Learnable Slice-to-Volume Reconstruction for Motion Compensation in Fetal Magnetic Resonance Imaging

Constantin Jehn[1], Johanna P. Müller[1], Bernhard Kainz[1,2]

[1]IDEA Lab, Friedrich–Alexander University Erlangen–Nürnberg, DE
[2]Imperial College London, SW7 2AZ, London, UK
constantin.jehn@fau.de

Abstract. Reconstructing motion-free 3D Magnetic Resonance Imaging (MRI) volumes of fetal organs comes with the challenge of motion artefacts due to fetal motion and maternal respiration. Current methods rely on iterative procedures of outlier removal, super-resolution (SR) and slice-to-volume registration (SVR). Long runtimes and missing volume preservation over multiple iterations are still challenges for widespread clinical implementation. We envision an end-to-end learnable reconstruction framework that enables faster inference times and that can be steered by downstream tasks like segmentation. Therefore, we propose a new hybrid architecture for fetal brain reconstruction, consisting of a fully differentiable pre-registration module and a CycleGAN model for 3D image-to-image translation that is pretrained on our custom-generated dataset of 209 pairs of low-resolution (LoRes) and high-resolution (HiRes) fetal brain volumes. Our results are evaluated quantitatively with respect to five different similarity metrics. We incorporate the Learned Perceptual Image Patch Similarity (LPIPS) metric and apply it to quantify volumetric image similarity for the first time in literature. Furthermore, we evaluate the model outputs qualitatively and conduct an expert survey to compare our method's reconstruction quality to an established approach.

1 Introduction

Abnormalities in fetal development can indicate pathological long-term problems. They can impact the survival rates in the perinatal period and beyond, making accurate means of fetal organ reconstruction essential to counselling during pregnancy [1]. Due to its superior soft-tissue contrast and the absence of ionising radiation, MRI is the modality of choice for detailed structural imaging, for example, for the fetal brain. During the imaging process, multiple stacks in orthogonal slicing directions are acquired. These stacks are subject to inter-slice and inter-stack motion artefacts resulting from fetal and maternal motion. The problem formulation can be condensed to the successful generation of HiRes reconstructions from corrupted input data. The majority of established reconstruction methods uses a pre-registration step that brings all input stacks into approximate alignment by 3D-3D registration, followed by an iterative procedure as it was introduced in [2]: outlier-rejection aims to reject voxels that are likely motion artefacts; SR updates combine the information from the different input stacks into a common reconstruction volume. SVR steps iteratively register the single slices from the input stacks to the most recent reconstruction volume. [3] proposes a surrogate measure for intra-stack motion corruption to select the optimal stack during pre-registration and publishes a GPU implementation with significantly lower runtime. A pre-registration step based on CNNs and a new parametrization for the SVR is proposed in [4]. Most recently, Transformers have been proposed to regress the transformation parameters for the SVR [5]. In this work we present an end-to-end learnable hybrid reconstruction framework that combines a pre-registration module and a Cycle-GAN model for volume-to-volume translation.

2 Material and methods

Hybrid reconstruction framework: Registration-based reconstruction algorithms have advanced the image quality of fetal MRI imaging. Yet, long runtimes resulting from the high-dimensional optimisation problem and room for improvement in volume preservation over the iterations remain open challenges in the field. Motivated by these shortcomings and the urge to steer the process by downstream tasks such as segmentation, we strive for an end-to-end learnable reconstruction pipeline. It offers the possibility to incorporate neural network architectures as part of the reconstruction pipeline and profit from advances in learnable SR, image-to-image translation and generative models. We propose a hybrid framework for fetal brain reconstruction. Its structure is outlined in Fig. 1: the fully differentiable pre-registration module P[1] transforms the input stacks into a latent reconstruction space. Subsequently, a CycleGAN model[2], pretrained on our custom dataset, translates the LoRes volume into a high-quality reconstruction volume.

Pre-registration: The preregistration step combines the incoming stacks into a first approximate common volume. The outline of the algorithm is provided in Fig. 2. It reimplements the pre-registration steps from [3], allowing gradients to backpropagate to registration parameters. The first step gives a surrogate measure for the motion artefacts within one stack. The method is based on Singular Value Decomposition and the assumption that neighbouring slices of continuous volumes are highly correlated. The stack with the least artefacts is selected as template, and the remaining stacks are aligned using 3D-3D registration. Subsequently, the common volume undergoes an SR update.

