DASeGAN: Domain Adaptation and Generalization for Medical Segmentation Tasks via Generative Adversarial Networks

Mario Parreño[1,*]**, Roberto Paredes**[1]**, and Alberto Albiol**[2]

[1]PRHLT Research Center, Universitat Politècnica de València, València, Spain
[2]iTeam Research Institute, Universitat Politècnica de València, València, Spain
[*]maparla@prhlt.upv.es

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, like segmentation, this is a common situation when data from new vendor devices or different hospitals have to be processed. In some cases, unlabeled data from the new domains are available in the training phase (domain adaptation), whereas in other cases, no training data is available for the new domains (domain generalization).

In this paper, we propose a framework for domain adaptation and generalization that maps images into a universal domain where the segmentation task is performed. The objective of using a universal domain is to unify the appearance of the images without regard to their respective original domain. To obtain this mapping, an image generator is trained to accomplish the following three objectives. First, it must generate images that are good for the task at hand. Second, it must generate realistic images, and finally, the origin domain of the images must be indistinguishable.

Our method has been validated on the M&Ms dataset, a multi-source dataset for unsupervised domain adaptation and generalization problems, outperforming previous methods. Specifically, our method significantly boosts the test results on the unlabeled and unseen domains without hurting the performance in labeled domains.

Introduction

In real-world problems, machine-learning methods are built with data from different, but related, source domains, which can then be deployed on a target domain with a different data distribution[1,2]. This is particularly common in medical image segmentation, where practitioners build their models with few labeled examples that are collected from different hospitals, with different machine vendors, and different acquisition protocols[3]. Ideally, these models should generalize to data coming from outside the training distribution (different vendors and protocols). In some cases, unlabeled data from a new location is available in advance, which provides a last opportunity to adapt the models before their deployment. On the other hand, in some cases, no data (labeled or unlabeled) is available from new locations, but the model should have enough generalization capabilities to provide an accurate result.

Deep Convolutional Neural Networks (CNNs) are currently the state of the art on almost all computer vision problems, like semantic segmentation among others. This superior performance is achieved by using large amounts of labeled data[4,5]. However, their performance can be significantly affected by slight differences at the source and target distributions[6]. To address this issue, several unsupervised domain adaptation and domain generalization methods have been proposed[7-10].

Since unlabeled data is relatively easy to collect, Unsupervised Domain Adaptation (UDA) methods aim to reduce the performance drop caused by changes between the source (training) and target (test) domains by using the available unlabeled samples as auxiliary information during training[11,12]. However, unlabeled data from the target domain is not always available. To tackle this situation, domain generalization methods try to learn domain agnostic representations by aligning the distributions of each domain at the pixel value[13] or feature level[14].

Although existing works[15] have improved the performance of domain adaptation and domain generalization tasks for semantic segmentation, most of them tackle these problems separately. We consider a more realistic scenario where adaptation and generalization can be tackled simultaneously.

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 data domains towards a new representation which is referred as universal domain in this work. The key elements of DASeGAN are: 1) a segmentation

MR Image   Translation
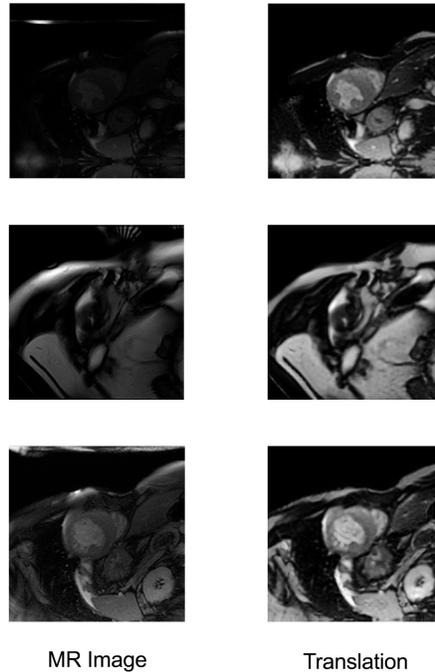
Figure 1. Domain translation to remove source domain discrepancies. The images on the left are original source Magnetic Resonance (MR) images. The images on the right are corresponding DASeGAN invariant translations.

network; 2) a universal translation network; 3) a domain discriminator network; and 4) a similarity consistency loss. The segmentation network takes the images and performs the segmentation task. The translation network converts data from different domains into a universal representation in which the discrepancies among the original domains are reduced. To do this, the translation network is trained adversarially by transforming a given image to a random source domain. This encourages the translation network to take into account all of the discrepancies from the source domains, pushing the translated image into a universal representation to fool the discriminator. With that, the discriminator is used in an adversarial manner to determine not only if whether input image is real, but also to distinguish which source domain the input image comes from. An important requirement of the translation network is that it must preserve the spatial structures of the image. To achieve this goal, a similarity consistency loss is used to enforce the spatial layout between the original and the translated images as shown in Figure 1. A key aspect of our approach is that the training process is much simpler compared to other algorithms that include multiple feature-level adversarial losses[6,16,17] and reverse cycle-consistent translation losses[18,19], for each pair-domains.

Our contributions.
We present an adversarial learning approach for domain generalization and domain adaptation problems.

