A Denoised Mean Teacher for Domain Adaptive Point Cloud Registration

Alexander Bigalke$^{(\boxtimes)}$ and Mattias P. Heinrich

Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
{alexander.bigalke,mattias.heinrich}@uni-luebeck.de

Abstract. Point cloud-based medical registration promises increased computational efficiency, robustness to intensity shifts, and anonymity preservation but is limited by the inefficacy of unsupervised learning with similarity metrics. Supervised training on synthetic deformations is an alternative but, in turn, suffers from the domain gap to the real domain. In this work, we aim to tackle this gap through domain adaptation. Self-training with the Mean Teacher is an established approach to this problem but is impaired by the inherent noise of the pseudo labels from the teacher. As a remedy, we present a denoised teacher-student paradigm for point cloud registration, comprising two complementary denoising strategies. First, we propose to filter pseudo labels based on the Chamfer distances of teacher and student registrations, thus preventing detrimental supervision by the teacher. Second, we make the teacher dynamically synthesize novel training pairs with noise-free labels by warping its moving inputs with the predicted deformations. Evaluation is performed for inhale-to-exhale registration of lung vessel trees on the public PVT dataset under two domain shifts. Our method surpasses the baseline Mean Teacher by 13.5/62.8%, consistently outperforms diverse competitors, and sets a new state-of-the-art accuracy (TRE = 2.31 mm). Code is available at https://github.com/multimodallearning/denoised_mt_pcd_reg.

Keywords: Point cloud registration · Domain adaptation · Mean Teacher

1 Introduction

Recent deep learning-based registration methods have shown great potential in solving medical image registration problems [10,14]. Most of these methods perform the registration based on the raw volumetric intensity images, e.g. [1,5,6,18,31]. By contrast, only a few recent works [13,21] operate on sparse, purely geometric point clouds extracted from the images, even though this representation promises multiple potential benefits, including computational efficiency, robustness against intensity shifts in the image domain, and anonymity

preservation. The latter, for instance, can facilitate public data access and federated learning, as exemplified by a recently released point cloud dataset of lung vessels [21] whose underlying CT scans are not publicly accessible. On the other hand, the sparsity of point clouds and the absence of intensity information make the registration problem more challenging. In particular, unsupervised learning with similarity metrics – as established for dense image registration [5,18] – was shown ineffective for deformable point cloud registration [21], as confirmed by our experiments. Since manual annotations for supervised learning are prohibitively costly, an alternative consists of training on synthetic deformations with known displacements [21], as known from dense registration [7,26]. The inevitable domain gap between synthetic and real deformations, however, involves the risk of suboptimal performance on real data. In this work, we aim to bridge this gap through domain adaptation (DA).

DA has widely been studied for classification and segmentation tasks [12], with popular techniques ranging from adversarial feature [11,25] or output [24] alignment to self-supervised feature learning [22]. However, these methods are insufficient for the specific characteristics of the registration problem, involving a more complex output space and requiring the detection of local correspondences. Instead, recent works adapted the Mean Teacher paradigm [23], previously established for domain adaptive classification [9] and segmentation [19], to the registration problem [3,16,29]. The basic idea is to supervise the learning student model with displacement fields (pseudo labels) provided by a teacher model, whose weights represent the exponential moving average of the student's weights. A significant limitation of this method, however, is the inevitable noise in the pseudo labels, potentially misguiding the adaptation process. Prior works addressed this problem by refining the pseudo labels [16] or weighting them according to model uncertainty, estimated through Monte Carlo dropout [29,30]. However, even refined pseudo labels remain inaccurate, and the proposed refinement strategy [16] assumes piecewise rigid motions of 3D objects and does not apply to complex deformations in medical applications. And weighting pseudo labels according to teacher uncertainty [29,30] does not explicitly consider the quality of the actual registrations, completely ignores the quality and certainty of the current student predictions, and can, therefore, not prevent detrimental supervision of the student through inferior teacher predictions.

