MIND THE domain GAP: unsupervised modality independent deformable domain adaptation*

Lasse Hansen[1 [0000−0003−3963−7052]] and Mattias P. Heinrich[1 [0000−0002−7489−1972]]

Institute of Medical Informatics, University of Luebeck, Germany
{heinrich,hansen}@imi.uni-luebeck.de
https://www.imi.uni-luebeck.de/en/research/medical-deep-learning-lab.html

Abstract. Many learning-based unsupervised domain adaptation methods have been proposed in computer vision and medical imaging, yet they may have certain limitations: 1) training and convergence becomes more difficult if the domain gap increases, 2) a large set of labelled source domain and unlabelled target domain scans are required and 3) coarse-to-fine solutions are not straightforward. Here we propose a fundamentally different approach that leverages modality independent neighbourhood descriptors (MIND) and deformation estimation to robustly bridge the domain gap even for challenging datasets or few labelled scans (30 in our experiments). Our approach is combined with the popular nnUNet framework and integrally enables coarse-to-fine learning and is fast in both training and inference.

Keywords: Multimodal · Descriptor · Fusion.

1 Introduction

1.1 Motivation

Supervised segmentation excels in many medical diagnostics and interventional tasks. Yet defining expert labels in different modalities is time-consuming and may require re-training of human annotators. Many learning-based unsupervised domain adaptation methods have been proposed in computer vision and medical imaging, yet they may have certain limitations: 1) training and convergence becomes more difficult if the domain gap increases, 2) a large set of labelled source domain and unlabelled target domain scans are required and 3) coarse-to-fine solutions are not straightforward.

* Supported by German Federal Ministry for Economic Affairs and Energy Grant Number FKZ 01MK20012B

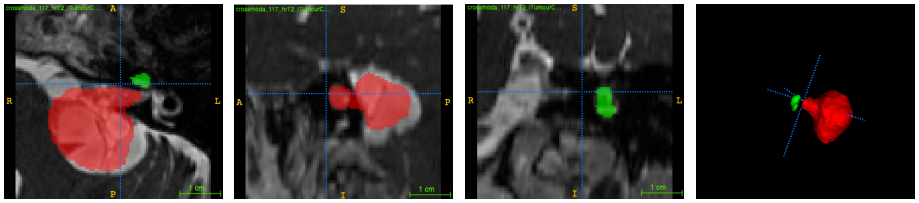

Fig. 1. This figure shows a representative example of the left hemisphere of a medium-to-moderately challenging case (#117) from the target training dataset. The overlaid segmentation is automatically obtained by cross-modal alignment and STAPLE.

2 Method and Experiments

The modality independent neighbourhood descriptor (MIND) and its extension to a self-similarity context(SSC) have been widely used with success for both intramodal and multimodal medical image registration [2]. MIND computes patch-based self-similarities within each scan and extracts a vector-valued descriptor for each voxel that is contrast invariant. It could be directly employed to bridge the domain gap, when sufficiently constraining the learning-based adaption. Here, we opt for more robustness, to avoid the use of excessive training data and hyper-parameter tuning, and use MIND for cross-modal registration.

2.1 Localiser:

For the given task of cochlea and vestibular schwannoma (VS) segmentation in high-resolution T2 MRI defining a suitable sub-volume of interest is a first difficulty. We resample both source and target domain to isotropic 1mm resolution and select fixed crops of size 64×64×96 voxels within the left and right hemisphere. This still means that the cochlea only comprise a few dozens of voxels usually around 0.01% of the volume. Hence, a further automatic localisation is necessary. We randomly select 30 source training images (15 with VS on left, and 15 on right side) and automatically register them to a subset of the target training scans both linearly and non-rigidly. Performance optimisation of classic discrete registration (deeds [1]) enabled us to reach sub-second 3D registration times with high accuracy. See https://github.com/mattiaspaul/deedsBCV for our source code. The propagated source labels are fused using the popular STAPLE algorithm (note that non-local intensity-based fusion is infeasible due to the appearance gap) [6]. Since, registration would be cumbersome at inference time, we train an nnUNet [3] for 50 epochs and in five folds that mimics the process and learns to deal with the noisy automatic labels on the target training data. Only a coarse segmentation mask is required, because subsequently we will refine both cross-domain registration and segmentation. Based on this mask an automatic crop of 48×48×48 mm is chosen (on both hemispheres using the centre-of-mass) with now a 0.5mm resolution.

