Harnessing Uncertainty in Domain Adaptation for MRI Prostate Lesion Segmentation

Eleni Chiou[1,2], Francesco Giganti[3,4], Shonit Punwani[5], Iasonas Kokkinos[2], and Eleftheria Panagiotaki[1,2]

[1] Centre for Medical Image Computing, UCL, London, UK
[2] Department of Computer Science, UCL, London, UK
[3] Department of Radiology, UCLH NHS Foundation Trust, London, UK
[4] Centre for Medical Imaging, Division of Medicine, UCL, London, UK
eleni.chiou.17@ucl.ac.uk

Abstract. The need for training data can impede the adoption of novel imaging modalities for learning-based medical image analysis. Domain adaptation methods partially mitigate this problem by translating training data from a related source domain to a novel target domain, but typically assume that a one-to-one translation is possible. Our work addresses the challenge of adapting to *a more informative target domain* where multiple target samples can emerge from a single source sample. In particular we consider translating from mp-MRI to VERDICT, a richer MRI modality involving an optimized acquisition protocol for cancer characterization. We explicitly account for the inherent uncertainty of this mapping and exploit it to generate multiple outputs conditioned on a single input. Our results show that this allows us to extract systematically better image representations for the target domain, when used in tandem with both simple, CycleGAN-based baselines, as well as more powerful approaches that integrate discriminative segmentation losses and/or residual adapters. When compared to its deterministic counterparts, our approach yields substantial improvements across a broad range of dataset sizes, increasingly strong baselines, and evaluation measures.

Keywords: Domain adaptation, Image synthesis, GANs, Segmentation, MRI

1 Introduction

Domain adaptation can be used to exploit training samples from an existing, densely-annotated domain within a novel, sparsely-annotated domain, by bridging the differences between the two domains. This can facilitate the training of powerful convolutional neural networks (CNNs) for novel medical imaging modalities or acquisition protocols, effectively compensating for the limited amount of training data available to train CNNs in the new domain.

The assumption underlying most domain adaptation methods is that one can align the two domains either by extracting domain-invariant representations (features), or by establishing a 'translation' between the two domains at the

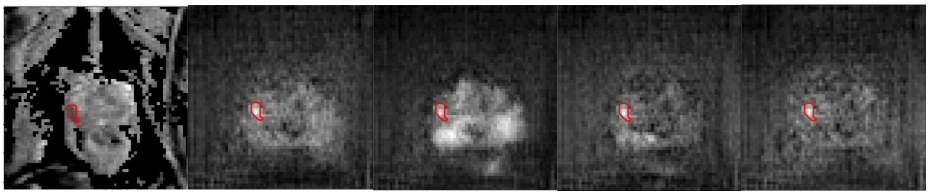

Fig. 1. One-to-many mapping from one mp-MRI image (left) to four VERDICT-MRI translations: our network can generate samples with both local and global structure variation, while at the same time preserving the critical structure corresponding to the prostate lesion, shown as a red circle. We note that the lesion area is annotated by a physician on the leftmost image, but is not used as input to the translation network - instead the translation network learns to preserve lesion structures thanks to the end-to-end discriminative training (details in text).

signal level, where in any domain the 'resident' and the translated signals are statistically indistinguishable.

In particular for medical imaging, [23] and [13] rely on adversarial training to align the feature distributions between the source and the target domain for medical image classification and segmentation respectively. Pixel-level distribution alignment is performed by [2, 11, 27, 28], who use CycleGAN [29] to map source domain images to the style of the target domain; they further combine the translation network with a task-specific loss to penalize semantic inconsistency between the source and the synthesized images. The synthesized images are used to train models for image segmentation in the target domain. Ouyang et al. [17] perform adversarial training to learn a shared, domain-invariant latent space which is exploited during segmentation. They show that their approach is effective in cases where target-domain data is scarce. Similarly, [26] embed the input images from both domains onto a domain-specific style space and a shared content space. Then, they use the content-only images to train a segmentation model that operates well in both domains. However, their approach does not necessarily preserve crucial semantic information in the content-only images.