Contributions. We introduce two complementary strategies to denoise the Mean Teacher for domain adaptive point cloud registration, addressing the above limitations (see Fig. 1). Both strategies are based on our understanding of an optimal student-teacher relationship. First, if the student's solution to a problem is superior to that of the teacher, good teachers should not insist on their solution but accept the student's approach. To implement this, inspired by a recent technique to filter pseudo labels for human pose estimation [2], we propose to assess the quality of both the teacher and student registrations with the Chamfer distance and to provide only those registrations of the teacher as supervision to the student that are more accurate. This approach differs from previous uncertainty-based methods [29,30] in two decisive aspects: 1) It explicitly assesses the quality
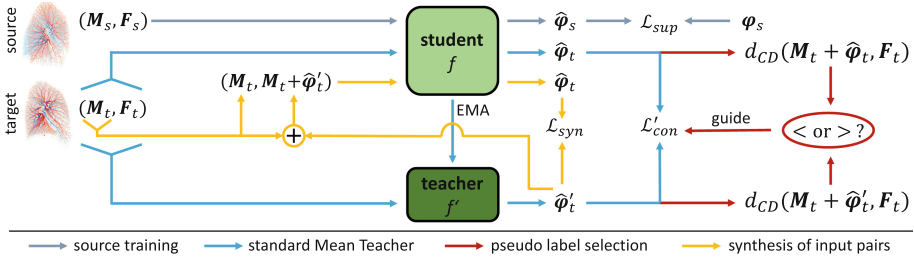
Fig. 1. Overview of our denoised Mean Teacher for domain adaptive registration. We overcome noisy supervision by the teacher with a novel pseudo label selection strategy and the synthesis of new training pairs with precisely known displacements.

of final registrations, using a model-free and objective measure with little computational overhead compared to multiple forward passes in Monte Carlo dropout. 2) The selection process considers both teacher and student predictions and can thus prevent detrimental supervision by the teacher. Our second strategy follows the intuition that good teachers should not pose problems to which they do not know the solution. Instead, they should come up with novel tasks with precisely known solutions. Consequently, we propose a completely novel teacher paradigm, where predicted deformations by the teacher are used to synthesize new training pairs for the student, consisting of the original moving inputs and their warps. These input pairs come with precise noise-free displacement labels and significantly differ from static hand-crafted synthetic deformations [21]. 1) The deformations are based on a real data pair that the teacher aims to align. 2) The deformations are dynamic and become more realistic as the teacher improves. Finally, we unify both strategies in a joint framework for domain adaptive point cloud registration. It is compatible with arbitrary geometric registration models, stable to train, and involves only a few hyper-parameters. We experimentally evaluate the method for inhale-to-exhale registration of lung vessel point clouds on the public PVT dataset [21], demonstrating substantial improvements over diverse competing methods and state-of-the-art performance.

2 Methods

2.1 Problem Setup and Standard Mean Teacher

In point cloud registration, we are given fixed and moving point clouds $F \in \mathbb{R}^{N_F \times 3}, M \in \mathbb{R}^{N_M \times 3}$ and aim to predict a displacement vector field $\varphi \in \mathbb{R}^{N_M \times 3}$ that spatially aligns M to F as $M + \varphi$. We address the task in a domain adaptation setting with training data comprising a labeled source dataset $\mathcal{S}$ of triplets (M_s, F_s, φ_s) and a shifted unlabeled target dataset $\mathcal{T}$ of tuples (M_t, F_t). While the formulation of our method is agnostic to the specific domain shift between $\mathcal{S}$ and $\mathcal{T}$, in this work, we generate the source samples on the

fly as random synthetic deformations of the target clouds using a fixed hand-crafted deformation function $def : \mathbb{R}^{N \times 3} \to \mathbb{R}^{N \times 3}$, i.e. source triplets are given as $(def(\boldsymbol{F}_t), \boldsymbol{F}_t, \boldsymbol{F}_t - def(\boldsymbol{F}_t))$ or $(def(\boldsymbol{M}_t), \boldsymbol{M}_t, \boldsymbol{M}_t - def(\boldsymbol{M}_t))$. Note that def preserves point correspondences enabling ground truth computation through point-wise subtraction. Given the training data, we aim to learn a function f that predicts deformation vector fields as $\hat{\boldsymbol{\varphi}} = f(\boldsymbol{M}, \boldsymbol{F})$ with optimal performance in the target domain.

