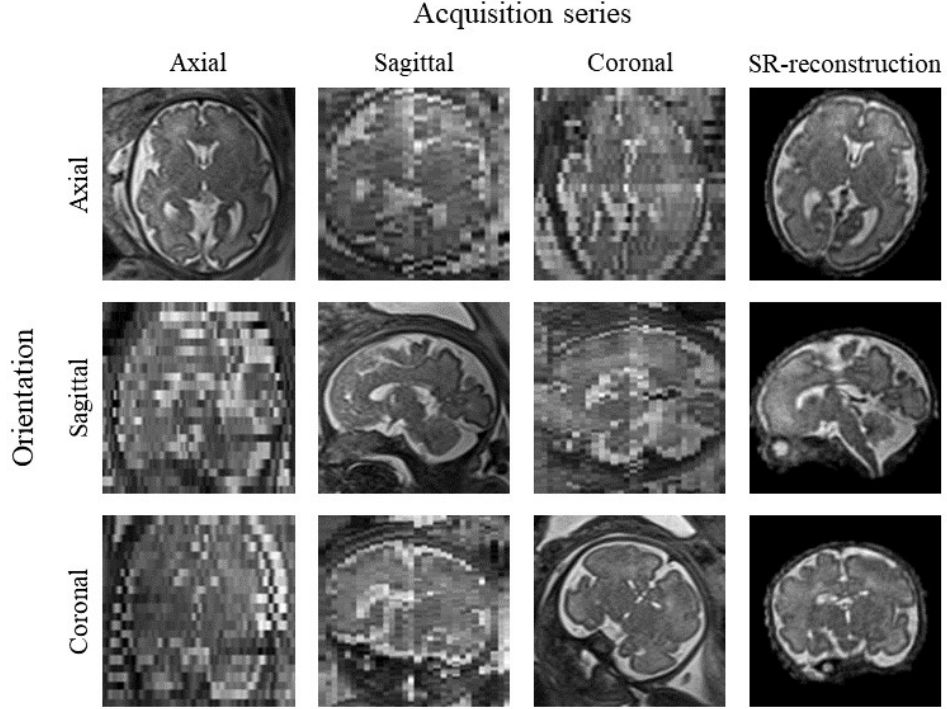


Figure 2.7: Example of the clinical acquisitions used in this project, showing three orthogonal acquisition series of the same subject (30 weeks gestational age), together with their SR-reconstruction made using the MIALSRTK pipeline.

Chapter 3

Results

3.1 Fluorescence microscopy

We tested the fluorescence microscopy pipeline on three volumes containing synthetic cellular data. The first contained only nuclei, the second only membranes, and the third contained both nuclei and membranes. Figure 3.1 shows an example of the obtained results, when training the network with our own datasets, generated using the pipeline presented in Figure 2.1. The PSF used to create the anisotropically blurred input volume was provided by Dr. Martin Weigert, but would be deduced from the microscope parameters in the case of real data.

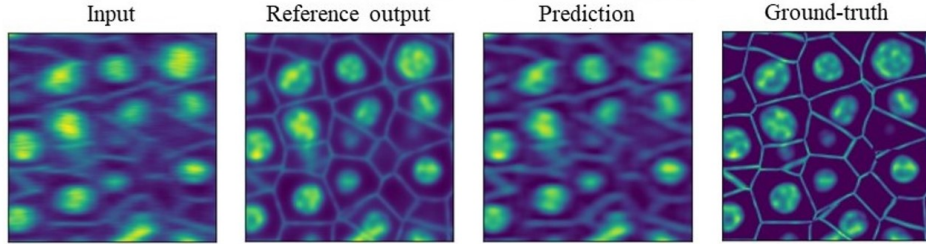


Figure 3.1: Sample xz -slice of the results obtained for simulated 3D fluorescence microscopy data. The input volume has been generated with the pipeline shown in Figure 2.1, the reference output corresponds to the isotropically blurred volume and the prediction is the result of applying the input volume to the trained network. The ground-truth is the unblurred numeric simulation of cell membranes and nuclei. The images shown here are from a zoomed region of the volume containing both nuclei and membranes.

From the images in Figure 3.1, we can see that the network slightly improves the resolution anisotropy of the input volume, but does not manage to reconstruct missing membranes and the prediction is still far from the reference output or the ground-truth. This fact becomes even more apparent when looking at the resulting PSNR values in Table 3.1. Besides the "Nuclei + Memb." volume, our values are much worse than those obtained in [7] on the same data. It should be noted that in some cases, specifically for "Membranes (ISO)" and "Nuclei + Memb. (ISO)", the PSNR values of the output volume are worse than those of the input volume, meaning that the network actually reduces the PSNR of the volume in some cases.

In order to see if the network would actually be capable of restoring isotropic resolution on the provided data, given the perfect circumstances, we trained it on the already blurred data provided by Dr. Martin Weigert, which was described as the input data used to produce the numbers shown in the respective columns of Table 3.1 and 3.2. As mentioned previously, this volume does not completely

Reference volume	Input	Output	Input ([7])	Output ([7])
Nuclei (ISO)	23.88	23.96	25.28	33.99
Nuclei (GT)	22.41	22.62	23.19	30.16
Membranes (ISO)	21.75	14.61	22.13	19.45
Membranes (GT)	16.11	16.65	15.98	30.26
Nuclei + Memb. (ISO)	26.43	24.91	27.91	25.15
Nuclei + Memb. (GT)	24.26	24.99	24.25	27.71

Table 3.1: PSNR values of our results and those obtained by Weigert et al. on simulated fluorescence microscopy data. The results in the two leftmost columns have been obtained by training and testing on our own datasets, generated with the pipeline shown in Figure 2.1. The reported PSNR values have been calculated for the entire volume, using the isotropically blurred volume (ISO) and the ground-truth volume (GT) as a reference.

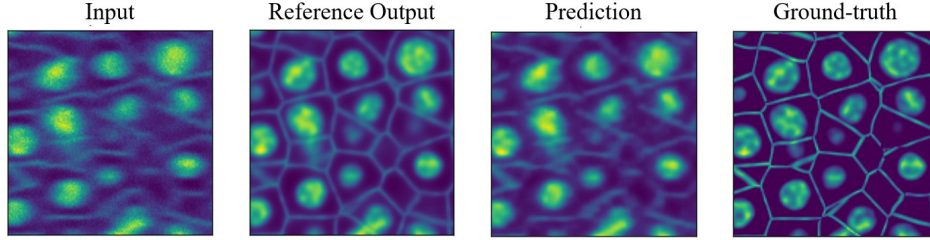


Figure 3.2: Sample xx -slice of the results obtained for simulated 3D fluorescence microscopy data. The input and reference images correspond to the data obtained by Dr. Martin Weigert and the prediction was made on the same dataset used for training. The images here are from a zoomed region of the volume containing both nuclei and membranes.

correspond to what we would expect as the input to the network. Looking at the leftmost image in Figure 3.2, we can see unblurred noise, which according to the diagram in Figure 2.1 should not be possible, as the noise is added before the upsampling step. Nevertheless, looking at the input PSNR values in Table 3.2, we can see that the PSNR of this volume almost perfectly corresponds to the one reported by [7], suggesting that this is in fact the volume that was used.