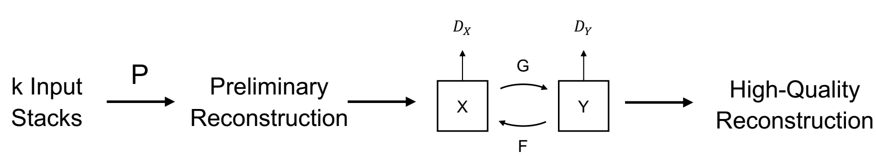

Fig. 1. The proposed hybrid architecture combines the preregistrator P with a pretrained CycleGAN model translating preliminary reconstructions into HiRes reconstructions.

[1] https://github.com/Constantin-Jehn/SVR_pytorch
[2] https://github.com/Constantin-Jehn/Cycle_GAN_3D

Image-to-image translation: The CycleGAN architecture is a variant of the GAN framework for unpaired image-to-image translation [6]. Unpaired models can be further improved by expert-approved HiRes reconstruction even without corresponding LoRes images. The model's structure is depicted in Fig. 1. The generators G and F translate between the set of LoRes volumes X and HiRes volumes Y, such that the results are not distinguishable from the target distribution. The discriminators D_X and D_Y aim to correctly classify true and synthetic samples. The key feature of the architecture is the additional cycle consistency loss, expressed in Eq. 1. It enforces $F(G(x)) \approx x$ and $G(F(y)) \approx y$, constraining the generators to mappings where input and output have a meaningful relationship. The complete objective function can be found in [6].

$$\mathcal{L}_{cyc} = \mathbb{E}_{x\sim p_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y\sim p_{data}(y)}[\|F(G(y)) - y\|_1] \qquad (1)$$

This additional constraint helps circumventing the introduction of random or artificial details during the image generation.

Data: We use a dataset containing 252 MRI studies from mothers in weeks 24-37 of their pregnancy acquired on a Philips Achieva 1.5 T and 3 T scanner. The volumes were acquired using single-shot fast spin echo T2-weighted sequences with half Fourier acquisition (ssFSE) and SENSE. The study was approved by the ethics committee at Imperial College London and the UK's NHS National Research Ethics Service.

Training details: Training the CycleGAN model requires two sets of volumes. However, our pre-registration module only generates approximate reconstructions. Therefore, we use an established slice-to-volume-reconstruction toolkit (SVRTK)[3] to generate surrugate HiRes reconstructions. All MRI stacks with manually generated brain segmentation masks are fed to SVRTK, and the HiRes reconstruction and its intermediate pre-registration result are extracted. The 209 MRI volumes with the respective segmentations are provided as input to our fully differentiable pre-registration module (FDPR). We include the 123 best FDPR pre-registrations, together with 86 pre-registration results from SVRTK (SVPR) in the LoRes set X such that the complete dataset is covered. The generator networks are three-dimensional 9-block ResNet architectures, and the discriminators are fully convolutional networks. Using the Adam optimiser [7], the model is trained over 200 epochs with a linear learning rate (LR) decay after 100 epochs. Three starting LRs, 0.0002, 0.0004 and 0.001, are compared. With a train-test split of 169-40, the training takes approx. 24 hours on one NVIDIA A6000 (48GB) GPU.

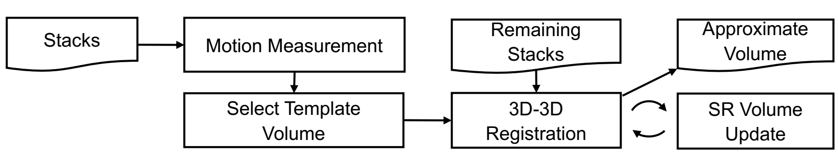

Fig. 2. The preregistration procedure selects the least motion-corrupted volume as a template. The remaining stacks are aligned by 3D-3D registration, followed by an SR update step.

