DASEGAN DOMAIN ADAPTATION FOR SEGMENTATION TASKS VIA GENERATIVE ADVERSARIAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

A weakness of deep learning methods is that they can fail when there is a mismatch between source and target data domains. In medical image applications, this is a common situation when data from new vendor devices or different hospitals is available. Domain adaptation techniques aim to fill this gap by generating mappings between image domains when unlabeled data from the new target domain is available. In other cases, no target domain data (labeled or unlabeled) is available during training. In this latter case, domain generalization methods focus on learning domain-invariant representations which are more robust to new domains. In this paper, a combination of domain adaptation and generalization techniques is proposed by leveraging domain-invariant image translations for image segmentation problems. This is achieved by adversarially training a generator that transforms source images to a universal domain. To preserve the semantic consistency between the source and universal domains a segmentation consistency loss between the source and universal predictions is used. Our method was validated on the M&Ms dataset, a multi-source unsupervised domain adaptation, and generalization problem, outperforming previous methods. Particularly, our method significantly boosts the test results over the unlabeled and unseen domains without hurting source-labeled domains.

1 INTRODUCTION

In real-world problems, machine-learning methods are built with data from different but related source domains, and then, are deployed on a target domain which can exhibit a completely new data distribution. This is particularly true in medical image segmentation (Esteva et al., 2019), where practitioners build their methods based on few labeled examples collected from different hospitals (Campello et al., 2021). Once their algorithms are trained with data from different medical machines, each with a particular configuration, the model should be ready to perform well in new hospitals with other machines without time to collect a high volume of labeled or unlabeled data, or even hospitals without or data rights to collect any kind of data.

Deep Convolutional Neural Networks (CNNs) have shown superior performance on almost all computer vision problems, as-is semantic segmentation, but require large amounts of labeled data to achieve a good performance. However, although when a large number of annotated images are used, slight variations from the network training domain can hurt its performance significantly (Tzeng et al., 2017). To address this issue, several unsupervised domain adaptation and domain generalization methods have been proposed.

Unsupervised domain adaptation (UDA) methods aim to reduce the performance gap between the training set (i.e. the source domain) and the final domain in which the model will be deployed (i.e. the target domain). In practice, unlabeled data is often relatively easy and cheap to collect. UDA approaches use samples from the target domain as auxiliary information during training to reduce the performance gap (Chen et al., 2019; Hoffman et al., 2018).

Unlabeled data from the target domain is not always available. To tackle this situation, domain generalization arises to further alleviate the problem of lack of target data. Samples from multiple source domains are used trying to learn domain agnostic representations, aligning the distributions of each domain at pixel value (He et al., 2021) or feature level (Zhao et al., 2019).

Although existing works have greatly boosted the performance of domain adaptation and domain generalization tasks for semantic segmentation, most of them tackle these problems individually. We consider a more realistic setting where sometimes there exists unlabeled target data, and other times data from the target domain will not be available.

Our work. In this paper, we present DaSeGAN, an adversarial domain adaptation and domain generalization method for semantic segmentation. DaSeGAN is an adversarial framework that reduces the discrepancies between the source domains towards a universal domain. Our approach consists of: 1) a segmentation task network ; 2) an universal translation network; 3) a domain discriminator network; 4) a dense prediction similarity consistency loss. The segmentation network takes the images to perform the segmentation task. A translation networks learns to remove all discrepancies between source domains with the feedback from a domain discriminator network towards an universal domain. The discriminator is trained to identify not only if the input image is real or it is generated by the translation network, but also learns to distinguish from which source domain the input image comes from. Then, the translation network is trained adversarially by transforming a given image to a random source domain. This raises that the translation takes into account all available sources of discrepancies from the source domain, pushing the generated image to an universal representation to fool the discriminator. This approach is computationally cheaper than other algorithms that include multiple feature-level adversarial losses and reverse cycle-consistent translation losses. We show that with only a single dense prediction similarity consistency loss is needed to preserve the spatial layout between original and translated images as shown in Figure 1.

