UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR MEDIZINISCHE INFORMATIK

IM FOCUS DAS LEBEN

MRT-Registration

MRT-Registration

Masterarbeit

verfasst am
Institut für Medizinische Informatik

im Rahmen des Studiengangs
Medizinische Informatik
der Universität zu Lübeck

vorgelegt von
Jan Meyer

ausgegeben und betreut von
Prof. Dr. Mattias Heinrich

mit Unterstützung von
Ziad Al-Haj Hemidi, Eytan Kats

Lübeck, den 1. Januar 2025

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Jan Meyer

Zusammenfassung

Irgendwas über MRT und Registration

Abstract

Something about MRI and registration

Acknowledgements

Danke an Mattias, Ziad und Eytan für die gute Betreuung.

Contents

1

Introduction

Introduction of stuff.

1.1 Contributions of the Thesis

We implemented stuff.

1.2 Related Work

There are many papers on image registration in general, however in the context of medical image registration with deep learning their number is reduced due to the specialized nature of the subject at hand. Yet, there are a couple of papers that give a good overview of the topic. They give a brief overview of registration methods, basics of deep learning with already existing networks for image registration as well as covering potential applications and challenges [1, 2, 3, 4, 5]. A lot of different approaches for medical image registration are based on *VoxelMorph* [6], such as both *Fourier-Net* [7] and its successor *Fourier-Net+* [8].

1.3 Structure of the Thesis

This Thesis contains a lot of stuff in different chapters.

2

Basics

In this chapter the basics of the thesis are explained.

2.1 Magnetic Particle Imaging

Here MRI is described.

2.2 Image Registration

Image registration is a challenging, yet important task for image processing. It can be described as the process of transforming different image datasets into one coordinate system with matched imaging contents [2]. In the medical field this can be used for clinical applications such as disease diagnosis and monitoring, image-guided treatment delivery, and post-operative assessment. Medical image registration is typically used to preprocess data for tasks like object detection (for e.g. tumor growth monitoring) and segmentation (for e.g. organ atlas creation) where variation in spatial resolution is common between modalities like CT and MRI and patients. Thus the performance of these methods is dependent on the quality of image registration [1].

Medical image registration was often done manually by clinicians, however, registration tasks are often challenging and the quality of manual alignments is dependent on the expertise of the user. These manual registrations are thus not only time consuming, but also hardly reproducible leading to high interobserver-variability. The need for automatic registration is very much apparent, but this task remained hard to solve for a long time, requiring a lot of computational power and time for computer algorithms to solve the problem. While neural networks also require a lot of computational power and time to train, they promise fast execution after training. With the rise of deep learning these network gained popularity and now pose a real alternative to conventional algorithms and manual registration [2]. We will discuss these new approaches in the next section, but first we need to formally define our problem.

In pair-wise image registration two images (F and M) are to be aligned, with F denoting the fixed and M the moving image. T is the desired spatial transformation that aligns the

two images. This can be posed as an optimization problem:

$$T' = \arg\max S(F, T(M)), \qquad (2.1)$$

with T' being the best transformation that maximizes the similarity S between the two images. This process is done iteratively improving estimates for the desired T, such that the defined similarity in the cost function is maximized [1].

Transformations can be categorized as rigid, affine, and deformable. A rigid transformation consists of rotation and translation; an affine transformation includes translations, rotations, scaling, and sheering; the two kinds of transformations are described as a 2D single matrix. Unlike rigid and affine transformation, deformable transformation is a high-dimension problem that we need to formulate by a 3D matrix for 2D deformable registration i.e., a so-called deformation field. While rigid and affine registration algorithms have already achieved good performance in many applications, deformable registration is still a challenging task due to its intrinsic complexity, particularly when the deformation is large. However, these are also the transformations most likely encountered in clinical practice as it can be utilized to fuse information from different modalities such as MRI and CT [4]. Additionally, deformable image registration can also be utilized for various computer-assisted interventions like biopsy [9] and (MRI-guided) radiotherapy [10, 11].

