GMM-CoRegNet: A Multimodal Groupwise Registration Framework Based on Gaussian Mixture Model

Zhenyu Li[1], Fan Yu[2,3], Jie Lu[2,3], and Zhen Qian[1,*]

[1] Institute of Intelligent Diagnostics, Beijing United-Imaging Research Institute of Intelligent Imaging, Beijing, China, zhen.qian@cri-united-imaging.com
[2] Department of Radiology and Nuclear Medicine, Xuanwu Hospital, Capital Medical University, Beijing, China
[3] Beijing Key Laboratory of Magnetic Resonance Imaging and Brain Informatics, Beijing, China

Abstract. Within-subject multimodal groupwise registration aims to align a group of multimodal images into a common structural space. Existing groupwise registration methods often rely on intensity-based similarity measures, but can be computationally expensive for large sets of images. Some methods build statistical relationships between image intensities and anatomical structures, which may be misleading when the assumption of consistent intensity-class correspondences do not hold. Additionally, these methods can be unstable in batch group registration when the number of anatomical structures varies across different image groups. To tackle these issues, we propose GMM-CoRegNet, a weakly supervised deep learning framework for within-subject multimodal groupwise registration. A prior Gaussian Mixture Model (GMM) consolidating the image intensities and anatomical structures is constructed using the label of reference image, then we derive a novel similarity measure for groupwise registration based on GMM and iteratively optimize the GMM throughout the training process. Notably, GMM-CoRegNet can register an arbitrary number of images simultaneously to a reference image only needing the label of reference image. We compared GMM-CoRegNet with state-of-the-art groupwise registration methods on two carotid datasets and the public BrainWeb dataset, demonstrated its superior registration performance even for the registration scenario of inconsistent intensity-class mappings.

Keywords: groupwise registration · multimodal · GMM.

1 Introduction

For a group of multimodal medical images corresponding to a subject, variations in image acquisition parameters or patient movement during the capture process often result in distortions or deformations of the anatomical structures. Consequently, the implementation of image registration methods becomes imperative

to align multimodal images into a common structural space. Previous registration methods have predominantly focused on pairwise registration [19,5,4,13,3], which align the moving images separately to a fixed image based on the similarity measure. For instance, Mutual Information (MI) is an information-theoretic measure widely employed in cross-modal image registration, particularly successful in assessing the statistical dependence between two images [11,18].

Due to the curse of dimensionality, extending the registration method to N images for $N \gg 2$ can be challenging. Existing groupwise registration methods usually estimate the desired spatial transformations by intensity-based similarity measures [9,14,21,16,10]. The accumulated pairwise estimates (APE) framework computes the sum of all pairwise mutual information (MI) from the warped images as similarity measure [21], but has a heavy computational burden grows quadratically with N. In [10], $\mathcal{X}$-CoReg introduces $\mathcal{X}$-Metric, which models the statistical dependencies between the appearance of multimodal images and their shared anatomical structures. Additionally, it expands a deep learning framework to concurrently execute image registration and segmentation tasks. However, certain structures may be visible only in a few specific modalities, and intensity-class mappings between image pixels and anatomical structures are not always one-to-one, where two pixels with the same intensity in one modality image may correspond to different structures and different intensities in other modalities. In such scenario, the effectiveness of MI or $\mathcal{X}$-Metric can't be guaranteed.

It is necessary to combine multimodal complementary information using a directly multimodal similarity measure guiding the deformation of each moving image. Orchard and Mann [14] achieve groupwise registration by modeling a multivariate Gaussian Mixture Model (GMM) with a fixed number of Gaussian components K for each group of images. Nevertheless, the number of anatomical structures in each subject may not be consistent, limiting its applicability in batch groupwise registration. In addition, small anatomical structures are susceptible to significant initial misalignment, the GMM constructed based on a single group of images may fail to achieve registration convergence.