The key point of our approach is that all of the data images are converted into a universal domain representation in which the segmentation task is performed. Our approach avoids the use of inverse translation networks by using an adversarial domain discriminator and spatial consistency loss between the segmentations obtained using the original images and the image representations in the universal domain.

Our framework can be applied to segmentation tasks with multiple source domains where the shape of instances and the spatial layout of different instances are similar across domains. Experimental results on the M&Ms dataset[2], which could be considered a gold standard for evaluation of domain adaptation and domain generalization approaches, demonstrates the effectiveness of our framework. The source code is available at *release after acceptance*.

1 Related Work

Unsupervised Domain Adaptation.
Domain Adaptation seeks to reduce the performance drop when source and target domains exhibit different data distributions. In the case of Unsupervised Domain Adaptation (UDA), unlabeled data from the target domain is available during training. UDA methods can be classified into two categories depending on the number of available labeled source domains: single-source domain adaptation (SDA)[20–22], and multi-source domain adaptation (MDA)[23–27].

SDA methods focus on the single-source scenario where models are generally trained using auxiliary losses to deal with the domain shift, such as reconstruction losses[28], discrepancy losses[29–31], and adversarial losses[32].

A more common situation is when multiple source domains are available. In this case, the simplest approach for MDA is to ignore the origin of the data and pool all of the data into a common general domain. However, it has been shown that this simple approach may not perform well in many scenarios[33]. More sophisticated, MDA methods make use of distribution shifts among available source domains to build more robust models. Despite the increasing interest in using MDA methods for semantic segmentation problems, most of them are still developed for image classification problems[30,34,35].

For the problem of semantic segmentation, Zhao *et al.*[14] propose learning a cycle GAN to translate images from each source domain to the target domain. Finally, a specific segmentation model for the target domain is trained using the translated source images. More recently, He *et al.*[13] presented a collaborative framework using an ensemble of models that are trained on source domains to generate pseudo labels for unlabeled target data, producing soft supervision from models that are trained on different source domains.

A common issue of MDA methods is that the number of translation and task-specific networks increases considerably as the number of data domains increases.

Domain Generalization.

The problem of domain generalization arises in situations where no training data (labeled or unlabeled) is available for a target domain. Domain generalization methods rely on the use of the available source domains to learn representations that are more invariant across the different domains or to learn models that are more robust to shifts in the input data distribution.

In order to reduce the biases between domains, Style Transfer methods seek invariant representations by reducing the appearance shifts. To achieve this goal, simple image normalization methods such as histogram matching are used[36]. More complex methods are built upon adaptive instance normalization (AdaIN) layers[37] allowing zero-shot style transfer to be performed between the source and the target domains[38,38,39].

More recent domain generalization approaches are based on the use of cycle GANs[40], which map data across source domains. In this case, the model is trained using both original and transformed images, i.e., cycle GANs are used as a powerful data augmentation method to increase the model robustness against changes in the input data distribution. In the case of semantic segmentation problems, the consistency of these augmentations is enforced by keeping the same segmentation labels in the original and transformed images[41].

Finally, instead of generating new data, feature-based domain generalization methods seek to remove texture discrepancies between source domains in order to learn disentangled representations that eliminate the style information[42].

Consistency constraint.

Consistency constraints allow neural networks to learn mappings between different domains without paired data. These constraints have been largely used in various domain adaptation schemes[12,43–45].

Consistency constraints can be applied at the pixel or feature levels. At the pixel level, the most common consistency constraint is image reconstruction cyclic loss[46]. This loss enforces the invertibility of the transformations so that when an image from a specific domain is translated to another, it should be possible to recover the original image from its translated counterpart. However, the complexity of this approach increases with the number of available domains as it requires as many cycle GANs as pairs of possible domains.

Consistency constraints can also be applied at the feature level. In this case, the models are forced to align the intermediate feature maps obtained for the original and corresponding translated images[11,14].

In our work, we present a novel consistency constraint that keeps the spatial layout of the different structures of interest by using a segmentation network and an additional synthetic universal domain. The use of a single additional domain is in contrast to other schemes that require multiple translation networks for each domain pair.

