

## CHAPTER 3

---

# DEFORMABLE IMAGE REGISTRATION WITH DEEP NEURAL NETWORKS

### *Highlights of the Chapter*

- *Registration with deep learning*
- *Supervised (Dual/ Weakly) transformation estimation*
- *Unsupervised (Similarity/ Feature based) transformation estimation*

As we know, the process of transforming different image datasets into one coordinate system is termed as Deformable Image registration (when the transformation is deformable), which has significant applications in medical image analysis. Deformable transformation models are required in several registration applications that often prohibit the use of traditional registration methods because of the computational expense associated. To address the potential shortcomings of classical deformable registration approaches, methods for deep neural network based registration has been extensively developed in recent years. Deep learning has changed the scenario of image registration. This chapter comprehensively explores the field of deep learning based deformable registration approaches including highlights of the challenges faced.

Deep learning can be assumed as a subset of machine learning that uses different kinds of neural network architectures consisting of a large number of layers and utilized for various applications [1]. Several publicly available libraries like Tensorflow, Keras, Pytorch etc are used for the architectures as well as strategies for training the neural networks. Deep learning has

achieved state-of-the-art performances in many computer vision tasks while it started its journey with the success of AlexNet in the ImageNet challenge of 2012 [2]. Initially, the performance of iterative, intensity-based registration was increased using deep learning [3, 4], followed by investigating the applications of reinforcement learning to registration [5, 6]. In order to develop faster registration methods using deep learning based transformation estimation, several challenges were faced, in which task specific procuring/ generating ground truth data was the notable one. This hurdleship motivated many groups to develop unsupervised frameworks for transformation estimation [7, 8]. Again this upgraded framework encountered a challenge of image similarity quantification to which current efforts using information theory-based similarity metrics, segmentations of anatomical structures, and generative adversarial network-like frameworks have shown promising results. The techniques of transformation estimation using deep neural networks for deformable image registration can be divided into two specific groups [9]: Supervised (Fully or Weakly) & Unsupervised (Similarity Metric or Feature constraint).

### **3.1. Supervised Transformation Estimation**

This objective for the development of these types of networks is that it could estimate the transformation that corresponds to optimal similarity in one step. However, the exclusive use of ground truth data for defining the loss function is one of the major challenges of fully supervised transformation estimation. Rather than using neural networks as iterative optimizers, performing deformable registration using it significantly speeds up the process. Methods that use supervision for registration are reviewed in this portion. First the discussion is done on the fully supervised approaches and then on dual/weakly supervised approaches. Later in this section, a description of notable works on supervised transformation estimation is presented in Table 3.1.

### 3.1.1. Fully Supervised Transformation Estimation

The deformation field is predicted in [10] using an FCN (consisting of an U-net-like architecture [11]) by Yang et al. and it is used to register 2D/3D intersubject brain MR volumes. For the input of the network, the pixel initial momentum values of the image volumes are utilized, large diffeomorphic metric mapping is used to provide a basis, and finally upgrading these values to compute the predicted deformation field. Rohe et al. [12] presented a U-Net based (modified) registration of 3D cardiac MR volumes using mesh segmentations to compute the reference transformation for a given image pair. Cao et al. [13] utilized image patches of a pair of 3D brain MR volumes as input to CNN in order to map their respective displacement vector. The registration is done using the deformation field achieved by combining the displacement vectors for a particular image. To perform the deformable registration of abdominal MR images, in [14], Lv et al. used a CNN while considering the deformation caused by respiration. This approach found to be better than non-motion corrected registrations and local affine registration. Sokooti et al. [15] performed an intra-subject registration of 3D chest CT images using a multi-scale CNN to predict the deformation field. In this article, random DVFs are used to augment the dataset. In [16], multi-scale, random transformations of aligned 3D lung CT image pairs are used (maintaining realistic image appearance) to eliminate the need for manually annotated ground truth data while also. Here, a 3DCNN is used by Eppenhof et al. to perform the deformable registration. Ito et al. [17] used a CNN on 3D brain MR volumes of ADNI dataset to learn plausible deformations for ground truth data generation. As the quality of the registrations using this framework is dependent on the quality of the ground truth registrations, it motivated the development of partially supervised/unsupervised approaches.

