Co-Unet-GAN: a Co-Learning Domain Adaptation Model on Echocardiography Segmentation

Junyang Cai[a], Christopher M. Haggerty[b], and Joshua V. Stough[a]

[a]Computer Science, Bucknell University, Lewisburg, PA;
[b]Translational Data Science and Informatics, Geisinger, Danville, PA

ABSTRACT

Convolutional neural networks (CNN) are a powerful deep learning method for medical image segmentation. However they often lack generalizability in clinical practice, as performance drops unhelpfully when models trained from a particular source domain are transferred to a different target domain (e.g. different vendor, acquisition parameters, protocols). To address this issue, domain adaptation has attracted increasing attention because it can minimize distribution differences among different but related domains. Extending from this prior work, we introduce Co-Unet-GAN: a co-learning domain adaptation and segmentation model addressing the domain shift problem. In this model, we train a Unet segmentation network and an image translation generative adversarial network (GAN) together to generalize performance across domains given supervised data only in the source domain. We evaluate our model on two large open echocardiography datasets, using the CAMUS set as supervised source domain and EchoNet-Dynamic as the unsupervised target. We obtain mean absolute error on ejection fraction of 9.67% on Co-Unet-GAN compared to 11.28% for a previously published Unet-GAN. Our Co-Unet-GAN for image translation and segmentation is a promising solution to the domain shift problem.

Keywords: Echocardiography, Segmentation, Domain Adaptation, Neural Networks

1. INTRODUCTION

Cardiovascular diseases(CVDs) are a leading cause of morbidity for most demographics in the United States, accounting for nearly one in every five deaths.[1] Echocardiography, ultrasound of the beating heart, is a widely used and readily available imaging modality for assessing CVDs through visualization of cardiac structure and function.[2] Precise delineations of left ventricular endocardium (LV_{endo}) in echo support the accurate derivation of ejection fraction (EF), the principal clinical index for quantifying cardiac function. Deep learning and convolutional neural network (CNN) techniques have recently shown considerable promise in auto-segmenting echocardiography.[3] However, when dealing with echocardiography data across different training and test domains, the trained CNN does not perform as well. It is also impractical to acquire enough patient data across all potential target domains.

Domain adaptation is a sub-discipline of machine learning which deals with scenarios in which a model trained on a source domain is used in the context of a different but related target domain. Many researchers in the medical imaging area have recently discovered the significance of using domain adaptation to reuse pre-trained models for some related domains.[4] Methods dealing with labeled data on source domain and unlabeled data on target domain have been classified as unsupervised domain adaptation.[5] Perone et al.[6] extend the method of unsupervised domain adaptation using self-ensembling on magnetic resonance images (MRI) and improve the generalization of the models. Chen et al.[7] present an unsupervised domain adaptation framework to adapt a segmentation network to an unlabeled target domain and perform well for bidirectional cross-modality adaptation between MRI and CT images.

Yan et al.[8] proposed a generic framework called Unet-GAN consisting of an unpaired generative adversarial network (GAN) for vendor adaptation and a Unet for object segmentation. The GAN model is based on Cycle-GAN,[9] which can translate an image from a source domain to a target domain in the absence of paired examples.

