Extensive Pixel-Level Augmentation for Cross-Modality Domain Adaptation

Maria Baldeon Calisto[1][0000−0001−9379−8151] and Susana K. Lai-Yuen[2][0000−0001−6330−2813]

[1] Universidad San Francisco de Quito, Diego de Robles s/n y Vía Interoceánica, Quito, Ecuador
[2] University of South Florida, Tampa, FL, USA mbaldeonc@usfq.edu.ec

Abstract. Convolutional neural networks (CNNs) have achieved great success in automating the segmentation of medical images. Nevertheless, when a trained CNN is tested on a new domain there is a performance degradation due to the distribution shift. In this work, we present an unsupervised Extensive Pixel-level Augmentation framework (EPA) for cross-modality domain adaptation. EPA implements a two-phase image- and feature-level adaptation method. In the first phase, the source domain images are mapped to target domain in pixel space using the CycleGAN, StAC-DA, and CUT translation models. In phase 2, a deeply supervised U-Net network is trained to segment the target images using a semi-supervised adversarial learning approach. In particular, a set of discriminator networks are trained to distinguish between the target and source domain segmentations, while the U-Net aims to fool them. EPA is tested on the task of brain structure segmentation from the Crossmoda 2022 Grand Challenge, achieving a competitive performance in the validation phase of the challenge.

Keywords: Domain Adaptation · Image Segmentation · Image-to-Image Translation.

1 Introduction

Image segmentation plays a critical role in various biomedical imaging applications. Convolutional Neural Networks (CNNs) have become the de-facto model for automatic segmentation given its unbeatable accuracy and stability. However, a CNNs performance often relies on large amounts of labeled data. Moreover, when a well-trained model is tested on a new domain, the test error increases significantly due to the domain shift.

To solve the problem of domain shift between the training set (source domain) and testing set (target domain), unsupervised domain adaptation techniques (UDA) have been proposed. UDA can be broadly classified into image-level adaptation methods [1], feature-level adaptation methods [2] , and combined image-level and feature-level adaptation methods [3].

In this work, we present an **E**xtensive **P**ixel-level **A**ugmentation (EPA) framework for unsupervised cross-modality domain adaptation. EPA is a two-phase

framework that implements a combined image-level and feature-level adaptation method. In the first phase, image-level adaptation is achieved by training a CycleGAN model [4], CUT model [5], and StAC-DA model [6] to translate the source domain images to target domain in pixel space. In this way, each model learns a different translation for a same source domain image, which extensively augments the diversity and size of the labeled dataset in the target domain. In the second phase, a deeply supervised U-Net is trained by randomly selecting a batch of translated source domain images from one of the three models. Moreover, the U-Net is also trained with target domain images by adding a set of discriminator networks that discriminate between translated source domain segmentations and target domain segmentations, achieving also a feature-level adaptation. EPA is evaluated on the crossMoDa 2022 Grand Challenge, which aims to segment two critical brain structures for the treatment planning of vestibular schwannoma. Our method has a competitive performance on the validation phase of the challenge.

2 Methods

EPA comprises of two phases as presented in Figure 1. In phase 1 we produce the image-level adaptation while in phase 2 we perform the feature-level adaptation. The phases are detailed next.

2.1 Phase I

In this phase the aim is to learn a set of functions that translate images from the source domain S to target domain T in pixel space. We are given N_s labeled source domain images $\left\{(x_i^S, y_i^S)\right\}_{i=1}^{N_s}$, and N_T unlabeled target domain images $\left\{x_j^T\right\}_{j=1}^{N_T}$. Given that the source domain images and target domain images are unpaired (corresponding examples from both domains are not available), there exists a space of possible mapping functions whose quality of translation depends on the objective functions being optimized during training. Our experiments show each image-to-image translation method provides a particular mapping that translates different structures of an image better. Hence, instead of choosing only one mapping method, three methods trained with distinct objective functions are applied to the source domain to increase the size and diversity of the translated dataset. This also works as a data augmentation strategy for phase 2. Specifically, CycleGAN, StAC-DA, and CUT are selected based on their excellent performance in medical imaging translation on different types of datasets. The hyperparameters of the models are set as established in the original works. The output of phase 1 are the three translated source domain datasets. Therefore, increasing the size of the labeled translated dataset by $3\times$.