### **3.1.2. Dual/Weakly Supervised Transformation Estimation**

In dual supervision, ground truth data and image similarity metric both are used to train a model, while the segmentation overlapping of corresponding anatomical structures is used to design the loss function for weak supervision. A hierarchical, dual supervised learning (by gap-filled U-Net) is used by Fan et al. [18] to predict the deformation field for 3D brain MR registration. Hu et al. [19] used two networks: local-net (for affine) and globalnet (for dense deformation field) and utilized label similarity to train their network to perform MR-TRUS registration. In [20], they combine these networks in an end-to-end framework. For predicting the deformation of MR-TRUS registration, simultaneous maximization of label similarity and minimization of an adversarial loss term is done in another work by Hu et al. [21]. On the basis of choosing proper hyper parameters, the predicted transformation is forced to generate a realistic image while using the adversarial loss as a regularization term. Hering et al. [22] introduced both segmentation overlap and edge-based normalized gradient fields distance to develop loss function for cardiac motion tracking in 2D cine-MR images. Limitations related to the trustworthiness and expense of ground truth is reduced by these transformation estimation methods. However, they still require manually annotated data (ground truth or segmentations).

### **3.2. Unsupervised Transformation Estimation**

As getting reliable ground truth remains a significant difficulty, researchers are motivated towards developing the unsupervised approaches. The spatial transformer network (STN) is one of the useful techniques. Several methods (including this thesis) use an STN for the deformations associated with registrations. This section encompasses the details of unsupervised methods using similarity metrics and feature representations. Notable works are presented in Table 3.2.

**Table 3.1.** Exploration of related literatures (supervised deformable transformation estimation)

<b>Ref. No.</b>	<b>Author</b>	<b>Registration Class</b>	<b>Supervision</b>	<b>Model</b>
[10]	Yang et al., 2016	Monomodal (Brain; MR)	Real Transforms	FCN
[12]	Rohe et al., 2017	Monomodal (Cardiac; MR)	Real Transforms	SVF-Net
[13]	Cao et al., 2017	Monomodal (Brain; MR)	Real Transforms	9-Layer CNN
[15]	Sokooti et al., 2017	Monomodal (Chest; CT)	Synthetic Transforms	RegNet
[14]	Lv et al., 2018	Monomodal (Abdominal; MR)	Real Transforms	CNN
[16]	Eppenhpf et al., 2018	Monomodal (Lung; CT)	Synthetic Transforms	U-Net
[17]	Ito et al., 2018	Monomodal (Brain; MR)	Synthetic Transforms	GoogleNet
[18]	Fan et al., 2019	Monomodal (Brain; MR)	Real Transforms + Similarity Met.	U-Net
[20]	Hu et al., 2018	Multimodal (Prostate; MR-US)	Segmentations	30-Layer FCN
[21]	Hu et al., 2018	Multimodal (Prostate; MR-US)	Segmentations + Adversarial Loss	GAN
[22]	Hering et al., 2019	Multimodal (Prostate; MR-US)	Segmentations + Similarity Metric	U-Net GAN

### 3.2.1. Similarity Metric-based Unsupervised Transformation Estimation

In [23], de Vos et al. presented a deformable registration of 4D cardiac cine-MR volumes using NCC to train an FCN while the moving volume is deformed by DVF. A multistage, multiscale approach is developed by them in [7] to perform unimodal registration using NCC and a bending-energy regularization term for training the networks on several datasets which predicts an affine transformation and subsequent coarse-to-fine deformations utilizing a B-Spline transformation model. A deformable intersubject registration of 3D brain MR volumes is performed in Li et al. [6] while the “self-supervising” training of FCN is done by employing NCC and several smoothing constraints. A neural network model is used in [24] to learn the relationship between image similarity metric values and TRE when registering CT image volumes. Similarly inspired, a general framework for unsupervised image registration is proposed by Balakrishnan et al. [25], which trained using predefined image similarity metrics. Later in another work by them [26], diffeomorphism is integrated to the transformer layer to compute a velocity field for obtaining an accurate deformation. The model in [27] minimizes the upper bound of the SSD (UB SSD) between the warped and fixed 3D brain MR images, while Sheikhsafari et al. [29] performed the deformable registration using learned latent representations from 2D cardiac cine-MR volumes. Stergios et al. [28] jointly estimated both linear (affine) and local (deformable) transformation using a CNN to register inhale–exhale pairs of lung MR volumes while an MSE term and regularization terms are employed for the loss function. A transfer learning-based approach is developed by Ferrante et al. [30], where the training and testing domains are different. A U-net like architecture and a STN is used to perform the unimodal registration of both X-ray and cardiac cine images using NCC as the primary loss term.

An intra-modality image similarity based training is applied by Cao et al. [31] to supervise the multimodal deformable registration of 3D pelvic CT/MR volumes. The loss function consists of the NCC between two warped images (warped using the ground truth and warped using the predicted transformation). Zhang et al. [32] introduces an inverse-consistent regularization term and an anti-folding regularization term to ensure that the symmetric diffeomorphic transformations do not result in folding, while utilizing the MSD for their unsupervised inverse-consistent deep network (ICNet). Mahapatra et al. [33] used loss function is consisting of conditional and cyclic constraints, in accordance with NMI, structural similarity index measure (SSIM), and a feature perceptual loss term (determined by the SSD between VGG outputs) to train a Generative Adversarial Network (GAN) for learning the plausible deformations implicitly on cardiac cine images and multimodal retinal images. In another work, Mahapatra et al. [34] proposed simultaneous segmentation and registration of chest X-rays by training a GAN framework using NMI, SSIM, and a feature perceptual loss term.