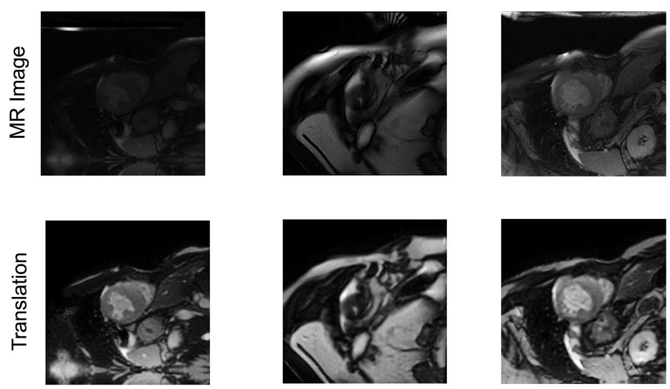

Figure 1: Domain translation towards remove source domain discrepancies. The top row shows original source Magnetic Resonance (MR) images. The bottom row shows corresponding DaSeGAN invariant translations.

Our contributions. We present an adversarial learning approach for domain generalization and domain adaptation which is applicable to all those segmentation tasks with multiple source domains where the shape of instances and spatial layout of different instances are almost the same in all domains. Experimental results demonstrate the effectiveness of our model in a real-world scenario where a bunch of domain adaptation and domain generalization approaches failed. Our model is tested on a target domain from which some unlabeled data is available, and without seeing any data from another target domain. The source code is available at *release after acceptance*.

2 Related Work

Unsupervised Domain Adaptation. Domain Adaptation seeks to reduce the performance gap between the training data and the target cases observed when networks are deployed. To handle this issue, Unsupervised Domain Adaptation (UDA) approaches address a common situation where some unlabeled data from the target domain is available. Currently, most of the UDA methods rely on generative adversarial networks (GANs) to translate images from the source domain to the target domain, and vice versa. For example, Huang et al. (2021) propose a translation regularization based on zero-shot style transfer to remove the appearance shift between source and target domains. He et al. (2021) present a collaborative framework using an ensemble of models trained on source domains to generate pseudo labels for unlabeled target data, producing soft supervision from models trained on different source domains. Although these methods have improved the performance under UDA constraints, they lack scalability as the number of source domains increases. In contrast, we only require a generator and discriminator network independently of the number of source and unlabeled target domains.

Domain Generalization. The problem of domain generalization arises in situations where unlabeled target samples are not available. Domain generalization methods rely on the use of multiple source domains to learn representations that are consistent across domains. A common approach is to learn transformations to map data between domains and then enforce invariance to these transformations as the label remains the same (Robey et al., 2021). These transformations are performed by using GANs and augment the training domains with data generated artificially (Zhou et al., 2020; Vandenhende et al., 2019). Alternatively, instead of generating new data, feature-based methods seek to remove texture discrepancies between source domains to learn disentangled representations that eliminate the style information (Nam et al., 2021).

Consistency constraint. Consistency constraints have been largely used in various domain adaptation schemes. This constraint allows the neural networks to learn mappings between different domains without paired data. Initially, the consistency constraint was defined at pixel level by setting an image reconstruction cyclic loss. This loss enforces that when an image from a specific domain is translated to another, it should be possible to arrive back at the original representation with a complimentary translation (Zhu et al., 2017). To do so, as many pairs of generators and discriminators as domains are used. The idea of matching distributions for domain adaptation was not limited only to the pixel level. Feature-level alignment, or enforcing the task networks to make the same predictions over an image and his translated version, had shown success tackling the domain-shift issue (Zhao et al., 2019; Chen et al., 2019).

In our work, we show that it is possible to keep the spatial layout of the different structures of interest by enforcing the task network to produce consistent predictions between the source domains and the universal one. This removes the overhead introduced by the reverse translation networks used in domain adaptation approaches while preserving image consistency.

