

# MvMM-RegNet: A new image registration framework based on multivariate mixture model and neural network estimation

Xinzhe Luo and Xiahai Zhuang <sup>\*</sup>

School of Data Science, Fudan University, Shanghai, China

**Abstract.** Current deep-learning-based registration algorithms often exploit intensity-based similarity measures as the loss function, where dense correspondence between a pair of moving and fixed images is optimized through backpropagation during training. However, intensity-based metrics can be misleading when the assumption of intensity class correspondence is violated, especially in cross-modality or contrast-enhanced images. Moreover, existing learning-based registration methods are predominantly applicable to pairwise registration and are rarely extended to groupwise registration or simultaneous registration with multiple images. In this paper, we propose a new image registration framework based on multivariate mixture model (MvMM) and neural network estimation. A generative model consolidating both appearance and anatomical information is established to derive a novel loss function capable of implementing groupwise registration. We highlight the versatility of the proposed framework for various applications on multimodal cardiac images, including single-atlas-based segmentation (SAS) via pairwise registration and multi-atlas segmentation (MAS) unified by groupwise registration. We evaluated performance on two publicly available datasets, i.e. MM-WHS-2017 and MS-CMRSeg-2019. The results show that the proposed framework achieved an average Dice score of  $0.871 \pm 0.025$  for whole-heart segmentation on MR images and  $0.783 \pm 0.082$  for myocardium segmentation on LGE MR images<sup>1</sup>.

## 1 Introduction

The purpose of image registration is to align images into a *common* coordinate space, where further medical image analysis can be conducted, including image-guided intervention, image fusion for treatment decision, and atlas-based segmentation [14]. In the last few decades, intensity-based registration has received considerable scholarly attention. Commonly used similarity measures comprise intensity difference and correlation-based methods for intra-modality registration, and information-theoretic metrics for inter-modality registration [10, 14, 18, 22, 23].

<sup>\*</sup> Corresponding author: zxh@fudan.edu.cn.

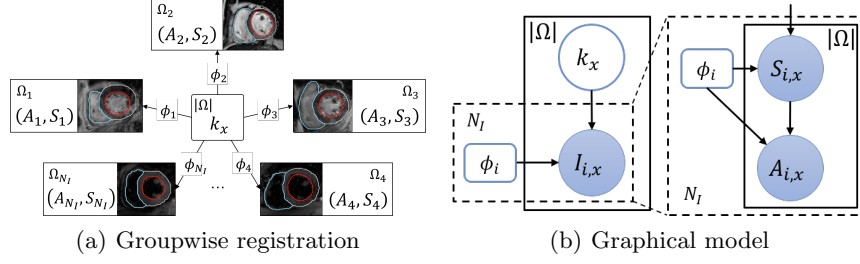
<sup>1</sup> Code is available from <https://zmiclab.github.io/projects.html>.

Recently, deep learning (DL) techniques have formulated registration as a parameterized mapping function, which not only made registration in one shot possible but achieved state-of-the-art accuracies [4, 8, 11, 24]. de Vos et al. [24] computed dense correspondence between two images by optimizing normalized cross-correlation between intensity pairs. While intensity-based similarity measures are widely used for intra-modality registration, there are circumstances when no robust metric, solely based on image appearance, can be applied. Hu et al. [11] therefore resorted to weak labels from corresponding anatomical structures and landmarks to predict voxel-level correspondence. Balakrishnan et al. [4] proposed leveraging both intensity- and segmentation-based metrics as loss functions for network optimization. More recently, Dalca et al. [8] developed a probabilistic generative model and derived a framework that could incorporate both of the intensity images and anatomical surfaces.