### **3.2.2. Feature-based Unsupervised Transformation Estimation**

Liu et al. [35] proposed a tensor-based MIND method for both unimodal and multimodal registration in which a principle component analysis-based network (PCANet) [36] is utilised. Multimodal pairs of brain MR images and thoracic CT inhale–exhale pair volumes are used for validating this approach. Registration of 2D brain and cardiac MRs is performed in Krebs et al. [37] and a stochastic latent space learning approach is employed to bypass the need for spatial regularization. The loss function for this method is defined by the negative log likelihood of the fixed image given the latent representation and the warped volume and KL divergence of the latent distribution from a prior distribution.

**Table 3.2.** Exploration of related literatures (unsupervised deformable transformation estimation)

<b>Ref. No.</b>	<b>Author</b>	<b>Registration Class</b>	<b>Loss Function</b>	<b>Model</b>
[23]	de Vos et al., 2017	Monomodal (Cardiac cine; MR)	NCC	DIRNet
[7]	de Vos et al., 2017	Monomodal (Cardiac cine; MR)	NCC	DLIR
[24]	Neylon et al., 2017	Monomodal (Head/ Neck; CT)	Predicted TRE	FCN
[27]	Ghosal et al., 2017	Monomodal (Brain; MR)	UB SSD	19-Layer FCN
[35]	Liu et al., 2017	Monomodal (Chest; CT) Multimodal (Brain; MR)	SSD MIND + PCANet o/p	FCN, PCA Net
[6]	Li et al., 2018	Monomodal (Brain; MR)	NCC	8-Layer FCN
[28]	Stergios et al., 2018	Monomodal (Lung; MR)	MSE	CNN
[29]	Eppenhof et al., 2018	Monomodal (Cardiac cine; MR)	MSE	8-Layer FCNet
[30]	Ferrante et al., 2018	Monomodal (Bone; X-ray) Monomodal (Cardiac cine; MR)	NCC	U-Net, STN
[31]	Cao et al., 2018	Multimodal (Pelvis; CT, MR)	NCC	U-Net
[32]	Zhang, 2018	Monomodal (Brain; MR)	MSD	ICNet
[33]	Mahapatra, 2018	Monomodal (Cardiac; MR) Monomodal (Retinal; FA)	NMI + SSIM + VGG outputs	GAN
[34]	Mahapatra et al., 2018	Monomodal (Bone; X-ray)	NMI + SSIM + VGG o/ps + BCE	GAN
[37]	Krebs et al., 2018	Monomodal (Brain; MR)	Log likelihood	cVAE STN
[25]	Balakrishnan et al., 2019	Monomodal (Brain; MR)	CC, MSE	VoxelMorph

This chapter summarizes the recent research works on the process of transformation estimation using learning based approaches. The survey helps in understanding that each application of deep neural networks has its own unique challenges and thus the frameworks should be designed and validated very carefully. The concepts gained from the literature review have further been utilized to achieve the objectives of the thesis.

## References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
2. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.
3. Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., & Komodakis, N. (2016, October). A deep metric for multimodal registration. In International conference on medical image computing and computer-assisted intervention (pp. 10-18). Springer, Cham.
4. Cheng, X., Zhang, L., & Zheng, Y. (2018). Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3), 248-252.
5. Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., & Comaniciu, D. (2017, February). An artificial agent for robust image registration. In Thirty-First AAAI Conference on Artificial Intelligence.
6. Ma, K., Wang, J., Singh, V., Tamersoy, B., Chang, Y. J., Wimmer, A., & Chen, T. (2017, September). Multimodal image registration with deep context reinforcement learning. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 240-248). Springer, Cham.
7. de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., & Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52, 128-143.