Intuitively, deformable image registration is an ill-posed problem, making it fundamentally different from other computer vision tasks such as object localization, segmentation or classification. Given two images, deformable image registration aims to find a spatial transformation that warps the moving image to match the fixed image as closely as possible. However, there is no ground-truth available for the desired deformation field and without enforcing any constraints on the properties of the spatial transformation, the resulting cost function is ill-conditioned and highly non-convex. In order to address the latter and ensure tractability, all image registration algorithms regularize the estimated deformation field, based on some prior assumptions on the properties of the underlying unknown deformation [1].

Many methods have been proposed for medical image registration to deal with the complex challenges of this task. Popular conventional registration methods include optical flow [12], demons [13] and many more. However, most of these still lack accuracy and computation speed, which makes newer deep learning approaches all the more interesting [3].

2.3 Deep Learning Architectures

Neural networks, despite the theoretical concepts being around for decades, have seen a meteoric rise in popularity over the last few years as constraints on computational power have been alleviated. Especially deep neural networks, which are often summarized under the term deep learning (DL). Recent years have witnessed an almost exponential growth in the development and use of DL algorithms, sustained thus far by rapid improvements in computational hardware (e.g. GPUs). Consequently, clinical applications requiring image classification, segmentation, registration, or object detection/localization, have

witnessed significant improvements in algorithmic performance, in terms of accuracy and/or efficiency [1]. The following network architecture are widely used for different tasks including medical image registration.
Some basic stuff about network training, testing and different architectures that are relevant for the later Chapters.

Convolutional Neural Networks

Convolutional neural networks (CNNs) are a type of deep neural networks with regularized multilayer perceptron, which are mainly used for image processing. CNNs use convolution operations instead of general matrix multiplications in typical neural networks. These convolutional filters make CNNs very suitable for visual signal processing. Because of their excellent feature extraction ability, CNNs are some of the most successful models for image analysis. Different variants of CNN have been proposed and have achieved the-state-of-art performances in various image processing tasks. A typical CNN usually consists of multiple convolutional layers, max pooling layers, batch normalization layers, sometimes dropout layers, a sigmoid or softmax layer. In each convolutional layer, multiple channels of feature maps are extracted by sliding trainable convolutional kernels across the input feature maps. Hierarchical features with high-level abstraction are extracted using multiple convolutional layers. These feature maps usually go through multiple fully connected layer before reaching the final decision layer. Max pooling layers are often used to reduce the image sizes and to promote spatial invariance of the network. Batch normalization is used to reduce internal covariate shift among the training samples. Weight regularization and dropout layers are used to alleviate data overfitting [3]. The loss function is often defined as the difference between the predicted and the target output. CNNs are usually trained by minimizing the loss via gradient back propagation using optimization methods like Adam [14].

U-Net

The U-Net [15] architecture is an extension of the typical CNNs structure typically used for image segmentation, however it can also be used for image registration tasks. It adopts symmetrical contractive and expansive paths with skip connections between them. The encoding blocks on the left extract important features from the image using convolution layers and max pooling, which are then stored in the latent space in the middle. From there it is reconstructed using upsampling and convolutions in the decoding blocks on the right. Additionally, skip connections are used to improve the spatial resolution of the segmentation. This architecture allows effective feature learning from a small number of training datasets [3].

Autoencoders

An autoencoder (AE) is a type of CNNs that learns to reconstruct an image from its input without supervision. AEs usually consists of an encoder which extracts the input features, which are stored a low-dimensional latent state space, similar to a U-Net, and a de-

coder which restore the original input from the latent space. To prevent an AE from learning an identity function, regularized autoencoders were invented, which can be used for e.g. denoising AEs. Variational AEs (VAEs) are generative models that learn latent representation using a variational approach, which constrains the variability of the outputs. VAEs can been used for anomaly detection and image generation [3].

Generative Adversarial Networks

Generative adversarial networks (GANs) consist of two competing networks, a generator and a discriminator. The generator is trained to generate artificial data that approximate a target data distribution from a low-dimensional latent space similar to an AE. The discriminator is trained to distinguish the artificial data from actual data. The discriminator encourages the generator to predict realistic data by penalizing unrealistic predictions via learning. Therefore, the discriminative loss could be considered as a dynamic network-based loss term. The generator and discriminator both are getting better during training to reach Nash equilibrium. In medical imaging, GANs have been used to perform image synthesis for inter- or intra-modality, such as MRI to synthetic CT and vise versa. In medical image registration, GANs are usually used to either provide additional regularization or translate multi-modal registration to unimodal registration [3].