2.2 Refinement:

Now, that we expect a good initial alignment, we can further refine the cross-modal alignment. For this purpose a we employ deeds again on the higher resolution crops. Again sub-second registration times are reached and STAPLE is used to reject inconsistent matches. In addition, we considered using an unsupervised CRF post-processing to improve the detail of segmentations, but found the improvements to be negligible. A second stage nnUNet is trained on the refined automatic (and less noisy) labels of the target training set. Hyperparameters (fixed crops, registration settings, ...) were determined by visual inspection and automated analysis of intermediate results (e.g. consensus of STAPLE output).

3 Experiments and Results

Experiments were conducted on a publicly available dataset comprising of 210 training brain MRI scans [5, 4] (105 T1 scans with VS and cochlea annotations, 105 T2 scans) and 32 validation T2 scans. All images were obtained on a 32-channel Siemens Avanto 1.5T scanner using a Siemens single-channel head coil. Contrast-enhanced T1-weighted imaging was performed with an MPRAGE sequence with in-plane resolution of 0.4×0.4 mm, in-plane matrix of 512×512, and slice thickness of 1.0 to 1.5 mm. High-resolution T2-weighted imaging was performed with a 3D CISS or FIESTA sequence in-plane resolution of 0.5x0.5 mm, in-plane matrix of 384x384 or 448x448, and slice thickness of 1.0 to 1.5 mm. Results were evaluated on the 32 validation scans using the automatic evaluation system at https://crossmoda.grand-challenge.org. With a mean Dice score of 0.5622 (VS: 0.6218, cochlea: 0.5026) and ASSDs of 1.9655 and 4.0430 for VS and cochlea, respectively, our proposed approach represents a robust baseline for registration based domain adaptation methods.

4 Discussion and Conclusion

We have presented a multimodal similarity based unsupervised cross-domain adaptation method that works very much differently than the vast majority of current deep learning based methods. One can relate the approach to CycleGAN (image translation), which also finds correspondences across modalities for unpaired data, but by using explicitly multimodal registration of numerous scan pairs our method can leverage powerful discrete optimisation schemes that enable realistic transformations. Our approach also avoids subtle appearance differences and artefacts still present in current translation techniques, by directly propagating segmentation labels. A downside to this method is that more powerful label fusion methods (non-local means or joint label fusion) are infeasible and we have to resort to the classic STAPLE algorithm. Despite very fast registration times for training, we avoid using registration within the target training or validation datasets (this was also discouraged by the challenge organisers) and train a fast nnUNet for inference on unseen data. Surprisingly, we found that

training the nnUNet with our noisy labels on target domain scans yielded little to no gain in qualitative accuracy. We hypothesise that the segmentation network learns an inherent bias of our transferred labels (e.g. a slight oversegmentation of cochlea) and is in its current form not capable enough to deal with label noise (e.g. down-weighting failure cases). We also realised that the number of labelled source training scans (15 for left and 15 for right hemisphere) was too little to avoid overfitting (despite using 120 unlabelled target domain crops). Due to the limited time, a number of promising improvements could not be fully evaluated, among others: multi-stage nnUNet training, increased training dataset and more extensive nonlinear registration (with outlier rejection or advanced label fusion).

5 Appendix

Semi-automatic localisation of hemisphere in target training data: We employ a STAPLE based performance estimation of rigid+nonlinear registration to automatically determine the hemisphere, which is affected by the tumour. This is important to ease the registration process and avoid unnecessary. In essence, we detect whether registering healthy or disease templates leads to higher confidence and use the average scores as decision metric. This works well in approx. 90% of cases. For a few handful of unclear predictions, we manually review and correct the side estimation. Note, that this is only done during training and the nnUNet prediction works completely automatically on unseen data.

References

1. Heinrich, M.P., Jenkinson, M., Brady, S.M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Transaction on Medical Imaging (TMI) **32**(7), 1239–48 (2013)
2. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 187–194 (2013)
3. Isensee, F., Jäger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods (2020)
4. Shapey, J., Kujawa, A., Dorent, R., Wang, G., Bisdas, S., Dimitriadis, A., Grishchuck, D., Paddick, I., Kitchen, N., Bradford, R., Saeed, S., Ourselin, S., Vercauteren, T.: Segmentation of vestibular schwannoma from magnetic resonance imaging: An open annotated dataset and baseline algorithm (2021). https://doi.org/10.7937/TCIA.9YTJ-5Q73
5. Shapey, J., Kujawa, A., Dorent, R., Wang, G., Dimitriadis, A., Grishchuk, D., Paddick, I., Kitchen, N., Bradford, R., Saeed, S.R., et al.: Segmentation of vestibular schwannoma from mri—an open annotated dataset and baseline algorithm. medRxiv (2021)
6. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. IEEE transactions on medical imaging **23**(7), 903–921 (2004)