

StAC-DA: Structure aware cross-modality domain adaptation framework with image and feature-level adaptation for medical image segmentation

DIGITAL HEALTH
Volume 10: 1–15
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241277440
journals.sagepub.com/home/dhj



Maria Baldeon-Calisto¹ , Susana K. Lai-Yuen²
and Bernardo Puente-Mejia¹

Abstract

Objective: Convolutional neural networks (CNNs) have achieved state-of-the-art results in various medical image segmentation tasks. However, CNNs often assume that the source and target dataset follow the same probability distribution and when this assumption is not satisfied their performance degrades significantly. This poses a limitation in medical image analysis, where including information from different imaging modalities can bring large clinical benefits. In this work, we present an unsupervised Structure Aware Cross-modality Domain Adaptation (StAC-DA) framework for medical image segmentation.

Methods: StAC-DA implements an image- and feature-level adaptation in a sequential two-step approach. The first step performs an image-level alignment, where images from the source domain are translated to the target domain in pixel space by implementing a CycleGAN-based model. The latter model includes a structure-aware network that preserves the shape of the anatomical structure during translation. The second step consists of a feature-level alignment. A U-Net network with deep supervision is trained with the transformed source domain images and target domain images in an adversarial manner to produce probable segmentations for the target domain.

Results: The framework is evaluated on bidirectional cardiac substructure segmentation. StAC-DA outperforms leading unsupervised domain adaptation approaches, being ranked first in the segmentation of the ascending aorta when adapting from Magnetic Resonance Imaging (MRI) to Computed Tomography (CT) domain and from CT to MRI domain.

Conclusions: The presented framework overcomes the limitations posed by differing distributions in training and testing datasets. Moreover, the experimental results highlight its potential to improve the accuracy of medical image segmentation across diverse imaging modalities.

Keywords

Unsupervised domain adaptation, medical image segmentation, generative adversarial networks, convolutional neural networks, image-level adaptation, feature-level adaptation

Submission date: 12 April 2024; Acceptance date: 6 August 2024

¹Departamento de Ingeniería Industrial, Colegio de Ciencias e Ingeniería, Instituto de Innovación en Productividad y Logística CATENA-USFQ, Universidad San Francisco de Quito, Quito, Ecuador

²Department of Industrial and Management Systems, University of South Florida, Tampa, FL, USA

Corresponding author:

Maria Baldeon Calisto, Departamento de Ingeniería Industrial, Colegio de Ciencias e Ingeniería, Instituto de Innovación en Productividad y Logística CATENA-USFQ, Campus Cumbayá, Casilla Postal 17-1200-841, Quito 170901, Ecuador.

Email: mbaldeonc@usfq.edu.ec



Introduction

Over the last years, deep convolutional neural networks (CNNs) have been successfully used for a variety of computer vision problems, such as image classification,¹ object recognition,² and segmentation.³ In the medical field, CNNs have been applied to develop intelligent methods that assist medical image diagnosis by segmenting anatomical structures, identifying tumors, studying electronic signals, among many others. Medical image segmentation is a critical step in computer-aided diagnosis, where deep learning models have thrived.⁴⁻⁶ Nevertheless, CNNs often rely on high-quality and large labeled training datasets to perform well. Given that annotating medical data are a time-consuming, tedious, and expensive process, acquiring a large medical dataset can be a challenge. Moreover, many supervised learning models assume that the training dataset (source dataset), and the test set (target dataset), follow the same probability distribution. This assumption is hardly met in medical images, where different acquisition protocols, imaging equipment, imaging modalities, and patient population produce a high variation across datasets.⁷ Research has even shown that the performance of a CNN degrades in proportion to the distribution difference between target and source domain.^{8,9}

Several techniques have been proposed to solve the problem of domain shift. The simplest solution is to perform transfer learning from the source to the target domain, and sample images from the target domain to fine-tune the network.¹⁰ Nonetheless, this method needs sufficient labeled samples from both domains, which can be restrictive due to the high cost or complex acquiring process. Domain adaptation approaches (DA) have also been presented, whose objective is to transfer knowledge across domains by learning domain-invariant transformations. In domain adaptation, it is assumed that the source domain dataset is annotated while the target dataset can be fully labeled, partially labeled, or completely unlabeled.⁷ The latter, also known as unsupervised domain adaptation (UDA), is especially relevant as the target domain is not required to be annotated and can broaden the applicability to different medical datasets. UDA methods are usually divided into two families, namely feature-level adaptation models¹¹ and image-level adaptation models.¹² Feature-level adaptation aims to align the feature space distribution of the source and target domain. Meanwhile, image-level adaptation methods reduce the gap between the two domains by aligning the data in pixel space. Recently, a third family of methods has been proposed, which combines image- and feature-level adaptation models to further reduce the domain shift.^{13,14} This type of method has been shown to provide a better segmentation performance on the target domain as feature and image adaptation are complementary perspectives.

While there has been relevant progress in the development of UDA models, most works focus on the problems

of natural image segmentation^{13,15,16} or medical image classification¹⁷⁻¹⁹. Given the complex nature and dimensionality of medical images, segmenting medical data is a more challenging task. The works devoted to medical image segmentation have mostly used a CycleGAN model²⁰ to map the source domain images to target domain and train the segmentation network with the translated source domain images.^{21,22} The challenge encountered with this approach is that by only using an adversarial loss to train the generators responsible for the translation, there is no guarantee that the original shape will be conserved during the mapping. Considering that the annotation correctness must be preserved during the transformation to succeed in a segmentation task, this is an issue that requires special attention. In Ref.^{23,24}, a loss function that encourages the preservation of anatomical structures during translation was proposed. Nevertheless, both works present solely an image-level adaptation method that might not be enough when the source and target dataset suffer from a severe distribution shift. Chen et al.¹⁴ presented an unsupervised cross-modality adaptation method that implements an image and feature alignment. However, the whole framework is trained end-to-end in one step, which is computationally and memory-intensive and can prohibit its application in high-resolution imagery or settings where powerful computer systems are not available.

In this work, we present Structure Aware Cross-modality Domain Adaptation (StAC-DA), an unsupervised StAC-DA framework for medical image segmentation. StAC-DA implements an image- and feature-level adaptation in a sequential two-step approach. The first step performs an image-level alignment, where images from the source domain are translated to the target domain in pixel space by implementing a CycleGAN-based model. The latter model includes a structure-aware module composed of two segmentation networks that preserves the shape of the anatomical structure during translation. The second step consists of a feature-level alignment. In this step, a U-Net network with deep supervision is trained with the transformed source domain images and target domain images in an adversarial manner to produce probable segmentations for the target domain. Furthermore, an auxiliary discriminator network that receives the predicted segmentations of the deep supervised layer is added to the model to improve the feature-level alignment. StAC-DA is purposely designed in a sequential manner to reduce the computational requirements during training. Furthermore, to prevent the loss of information between the two steps, transfer learning is applied from the architectures from Step 1 to Step 2. The proposed framework is evaluated on the problem of bidirectional cardiac substructure segmentation from the Multi-Modality Whole Heart Segmentation Challenge dataset.²⁵ We validate the proposed method in unpaired Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) images by adapting images from MRI to CT domain, and from CT to MRI domain.

The experimental results show that StAC-DA outperforms leading unsupervised domain adaptation and is ranked first in the segmentation of the ascending aorta when adapting from MRI to CT domain and from CT to MRI domain. Moreover, StAC-DA is ranked third in the segmentation of the left ventricle blood cavity and left ventricle myocardium in the adaptation from the MRI to CT domain.

The contributions of this work are threefold. First, we propose an image- and feature-level adaptation framework for UDA that through its two-step implementation preserves the semantic information and enhances domain alignment. Secondly, we present a structure-aware CycleGAN-based model that performs an image-level alignment to emphasize shape consistency by including a source and target domain segmentation network in the model. Finally, we propose a validation loss function, based on what we have denominated the class area ratio metric, to monitor the performance of the network on the unlabeled target dataset.

The remainder of the paper is organized as follows. Section Related work provides an overview of related work on UDA models for medical image segmentation. Section Method presents the two-step StAC-DA framework and the experimental methodology. Section Results provides ablation studies and benchmark results obtained on the unsupervised cardiac segmentation task. Section Discussion contains the discussion, and Section Conclusion presents the conclusions.

Related work

In this section, we provide a review of works on UDA for medical image segmentation classified by feature-level adaptation methods, image-level adaptation methods, and combined image- and feature-level adaptation methods.