Figure 3.2 shows the results of training and prediction directly on the data provided by Dr. Martin Weigert. Comparing these results with the ones from Figure 3.1, we can see that in this case, the network is able to better restore isotropic resolution than in the previous one. However, the network is still not able to restore missing membranes, and the image is still heavily blurred. Looking at the numbers in Table 3.2, corresponding to the PSNR values when using the already blurred volumes, we can see a general improvement in the output PSNR, except for "Membranes (GT)". Although the PSNR is much higher, the numbers do not correspond at all to the ones obtained by [7]. Interestingly, we always obtain better results when comparing the output to the isotropically blurred volume (ISO) than when comparing them to the Ground-truth (GT), while Weigert et al. only reported better results for ISO with the "Nuclei" volume.

3.2 Fetal brain MRI

3.2.1 Simulated data

In terms of simulated data, we used numerical acquisition series of four simulated subjects of gestational ages 26, 30 and 33 weeks as described in Section 2.3.1. Here, we will present two examples of the isotropic resolution prediction results. Each figure shows the middle slices in the three orthogo-

Reference volume	Input	Output	Input ([7])	Output ([7])
Nuclei (ISO)	25.28	31.65	25.28	33.99
Nuclei (GT)	23.19	26.41	23.19	30.16
Membranes (ISO)	22.13	27.42	22.13	19.45
Membranes (GT)	15.98	16.54	15.98	30.26
Nuclei + Memb. (ISO)	28.30	31.88	27.91	25.15
Nuclei + Memb. (GT)	24.65	25.33	24.25	27.71

Table 3.2: PSNR values of our results and those obtained by Weigert et al. on simulated fluorescence microscopy data. The results in the two leftmost columns have been obtained by training and testing directly on the already blurred data provided by Dr. Martin Weigert. The reported PSNR values have been calculated for the entire volume, using the isotropically blurred volume (ISO) and the ground-truth volume (GT) as a reference.

nal directions 'Axial', 'Sagittal' and 'Coronal'. The left column shows the original MRI acquisition simulation with anisotropic resolution, which we use as the input data, the middle column shows the output of our SAIR pipeline and the right column shows the ground-truth volume, which was used as a reference. Both the original and predicted volumes have been aligned to the ground-truth using volume-to-volume rigid-body registration. The name of the acquisition refers to the in-plane orientation, which can be clearly seen in the 'Input data' column, where the slice of the in-plane orientation has 2D-isotropic resolution, while the other two have a lower resolution in one axis. Table 3.3 shows the numerical results of applying SAIR to a selected acquisition on three subjects. The subject of 30 weeks gestational age corresponds to Figure 3.3 and the subject of 26 weeks gestational age corresponds to Figure 3.4. From the table, we can see that only the subject of 30 weeks gestational age sees any improvement in MSE after applying the SAIR pipeline. However, as mentioned in the fluorescence microscopy case, we can clearly see from the images in Figures 3.3 and 3.4 that these numbers are not representative of the visual resolution improvement and should not be used as a definitive metric for the pipeline's performance.

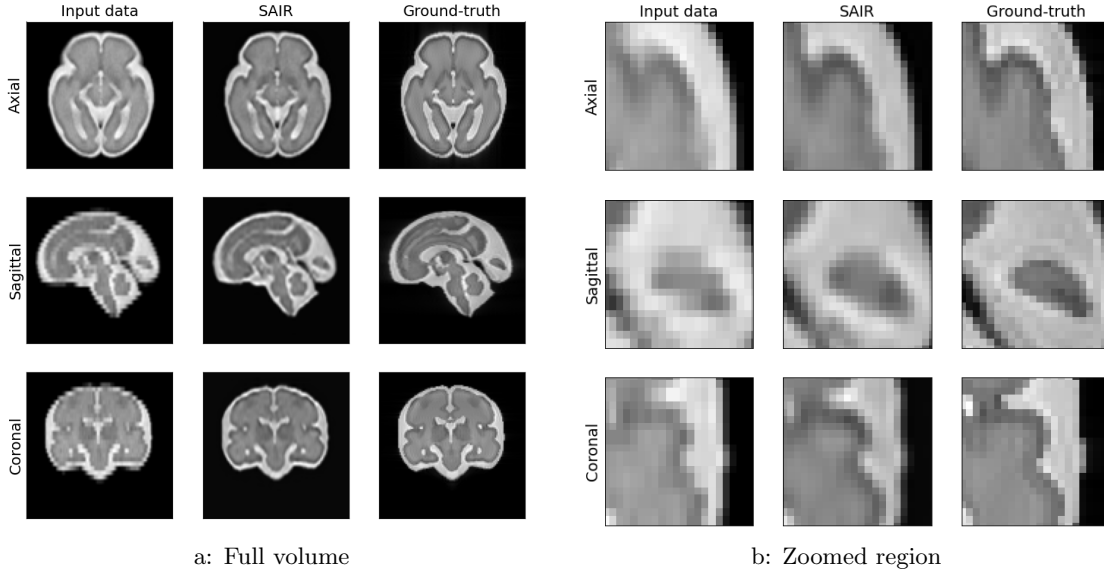


Figure 3.3: SAIR output on an axial acquisition series of a simulated subject of 30 weeks gestational age. The simulation of this acquisition was made without fetal motion.

The isotropic resolution reconstruction on an axial acquisition series of a simulated subject of 30

Age (weeks)	Input data	SAIR
26	-23.84	-20.88
30	-25.74	-26.81
33	-22.54	-21.83

Table 3.3: MSE values in dB compared to the ground-truth volume of three subjects with different gestational ages.

weeks gestational age, shown in Figure 3.3, is a nice example of how well our pipeline can perform in the right circumstances. As can be seen from the images, our SAIR pipeline is able to reconstruct isotropic resolution with very high accuracy. Looking only at the middle column, it is almost impossible to determine which corresponds to the in-plane slice. Looking at the sagittal and coronal slices, we can clearly see the resolution enhancement from the input to the output. The brain structures are better defined and more distinguishable after applying SAIR, and the overall shape of the brain is much closer to the ground-truth. Even in the in-plane slices, we can see some improvements in terms of reduced blurring and better tissue contrast. This specific volume has been simulated without any fetal motion during the acquisition process. As a consequence, there are no motion artefacts, like slice misalignment or intensity inhomogeneities. In these conditions, our SAIR pipeline performs exceptionally well considering that we only use one single volume.