Corresponding author: jc092@bucknell.edu, joshua.stough@bucknell.edu

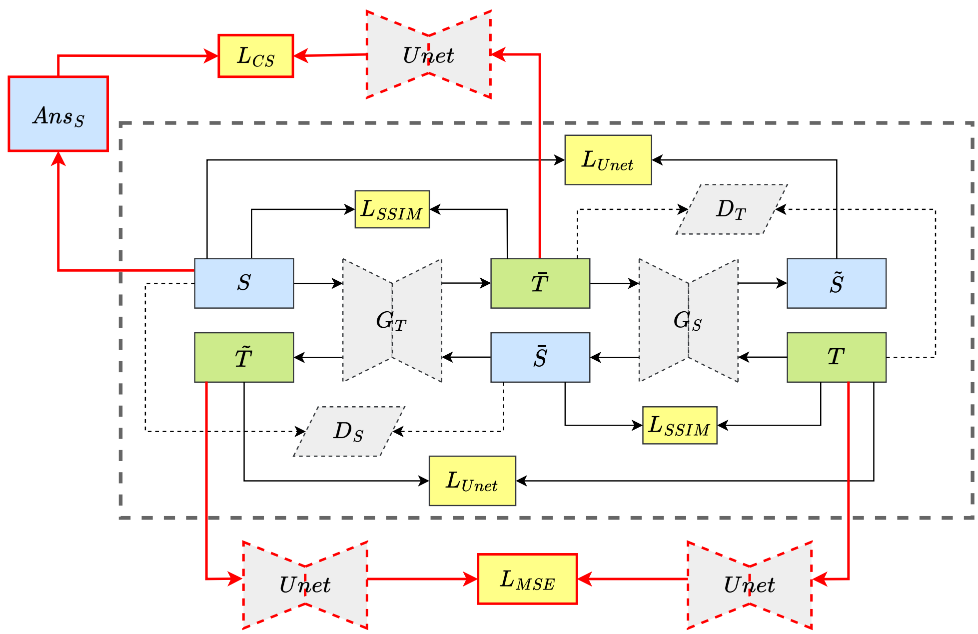

Figure 1. **Co-Unet-GAN Model Architecture:** We introduce an extension of the original Unet-GAN model[8] (represented by the features within the black dotted box). By adding two new loss functions to the Unet training procedure (highlighted in red pathways) we perform a co-learning procedure of the Unet and GAN models between the source (blue boxes) and target (green boxes) domains. Loss functions are represented in yellow color and the models are in grey color. S is source domain, $\bar{T}$ is translated target domain, $\tilde{S}$ is recovered source domain; T is target domain, $\bar{S}$ is translated source domain, $\tilde{T}$ is recovered target domain. The G_T, G_S, D_T, D_S makes the Cycle-GAN model.

Yan et al. modified the Cycle-GAN model by adding loss functions which can keep the anatomical structure of the translated image consistent with the original image and can minimize feature distinction of the recovered image from original image. Their model achieves improved segmentation results across vendors on cardiac cine MRI and provides an annotation-free solution to the cross-vendor medical image segmentation problem. However, the Unet segmentation is fixed and used only as a feature extractor, and thus does not become generalized on the translated source domain.

In this work, we introduce Co-Unet-GAN to further improve the performance of such a domain adaptation model on medical image datasets. We extend Unet-GAN by introducing a co-learning process of the Unet model during the training of the GAN network. With the co-learning process, we can ensure not only that the GAN is working to translate images from one domain to the other, but also that the Unet model generalizes performance on target images translated into the source domain. In section 2, we introduce the architecture of our model and additional loss functions to make co-learning scheme possible. In section 3, we evaluate the performance of our model on echocardiography using two independent benchmark datasets, CAMUS[10] and EchoNet,[2] and compare the performance of our Co-Unet-GAN model against both a non-adapted Unet and the previously published Unet-GAN model.

2. ARCHITECTURE AND METHODS

Image translation and segmentation are two complementary tasks in segmenting across domains. In Yan et al.,[8] the authors introduce the Unet-GAN model to deal with these two tasks separately. We propose Co-Unet-GAN, which modifies Unet-GAN model by introducing new loss functions to cooperate the training of these two tasks and improve the segmentation performance on translated source domain. In this section we elucidate these two tasks in Co-Unet-GAN and we explain the evaluation and the workflow of our model.

Model Architecture: The image translation task in Co-Unet-GAN is based on CycleGAN[9] (see Fig. 1). It contains two generators and two discriminators. The discriminators encourage the generators to translate images from one domain into outputs indistinguishable from another domain. For example, generator G_T takes inputs from the source domain S and outputs images in $\bar{T}$, representing translated source images as if they are in the target domain T. Then G_S takes inputs from $\bar{T}$ and outputs images in $\tilde{S}$ that is the recovered source domain. The generators are trained with a cycle-consistence loss to ensure images from S and $\tilde{S}$ (and similarly T and $\tilde{T}$) are close to each other, while the discriminators optimize on an adversarial loss to ensure they can distinguish between images from $\bar{S}$ and S (or $\bar{T}$ and T).