2 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 i.i.d distributions (e.g., from independent hospitals), apart from differences in appearance, the shape of structures of interest and the spatial layout of each structure is similar in all domains. The rest of this section is organized as follows. First, we provide an overview of our DASeGAN method in Section 2.1. Then, in Section 2.2 we describe the losses that we used to train DASeGAN framework. Finally, network architecture and training details are provided in Section 2.3.
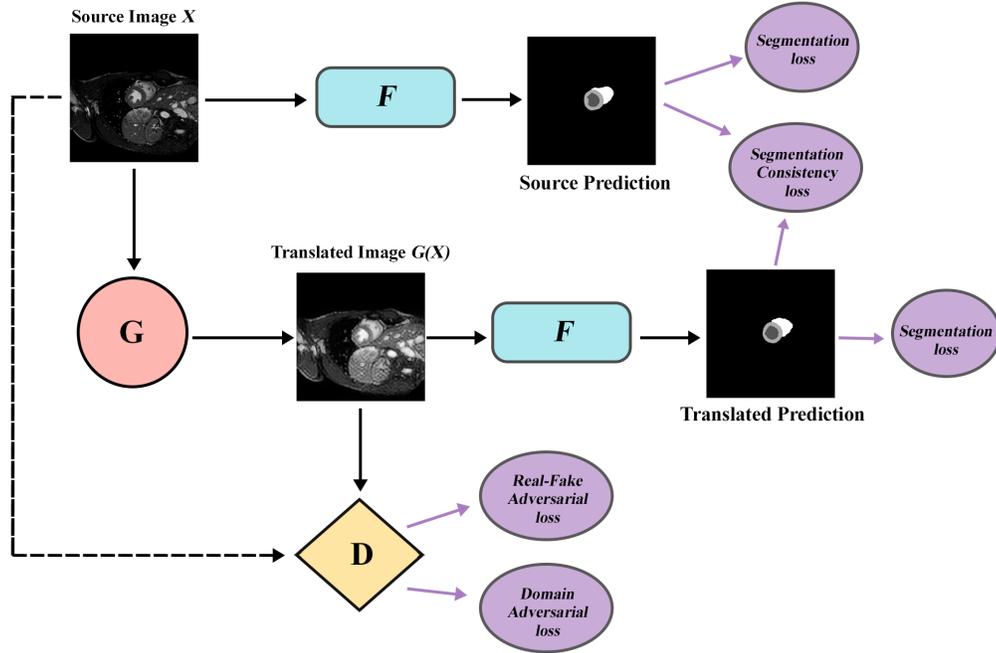
Figure 2. DASeGAN framework overview. By only using the signals of the segmentation network F, we can maintain the consistency of the structures of interest between the initial source domain images X and their translated version $G(X)$. Given a source domain image X, the translation network G learns to translate it to a random source domain $G(X)$, pushing the generator to learn a mapping that removes the discrepancies between these domains.

2.1 Method Overview

At the time of training, we are provided with data from multiple labeled data domains $\mathscr{L} = \{l_1, l_2...l_m\}$ and unlabeled data domains $\mathscr{U} = \{u_1, u_2...u_n\}$, where m and n are the number of labeled and unlabeled domains, respectively. Let $\mathscr{S} = \mathscr{L} \cup \mathscr{U}$ be the set of data source domains available for training.

At validation and test times, the models are evaluated in a set of target domains $\mathscr{T} = \mathscr{S} \cup \{t_1, t_2,...t_p\}$, where p is the number of data domains that are not available for training.

Data domains in $\mathscr{U}$ are useful to evaluate unsupervised domain adaptation, whereas domains t_i, $1 \leq i \leq p$, are useful to evaluate the model domain generalization.

Our main goal is to learn a model F that correctly predicts the semantic segmentation masks from all of the domains in $\mathscr{T}$. Figure 2 illustrates the framework used to accomplish this goal.

The key element of this framework is the translation network G. The task of G is to reduce the discrepancies between data domains (data distribution shifts) by translating data images into a common (universal) domain where the segmentation can be performed more effectively. G is trained adversarially by using a discriminator network D. The discriminator D has two main objectives. First, it must distinguish whether a given image is original (real) or translated (fake). Second, it must predict the data domain label from which the original image was sampled.

A common issue of GAN frameworks is the mode collapse where the generative network does not preserve the image structures of the image by finding an easy way to fool the discriminator. In the context of domain translation, this problem is tackled by the use of cycle consistency losses[47]. However, this requires as many independent cycle GANs as source domains pair combinations.

In our methodology, we prevent this mode collapse by applying the same segmentation model F to the original and translated images. Then, a consistency loss is used to preserve the spatial structures of the image, as shown in Figure 2.

2.2 Training losses

Segmentation loss.

For semantic segmentation, we define the segmentation loss L_{seg} as the K-way cross-entropy loss between the segmentation predictions and the corresponding ground truth:

$$L_{seg}(G,F,\mathscr{L}) = -\mathbb{E}_{(X,Y)\sim\mathscr{L}}\left[\sum_{k=1}^{K}\mathbb{1}_{[k=Y]}\log(\sigma(F^{(k)}(X))) +\lambda_{qual}\sum_{k=1}^{K}\mathbb{1}_{[k=Y]}\log(\sigma(F^{(k)}(G(X))))\right] \tag{1}$$

where K is the number of different semantic classes, σ denotes the softmax function, (X,Y) are the source and corresponding label images sampled from a labeled data domain in $\mathscr{L}$, and $G(X)$ is the translated image. The hyperparameter λ_{qual} balances the values of the two terms of L_{seg}. The value of λ_{qual} is increased during the training process to compensate for the reduced quality of $G(X)$ at the beginning of the training.

Segmentation Consistency loss
This loss enforces the consistency between the predictions made by F using the original X and translated images $G(X)$. We encourage this consistency by using an L1 norm as the loss function:

$$L_{consis}(G,F,\mathscr{S}) = \mathbb{E}_{X\sim\mathscr{S}}\left[\|\,F(G(X)) - F(X)\,\|_1\right] \tag{2}$$

It is important to note that X is picked from $\mathscr{S} = \mathscr{L}\cup\mathscr{U}$ so that X could be an unsupervised image. Therefore, while the previous segmentation loss can only be applied to images sampled from $\mathscr{L}$, this segmentation consistency loss is also applied to images where the segmentation label image Y is not available.