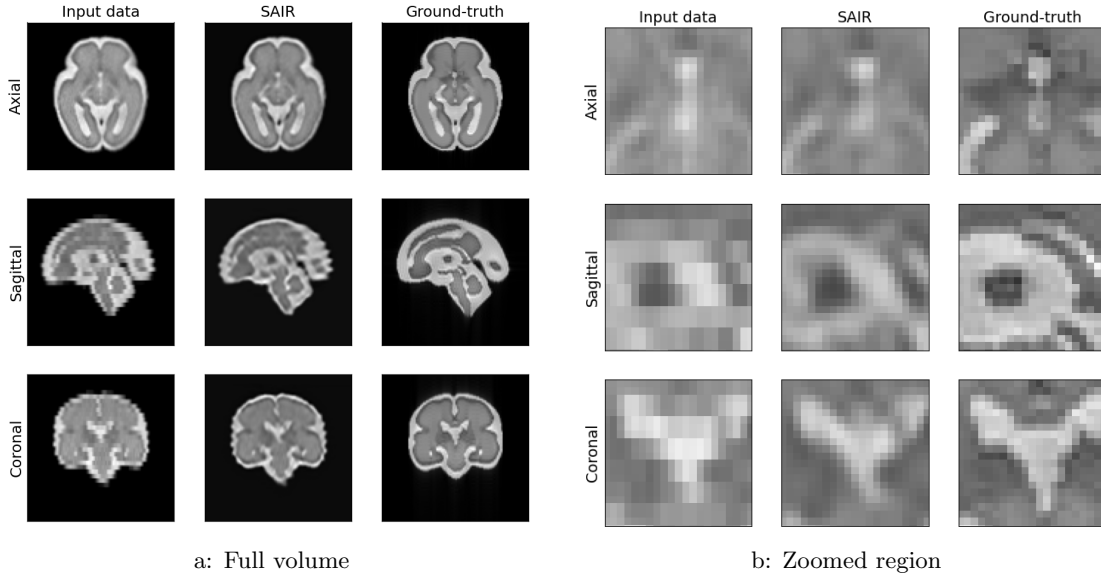


Figure 3.4: SAIR output on an axial acquisition series of a simulated subject of 26 weeks gestational age.

Figure 3.4 shows the SAIR results on an axial acquisition of a simulated subject of 26 weeks gestational age. This simulation included artificial movement of the fetus during acquisition, leading to motion artefacts, as is often the case in fetal MRI data. From this example, we can see that even in such non-ideal cases, our SAIR pipeline is able to restore isotropic resolution quite well. We can see a significant improvement in the sharpness and contrast of the output volume. We can also see that the motion artefacts have a visible influence on the reconstruction. Since the SAIR pipeline doesn't use any form of non-rigid registration to perform slice-to-volume alignment, it is not able to correct for slice misalignment, which is prominent in this acquisition.

3.2.2 Clinical Acquisitions

We used real fetal brain MR acquisitions of five subjects with different levels of motion and three gestational ages to verify the real-world applicability of the SAIR pipeline. We will show the results on two different subjects. The layout of the figures is the same as in Section 3.2.1, with the only exception being the ground-truth volume, which is replaced by the super-resolution reconstruction generated with the MIALSRTK pipeline. We don't expect our pipeline to perform as well as the MIALSRTK reconstructions, as they use many more volumes. Because we don't have a ground-truth volume for real data, however, we use them as a reference for visual analysis.

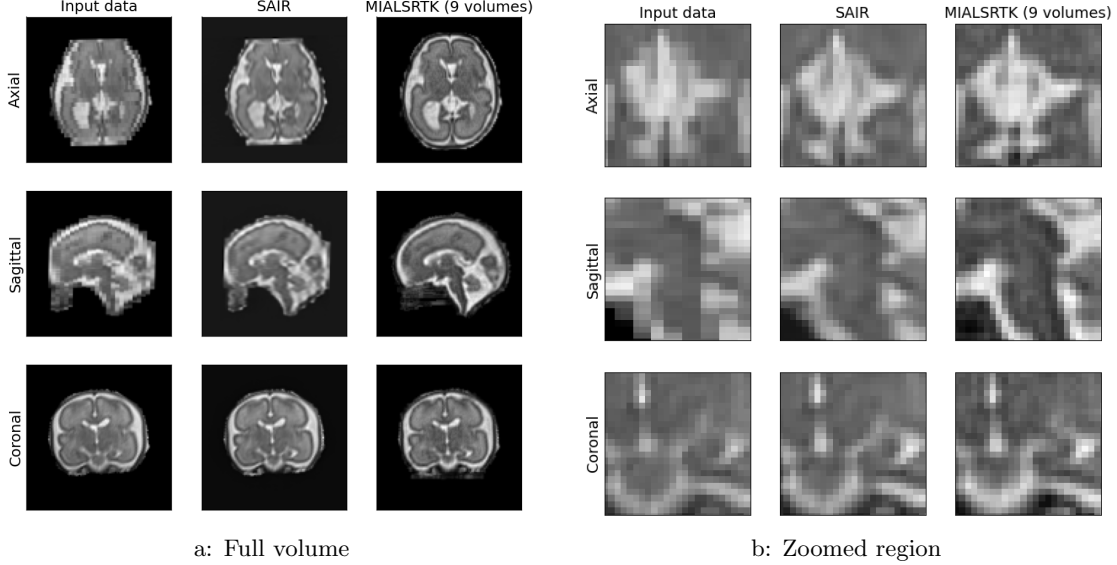


Figure 3.5: SAIR output on a coronal acquisition series of a subject of 30 weeks gestational age with little motion artefacts.

Figure 3.5 shows the SAIR results on a coronal acquisition series of a subject of 30 weeks gestational age, which was specifically chosen for its lack of motion artefacts. From the images, we can see that our pipeline is capable of restoring isotropic resolution on real fetal MRI data. As in the simulated subjects, the SAIR output shows much better sharpness in the through-plane direction and the resulting volume is visually very close to the MIALSRTK reconstruction, which used 9 volumes for this reconstruction.

In Figure 3.6 we can see the results of applying the SAIR pipeline on a coronal acquisition of a subject of 26 weeks gestational age. This acquisition contains some movement artefacts near the top of the head and some intensity inhomogeneities throughout the entire volume. Even in this case, SAIR is able to restore isotropic resolution in the whole volume. In the output, the edges of the different brain regions are much better defined, the contrast is improved, and even small details are visible. Again, the results are not very far from the MIALSRTK reconstruction, which used 4 different volumes to generate the shown results.

These two examples demonstrate the capability of the SAIR pipeline to perform well on real data, even in the presence of some motion artefacts, and using only one single volume to train the network.