In this paper, we focus on modeling the similarity measure based on a multivariate GMM and construct a deep learning groupwise registration framework. Different from [14], we model the global multivariate Gaussian distributions corresponding to different anatomical structures by all training subjects, utilizing the label of reference image in each subject as weakly supervised information. Then the GMM constructed from specific structure numbers $K^{'}$ for each group of images can serve as prior information, guiding the groupwise registration process. The main contributions of this work can be summarized as follows:

1) We introduce GMM-CoRegNet, a weakly supervised deep learning framework for within-subject multimodal groupwise registration. This framework can register an arbitrary number of images simultaneously to a reference image and achieve weak supervision by only requiring the label of the reference image.

2) The GMM corresponding to anatomical structures is modeled using the label of reference image as prior information. We derive a novel similarity measure

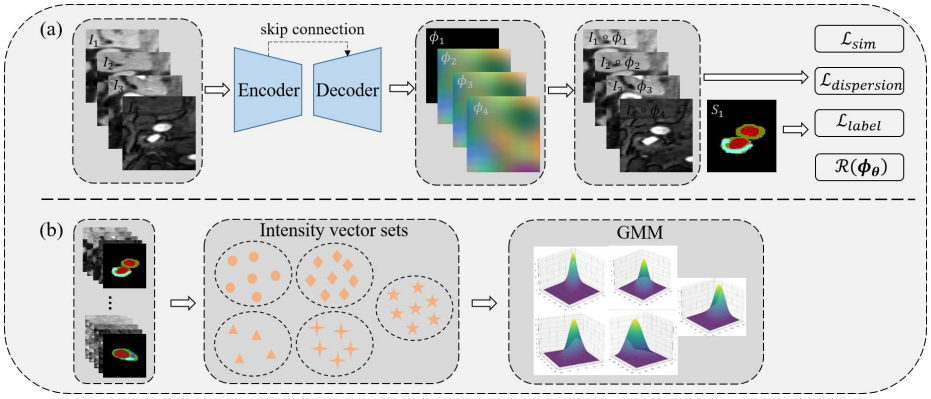

Fig. 1: (a) Groupwise registration framework of our method when $N = 4$. (b) Schematic diagram of the process of constructing the prior Gaussian Mixture Model when $K = 5$. For ease of visualization, each N-dimensional Gaussian distribution is shown in two dimensional space.

for groupwise registration based on GMM and iteratively optimize the GMM throughout the training process.

3) We demonstrate the superiority of GMM-CoRegNet over previous SOTA methods on two carotid datasets and the publicly available BrainWeb dataset.

2 Method

Assuming the training set has M subjects, let $\boldsymbol{I} = \{I_i \mid_{i=1,\dots,N}\}$ be the group of N multimodal images of the same subject. Groupwise registration aims to align images $\boldsymbol{I}$ to a common space Ω. In this paper, treating each group of images as a whole, we align $N-1$ images to a reference image, with $\Omega_{ref} = \Omega$. Assuming that within each subject, only the reference image is accompanied by a label image denoted as S_{ref}. Each location $x \in \Omega$ corresponds to a label value k, $k \in K$, K is the set of labels. Our method models prior GMM leveraging S_{ref} as weak supervision and proposes a similarity measure, helping to derive spatial transformation fields $\boldsymbol{\phi} = \{\phi_i \mid_{i=1,\dots,N}\}$ that map all moving images to the reference image.

Specifically, we designate T1 as the reference image in each $\boldsymbol{I}$, thus defining the common space as Ω_1, with ϕ_1 equals to zero. Detailed implementations are given in the following sections.

2.1 Gaussian Mixture Model

For groupwise registration, given a set of N multimodal images observed, the pixel at location x has N intensity values associated with it. We can represent $I_i = (v_{\omega,i})_{\omega \in \Omega_i}$, then the resampled intensity vector can be denoted as $\boldsymbol{v}_x^{\boldsymbol{\phi}}$, i.e.

$\boldsymbol{v}_x^{\boldsymbol{\phi}} = \left[v_{x,1}^{\phi_1}, ..., v_{x,N}^{\phi_N}\right]$, where $v_{x,i}^{\phi_i} = I_i \circ \phi_i(x)$ represents the pixel value of I_i at location $x \in \Omega$ after applying the spatial transformation with parameters ϕ_i.