Meanwhile, in the literature several studies have suggested coupling registration with segmentation, in which image registration and tissue classification are performed simultaneously within the same model [2, 6, 20, 25]. However, the search for the optimal solution of these methods usually entails computationally expensive iterations and may suffer from problems of parameter tuning and local optimum. A recent study attempted to leverage registration to perform Bayesian segmentation on brain MRI with an unsupervised deep learning framework [9]. Nevertheless, it can still be difficult to apply unsupervised intensity-based approaches to inter-modality registration or to datasets with poor imaging quality and obscure intensity class correspondence. Besides, previous DL-integrated registration methods have mainly focused on pairwise registration and are rarely extended to groupwise registration or simultaneous registration with multiple images.

In this paper, we consider the scenario in which multiple images from various modalities need to be co-registered simultaneously onto a *common* coordinate space, which is set onto a reference subject or can be implicitly assumed during groupwise registration. To this end, we propose a probabilistic image registration framework based on both multivariate mixture model (MvMM) and neural network estimation, referred to as MvMM-RegNet. The model incorporates both types of information from the appearance and anatomy associated with each image subject, and explicitly models the correlation between them. A neural network is then employed to estimate likelihood and achieve efficient optimization of registration parameters. Besides, the framework provides posterior estimation for MAS on novel test images.

The main contribution of this work is four-fold. First, we extend the conventional MvMM for image registration with multiple subjects. Second, a DL-integrated groupwise registration framework is proposed, with a novel loss function derived from the probabilistic graphical model. Third, by modelling the relationship between appearance and anatomical information, our model outperforms previous ones in terms of segmentation accuracy on cardiac medical images. Finally, we investigate two applications of the proposed framework on cardiac image segmentation, i.e. SAS via pairwise registration and MAS unified



**Fig. 1.** (a) Groupwise registration framework, (b) Graphical representation of the proposed generative model, where random variables are in circles, deterministic parameters are in boxes, observed variables are shaded and plates indicate replication.

by groupwise registration, and achieve state-of-the-art results on two publicly available datasets.

## 2 Methods

Groupwise registration aims to align every subject in a population to a *common* coordinate space  $\Omega$  [5, 7], referred to as the *common space* [25]. Assume we have  $N_I$  *moving* subjects  $\mathbf{I} = \{I_i\}_{i=1}^{N_I}$ , of which each is defined on spatial domain  $\Omega_i$ . For each subject  $I_i$ , we can observe its appearance  $A_i$  from medical imaging as well as labels of anatomical structures  $S_i$  in various cases for image registration tasks. Thus, we can formulate  $I_i = (A_i, S_i)$  as a pair of appearance and anatomical observations for each subject.

Associated with the *moving* subjects is a set of spatial transforms  $\phi$  that map points from the common space to counterparts in each subject space:

$$\phi = \{\phi_i : y_i = \phi_i(x), i = 1, \dots, N_I\}, \quad (1)$$

where  $x \in \Omega$ ,  $y_i \in \Omega_i$ . The framework is demonstrated in Fig. 1(a).

### 2.1 Multivariate mixture model

The proposed method builds on a generative model of the appearance and anatomical information over a population of subjects. The likelihood function is computed as a similarity measure to drive the groupwise registration process.

For spatial coordinates in the common space, an exemplar atlas can be determined *a priori*, providing anatomical statistics of the population regardless of their corresponding appearances through medical imaging. For notational convenience, we denote tissue types using label values  $k_x$ , where  $k \in K$ ,  $K$  is the set of labels, with its prior distribution defined as  $\pi_{kx} = p(k_x)$ . Assuming *independence* of each location, the likelihood can be written as  $\mathcal{L}(\phi|\mathbf{I}) = \prod_{x \in \Omega} p(\mathbf{I}_x|\phi)$ . Moreover, by summing over all states of the hidden variable  $k_x$ , we have

$$\mathcal{L}(\phi|\mathbf{I}) = \prod_{x \in \Omega} \sum_{k \in K} \pi_{kx} p(\mathbf{I}_x|k_x, \phi). \quad (2)$$