3.2.3 Using multiple volumes

Using multiple acquisition series of the same subject increases the amount of information available for the reconstruction. To test if this improves the quality of the reconstruction over using only one volume, and to evaluate which of the presented multi-volume pipelines (MTSAIR, MPSAIR and MVSAIR) uses the additional information most efficiently, we applied all presented pipelines to the same acquisitions

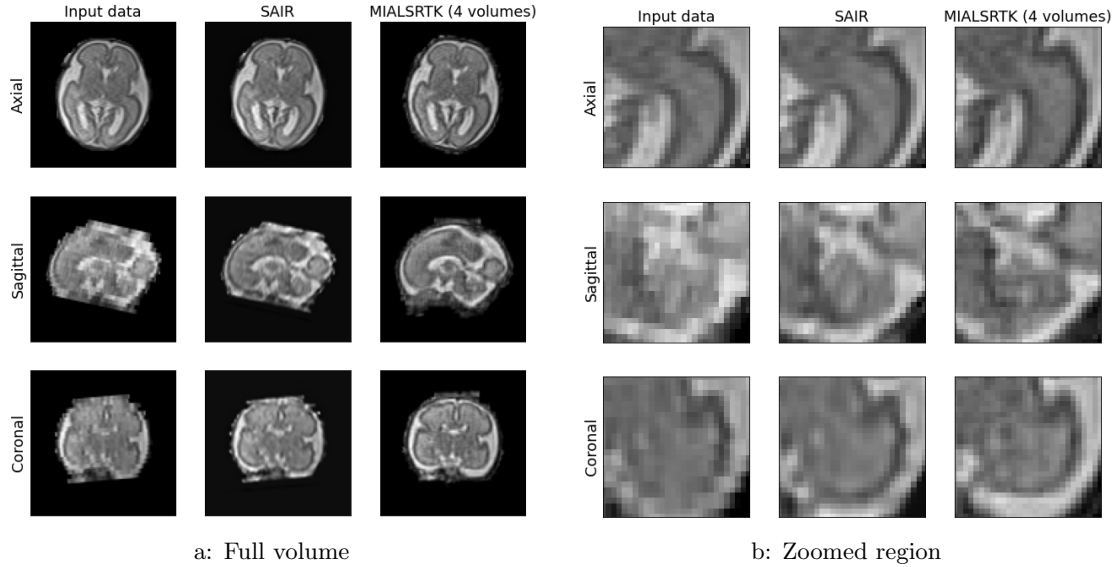


Figure 3.6: SAIR output on an axial acquisition series of a subject of 26 weeks gestational age.

of a simulated subject of 30 weeks gestational age without any motion artefacts. The resulting volumes can be seen in Figure 3.7, where the first column corresponds to the axial acquisition of this subject, the next four show the results of the different pipelines, and the last contains the simulated ground-truth volume. In addition, we provide the MSE values of all the results compared to the ground-truth volume in Table 3.4.

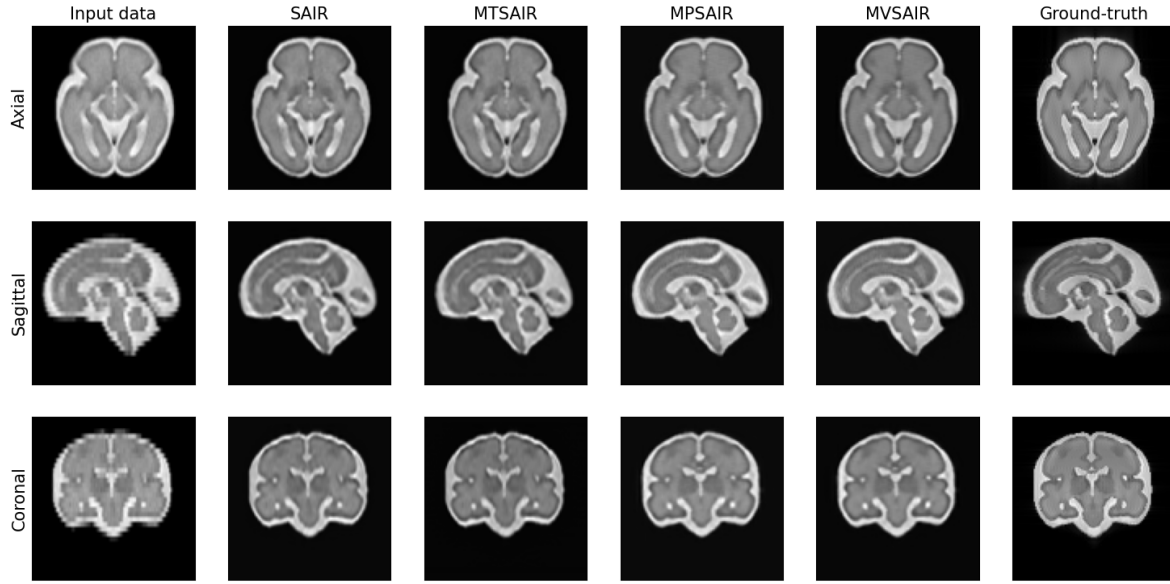
	Input data	SAIR	MTSAIR	MPSAIR	MVSAIR
MSE (dB)	-25.74	-26.81	-25.18	-26.02	-26.43

Table 3.4: MSE values corresponding to the images in Figure 3.7).

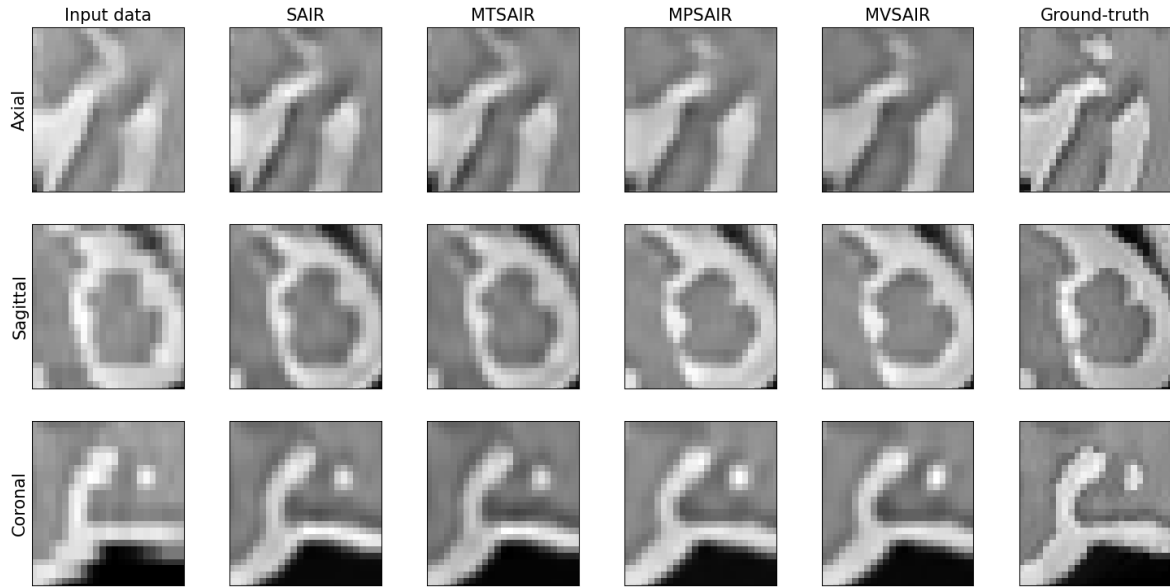
From the images in Figure 3.7, we can see that all pipelines perform very well at reconstructing isotropic resolution. However, looking at the zoomed regions in detail, MTSAIR seems to not contribute any significant improvement in visual quality or anatomical accuracy over the simple SAIR pipeline. Especially compared with both MPSAIR and MVSAIR, which are able to recover more details and seem sharper in general. Between MPSAIR and MVSAIR, there is no visually notable difference.