Domain Adversarial loss
The mapping function G seeks to reduce the discrepancies between the different source domains in $\mathscr{S}$ by translating the images into a common (universal) domain. In order to learn that mapping, we introduce an adversarial discriminator with cross-entropy:

$$L_{domain}(G,D,\mathscr{S}) = \mathbb{E}_{(X,y_s)\sim\mathscr{S}}\left[\log(\sigma(D^{(y_s)}(X)))\right] + \mathbb{E}_{(X,y_s)\sim\mathscr{S}}\left[\log(\sigma(D^{(\hat{y}_s)}(G(X))))\right] \tag{3}$$

where y_s is the label that indicates the source domain in $\mathscr{S}$ from which X was sampled.

The objective of G is to generate images $G(X)$ that the discriminator D cannot determine where $\mathscr{S}$ comes from. To do this, when the generator G is trained, a random source domain label $\hat{y}_s$, $\hat{y}_s\neq y_s$ is used as the target. By using a random domain label $\hat{y}_s$ at every batch, we enforce the generator to translate the source images to an equiprobable universal domain where the discrepancies from each domain are removed in order to fool the discriminator the maximum number of times.

Real-Fake adversarial loss.
The same discriminator network D is also trained to spot real and fake images (i.e., original images X and artificially generated images $G(X)$) by using a binary cross-entropy loss. This is a complimentary loss that enforces G to generate images that seem real and not artificially generated:

$$L_{fake}(G,D,\mathscr{S}) = \mathbb{E}_{X\sim\mathscr{S}}\left[\log(\sigma(D(X)))\right] + \mathbb{E}_{X\sim\mathscr{S}}\left[\log(1-\sigma(D(G(X))))\right] \tag{4}$$

The use of L_{fake} prevents G from including artifacts or peculiarities of specific cases.

Full Objective.
Our full objective is:

$$L_{DASeGAN}(G,D,F,\mathscr{S}) = \lambda_{seg}\cdot L_{seg} + \lambda_{consis}\cdot L_{consis} + \lambda_{domain}\cdot L_{domain} + \lambda_{fake}\cdot L_{fake} \tag{5}$$

where λ_{seg}, λ_{consis}, λ_{domain}, and λ_{fake} are the hyperparameters that are used to control the relative importance of the respective loss terms. The $L_{DASeGAN}$ loss is used to solve the following problem:

$$G^*, F^*_{seg}, D^* = \arg\min_{G,F}\max_{D} L_{DASeGAN} \tag{6}$$

2.3 Network Architecture and Training Details

We implemented and trained DASeGAN framework using Pytorch. We adopted the U-Net[48] architecture as the segmentation network with ResNet-34[49] as the backbone. The generator structure consisted of nine residual blocks, each of which was composed of a convolutional layer, instance normalization, and a ReLU activation. For the discriminator architecture, we used a PatchGAN[50], which aims to classify whether overlapping image patches are real or fake. All of the networks were randomly initialized. The networks were trained using 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 kept the same learning rate for the first 40 epochs and linearly decayed the rate to zero over the next 20 epochs. The images were center cropped and resized to 256×256 pixels.

We used two kinds of data augmentation techniques to extract more information from the original dataset and to enhance the model performance. These transformations incorporate the typical variance present in the dataset and can be divided into spatial transformations and intensity driven-modifications. We applied the following spatial transformations to all of the source images: elastic deformations with a gaussian filter of 177 pixels of standard deviation; horizontal and vertical flips; up to 45-degree rotations; and scalings and image shifts within a 20% range of the image size. To preserve coherence, we made spatial transformations by applying the same geometrical transformations to both images and masks. Meanwhile, intensity transformations were only applied to translated images $G(X)$ in order to regularize the segmentation network against transformations of that kind. These included: histogram matching with random samples; changes in brightness and saturation; and contrast modifications of up to 25% of the image.

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 capabilities, the results are stratified by the domain and the heart substructure of interest. We trained our framework in parallel with two NVIDIA RTX 2080 GPUs with 8 GB memory each.

3 Experimental Results

3.1 Dataset: Open M&Ms

We evaluate DASeGAN under medical image segmentation settings. In this scenario, there are abundant unlabeled data from multiple source domains, but there are few labeled samples since manual segmentation is a time-consuming task that must be done by an expert clinician. Furthermore, the model may be deployed on new domains where no training data is available (labeled or unlabeled) due to issues with data-right.

The Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&Ms)[2] provides a reference dataset for the community to build and assess future generalizable models in CMR segmentation. M&Ms presents a hard challenge since it assesses model generalization on unknown CMR vendors (see figure 3). To do this, the dataset was composed of 375 patients with hypertrophic and dilated cardiomyopathies as well as healthy subjects. All of 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).

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 were removed from the 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 were calculated taking the corresponding subset of cases for a fair comparison. The models were trained using the provided training partition and were then evaluated using the test partition.