Given the common-space anatomical structures, the multivariate mixture model assumes *conditional independence* of the moving subjects, namely

$$p(\mathbf{I}_x | k_x, \phi) = \prod_{i=1}^{N_I} p(I_{i,x} | k_x, \phi_i) = \prod_{i=1}^{N_I} p(I_{i,y_i} | k_x), \quad (3)$$

where  $I_{i,y_i}$  denotes a patch of observations centred at  $y_i = \phi_i(x)$ . Given anatomical structures of each subject, one can further assume its appearance is *conditional independent* of the groupwise anatomy, i.e.  $p(A_{i,y_i} | S_{i,y_i}, k_x) = p(A_{i,y_i} | S_{i,y_i})$ . Hence, we can further factorize the conditional probability into

$$p(I_{i,y_i} | k_x) = p(A_{i,y_i} | S_{i,y_i}) p(S_{i,y_i} | k_x). \quad (4)$$

Accordingly, the log-likelihood is given by

$$l(\phi | \mathbf{I}) = \sum_{x \in \Omega} \log \left\{ \sum_{k \in K} \pi_{kx} \prod_{i=1}^{N_I} p(A_{i,y_i} | S_{i,y_i}) p(S_{i,y_i} | k_x) \right\}. \quad (5)$$

In practice, we optimize the negative log-likelihood as a dissimilarity measure to obtain the desired spatial transforms  $\hat{\phi}$ . The graphical representation of the proposed model is shown in Fig. 1(b).

## 2.2 The conditional parameterization

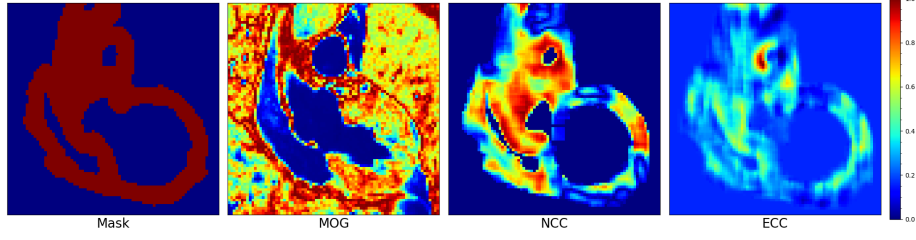
In this section, we specify in detail the conditional probability distributions (CPDs) for a joint distribution that factorizes according to the Bayesian network structure represented in Fig. 1(b).

**Spatial prior.** One way to define the common-space spatial prior is to average over a cohort of subjects [2], and the resulting probabilistic atlas is used as a reference. To avoid bias from a fixed reference and consider the population as a whole, we simply adopt a flat prior over the common space, i.e.  $\pi_{kx} = c_k, \forall x \in \Omega$  satisfying  $\sum_{k \in K} c_k = 1$ , where  $c_k$  is the weight to balance each tissue class.

**Label consistency.** Spatial alignment of a group of subjects can be measured by their label consistency, defined as the joint distribution of the anatomical information  $p(\mathbf{S}_{i,y_i}, k_x)$ , where  $\mathbf{S}_{i,y_i} = \{S_{i,y_i}\}_{i=1}^{N_I}$ . Each CPD  $p(S_{i,y_i} | k_x)$  gives the likelihood of the anatomical structure around a subject location being labelled as  $k_x$ , conditioned on the transform that maps from the common space to each subject space. We model it efficiently by a local Gaussian weighting:

$$p(S_{i,y_i} | k_x) = \sum_{z \in \mathcal{N}_{y_i}} w_z \cdot \delta(S_i(z) = k_x), \quad (6)$$

where  $\delta(\cdot)$  is the Kronecker delta function,  $\mathcal{N}_{y_i}$  defines a neighbourhood around  $y_i$  of radius  $r_s$  and  $w_z$  specifies the weight for each voxel within the neighbourhood. This formulation is equivalent to applying Gaussian filtering using an isotropic standard deviation  $\sigma_s$  to the segmentation mask [11], where we set  $r_s = 3\sigma_s$ .