3 Methodology

We consider the problem of multi-source unsupervised domain adaptation and generalization for semantic segmentation. We observe that, although the images of different source domains are sampled from different hospitals, i.e., from different i.i.d distributions, apart from differences in appearance, the shape of structures of interest and spatial layout of each structure is similar in all domains. In this section, we first provide an overview of our method. Then, we describe the loss function to enforce the generator network to perform semantically consistent translations. Finally, we explain how the discriminator network pushes the generator to translate the images to a universal domain and other losses that facilitate the network training.

3.1 Method Overview

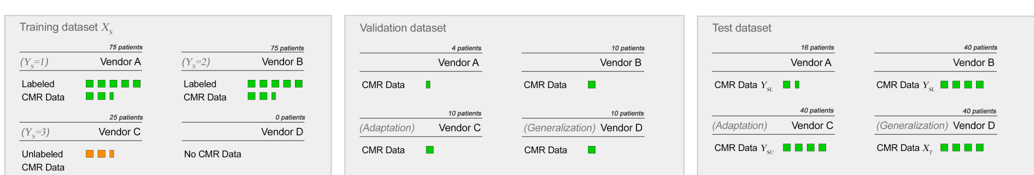

Figure 2: Open M&Ms data splitting method. Each colored square represents 10 subjects of the dataset (a CMR cine sequence).

We are provided with data from multiple source domains X_S with their corresponding domain label Y_S. Given a source domain it can be composed by a set of images X_{SL} and their corresponding mask label Y_{SL}, or unlabeled images X_{SU} (to test the domain adaptation performance). Furthermore, at test time we are provided with one or more unlabeled domains X_T that are only available at test time (to measure domain generalization). Our goal is to learn a model F that correctly predicts the labels from the unlabeled domains Y_{SU} and the unknown domains Y_T, without harming the performance on the labeled source domains Y_{SL}. Figure 2 shows our dataset overview.

To accomplish this task, we propose a universal translation framework G that seeks to reduce the discrepancies between source domains, regardless of whether they are labeled or not. We begin by learning a task model that can perform the task at samples from source domains X_{SL}. Using this base task model, it can be observed a performance degradation over the unlabeled domains X_{SU} and X_T. To reduce this performance gap between labeled and unlabeled domains, we use 1) a translation network $G_{S\to U}$ that removes the discrepancies and artifacts that appear in source domains, and 2) a discriminator network D that tries to distinguish from which source domain the translated image comes from. Figure 3 shows the overall architecture.

