Using Out-of-the-Box Frameworks for Unpaired Image Translation and Image Segmentation for the crossMoDA Challenge

Jae Won Choi[1,2]

[1] College of Medicine, Seoul National University, South Korea
[2] Department of Radiology, Armed Forces Yangju Hospital, South Korea

Abstract. The purpose of this study is to apply and evaluate out-of-the-box deep learning frameworks for the crossMoDA challenge. We use the CUT model (https://github.com/taesungp/contrastive-unpaired-translation) for domain adaptation from contrast-enhanced T1 MR to high-resolution T2 MR. As data augmentation, we generated additional images with vestibular schwannomas with lower signal intensity. For the segmentation task, we use the nnU-Net framework (https://github.com/MIC-DKFZ/nnUNet). Our final submission achieved a mean Dice score of 0.8299±0.0465 in the validation phase.

1 Introduction

Many publications on deep learning in medical image analysis focus on novel network architectures or training workflows to enhance performance. However, specialized methods make it difficult for other researchers to reproduce the published results or apply them to different datasets or tasks. Thus, publicly available generic models may be a good choice of methods, especially in medical imaging where reproducibility and generalizability are critical to be actually used in clinical practice [5]. The purpose of this study is to apply and evaluate out-of-the-box deep learning frameworks for the crossMoDA (Cross-Modality Domain Adaptation for Medical Image Segmentation) challenge (https://crossmoda.grand-challenge.org/). We use CUT (https://github.com/taesungp/contrastive-unpaired-translation) [6], a generic model for unpaired image-to-image translation based on patchwise contrastive learning and adversarial learning, to adapt contrast-enhanced T1-weighted MR images (*ceT1* domain) to domain to high-resolution T2-weighted MR images (*hrT2* domain). For the segmentation task in the hrT2 domain, we utilize nnU-Net (https://github.com/MIC-DKFZ/nnUNet) [1], a framework that showed state-of-the-art performance in multiple medical image segmentation challenges [3,2].

2 Methods

Since we focus on applications of the publicly available frameworks, there is no modification to the mathematical setting or algorithm of the original works.

All implementations were performed with PyTorch [7] (version 1.7.1) on Nvidia RTX 3090 GPUs (single GPU training).

2.1   Data

The official training set includes ceT1 images with segmentation labels from 105 patients and hrT2 images without labels from a separate set of 105 patients. The vestibular schwannomas (label 1) and cochleas (label 2) were manually segmented in consensus by the treating neurosurgeon and physicist using both the ceT1 and hrT2 images [8,9]. The official validation set includes hrT2 MR images of 32 patients. As stated in the official challenge rules, no additional data was included.

2.2   Preprocessing

Since the voxel spacings of the given training data are heterogeneous, we resample all cases to common voxel shaping of $0.6 \times 0.6 \times 1.0$ mm. Labels were also interpolated likewise for the ceT1 domain. For each case, the input volume is scaled to [0.0, 1.0]. Then, a center z-axis is calculated as the average of x and y coordinates of voxels with intensity higher than the 75th percentile of the whole volume. We crop the input volume with a size of 256×256 pixels in xy-plane around the center z-axis, resulting in an image shape of $256 \times 256 \times N$ voxels. Finally, we slice the volume data along the z-axis to acquire N images with the size of 256×256 pixels because the CUT model only supports 2D images.

2.3   Domain Adaptation

We employ two model configurations in the official PyTorch implementation of CUT, CUT and FastCUT, to train models to perform domain adaptation from ceT1 to hrT2 domain on the training set using the default options except that no resizing or cropping is performed and the number of epochs with the initial learning rate and the number of epochs with decaying learning rate are both set to 25.

   We apply the trained domain adaptation model on all ceT1 images in the training set to acquire fake hrT2 images. The generated fake hrT2 images are stacked along the z-axis to reconstruct a volume data for each case in the training set.
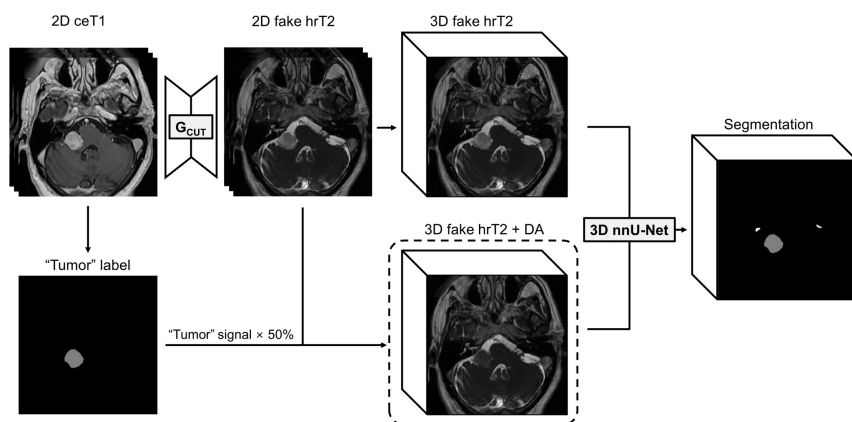
2.4   Segmentation

We use the default 3D full resolution U-Net configuration of the nnU-Net framework for training and inference for the segmentation task except that the total epochs for training was set to 250.