**Fig. 2.** Visualization of different appearance models computed from a coronal view of a whole heart MR image subject at background areas, where "Mask", "MOG", "NCC" and "ECC" denote appearance model using ROI mask, mixture of Gaussians, normalized cross correlation and entropy cross correlation, respectively. For comparison, values are normalized to intervals between 0 and 1.

**Appearance model.** Finally, we seek to specify the term  $p(A_{i,y_i}|S_{i,y_i})$ . A common approach adopted by many tissue classification algorithms [2, 9, 13, 16, 25] is to model this CPD as a mixture of Gaussians (MOG), where intensities of the same tissue type should be clustered and voxel locations are assumed independent. Nevertheless, we hypothesize that using such an appearance model can mislead the image registration when the assumption of intensity class correspondence is violated, due to poor imaging quality, particularly in cross-modality or contrast enhanced images [19]. Let  $\nabla(\cdot)$  and  $\|\cdot\|$  be the voxel-wise gradient and Euclidean-norm operators, respectively. A vanilla means is to use a mask around the ROI boundaries:

$$p(A_{i,y_i}|S_{i,y_i}) = \begin{cases} 1 & \text{if } \exists z \in \mathcal{N}_{y_i} \text{ s.t. } \nabla\|S_i(z)\| > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

which ignores the appearance information. However, we argue that a reasonable CPD design should reflect fidelities of medical imaging and serve as a voxel-wise weighting factor for likelihood estimation. Thus, we formalize a CPD that 1) is defined with individual subjects, 2) is zero on voxels distant to the ROIs, 3) has increasing values at regions where appearance and anatomy have consistent rate of change. Therefore, we speculate that voxels with concordant gradient norms between appearance and anatomy are more contributory to determining the spatial correspondence. Based on these principles, one can estimate the CPD as a Gibbs distribution computed from an energy function or negative similarity measure between gradient-norm maps of appearance and anatomy, i.e.

$$p(A_{i,y_i}|S_{i,y_i}) = \frac{1}{Z} \exp \{-E(\|\nabla A_{i,y_i}\|, \|\nabla S_{i,y_i}\|)\}, \quad (8)$$

where  $Z$  is the normalization factor and  $E(\cdot)$  can be the negative normalized cross-correlation (NCC) [3] or negative entropy correlation coefficient (ECC) [17]. Fig. 2 visualises the different appearance models.

### 2.3 Neural network estimation

We formulate a neural network  $g_{\theta}(\cdot)$  parameterized by  $\theta$  that takes as input a group of  $N_I$  images to predict the deformation fields, based on a 3D UNet-style architecture designed for image registration [11]. To discourage non-smooth displacement, we resort to bending energy as a deformation regularization term and incorporate it into the loss function [11, 24]. Hence, the final loss function for network optimization becomes

$$Loss(\theta; \mathbf{I}) = -l(\phi_{\theta}|\mathbf{I}) + \lambda \cdot R(\phi_{\theta}), \quad (9)$$

where  $R(\cdot)$  denotes the deformation regularization term and  $\lambda$  is a regularization coefficient.

### 2.4 Applications

In this section, we present two applications from the proposed MvMM-RegNet framework, which are validated in our experiments.

**Pairwise MvMM-RegNet for SAS.** Pairwise registration can be considered as a specialization of groupwise registration where the number of subjects equals two and one of the spatial coordinate transforms is the identity mapping. We will demonstrate the registration capacity of our model by performing pairwise registration on a real clinical dataset, referred to as pMvMM-RegNet.

