

Comparing Unsupervised Domain Adaptation and Style-Transfer Methods in CrossMoDA Challenge

Arseniy Belkov ^{*1}, Boris Shirokikh ^{*2}, and Mikhail Belyaev²

¹ Moscow Institute of Physics and Technology, Moscow, Russia

² Skolkovo Institute of Science and Technology, Moscow, Russia

`boris.shirokikh@skoltech.ru`

Abstract. Domain shift is one of the most challenging problems in medical image processing. Domain Adaptation (DA) methods are widely used to overcome the problem of differently distributed train and test data. Due to the lack of annotated medical data, unsupervised DA methods are considered a significant problem. In this paper, we provide the results of implementing and comparing two unsupervised DA approaches within the CrossMoDA Challenge 2021³. We compare the approaches of unsupervised DA via backpropagation and image-to-image translation. Here, we consider GANs for image-to-image translation as the alternative approach to tackle the domain shift problem by translating one domain into another. We made our solution available via Docker container uploaded to Docker Hub⁴.

Keywords: Deep Learning, Domain Adaptation, MRI, Segmentation

1 Data and preprocessing

We implement, train and compare our algorithms within the Cross-Modality Domain Adaptation Challenge 2021. We use the data solely from the challenge datasets [3, 4].

In all experiments we preserve the same data preprocessing pipeline:

1. We rescale the images to 0.46875×0.46875 mm axial resolution.
2. We crop the $(224, 224)$ pixel shape in the axial plain from the center of the image. Here, we assume all annotated structures to be located in the central part of the images.
3. In training data, we leave only 32 consequent axial slices that are centered on the annotated structures. For target domain we leave the bottom 55 axial slices. Removing the slices without annotated structures reduces the memory consumption and helps the models to focus on the regions of interest.

* equal contribution

³ <https://crossmoda.grand-challenge.org/>

⁴ <https://hub.docker.com/r/shboris/crossmoda-ira>

4. We normalize image, so all voxel values are in range of $[0, 1]$.

An example of preprocessed source domain image is given in the Fig. 1.

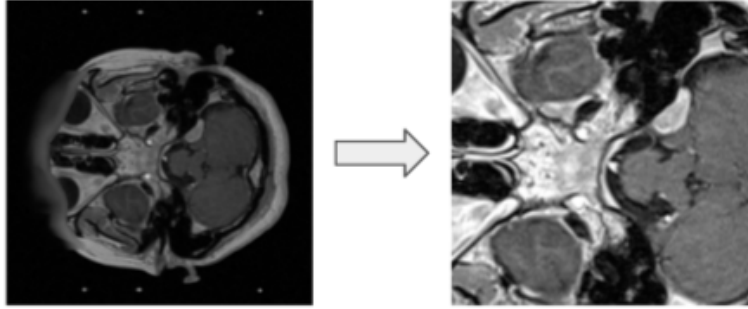


Fig. 1: The result of preprocessing of the source domain image.

2 Augmentation

Firstly, we find *gamma-transform* to improve the results both on the train set cross-validation as well as in the simple transferring of the source-based model on the validation data. During the training we apply gamma-transform with the probability of 0.7 and with $\gamma = e^{0.1x}$, where x is sampled from $\mathcal{N}(0, 1)$, to images from both source and target domain.

We additionally apply *histogram matching* and *Gaussian blur* (with $\sigma = 0.7$) to the target domain images. During the training, target images are being matched to the random source samples. Similarly to the gamma-transform, we found these techniques to improve the validation result. An example of the applied target domain transformations is shown in the Fig. 2.

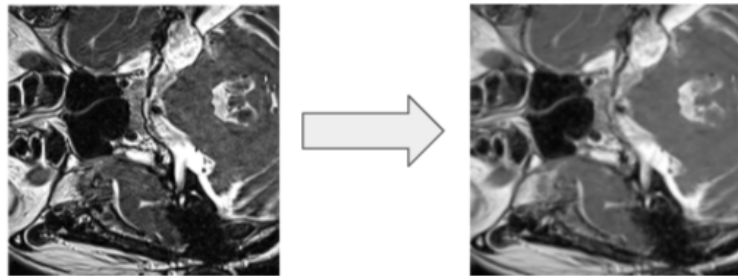


Fig. 2: The result of applying histogram matching and Gaussian blur to the target domain image.

3 Methods and results

We train two separated models to solve the binary segmentation task for *cochlea* and *schwannoma* respectively. Further, we simply merge the predictions of these two models to obtain the final result.

In all our experiments we use the modified 3D U-Net [8] as a backbone architecture for segmentation. We detail our architecture in the Fig. 3.