Looking at the numbers in Table 3.4, it would seem that the simple SAIR pipeline produces the best results, with MVSAIR being the best of the multi-volume pipelines. These numbers, again, do not agree with the visual analysis as we have seen before, so they should be taken with a grain of salt.

From the visual analysis, it seems like the important aspect of using multiple volumes is the combination of multiple reconstructed volumes during prediction, regardless if these reconstructions have been made by a network trained on multiple volumes or only on one. Training the network with multiple volumes does not seem to improve the final reconstruction. If it did, one would expect MTSAIR to produce better results than SAIR, and MVSAIR should be better than MPSAIR, which is both not the case.



a: Full volume



b: Zoomed region

Figure 3.7: Comparison of the different pipelines using multiple volumes of a simulated subject of 30 weeks gestational age. SAIR is using only one volume (axial) for training and prediction. MTSAIR, MPSAIR and MVSAIR are using three volumes (axial, sagittal and coronal). MTSAIR uses one single network that was trained on the three volumes, MPSAIR uses three networks, each trained on one of the three volumes and combines their results, and MVSAIR uses one network trained on the three volumes and combines their results. For simplicity, only the axial acquisition series is shown as the input data.

Chapter 4

Discussion

4.1 Fluorescence microscopy

Reproducing the work in [7] was meant as an easy introduction to the topic of anisotropic resolution reconstruction in 3D imaging. As mentioned in Section 2.1 however, this task turned out to be much more difficult than expected. One of the reasons for this were unclear descriptions, which forced us to experiment with different variations of the proposed pipeline. In particular, it was unclear if the authors added additional noise during the training data generation, after the upsampling operator $U_{x,r}$ (cf. Figure 2.1), which would seem to be the case in the data that we obtained from Dr. Martin Weigert. This however, would not match with the description in the paper, where they specifically omit the addition of noise (cf. (1.1)). Furthermore, the description often mentions operations on the slices of a volume, leading us to believe that they would perform these operations on the 2D slices instead of the 3D volume, which caused some confusion. It was also challenging to recreate the "isotropic average" of the PSF, which came without any explanation in the paper. In addition to these missing and misleading explanations, the task was further complicated by some misunderstandings on our part, like the misinterpretation of the exact neural network architecture, which we first implemented without the additional top-level skip-connection. This effort would have been simplified by the availability of the source code. Unfortunately, it was no longer available in its original form because it had been integrated into the *CSBDeep* package.

From the results presented in Section 3.1 we can assert that we didn't manage to reproduce the desired numbers. Nevertheless, there are some interesting observations to be discussed. Firstly, as we teach the network to reconstruct an isotropic blurring from an anisotropic blurring, the results should be closer to the isotropically blurred volume (ISO) than to the ground-truth (GT). From Tables 3.1 and 3.2, however, we can see that this is not the case for the results in [7], except for the "Nuclei" volume. In the "Membranes" volume, for example, the PSNR compared to the GT volume is 11 dB higher than compared to the ISO volume, even though the input PSNR is more than 6 dB higher for the ISO volume. Looking more closely at the values, we can see that the output PSNR of both the "Membranes" and "Nuclei + Memb." volumes are lower than the input PSNR when comparing it to the ISO volume. When looking at Figure 2 of [7], we can find that these numbers do not correctly represent the visual appearance of the volumes. They are in fact much more similar to the ISO than to the GT volume, and definitely improve in visual quality compared to the input volumes. This shows that it is very difficult to find a metric in image processing that accurately represents the visual characteristics of an image [39], which is the reason why we mainly focused on the visual analysis in this project.

This initial process was very important for the continuation of the project. Although we did not manage to reproduce the results in [7], we gained the necessary knowledge to tackle the novel part of this project.

4.2 Fetal brain MRI

With the knowledge gained from working on 3D fluorescence microscopy, we quickly managed to adapt the pipeline for use in fetal brain MRI, and we were able to produce promising results on synthetic data early on. The results became even better after implementing the use of multiple rotated volumes both during training and prediction, which was definitely one of the most important additions to the pipeline. As we have seen from the results, the SAIR pipeline delivers impressive results, both on simulated and real data, especially if the data contains no motion artefacts. The reduced performance in the presence of motion artefacts is no surprise, because the SAIR pipeline does not include any form of motion compensation. In contrast to most existing approaches at restoring isotropic resolution in MRI, we completely focused on the resolution restoration part of the problem. The reason being, on the one hand, that there already exists an abundance of motion compensation algorithms, and on the other hand, that it was simply outside the scope and time frame of this project. It is clear that without any kind of motion compensation, SAIR will never be able to perform as well as MIALSRTK, for example, but this was never the goal in the first place. Realistically, even with motion compensation SAIR is designed to use only a single volume for the reconstruction, compared to MIALSRTK, which ideally uses around 9 volumes. Because SAIR uses only a single volume, it is not possible to create a standalone iterative algorithm with motion compensation and registration. However, the idea of the SAIR pipeline is not to replace the other pipelines, but to enhance them. For example, SAIR could be used instead of a simple upsampling step, increasing the quality of the input volume to other pipelines and improving the final reconstruction result. It could even be used as the very first step in a pipeline, before motion compensation, to improve the motion compensation itself.

The multi-volume pipelines, especially MPSAIR and MVSAIR, proved to deliver even better results than SAIR by using multiple volumes of the same subject. MPSAIR and MVSAIR could be used in the same manner as SAIR, to enhance other existing pipelines. For example, the final FBA step of MPSAIR (cf. Figure 2.6) could simply be replaced by a more sophisticated volume-combining method. However, they also show promising potential for a novel iterative algorithm when paired with a motion compensation step. For example, such an algorithm could consist of simply applying motion compensation, then using MPSAIR (because it's faster to train than MVSAIR and delivers almost the same results) to reconstruct the volume with isotropic resolution, then we would use the reconstructed volume to improve the initial motion compensation and repeat the process until we converge to the best possible output. For this, we could use any number of volumes from the same subject, as does MIALSRTK, with more volumes potentially resulting in a better reconstruction.

In future work, besides motion compensation, another potential improvement could be to change the mathematical model of the MRI acquisition process (1.2) to a more realistic approximation, which completely or partially operates in the Fourier domain. This could provide a more accurate acquisition simulation on the in-plane slices, and improve network performance. Moreover, one could experiment with other network architectures, as we have seen that other related pipelines [25, 26] use networks focused on multiple residual components.