GMM Definition. The use of GMM to capture the probability density function of image data is a well established approach [6,2]. Inspired by the work in [14] and under the assumption of GMM [12], our method treats each anatomical structure as a component k of the GMM follows multivariate Gaussian distribution. Then the probability density function of $\boldsymbol{v}_x^{\boldsymbol{\phi}}$ can be expressed as:

$$p(\boldsymbol{v}_x^{\boldsymbol{\phi}}) = \textstyle\sum_{k \in K} \pi_k \mathcal{N}(\boldsymbol{v}_x^{\boldsymbol{\phi}} \mid \mu_k, \Sigma_k) \tag{1}$$

$$\mathcal{N}(\boldsymbol{v}_x^{\boldsymbol{\phi}} \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{(\boldsymbol{v}_x^{\boldsymbol{\phi}} - \mu_k)^T \Sigma_k^{-1} (\boldsymbol{v}_x^{\boldsymbol{\phi}} - \mu_k)}{2}} \tag{2}$$

where μ_k and Σ_k are the mean and covariance matrix of component k in the feature space, and $\pi_k = p(k)$ is the prior probability of class k. Thus the distribution $p(\boldsymbol{v}_x^{\boldsymbol{\phi}} \mid k)$ denotes the conditional prior probability distribution of an intensity vector $\boldsymbol{v}_x^{\boldsymbol{\phi}}$ categorized to the component k, which can be expressed as:

$$p(\boldsymbol{v}_x^{\boldsymbol{\phi}} \mid k) = \mathcal{N}(\boldsymbol{v}_x^{\boldsymbol{\phi}} \mid \mu_k, \Sigma_k) \tag{3}$$

Parameters Calculation. Let $\Omega^k \in \Omega$ represent the coordinate space where structure k is located, utilizing S_{ref} of each subject, the intensity vector set corresponding to structure k can be extracted as $\boldsymbol{U}_j^k = \boldsymbol{v}_{x^k}^{\boldsymbol{\phi}}$, where $x^k \in \Omega^k$, and $j \in M$. As not every subject has K anatomical structures, let $K^{'}$ as the set of labels corresponding to specific subject, thus each subject can obtain the intensity vector sets based on $k \in K^{'}$.

As shown in Fig. 1(b), to avoid bias from a fixed subject and consider the population of train set as a whole to get GMM as prior information, we merge the intensity vector sets extracted from the M training subjects into K overall intensity vector sets. The $\boldsymbol{U}^k = \left[\boldsymbol{v}_{x^k,1}^{\boldsymbol{\phi}}, ..., \boldsymbol{v}_{x^k,n_k}^{\boldsymbol{\phi}}\right]^{\mathrm{T}}$ corresponding to each component k is represented as an array of size $[n_k, N]$, where n_k denotes the number of the pixels corresponding to structure k across the M training subjects. Then the parameters of each Gaussian component k can be calculated by:

$$\pi_k = \frac{n_k}{\sum_{k \in K} n_k}, \mu_k = \frac{1}{n_k} \textstyle\sum_{i=1}^{n_k} \boldsymbol{v}_{x^k,i}^{\boldsymbol{\phi}}, \Sigma_k = \sum_{i=1}^{n_k} (\boldsymbol{v}_{x^k,i}^{\boldsymbol{\phi}} - \mu_k)(\boldsymbol{v}_{x^k,i}^{\boldsymbol{\phi}} - \mu_k)^T \tag{4}$$

Before performing the groupwise registration, deformation fields $\boldsymbol{\phi}$ are equal to zero, thus the initialization parameters for each multivariate Gaussian distribution can be directly computed using the undeformed training dataset. As $\boldsymbol{\phi}$ iteratively update, we update the parameters in Equation (4) using the warped training set images every 10 training epochs.

2.2 GMM-based similarity measure

Given the label image S_{ref} in each subject as weakly supervised information, the specific number of intensity vectors corresponding to structure k can be

determined as $n_{k,j}$, where $k \in K^{'}$, $j \in M$ and $\sum_{k \in K'} n_{k,j} = n_j$. Then the label proportion for each subject can be defined as $\pi_{k,j} = \frac{n_{k,j}}{n_j}$.