Feature-level adaptation methods

Feature-level adaptation methods transform the source and target domain data from their original feature space to a new shared and aligned feature space. The alignment is usually achieved by minimizing a distance measure such as maximum mean discrepancy,²⁶ correlation distance,²⁷ or adversarial discriminator accuracy.²⁸ In,¹¹ Dou et al. proposed a plug-and-play adversarial domain adaptation network that aligns the target and source domain in feature space at multiple scales. Using adversarial learning, two discriminators are built to distinguish multilevel features and predicted segmentation masks on the two domains. Degel et al.²⁹ presented a combined deep-learning-based approach that incorporates shape prior information and a domain discriminator to encourage feature domain-invariance across datasets. Kamnitsas et al.³⁰ developed a multilevel feature adaptation method to derive domain-invariant features with a multiconnected domain discriminator. Although feature-level adaptation

methods are efficient, they can fail to capture low-level appearance variance and do not enforce semantic consistency.

Image-level adaptation methods

In image-level adaptation methods, the images from one domain are transferred to another domain through a pixel-to-pixel transformation. The majority of methods achieve image translation through the application of generative models such as generative adversarial networks (GANs).³¹ Zhao et al.¹² proposed a modified U-Net to synthesize MR brain images from CT images using a paired co-registered image dataset. Afterwards, a MALP-EM network³² is applied to segment the whole brain from the synthetic MR images. Nie et al.³³ proposed a context-aware GAN that generates CT images from MR images. An image-gradient-difference based loss is presented to alleviate the blurriness of the generated CT image. Tomar et al.²⁴ presented a self-attentive spatial adaptive normalization method that introduces a self-attention module that focuses on the anatomical structures of organs to improve the image translation task. Although these methods have shown promising results, most of them work well mainly in datasets with a limited domain shift and can lose semantic content.

Combined feature-level and image-level adaptation methods

Recently, works have proposed using a hybrid image- and feature-level adaptation method to mitigate severe domain shift. Chen et al.¹⁴ presented a synergistic framework for cardiac image segmentation. The appearance of the source images is translated to the target domain with a cycle-consistent GAN (CycleGAN)²⁰ while simultaneously a two-stream CNN is trained with a domain discriminator to reduce the domain gap in feature space. Yan et al.³⁴ applied a CycleGAN with a modified loss, which includes image and feature-level similarity, to transform target images to source domain. Afterward, a U-net network is fully trained in the source domain and used for inference. Cui et al.³⁵ proposed a GAN-based bidirectional adaptive framework, which applies a CycleGAN-based process to translate the images from the source domain to the target domain. During the image synthesis and semantic prediction, the networks share the same encoder. Furthermore, a self-attention mechanism and spectral normalization are included in the generator, encoder, and discriminator networks to enhance the authenticity of the generated target domain images. In Ref.³⁶, we presented a sequential image- and feature-level adaptation method for brain MRI segmentation. In the first step, a CycleGAN model is implemented for image translation between the source and target domain. In the second step, a U-Net network is trained to

segment the target domain images in an adversarial manner using information about the shape, texture, and contour of the predicted segmentation. StAC-DA is a significant extension of that work. Specifically, we introduce a new structure-aware CycleGAN-based model in Step 1 to encourage annotation correctness during translation. Furthermore, we enhance the validation by performing experiments on the bidirectional segmentation of a cardiac dataset and provide a more detailed description of the framework.

Method

In this predictive study, we propose the StAC-DA framework. StAC-DA is an unsupervised structure-aware cross-modality domain adaptation method composed of two sequential steps during training, as shown in Figure 1. In Step 1, the images of the source domain are translated to the target domain using a CycleGAN-based model with a structure-aware module to preserve the shape of the anatomical structures during the domain translation. This step converts the source domain images into target style-like images. In Step 2, a feature-level adaptation method is proposed by training a U-Net segmentation network with the translated source domain images and the target

domain images using an adversarial training scheme. The objective is to produce probable segmentations for the target domain that follow the same probability distribution as the ground truth segmentations from the source domain. Since the U-Net network is trained with synthesized target domain images, only the trained U-Net architecture is necessary to produce the predictions during inference.

In the following subsections, we describe Steps 1 and 2 of the proposed framework, the validation loss function implemented to monitor the performance of the network, the dataset used for the experiments, training details, and the quantitative metrics for evaluation. The computational experiments were performed at Universidad San Francisco de Quito, Ecuador, from August 2021 to August 2022.

Step 1: structure-aware image-level adaptation

In this step, the aim is to learn a structure-aware mapping network that translates the images from the source domain S to the target domain T in terms of visual appearance. It is assumed that the morphology of the anatomical structures is invariant to changes in the image domain (medical images from different modalities have the same

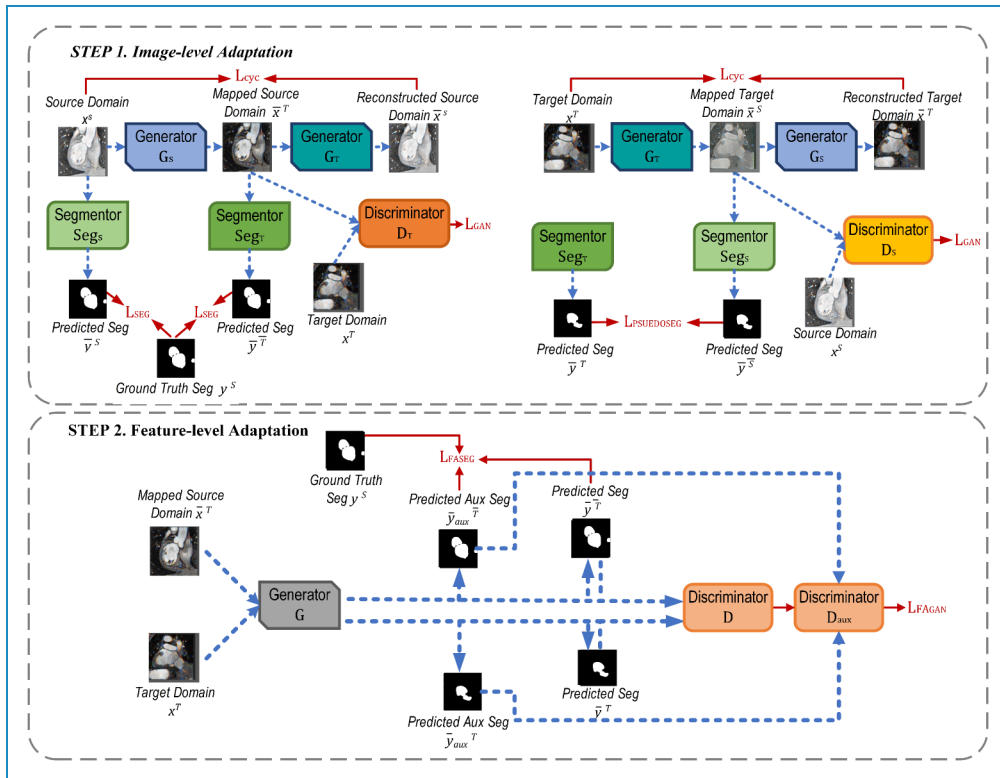


Figure 1. StAC-DA framework is composed of two sequential steps during training. Step 1 performs a structure-aware image-level adaptation by implementing a CycleGAN-based model. Step 2 implements a feature-level adaptation by training a deeply supervised U-Net network with an adversarial training scheme.

semantic information). Thus, the translated source domain images can approximate the target domain distribution and be used to train a target domain segmentation network. Since the accuracy of the segmentation task is dependent on the shape of the region of interest (ROI), it is indispensable that the original structure of the ROI remains preserved during the transformation. For this step, a CycleGAN-based model with a structure-aware network is proposed.