**Groupwise MvMM-RegNet for MAS.** During multi-atlas segmentation (MAS), multiple expert-annotated images with segmented labels, called atlases, are co-registered to a target space, where the warped atlas labels are combined by label fusion [12]. Delightfully, our model provides a unified framework for this procedure through groupwise registration, denoted as gMvMM-RegNet. By setting the *common space* onto the target as the *reference space*, we can derive the following segmentation formula:

$$\begin{aligned} \hat{S}_T(x) &= \operatorname{argmax}_{k \in K} p(k_x | \mathbf{I}_x, \phi) \\ &= \operatorname{argmax}_{k \in K} \left\{ \pi_{k_x} \prod_{i=1}^{N_I} p(I_{i,x} | k_x, \phi_i) \right\}. \end{aligned} \quad (10)$$

In practice, the MAS result with  $N_I \times t$  atlases can be generated from  $t$  times of groupwise registration over  $N_I$  subjects followed by label fusion using Eq. (10).

## 3 Experiments and Results

In this section, we investigate two applications of the proposed framework described in Section 2.4. In both of the two experiments, the neural networks

**Table 1.** Average substructure Dice and Hausdorff distance (HD) of MR-to-MR and CT-to-MR inter-subject registration, with \* indicating statistically significant improvement given by a Wilcoxon signed-rank test ( $p < 0.001$ ).

Methods	<i>MR-to-MR</i>		<i>CT-to-MR</i>	
	Dice	HD (mm)	Dice	HD (mm)
WeaklyReg	$0.834 \pm 0.031$	$17.45 \pm 2.482$	$0.842 \pm 0.033$	$17.99 \pm 2.681$
Baseline-MOG	$0.832 \pm 0.027$	$19.65 \pm 2.792$	$0.851 \pm 0.028^*$	$19.03 \pm 2.564$
Baseline-Mask	$0.840 \pm 0.028^*$	$16.91 \pm 2.374^*$	$0.851 \pm 0.032$	$17.51 \pm 2.687^*$
Baseline-ECC	$0.844 \pm 0.026^*$	$16.69 \pm 2.355^*$	$0.850 \pm 0.032$	$17.70 \pm 2.659$
Baseline-NCC	$0.847 \pm 0.028^*$	$16.83 \pm 2.422$	$0.850 \pm 0.032$	$17.78 \pm 2.721$
MVF-MvMM	Dice= $0.871 \pm 0.025$		HD (mm)= $17.21 \pm 4.408$	

were trained on a NVIDIA<sup>®</sup> RTX<sup>™</sup> 2080 Ti GPU with the spatial transformer module adapted from open-source code in VoxelMorph [4], implemented in TensorFlow [1]. The Adam optimizer was adopted [15], with a cyclical learning rate bouncing between  $1e-5$  and  $1e-4$  to accelerate convergence and avoid shallow local optima [21].

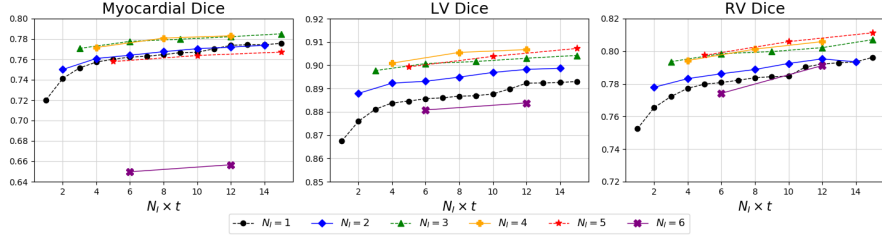
### 3.1 pMvMM-RegNet for SAS on whole heart MRI

**Materials and baseline.** This experiment was performed on the MM-WHS challenge dataset, which provides 120 multi-modality whole-heart images from multiple sites, including 60 cardiac CT and 60 cardiac MRI [26,27], of which 20 subjects from each of the modalities were selected as training data. Intra- (MR-to-MR) and inter-modality (CT-to-MR) but inter-subject registration tasks were explored on this dataset, resulting in 800 propagated labels in total for 40 test MR subjects.