These methods rely on the strong assumption that the two domains can be aligned - our work shows that accuracy gains can be obtained by acknowledging that this can often be only partially true, and mitigating the resulting challenges. As a natural image example, an image taken at night can have many day-time counterparts, revealed by light; similarly in medical imaging, a better imaging protocol can reveal structures that had previously passed unnoticed. In technical terms, the translation can be one-to-many, or, stated in probabilistic terms, multi-modal [10, 14, 30]. Using a one-to-one translation network in such a setting can harm performance, since the translation may predict the mean of the underlying multi-modal distribution, rather than provide diverse, realistic samples from it.

In our work we accommodate the inherent uncertainty in the cross-domain mapping and, as shown in Figure 1, generate multiple outputs conditioned on a

single input, thereby allowing for better generalization of the segmentation network in the target domain. As in recent studies [2,11,27,28], we use GANs [6] to align the source and target domains, but go beyond their one-to-one, deterministic mapping approaches. In addition, inspired by [2,8,11,27], we enforce semantic consistency between the real and synthesized images by exploiting source-domain lesion segmentation supervision to train target-domain networks operating on the synthesized images. This results in training networks that can generate diverse outputs while at the same time preserving critical structures - such as the lesion area in Figure 1. We further accommodate the statistical discrepancies between real and synthesized data by introducing residual adapters (RAs) [5,22] in the segmentation network. These capture domain-specific properties and allow the segmentation network to generalize better across the two domains.

We demonstrate the effectiveness of our approach in prostate lesion segmentation and an advanced diffusion weighted (DW)-MRI method called VERDICT-MRI (Vascular, Extracellular and Restricted Diffusion for Cytometry in Tumors). VERDICT-MRI is a non-invasive imaging technique for cancer microstructure characterisation [12,18,20]. The method has been recently in clinical trial to supplement standard multi-parametric (mp)-MRI for prostate cancer diagnosis. Compared to the naive DW-MRI from mp-MRI acquisitions, VERDICT-MRI has a richer acquisition protocol to probe the underlying microstructure and reveal changes in tissue features similar to histology. A recent study [12] has demonstrated that the intracellular volume fraction (FIC) maps obtained with VERDICT-MRI differentiate better benign and malignant lesions compared to the apparent diffusion coefficient (ADC) map obtained with the naive DW-MRI from mp-MRI acquisitions. However, the limited amount of available labeled training data does not allow the training of robust deep neural networks that could directly exploit the information in the raw VERDICT-MRI [4]. On the other hand, large scale clinical mp-MRI datasets exist [1,15]. As shown experimentally, our approach largely improves the generalization capabilities of a lesion segmentation model on VERDICT-MRI by exploiting labeled mp-MRI data.

2 Method

Our approach relies on a unified network for cross-modal image synthesis and segmentation, that is trained end-to-end with a combination of objective functions. As shown in Figure 2, at the core of this network is an image-to-image translation network that maps images from the source ('S') to the target ('T') domain. The translation network is trained in tandem with a segmentation network that operates in the target domain, and is trained with both the synthesized and the few real annotated target-domain images. Beyond these standard components, our approach relies on three additional components: firstly, we sample a latent variable from a Gaussian distribution when translating to the target domain; this represents structures that cannot be accounted by a deterministic mapping, and can result in one-to-many translation when needed. Secondly, we introduce residual adapters (RAs) to a common backbone network for se-

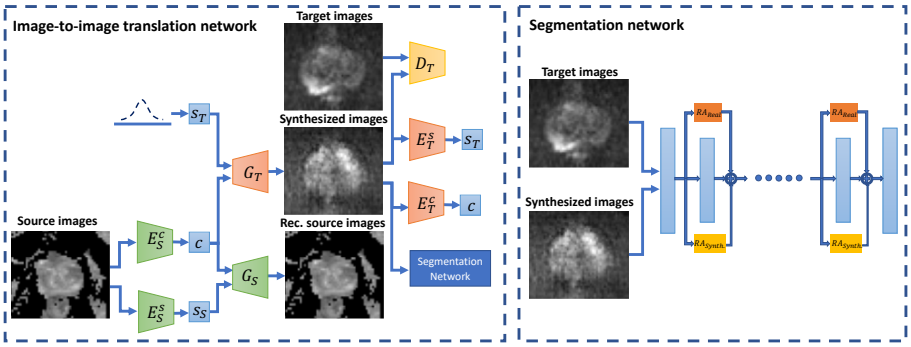

Fig. 2. Overview of our domain adaptation framework: we train a noise-driven domain translation network in tandem with a discriminatively supervised segmentation network in the target domain; GAN-type losses align the translated samples with the target distribution, while residual adapters allow the segmentation network to compensate for remaining discrepancies. Please see text for details.