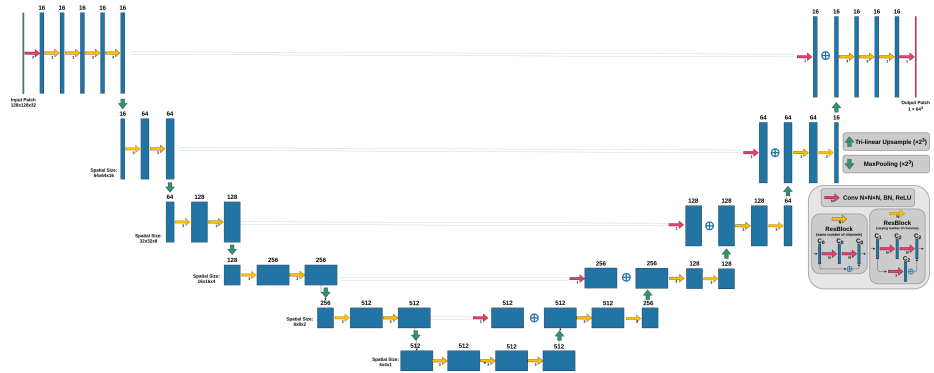


Fig. 3: Our implementation of 3D U-Net architecture.

Simple transferring of the models trained in source domain only gives us close to zero validation Dice Scores. Therefore, we continue to develop and validate domain adaptation approaches.

3.1 Unsupervised DA with deep adversarial neural networks

Authors of [1] have proposed to use the Deep Adversarial Neural Network (DANN) approach in unsupervised DA setup. However, there is no consensus of which layers to stem the adversarial head from. We use the intuition developed by the authors of [6] and stem the adversarial head from the first layers. The outputs of the first six layers (the original resolution path of our backbone) are concatenated and passed through ResNet-like classification head.

The model is trained with Adam optimizer for 300 epochs and with batch size of 6. For the first 80 epochs we set $\alpha = 0$, thus none of the gradients from the adversarial head influences the backbone parameters. From the epoch 80-th, we started changing α as in the original paper [1]. We use two cosine one-cycle learning rate schedulers: one for the first 80 epochs and the other 220 epochs. The maximum learning rate is set to $2 \cdot 10^{-4}$. The training takes about 8 hours on a 16GB nVidia Tesla V100 GPU on Zhores supercomputer [5].

3.2 Unsupervised DA with image-to-image translation

Secondly, we adapt the image-to-image translation approach of [7]. Both of our generators \mathcal{G} ($ceT1$ to $hrT2$) and \mathcal{F} ($hrT2$ to $ceT1$) have the same backbone architecture as our binary segmentation models. The networks \mathcal{G} and \mathcal{F} and the corresponding discriminators are trained for 300 epochs with the Adam optimizer and with the cosine one-cycle learning rate scheduler at the maximum learning rate of $2 \cdot 10^{-4}$. The training takes about 24 hours on a 40GB nVidia Tesla A100 GPU.

Further, we train the binary segmentation models on the generated $\mathcal{G}(ceT1)$ data. We add Dice Loss [2] as a consistency loss between $hrT2$ and $\mathcal{G}(\mathcal{F}(hrT2))$ to enforce the domain invariance. Contrary to the previously described binary segmentation training, we do not use neither *histogram matching* nor *Gaussian blur* on the target domain. Here, we train the segmentation models for 100 epochs with the same scheduler and optimizer.

3.3 Results on the validation stage

With the DANN approach, we achieved the mean Dice Score of 0.12 in schwannoma segmentation and 0.3 in cochlea segmentation. With the image-to-image translation, we achieved the mean Dice Score of 0.08 in schwannoma segmentation and 0.22 in cochlea segmentation.

References

1. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation (2015)
2. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. IEEE (2016)
3. 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. medRxiv (2021). <https://doi.org/10.1101/2021.08.04.21261588>, <https://www.medrxiv.org/content/early/2021/08/08/2021.08.04.21261588>
4. 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)
5. Zacharov, I., Arslanov, R., Gunin, M., Stefonishin, D., Pavlov, S., Panarin, O., Maliutin, A., Rykovanov, S.G., Fedorov, M.: 'zhores' – petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in skolkovo institute of science and technology. Open Engineering **9**, 512 – 520 (2019)
6. Zakazov, I., Shirokikh, B., Chernyavskiy, A., Belyaev, M.: Anatomy of domain shift impact on u-net layers in mri segmentation (2021)
7. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks (2020)

8. Özgün Çiçek, Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation (2016)