In Co-Unet-GAN, we keep the same image translation loss functions developed by Yan et al.[8] These loss functions combine information at both the image and feature levels. The structural similarity index measure (SSIM)[11] can ensure the translated image (i.e., in $\bar{S}$ or $\bar{T}$) and original image are close in luminance, contrast and structure. By assigning a much higher weight on structure difference, L_{SSIM} can ensure the generators generate structure-similar images. The encoder path of the Unet is regarded as an effective feature extractor, highlighting features most relevant for the segmentation task. The L_{Unet} (mean squared error) calculated between the original and recovered image (S and $\tilde{S}$, T and $\tilde{T}$) can encourage the similarity with respect to features specific to this segmentation task.

For the segmentation task, we use a well-validated Unet model from Stough et al.[12,13] In Co-Unet-GAN, we alternate the training of Unet with the training of GAN; the Unet becomes generalized across domains by integrating the image translation and segmentation tasks. The idea of alternating training strategy comes from GAN training, during which the generators are fixed when training the discriminators and vice versa. We add a third section in the training process of the overall network, during which the Unet is training while the weights of generators and discriminators are fixed.

The Unet segmentation model is optimized according to two additional loss functions, L_{CS} and L_{MSE}. L_{CS} calculates the cross-entropy loss of the segmentation result from images from generated target domain ($\bar{T}$) image and the ground truth. Since the SSIM loss ensures that the generated target image have the same anatomical structure as the original image, we expect the segmentation results of these two images are similar. Thus, we are able to use the ground truth for the original image from the source domain to compare with the segmentation result of the generated target domain.

L_{MSE} calculates the mean squared error of the segmentation results from both the target domain and the recovered target domain (T and $\tilde{T}$). Since we don't have the ground truth for the target domain, we can only use the difference between two segmentation results to judge the performance of our Unet on the target domain. With this loss function, our Unet model can be slightly better at segmenting target domain image and work together with GAN model to overcome the domain shift problem. It is important that these two loss functions work together. The second loss will mislead the model if we don't add the first meaningful loss function because applying this loss alone will make Unet model learn that just producing empty segmentation result will produce perfect zero loss. L_{CS} and L_{MSE} are both needed, as L_{MSE} alone would produce a degenerate solution of empty segmentation results.

Evaluation and Workflow: To evaluate our method, we compare ejection fraction, which is based on segmentation of LV_{endo}. We calculate the LV_{endo} volume through the Simpson's biplane method of disk,[14] which approximates the left ventricle as contiguous elliptical cylinders. The ejection fraction of a patient is the percentage change in LV_{endo} from end-diastolic (ED, maximum volume) to end-systolic (ES, maximum contraction), $\frac{\text{EDV-ESV}}{\text{EDV}} \times 100\%$.

The training and testing process works as follows. First, we pre-train the Unet model on the source domain to ensure good initial performance. Then we pass the pre-trained Unet model into Co-Unet-GAN and set appropriate hyper parameters to train the network. During the first 5 epochs of training, the Unet model is fixed to allow the GAN generators time to produce plausible image translation. After that initial warmup, the three-stage alternating process is used (see above). After training, we save the trained Unet model and use the GAN to translate images from the target domain to source domain. Since we find that co-learning strategy does not train our Unet model to be fully specialized on target domain, we use the trained Unet model to segment the translated source domain images ($\bar{S}$).