Baseline Mean Teacher. To solve the problem, the standard Mean Teacher framework [3,9,23] employs two identical networks, denoted as the student f and teacher f', with parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. While the student's weights $\boldsymbol{\theta}$ are optimized through gradient descent, the teacher's weights correspond to the exponential moving average (EMA) of the student and are updated as $\boldsymbol{\theta}'_i = \alpha \boldsymbol{\theta}'_{i-1} + (1-\alpha) \boldsymbol{\theta}_i$ at iteration i with momentum α. Meanwhile, the student is trained by minimizing

$$\mathcal{L}(\boldsymbol{\theta}) = \lambda_1 \underbrace{\|f(\boldsymbol{M}_s, \boldsymbol{F}_s) - \boldsymbol{\varphi}_s\|_2^2}_{\mathcal{L}_{\text{sup}}} + \lambda_2 \underbrace{\|f(\boldsymbol{M}_t, \boldsymbol{F}_t) - f'(\boldsymbol{M}_t, \boldsymbol{F}_t)\|_2^2}_{\mathcal{L}_{\text{con}}} \qquad (1)$$

consisting of the supervised loss $\mathcal{L}_{\text{sup}}$ on source data and the consistency loss $\mathcal{L}_{\text{con}}$ on target data, weighted by λ_1 and λ_2. $\mathcal{L}_{\text{con}}$ guides the learning of the student in the target domain with pseudo-supervision from the teacher, which, as a temporal ensemble, is expected to be superior to the student. Nonetheless, predictions by the teacher can still be noisy and inaccurate, limiting the efficacy of the adaptation process.

2.2 Chamfer Distance-Based Filtering of Pseudo Labels

In a worst-case scenario, the student might predict an accurate displacement field $\hat{\boldsymbol{\varphi}}_t$, which is strongly penalized by the consistency loss due to an inaccurate teacher prediction $\hat{\boldsymbol{\varphi}}'_t$. To prevent such detrimental supervision, we aim to select only those teacher predictions for supervision that are superior to the corresponding student predictions, which, however, is complicated by the absence of ground truth. We, therefore, propose to assess the quality of student and teacher registrations by measuring the similarity/distance between fixed and warped moving clouds, with higher similarities/lower distances indicating more accurate registrations. Among existing similarity measures, we opt for the symmetric Chamfer distance [28], which computes the distance between two point clouds $\boldsymbol{X}, \boldsymbol{Y}$ as

$$d_{\text{CD}}(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{\boldsymbol{x} \in \boldsymbol{X}} \min_{\boldsymbol{y} \in \boldsymbol{Y}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \sum_{\boldsymbol{y} \in \boldsymbol{Y}} \min_{\boldsymbol{x} \in \boldsymbol{X}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \qquad (2)$$

While we experimentally found the Chamfer distance insufficient as a direct loss function – presumably due to sparse differentiability and susceptibility to local minima, we still observed a strong correlation between Chamfer distance and actual registration error, making it a suitable choice for our purposes. We also explored other measures (Laplacian curvature [28], Gaussian MMD [8]), which

proved slightly inferior (Supp., Table 2). Formally, we thus measure the quality of the student prediction $\hat{\boldsymbol{\varphi}}_t = f(\boldsymbol{M}_t, \boldsymbol{F}_t)$ as $d_{\mathrm{CD}}(\boldsymbol{M}_t + \hat{\boldsymbol{\varphi}}_t, \boldsymbol{F}_t)$ and analogously for the teacher prediction $\hat{\boldsymbol{\varphi}}_t'$. We then define our indicator function

$$I(\hat{\boldsymbol{\varphi}}_t, \hat{\boldsymbol{\varphi}}_t') = \begin{cases} 1 & d_{\mathrm{CD}}(\boldsymbol{M}_t + \hat{\boldsymbol{\varphi}}_t', \boldsymbol{F}_t) < d_{\mathrm{CD}}(\boldsymbol{M}_t + \hat{\boldsymbol{\varphi}}_t, \boldsymbol{F}_t) \\ 0 & \text{else} \end{cases} \tag{3}$$

and reformulate the consistency loss in Eq. 1 as

$$\mathcal{L}_{\mathrm{con}}' = I(\hat{\boldsymbol{\varphi}}_t, \hat{\boldsymbol{\varphi}}_t') \cdot \|f(\boldsymbol{M}_t, \boldsymbol{F}_t) - f'(\boldsymbol{M}_t, \boldsymbol{F}_t)\|_2^2 \tag{4}$$