8. Li, H., & Fan, Y. (2018, April). Non-rigid image registration using self-supervised fully convolutional networks without training data. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 1075-1078). IEEE.
9. Haskins, G., Kruger, U., & Yan, P. (2020). Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1), 8.
10. Yang, X., Kwitt, R., & Niethammer, M. (2016). Fast predictive image registration. In *Deep Learning and Data Labeling for Medical Applications* (pp. 48-57). Springer, Cham.
11. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
12. Rohé, M. M., Datar, M., Heimann, T., Sermesant, M., & Pennec, X. (2017, September). SVF-Net: Learning deformable image registration using shape matching. In *International conference on medical image computing and computer-assisted intervention* (pp. 266-274). Springer, Cham.
13. Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., & Shen, D. (2017, September). Deformable image registration based on similarity-steered CNN regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 300-308). Springer, Cham.
14. Lv, J., Yang, M., Zhang, J., & Wang, X. (2018). Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study. *The British journal of radiology*, 91(xxxx), 20170788.
15. Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B. P., Išgum, I., & Staring, M. (2017, September). Nonrigid image registration using multi-scale 3D convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 232-239). Springer, Cham.
16. Eppenhof, K. A., & Pluim, J. P. (2018). Pulmonary CT registration through supervised learning with convolutional neural networks. *IEEE transactions on medical imaging*, 38(5), 1097-1105.
17. Ito, M., & Ino, F. (2018, January). An Automated Method for Generating Training Sets for Deep Learning based Image Registration. In *BIOIMAGING* (pp. 140-147).
18. Fan, J., Cao, X., Yap, P. T., & Shen, D. (2019). BIRNet: Brain image registration using dual-supervised fully convolutional networks. *Medical image analysis*, 54, 193-206.

19. Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C. M., ... & Vercauteren, T. (2018, April). Label-driven weakly-supervised learning for multimodal deformable image registration. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 1070-1074). IEEE.
20. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., ... & Ourselin, S. (2018). Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis*, 49, 1-13.
21. Hu, Y., Modat, M., Gibson, E., Ghavami, N., Bonmati, E., Moore, C. M., ... & Vercauteren, T. (2018, April). Label-driven weakly-supervised learning for multimodal deformable image registration. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 1070-1074). IEEE.
22. Hering, A., Kuckertz, S., Heldmann, S., & Heinrich, M. P. (2019). Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking. In *Bildverarbeitung für die Medizin 2019* (pp. 309-314). Springer Vieweg, Wiesbaden.
23. de Vos, B. D., Berendsen, F. F., Viergever, M. A., Staring, M., & Išgum, I. (2017). End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 204-212). Springer, Cham.
24. Neylon, J., Min, Y., Low, D. A., & Santhanam, A. (2017). A neural network approach for fast, automated quantification of DIR performance. *Medical physics*, 44(8), 4126-4138.
25. Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., & Dalca, A. V. (2019). Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8), 1788-1800.
26. Dalca, A. V., Balakrishnan, G., Guttag, J., & Sabuncu, M. R. (2018, September). Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 729-738). Springer, Cham.
27. Ghosal, S., & Ray, N. (2017). Deep deformable registration: Enhancing accuracy by fully convolutional neural net. *Pattern Recognition Letters*, 94, 81-86.
28. Stergios, C., Mihir, S., Maria, V., Guillaume, C., Marie-Pierre, R., Stavroula, M., & Nikos, P. (2018). Linear and deformable image registration with 3d convolutional neural networks. In

Image Analysis for Moving Organ, Breast, and Thoracic Images (pp. 13-22). Springer, Cham.

29. Eppenhof, K. A., & Pluim, J. P. (2018). Error estimation of deformable image registration of pulmonary CT scans using convolutional neural networks. *Journal of medical imaging*, 5(2), 024003.
30. Ferrante, E., Oktay, O., Glocker, B., & Milone, D. H. (2018, September). On the adaptability of unsupervised CNN-based deformable image registration to unseen image domains. In *International Workshop on Machine Learning in Medical Imaging* (pp. 294-302). Springer, Cham.
31. Cao, X., Yang, J., Wang, L., Xue, Z., Wang, Q., & Shen, D. (2018, September). Deep learning based inter-modality image registration supervised by intra-modality similarity. In *International Workshop on Machine Learning in Medical Imaging* (pp. 55-63). Springer, Cham.
32. Zhang, J. (2018). Inverse-consistent deep networks for unsupervised deformable image registration. arXiv preprint arXiv:1809.03443.
33. Mahapatra, D. (2018). GAN Based Medical Image Registration. arXiv preprint arXiv:1805.02369.
34. Mahapatra, D., Ge, Z., Sedai, S., & Chakravorty, R. (2018, September). Joint registration and segmentation of xray images using generative adversarial networks. In *International Workshop on Machine Learning in Medical Imaging* (pp. 73-80). Springer, Cham.
35. Liu, Q., & Leung, H. (2017, July). Tensor-based descriptor for image registration via unsupervised network. In *2017 20th International Conference on Information Fusion (Fusion)* (pp. 1-7). IEEE.
36. Chan, T. H., Jia, K., Gao, S., Lu, J., Zeng, Z., & Ma, Y. (2015). PCANet: A simple deep learning baseline for image classification?. *IEEE transactions on image processing*, 24(12), 5017-5032.
37. Krebs, J., Mansi, T., Mailhé, B., Ayache, N., & Delingette, H. (2018). Learning structured deformations using diffeomorphic registration. arXiv preprint arXiv:1804.07172.