2.4   Deep Learning for Image Registration

Recently, there has been a surge in the use of deep learning based approaches for medical image registration. Their success is largely due to their ability to perform fast inference, and the flexibility to leverage auxiliary information such as anatomical masks as part of the training process. The most effective methods, such as *VoxelMorph* [6], typically employ a U-Net style architecture to estimate dense spatial deformation fields. These methods require only one forward pass during inference, making them orders of magnitude faster than traditional iterative methods. Following the success of *VoxelMorph*, numerous deep neural networks have been proposed for various registration tasks [8]. Other approaches also utilize CNNs, AEs and GANs. Typical strategies are discussed in more detail in the following sections.

Supervised Registration

Supervised Registration describes training a network with a ground truth displacement field that is either real or synthetic. Thus the loss can easily be calculated as the difference in the displacement fields of the network and ground truth. These methods have achieved notable results with real displacement fields as supervision. However, this is limited by the size and the diversity of the dataset. The real displacement field are also often calculated by conventional algorithms, which might limit their effectiveness for difficult problems. Fully supervised methods are widely studied and have notable results, but the generation of real or synthetic displacement fields is hard, and these displacements fields might be different from the real ground truth, which can impact the accuracy and efficiency of these kinds of methods [4].

Unsupervised Registration

As the preparation of the ground truth displacement field for supervised methods is inconvenient, limitations in generalizing results in different domains and various registration tasks are inevitable. Thus, unsupervised registration has a more convenient training process with paired images as inputs, but without a ground truth. Generally, unsupervised learning consists of similarity-based and GAN-based methods [4].

3

Methodology

In this chapter the main part of the actual work is discussed.

3.1 Data

Describe the data used for evaluation.

3.2 Network Architecture

Describe the network being used.

Fourier-Net

The *Fourier-Net* is new approach that aims to learn a low-dimensional representation of the displacement field in a band-limited Fourier domain instead of the full field in the spatial domain. This band-limited representation is then decoded by our devised model-driven decoder to the dense, full-resolution displacement field in the spatial domain. These changes allow for fewer parameters and computational operations, resulting in faster inference speeds [7]. The architecture is based on the U-Net [15], like most deep registration approaches, but replaces the expansive path with a parameter-free model-driven decoder as mentioned before.
The encoder of *Fourier-Net* consists of a CNN that is based on the networks in [16, 17] and takes two images (fixed and moving) as inputs. The output is a displacement field of the images that is then converted from the spatial domain into the Fourier domain via an discrete Fourier transformation (DFT) and band-limiting by low-pass filtering the displacement field. From there, the band-limiting representation is padded with zeros to the full resolution of the original displacement field. The field is then recovered by using the inverse DFT (iDFT) to convert it back into the spatial domain. This displacement field is then used to warp the moving image into the fixed image. For this a warping layer similar to [6] is used. Optionally, squaring and scaling layers [18] can also be added

before warping the image to encourage a diffeomorphism in final deformation. A diffeomorphic deformation is defined as a smooth and invertible deformation, thus the output of the iDFT layer can be regarded as a stationary velocity field denoted by v instead of the displacement field θ. In group theory, v is a member of Lie algebra, and it can be exponentiated to obtain the diffeomorphic deformation. A diagram of this can be seen in Figure 3.1. After the moving image has been warped the unsupervised loss $\mathcal{L}$ can be calculated as follows:

$$\mathcal{L}(\Theta) = \min\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{Sim}(I_{M_i}\circ(\phi_i(\Theta)+\mathrm{Id})-I_{F_i})\right)+\lambda\cdot\left(\frac{1}{N}\sum_{i=1}^{N}\|\nabla\phi_i(\Theta)\|_2^2\right)\right],$$

(3.1)

where $\circ$ denotes the warping operation, N the number of training pairs with moving images I_{M_i} and fixed images I_{F_i}, Θ the network parameters, ϕ the displacement field, Id the identity grid and ∇ the first order gradient. The two terms of the loss, $\mathcal{L}_{Sim}$ determining the similarity between warped moving images and fixed images via MSE or NCC, and the second term defining the smoothness of the displacement fields, are balanced using the scalar parameter λ.