An optimal weighting of bending energy could lead to a low registration error, when maintaining the global smoothness of the deformations. To be balanced, we set  $\lambda = 0.5$  as the default regularization strategy<sup>2</sup>. We analysed different variants of the appearance model described in Section 2.2, i.e. "MOG", "Mask", "NCC" and "ECC", and compared with a reimplementation of [11], known as "WeaklyReg", which exploited the Dice similarity metric for weakly-supervised registration. In addition, with the propagated labels obtained from pairwise registrations, we evaluated the performance of MAS by applying a simple majority vote to the results, denoted as "MVF-MvMM".

**Results and discussion.** Table 1 presents the Dice statistics of both intra- and inter-modality registration tasks on the MM-WHS dataset. With increasingly plausible modelling of the relationship between appearance and anatomy, we have observed better registration accuracy especially for MR images, indicating efficacy of the proposed framework. Fusing labels by majority vote

<sup>2</sup> See Fig. 1 in the supplementary material for an empirical result.



**Fig. 3.** Dice scores of MAS results using  $N_I \times t$  atlases, where  $N_I$  denotes the number of subjects used in each groupwise registration and  $t$  counts the number of groupwise registrations performed before label fusion.

(“MVF-MvMM”) can produce a better segmentation accuracy, reaching an average Dice score of  $0.871 \pm 0.025^3$ , comparable to the inter-observer variability of  $0.878 \pm 0.014$  reported in [26].

### 3.2 gMvMM-RegNet for MAS on LGE CMR

**Materials and baseline.** In this experiment, we explored MAS with the application of Eq. (10) on MS-CMRSeg challenge dataset [25]. The dataset consists of 45 patients scanned using three CMR sequences, i.e. the LGE, T2 and bFFSP, from which 20 patients were chosen in random for training, 5 for validation and 20 for testing. We implemented inter-subject and inter-modality groupwise registration and evaluated the MAS results on LGE CMR images.

A 2D version of the network architecture described in Section 2.3 was devised to jointly predict the deformation fields for  $N_I$  atlases by optimizing Eq. (9). The MAS result was generated by  $t$  times of groupwise registration over  $N_I$  randomly sampled subjects followed by label fusion using Eq. (10).

**Results and discussion.** The comparison between SAS and MAS highlights that more accurate and realistic segmentation is generated by groupwise registration than pairwise registration, especially for apical and basal slices<sup>4</sup>. Fig. 3 further reports the mean Dice scores for each cardiac substructure obtained from MAS using  $t$  times of groupwise registration with  $N_I$  subjects. With a fixed total number of atlases, label fusion on 2D slices resulting from groupwise registration outperforms those from conventional pairwise registration, reaching the average myocardium Dice score of  $0.783 \pm 0.082$ . However, we also observe decline in accuracy when having a large number of subjects ( $N_I \geq 5$ ) to be groupwise registered. This discrepancy could be attributed to the lack of network parameters compromising the predicted deformations.

<sup>3</sup> See Fig.2 in the supplementary material for evaluation statistics on all cardiac substructure.

<sup>4</sup> See Fig. 3 in the supplementary material for visualization of the segmentation results.



## 4 Conclusion

In this work, we propose a probabilistic image registration framework based on multivariate mixture model and neural network estimation, coupling group-wise registration and multi-atlas segmentation in a unified fashion. We have evaluated two applications of the proposed model, i.e. SAS via pairwise registration and MAS unified by groupwise registration, on two publicly available cardiac image datasets and compared with state-of-the-art methods. The proposed appearance model along with MvMM has shown its efficacy in realizing registration on cardiac medical images characterizing inferior intensity class correspondence. Our method has also proved its superiority over conventional pairwise registration algorithms in terms of segmentation accuracy, highlighting the advantage of groupwise registration as a subroutine to MAS.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China (grant no. 61971142).