Bibliography

- [1] T. W. Redpath, “Signal-to-noise ratio in MRI,” *The British Journal of Radiology*, vol. 71, no. 847, pp. 704–707, 1998, pMID: 9771379. [Online]. Available: <https://doi.org/10.1259/bjr.71.847.9771379>
- [2] A. Nwaneshiudu, C. Kuschal, F. Sakamoto, R. Rox Anderson, K. Schwarzenberger, and R. Young, “Introduction to confocal microscopy,” *Journal of Investigative Dermatology*, vol. 132, no. 12, pp. 1–5, Dec. 2012.
- [3] P. A. Santi, “Light Sheet Fluorescence Microscopy: A Review,” *Journal of Histochemistry & Cytochemistry*, vol. 59, no. 2, pp. 129–138, 2011, pMID: 21339178. [Online]. Available: <https://doi.org/10.1369/0022155410394857>
- [4] M. Minsky, “Microscopy apparatus,” U.S. Patent US3013467A, Dec. 19, 1961. [Online]. Available: <https://patents.google.com/patent/US3013467A/en>
- [5] K. Greger, J. Swoger, and E. H. K. Stelzer, “Basic building units and properties of a fluorescence single plane illumination microscope,” *Review of Scientific Instruments*, vol. 78, no. 2, p. 023705, 2007. [Online]. Available: <https://doi.org/10.1063/1.2428277>
- [6] H. Jelínková and J. Šulc, “2 - Laser characteristics,” in *Lasers for Medical Applications*, ser. Woodhead Publishing Series in Electronic and Optical Materials, H. Jelínková, Ed. Woodhead Publishing, 2013, pp. 17–46. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780857092373500023>
- [7] M. Weigert, L. Royer, F. Jug, and G. Myers, “Isotropic Reconstruction of 3D Fluorescence Microscopy Images Using Convolutional Neural Networks,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Janin, D. L. Collins, and S. Duchesne, Eds. Cham: Springer International Publishing, 2017, pp. 126–134.
- [8] C. Weisstanner, G. Kasprian, G. M. Gruber, P. C. Brugger, and D. Prayer, “MRI of the Fetal Brain,” *Clinical Neuroradiology*, vol. 25, no. 2, pp. 189–196, Oct 2015. [Online]. Available: <https://doi.org/10.1007/s00062-015-0413-z>
- [9] Q. Gong, N. Roberts, A. Garden, and G. Whitehouse, “Fetal and Fetal Brain Volume Estimation in the Third Trimester of Human Pregnancy Using Gradient Echo MR Imaging,” *Magnetic Resonance Imaging*, vol. 16, no. 3, pp. 235–240, 1998. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0730725X97002816>
- [10] K. P. Cosgrove, C. M. Mazure, and J. K. Staley, “Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry,” *Biological Psychiatry*, vol. 62, no. 8, pp. 847–855, 2007, bipolar Disorder and OCD: Circuitry of Impulsive and Compulsive Behaviors. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0006322307001989>

- [11] M. R. Patel, R. A. Klufas, R. A. Alberico, and R. R. Edelman, "Half-fourier acquisition single-shot turbo spin-echo (HASTE) MR: comparison with fast spin-echo MR in diseases of the brain." *American Journal of Neuroradiology*, vol. 18, no. 9, pp. 1635–1640, 1997. [Online]. Available: <http://www.ajnr.org/content/18/9/1635>
- [12] R. C. Semelka, N. L. Kelekis, D. Thomasson, M. A. Brown, and G. A. Laub, "HASTE MR imaging: Description of technique and preliminary results in the abdomen," *Journal of Magnetic Resonance Imaging*, vol. 6, no. 4, pp. 698–699, 1996. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.1880060420>
- [13] O. A. Glenn, "MR imaging of the fetal brain," *Pediatric Radiology*, vol. 40, no. 1, p. 68, Nov 2009. [Online]. Available: <https://doi.org/10.1007/s00247-009-1459-3>
- [14] S. Jiang, H. Xue, A. Glover, M. Rutherford, D. Rueckert, and J. Hajnal, "MRI of Moving Subjects Using Multislice Snapshot Images With Volume Reconstruction (SVR): Application to Fetal, Neonatal, and Adult Brain Studies," *IEEE Transactions on Medical Imaging*, vol. 26, no. 7, pp. 967–980, 2007.
- [15] S. Jiang, H. Xue, A. Glover, M. Rutherford, D. Rueckert, and J. V. Hajnal, "MRI of Moving Subjects Using Multislice Snapshot Images With Volume Reconstruction (SVR): Application to Fetal, Neonatal, and Adult Brain Studies," *IEEE Transactions on Medical Imaging*, vol. 26, no. 7, pp. 967–980, 2007.
- [16] B. Kainz, M. Steinberger, W. Wein, M. Kuklisova-Murgasova, C. Malamateniou, K. Keraudren, T. Torsney-Weir, M. Rutherford, P. Aljabar, J. V. Hajnal, and D. Rueckert, "Fast Volume Reconstruction From Motion Corrupted Stacks of 2D Slices," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1901–1913, 2015.
- [17] A. Gholipour, J. A. Estroff, and S. K. Warfield, "Robust Super-Resolution Volume Reconstruction From Slice Acquisitions: Application to Fetal Brain MRI," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1739–1758, 2010.
- [18] K. Kim, P. A. Habas, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme, "Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation," *IEEE Trans Med Imaging*, vol. 29, no. 1, pp. 146–158, Jan 2010.
- [19] M. Kuklisova-Murgasova, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, and J. A. Schnabel, "Reconstruction of fetal brain MRI with intensity matching and complete outlier removal," *Med Image Anal*, vol. 16, no. 8, pp. 1550–1564, Dec 2012.
- [20] F. Rousseau, O. A. Glenn, B. Iordanova, C. Rodriguez-Carranza, D. B. Vigneron, J. A. Barkovich, and C. Studholme, "Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images," *Acad Radiol*, vol. 13, no. 9, pp. 1072–1081, Sep 2006.
- [21] M. Ebner, G. Wang, W. Li, M. Aertsen, P. A. Patel, R. Aghwane, A. Melbourne, T. Doel, S. Dymarkowski, P. De Coppi, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI," *NeuroImage*, vol. 206, p. 116324, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053811919309152>
- [22] S. Tourbier, X. Bresson, P. Hagmann, J.-P. Thiran, R. Meuli, and M. B. Cuadra, "An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization," *NeuroImage*, vol. 118, pp. 584–597, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053811915005157>