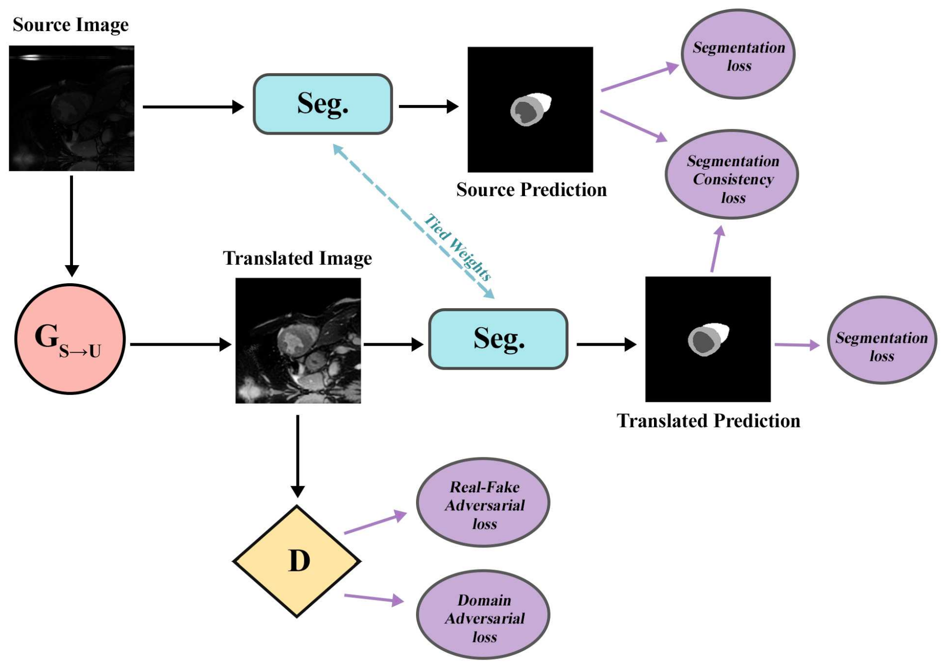

Figure 3: DaSeGAN framework overview. By only using the signals of the task network, we can maintain the consistency of the structures of interest between the initial source domain images and their translated version. Given a source domain image, the translation network learns to translate it to a random source domain, pushing the generator to learn a mapping that removes the discrepancies between these domains.

3.2 SEGMENTATION CONSISTENCY LOSS

A common approach to enforce the translation network to maintain the consistency between the source domains and their translated version is to introduce an inverse mapping. Since our objective is to find a universal domain that mitigates discrepancies from each source domain, we would need to apply as many inverse translation networks as source domains. If we want to take advantage of all source domains, the overhead introduced by these inverse translation networks increases linearly, which could make the training infeasible. To remove this overhead, our key insight is to only rely on the task network. While images from the universal domain should suppress inconsistencies and artifacts across source domains, the structure of the regions of interest has to be the same. We propose to use the task network to, given an image, enforce to identify the same areas of interest on both source and universal domains, i.e. the task prediction maps should be similar.

For each image from a source domain X_S and its translated version $G(X_S)$, we apply a segmentation consistency loss that enforces the task predictions of the translations $F(G(X_S))$ to be structurally similar to its source version $F(X_S)$. We encourage this behavior by using an L1 norm as loss as:

$$\mathcal{L}_{consis}(G) = \mathbb{E}_{X_S \sim X_U}[\| F(G(X_S)) - F(X_S) \|_1] \tag{1}$$

This limits the applicability of our approach to problems where some kind of structure must be found to preserve the translation consistency, but as a benefit, we are able to maintain the important structures, i.e. the structures that we are looking for with the task network, without introducing extra networks to conduct inverse mappings.

3.3 Domain Adversarial loss

Our purpose is to learn a mapping function $G_{S \to U}$ that given a series of source domains it reduces the discrepancies between them towards a universal domain. For that mapping, we introduce an adversarial discriminator with cross-entropy loss as objective

$$\mathcal{L}_{adv}(D, G, X_S, Y_S) = -\mathbb{E}_{(x_s, y_s) \sim (X_S, Y_S)} \sum_{k=1}^{K} \mathbb{1}_{[k != y_s]} log(\sigma(D^{(k)}(G(x_s)))) \tag{2}$$

where K denotes the number of source domains and σ represents the softmax function.The objective of G is to generate images $G(S)$ that the discriminator D cannot distinguish from which domain comes from. By taking a random source domain as label y_s, we conduct the generator G to learn a mapping that learns a invariant representation.

3.4 Other losses and Objective function

We also use several other losses in order to regularize the generator, train the task network and train the discriminator.

Image adversarial loss. When training the generator, we add an image adversarial loss to regularize the universal adversarial loss term. This loss enforces the generator to not include artifacts or peculiarities of each domain:

$$\mathcal{L}_{img}(D, G, X_S, Y_S) = -\mathbb{E}_{(x_s, y_s) \sim (X_S, Y_S)} \sum_{k=1}^{K} \mathbb{1}_{[k != y_s]} log(\sigma(D^{(k)}(G(x_s)))) \tag{3}$$

making G to generate images that seem real and not artificially generated.

Task objective. For semantic segmentation, we define the task loss $\mathcal{L}_{task}$ as the cross-entropy loss between the task predictions and the corresponding ground truth. To encourage the task network to use the available unlabeled source domains, we apply a self-ensemble approach by using the task predictions over the translated images as ground truth. As translated images quality improve as the framework training progresses, we set a hyperparameter λ_{tasku} to control the importance of task loss over unlabeled domains.