The CycleGAN-based model consists of two modules, the image synthesis module, and the structure-aware module. The image synthesis module is made up of the source domain generator network G_S , target domain generator network G_T , source domain discriminator network D_S , and target domain discriminator network D_T . Using adversarial training, G_S and D_S learn to transform images from domain S to domain T , while G_T and D_T learn to translate images from domain T to domain S . Let $\{x^s, y^s\} = \{(x_i^s, y_i^s) | i = 1, \dots, N_s\}$ represent the images and ground truth from domain S , and $\{x^T\} = \{(x_j^T) | j = 1, \dots, N_t\}$ be the unlabeled images from domain T . Then, G_S is the mapping network that translates the source domain images x^S to synthetic target domain images \hat{x}^T , $G_S = x^S \rightarrow \hat{x}^T$. Meanwhile, D_S tries to distinguish real target domain images x^T from synthetic ones \hat{x}^T . Therefore, D_S and G_S compete in a two-player minimax game where D_S maximizes the objective function shown in equation (1), while G_S minimizes it. However, instead of using the original cross-entropy loss, a Least Squares GAN³⁷ objective function is implemented for a more stable training.

$$L_{\text{GAN}}(G_S, D_S) = \mathbb{E}_{x^T \sim T} [\log D_S(x^T)] + \mathbb{E}_{x^S \sim S} [\log(1 - D_S(G_S(x^S)))] \quad (1)$$

G_T and D_T are trained in a similar manner, where G_T translates the target domain images x^T to source domain \hat{x}^S ($G_T = x^T \rightarrow \hat{x}^S$), and D_T aims to discriminate between x^S and \hat{x}^S . The generator networks have a ResNet architecture,³⁸ while the discriminators have a PatchGAN structure.³⁹

Furthermore, to avoid model collapse during training and incentivize the mapped images \hat{x}^T and \hat{x}^S to return to their original distribution x^S and x^T after applying G_T and G_S , respectively, a cycle consistency loss L_{cyc} as displayed in equation (2) is added. Where $*$ corresponds to an L1 loss as it has shown to have better visual results than an L2 loss.

$$L_{\text{cyc}}(G_S, G_T) = \mathbb{E}_{x^S \sim S} [G_T(G_S(x^S)) - x^S] + \mathbb{E}_{x^T \sim T} [G_S(G_T(x^T)) - x^T] \quad (2)$$

Finally, an identity loss is also included to encourage the generators to preserve the color composition during mapping. The identity loss also regularizes the generator to approximate an identity function when images are

translated to the same domain. The loss is presented in equation (3), where $*$ represents an L1 loss.

$$L_{\text{identity}}(G_S, G_T) = \mathbb{E}_{x^S \sim S} [G_T(x^S) - x^S] + \mathbb{E}_{x^T \sim T} [G_S(x^T) - x^T] \quad (3)$$

The second module of the CycleGAN based model is the structure-aware module. The structure-aware module is composed of a source domain segmentation network Seg_S , and a target domain segmentation network Seg_T . The segmentation networks are connected directly to the system by receiving as input the real and synthetic images as shown in Figure 1. Seg_S is trained to segment the source domain images in a supervised manner using the labeled source domain dataset $\{x^s, y^s\}$. Similarly, Seg_T is trained to segment the target domain images using the translated source domain images $\{\hat{x}^T, y^s\}$. The loss function being minimized during the training of both networks is a linear combination of the Dice coefficient and cross-entropy loss, as presented in Eq. 4.

$$L_{\text{seg}}(\text{Seg}_S, \text{Seg}_T, G_S) = \beta \sum_c \alpha_c \left(1 - \frac{2 \sum_i \hat{y}_{ic} y_{ic}}{\sum_i \hat{y}_{ic} + \sum_i y_{ic}} \right) - (1 - \beta) \sum_c \sum_i (y_{ic} \log(\hat{y}_{ic})) \quad (4)$$

where y_{ic} and \hat{y}_{ic} are the ground-truth label and the predicted probability of pixel i being part of class c , respectively. α_c is a weight parameter for class c in the dice loss, and β a weight parameter for the dice loss. Since the loss of Seg_T depends on the geometric invariance during the translation from source to target domain, this network encourages G_S to preserve the anatomical structure of the region being segmented.

Finally, Seg_S and Seg_T are also trained in an unsupervised manner using a pseudo segmentation loss. The target domain set $\{x^T\}$ and translated target domain set $\{\hat{x}^S\}$ are sent to Seg_T and Seg_S , respectively, to produce the target domain predicted segmentation \hat{y}_T and source domain predicted segmentation \hat{y}_S . Since Seg_S is being trained with the labeled source domain dataset, \hat{y}_S provides a better estimation of the ground truth segmentation. Hence \hat{y}_S is considered the pseudo ‘‘ground truth’’ segmentation, and \hat{y}_T the predicted segmentation. The difference between \hat{y}_S and \hat{y}_T is penalized using the same segmentation loss function from equation (4). Moreover, to incentivize the segmentations from both domains to be consistent, an L1 loss between \hat{y}_T and \hat{y}_S is also added. The pseudo segmentation loss is displayed in equation (5). Both losses force the predicted segmentations from the original and translated images to be the same. Therefore, incentivizing G_T to preserve the anatomical structure information during translation and allowing Seg_S and

Seg_T to recognize the same regions.

$$L_{\text{pseudoseg}}(\text{Seg}_S, \text{Seg}_T, G_T) = \beta \sum_c \alpha_c \left(1 - \frac{2 \sum_i \hat{y}_{icT} \hat{y}_{icS}}{\sum_i \hat{y}_{icT} + \sum_i \hat{y}_{icS}} \right) - (1 - \beta) \sum_c \sum_i (\hat{y}_{ic} \log(\hat{y}_{ic})) + \|\hat{y}_T - \hat{y}_S\| \quad (5)$$

The full objective functions being optimized in the CycleGAN based method is composed of six loss function as presented in equation (6), where λ_i is the weight for term i .

$$L_{\text{adv}}(G_S, G_T, D_S, D_T, \text{Seg}_S, \text{Seg}_T) = \mathcal{L}_{\text{GAN}}(G_S, D_S) + \mathcal{L}_{\text{GAN}}(G_T, D_T) + \lambda_1 L_{\text{cyc}}(G_S, G_T) + \lambda_2 L_{\text{identity}}(G_S, G_T) + \lambda_3 L_{\text{seg}}(\text{Seg}_S, \text{Seg}_T, G_S) + \lambda_4 L_{\text{pseudoseg}}(\text{Seg}_S, \text{Seg}_T, G_T) \quad (6)$$

Following the recommendations of Ref.²⁰, in all experiments the lambdas are set to $\lambda_1 = 10$, $\lambda_2 = 2.5$. Furthermore, in the segmentation losses, the lambdas are set to $\lambda_3 = 0.1$ and $\lambda_4 = 0.1$. The model is trained alternatively by first fixing D_S and D_T , and training $G_S, G_T, \text{Seg}_S, \text{Seg}_T$. Afterwards, the generators and

segmentors are fixed, and the discriminators updated. The segmentation networks have a U-Net architecture.⁴⁰

Step 2: feature-level adaptation

In Step 2, a feature-level adaptation method in the semantic prediction space is implemented by using an adversarial training scheme. This phase is especially necessary when there is a severe domain gap between the target and source images. After finishing Step 1, the G_S network is utilized to translate the source domain images $\{x^s, y^s\}$ to target domain $\{\hat{x}^T, y^s\}$. Then, a segmentation network G takes the role of a generator and receives as input the mapped source domain images \hat{x}^T and target domain images x^T and produces their predicted segmentations $G(\hat{x}^T) = \hat{y}^{\hat{T}}$ and $G(x^T) = \hat{y}^T$. A discriminator D is fed the two predictions $\hat{y}^{\hat{T}}$ and \hat{y}^T , and is trained to discriminate between the two. Meanwhile, G aims to fool the discriminator by learning domain invariant features and producing segmentations that follow the same the distribution. The loss function being optimized is presented in equation (7), where G looks to minimize it and D to maximize it.

$$L_{\text{FAGAN}}(G, D) = \mathbb{E}_{x^T \sim T} [\log D_s(\hat{y}^{\hat{T}})] + \mathbb{E}_{x^s \sim S} [\log(1 - D_s(\hat{y}^T))] \quad (7)$$

As with the previous step, a least squares objective function is implemented to stabilize training. The generator network has a U-Net architecture and the discriminator a PatchGAN