[3] `https://github.com/SVRTK/SVRTK`

3 Results

Quantitative results: Based on image similarity metrics, volumes generated by the CycleGAN model can be compared to target images from set Y, which allows quantifying the model's image-to-image translation performance. To facilitate comparisons, we assess the model with respect to four different global metrics: the peak signal-to-noise ratio (PSNR), the normalised mean square error (NMSE), the structural similarity index measure (SSIM) and normalized cross-correlation (NCC). Moreover, we incorporate the Learned Perceptual Image Patch Similarity (LPIPS) as a metric that has been shown to align better with human-perceived image similarity than standard metrics [8]. While LPIPS has been developed for 2D RGB-images, we aim to quantify similarity on volumetric one-channel data. To apply the 2D metric to our MRI volumes, all 2D slices in three orthogonal directions are extracted, and the channel values are assigned identically to all three RGB channels. We take the mean over all LPIPS values for the individual slices as the final measure. The mean results, with respect to the similarity metrics, on the subsets SVPR, FDPR and their union PR, are reported Table 1 for three different learning rates (0.0002, 0.0004 and 0.001). Each mean value is supplemented by the standard error of the mean (s.e.m.) $\frac{\sigma}{\sqrt{N}}$, with standard deviation σ and the number samples N, to quantify the variability of the mean values taking the size of the test data set into account. The rows SVPR, FDPR and PR report the mean similarity in their subset of X to its HiRes counterpart in Y. The three rows below state the mean similarity of the image-translation result of the generator G to the HiRes volumes for different LR on the subset name above in the table. Firstly, we can observe that the LR of 0.0004 performs best on three out of five metrics on the complete set. The PSNR measures a significant improvement on the complete set, mainly induced by the SVPR set. NMSE, SSIM and NCC indicate no significant improvement.The LPIPS metric improves over all the sets, which also aligns with the qualitative analysis revealing a similarity increase of the CycleGAN outputs to its targets. For the following evaluation steps, the results of the model trained with LR of 0.0004 are considered.

Qualitative results: Fig. 3 presents two example reconstructions taken from the test set. The upper row represents the preregistration inputs for the CycleGAN model, the second row depicts the CycleGAN result, and the last row visualises SVRTK's outputs for reference. For FDPR and SVPR inputs, the CycleGAN model creates reconstructions with significantly improved image quality and increased visual similarity to the SVRTK result. The CycleGAN reconstructions with SVPR inputs are more similar to the target image, which can be explained by the input being an intermediate result of the target volume's reconstruction.

Expert survey: Any reconstruction is the result of either a complex optimisation problem or a deep neural network (NN), and therefore we have to acknowledge the absence of reliable ground truth data. To assess the reconstruction quality of our model compared to SVRKT, we conducted a survey among 16 domain experts. The survey is designed as two alternative-forced-choice (2AFC) test. The experts were shown three orthogonal views of a CycleGAN reconstruction and an SVRTK reconstruction of the same input. They had to answer which reconstruction they would prefer for interpretation. The indication of the methods was not given, and the answers were shuffled randomly over the survey. Fig. 4 shows that 19 out of 25 reconstruction comparisons were answered

Tab. 1. Image-to-image translation results with learning rates of 0.0002, 0.0004 and 0.001. For all rows, the result of three iterations of SVRTK is taken as a reference. The mean values, over the test set data, for different metrics are reported with standard error of the mean as a variability measure.

References	PSNR ↑	NMSE ↓	SSIM ↑	| NCC | ↑	LPIPS ↓
SVPR	23.11 ± 0.60	0.110 ± 0.010	**0.93** ± 0.010	**0.94** ± 0.009	0.10 ± 0.0064
G(lr=0.0002)	27.01 ± 0.40	0.069 ± 0.004	0.92 ± 0.010	0.91 ± 0.011	0.08 ± 0.0062
G(lr=0.0004)	**27.80** ± 0.27	**0.064** ± 0.002	0.92 ± 0.009	0.91 ± 0.011	**0.08** ± 0.0059
G(lr=0.001)	23.36 ± 0.23	0.140 ± 0.005	0.82 ± 0.011	0.85 ± 0.015	0.16 ± 0.0070
FDPR	21.80 ± 0.28	0.170 ± 0.009	0.82 ± 0.010	0.83 ± 0.010	0.19 ± 0.0060
G(lr=0.0002)	22.06 ± 0.28	0.181 ± 0.009	0.81 ± 0.011	0.83 ± 0.010	0.15 ± 0.0061
G(lr=0.0004)	22.10 ± 0.26	0.190 ± 0.009	0.82 ± 0.010	0.83 ± 0.010	0.14 ± 0.0059
G(lr=0.001)	21.99 ± 0.23	0.180 ± 0.006	0.79 ± 0.008	0.82 ± 0.099	0.19 ± 0.047
PR	22.39 ± 0.33	0.140 ± 0.008	0.87 ± 0.011	0.88 ± 0.069	0.15 ± 0.0085
G(lr=0.0002)	24.29 ± 0.46	0.132 ± 0.010	0.86 ± 0.011	0.87 ± 0.010	0.12 ± 0.0067
G(lr=0.0004)	24.66 ± 0.49	0.144 ± 0.012	0.87 ± 0.011	0.87 ± 0.010	0.11 ± 0.0065
G(lr=0.001)	22.61 ± 0.12	0.160 ± 0.005	0.81 ± 0.007	0.83 ± 0.009	0.17 ± 0.0046