- [23] S. Tourbier, C. Velasco-Annis, V. Taimouri, P. Hagmann, R. Meuli, S. K. Warfield, M. Bach Cuadra, and A. Gholipour, “Automated template-based brain localization and extraction for fetal brain MRI reconstruction,” *NeuroImage*, vol. 155, pp. 460–472, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053811917302938>
- [24] S. Tourbier, P. De Dumast, H. Kebiri, P. Hagmann, and M. Bach Cuadra, “Medical-Image-Analysis- Laboratory/mialsuperresolutiontoolkit: MIAL Super- Resolution Toolkit v2.0.3,” Dec. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.5803816>
- [25] C. Zhao, B. E. Dewey, D. L. Pham, P. A. Calabresi, D. S. Reich, and J. L. Prince, “SMORE: A Self-Supervised Anti-Aliasing and Super-Resolution Algorithm for MRI Using Deep Learning,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 805–817, 2021.
- [26] J. Xu, E. Abaci Turk, P. E. Grant, P. Golland, and E. Adalsteinsson, “STRESS: Super-Resolution for Dynamic Fetal MRI Using Self-supervised Learning,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 197–206.
- [27] H. Kebiri, E. J. Canales Rodríguez, H. Lajous, P. de Dumast, G. Girard, Y. Alemán-Gómez, M. Koob, A. Jakab, and M. B. Cuadra, “Through-plane super-resolution with autoencoders in diffusion magnetic resonance imaging of the developing human brain,” *bioRxiv*, 2021. [Online]. Available: <https://www.biorxiv.org/content/early/2021/12/07/2021.12.06.471406>
- [28] N. M. Gottschling, V. Antun, B. Adcock, and A. C. Hansen, “The troublesome kernel: why deep learning for inverse problems is typically unstable,” 2020. [Online]. Available: <https://arxiv.org/abs/2001.01258>
- [29] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” 2015. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [30] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep Convolutional Neural Network for Inverse Problems in Imaging,” *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.
- [31] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-P. Thiran, “CNN-Based Image Reconstruction Method for Ultrafast Ultrasound Imaging,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 4, pp. 1154–1168, 2022.
- [32] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” 2014. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [33] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: <https://doi.org/10.1038/s41586-020-2649-2>
- [34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>

- [35] M. Delbracio and G. Sapiro, “Removing Camera Shake via Weighted Fourier Burst Accumulation,” *CoRR*, vol. abs/1505.02731, 2015. [Online]. Available: <http://arxiv.org/abs/1505.02731>
- [36] H. Lajous, C. W. Roy, T. Hilbert, P. de Dumast, S. Tourbier, Y. Alemán-Gómez, J. Yerly, T. Yu, H. Kebiri, K. Payette, J.-B. Ledoux, R. Meuli, P. Hagmann, A. Jakab, V. Dunet, M. Koob, T. Kober, M. Stuber, and M. Bach Cuadra, “A Fetal Brain magnetic resonance Acquisition Numerical phantom (FaBiAN),” *Scientific Reports*, vol. 12, no. 1, p. 8682, May 2022. [Online]. Available: <https://doi.org/10.1038/s41598-022-10335-4>
- [37] H. Lajous, C. W. Roy, J. Yerly, and M. Bach Cuadra, “Medical-Image-Analysis-Laboratory/FaBiAN: FaBiAN v1.1,” Oct. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.5599311>
- [38] H. Lajous, T. Hilbert, C. W. Roy, J.-B. Ledoux, V. Dunet, M. Koob, and M. Bach Cuadra, “Dataset A Fetal Brain magnetic resonance Acquisition Numerical phantom (FaBiAN),” May 2022, This work was supported by the Swiss National Science Foundation through grants 182602, 141283 and 173129. We acknowledge the expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole Polytechnique Fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG). [Online]. Available: <https://doi.org/10.5281/zenodo.6477946>
- [39] T. Yu, T. Hilbert, G. F. Piredda, A. Joseph, G. Bonanno, S. Zenkhri, P. Omoumi, M. B. Cuadra, E. J. Canales-Rodríguez, T. Kober, and J.-P. Thiran, “Validation and Generalizability of Self-Supervised Image Reconstruction Methods for Undersampled MRI,” 2022. [Online]. Available: <https://arxiv.org/abs/2201.12535>
- [40] “Quantitative MRI - NIST,” <https://www.nist.gov/programs-projects/quantitative-mri>, (Accessed on 07.07.2022).

Chapter 5

Appendix

5.1 Additional work

The very first simulated data that we used was a handcrafted numerical imitation of the NIST/ISMRM system phantom [40], presented in Figure 5.1. The NIST/ISMRM system phantom is a spherical case, containing several spheres of different materials, which each produce a unique intensity in the MRI scanner and is used to calibrate or test the scanner. This simulated data could provide the first quantitative results and helped to test and improve the data processing pipeline. We could also use it to perform some basic first analyses of the network parameters, but it was heavily limited in terms of diversity and realism. For that reason, we quickly moved on to more sophisticated simulations of the fetal brain provided by the FaBiAN toolkit.

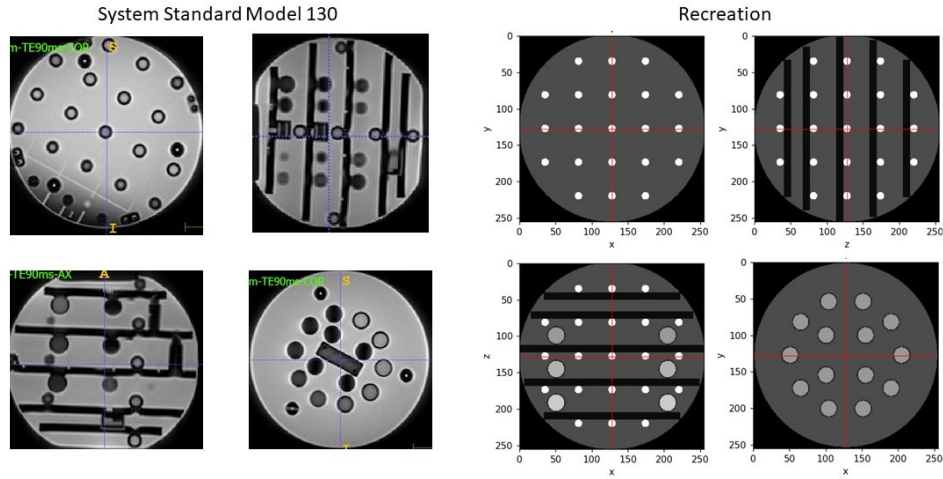


Figure 5.1: MRI images of the NIST/ISMRM system phantom and their numerical recreation.

5.2 Supplementary figures

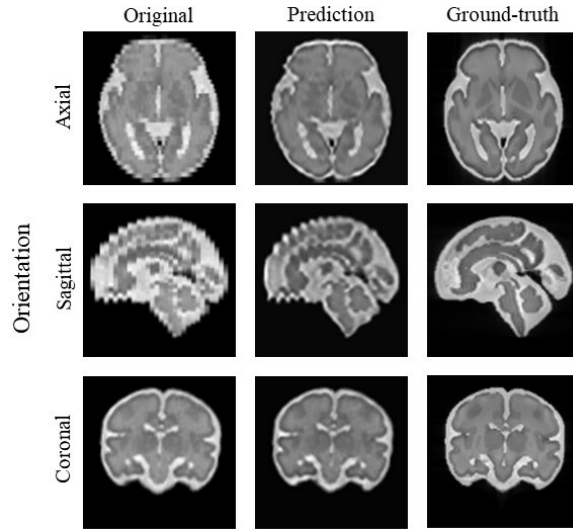


Figure 5.2: SAIR output on a coronal acquisition series of a simulated subject of 33 weeks gestational age.