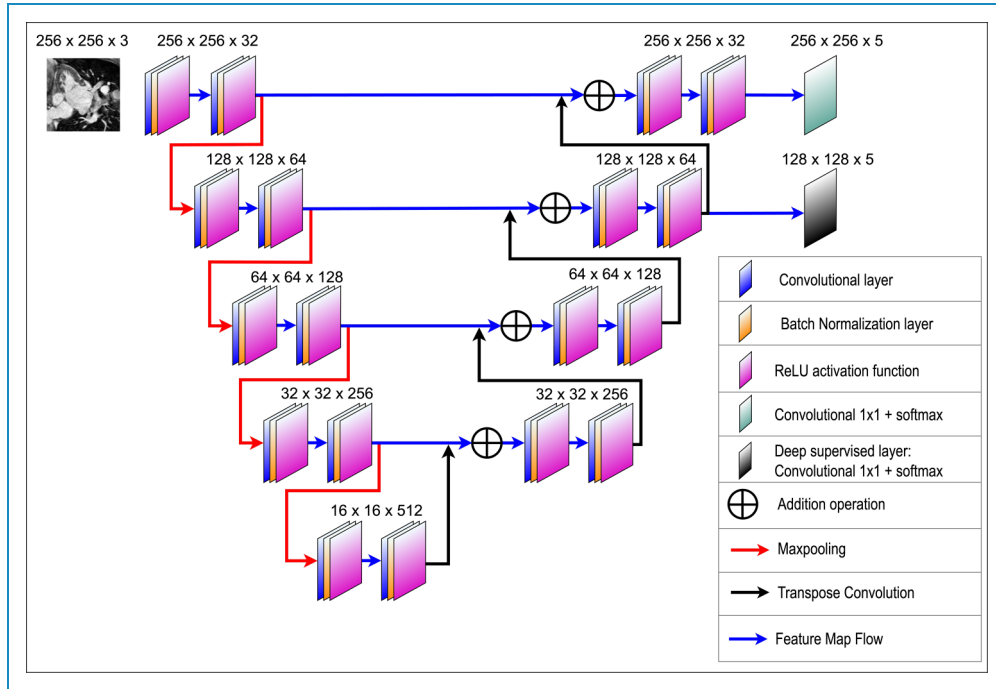


Figure 2. Deeply supervised U-Net network trained in Step 2 to segment the unlabeled target domain images. The numbers over the convolutional blocks correspond to the height, width, and number of feature maps.

structure. To help the U-Net learn rich hierarchical features, a deep supervised layer with an auxiliary segmentation loss⁴¹ is added in the second-last up-sampling block. In Figure 2, the architecture of the implemented U-Net is presented. Furthermore, to increase the adaptation in the low-level feature space, an auxiliary discriminator network D_{aux} is added to the system. Let \hat{y}_{aux}^T be the deep supervised layer predicted segmentation from \hat{x}^T , and \hat{y}_{aux}^T the deep supervised layer predicted segmentation from x^T . D_{aux} is trained to discriminate between \hat{y}_{aux}^T and \hat{y}_{aux}^T by maximizing equation (7). D_{aux} has a PatchGAN architecture.

In addition to the adversarial training, on each iteration the U-Net is also trained in a supervised manner to segment the mapped source domain images $\{\hat{x}^T, y^s\}$. The loss function minimized in this step is a linear combination of the dice loss and cross entropy loss as shown in equation 8, where L_{seg} refers to the segmentation loss in equation 4, and λ_{aux} to the weight for the auxiliary segmentation loss of the deep supervised layer. In our work, λ_{aux} is set to 0.1.

$$L_{\text{FAseg}}(G) = \sum_{n=1}^{N_c} L_{\text{seg}}(y^s, \hat{y}^T) + \lambda_{\text{aux}} L_{\text{seg}}(y^s, \hat{y}_{\text{aux}}^T) \quad (8)$$

The training is done by first fixing D and training G in an adversarial and supervised manner, and then setting G and training D . During inference, only the U-Net architecture trained in this step is used to produce the predicted segmentations in target domain.

Monitoring validation metric

Since there are no labeled target domain images, it is a challenge to select the best weights for testing. Hence, we propose a pseudo validation loss function based on the segmentation area and dice coefficient to monitor the performance of the network on the target dataset. First, the ground truth segmentations from the source domain are used to calculate the average number of pixels per class c and slice (AvgPix_c) as shown in equation (9).

$$\text{AvgPix}_c = \frac{\sum_{n_s} \sum_i y_{ic}^s}{n_s} \quad (9)$$

where n_s are the number of slices on the source domain, and y_{ic}^s the pixels on the source domain's ground truth that have a value of 1 in class c . Since the morphology of the anatomical structures should be consistent across the different imaging modalities, AvgPix_c is a good approximator of the average number of pixels from class c that should be segmented on the target domain. Therefore, to validate the network during training, the average number of pixels predicted to be part of class c on the target domain segmentation are calculated and divided by AvgPix_c . If the structures being segmented on the target domain are similar in area to the structures on the source domain's ground truth,

this metric should have a value close to 1. On the other hand, if the network is under-segmenting the ROIs, the value of the metric should be less than 1, while over-segmenting the region would receive a value more than 1. Moreover, to include information about the segmentation performance of the network, in the function we add the dice loss calculated on the translated source domain images. The weights that minimize the function displayed in equation (10) are selected for testing, where n_T are the number of slices in the in the target domain, and \hat{y}_{ic}^T the predicted pixels in the target domain that correspond to class c .

$$\begin{aligned} \text{Pseudo validation loss} = \sum_c \left| 1 - \frac{\sum_{n_T} \sum_i \hat{y}_{ic}^T}{n_T \text{AvgPix}_c} \right| \\ + \sum_c \left(1 - \frac{2 \sum_i \hat{y}_{ic}^s y_{ic}^s}{\sum_i \hat{y}_{ic}^s + \sum_i y_{ic}^s} \right) \end{aligned} \quad (10)$$

Dataset and preprocessing

The proposed StAC-DA is evaluated on the task of bidirectional segmentation of cardiac structures on MRI and CT imaging modalities from the Multi-Modality Whole Heart Segmentation²⁵ (MMWHS) dataset. The MMWHS dataset consists of 20 MRI and 20 CT whole cardiac volumes. The MRI and CT images are unpaired and collected from distinct patient cohorts. For both modalities, the ground truth segmentations for the ascending aorta (AA), left atrium blood cavity (LA-blood), left ventricle blood cavity (LV-blood), and myocardium of the left ventricle (LV-my) are provided. Following the work of,²⁴ the images of 16 subjects are used for training, and images from 4 subject for testing. Furthermore, the preprocessing proposed by Ref.¹⁴ is used, where the central heart region of the image is first cropped to size of 256×256 pixels. Then, a histogram filtering at percentile 2 and 98 is applied. Finally, the pixel values are normalized to zero mean and unit variance.

Training details

We perform two types of experiments. First, we evaluate the effectiveness of the method by performing the adaptation from the MRI to the CT domain (MRI \rightarrow CT). Then, the adaptation is done from the CT to MRI domain (MRI \rightarrow CT). In both experiments, Step 1 is run for 200 epochs with $\beta = 0.25$ and $\alpha_c = 0.20$ on the L_{seg} and $L_{\text{pseudoseg}}$ loss functions. The generators, discriminators, and segmentation networks are optimized with the Adam optimizer with beta 1:0.5 and beta 2: 0.999, a learning rate of 5×10^{-5} , weight decay of 2×10^{-4} , and batch size of 5. Furthermore, to improve the convergence speed of

the CycleGAN based model, a U-Net network is pretrained with the source domain images on a supervised manner for 200 epochs and the weights uploaded to the Seg_S segmentation network. In Figure 3, we present the qualitative results from the proposed structure-aware translation model from step 1 and the CycleGAN model. The figure demonstrates that the proposed network successfully translates the images from the source to the target domain while preserving the shape and characteristics of the anatomical structures. In contrast, the CycleGAN model tends to produce blurry and low-contrast translations where the shape of the anatomical regions is altered.

In Step 2, the U-Net model is trained for 100 epochs. The discriminators and U-Net network are optimized with the Adam optimizer and a learning rate of 1×10^{-5} . Similar to Step 1, an $\beta = 0.25$ and $\alpha_c = 0.20$ are used on the $L_{F\text{Aseg}}$ loss function. Finally, in this step, real-time data augmentation is employed to improve the generalization of the model. Specifically, random rotation, horizontal flipping, and vertical flipping are applied with a 0.5 probability. StAC-DA is implemented using Python 3.8 and the Pytorch framework. The experiments are carried out on a DGX station with 4 V100 GPUs with 32 GB, 2.2 GHz CPU, and 256 GB RAM.