Let $(\boldsymbol{v}_{x_j}^{\boldsymbol{\phi}})_{x_j \in \Omega}$ represents as the intensity vector of subject j, according to the Bayesian formula, the posterior probability distribution can be written as:

$$p(k \mid \boldsymbol{v}_{x_j}^{\boldsymbol{\phi}}) = \frac{p(\boldsymbol{v}_{x_j}^{\boldsymbol{\phi}} \mid k)p(k)}{p(\boldsymbol{v}_{x_j}^{\boldsymbol{\phi}})} = \frac{\mathcal{N}(\boldsymbol{v}_{x_j}^{\boldsymbol{\phi}} \mid \mu_k, \Sigma_k)\pi_{k,j}}{\sum_{k \in K'} \mathcal{N}(\boldsymbol{v}_{x_j}^{\boldsymbol{\phi}} \mid \mu_k, \Sigma_k)\pi_{k,j}} \tag{5}$$

Therefore, our similarity measure based on GMM can be derived using the cross-entropy between the one-hot label S_{ref} and the negative logarithm of the posterior probability function:

$$\mathcal{L}_{sim} = -\sum_{k \in K'} \sum_{x_j \in \Omega} s_j log(p(k \mid \boldsymbol{v}_{x_j}^{\boldsymbol{\phi}})) = -\sum_{k \in K'} \sum_{x_{jk} \in \Omega^k} log(p(k \mid \boldsymbol{v}_{x_{jk}}^{\boldsymbol{\phi}})) \tag{6}$$

where s_j is the sign function equal to 1 if $x_{jk} \in \Omega^k$ otherwise 0. Therefore, we partition the common space Ω into $K^{'}$ anatomical structure regions and accomplish groupwise registration by aligning all component regions through the similarity function in Equation (6).

2.3 Neural network estimation

As shown in Fig. 1(a), we design a neural network framework parameterized by $\boldsymbol{\theta}$ to predict the deformation fields $\boldsymbol{\phi}$ for a set of N multimodal images as input. The neural network adopts the backbone architecture as described in VoxelMorph [3]. Specifically, multimodal images are concatenated as a single N-channel image input, and the output channel of the last convolutional layer of VoxelMorph is modified to obtain $N-1$ deformation fields. To guarantee the smoothness of the deformation fields, we employ the bending energy $\mathcal{R}(\boldsymbol{\phi_\theta})$ as a regularization term for deformation and incorporate it into the loss function.

2.4 Auxiliary loss function

As a group of images gradually optimizes alignment, intensity vectors corresponding to the same structure will become more concentrated in the feature space [20]. We measure the dispersion of the feature space by calculating the variance of the intensity vector set corresponding to each anatomical structure:

$$\mathcal{L}_{dispersion} = \sum_{k \in K'} \frac{1}{n_{k,j}} \sum_{x_{jk} \in \Omega^k} (\boldsymbol{v}_{x_{jk}}^{\boldsymbol{\phi}} - \mu_{k,j})^2 \tag{7}$$

where $\mu_{k,j} = \frac{1}{n_{k,j}} \sum_{x_{jk} \in \Omega^k} \boldsymbol{v}_{x_{jk}}^{\boldsymbol{\phi}}$ represents for the mean of intensity vector set corresponding to label k of subject j.

Furthermore, using Equation (5), the label value for location $x \in \Omega$ can be predicted by the maximum probability. We further design a regularization function computing the neighborhood similarity of the predicted label map S_{pred}:

$$\mathcal{L}_{label} = \sum_{x \in \Omega} \frac{1}{n} \sum_{i \in n} f(s_x, s_{x_i}) \tag{8}$$

