Learning on MIND features and noisy labels from image registration

C. N. Kruse[1], L. Hansen[1], and M. P. Heinrich[1]

Institute of Medical Informatics, University of Lübeck, Germany

Abstract. We participate in the crossMoDa 2021 challenge with a combination of a classic handcrafted feature extractor, namely modality invariant neighborhood descriptors and a Deeplab segmentation model. We further improve on the initial scores by using noisy labels for the target domain obtained by non-linear image registration and a staple combination of multiple labels. We reach a final score of 0.5837 ± 0.0712.

Keywords: MIND feature, DeepLab

1 Introduction

Deep learning models in general require large amounts of annotated data to reach performance levels that are equal or better than classical methods. For deep learning in medical image analysis, datasets with a large amount of annotated data are scarce and expensive to generate as annotations generally have to be done by a trained radiologist. Furthermore, the same anatomic structures are often analyzed with different image modalities, such as computer tomography (CT), magnetic resonance (MR) imaging , ultrasound and many more. Generating annotations for each of these image modalities reduces the availability of suitable data for different deep learning tasks. Therefore, transferring labels between imaging modalities, or training models that perform equally well on multiple imaging domains is a valuable goal in medical deep learning. Recent approaches for this domain transfer include direct image translation for example using generative adversarial networks [1], or self-supervised domain adaptation, for example descriptor learning [2].

The above approaches are all trained with a specific target domain, which is available during training time. We overcome this limitation by employing a classic feature extraction as a first step in our model pipeline. With this we can train a model with only source data available at training time, that performs well on other image modalities. The details and results of our approach are presented below.

With a specific target domain present we can further improve on the method introduced above by employing additional training data either in an unsupervised manner e.g. by using auxiliary tasks or, as we chose for this challenge, by using noisy Labels we obtained from multi-modal image registration.

2 Method

We combine a classic modality invariant feature extraction model and deep learning to enable modality transfer of the learned model.

The first step in our segmentation pipeline is to calculate modality invariant neighborhood descriptors (MIND) [4]. Using these MIND features as input for the deep learning part of the segmentation pipeline greatly reduces the remaining domain gap as shown below. For the deep learning part of the pipeline we use a DeepLab v3 model with a MobileNet backbone and atrous spatial pyramid pooling [5]. We choose the DeepLab architecture rather than the commonly used U-Net, because the backbone features can be used for further experiments, for example unsupervised domain adaptation. In the U-Net the skip connections limit the further usability of learned features.

To improve performance on the given T2 target domain, we extend the training data by target domain images with noisy labels obtained from image registration.

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 [3]) 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) [7]

3 Preprocessing

The method is tested on the crossMoDa 2021 challenge dataset [6]. We use all images from the source and the target domain in our training routine. To avoid unnecessary difficulty for the cross-domain task we first resample the source data to 0.5x0.5 mm^2 in-plane resolution to match the target domain data. We further normalize all images to zero mean and unit standard deviation.

We saw in our Experiments, that reducing the resolution, as is often done for memory-intensive tasks, is counter productive in this case because of the small structures of the Cochlea. To save GPU memory and training time we therefore crop all data to 50% of their edge length and take only part of the slices resulting in full resolution images with sizes of 192x192x64 pixels. We designed this crop to ensure all annotation are still presented in the data.

The model is exclusively trained and used on these cropped data. The necessary padding of the modeled annotations for inference is done as a post-processing step.

4 Training routine

The model is trained for 2000 epochs with an Adam optimizer with a learning rate of 0.001 and a batch size of one due to GPU memory limitations. The

optimizer is initialized with class weights as the normalized inverse square root of the segmentation voxel bin count. We use affine and random noise augmentations to reduce overfitting and increase stability.

The hyper-parameters were chosen from experience with this approach on abdominal organ segmentation and not further optimized.

5 Results

On the challenge validation set we achieved a dice score of 0.5837 ± 0.0712 as posted on the leader board. Upon closer examination we see that the Cochlea are reliably segmented but with low max. scores with a mean of 0.5236 and min. and max. of 0.3474 and 0.6451, respectively. The vestibular schwannoma on the other hand has high max. scores but is segmented less reliable with a mean of 0.6437 and min. and max. scores of 0.1905 and 0.8649, respectively.

We also tested the MIND+Deeplab model without target domain training and achieved dice scores of 0.5145 ± 0.1480. Notably, the MIND+Deeplab model achieved similar max. scores without the noisy labels but completely failed in a few cases for the VS with dices scores of 0.0. This shows the increased stability when incorporating training domain data into the training.

6 Conclusion

We applied the MIND+DeepLab to brain segmentation for the first time. For a first try we achieved a decent result with an overall dice score of 0.5837 ± 0.0712. The method effectively minimizes the domain gap between the analyzed MR modalities. The model was, however, originally designed for abdominal organ segmentations and thus not optimized for brain structures especially small ones like the Cochlea. This was also shown by the relatively low dice scores on the Cochlea segmentation. For the larger VS the model performed better with max. dices scores of 0.8649, yet less stable. Due to time limitations we could not exactly identify the origin of this instability. One possible reasons is the perceptive field and the cropping in preprocessing which might confuse the model in some cases. Another possibility is that the domain transfer is still problematic in some cases, which might be improved using auxiliary task learning or different augmentation techniques.

References

1. Armanious, K., Jiang, C., Fischer, M., Küstner, T., Nikolaou, K., Gatidis, S., Yang, B.: MedGAN: Medical Image Translation using GANs. Computerized Medical Imaging and Graphics **79**, 101684 (Jan 2020). https://doi.org/10.1016/j.compmedimag.2019.101684, `http://arxiv.org/abs/1806.06397`, arXiv: 1806.06397
2. Blendowski, M., Heinrich, M.: Learning interpretable multi-modal features for alignment with supervised iterative descent. In: MIDL (2019)

3. 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)
4. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. pp. 187–194. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
5. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
6. Shapey, J., Kujawa, A., Dorent, R., Wang, G., Dimitriadis, A., Grishchuk, D., Paddick, I., Kitchen, N., Bradford, R., Saeed, S.R., Bisdas, S., Ourselin, S., Vercauteren, T.: Segmentation of vestibular schwannoma from mri — an open annotated dataset and baseline algorithm. Scientific Data (2021), in press. Preprint available at medRXiv:10.1101/2021.08.04.21261588
7. 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)