Discriminator objective. The discriminator network is trained to give the appropriate feedback to the generator to let him learn. To accomplish that, it plays an adversarial training challenging the generator and pushing it to do it better. It learns to spot real and fake, i.e. generated images, by using a binary cross-entropy loss. In addition, we use cross-entropy loss to train the discriminator to identify from which source domain an image comes.

3.5 Network Architecture and Training Details

We implement and train DaSeGAN framework using Pytorch. We adopt the U-Net (Ronneberger et al., 2015) architecture as task network with ResNet-34 (He et al., 2015) as backbone. The generator structure consists of nine residual blocks, each of which is composed of a convolutional layer, instance normalization, and a ReLU activation. For the discriminator architecture, we use a PatchGAN (Isola et al., 2017), which aims to classify whether overlapping image patches are real or fake. All networks were randomly initialized. The networks are trained with Adam optimizer with a starting learning rate of 0.001, β_1 of 0.5, β_2 of 0.999, and a batch size of 16 images. We keep the same learning rate for the first 40 epochs and linearly decay the rate to zero over the next 20 epochs. Images were center cropped and resized to 256×256 pixels. Following previous works, we evaluated DaSeGAN performance on the test set obtaining the Dice similarity index associated with the segmentation. To analyze domain generalization and domain adaptation results are stratified by domain and heart substructure of interest. We train our framework in parallel with two NVIDIA RTX 2080 GPUs with 8 GB memory each.

4 Experimental Results

4.1 Dataset: Open M&Ms

The Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&Ms) (Campello et al., 2021) provides a reference dataset for the community to build and assess future generalizable models in CMR segmentation. To do this, the dataset is composed of 375 patients with hypertrophic and dilated cardiomyopathies as well as healthy subjects. All the subjects were scanned in clinical centers in three different countries (Spain, Germany, and Canada) using four different magnetic resonance scanner vendors (Canon, Siemens, General Electric, and Philips). The CMR images were segmented by experienced clinicians from the respective institutions, including contours for the left (LV) and right ventricle (RV) blood pools as well as for the left ventricular myocardium (MYO). M&Ms presents a hard challenge as it assesses model generalization on unknown CMR vendors.

Due to data rights problems, M&Ms was released as a subset from the M&Ms Challenge dataset, where 20 cases were removed from the test partition, and 10 cases from validation partition. We denote this subset composed of 345 patients as Open M&Ms. The authors of M&Ms provide detailed metrics from all of the participants for every segmentation case. Therefore, for Open M&Ms, the metrics are calculated taking the corresponding subset of cases for a fair comparison. Models were trained using the provided training partition and were then evaluated using the test partition.

4.2 Results

We evaluate DaSeGAN under the medical image segmentation framework. In this scenario, there are abundant unlabeled data from multiple source domains, but there are few labeled samples as manual segmentation is a time-consuming task that must be done by an expert clinician. Furthermore, it can be required to deploy the model under an unknown new domain, as is not always possible to collect samples due to data-right problems.

We compare our approach with top-5 results on the M&Ms Challenge (Full et al., 2021; Zhang et al., 2021; Ma, 2021; Parreño et al., 2021; Kong & Shadden, 2021). We recalculated their publicly available scores over the released Open M&Ms subset. The bottom block of Table 1 shows the task network results, and when using our task network and when this task network is embedded in the DaSeGAN framework. Also, we compare the results when using the source image and when the task network uses the translation path. We can observe how DaSeGAN maintains its performance over the source labeled domains, Vendors A and B, with negligible improvements. Studying the domain adaptation capabilities over the unlabeled source domain, Vendor C, we can see a 1.5% improvement to the baseline method. Finally, the major performance improvement appears when addressing the domain generalization problem, with the unseen domain Vendor D, boosting segmentation indices by 2% on average.