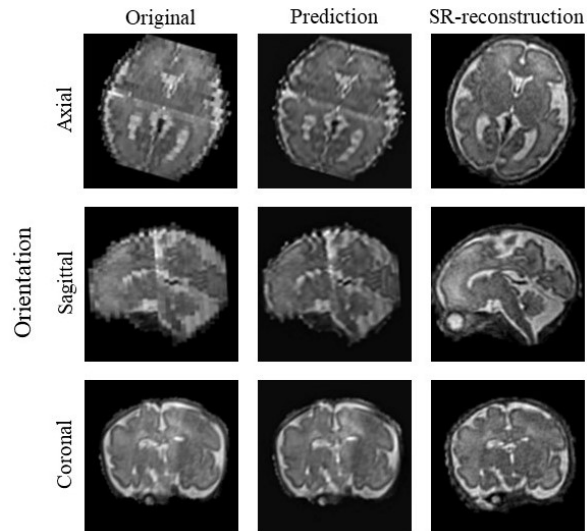


Figure 5.3: SAIR output on a coronal acquisition series of a subject of 30 weeks gestational age.

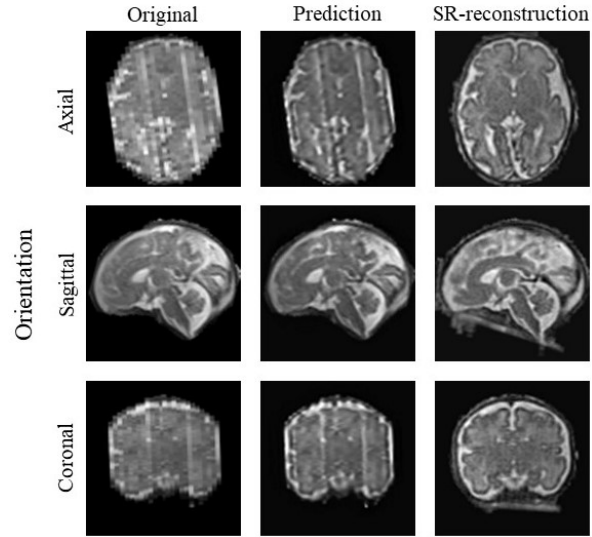


Figure 5.4: SAIR output on a sagittal acquisition series of a subject of 33 weeks gestational age.

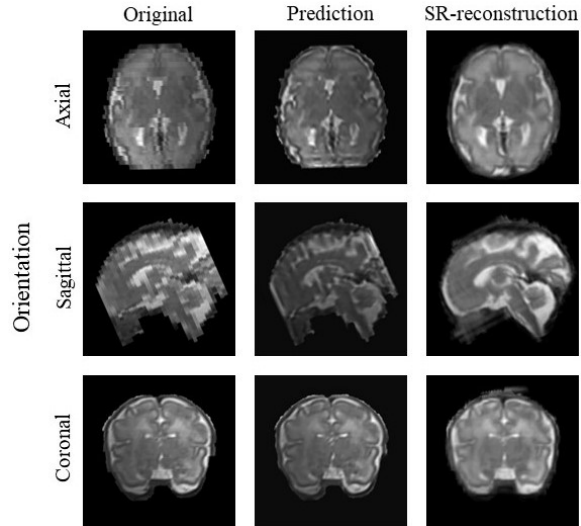


Figure 5.5: SAIR output on a coronal acquisition series of a subject of acquired at $B_0 = 3$ T.

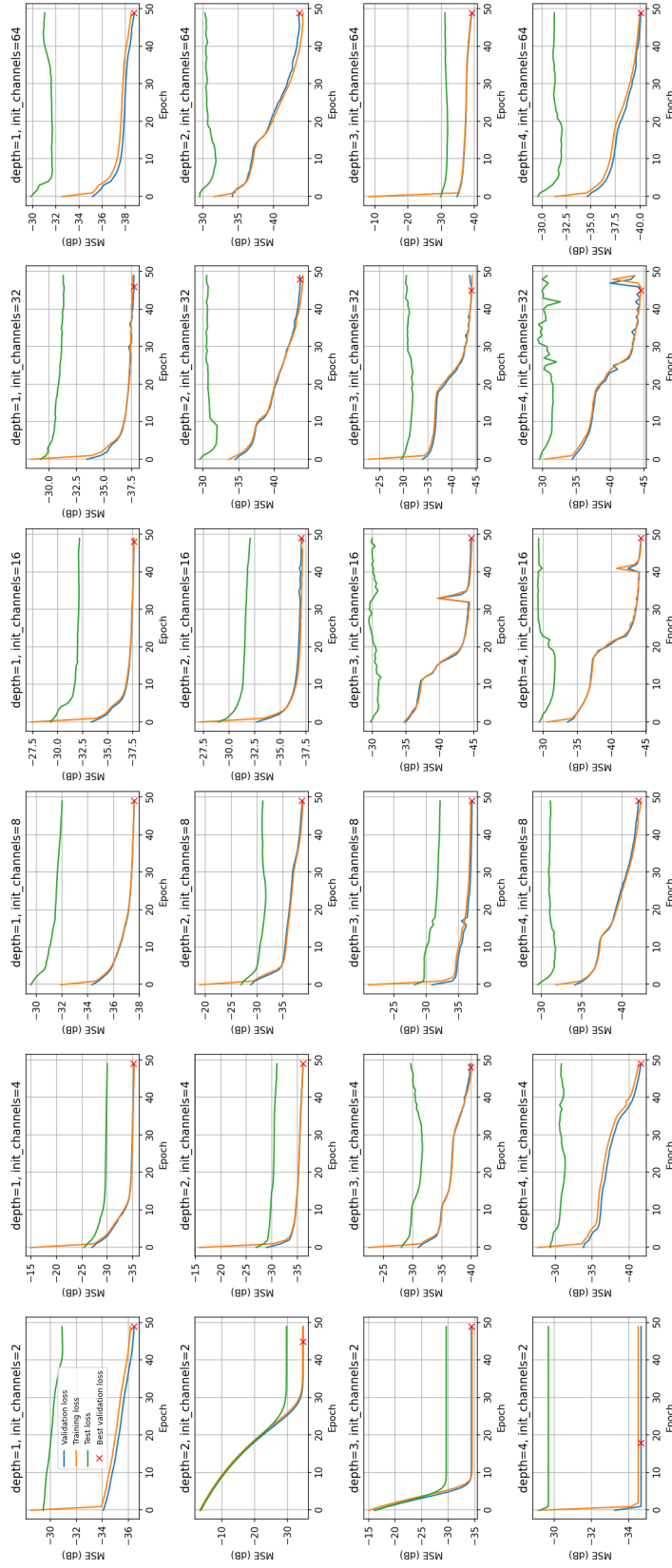


Figure 5.6: Training histories of a grid search test on the network depth and initial channels using synthetic MRI data.