Evaluation metrics

The Dice similarity coefficient (Dice) and average symmetric surface distance metric (ASSD) are employed to quantitatively evaluate the segmentation performance of the models. The Dice coefficient is an overlap-based metric that measures the intersection between the ground truth segmentation and predicted segmentation. The ASSD distance is a spatial-based metric that calculates the average of all distances between points on the ground truth's boundary surface to points on the predicted segmentation's boundary surface. Let Y be the ground truth segmentation, Y_1 the pixels from the ground truth that are part of the foreground, and Y_0 the pixels from the background. Moreover, let \hat{Y} be the predicted segmentation, \hat{Y}^1 the pixels predicted to be part of the foreground, and \hat{Y}^0 the pixels from the background. The Dice coefficient is calculated as shown in equation (11), and the ASSD as presented in equation (12). The evaluation is performed on the subject-level segmentation volume.

$$\text{Dice}(Y, \hat{Y}) = \frac{2|Y^1 \cap \hat{Y}^1|}{|Y^1| + |\hat{Y}^1|} \quad (11)$$

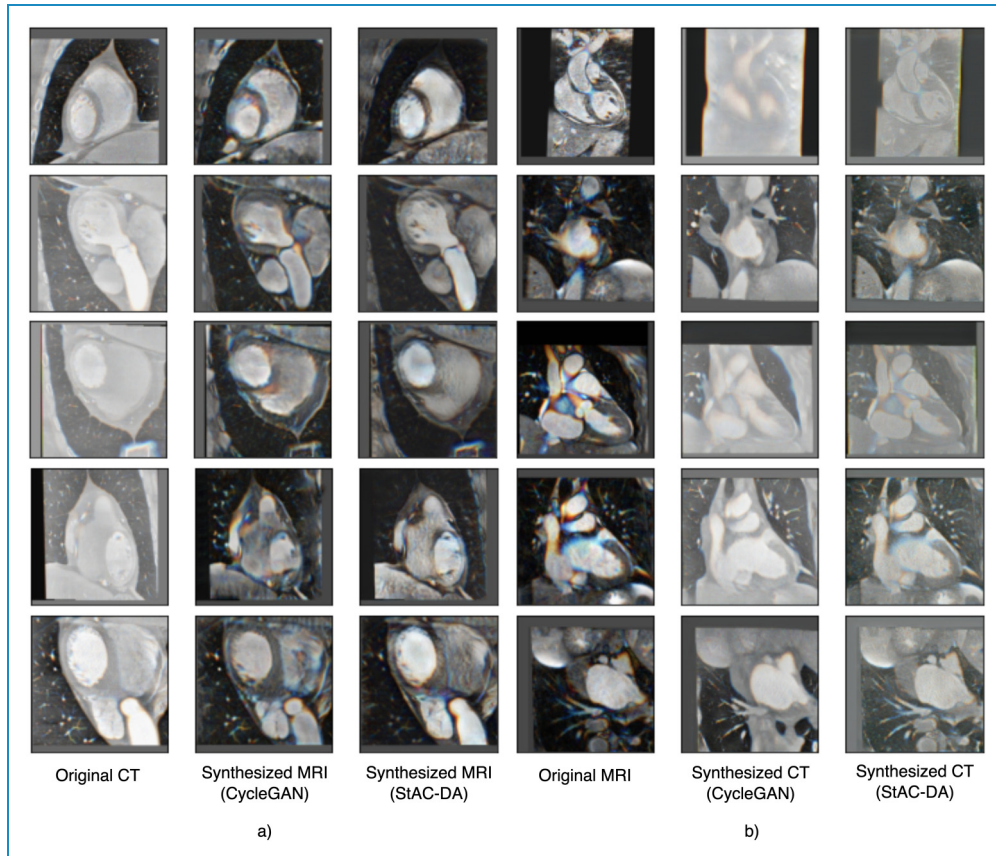


Figure 3. Examples of the translation results using the CycleGAN model and structure-aware model from Step 1. (a) Adaptation from CT to MRI domain. (b) Adaptation from MRI to CT domain. The proposed structure-aware model successfully translates the images between domains while keeping the structure of the anatomical region.

$$\text{ASSD}(Y, \hat{Y}) = \frac{1}{|Y||\hat{Y}|} \left(\sum_{y \in Y^1} \min_{\hat{y} \in \hat{Y}^1} \|y - \hat{y}\| + \sum_{\hat{y} \in \hat{Y}^1} \min_{y \in Y^1} \|y - \hat{y}\| \right) \quad (12)$$

Results

Ablation studies

The proposed StAC-DA model is composed of a structure-aware CycleGAN-based model in Step 1 and a segmentation network with feature-level alignment in Step 2. In this section, we perform ablation studies to evaluate the contribution of each of these steps. For this objective, we train four models for the adaptation from MRI to CT domain. The results are displayed in Table 1. In the first model, we implement the CycleGAN model²⁰ to translate the images from MRI to CT modality, and train a U-Net network on the translated CT images in a supervised manner. In the second model, we apply the proposed structure-aware CycleGAN-based model for image translation and use the trained Seg_T network for the segmentation of the target domain images. This model is denoted as “Step1 StAC-DA.” In the third model, we train the structure-aware CycleGAN-based model to synthesize CT images from MRI images and use the translated CT images to train a U-Net network in a supervised manner. This model is designated as “Step1 StAC-DA + U-Net.” In the last model, we test the proposed StAC-DA model where we add the adversarial training for feature alignment to the previously described “Step1 StAC-DA + U-Net.”

The results demonstrate that including the structure-aware module in the CycleGAN network helps translate the overall shape of the heart better as the segmentation network is able to identify and segment the different cardiac regions with higher accuracy. Particularly, “Step 1 StAC-DA” and “Step 1 StAC-DA + U-Net” have an important improvement over CycleGAN in the segmentation of ascending aorta, left ventricle blood cavity, and myocardium of the left ventricle when considering Dice

and ASSD metrics. The experiments also show that training a segmentation network with the synthesized images obtained from the structure-aware CycleGAN-based network increases the performance in comparison to using the Seg_T network obtained in Step 1. This behavior can be attributed to the complexity of the optimization problem being solved in the CycleGAN-based network. Specifically, Seg_T is trained with the other five networks while optimizing six objective functions that not only aim to improve segmentation performance but also image translation. Meanwhile, the U-Net network is trained exclusively to minimize the segmentation loss which seems to enhance its segmentation capability. Finally, the results also prove that including the feature-level adaptation through the discriminators networks further boosts performance.

Benchmark results

Our method is compared against six leading unsupervised domain adaptation networks: U-GAT-IT model,⁴² PnP-Ada-Net model,¹¹ SynSeg-Net,⁴³ AdaOutput,¹⁵ Cycada,¹³ and SIFA model.¹⁴ Moreover, we provide a performance upper bound to measure the performance gap by training the U-Net network in a supervised manner on the target domain (denoted as “UpperB U-Net”). To evidence the domain shift, a performance lower bound is also presented (denoted as “LowerB U-Net”) by using the U-Net trained on the source domain to predict the segmentation on the target domain images without any adaptation method. Finally, the results obtained after applying only Step 1 of the proposed framework and using the Seg_T U-Net segmentation network for the segmentation on the target domain are also shown as “Step 1 StAC-DA.” In Table 2 the results from the MRI → CT adaptation are presented, while in Table 3 the results from the CT → MRI adaptation. As can be seen from the quantitative evaluation, the proposed model outperforms other models in the segmentation of the ascending aorta (AA). In the MRI → CT adaptation, StAC-DA is ranked first in terms

Table 1. Ablation study of the components of StAC-DA.

Method	AA		LA-blood		LV-blood		LV-myocardium	
	Dice	ASSD	Dice	ASSD	Dice	ASSD	Dice	ASSD
CycleGAN	0.57	19.63	0.66	15.98	0.45	12.46	0.31	9.58
Step 1 StAC-DA	0.71	18.48	0.64	13.37	0.55	11.84	0.43	8.91
Step 1 StAC-DA + U-Net	0.73	7.67	0.59	13.97	0.68	7.25	0.50	7.67
StAC-DA (proposed)	0.82	12.84	0.63	13.25	0.73	5.15	0.53	7.25

In bold is the performance of the proposed model.

Table 2. Performance comparison of the proposed method (StAC-DA) and leading unsupervised domain adaptation methods for cardiac structure segmentation from MRI to CT domain (MRI \rightarrow CT).

Method	AA		LA-blood		LV-blood		LV-myo	
	Dice	ASSD	Dice	ASSD	Dice	ASSD	Dice	ASSD
UpperB U-Net	0.83	2.3	0.92	1.93	0.92	1.7	0.85	1.8
LowerB U-Net	0.13	23.0	0.01	65.6	0	70.4	0	57.7
U-GAT-IT	0.68	12.0	0.66	13.7	0.55	8.9	0.39	8.9
Pnp-Ada-Net	0.74	12.8	0.69	6.3	0.62	17.4	0.51	14.7
SynSeg-Net	0.72	11.7	0.69	7.8	0.52	7.0	0.41	9.2
AdaOutput	0.65	17.9	0.76	5.5	0.54	5.9	0.44	8.9
Cycada	0.73	9.6	0.77	8.0	0.62	9.6	0.45	10.5
SIFA	0.81	7.9	0.80	6.2	0.74	5.5	0.62	8.5
Step 1 StAC-DA	0.71	18.5	0.64	13.4	0.55	11.8	0.43	8.9
StAC-DA (our model)	0.82	12.8	0.63	13.2	0.73	5.2	0.53	7.3

Note. The values presented for competing models are as reported by Ref.¹⁴ in the published paper. Values in bold represent the best performance.