When using the squaring and scaling layers, thus making the deformation of the moving image diffeomorphic, the loss can be modified by replacing the displacement field θ with the velocity field v:

$$\mathcal{L}(\Theta) = \min\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{Sim}(I_{M_i}\circ Exp(v_i(\Theta)-I_{F_i}))+\lambda\cdot\left(\frac{1}{N}\sum_{i=1}^{N}\|\nabla v_i(\Theta)\|_2^2\right)\right].$$
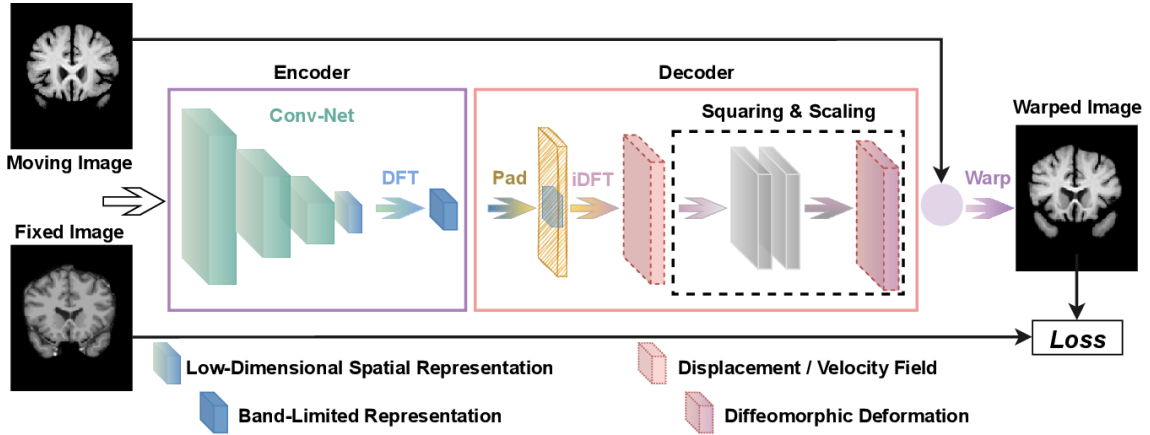
(3.2)

Figure 3.1: Architecture of *Fourier-Net* taken from [7].

Fourier Net+

The *Fourier-Net+*, as the name suggests, is an extension of Fourier-Net which takes the band-limited spatial representation of the images as input, instead of their original full-resolution counterparts. This leads to further reduction in the number of convolutional

layers in the contracting path of the network, resulting in a decrease of parameters, memory usage, and computational operations. This makes *Fourier-Net+* even more efficient than its older predecessor [8].

As seen in Figure 3.2, the network architecture is almost the same as for *Fourier-Net* (see Figure 3.1 for reference). However, while the decoder, and thus the loss function, remain the same, the encoder is slightly altered to make the network even more efficient. For this, similarly to the decoder, a DFT is used, however this time the idea is applied to the input images. These are first transformed into the Fourier domain, then low-pass filtered and finally reconstructed from their band-limited representation back into the spatial domain via an iDFT. The two images, now compressed, are the input for the encoder of *Fourier-Net*, meaning the CNN and following DFT. However, due to the band-limiting before the CNN, the latter can be made much more light-weight, thus reducing computational cost. This is visualized in Figure 3.3. Thus, *Fourier-Net+* too is overall lighter than the baseline *Fourier-Net* in terms of the number of parameters and computations. However, such a light network may face limitations in accurately capturing complex deformations. To counter this potential weakness, the authors propose a cascaded version of *Fourier-Net+*, which uses multiple versions of *Fourier-Net+* cascaded one after the other to achieve a better overall displacement field [8]. A schematic for this can be seen in Figure 3.4.
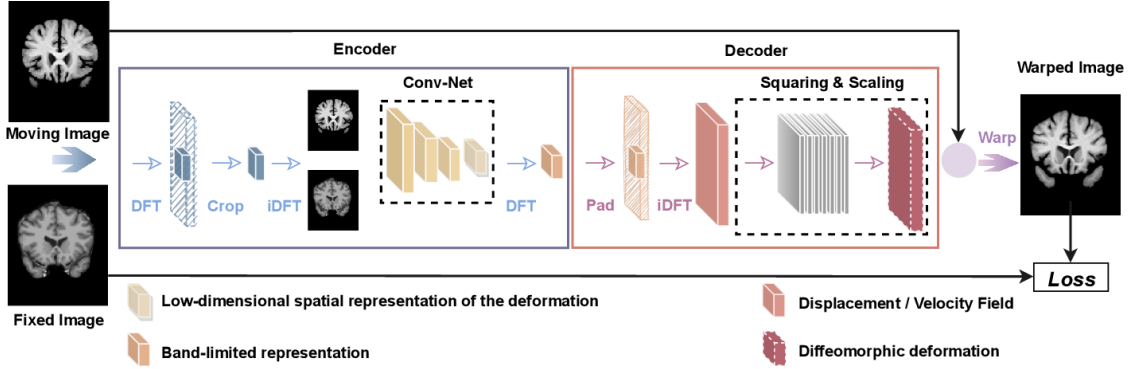