2.3   Synthesizing Inputs with Noise-Free Supervision

While the above filtering strategy mitigates detrimental supervision, the selected pseudo labels are still inaccurate. Therefore, we complement the strategy with a novel teacher paradigm, where the teacher dynamically synthesizes new training pairs with precisely known displacements for supervision. Specifically, given a teacher prediction $\hat{\boldsymbol{\varphi}}_t' = f'(\boldsymbol{M}_t, \boldsymbol{F}_t)$, we do not only use it to supervise the student on the same input pair but also generate a new input sample $(\boldsymbol{M}_t, \boldsymbol{M}_t + \hat{\boldsymbol{\varphi}}_t')$ by warping $\boldsymbol{M}_t$ with $\hat{\boldsymbol{\varphi}}_t'$. The underlying displacement field is naturally precisely known, enabling noise-free training of the student by minimizing

$$\mathcal{L}_{\mathrm{syn}} = \|f(\boldsymbol{M}_t, \boldsymbol{M}_t + \hat{\boldsymbol{\varphi}}_t') - \hat{\boldsymbol{\varphi}}_t'\|_2^2 \tag{5}$$

To our knowledge, there is no prior work with a similarly "generative" teacher model. Altogether, we train the student network by minimizing the loss

$$\mathcal{L}(\boldsymbol{\theta}) = \lambda_1 \mathcal{L}_{\mathrm{sup}} + \lambda_2 \mathcal{L}_{\mathrm{con}}' + \lambda_3 \mathcal{L}_{\mathrm{syn}} \tag{6}$$

Technical Details. The synthesized input pairs $(\boldsymbol{M}_t, \boldsymbol{M}_t + \hat{\boldsymbol{\varphi}}_t')$ exhibit exact point correspondence, i.e., for each point in $\boldsymbol{M}_t$ exists a corresponding point in $\boldsymbol{M}_t + \hat{\boldsymbol{\varphi}}_t'$. That is usually not the case for real data pairs and thus introduces another domain shift, which prevented proper convergence in our initial experiments. To overcome the problem, we exploit that the original point clouds in a dataset, denoted as $\boldsymbol{M}_{t*}$, usually comprise more points than the subsampled clouds $\boldsymbol{M}_t$ that are fed to the network. Given predicted displacements $\hat{\boldsymbol{\varphi}}_t'$ for $\boldsymbol{M}_t$, we interpolate the displacement vectors to $\boldsymbol{M}_{t*}$ with an isotropic Gaussian kernel, yielding $\hat{\boldsymbol{\varphi}}_{t*}'$. The final input pair is then obtained by sampling disjoint point subsets from $(\boldsymbol{M}_{t*}, \boldsymbol{M}_{t*} + \hat{\boldsymbol{\varphi}}_{t*}')$, excluding one-to-one correspondences.

3   Experiments

3.1   Experimental Setup

Datasets. We evaluate our method for inhale-to-exhale registration of lung vessel point clouds on the public PVT dataset [21] (https://github.com/uncbiag/

robot, License: CC BY-NC-SA 3.0). The dataset comprises 1,010 such data pairs, which were extracted from lung CT scans as part of the IRB-approved COPDGene study (NCT00608764). Ten of these scan pairs are cases from the Dirlab-COPDGene dataset [4] and thus annotated with 300 landmark correspondences. We use these cases as the test set and split the remaining unlabeled pairs into 800 cases for training and 200 for validation (on synthetic deformations only). The original point clouds in the dataset have a very high resolution (∼100k points), making the processing with deep networks computationally costly. Therefore, we extract distinctive keypoints by local density estimation followed by non-maximum suppression. We extract two sets of such keypoints for each cloud: one with the ∼8k most distinctive points for inference, and another with ∼16k points, from which we randomly sample subsets during training for increased variability (see Sect. 2.3, technical details). Finally, we pre-align each pair by matching the mean and standard deviation of the coordinates.

Implementation Details. The registration network f is implemented as the default 4-scale architecture of PointPWC-Net [28], operating on 8192 points per cloud. Following [28], we implement $\mathcal{L}_{\mathrm{sup}}$, $\mathcal{L}'_{\mathrm{con}}$, and $\mathcal{L}_{\mathrm{syn}}$ as multi-scale losses. Optimization is performed with the Adam optimizer. We first pre-train the network on source data (batch size 4) for 160 epochs and subsequently minimize the joint loss (Eq. 6) for 140 epochs, both with a constant learning rate of 0.001, which requires up to 11 GB and 13/23 h on an RTX2080. For joint optimization, we use mixed batches of 4 source and 4 target samples, set $\lambda_1 = \lambda_2 = \lambda_3 = 10$, and the EMA-parameter to $\alpha = 0.996$. While the original PVT data pairs represent the target domain in all experiments, we consider two variants of the function def to synthesize source data pairs: a realistic task-specific 2-scale random field similar to [21] and a simple rigid transformation. This enables us to evaluate our method under two differently severe domain shifts. Since real validation data are unavailable, hyper-parameters of all compared methods were tuned in a synthetic adaptation scenario, with the rigid deformations in the source and the 2-scale random field deformations in the target domain. For further implementation details, we refer to our public code.