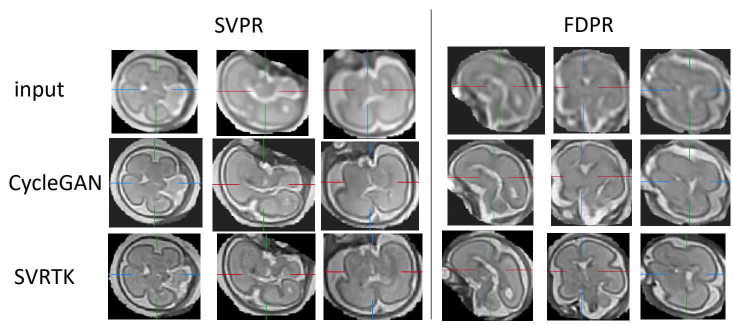

Fig. 3. The first row shows the LoRes input volumes, and the second row the CycleGAN volume-to-volume translation results. The third row gives SVRTK's reconstruction as a reference.

in favour of SVRTK, by the experts. There are only two edge cases where the CycleGAN significantly outperforms the reference method.

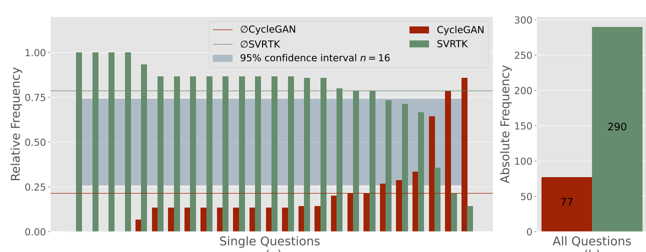

Fig. 4. In the expert survey, 19 out of 25 questions were answered significantly in SVRTK's favour. There are two edge cases where the CycleGAN model significantly outperforms SVRTK.

4 Discussion

Interpretation and evaluation: From the quantitative analysis and, particularly, the PSNR and LPIPS results, we observe an increase in image similarity of model outputs. This is further supported by the presented example reconstructions, showing enhanced reconstruction quality compared to the model inputs while preserving their key features. Hence, we can state that CycleGAN models can improve preregistration quality and should be considered for future work on end-to-end learnable reconstruction frameworks. The expert survey showed that the developed method is not yet on the level of specialized SVR reconstruction algorithms, which might be a result of observer bias and hidden confounders that need to be investigated in future work. Effects on downstream data analysis and diagnosis will also provide an intriguing direction for further work.

Outlook: A major limitation of this work is the generated training data for the model. Working with real data has the advantage of a realistic reconstruction setting. However, the absence of true HiRes data and the necessity to construct the target set Y from the output of another framework leaves us with an implicit upper bound of the method, which is predefined by the chosen method. Hence, we propose to carry on the presented work with HiRes volumes from acquisition with little or no fetal motion and introducing synthetic motion artefacts as in [5] to construct the LoRes set X. Additionally, a canonical atlas representation of the volumes in X could allow the translation model to transfer morphological details more precisely, as they would be spatially closer across the different volumes in the training set.

Acknowledgements: The authors gratefully acknowledge the HPC resources provided by (NHR@FAU ID PatRo-MRI). NHR funding is provided by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by the German Research Foundation (DFG) – 440719683. Data had been provided by the iFind project, Wellcome Trust IEH 102431.

References

1. Gholipour A, et al. A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. Scientific reports 7.1 (2017), p. 476.
2. Kuklisova-Murgasova M, et al. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. Medical Image Analysis 16 (8 2012), pp. 1550–1564.
3. Kainz B, et al. Fast volume reconstruction from motion corrupted stacks of 2D slices. IEEE Transactions on Medical Imaging 34 (9 2015), pp. 1901–1913.
4. Hou B, et al. 3-D reconstruction in canonical co-ordinate space from arbitrarily oriented 2-D images. IEEE transactions on medical imaging 37.8 (2018), pp. 1737–1750.
5. Xu J, et al. SVoRT: iterative transformer for slice-to-volume registration in fetal brain MRI. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer. 2022, pp. 3–13.
6. Zhu JY, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision. 2017, pp. 2223–2232.
7. Kingma DP, et al. Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014).
8. Zhang R, et al. The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 586–595.