Table 3. Performance comparison of the proposed method (StAC-DA) and leading unsupervised domain adaptation methods for cardiac structure segmentation from CT to MRI domain (CT \rightarrow MRI).

Method	AA		LA-blood		LV-blood		LV-myo	
	Dice	ASSD	Dice	ASSD	Dice	ASSD	Dice	ASSD
UpperB U-Net	0.85	3.3	0.87	1.83	0.93	1.2	0.81	1.7
LowerB U-Net	0	50	0	45.4	0	48.8	0	39.3
U-GAT-IT	0.55	16.5	0.39	12.1	0.69	7.6	0.49	7.0
Pnp-Ada-Ne	0.44	11.4	0.47	14.5	0.78	4.5	0.49	5.3
SynSeg-Net	0.41	8.6	0.58	10.7	0.64	5.4	0.37	5.9
AdaOutput	0.61	5.7	0.40	8.0	0.72	4.6	0.36	4.6
Cycada	0.61	7.7	0.44	13.9	0.78	4.8	0.48	5.2
SIFA	0.65	7.3	0.62	7.4	0.79	3.8	0.47	4.4
Step 1 StAC-DA	0.54	10.2	0.40	14.3	0.50	9.6	0.18	9.2
StAC-DA (our model)	0.68	7.4	0.39	11.5	0.68	8.4	0.44	7.8

Note. The values presented for competing models are as reported by Ref.¹⁴ in the published paper. Values in bold represent the best performance.

of the Dice score and is only 0.01 below the U-Net trained in a supervised manner on the target domain. Similarly, using the Dice score and ASSD metric, the proposed model is ranked first and third, respectively, in the AA segmentation on the CT \rightarrow MRI adaptation. In the segmentation of the left ventricle blood cavity on the MRI \rightarrow CT domain adaptation, StAC-Da is also first in terms of ASSD and second in relation to the Dice score. Finally, in the segmentation of the left ventricle myocardium on the MRI \rightarrow CT domain adaptation, we are ranked second in relation to the Dice score and first in ASSD. It is interesting to see that the Seg_T segmentation network obtained from Step 1 produces competitive segmentations on the target domain. For example, in the segmentation of the AA on the MRI \rightarrow CT adaptation, the Seg_T obtains a Dice score of 0.71, which outperforms the U-GAT-IT and AdaOutput model. However, as discussed in Sec. 4.4 this adaptation method is not sufficient by itself to produce the best-performing segmentations. The feature-level adaptation proposed in Step 2 is indispensable to improve the segmentation performance on all cardiac substructures. Another interesting observation is the degradation in performance all models suffer when adapting from the CT \rightarrow MRI domain. MRIs usually offer greater contrast, image clarity, and detail than CT imaging.⁴⁴ Hence, when reconstructing MRIs from CT some information is lost, which reduces the accuracy of the segmentation models.

In Figures 4 and 5 a qualitative evaluation of the proposed model is shown on the MRI \rightarrow CT and CT \rightarrow MRI adaptation respectively. Furthermore, the segmentation achieved with the U-Net without adaptation and the Seg_T network after Step 1 is also included for comparison. StAC-DA produces good segmentations that considerably reduce the performance degradation evidenced in the predictions of the U-Net without adaptation. Step 1 from the StAC-DA framework achieves an acceptable performance; however, in some slices where the domain shift is severe, the model does not correctly identify the regions of interest. The feature-level adaptation implemented in Step 2 helps improve the segmentation performance in these slices.

Computational requirements

StAC-DA is designed sequentially to reduce computational requirements during training. In Table 4, we present the GPU and RAM memory used in each step during training, using a batch size of 1, as well as the number of trainable parameters. For comparison, we have also included the computational requirements of the unified model proposed by Chen et al.¹⁴ In terms of GPU and RAM usage, the entire StAC-DA framework has lower requirements. Furthermore, as each step is trained independently, the model can be trained in facilities with less powerful GPUs and RAM memory. Regarding the number of trainable parameters,

SIFA is smaller. Nevertheless, since the entire StAC-DA framework is not processed simultaneously, it is unnecessary to save all the models in Steps 1 and 2. Specifically, after training Step 1, only the Generator network is needed in Step 2 to transfer the images from the source to target domain. The Generator network has 7.8×10^6 trainable parameters. Lastly, after Step 2, only the Segmentation network is needed for inference, which has 6.9×10^6 parameters. Hence, for deployment, only the Segmentation network needs to be kept. While it is advantageous to train only once a unified model as in Chen et al.,¹⁴ scenarios with limited computational resources for training can prohibit its application.

Discussion

Deep learning models have been shown to excel in various complex tasks when large amounts of data are available. Nevertheless, when the networks are tested in data that does not follow the training distribution, their performance can significantly degrade. This challenge is of special interest in the medical imaging community, where labeling images is very costly, and due to the different imaging modalities and acquisition protocols, the testing data can differ significantly from the training set. Developing unsupervised cross-modality models is advantageous for the application of AI in medical settings because it decreases the need to obtain costly labeled data while exploiting to the fullest the unlabeled data from the target domain. In this work, we presented StAC-DA, an unsupervised structure-aware domain adaptation framework for cross-modality medical image segmentation. The proposed framework is comprised of a two-step image and feature-level adaptation that importantly reduces the performance degradation when moving from two different imaging modalities.

StAC-DA has been tested on two domain adaptation tasks from a publicly available cross-modality cardiac segmentation challenge. We implemented the proposed framework for domain adaptation from MRI to CT imaging modality (MRI \rightarrow CT), as well as from CT to MRI modality (CT \rightarrow MRI). As shown in Tables 2 and 3, the segmentation network without any domain adaptation method suffered from a severe domain shift. On average, the U-Net without adaptation had a 0.035 dice score when moving from MRI \rightarrow CT domain, and a 0 dice score when translating between CT \rightarrow MRI domain. After applying the proposed image and feature-level adaptation framework, the performance improved considerably. StAC-DA attained an average 0.68 dice score when translating between MRI \rightarrow CT domain, and an average 0.55 dice score between the CT \rightarrow MRI domain. A specially good performance is achieved in the segmentation of the ascending aorta on the MRI \rightarrow CT domain adaptation, where StAC-DA got a 0.82 dice score that is only 0.01 lower

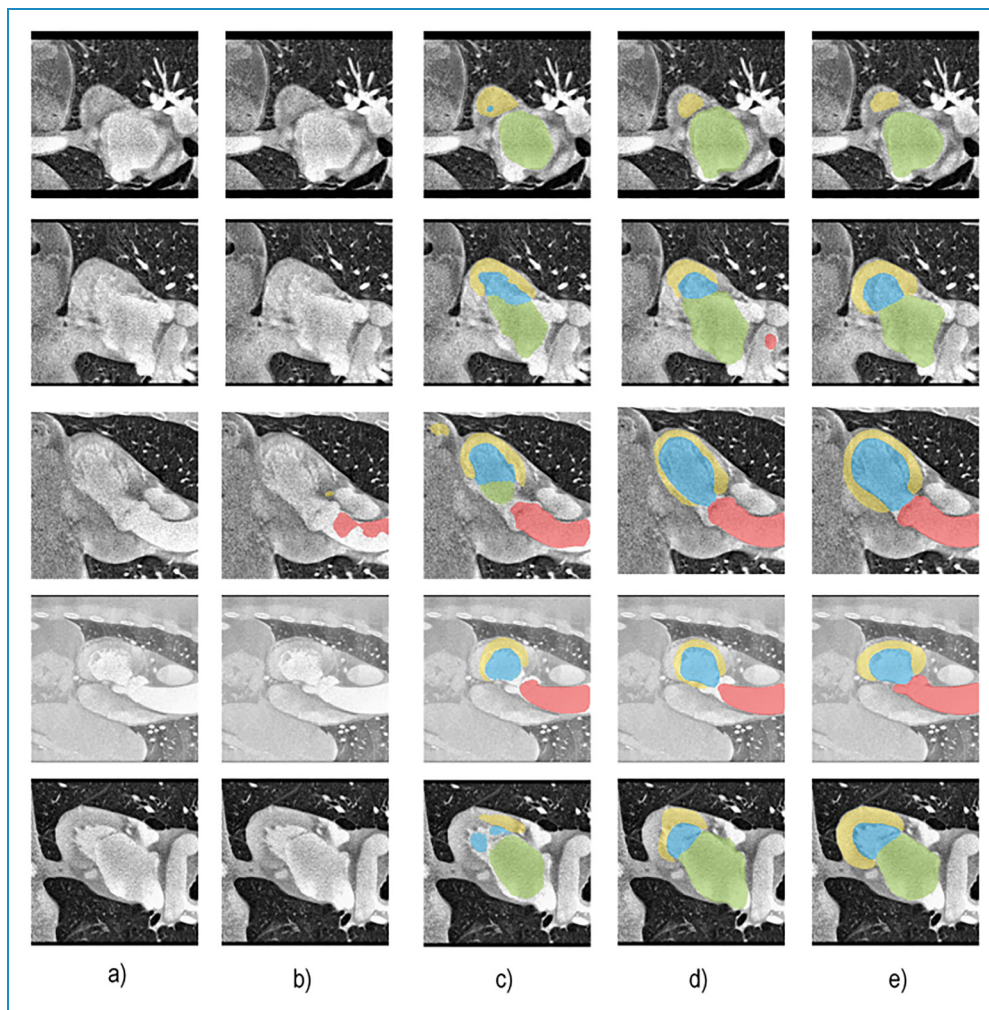


Figure 4. Segmentation results after the MRI \rightarrow CT adaptation. (a) Original CT image. (b) Segmentation results of the U-Net without adaptation. (c) Segmentation results after applying Step 1 of the proposed framework. (d) Segmentation results of the proposed StAC-DA. (e) Ground truth segmentation.