Comparison Methods. 1) The source-only model is exclusively trained on source data without DA. 2) We adopt the standard Mean Teacher [3]. 3) An uncertainty-aware Mean Teacher (UA-MT), similar to [29,30]. 4) As proposed in [28], we performed purely unsupervised training on target data with a Chamfer loss. However, consistent with the findings in [21], this approach could not converge for complex geometric lung structures. Instead, we use the Chamfer loss on target data as an additional loss to complement supervised source training. 5) We guide the learning on target data with the cycle-consistency method from [17]. 6) As a classical algorithm, we adapt sLBP [13]. 7) We collect the results of two current SOTA methods, S-Robot and D-Robot, from [21], which combine deep networks (Point U-Net, PointPWC-Net), trained on synthetic deformations, with optimal transport modules. Note, however, that the experimental setup in

Table 1. Quantitative results on the PVT dataset, reported as mean TRE and 25/75% percentiles in mm and SDlogJ. † indicates a deviating experimental setup (Sect. 3.1).

| Method	def = 2-scale rnd. field				def = rigid			
	TRE	25%	75%	SDlogJ	TRE	25%	75%	SDlogJ
initial	23.32	13.22	31.61	–	23.32	13.22	31.61	–
pre-align	12.83	8.25	16.68	–	12.83	8.25	16.68	–
sLBP [13]	3.62	1.24	3.29	0.038	3.62	1.24	3.29	0.038
S-Robot† [21]	5.48	2.86	7.14	N/A	N/A	N/A	N/A	N/A
D-Robot† [21]	2.86	1.25	3.11	N/A	N/A	N/A	N/A	N/A
source-only	4.50	1.62	5.49	0.034	7.62	3.12	11.15	0.019
Chamfer loss [28]	3.96	1.47	4.43	0.036	4.18	1.54	5.27	0.043
cycle-consistency [17]	3.93	1.48	4.36	0.035	6.47	2.43	9.08	0.029
Mean Teacher [3]	2.67	1.33	3.12	**0.028**	6.40	2.42	9.50	**0.013**
UA-MT [29]	2.58	1.28	3.04	0.029	5.71	1.83	8.77	0.015
ours w/o $\mathcal{L}_{\mathrm{syn}}$	2.49	1.23	2.88	0.030	2.57	1.22	2.93	0.027
ours w/o $\mathcal{L}'_{\mathrm{con}}$	2.96	1.27	3.29	0.035	3.00	1.21	3.39	0.034
ours	**2.31**	**1.16**	**2.66**	0.034	**2.38**	**1.12**	**2.66**	0.033

[21] slightly differs from our setting in terms of more input points (60k vs. 8k) and additional input features (vessel radii), thus accessing more information.

Metrics. We interpolate the predicted displacements from the moving input cloud to the annotated moving landmarks with an isotropic Gaussian kernel ($\sigma = 5\,\mathrm{mm}$) and measure the target registration error (TRE) with respect to the fixed landmarks. To assess the smoothness of the predictions, we interpolate the sparse displacement fields to the underlying image grid and measure the standard deviation of the logarithm of the Jacobian determinant (SDlogJ).