mantic segmentation, allowing the discriminative training to accommodate any remaining discrepancies between the real and synthesized target domain images. Finally, we use a dual translation network from the target to the source domain, allowing us to use cycle-consistency in domain adaptation [10, 16, 29]; the cycle constraint allows us to disentangle the deterministic, transferable part from the stochastic, non-transferable part, which is filled in by Gaussian sampling, as mentioned earlier.

2.1 Problem formulation

Having provided a broad outline of our method, we now turn to a more detailed technical description. We consider the problem of domain adaptation in prostate lesion segmentation. We assume that the source domain, $\mathcal{X}_S$, contains N_S images, $x_S \in \mathcal{X}_S$, with associated segmentation masks, $y_S \in \mathcal{Y}_S$. Similarly, the sparsely labeled target domain, $\mathcal{X}_T$, consists of N_T images, $x_T \in \mathcal{X}_T$. A subset $\tilde{\mathcal{X}}_T$ of $\mathcal{X}_T$ comes with associated segmentation masks, $y_T \in \mathcal{Y}_T$. The proposed framework consists of two main components, i.e. an image-to-image translation network and a segmentation network described below.

Segmentation Network

The segmentation network (Fig. 2), Seg, operates on image-label pairs of both real, $\mathcal{X}_T$, and synthesized data, $\mathcal{X}_{S\to T}$, translated from source to target. An encoder-decoder network [3, 24] is the main backbone which serves both domains. To compensate further for differences in the feature statistics of real and synthesized data we install residual adapter modules [22] in parallel to each of the convolutional layer of the backbone. Introducing residual adapters ensures

that most of the parameters stay the same with the network, but also that the new unit introduces a small, but effective modification that accommodates the remaining statistical discrepancies of the two domains.

More formally, let ϕ_l be a convolutional layer in the segmentation network and $\mathbf{F}^l \in \mathbb{R}^{k\times k\times C_i\times C_o}$ be a set of filters for that layer, where $k \times k$ is the kernel size and C_i, C_o are the number of input and output feature channels respectively. Let also $\mathbf{Z}_i^l \in \mathbb{R}^{1\times 1\times C_i\times C_o}$ be a set of domain-specific residual adapter filters of domain i, where $i \in \{1, 2\}$, installed in parallel with the existing set of filters $\mathbf{F}_l$. Given an input tensor $\mathbf{x}_l \in \mathbb{R}^{H\times W\times C_i}$, the output $\mathbf{y}_l \in \mathbb{R}^{H\times W\times C_o}$ of layer l is given by

$$\mathbf{y}_l = \mathbf{F}^l * \mathbf{x} + \mathbf{Z}_i^l * \mathbf{x}. \tag{1}$$

We train the segmentation network by optimizing the following objective

$$\begin{aligned}\mathcal{L}_{Seg}(Seg, \tilde{\mathcal{X}_T}, \mathcal{Y}_T, \mathcal{X}_{S\to T}, \mathcal{Y}_S) = \\ \mathcal{L}_{DSC}(Seg, \tilde{\mathcal{X}_T}, \mathcal{Y}_T) + \mathcal{L}_{DSC}(Seg, \mathcal{X}_{S\to T}, \mathcal{Y}_S).\end{aligned} \tag{2}$$

The dice loss, $\mathcal{L}_{DSC}$, is given by

$$\mathcal{L}_{DSC}(Seg, \mathcal{X}, \mathcal{Y}) = -\frac{2\sum_{(\mathbf{x},\mathbf{y})\in(\mathcal{X},\mathcal{Y})}\sum_{k=1}^{K} Seg(\mathbf{x})_k \mathbf{y}_k}{\sum_{(\mathbf{x},\mathbf{y})\in(\mathcal{X},\mathcal{Y})}\sum_{k=1}^{K}(Seg(\mathbf{x})_k^2 + \mathbf{y}_k^2)}, \tag{3}$$

where K the number of voxels in the input images. We adopt this objective function since it is a differentiable approximation of a criterion that is well-adapted to our task.