2.2 Phase II

In phase 2, a feature level adaptation is achieved by training a deeply supervised U-Net through a semi-supervised training approach. The implemented UNet

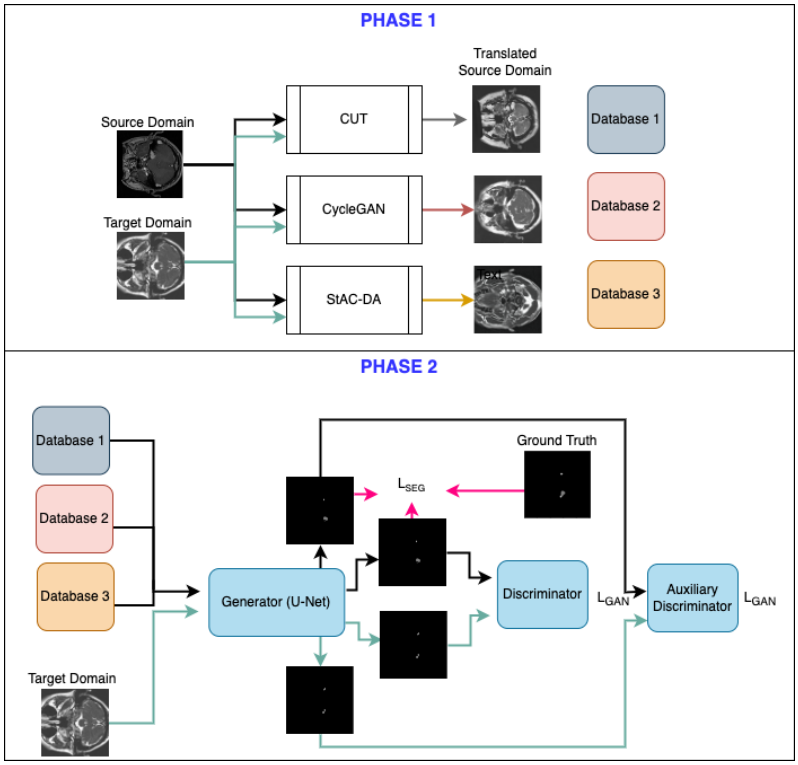

Fig. 1. Proposed EPA framework, which implements an image- and feature-level adaptation in two phases.

has an additional output layer in the second-last up-sampling block to increase the discriminativeness of the learned features. The network is first trained in a supervised manner by randomly selecting a batch of images from the three translated datasets. The soft dice loss function is optimized during training. Moreover, an auxiliary segmentation loss is added to the final loss to include the predictions of the deeply supervised layer.

Afterwards, a batch from the target domain images is sampled and sent to the U-Net to produce a predicted segmentation $\widehat{y}^T$. A discriminator network D receives $\widehat{y}^T$ and the predictions from the translated source domain $\widehat{y}^S$, and is trained to distinguish between both. Meanwhile, the U-Net takes the role of the generator and aims to fool the discriminator by producing segmentations that follow a similar distribution independent of their input domain. Since we assume the shape of the anatomical structures are not affected by the domain shift, this step forces the U-Net to learn invariant features from the source and target domain. The function being optimized is presented in Eq. 1, where the U-Net looks to minimize it while the discriminator to maximize it.

$$\mathcal{L}_{GAN}(G, D) = E_{x^t \sim T}[logD(\widehat{y}^S)] + E_{x^s \sim S}[log(1 - D(\widehat{y}^T))]\,, \tag{1}$$

To further increase the adaptation in feature space, an auxiliary discriminator network is added to the system, which receives the predictions from the deeply supervised layer. The auxiliary discriminator is trained in the same manner as the main discriminator.

3 Results

The proposed EPA framework is evaluated on the task of brain structure segmentation from the Crossmoda 2022 grand challenge[7, 8]. The goal is the segmentation of the tumour and cochlea, which are two key structures involved in the follow-up and treatment planning of vestibular schwannoma. The training dataset comprises of 210 labeled contrast-enhanced T1-weighted MRIs (source dataset) and 210 unlabeled high-resolution T2-weighted MRIs (target dataset). The validation set is composed of 64 unlabeled hrT2 MRIs. The datasets are obtained from two different medical centers, having a different spatial resolution and size.

