A novel Mean Teacher framework for domain adaptive lung registration

Anonymous authors

Anonymous institute
****@****

Abstract. Recent deep learning-based registration models achieve excellent results but require plenty of labeled training data and suffer from domain shifts between training and test data. As a remedy, we present a novel method for domain adaptive image registration. We propose to reduce the domain shift through self-ensembling and embed a keypoint-based registration model into the Mean Teacher paradigm. We extend the Mean Teacher to the registration problem by 1) adapting the stochastic augmentation scheme and 2) combining learned feature extraction with differentiable optimization. This enables us to guide the learning process in the unlabeled target domain by enforcing consistent predictions of the learning student and the temporally averaged teacher model. We evaluate the method for exhale-to-inhale lung CT registration under two challenging adaptation scenarios. Our method consistently improves on the baseline model by 44 %/47 % while even matching the accuracy of models trained on target data. Source code is available at https://anonymous.4open.science/r/reg-da-mean-teacher.

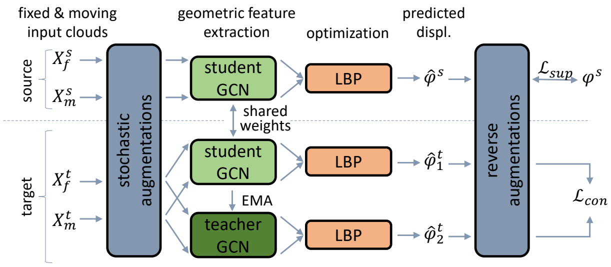
Keywords: Registration · Domain adaptation · Mean Teacher.

1 Introduction

Recent deep learning-based registration methods achieve excellent results but require costly labeled training data and often generalize poorly to shifted domains. Domain adaptation [9] can overcome these problems by adapting a model from a labeled source to an unlabeled target domain. Domain adaptive registration, however, has rarely been studied and remains a challenging problem. In particular, popular techniques like the alignment of global feature vectors [4] are insufficient for registration, which requires the identification of local correspondences. We propose to address domain adaptive registration through self-ensembling [3] and embed a registration model into the Mean Teacher paradigm [8]. We adjust the Mean Teacher framework accordingly and present the first Mean Teacher that combines learned feature extraction with differentiable optimization.

2 Methods

We build our method on a keypoint-based model [5], which implements image



Method	COPD →L2R	4D CT →COPD
initial	10.24	11.99
VM++ source	4.34	3.55
VM++ target	3.09	2.46
source-only	4.96	4.02
target-only	2.61	2.26
ours	2.64	2.01

Fig. 1. Left: Overview of our method. Right: Results for both adaptation scenarios, reported as target registration error in mm at available landmark correspondences.

registration as the geometric alignment of Foerstner keypoints from fixed and moving input scans. The model combines a graph convolutional network (GCN) for learned geometric feature extraction with differentiable loopy belief propagation (LBP) for alignment. To adapt the model to a shifted target domain, we propose a novel self-ensembling framework based on the Mean Teacher paradigm (see Fig. 1, left). The framework includes two GCNs—a student and a teacher GCN with weights $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. While the teacher's weights are the exponential moving average (EMA) of the student's weights, the student is optimized by minimizing the loss function

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{\theta}',\mathcal{S},\mathcal{T}) = \mathcal{L}_{sup}(\boldsymbol{\theta};\mathcal{S}) + \mathcal{L}_{con}(\boldsymbol{\theta};\boldsymbol{\theta}',\mathcal{T}) \quad (1)$$

It includes a supervised loss $\mathcal{L}_{sup}$ on labeled source data $\mathcal{S}$ and a consistency loss $\mathcal{L}_{con}$ on unlabeled target data $\mathcal{T}$. The latter encourages consistent predictions by teacher and student stream for different augmentations of the same input pair. Our implementation of $\mathcal{L}_{con}$ differs from the standard Mean Teacher in two decisive aspects. First, we incorporate inverse geometric transformations into the augmentation scheme to align predictions in both streams. Second, we do not impose consistency at the output of the learning network but after LBP. That way, the adaptation process can benefit from the regularizing effect of LBP.

3 Experiments and conclusion

Setup. We evaluate our method for exhale-to-inhale lung CT registration under two adaptation scenarios. 1) We use the DIR-Lab COPD dataset [1] (10 labeled scan pairs) as source and the Learn2Reg (L2R) Task 2 dataset [7] as target domain (12 unlabeled pairs for training, official test split of 10 scan pairs for evaluation). The domain shift consists in exhale scans from the target domain exhibiting a cropped field of view such that upper and lower parts of the lungs are partially cut off. 2) We adapt from the DIR-Lab 4D CT dataset [2] (10 labeled pairs) as the source to the COPD dataset as the target domain (5-fold cross-validation over 10 scan pairs). Here, the domain gap consists in different breathing types—shallow resting breathing (4D CT) vs actively forced full inhalation and exhalation (COPD).

Baselines. We train the baseline model without adaptation techniques on source data only as lower bound and on labeled target data only as upper bound. This is repeated for Voxelmorph++ [6] as an intensity-based baseline model.

Results are shown in Fig. 1 (right) and reveal consistent findings under both adaptation scenarios. First, significant gaps between source-only and target-only models demonstrate the severity of the domain gaps. Second, our proposed method effectively adapts the baseline model to the target domain, improving on the source-only model by 47 %/44 % while matching or even surpassing the performance of the target-only model.

In conclusion, our work reveals great potential of the Mean Teacher framework for learning-based image registration and its capability to improve feature learning in the absence of labels. Thus, the Mean Teacher could become a pivotal tool to overcome the prevailing scarcity of manual annotations.

References

1. Castillo, R., Castillo, E., Fuentes, D., Ahmad, M., Wood, A.M., Ludwig, M.S., Guerrero, T.: A reference dataset for deformable image registration spatial accuracy evaluation using the copdgene study archive. Physics in Medicine & Biology **58**(9), 2861 (2013)
2. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in Medicine & Biology **54**(7), 1849 (2009)
3. French, G., Mackiewicz, M., Fisher, M.: Self-ensembling for visual domain adaptation. In: International Conference on Learning Representations (2018)
4. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. pp. 1180–1189. PMLR (2015)
5. Hansen, L., Heinrich, M.P.: Deep learning based geometric registration for medical images: How accurate can we get without visual features? In: International Conference on Information Processing in Medical Imaging. pp. 18–30. Springer (2021)
6. Heinrich, M.P., Hansen, L.: Voxelmorph++ going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation. arXiv preprint arXiv:2203.00046 (2022)
7. Hering, A., Murphy, K., van Ginneken, B.: Learn2reg challenge: Ct lung registration - training data (May 2020). https://doi.org/10.5281/zenodo.3835682, https://doi.org/10.5281/zenodo.3835682
8. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
9. Wang, M., Deng, W.: Deep visual domain adaptation: A survey. Neurocomputing **312**, 135–153 (2018)