The employed M&Ms dataset analyzed during the current study is available in the M&Ms Challenge repository. All the data is included in "Multi-Centre, Multi-Vendor, and Multi-Disease Cardiac Segmentation: The M&Ms Challenge"[2] published article, and can be downloaded under a reasonable request at https://www.ub.edu/mnms/. Furthermore, we confirm that all experiments were performed in accordance with relevant guidelines and regulations and consent was obtained from all participants.

3.2 Results

We compare our approach with the top-5 results on the M&Ms Challenge[51–55]. We recalculated their publicly available scores on the released Open M&Ms subset, which are shown in the top block in Table 1. The middle block in Table 1 shows our main results. First, a baseline result is obtained only using the segmentation network F trained with the labeled source domains $\mathscr{L}$. Next, we add the results when the segmentation network F is embedded in the DASeGAN framework show in Figure 2. As it is shown in this diagram, it is possible to retrieve the segmentation masks by using either the original image X (top branch), or the translated image $G(X)$ (middle branch). The results obtained using each of these branches are named DASeGAN-S and
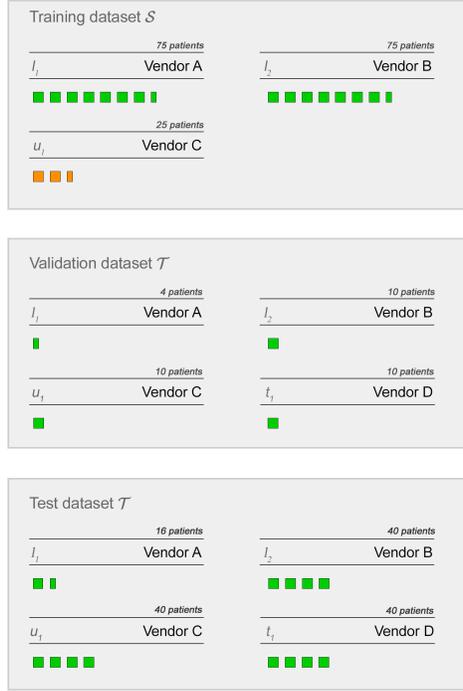
Figure 3. Open M&Ms dataset. Each colored square represents 10 subjects of the dataset (a CMR cine sequence). Green square represent labeled data; orange squares represent unlabeled data.

Method	Vendor A			Vendor B			Vendor C			Vendor D		
	LV	MYO	RV	LV	MYO	RV	LV	MYO	RV	LV	MYO	RV
Robustness[51]	0.937	**0.853**	**0.908**	0.914	0.876	**0.887**	0.903	0.841	0.883	0.909	0.838	0.882
L. P.[52]	0.931	0.848	0.905	**0.915**	0.872	0.886	0.898	0.833	0.876	0.902	0.826	0.870
H. M.[53]	0.928	0.839	0.899	0.913	0.867	0.879	0.894	0.826	0.873	0.897	0.824	0.870
Backprop[54]	0.934	0.836	0.901	0.913	0.867	0.879	0.905	0.832	0.870	0.918	0.833	0.816
Aug.[55]	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**
$\lambda_{fake} = 0$ (DASeGAN-T)	0.935	0.850	0.902	0.910	0.874	0.886	0.906	0.845	0.892	0.911	0.835	0.882
$\lambda_{consis} = 0$ (DASeGAN-T)	0.897	0.826	0.779	0.863	0.796	0.803	0.839	0.801	0.814	0.830	0.744	0.726

Table 1. Quantitative results comparison using the Dice similarity on M&Ms dataset. The 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 the best results in **bold**. The baseline represents the classical approach of training the segmentation network F with labeled data only. DASeGAN-S and DASeGAN-T represent when using the source image and their translated version, respectively. Finally, two ablation studies are carried out to study the influence of the training losses in the DASeGAN framework.

DASeGAN-T respectively in Table 1. Finally, two ablation studies are shown in the last block of Table 1 using the DASeGAN-T approach. In the first ablation, we study the influence of the real-Fake adversarial loss. In the second ablation, we study the influence of the segmentation consistency loss.

At first glance, it can observed how DASeGAN-S and DASeGAN-T maintain their performance on the source labeled domains (Vendors A and B) with negligible improvements (i.e., do not hurt the baseline performance). Studying the domain adaptation capabilities on the unlabeled source domain (Vendor C), we can see a 1.5% improvement to the baseline method by the DASeGAN-T approach. For the case of the domain generalization problem (Vendor D), the DaSeGAN-T improves the segmentation results by 2% on average compared to the baseline case, achieving the best results over all the compared methods.

This superior performance of DASeGAN-S and DASeGAN-T on the unlabeled (unseen) datasets, when comparing with the baseline method, suggests that the DASeGAN framework acts like other self-ensemble approaches by enforcing the source and translated mask predictions to be consistent[10,56].

A detailed comparison between DASeGAN-S and DASeGAN-T methods shows that the differences on the unlabeled domain (Vendor C) are minimal. 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.

The ablation studies shown in the bottom block of Table 1 show the importance of the consistency segmentation loss L_{consis}. Notice that when $\lambda_{consis} = 0$ the segmentation network F can not take any advantage with the use of unlabeled data during the training. Finally, the use of the real-fake loss L_{fake} adds a marginal improvement since it helps the generator to create more realistic translated images.