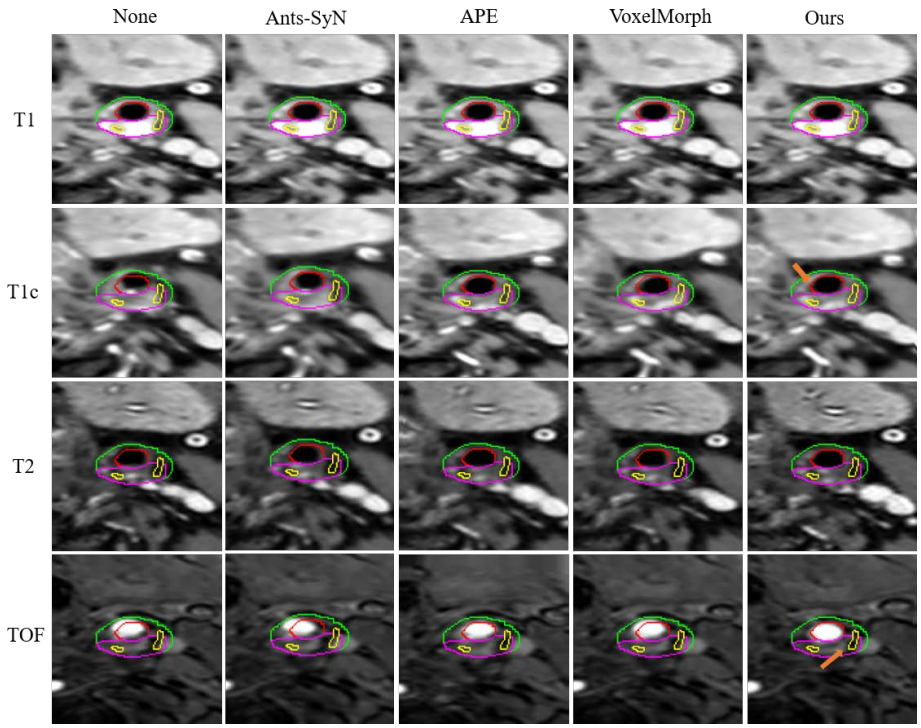

Fig. 2: Example images (with segmentation contours of T1 overlaid) before and after registration by compared methods on the proprietary carotid dataset. Orange arrows indicate regions where our method outperforms baseline methods.

where s_x represents the predicted label value of S_{pred} at location x, x_i refers to the neighborhood of x, $f(s_x, s_{x_i})$ equals to 0 if $s_x = s_{x_i}$ otherwise 1. Specifically, for 3D images and 2D images, n denotes the six-neighborhood and four-neighborhood, respectively. Hence, the final loss function of our method becomes:

$$\mathcal{L}_{total} = \mathcal{L}_{sim} + \gamma_1 \mathcal{L}_{dispersion} + \gamma_2 \mathcal{L}_{label} + \gamma_3 \mathcal{R}(\boldsymbol{\phi_\theta}) \tag{9}$$

γ_1, γ_2, γ_3 indicate the coefficient of $\mathcal{L}_{dispersion}$, $\mathcal{L}_{label}$ and $\mathcal{R}(\boldsymbol{\phi_\theta})$, respectively.

3 Experiments and Results

3.1 Datasets and Preprocessing

Proprietary Carotid Dataset. This dataset includes misaligned 3D T1, T1c, T2, and TOF carotid MR scans from our collaborative hospital. Experienced physicians annotated the label image on the T1 scan, identifying vascular lumen (VL), calcification (Cal), lipid (Lip), hemorrhage (Hem) and the rest area of the vascular wall(VM). The dataset was randomly divided into training/validation/test

Table 1: Quantitative evaluation results of compared methods on two carotid artery datasets evaluated by DSC (%). The bolded numbers denote the highest scores.

Methods	Carotid simulation dataset						Proprietary carotid dataset					
	Reg DSC (%)						Seg DSC (%)					
	Avg	VL	VM	Cal	Lip	Hem	Avg	VL	VM	Cal	Lip	Hem
None	70.18	83.52	71.55	44.97	58.89	73.97	54.21	81.10	61.61	27.22	28.86	49.29
Ants-SyN [1]	77.67	87.52	78.58	56.42	64.02	81.49	59.77	83.19	63.76	33.78	32.09	53.48
APE [21]	78.11	**88.77**	76.34	58.33	66.34	82.64	60.27	83.83	64.28	34.44	30.32	54.39
VoxelMorph [3]	76.71	87.47	77.80	53.35	60.28	81.30	60.55	84.67	64.61	32.77	33.77	52.61
$\mathcal{X}$-CoReg [10]	75.22	87.54	77.79	50.67	63.31	78.47	-	-	-	-	-	-
Ours	**79.41**	88.06	**80.20**	**60.29**	**67.52**	**85.45**	**62.88**	**85.69**	**68.07**	**36.71**	**36.09**	**55.70**

sets with 160/25/45 subjects. All volumes were resampled to $1 \times 1 \times 1\ mm^3$, cropped to $128 \times 128 \times 32$, and normalized to the range of $[0, 1]$.