After DaSeGAN training, we analyze the results when using source images (DaSeGAN-S) and when using their translated version (DaSeGAN-T). Regarding the results for domain adaptation, Vendor

C, the difference between using the source image and his translation is minimal. That suggests that the boost of DaSeGAN over domain adaptation is because DaSeGAN acts like other self-ensemble approaches by enforcing the source and translated mask predictions to be consistent. In contrast, with the unknown domain, Vendor D, the translation version DaSeGAN-T achieves significant improvement concerning the baseline approach and DaSeGAN-S. This result also validates the effectiveness of our proposed method for domain generalization, reducing the segmentation gap between domains.

Figure 4 shows qualitative results stratified by scanner vendor. We show DaSeGAN translation examples and compare the results using the baseline method, DaSeGAN-S, and DaSeGAN-T. At first glance, we can see how the translation network tries to equalize and enhance the images from different domains to look similar. More interestingly, we can see how DaSeGAN removes artifacts as they are not consistent across domains, suppressing a spot that had captured all the intensity and altered the image of vendor A (upper left corner). Regarding the predictions, most of the improvement can be seen at the unlabeled and unknown domains, Vendor C and D, where DaSeGAN-T gets the best results.

Table 1: Quantitative results comparison. Dice similarity indices by state-of-the-art methods. Results were stratified by scanner vendor and heart substructure. There are labeled data from Vendor A and B, unlabeled data from Vendor C, and no data from Vendor D. We denote best results as **bold**. Baseline represents the classical approach training the task network with labeled data only. DaSeGAN-S and DaSeGAN-T represent when using the source image and their translated version, respectively.

Method	Vendor A			Vendor B			Vendor C			Vendor D		
	LV	MYO	RV	LV	MYO	RV	LV	MYO	RV	LV	MYO	RV
Full et al. (2021)	0.937	**0.853**	**0.908**	0.914	0.876	**0.887**	0.903	0.841	0.883	0.909	0.838	0.882
Zhang et al. (2021)	0.931	0.848	0.905	**0.915**	0.872	0.886	0.898	0.833	0.876	0.902	0.826	0.870
Ma (2021)	0.928	0.839	0.899	0.913	0.867	0.879	0.894	0.826	0.873	0.897	0.824	0.870
Parreño et al. (2021)	0.934	0.836	0.901	0.913	0.867	0.879	0.905	0.832	0.870	0.918	0.833	0.816
Kong & Shadden (2021)	0.924	0.826	0.876	0.910	0.858	0.870	0.890	0.817	0.819	0.902	0.820	0.882
Baseline	0.934	0.850	0.902	0.911	0.870	0.883	0.895	0.834	0.874	0.902	0.830	0.874
DaSeGAN-S	0.935	0.848	0.901	0.913	0.872	0.885	0.906	0.845	0.890	0.908	0.833	0.879
DaSeGAN-T	**0.938**	**0.853**	0.907	0.913	**0.877**	**0.887**	**0.909**	**0.847**	**0.895**	**0.920**	**0.844**	**0.890**

5 Limitations and Discussion

Although our method can be used in many computer vision problems, the DaSeGAN framework needs labeled data from at least one source domain. Specifically, it needs to know to detect the structures of interest of the main task to achieve semantic consistency between translations and source images. In this sense, self-supervised learning approaches work on a pretext task without involving any human annotation to generate pseudo labels that can be used as supervisory signals. It may be interesting to explore self-supervised learning approaches to learn consistent representations and extend our method to other machine learning problems. In this sense, for example, we could replace the task network with an edge detector and maintain the general structure of the images by using the extracted edges (Li et al., 2016).

6 Conclusions

In this paper, we present a simple but effective domain adaptation and generalization framework for semantic segmentation. We show how our method can maintain the consistency between source and translated images by only using the task predictions, removing the need for inverse mappings and supplementary networks. Once we achieve semantic consistency, our method scales to unlimited source domains without any memory constraint. Finally, DaSeGAN demonstrates his effectiveness in the M&Ms dataset, a multisource domain segmentation problem.