Diverse Image-to-Image Translation Network

Recently, several studies [10, 30] have pointed out that cross-domain mapping is inherently multi-modal and proposed approaches to produce multiple outputs conditioned on a single input. Here we use MUNIT [10] to illustrate the key idea. As it is illustrated in Figure 2 the image-to-image translation network consists of content encoders E_S^c, E_T^c, style encoders E_S^s, E_T^s, generators G_S, G_T and domain discriminators D_S, D_T for both domains. The content encoders E_S^c, E_T^c map images from the two domains onto a domain-invariant content space $\mathcal{C}$ ($E_S^c : \mathcal{X}_S \to \mathcal{C}$, $E_T^c : \mathcal{X}_T \to \mathcal{C}$) and the style encoders E_S^s, E_T^s map the images onto domain-specific style spaces $\mathcal{S}_S$ ($E_S^s : \mathcal{X}_S \to \mathcal{S}_S$) and $\mathcal{S}_T$ ($E_T^s : \mathcal{X}_T \to \mathcal{S}_T$). The content code can be understood as the underlying anatomy which is the information that we want transfer during the translation while the style codes capture information related to the imaging modalities. Image-to-image translation is performed by combining the content code extracted from a given input and a random style code sampled from the target-style space. For instance, to translate an image $x_S \in \mathcal{X}_S$ to $\mathcal{X}_T$ we first extract its content code $c = E_S^c(x_S)$. The generator G_T uses the extracted content code c and a randomly drawn style code $s_T \in \mathcal{S}_T$ to produce the final output $x_{S\to T} = G_T(c, s_T)$. By

sampling random style codes from the style spaces $\mathcal{S}_S$ and $\mathcal{S}_T$ the generators G_S and G_T are able to produce diverse outputs. We train the networks with a loss function that consists of domain adversarial, self-reconstruction, latent reconstruction, cycle-consistency and segmentation losses.

Domain adversarial loss. We utilize GANs to match the distribution between the synthesized and the real images of the two domains. The adversarial discriminators D_T, D_S aim at discriminating between real and synthesized images, while the generators G_T, G_S aim at generating realistic images that fool the discriminators. For G_T and D_T the GAN loss is defined as

$$\begin{aligned}&\mathcal{L}^T_{GAN}(E^c_S, G_T, D_T, \mathcal{S}_T, \mathcal{X}_S) = \\ &\quad \mathbb{E}_{x_S\sim\mathcal{X}_S, s_T\sim\mathcal{S}_T}[\log(1 - D_T(G_T(E^c_S(x_S), s_T)))] + \mathbb{E}_{x_T\sim\mathcal{X}_T}[\log(D_T(x_T))].\end{aligned} \tag{4}$$

Self-reconstruction loss. Given the encoded content and style codes of a source-domain image the generator G_S should be able to decode them back to the original one.

$$\mathcal{L}^S_{recon}(G_S, E^s_S, E^c_S, \mathcal{X}_S) = \mathbb{E}_{x_S\sim\mathcal{X}_S}[\|G_S(E^c_S(x_S), E^s_S(x_S)) - x_S\|_1]. \tag{5}$$

Latent reconstruction loss. To encourage the translated image to preserve the content of the source image, we require that a latent code c sampled from the latent distribution can be reconstructed after decoding and encoding.

$$\begin{aligned}&\mathcal{L}^{c_S}_{recon}(E^c_S, G_T, E^c_T, \mathcal{X}_S, S_T) = \\ &\quad \mathbb{E}_{x_S\sim\mathcal{X}_S, s_T\sim\mathcal{S}_T}[\|E^c_T(G_T(E^c_S(x_s), s_T)) - E^c_S(x_s)\|_1].\end{aligned} \tag{6}$$

Similarly, to align the style representation with a Gaussian prior distribution, we enforce the same constrain for the latent style code.