3. EXPERIMENTAL RESULTS

We have two datasets of echocardiogram images: CAMUS and EchoNet-Dynamic. The publicly-released portion of the CAMUS dataset consists of 450 patients, two annotated (ED/ES) phases per view, totalling 900

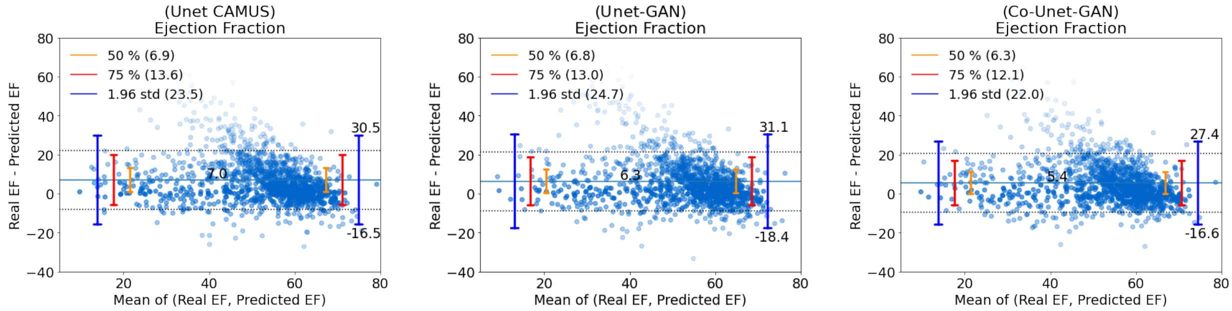

Figure 2. Bland-Altman plots comparing different model segmentation results on predicted ejection fraction compared to ground truth. From left to right: Unet CAMUS, Unet-GAN, and Co-Unet-GAN, with 50% percentile (yellow), 75% percentile (red) and $\pm 1.96\sigma$ (blue). More data points are within reported inter-rater variability for 2D echocardiography (dotted black) from left to right.

echocardiographic frames and corresponding LV_{endo} label masks. EchoNet-Dynamic data contains 10030 patient echocardiograms with one or more heartbeats in apical four-chamber view. For each video, LV_{endo} is manually annotated by clinicians in one ED and one ES frame. In the experiment, we use CAMUS as source domain and EchoNet-Dynamic as target domain. Since EchoNet-Dynamic only has annotation for LV_{endo}, we only use LV_{endo} to evaluate the performance of the model.

The Unet is pre-trained on a 90-10 split of the CAMUS dataset. We use the Ouyang et al.'s split of train, validate, and test sets for EchoNet-Dynamic. We take 900 images from CAMUS and randomly selected 900 images from EchoNet-Dynamic training set to train Co-Unet-GAN model. We test the model by 1264 Echonet-Dynamic test set patients and assess its performance in LV_{endo} segmentation by EF estimation. The Co-Unet-GAN model is trained for 60 epochs with initial learning rate 2×10^{-4} and it will start to decay from 30 epochs. The losses for SSIM, Unet, CS, MSE are weighted by 5, 25, 2, 10 respectively. Developed in PyTorch, the model requires $\sim$300s to train for one epoch on Nvidia Titan RTX.

On 1264 test echocardiograms in EchoNet, Co-Unet-GAN leads to LV_{endo} segmentation with mean absolute error of 9.67% in ejection fraction, compared to 11.28% on Unet-GAN and 11.47% on the non-adapted U-Net trained on CAMUS. We further compare against the values reported by these three models using Bland-Altman analysis[15] (see Fig. 2). We obtain bias$\pm 1.96\sigma$ of 6.29%$\pm$25.99 on Co-Unet-GAN, compared to 8.56%$\pm$36.22 on Unet-GAN and 8.68%$\pm$38.77 on non-adapted Unet trained on CAMUS. The result shows that Co-Unet-GAN improves upon the segmentation performance of these other models.

4. DISCUSSION

In this work we introduce Co-Unet-GAN, a co-learning domain adaptation model for domain shift problems in medical imaging segmentation. Building upon the prior Unet-GAN model with limited connection between Unet and GAN, our Co-Unet-GAN includes additional Unet training to make it more generalized on both domains during the training of GAN. We achieve good results on the domain shift from CAMUS to EchoNet-Dynamic and Co-Unet-GAN outperforms previous models. These results explicitly show that this model leads to improved performance of the Unet compared with a) simply applying the Unet trained on source domain data and tested on target domain data and b) applying the non-co-learned Unet to the translated source domain images, which collectively demonstrates the value of the co-learning procedure we have introduced. Since we only need same amount of data in both domains, our technique also holds promise for the problem of limited amount of available labeled data that is not enough for training directly on.