than the dice score achieved by the segmentation network trained in a supervised manner with the labeled target domain images. In comparison with the leading methods in UDA, StAC-DA had a competitive performance being ranked first in the segmentation of the ascending aorta.

An interesting observation is that although all competing models seem to reduce the domain shift, there is still a performance gap, particularly in the CT \rightarrow MRI domain adaptation task. This behavior can be due to the characteristics of the imaging modalities. The left ventricle blood cavity and left ventricle myocardium had a low contrast to the surrounding tissues on the CT imaging, hence when transferring to the MRI domain the contrast information could not be recovered. This is an important point to consider when developing cross-modality models, to obtain a good result the source domain images must have a good quality and contrast so that the information is kept during translation.

A limitation of the proposed framework is applying 2D CNNs for image translation and segmentation. Although 2D CNNs are able to capture intra-slice information, they do not fully exploit volumetric information. Hence, some important 3D features that can boost performance might be missed. Since training the 2D structure-aware CycleGAN-based model is already computationally and time intensive, a future direction can be using a 3D segmentation CNN for Step 2 or using 3D patches instead of the whole 3D image for Steps 1 and 2. Furthermore, when utilizing the monitoring validation metric we are assuming the morphology of anatomical structures is consistent across the imaging modalities. However, if the population in the source domain is considerably different from that of the target domain the metric might not be applicable. Lastly, although the proposed framework aims to reduce the computational cost by dividing the training process into two steps, the memory usage can still be limiting with large

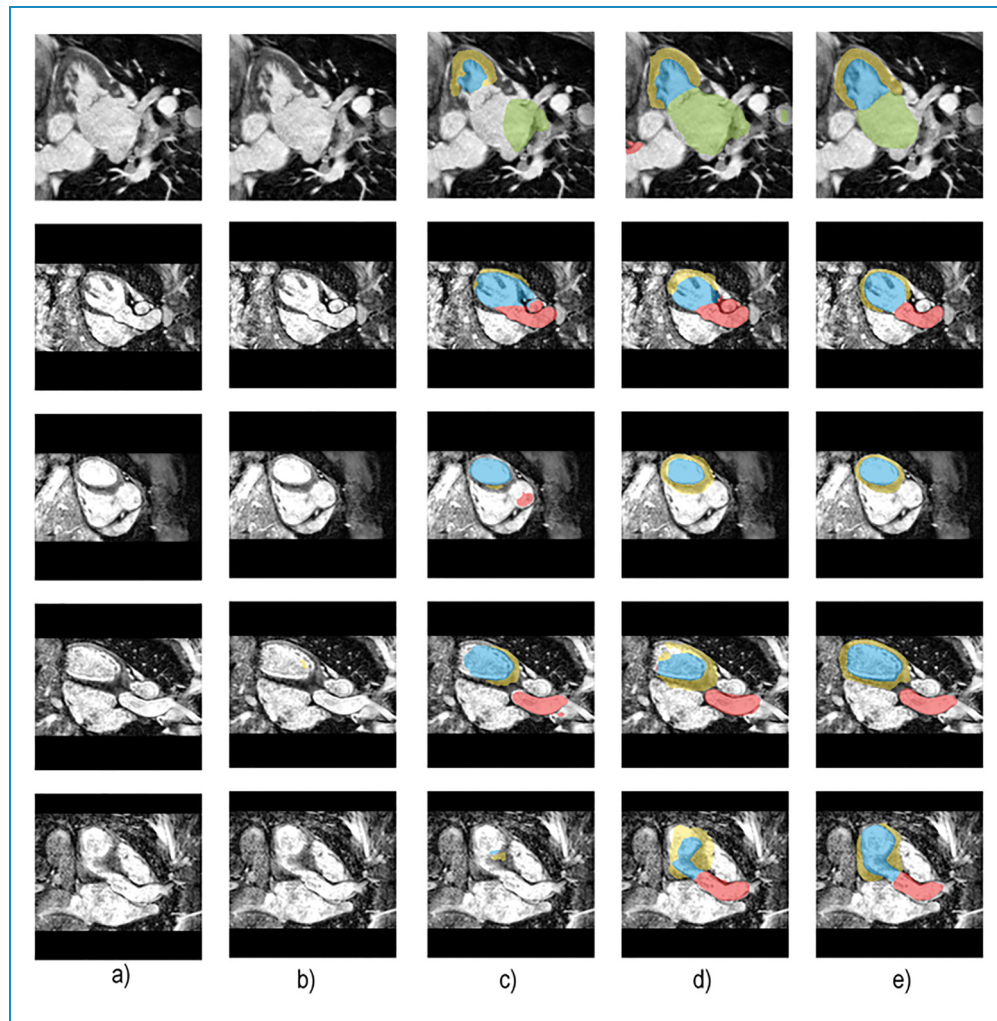


Figure 5. Segmentation results after the CT \rightarrow MRI adaptation. (a) Original CT image. (b) Segmentation results of the U-Net without adaptation. (c) Segmentation results after applying Step 1 of the proposed framework. (d) Segmentation results of the proposed StAC-DA. (e) Ground truth segmentation.

Table 4. Computational requirements of StAC-DA and SIFA during training with a batch size of 1.

Method	GPU usage (MiB)	RAM usage (MiB)	Number of trainable parameters
StAC-DA (Step 1)	3432	3293	35.1×10^6
StAC-DA (Step 2)	1458	3259	12.5×10^6
StAC-DA (total)	4890	6552	47.6×10^6
SIFA	5900	9694	43.3×10^6

datasets, high-resolution images or more complex segmentation tasks. To address these issues, future work could explore optimization techniques like model pruning,

distributed training strategies, and scalable architectures, potentially broadening the framework’s applicability.

Conclusion

In this work, we present StAC-DA, an unsupervised structure-aware cross-modality domain adaptation framework for medical image segmentation. StAC-DA is composed of two sequential steps. First performs a structure-aware image-level adaptation, where images from the source domain are mapped to the target domain through a CycleGAN-based model. The latter includes a segmentation network that preserves the anatomical structures during translation. In the second step, a feature-level adaptation is applied by training a deeply supervised U-Net architecture in an adversarial manner to produce probable segmentations for the target domain. StAC-DA is evaluated on the task of bidirectional

cardiac substructure segmentation from the Multi-Modality Whole Heart Segmentation Challenge dataset. The experiments demonstrate that the proposed model has a very competitive performance, being ranked first in the segmentation of the ascending aorta when adapting from MRI to CT domain and from the CT to MRI domain.

Acknowledgments: This work was supported in part by POLIGRANT No. 17376, Colegio de Ciencias e Ingenierías, USFQ. The authors thank to the Applied Signal Processing and Machine Learning Research Group of USFQ for providing the computing infrastructure (NVIDIA DGX workstation) to implement and execute the developed source code.

Data availability statement: The dataset used in this study is publicly available in MMWHS challenge's website <https://zmiclab.github.io/zxh/0/mmwhs/>.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article. This work was supported by the Universidad San Francisco de Quito (grant number: 17376).