$$\begin{aligned}&\mathcal{L}^{s_T}_{recon}(E^c_S, G_T, E^s_T, \mathcal{X}_S, S_T) = \\ &\quad \mathbb{E}_{x_S\sim\mathcal{X}_S, s_T\sim\mathcal{S}_T}[\|E^s_T(G_T(E^c_S(x_s), s_T)) - s_T)\|_1].\end{aligned} \tag{7}$$

Cycle-consistency loss. To facilitate training we enforce cross-cycle consistency which implies that if we translate an image to the target domain and then translate it back to the source domain using the extracted source-domain style code, we should be able to obtain the original image.

$$\begin{aligned}&\mathcal{L}^S_{cyc}(E^c_S, E^s_S, G_T, E^c_T, G_S, \mathcal{X}_S, S_T) = \\ &\quad \mathbb{E}_{x_S\sim\mathcal{X}_S, s_T\sim\mathcal{S}_T}[\|G_S(E^c_T(G_T(E^c_S(x_S), s_T)), E^s_S(x_S)) - x_S\|_1].\end{aligned} \tag{8}$$

$\mathcal{L}^S_{GAN}$, $\mathcal{L}^T_{recon}$, $\mathcal{L}^{c_T}_{recon}$, $\mathcal{L}^{s_S}_{recon}$, $\mathcal{L}^T_{cyc}$ are defined in a similar way.

Segmentation loss. To enforce the generator to preserve the lesions, we enrich the network with segmentation supervision on the synthesized images. The segmentation loss on the synthesized images is given by

$$\mathcal{L}^{Synth}_{Seg}(Seg, G_T, E^c_S, \mathcal{X}_S, \mathcal{Y}_S, \mathcal{S}_T) = \mathcal{L}_{DSC}(Seg, G_T(E^c_S(\mathcal{X}_S), S_T), \mathcal{Y}_S). \tag{9}$$

The full objective is given by

$$\begin{aligned}\min_{E_S^c,E_S^s,E_T^c,E_T^s,G_S,G_T}\max_{D_S,D_T}\ &\lambda_{GAN}(\mathcal{L}_{GAN}^S+\mathcal{L}_{GAN}^T)+\lambda_x(\mathcal{L}_{recon}^S+\mathcal{L}_{recon}^T)\\&+\lambda_c(\mathcal{L}_{recon}^{c_S}+\mathcal{L}_{recon}^{c_T})+\lambda_s(\mathcal{L}_{recon}^{s_S}+\mathcal{L}_{recon}^{s_T})\\&+\lambda_{cyc}(\mathcal{L}_{cyc}^S+\mathcal{L}_{cyc}^T)+\mathcal{L}_{Seg}^{Synth},\end{aligned}\tag{10}$$

where λ_{GAN}, λ_x, λ_c, λ_s, λ_{cyc} are weights that control the importance of each term.

2.2 Implementation details

We implement our model using Pytorch [21]. The content encoders consist of several convolutional layers and residual blocks followed by instance normalization [25]. The style encoders consist of convolutional layers followed by fully connected layers. The decoders include residual blocks followed by upsampling and convolutional layers. The residual blocks are followed by adaptive instance normalization (AdaIN) [9] layers to adjust the style of the output image. The affine parameters of AdaIN are generated by a multilayer perceptron from a given style code. The discriminators consist of several convolutional layers. The encoder of the segmentation network is a standard ResNet [7] consisting of several convolutional layers while the decoder consists of several upsampling and convolutional layers. For training we use Adam optimizer, a batch size of 32 and a learning rate of 0.0001. We make our code available at https://github.com/elchiou/DA.

2.3 Datasets

VERDICT-MRI: We use VERDICT-MRI data collected from 60 men with a suspicion of cancer. VERDICT-MRI images were acquired with pulsed-gradient spin-echo sequence (PGSE) using an optimized imaging protocol for VERDICT prostate characterization with 5 b-values (90, 500, 1500, 2000, 3000 s/mm^2) in 3 orthogonal directions [19]. Images with b = 0 s/mm^2 were also acquired before each b-value acquisition. The DW-MRI sequence was acquired with a voxel size of $1.25 \times 1.25 \times 5$ mm^3, 5 mm slice thickness, 14 slices, and field of view of 220×220 mm^2 and the images were reconstructed to a 176×176 matrix size. A dedicated radiologist, highly experienced in prostate mp-MRI, contoured the lesions on VERDICT-MRI using mp-MRI for guidance.