   The fake hrT2 volumes and labels from the corresponding ceT1 images from the training set are used for training segmentation models. We hereafter refer

to the nnU-Net model trained using fake hrT2 images generated by our trained CUT model as simply CUT, and likewise for FastCUT.

On MR T2 imaging, vestibular schwannomas are generally hyperintense but some tumors can show heterogeneous signal intensity [4]. To introduce heterogeneity of tumor signals to mimic such clinical characteristics, we generate additional training data by reducing the signal intensity of the labeled vestibular schwannomas by 50% (hereafter referred to as DA). Thus, with DA, 210 cases were used as training data instead of 105 cases. We evaluate segmentation results of models trained on the original training data and the data with DA.

Fig. 1. Overview of our implementation of unpaired image translation with CUT and segmentation with nnU-Net. Training data is augmented by generating images with tumor signals reduced by 50% (referred to as DA).

3   Results

All results are obtained via the validation leaderboard of the crossMoDA challenge. Mean Dice scores are used to compare experiments, although other metrics including Dice scores and average symmetric surface distances (ASSD) for each label are also provided. Although five-fold cross-validation is recommended for optimal usage of nnU-Net, due to lack of time and computing power considering the challenge environment, segmentation models trained on the first fold are tested on the validation set for comparison of CUT and FastCUT (Table 1).

All metrics showed better results for CUT compared to FastCUT. Therefore, we choose CUT for running full five-fold cross-validations and evaluating the effect of DA. Table 2 shows results from ensembles of five-fold cross-validations of CUT trained with and without DA.

Table 1. Comparison of results on the validation set between segmentation models trained on images generated by CUT and FastCUT. Results are based on the first fold of five-fold cross-validation workflow in the nnU-Net framework.

Experiment	Mean Dice	Tumor Dice	Tumor ASSD	Cochlea Dice	Cochlea ASSD
CUT	**0.7606±0.1357**	0.7098±0.2663	2.5870±4.2764	0.8113±0.0293	0.1829±0.0420
FastCUT	0.7377±0.1553	0.6681±0.2972	4.5624±5.9722	0.8075±0.0408	0.1994±0.0577

Table 2. Comparison of results on the validation set of CUT trained with and without DA. Results are based on five-fold cross-validations.

Experiment	Mean Dice	Tumor Dice	Tumor ASSD	Cochlea Dice	Cochlea ASSD
CUT w/o DA	0.7703±0.1428	0.7217±0.2817	1.6655±1.8147	0.8188±0.0219	0.1765±0.0340
CUT w/ DA	**0.8299±0.0465**	0.8375±0.0834	1.2940±1.2373	0.8223±0.0235	0.1720±0.0369

We submitted the ensemble of five-fold cross-validations of CUT with DA as the final submission for the challenge. The evaluation metrics showed improvements with DA not only for vestibular schwannomas but also for cochlea even though DA involved only altering the signal intensity of the vestibular schwannomas.

4   Discussion

With a mean Dice score of 0.8299, our method ranks among the top 10 on the validation leaderboard on the submission deadline. Our results show publicly available generic deep learning frameworks can achieve a certain degree of performance in medical imaging without a novel network or methodology. This study involves a limited range of hyperparameters due to the circumstance of a challenge. Further experiments on different preprocessing and data augmentation may enhance performance. Also, comparison with other out-of-the-box frameworks is warranted.

References

1. Isensee, F., Jaeger, 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 **18**(2), 203–211 (2021)
2. Isensee, F., Maier-Hein, K.H.: An attempt at beating the 3d u-net. arXiv preprint arXiv:1908.02182 (2019)
3. Isensee, F., Petersen, J., Kohl, S.A., Jäger, P.F., Maier-Hein, K.H.: nnu-net: Breaking the spell on successful medical image segmentation. arXiv preprint arXiv:1904.08128 **1**, 1–8 (2019)
4. Lin, E., Crane, B.: The management and imaging of vestibular schwannomas. American Journal of Neuroradiology **38**(11), 2034–2043 (2017)
5. Park, S.H., Han, K.: Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology **286**(3), 800–809 (2018)

6. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation (2020)
7. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
8. 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, https://wiki.cancerimagingarchive.net/x/PZwvB
9. 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