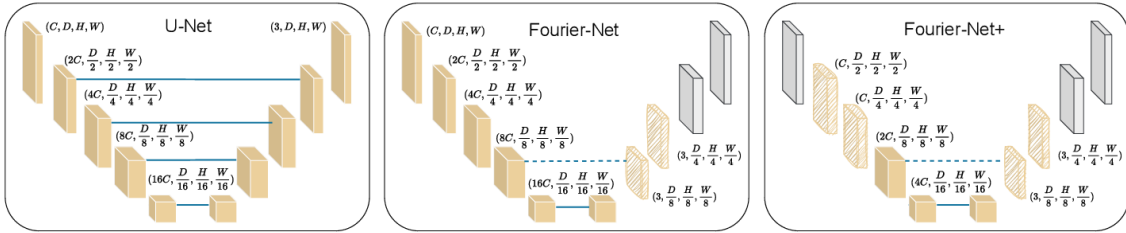Figure 3.2: Architecture of *Fourier-Net+* taken from [8].

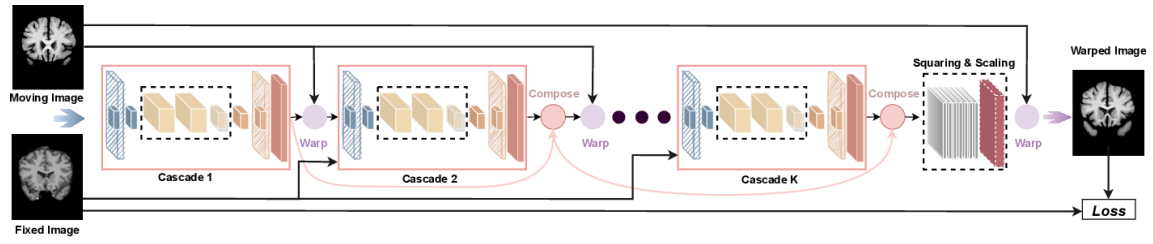Figure 3.3: Architecture of the CNN for a typical U-Net, *Fourier-Net* and *Fourier-Net+* taken from [8].

Figure 3.4: Cascaded version of *Fourier-Net+* taken from [8].

3.3 Evaluation

Describe the evaluation methods being used.

4

Results and Discussion

Here go the results with the discussion.

5

Conclusion

Summery of all stuff...

Bibliography

[1] Chen, X., Diaz-Pinto, A., Ravikumar, N., and Frangi, A. Deep learning in medical image registration. In: *Progress in Biomedical Engineering*, Dec. 2020. ISSN: 2516-1091. DOI: 10.1088/2516-1091/abd37c.

[2] Haskins, G., Kruger, U., and Yan, P. Deep learning in medical image registration: a survey. In: *Machine Vision and Applications* 31(1–2), Jan. 2020. ISSN: 1432-1769. DOI: 10.1007/s00138-020-01060-x.

[3] Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., and Yang, X. Deep learning in medical image registration: a review. In: *Physics in Medicine & Biology* 65(20):20TR01, Oct. 2020. ISSN: 1361-6560. DOI: 10.1088/1361-6560/ab843e.