Carotid Simulation Dataset. Leveraging clinical literature [7,17] and our proprietary carotid dataset, we utilized the pre-processed label image of each subject to assign specific intensity values to anatomical structure regions across the four modalities. This aimed to create simulated multimodal images that closely resembled the actual intensity distributions of the structures. The division ratio of this dataset was the same as that of the proprietary carotid dataset.

BrainWeb Dataset.[4] This dataset offers simulated T1, T2, and PD MR scans. Adhering to the procedures detailed in [10], we preprocessed the dataset by selecting the middle slice from each image, cropping to a size of 192×160, and normalizing to the range of $[0, 1]$. Anatomical structures encompass cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM). Further, FFDs were applied to introduce misalignments. The preprocessed dataset was subsequently partitioned into 252/36/72 for training/validation/test set.

3.2 Experimental Setups

Compared Methods. Two iterative methods Ants-SyN [1], APE [21] and two deep learning models VoxelMorph [3], $\mathcal{X}$-CoReg [10] were compared with our GMM-CoRegNet. For Ants-SyN, we utilize MI as similarity measure and align the moving images separately to the reference image. For VoxelMorph, we extend its framework to groupwise registration (the same as introduced in section 2.3), and calculate the sum of MI between each pair of moving image and reference image as similarity measure. Regarding $\mathcal{X}$-CoReg, its deep learning framework with weakly supervised strategy was employed, utilizing only the label of T1.

Implementation Details. GMM-CoRegNet was implemented using PyTorch [15] on an NVIDIA A800 GPU. Training utilized the Adam optimizer [8] with a learning rate of 1e-4 and a batch size of 1 for 500 epochs. γ_1, γ_2, and γ_3 were set to 2, 1, 100 for two carotid datasets and 5, 1, 100 for the BrainWeb dataset. we set K to 6 and 4 for the carotid simulation dataset and the BrainWeb dataset,

[4] `https://brainweb.bic.mni.mcgill.ca/`

respectively, for both $\mathcal{X}$-CoReg and GMM-CoRegNet. In the proprietary carotid dataset without background suppression, K was specifically set to 5.

Evaluation Metrics. We computed the Dice similarity coefficient (DSC) averaged over all pairwise combinations of labels for registered images on the carotid simulation dataset and BrainWeb dataset. Given that the proprietary carotid dataset has labels only for the T1 modality, we assumed that a higher alignment degree of multimodal images would lead to more accurate segmentation results. After aligning the proprietary carotid dataset using various registration methods, we maintained the same dataset partition ratios and employed TransBTS [22] as the segmentation algorithm. Subsequently, the DSC of the segmentation results on the test set was calculated as the evaluation metric.

3.3 Results

Comparison with SOTA. Table 1 presents registration accuracy on two carotid datasets. For carotid simulation dataset, our method yielded optimal groupwise registration results for 4 of the 5 anatomical structures, except the DSC on the vascular lumen was slightly lower than APE. Remarkably, $\mathcal{X}$-CoReg exhibited suboptimal registration performance on carotid plaques, which can be attributed to the non-uniform intensity-class mappings between the anatomical structure and appearance. For instance, calcifications only exhibit distinguishable intensity distribution in the TOF modality. As for proprietary carotid dataset, $\mathcal{X}$-CoReg is not suitable for this experiment due to unablely constructing the accurate statistical relationship between single background label and complex background appearance without background suppression. Our method achieved the highest segmentation improvement across all anatomical structures, demonstrating excellent registration performance on clinical carotid artery images even without background suppression. In addition, experiments on the BrainWeb dataset demonstrated that our method worked consistently better than compared methods for multimodal brain images registration with superior DSC of 76.02. [5] Fig. 2 shows the example of qualitative comparison on the proprietary carotid dataset, demonstrating our method could achieve better alignment for both large-scale anatomy and local regions in most anatomical structures.