## References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I.J., Harp, A., Irving, G., Isard, M., Jia, Y., Józefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D.G., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P.A., Vanhoucke, V., Vasudevan, V., Viégas, F.B., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *ArXiv abs/1603.04467* (2015)
2. Ashburner, J., Friston, K.J.: Unified segmentation. *NeuroImage* **26**, 839–851 (2005)
3. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* **12** **1**, 26–41 (2008)
4. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Gutttag, J.V., Dalca, A.V.: Voxel-morph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging* **38**, 1788–1800 (2019)
5. Balci, S.K., Golland, P., Shenton, M.E., Wells, W.M.: Free-form b-spline deformation model for groupwise registration. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* **10** **WS**, 23–30 (2007)
6. Bhatia, K.K., Aljabar, P., Boardman, J.P., Srinivasan, L., Murgasova, M., Counsell, S.J., Rutherford, M.A., Hajnal, J.V., Edwards, A.D., Rueckert, D.: Groupwise combined segmentation and registration for atlas construction. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* **10** **Pt 1**, 532–40 (2007)

7. Bhatia, K.K., Hajnal, J.V., Hammers, A., Rueckert, D.: Similarity metrics for groupwise non-rigid registration. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* **10 Pt 2**, 544–52 (2007)
8. Dalca, A.V., Balakrishnan, G., Guttag, J.V., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis* **57**, 226–236 (2019)
9. Dalca, A.V., Yu, E.M., Golland, P., Fischl, B., Sabuncu, M.R., Iglesias, J.E.: Unsupervised deep learning for bayesian brain mri segmentation. In: *MICCAI* (2019)
10. Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. *Physics in medicine and biology* **46 3**, R1–45 (2001)
11. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., Ourselin, S., Noble, J.A., Barratt, D.C., Vercauteren, T.: Weakly-supervised convolutional neural networks for multimodal image registration. *Medical Image Analysis* **49**, 1–13 (2018)
12. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: A survey. *Medical image analysis* **24 1**, 205–219 (2014)
13. Iglesias, J.E., Sabuncu, M.R., Leemput, K.V.: A unified framework for cross-modality multi-atlas segmentation of brain mri. *Medical Image Analysis* **17**, 1181–1191 (2013)
14. Khalil, A., Ng, S.C., Liew, Y.M., Lai, K.W.: An overview on image registration techniques for cardiac diagnosis and treatment. *Cardiology research and practice* (2018)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR abs/1412.6980* (2014)
16. Leemput, K.V., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of mr images of the brain. *IEEE Transactions on Medical Imaging* **18**, 897–908 (1999)
17. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging* **16**, 187–198 (1997)
18. Mäkelä, T., Clarysse, P., Sipilä, O., Pauna, N., Pham, Q.C., Katila, T., Magnin, I.E.: A review of cardiac image registration methods. *IEEE Transactions on Medical Imaging* **21**, 1011–1021 (2002)
19. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging* **22**, 986–1004 (2003)
20. Pohl, K.M., Fisher, J.W., Grimson, W.E.L., Kikinis, R., Wells, W.M.: A bayesian model for joint segmentation and registration. *NeuroImage* **31**, 228–239 (2006)
21. Smith, L.N.: Cyclical learning rates for training neural networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* pp. 464–472 (2015)
22. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging* **32**, 1153–1190 (2013)
23. Viergever, M.A., Maintz, J.B.A., Klein, S., Murphy, K., Staring, M., Pluim, J.P.W.: A survey of medical image registration - under review. *Medical image analysis* **33**, 140–144 (2016)
24. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis* **52**, 128143 (2018)

25. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**, 2933–2946 (2019)
26. Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M.P., Oster, J., Wang, C., Smedby, Ö., Bian, C., Yang, X., Heng, P.A., Mortazi, A., Bagci, U., Yang, G., Sun, C., Galisot, G., Ramel, J.Y., Brouard, T., Tong, Q., Si, W., Liao, X., Zeng, G., Shi, Z., Zheng, G., Wang, C., MacGillivray, T., Newby, D.E., Rhode, K.S., Ourselin, S., Mohiaddin, R., Keegan, J., Firmin, D.N., Yang, G.: Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. In: *Medical Image Analysis* (2019)
27. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. *Medical image analysis* **31**, 77–87 (2016)