References

Víctor M. Campello, Polyxeni Gkontra, Cristian Izquierdo, Carlos Martín-Isla, Alireza Sojoudi, Peter M. Full, Klaus Maier-Hein, Yao Zhang, Zhiqiang He, Jun Ma, Mario Parreño, Alberto Albiol,

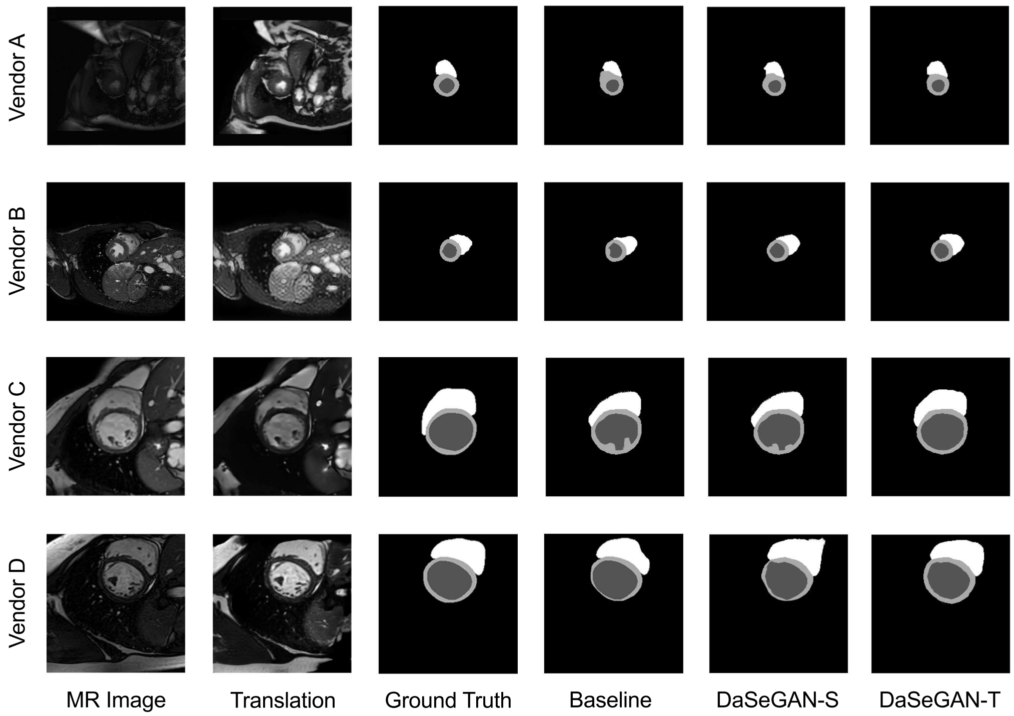

Figure 4: Qualitative results. Comparison of the segmentation results with baseline task network (Baseline), and when incorporating it to DaSeGAN approach evaluating the source image (DaSeGAN-S) and their translated version (DaSeGAN-T).

Fanwei Kong, Shawn C. Shadden, Jorge Corral Acero, Vaanathi Sundaresan, Mina Saber, Mustafa Elattar, Hongwei Li, Bjoern Menze, Firas Khader, Christoph Haarburger, Cian M. Scannell, Mitko Veta, Adam Carscadden, Kumaradevan Punithakumar, Xiao Liu, Sotirios A. Tsaftaris, Xiaoqiong Huang, Xin Yang, Lei Li, Xiahai Zhuang, David Viladés, Martín L. Descalzo, Andrea Guala, Lucia La Mura, Matthias G. Friedrich, Ria Garg, Julie Lebel, Filipe Henriques, Mahir Karakas, Ersin Çavuş, Steffen E. Petersen, Sergio Escalera, Santi Seguí, José F. Rodríguez-Palomares, and Karim Lekadir. Multi-centre, multi-vendor and multi-disease cardiac segmentation: The m amp;ms challenge. *IEEE Transactions on Medical Imaging*, pp. 1–1, 2021. doi: 10.1109/TMI.2021.3090082.