Ablation Study. We analyzed the effectiveness of each auxiliary loss function by excluding one of them from GMM-CoRegNet on the carotid simulation dataset(average DSC of 78.23/78.45 for excluding $\mathcal{L}_{dispersion}/\mathcal{L}_{label}$, respectively)[6]. The results indicated that both auxiliary loss functions contribute to the registration performance improvement. Even with the removal of one auxiliary loss function, the performance of our method remains competitive with the best-performing baseline model (highest average DSC for APE: 0.7811).

[5] See Table.1 in the supplementary material for results on all brain substructure.

[6] See Table.2 in the supplementary material for detailed ablation study results.

4 Conclusion

In this study, we introduced a deep learning framework based on GMM for multimodal groupwise registration. This framework leverages the label of the reference image as weak supervision and incorporates the prior distribution of GMM to design a multimodal similarity measure, guiding the groupwise registration process. The experiments on two carotid datasets and Brainweb dataset demonstrated that our framework not only achieved better registration performance compared to the baseline methods in multimodal images registration with inconsistent intensity mappings, but also applied effectively to conventional groupwise registration tasks, such as multimodal brain images registration.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China under Grant No. 82130058.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Avants, B.B., Tustison, N., Song, G., et al.: Advanced normalization tools (ants). Insight j **2**(365), 1–35 (2009)
2. Balafar, M.: Gaussian mixture model based segmentation methods for brain mri images. Artificial Intelligence Review **41**, 429–439 (2014)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
4. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. pp. 729–738. Springer (2018)
5. 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**, 128–143 (2019)
6. Greenspan, H., Ruf, A., Goldberger, J.: Constrained gaussian mixture model framework for automatic segmentation of mr brain images. IEEE transactions on medical imaging **25**(9), 1233–1245 (2006)
7. Kassem, M., Florea, A., Mottaghy, F.M., van Oostenbrugge, R., Kooi, M.E.: Magnetic resonance imaging of carotid plaques: current status and clinical perspectives. Annals of translational medicine **8**(19) (2020)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2), 236–250 (2005)
10. Luo, X., Zhuang, X.: $\mathcal{X}$-metric: An n-dimensional information-theoretic framework for groupwise registration and deep combined computing. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
11. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE transactions on Medical Imaging **16**(2), 187–198 (1997)
12. McLachlan, G.J., Basford, K.E.: Mixture models: Inference and applications to clustering, vol. 38. M. Dekker New York (1988)
13. Mok, T.C., Chung, A.: Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4644–4653 (2020)
14. Orchard, J., Mann, R.: Registering a multisensor ensemble of images. IEEE Transactions on Image Processing **19**(5), 1236–1247 (2009)
15. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
16. Polfliet, M., Klein, S., Huizinga, W., Paulides, M.M., Niessen, W.J., Vandemeulebroucke, J.: Intrasubject multimodal groupwise registration with the conditional template entropy. Medical image analysis **46**, 15–25 (2018)
17. Saba, L., Cau, R., Murgia, A., Nicolaides, A.N., Wintermark, M., Castillo, M., Staub, D., Kakkos, S.K., Yang, Q., Paraskevas, K.I., et al.: Carotid plaque-rads: a novel stroke risk classification system. Cardiovascular Imaging **17**(1), 62–75 (2024)

18. Snaauw, G., Sasdelli, M., Maicas, G., Lau, S., Verjans, J., Jenkinson, M., Carneiro, G.: Mutual information neural estimation for unsupervised multi-modal registration of brain images. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 3510–3513. IEEE (2022)
19. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE transactions on medical imaging **32**(7), 1153–1190 (2013)
20. Spiclin, Ž., Likar, B., Pernus, F.: Groupwise registration of multimodal images by an efficient joint entropy minimization scheme. IEEE Transactions on Image Processing **21**(5), 2546–2558 (2012)
21. Wachinger, C., Navab, N.: Simultaneous registration of multiple images: similarity metrics and efficient optimization. IEEE transactions on pattern analysis and machine intelligence **35**(5), 1221–1233 (2012)
22. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J.: Transbts: Multimodal brain tumor segmentation using transformer. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 109–119. Springer (2021)