Figure 4 shows the qualitative results stratified by scanner vendor. We show DASeGAN translation examples and compare the results using the baseline method, DASeGAN-S, and DASeGAN-T. It can be observed how the translation network tries to equalize and enhance the images from different domains to look similar. More interestingly, it can be observed how the DASeGAN translation network is able to remove artifacts as long as they are not consistent across domains (see the upper left corner of the original and translated images of vendor A in Figure 4). Figure 1 shows another representative example where DASeGAN can non-linearly correct and balance the brightness levels and is also able to suppress the artifacts that were causing intensity problems. Compared to other methods (Table 1) it can be seen that DASeGAN-T outperforms all other approaches in the unlabeled and unknown domains (Vendor C and D), where there was greatest room for improvement.

4 Limitations and Discussion

Although our method can be used in many computer vision problems, the DASeGAN framework needs labeled segmented data from at least one source domain. Specifically, it needs to know how to segment the structures of interest of the main task in order to achieve semantic consistency between translations and source images. Thus, 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 in order to learn consistent representations and to extend our method to other machine-learning problems. For instance, our method could be easily adapted to classification problems by replacing segmentation labels with automatically extracted image edges in order to maintain the general spatial structure of the images[57].

5 Conclusions

In this paper, we present a simple but effective domain adaptation framework for semantic segmentation, which also improves the performance on unseen domains (domain generalization). We show how our method maintains the consistency between source and translated images by only using the task predictions and an adversarial domain discriminator, removing the need for inverse mappings and supplementary networks commonly found in the literature. An important advantage of our method is that it scales to unlimited number of source domains without increasing the complexity since we use the same segmentation network over a universal domain to obtain the segmentation masks. Finally, we demonstrate the effectiveness of DASeGAN by using the M&Ms dataset, a multi-source domain segmentation problem. Thus, DASeGAN can improve the domain adaptation and generalization capabilities of the base segmentation network by outperforming previous results in unlabeled and unseen domains, while maintaining segmentation network performance on labeled domains.

References

1. Tao, Q. *et al.* Deep learning–based method for fully automatic quantification of left ventricle function from cine mr images: A multivendor, multicenter study. *Radiology* **290**, 180513, DOI: 10.1148/radiol.2018180513 (2018).
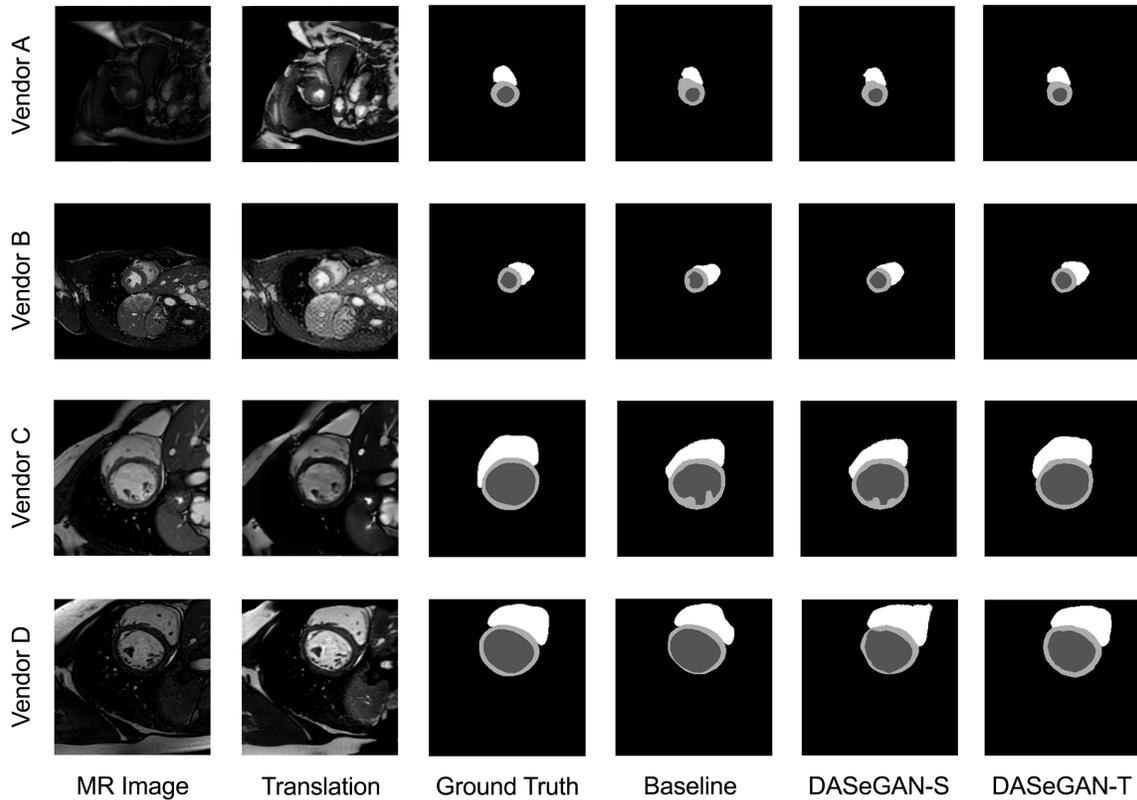
Figure 4. Qualitative results. Comparison of the segmentation results of the baseline segmentation network (Baseline) with the DASeGAN approach when using the source image (DASeGAN-S) and its translated version (DASeGAN-T).