3.2   Results

Quantitative results are shown in Table 1 and reveal the following insights: 1) The source-only model benefits from realistic synthetic deformations in the source domain, yielding a 40.9% lower TRE. 2) The standard Mean Teacher proves effective under the weaker domain shift (-40.7% TRE compared to source-only) but only achieves a slight improvement of 16.0% in the more challenging scenario, where pseudo labels by the teacher are naturally noisier, in turn limiting the efficacy of the adaptation process. 3) Our proposed strategy to filter pseudo labels (ours w/o $\mathcal{L}_{\mathrm{syn}}$) improves the standard teacher and its uncertainty-aware extension, particularly notable under the more severe domain shift ($-59.8/-55.0\%$ TRE). 4) Synthesizing novel data pairs with the teacher (ours w/o $\mathcal{L}'_{\mathrm{con}}$) alone is
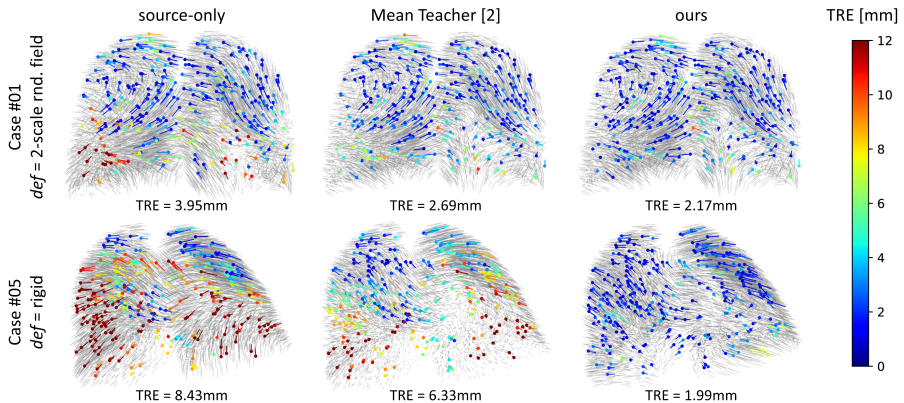
Fig. 2. Qualitative results on two sample cases of the PVT dataset. Predicted displacement fields are shown in gray. Colored dots and lines represent moving landmarks and their interpolated flow, with colors encoding the TRE (clamped to 12 mm). (Color figure online)

slightly inferior to the standard teacher for realistic deformations in the source domain but substantially superior for simple rigid transformations. 5) Combining our two strategies yields further considerable improvements to TREs of 2.31 and 2.38 mm, demonstrating their complementarity. Thus, our method improves the standard Mean Teacher by 13.5/62.8%, outperforms all competitors by statistically significant margins ($p < 0.001$ in a Wilcoxon signed-rank test), and sets a new state-of-the-art accuracy. Remarkably, our method achieves almost the same accuracy for simple rigid transformations in the source domain as for complex, realistic deformations. Thus, it eliminates the need for designing task-specific deformation models, which requires strong domain knowledge. Qualitative results are presented in Fig. 2 and Supp., Fig. 3, demonstrating accurate and smooth deformation fields by our method, as confirmed by the SDlogJ in Table 1, which takes small values for all methods.

4   Conclusion

Our work addressed domain adaptive point cloud registration to bridge the gap between synthetic source and real target deformations. Starting from the established Mean Teacher paradigm, we presented two novel strategies to tackle the noise of pseudo labels from the teacher model, which is a persistent, significant limitation of the method. Specifically, we 1) proposed to prevent detrimental supervision through the teacher by filtering pseudo labels according to Chamfer distances of student and teacher registrations and 2) introduced a novel teacher-student paradigm, where the teacher synthesizes novel training data pairs with perfect noise-free displacement labels. Our experiments for lung vessel registration on the PVT dataset demonstrated the efficacy of our method under two

scenarios, outperforming the standard Mean Teacher by up to 62.8% and setting a new state-of-the-art accuracy (TRE = 2.31 mm). As such, our method even favorably compares to popular image-based deep learning methods (VoxelMorph [1] and LapIRN [18], e.g., achieve TREs of 7.98 and 4.99 mm on the original DIR-Lab CT images) but lags behind conventional image-based optimization methods [20] with 0.83 mm TRE. But while the latter require run times of several minutes to process the dense intensity scans with 30M+ voxels, our method processes sparse, purely geometric point clouds with 8k points only, enabling anonymity-preservation and extremely fast inference within 0.2 s. In this light, we see two significant potential impacts of our work: First, our method could generally advance purely geometric keypoint-based medical registration, previously limited by the inefficacy of unsupervised learning with similarity metrics. In particular, medical point cloud registration, currently primarily focusing on lung anatomies, still needs to be investigated for other anatomical structures (abdomen, brain) in future work, which might benefit from our generic approach. Second, our method is conceptionally transferable to dense image registration (e.g., intensity-based similarity metrics [15,27] can replace the Chamfer distance). In this context, it appears of great interest to revisit learning from synthetic deformations [7] within a DA setting or to combine our method with unsupervised learning under metric supervision.