Guarantor: MB-C

ORCID iD: Maria Baldeon-Calisto  <https://orcid.org/0000-0001-9379-8151>

References

- Gour M and Jain S. Uncertainty-aware convolutional neural network for COVID-19 X-ray images classification. *Comput Biol Med* 2022; 140: 105047.
- Sasibhooshan R, Kumaraswamy S and Sasidharan S. Wavnet—visual saliency detection using discrete wavelet convolutional neural network. *J Vis Commun Image Represent* 2021; 79: 103236.
- Baldeon-Calisto M and Lai-Yuen S. EMONAS-Net: efficient multiobjective neural architecture search using surrogate-assisted evolutionary algorithm for 3D medical image segmentation. *Artif Intell Med* 2021; 119: 102154.
- Li Y, Yang J, Ni J, et al. TA-Net: triple attention network for medical image segmentation. *Comput Biol Med* 2021; 137: 104836.
- Calisto MB and Lai-Yuen SK. AdaEn-Net: an ensemble of adaptive 2D–3D fully convolutional networks for medical image segmentation. *Neural Netw* 2020; 126: 76–94.
- Baldeon-Calisto M and SK L-Y. AdaResU-Net: multiobjective adaptive convolutional neural network for medical image segmentation. *Neurocomputing* 2020; 392: 325–340.
- Choudhary A, Tong L, Zhu Y, et al. Advancing medical imaging informatics by deep learning-based domain adaptation. *Yearb Med Inform* 2020; 29: 129–138.
- Guan H and Liu M. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering* 2021; 69(3): 1173–1185.
- Torralba A and Efros AA. Unbiased look at dataset bias. In: *CVPR 2011*, Colorado Springs, CO, USA, 2011, pp.1521–1528. IEEE.
- Dai C, Mo Y, Angelini E, et al. Transfer learning from partial annotations for whole brain segmentation. In: Wang Q, Milletari F and Nguyen HV, et al. (eds) *Domain adaptation and representation transfer and medical image learning with less labels and imperfect data*. Cham: Springer International Publishing, 2019, pp.199–206.
- Dou Q, Ouyang C, Chen C, et al. PnP-AdaNet: plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation. *arXiv:181207907 [cs]*, <http://arxiv.org/abs/1812.07907> (2018, accessed 18 August 2021).
- Zhao C, Carass A, Lee J, et al. Whole brain segmentation and labeling from CT using synthetic MR images. In: *International workshop on machine learning in medical imaging*, Quebec City, QC, Canada: Springer, 2017, pp.291–298.
- Hoffman J, Tzeng E, Park T, et al. CyCADA: cycle-consistent adversarial domain adaptation. In: *International Conference on Machine Learning*, Stockholm, Sweden, 10–15 July 2018, pp.1989–1998. Cambridge, MA, USA: PMLR.
- Chen C, Dou Q, Chen H, et al. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans Med Imaging* 2020; 39: 2494–2505.
- Tsai Y-H, Hung W-C, Schuster S, et al. Learning to adapt structured output space for semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, USA, June 2018, pp.7472–7481. New York City, New York, USA: IEEE.
- Liu Y, Zhang W and Wang J. Source-free domain adaptation for semantic segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Nashville, TN, USA, 20–25 June 2021, pp.1215–1224. New York City, New York, USA: IEEE.
- Fan L, Gong X and Guo Y. General multiscenario ultrasound image tumor diagnosis method based on unsupervised domain adaptation. *Ultrasound Med Biol* 2023; 49: 2291–2301.
- Gao Y, Zhang Y, Cao Z, et al. Decoding brain states from fMRI signals by using unsupervised domain adaptation. *IEEE J Biomed Health Inform* 2020; 24: 1677–1685.
- Zhang J, Liu M, Pan Y, et al. Unsupervised conditional consensus adversarial network for brain disease identification with structural MRI. In: Suk H-I, Liu M and Yan P, et al. (eds) *Machine learning in medical imaging*. Cham: Springer International Publishing, 2019, pp.391–399.
- Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp.2242–2251.
- Gademayr M, Gupta L, Appel V, et al. Generative adversarial networks for facilitating stain-independent supervised and

- unsupervised segmentation: a study on kidney histology. *IEEE Trans Med Imaging* 2019; 38: 2293–2302.
22. Jue J, Jason H, Neelam T, et al. Integrating cross-modality hallucinated MRI with CT to aid mediastinal lung tumor segmentation. In: Shen D, Liu T and Peters TM, et al. (eds) *Medical image computing and computer assisted intervention – MICCAI 2019*. Cham: Springer International Publishing, 2019, pp.221–229.
 23. Cai J, Zhang Z, Cui L, et al. Towards cross-modal organ translation and segmentation: a cycle- and shape-consistent generative adversarial network. *Med Image Anal* 2019; 52: 174–184.
 24. Tomar D, Lortkipanidze M, Vray G, et al. Self-attentive spatial adaptive normalization for cross-modality domain adaptation. *IEEE Transactions on Medical Imaging* 2021; 40: 2926–2938.
 25. Zhuang X and Shen J. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Med Image Anal* 2016; 31: 77–87.
 26. Bousmalis K, Trigeorgis G, Silberman N, et al. Domain Separation Networks. In: *Advances in neural information processing systems*. Barcelona, Spain: Curran Associates, Inc., 2016. <https://proceedings.neurips.cc/paper/2016/hash/45fbc6d3e05ebd93369ce542e8f2322d-Abstract.html> (accessed 24 February 2022).
 27. Sun B and Saenko K. Deep CORAL: correlation alignment for deep domain adaptation. In: Hua G and Jégou H (eds) *Computer vision – ECCV 2016 workshops*. Cham: Springer International Publishing, 2016, pp.443–450.
 28. Ganin Y, Ustinova E, Ajakan H, et al. Domain-Adversarial training of neural networks. In: Csurka G (ed) *Domain adaptation in computer vision applications*. Cham: Springer International Publishing, 2016, pp.189–209.
 29. Degel MA, Navab N, Albarqouni S, et al. Domain and geometry agnostic CNNs for left atrium segmentation in 3D ultrasound. In: Frangi AF, Schnabel JA and Davatzikos C (eds) *Medical image computing and computer assisted intervention – MICCAI 2018*. Cham: Springer International Publishing, 2018, pp.630–637.
 30. Unsupervised Domain Adaptation in Brain Lesion Segmentation with Adversarial Networks. <https://www.springerprofessional.de/en/unsupervised-domain-adaptation-in-brain-lesion-segmentation-with/12337964> (accessed 18 August 2021).
 31. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM* 2020; 63: 139–144.
 32. Kamnitsas K, Ledig C, Newcombe VF, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017; 36: 61–78.
 33. Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-aware generative adversarial networks. In: Descoteaux M, Maier-Hein L and Franz A, et al. (eds) *Medical image computing and computer assisted intervention – MICCAI 2017*. Cham: Springer International Publishing, 2017, pp.417–425.
 34. Yan W, Wang Y, Gu S, et al. The domain shift problem of medical image segmentation and vendor-adaptation by unet-GAN. In: Shen D, Liu T and Peters TM, et al. (eds) *Medical image computing and computer assisted intervention – MICCAI 2019*. Cham: Springer International Publishing, 2019, pp.623–631.
 35. Cui H, Yuwen C, Jiang L, et al. Bidirectional cross-modality unsupervised domain adaptation using generative adversarial networks for cardiac image segmentation. *Comput Biol Med* 2021; 136: 104726.
 36. Baldeon-Calisto M and Lai-Yuen SK. C-MADA: unsupervised cross-modality adversarial domain adaptation framework for medical image segmentation. In: *Medical imaging 2022: image processing*. San Diego, CA, USA: SPIE, 2022, pp.971–978.
 37. Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*, Venice, Italy, October 2017, pp.2794–2802. New York City, New York, USA: IEEE.
 38. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, Nevada, USA, June 2016, pp.770–778. New York City, New York, USA: IEEE.
 39. Isola P, Zhu J-Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Hawaii, USA, July 2017, pp.1125–1134. New York City, New York, USA: IEEE.
 40. Ronneberger O, Fischer P and Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, Munich, Germany, October 2015, pp.234–241. Cham: Springer.
 41. Lee C-Y, Xie S, Gallagher P, et al. Deeply-supervised nets. In: *Artificial intelligence and statistics*, San Diego, California, USA, May 2015, pp.562–570. MA, USA: PMLR.
 42. Kim J, Kim M, Kang H, et al. U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. <https://openreview.net/forum?id=BJIZ5ySKPH> (2019, accessed 4 February 2022).
 43. Huo Y, Xu Z, Moon H, et al. Synseg-net: synthetic segmentation without target modality ground truth. *IEEE Trans Med Imaging* 2018; 38: 1016–1025.
 44. Scully PR, Bastarriga G, Moon JC, et al. Myocardial extracellular volume quantification by cardiovascular magnetic resonance and computed tomography. *Curr Cardiol Rep* 2018; 20: 15.
-