2. Campello, V. M. e. a. Multi-centre, multi-vendor and multi-disease cardiac segmentation: The m amp;ms challenge. *IEEE Transactions on Med. Imaging* 1–1, DOI: 10.1109/TMI.2021.3090082 (2021).

3. Esteva, A. *et al.* A guide to deep learning in healthcare. *Nat. Medicine* **25**, 24–29, DOI: 10.1038/s41591-018-0316-z (2019).

4. Li, X. *et al.* Spatial pyramid based graph reasoning for semantic segmentation. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8950–8959 (2020).

5. Fu, J. *et al.* Scene segmentation with dual relation-aware attention network. *IEEE Transactions on Neural Networks Learn. Syst.* (2020).

6. Tzeng, E., Hoffman, J., Saenko, K. & Darrell, T. Adversarial discriminative domain adaptation. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, 2962–2971, DOI: 10.1109/CVPR.2017.316 (2017). 1702.05464.

7. Tang, H. & Jia, K. Discriminative Adversarial Domain Adaptation. In *Association for the Advancement of Artificial Intelligence (AAAI)* (2020).

8. Lee, S., Kim, D., Kim, N. & Jeong, S.-G. Drop to Adapt: Learning Discriminative Features for Unsupervised Domain Adaptation. In *The IEEE International Conference on Computer Vision (ICCV)* (2019).

9. Kang, G., Jiang, L., Yang, Y. & Hauptmann, A. G. Contrastive Adaptation Network for Unsupervised Domain Adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4893–4902 (2019).

10. Perone, C. S., Ballester, P. L., Barros, R. C. & Cohen-Adad, J. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. *NeuroImage* **194**, 1–11 (2019).

11. Chen, Y.-C., Lin, Y.-Y., Yang, M.-H. & Huang, J.-B. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).

12. Hoffman, J. *et al.* CyCADA: Cycle-Consistent Adversarial Domain Adaptation. In Dy, J. & Krause, A. (eds.) *Proceedings of the 35th International Conference on Machine Learning*, vol. 80 of *Proceedings of Machine Learning Research*, 1989–1998 (PMLR, 2018).

13. He, J., Jia, X., Chen, S. & Liu, J. Multi-source domain adaptation with collaborative learning for semantic segmentation. In *CVPR* (2021).

14. Zhao, S. *et al.* Multi-source domain adaptation for semantic segmentation. In *Advances in Neural Information Processing Systems*, vol. 32 (2019). 1910.12181.

15. Kouw, W. M. & Loog, M. A review of domain adaptation without target labels. *IEEE Transactions on Pattern Analysis & Mach. Intell.* **43**, 766–785, DOI: 10.1109/TPAMI.2019.2945942 (2021).

16. Kamnitsas, K. *et al.* Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. 597–609, DOI: 10.1007/978-3-319-59050-9_47 (2017).

17. Joyce, T., Chartsias, A. & Tsaftaris, S. Deep multi-class segmentation without ground-truth labels. In *Medical Imaging with Deep Learning* (2018).

18. Modanwal, G., Vellal, A., Buda, M. & Mazurowski, M. MRI image harmonization using cycle-consistent generative adversarial network. 36, DOI: 10.1117/12.2551301 (2020).

19. Gao, Y., Liu, Y., Wang, Y., Shi, Z. & Yu, J. A Universal Intensity Standardization Method Based on a Many-to-One Weak-Paired Cycle Generative Adversarial Network for Magnetic Resonance Images. *IEEE transactions on medical imaging* **38**, 2059–2069, DOI: 10.1109/TMI.2019.2894692 (2019).

20. Tsai, Y.-H. *et al.* Learning to adapt structured output space for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

21. Yang, Y. & Soatto, S. Fda: Fourier domain adaptation for semantic segmentation. *2020 IEEE/CVF Conf. on Comput. Vis. Pattern Recognit. (CVPR)* 4084–4094 (2020).

22. Li, Y., Yuan, L. & Vasconcelos, N. Bidirectional learning for domain adaptation of semantic segmentation. *2019 IEEE/CVF Conf. on Comput. Vis. Pattern Recognit. (CVPR)* 6929–6938 (2019).

23. Mansour, Y., Mohri, M. & Rostamizadeh, A. Multiple source adaptation and the rényi divergence. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, 367–374 (AUAI Press, Arlington, Virginia, USA, 2009).

24. Gholami, B., Sahu, P., Rudovic, O., Bousmalis, K. & Pavlovic, V. Unsupervised multi-target domain adaptation: An information theoretic approach. *IEEE Transactions on Image Process.* **29**, 3993–4002, DOI: 10.1109/TIP.2019.2963389 (2020).

25. Xu, R., Chen, Z., Zuo, W., Yan, J. & Lin, L. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018* (2018).

26. Zhao, S. *et al.* Multi-source distilling domain adaptation. *Proc. AAAI Conf. on Artif. Intell.* **34**, 12975–12983, DOI: 10.1609/aaai.v34i07.6997 (2020).