Yun-Chun Chen, Yen-Yu Lin, Ming-Hsuan Yang, and Jia-Bin Huang. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature Medicine*, 25(1):24–29, 2019. ISSN 1546-170X. doi: 10.1038/s41591-018-0316-z. URL `https://doi.org/10.1038/s41591-018-0316-z`.

Peter M Full, Fabian Isensee, Paul F Jäger, and Klaus Maier-Hein. Studying Robustness of Semantic Segmentation Under Domain Shift in Cardiac MRI. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, pp. 238–249, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68107-4.

Jianzhong He, Xu Jia, Shuaijun Chen, and Jianzhuang Liu. Multi-source domain adaptation with collaborative learning for semantic segmentation. In *CVPR*, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.

Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. CyCADA: Cycle-Consistent Adversarial Domain Adaptation. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1989–1998. PMLR, 2018. URL `http://proceedings.mlr.press/v80/hoffman18a.html`.

Xiaoqiong Huang, Zejian Chen, Xin Yang, Zhendong Liu, Yuxin Zou, Mingyuan Luo, Wufeng Xue, and Dong Ni. Style-Invariant Cardiac Image Segmentation with Test-Time Augmentation. In Esther Puyol Anton, Mihaela Pop, Maxime Sermesant, Victor Campello, Alain Lalande, Karim Lekadir, Avan Suinesiaputra, Oscar Camara, and Alistair Young (eds.), *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, pp. 305–315, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68107-4.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

Fanwei Kong and Shawn C Shadden. A Generalizable Deep-Learning Approach for Cardiac Magnetic Resonance Image Segmentation Using Image Augmentation and Attention U-Net. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, pp. 287–296, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68107-4.

Yin Li, Manohar Paluri, James M. Rehg, and Piotr Dollár. Unsupervised learning of edges. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1619–1627, 2016. doi: 10.1109/CVPR.2016.179.

Jun Ma. Histogram Matching Augmentation for Domain Adaptation with Application to Multi-centre, Multi-vendor and Multi-disease Cardiac Image Segmentation. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, pp. 177–186, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68107-4.

Hyeonseob Nam, HyunJae Lee, Jongchan Park, Wonjun Yoon, and Donggeun Yoo. Reducing domain gap by reducing style bias. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

Mario Parreño, Roberto Paredes, and Alberto Albiol. Deidentifying MRI Data Domain by Iterative Backpropagation. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, pp. 277–286, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68107-4.

Alexander Robey, George J. Pappas, and Hamed Hassani. Model-Based Domain Generalization. *arXiv preprint arXiv:2102.11436*, 2021. URL `http://arxiv.org/abs/2102.11436`.

O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pp. 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, volume 2017-January, pp. 2962–2971, 2017. ISBN 9781538604571. doi: 10.1109/CVPR.2017.316.

Simon Vandenhende, Bert De Brabandere, Davy Neven, and Luc Van Gool. A three-player GAN: Generating hard samples to improve classification networks. In *Proceedings of the 16th International Conference on Machine Vision Applications, MVA 2019*, pp. 1–6, 2019. ISBN 9784901122184. doi: 10.23919/MVA.2019.8757893.

Yao Zhang, Jiawei Yang, Feng Hou, Yang Liu, Yixin Wang, Jiang Tian, Cheng Zhong, Yang Zhang, and Zhiqiang He. Semi-supervised Cardiac Image Segmentation via Label Propagation and Style Transfer. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, pp. 219–227, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68107-4.

Sicheng Zhao, Bo Li, Xiangyu Yue, Yang Gu, Pengfei Xu, Runbo Hu, Hua Chai, and Kurt Keutzer. Multi-source domain adaptation for semantic segmentation. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Deep Domain-Adversarial Image Generation for Domain Generalisation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:13025–13032, 2020. doi: 10.1609/aaai.v34i07.7003.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.