[4] Zou, J., Gao, B., Song, Y., and Qin, J. A review of deep learning-based deformable medical image registration. In: *Frontiers in Oncology* 12, Dec. 2022. ISSN: 2234-943X. DOI: 10.3389/fonc.2022.1047215.

[5] Chen, J., Liu, Y., Wei, S., Bian, Z., Subramanian, S., Carass, A., Prince, J. L., and Du, Y. *A survey on deep learning in medical image registration: new technologies, uncertainty, evaluation metrics, and beyond*. 2023. DOI: 10.48550/ARXIV.2307.15615.

[6] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. Voxel-Morph: A Learning Framework for Deformable Medical Image Registration. In: *IEEE Transactions on Medical Imaging* 38(8):1788–1800, Aug. 2019. ISSN: 1558-254X. DOI: 10.1109/tmi.2019.2897538.

[7] Jia, X., Bartlett, J., Chen, W., Song, S., Zhang, T., Cheng, X., Lu, W., Qiu, Z., and Duan, J. Fourier-Net: Fast Image Registration with Band-Limited Deformation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 1. 2023, pp. 1015–1023.

[8] Jia, X., Thorley, A., Gomez, A., Lu, W., Kotecha, D., and Duan, J. *Fourier-Net+: Leveraging Band-Limited Representation for Efficient 3D Medical Image Registration*. 2023. DOI: 10.48550/ARXIV.2307.02997.

[9] Tam, A. L., Lim, H. J., Wistuba, I. I., Tamrazi, A., Kuo, M. D., Ziv, E., Wong, S., Shih, A. J., Webster, R. J., Fischer, G. S., et al. Image-Guided Biopsy in the Era of Personalized Cancer Care: Proceedings from the Society of Interventional Radiology Research Consensus Panel. In: *Journal of Vascular and Interventional Radiology* 27(1):8–19, Jan. 2016. ISSN: 1051-0443. DOI: 10.1016/j.jvir.2015.10.019.

[10] Chen, A. M., Hsu, S., Lamb, J., Yang, Y., Agazaryan, N., Steinberg, M. L., Low, D. A., and Cao, M. MRI-guided radiotherapy for head and neck cancer: initial clinical experience. In: *Clinical and Translational Oncology* 20(2):160–168, June 2017. ISSN: 1699-3055. DOI: 10.1007/s12094-017-1704-4.

[11] Rigaud, B., Simon, A., Castelli, J., Lafond, C., Acosta, O., Haigron, P., Cazoulat, G., and Crevoisier, R. de Deformable image registration for radiation therapy: princi-

ple, methods, applications and evaluation. In: *Acta Oncologica* 58(9):1225–1237, June 2019. ISSN: 1651-226X. DOI: 10.1080/0284186x.2019.1620331.

[12] Yang, D., Li, H., Low, D. A., Deasy, J. O., and Naqa, I. E. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. In: *Physics in Medicine and Biology* 53(21):6143–6165, Oct. 2008. ISSN: 1361-6560. DOI: 10.1088/0031-9155/53/21/017.

[13] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. Diffeomorphic demons: Efficient non-parametric image registration. In: *NeuroImage* 45(1):S61–S72, Mar. 2009. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2008.10.040.

[14] Kingma, D. P. and Ba, J. *Adam: A Method for Stochastic Optimization*. 2014. DOI: 10.48550/ARXIV.1412.6980.

[15] Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241. ISBN: 9783319245744. DOI: 10.1007/978-3-319-24574-4_28.

[16] Zhang, J. *Inverse-Consistent Deep Networks for Unsupervised Deformable Image Registration*. 2018. DOI: 10.48550/ARXIV.1809.03443.

[17] Mok, T. C. and Chung, A. C. Fast Symmetric Diffeomorphic Image Registration with Convolutional Neural Networks. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020. DOI: 10.1109/cvpr42600.2020.00470.

[18] Dalca, A. V., Balakrishnan, G., Guttag, J., and Sabuncu, M. R. Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration. In: *Lecture Notes in Computer Science*. Springer International Publishing, 2018, pp. 729–738. ISBN: 9783030009281. DOI: 10.1007/978-3-030-00928-1_82.