27. Zhao, S., Li, B., Reed, C., Xu, P. & Keutzer, K. Multi-source domain adaptation in the deep learning era: A systematic survey. *ArXiv* **abs/2002.12169** (2020).

28. Saito, K., Watanabe, K., Ushiku, Y. & Harada, T. Maximum classifier discrepancy for unsupervised domain adaptation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3723–3732, DOI: 10.1109/CVPR.2018.00392 (2018).

29. Sun, B. & Saenko, K. Deep CORAL: Correlation alignment for deep domain adaptation. *Lect. Notes Comput. Sci. (including subseries Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **9915 LNCS**, 443–450, DOI: 10.1007/978-3-319-49409-8_35 (2016). 1607.01719.

30. Peng, X. *et al.* Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, 1406–1415 (2019).

31. Zhang, Y., Liu, T., Long, M. & Jordan, M. I. Bridging theory and algorithm for domain adaptation. In *ICML* (2019).

32. Zhang, Y., Tang, H., Jia, K. & Tan, M. Domain-symmetric networks for adversarial domain adaptation. *2019 IEEE/CVF Conf. on Comput. Vis. Pattern Recognit. (CVPR)* 5026–5035 (2019).

33. Riemer, M. *et al.* Learning to learn without forgetting by maximizing transfer and minimizing interference. *Int. Conf. on Learn. Represent.* (2019).

34. Liu, H., Shao, M., Ding, Z. & Fu, Y. Structure-preserved unsupervised domain adaptation. *IEEE Transactions on Knowl. Data Eng.* **31**, 799–812, DOI: 10.1109/TKDE.2018.2843342 (2019).

35. Zhao, H. *et al.* Adversarial Multiple Source Domain Adaptation. In Bengio, S. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 31 (Curran Associates, Inc., 2018).

36. Zhang, Y. *et al.* Semi-supervised cardiac image segmentation via label propagation and style transfer. In Puyol Anton, E. *et al.* (eds.) *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, 219–227 (Springer International Publishing, Cham, 2021).

37. Huang, X. & Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV* (2017).

38. Huang, X. *et al.* Style-Invariant Cardiac Image Segmentation with Test-Time Augmentation. In Puyol Anton, E. *et al.* (eds.) *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, 305–315 (Springer International Publishing, Cham, 2021).

39. Li, L. *et al.* Random style transfer based domain generalization networks integrating shape and spatial information. In Puyol Anton, E. *et al.* (eds.) *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, 208–218 (Springer International Publishing, Cham, 2021).

40. Modanwal, G., Vellal, A. & Mazurowski, M. Normalization of breast MRIs using Cycle-Consistent Generative Adversarial Networks. *ArXiv* **abs/1912.0** (2019).

41. Robey, A., Pappas, G. J. & Hassani, H. Model-Based Domain Generalization. *arXiv preprint arXiv:2102.11436* (2021). 2102.11436.

42. Nam, H., Lee, H., Park, J., Yoon, W. & Yoo, D. Reducing domain gap by reducing style bias. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021).

43. Li, J. *et al.* Cycle-consistent Conditional Adversarial Transfer Networks. In *ACM MM* (ACM, 2019).

44. Xie, X. *et al.* Self-Supervised CycleGAN for Object-Preserving Image-to-Image Domain Adaptation. In Vedaldi, A., Bischof, H., Brox, T. & Frahm, J.-M. (eds.) *Computer Vision – ECCV 2020*, 498–513 (Springer International Publishing, Cham, 2020).

45. Köksal, A. & Lu, S. RF-GAN: A Light and Reconfigurable Network for Unpaired Image-to-Image Translation. In Ishikawa, H., Liu, C.-L., Pajdla, T. & Shi, J. (eds.) *Computer Vision – ACCV 2020*, 542–559 (Springer International Publishing, Cham, 2021).

46. Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on* (2017).

47. Dar, S. U. *et al.* Image Synthesis in Multi-Contrast MRI With Conditional Generative Adversarial Networks. *IEEE Transactions on Med. Imaging* **38**, 2375–2388 (2019).

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

49. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *2016 IEEE Conf. on Comput. Vis. Pattern Recognit. (CVPR)* 770–778 (2015).

50. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. *CVPR* (2017).

51. Full, P. M., Isensee, F., Jäger, P. F. & Maier-Hein, K. 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*, 238–249 (Springer International Publishing, Cham, 2021).

52. Zhang, Y. *et al.* 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*, 219–227 (Springer International Publishing, Cham, 2021).

53. Ma, J. 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*, 177–186 (Springer International Publishing, Cham, 2021).

54. Parreño, M., Paredes, R. & Albiol, A. Deidentifying MRI Data Domain by Iterative Backpropagation. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*, 277–286 (Springer International Publishing, Cham, 2021).

55. Kong, F. & Shadden, S. C. 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*, 287–296 (Springer International Publishing, Cham, 2021).

56. Zou, Y., Yu, Z., Kumar, B. V. K. V. & Wang, J. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. 297–313 (Springer International Publishing, 2018).

57. Li, Y., Paluri, M., Rehg, J. M. & Dollár, P. Unsupervised learning of edges. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1619–1627, DOI: 10.1109/CVPR.2016.179 (2016).