The Unet model we use in the experiment does not reach highest current standard. We expect that a more recent Unet model[16] with better segmentation result would not affect the relative performance, and further improve Co-Unet-GAN model on domain generalization. In further work, we look to produce cross-domain results that are within inter-rater variability. We also want to expand this co-learning strategy on other medical images and other segmentation models.

REFERENCES

[1] Virani, S. S. et al., "Heart disease and stroke statistics—2020 update: a report from the american heart association," *Circulation* **141**(9), e139–e596 (2020). https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000000757.

[2] Ouyang, D., He, B., et al., "Video-based ai for beat-to-beat assessment of cardiac function," *Nature* **580**(7802), 252–256 (2020). https://www.nature.com/articles/s41586-020-2145-8.

[3] Chen, C., Qin, C., et al., "Deep learning for cardiac image segmentation: a review," *Frontiers in Cardiovascular Medicine* **7**, 25 (2020).

[4] Guan, H. and Liu, M., "Domain adaptation for medical image analysis: a survey," *IEEE Transactions on Biomedical Engineering* **69**(3), 1173–1185 (2021). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9557808.

[5] Ganin, Y. and Lempitsky, V., "Unsupervised domain adaptation by backpropagation," in [*International conference on machine learning*], 1180–1189, PMLR (2015). http://proceedings.mlr.press/v37/ganin15.pdf.

[6] Perone, C. S., Ballester, P., et al., "Unsupervised domain adaptation for medical imaging segmentation with self-ensembling," *NeuroImage* **194**, 1–11 (2019). https://www.sciencedirect.com/science/article/pii/S1053811919302034.

[7] Chen, C., Dou, Q., et al., "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE transactions on medical imaging* **39**(7), 2494–2505 (2020). https://ieeexplore.ieee.org/iel7/42/4359023/08988158.pdf.

[8] Yan, W., Wang, Y., et al., "The domain shift problem of medical image segmentation and vendor-adaptation by unet-gan," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 623–631, Springer (2019). https://arxiv.org/pdf/1910.13681.

[9] Zhu, J.-Y., Park, T., et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in [*Proceedings of the IEEE international conference on computer vision*], 2223–2232 (2017). https://openaccess.thecvf.com/content_ICCV_2017/papers/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.pdf.

[10] Leclerc, S., Smistad, E., et al., "Deep learning for segmentation using an open large-scale dataset in 2d echocardiography," *IEEE transactions on medical imaging* **38**(9), 2198–2210 (2019). https://ieeexplore.ieee.org/iel7/42/4359023/08649738.pdf.

[11] Wang, Z., Bovik, A. C., et al., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1284395.

[12] Stough, J. V., Raghunath, S., et al., "Left ventricular and atrial segmentation of 2d echocardiography with convolutional neural networks," in [*Medical Imaging 2020: Image Processing*], **11313**, 32–38, SPIE (2020). https://doi.org/10.1117/12.2547375.

[13] Zhang, X., Cerna, A. E. U., et al., "Generalizability and quality control of deep learning-based 2d echocardiography segmentation models in a large clinical dataset," *The International Journal of Cardiovascular Imaging* , 1–13 (2022). https://link.springer.com/article/10.1007/s10554-022-02554-7.

[14] Folland, E., Parisi, A., et al., "Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. a comparison of cineangiographic and radionuclide techniques.," *Circulation* **60**(4), 760–766 (1979). https://www.ahajournals.org/doi/pdf/10.1161/01.CIR.60.4.760.

[15] Bland, J. M. and Altman, D. G., "Comparing methods of measurement: why plotting difference against standard method is misleading," *The lancet* **346**(8982), 1085–1087 (1995). https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(95)91748-9.pdf.

[16] Tran, T., Stough, J. V., et al., "Bayesian optimization of 2d echocardiography segmentation," in [*2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*], 1007–1011, IEEE (2021). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9433868.