In phase 1, the CycleGAN, StAC-DA, and CUT models are trained for 100 epochs and the weights that reduce the G_S loss selected for mapping the source domain images. In phase 2, two U-Nets are trained for 200 epochs, where each U-Net focuses solely in the segmentation of one brain substructure (i.e., tumour or cochlea). Training two architectures produced a higher segmentation accuracy than having only one. The weights that minimize the translated source domain´s validation loss are selected for evaluation.

3.1 Validation Results

The evaluation of the 64 validation images is carried via online submission to the challenge. The evaluation metrics measured are the Dice Similarity Coefficient (DSC) and Average Symmetric Surface Distance (ASSD). In Table 1 we present the results. First, we train the deeply supervised U-Net using only one of the translated datasets (displayed as method+U-Net in the table). The experiments show that StAC-DA is superior when mapping the tumour area (VS), while CycleGAN is better in the translation of the cochlea. If we ensemble the three trained models (displayed as Ensemble+U-Net in the table), the performance improves in terms of the DSC and ASSD on both substructures. However, when we train the U-Net with the 3× augmented dataset, there is a significant improvement over the individual models, models trained with two datasets, and also over the ensemble. This shows that the diversity achieved with the three translated datasets enhances the recognition and generalization of the model to the new domain. Lastly, the proposed EPA framework has the best results demonstrating that image-level and feature-level adaptation methods are complementary and necessary in datasets with a high domain shift. For the final submission, the predictions of an ensemble of EPA models using a five-fold

training scheme is used and the results shown under EPA ensemble, which show a very good performance in both substructures.

Table 1. Results on the validation set.

Method	VS DSC	VS ASSD	Cochlea DSC	Cochlea ASSD
EPA Ensemble	**0.703 ± 0.233**	**3.595 ± 11.136**	**0.502 ± 0.046**	**16.648 ± 1.337**
EPA	**0.679±0.254**	**4.417±12.641**	**0.496±0.051**	**16.669±1.330**
Phase1+U-Net	0.577±0.295	5.585±13.422	0.476±0.067	16.751±1.361
Ensemble+U-Net	0.499±0.312	9.541±17.967	0.366±0.076	16.750±1.378
StAC-DA+U-Net	0.410±0.305	11.968±19.566	0.213±0.138	18.275±1.773
CUT+U-Net	0.391±0.324	12.677±19.269	0.338±0.106	16.972± 1.398
CycleGAN+U-Net	0.383±0.308	11.895±20.938	0.362±0.109	16.893±1.420

Acknowledgment

Authors thank the Applied Signal Processing and Machine Learning Research Group of USFQ for providing the computing infrastructure (NVidia DGX workstation) to implement and execute the developed source code.

References

1. Huo, Y., Xu, Z., Moon, H., Bao, S., Assad, A., Moyo, T. K., ... Landman, B. A. : Synseg-net: Synthetic segmentation without target modality ground truth. IEEE transactions on medical imaging **38**(4), 1016–1025 (2018)
2. Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., Heng, P. A. : Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. EEE Access **7**, 99065–99076 (2019)
3. Baldeon Calisto, M., Lai-Yuen, S. K.: C-MADA: unsupervised cross-modality adversarial domain adaptation framework for medical image segmentation.In: Medical Imaging 2022: Image Processing, SPIE, vol. 12032, pp. 971–978, San Diego, California (2022) https://doi.org/https://doi.org/10.1117/12.2611499
4. Zhu, J. Y., Park, T., Isola, P., Efros, A. A. : Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232 (2017)
5. Park, T., Efros, A. A., Zhang, R., Zhu, J. Y. : Contrastive learning for unpaired image-to-image translation. In: European conference on computer vision, pp. 319–345, Springer (2020)
6. Baldeon Calisto, M., Lai-Yuen, S. K., Puente-Mejia, B. : Stac-Da: Structure Aware Cross-Modality Domain Adaptation Framework with Image and Feature Level Adaptation for Medical Image Segmentation. SSRN 4075460 (2022)

7. 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 Magnetic Resonance Imaging: An Open Annotated Dataset and Baseline Algorithm. Scientific Data **8**(1), 1–6 (2021)
8. Dorent, R., Kujawa, A., Ivory, M., Bakas, S., Rieke, N., Joutard, S., ... Vercauteren, T. : CrossMoDA 2021 challenge: Benchmark of Cross-Modality Domain Adaptation techniques for Vestibular Schwnannoma and Cochlea Segmentation. arXiv preprint arXiv:2201.02831 (2022)