References

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)
2. Bigalke, A., Hansen, L., Diesel, J., Hennigs, C., Rostalski, P., Heinrich, M.P.: Anatomy-guided domain adaptation for 3D in-bed human pose estimation. arXiv preprint arXiv:2211.12193 (2022)
3. Bigalke, A., Hansen, L., Heinrich, M.P.: Adapting the mean teacher for keypoint-based lung registration under geometric domain shifts. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 280–290. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_27
4. Castillo, R., et al.: A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive. Phys. Med. Biol. **58**(9), 2861 (2013)
5. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: TransMorph: transformer for unsupervised medical image registration. Med. Image Anal. **82**, 102615 (2022)
6. 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. Med. Image Anal. **52**, 128–143 (2019)
7. Eppenhof, K.A., Pluim, J.P.: Pulmonary CT registration through supervised learning with convolutional neural networks. IEEE Trans. Med. Imaging **38**(5), 1097–1105 (2018)

8. Feydy, J.: Geometric data analysis, beyond convolutions. Ph.D. thesis, Université Paris-Saclay Gif-sur-Yvette, France (2020)

9. French, G., Mackiewicz, M., Fisher, M.: Self-ensembling for visual domain adaptation. In: International Conference on Learning Representations (2018)

10. Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X.: Deep learning in medical image registration: a review. Phys. Med. Biol. **65**(20), 20TR01 (2020)

11. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning, pp. 1180–1189. PMLR (2015)

12. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Trans. Biomed. Eng. **69**, 1173–1185 (2021)

13. Hansen, L., Heinrich, M.P.: Deep learning based geometric registration for medical images: how accurate can we get without visual features? In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 18–30. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_2

14. Haskins, G., Kruger, U., Yan, P.: Deep learning in medical image registration: a survey. Mach. Vis. Appl. **31**(1), 1–18 (2020)

15. Heinrich, M.P., et al.: Mind: modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. **16**(7), 1423–1435 (2012)

16. Jin, Z., Lei, Y., Akhtar, N., Li, H., Hayat, M.: Deformation and correspondence aware unsupervised synthetic-to-real scene flow estimation for point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7233–7243 (2022)

17. Mittal, H., Okorn, B., Held, D.: Just go with the flow: self-supervised scene flow estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11177–11185 (2020)

18. Mok, T.C.W., Chung, A.C.S.: Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 211–221. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_21

19. Perone, C.S., Ballester, P., Barros, R.C., Cohen-Adad, J.: Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. Neuroimage **194**, 1–11 (2019)

20. Rühaak, J., et al.: Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration. IEEE Trans. Med. Imaging **36**(8), 1746–1757 (2017)

21. Shen, Z., et al.: Accurate point cloud registration with robust optimal transport. In: Advances in Neural Information Processing Systems, vol. 34, pp. 5373–5389 (2021)

22. Sun, Y., Tzeng, E., Darrell, T., Efros, A.A.: Unsupervised domain adaptation through self-supervision. arXiv preprint arXiv:1909.11825 (2019)

23. Tarvainen, A., Valpola, H.: Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems, vol. 30 (2017)

24. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7472–7481 (2018)

25. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167–7176 (2017)

26. Uzunova, H., Wilms, M., Handels, H., Ehrhardt, J.: Training CNNs for image registration from few samples with model-based data augmentation. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 223–231. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_26

27. de Vos, B.D., van der Velden, B.H., Sander, J., Gilhuijs, K.G., Staring, M., Išgum, I.: Mutual information for unsupervised deep learning image registration. In: Medical Imaging 2020: Image Processing, vol. 11313, pp. 155–161. SPIE (2020)

28. Wu, W., Wang, Z.Y., Li, Z., Liu, W., Fuxin, L.: PointPWC-Net: cost volume on point clouds for (self-)supervised scene flow estimation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12350, pp. 88–107. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58558-7_6

29. Xu, Z., et al.: Double-uncertainty guided spatial and temporal consistency regularization weighting for learning-based abdominal registration. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 14–24. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_2

30. Yu, L., Wang, S., Li, X., Fu, C.-W., Heng, P.-A.: Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 605–613. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_67

31. Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al.: Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10600–10610 (2019)