DW-MRI from mp-MRI acquisition: We use DW-MRI data from the ProstateX challenge dataset [15] which consists of training mp-MRI data acquired from 204 patients. The DW-MRI data were acquired with a single-shot echo planar imaging sequence with a voxel size of $2 \times 2 \times 3.6$ mm^3, 3.6 mm slice thickness. Three b-values were acquired ($50, 400, 800$ s/mm^2), and subsequently, the ADC map and a b-value image at b = 1400 s/mm^2 were calculated by the scanner software. In this study, we use DW-MRI data from 80 patients. Since the ProstateX dataset provides only the position of the lesion, a dedicated radiologist manually annotated the lesions on the ADC map using as reference the provided position of the lesion.

Table 1. Average recall, precision, dice similarity coefficient (DSC), and average precision (AP) across 5 folds. The results are given in mean (±std) format.

Model	Recall	Precision	DSC	AP
VERDICT-MRI only	67.1 (±14.2)	59.6 (±11.5)	62.4 (±13.4)	63.5 (±13.1)
Finetuning	68.4 (±12.4)	62.5 (±13.5)	64.7 (±11.2)	65.8 (±14.7)
RAs	66.6 (±11.6)	67.0 (±8.8)	65.7 (±10.2)	66.6 (±12.6)
MUNIT	65.2 (±10.2)	64.2 (±13.7)	64.4 (±11.3)	68.2 (±12.0)
CycleGAN + $\mathcal{L}_{Seg}^{Synth}$	64.5 (±10.4)	66.1 (±10.1)	64.8 (±8.7)	70.1 (±9.8)
CycleGAN + $\mathcal{L}_{Seg}^{Synth}$ + RAs	60.9 (±10.7)	**74.0** (±11.8)	66.6 (±13.6)	71.6 (±11.3)
MUNIT + $\mathcal{L}_{Seg}^{Synth}$ (Ours)	**71.8** (±7.8)	68.0 (±6.8)	69.8 (±7.9)	73.5 (±8.1)
MUNIT + $\mathcal{L}_{Seg}^{Synth}$ + RAs (Ours)	69.2 (±8.6)	71.2 (±9.7)	**69.9** (±9.0)	**75.4** (±9.7)

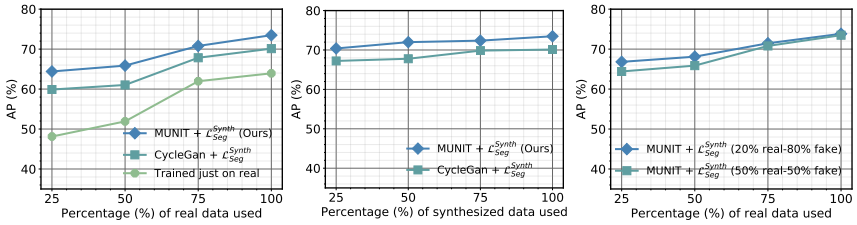

Fig. 3. Impact of the ratio of synthesized to real data on the performance. (Right) Average precision (AP) as a function of the percentage of real samples used given a constant number of synthesized ones. (Middle) AP as a function of the number of synthesized examples used given a constant number of real ones. (Left) AP as a function of the percentage of real data used given a constant number of synthesized ones. Here, the ratio of real to synthesized data in a mini-batch also varies during training.

3 Results

In this section we evaluate the performance of our approach and the impact of the ratio of synthesized to real data on the performance. In the supplementary material we provide qualitative results and quantitative results related to the effect of sampling random style codes on the performance.

Performance evaluation. We first compare our approach to several baselines. i)VERDICT-MRI only: we train the segmentation network only on VERDICT-MRI. ii) Finetuning: we pre-train on mp-MRI and then perform finetuning using the VERDICT-MRI data. iii) RAs: we pre-train on mp-MRI, then we install RAs in parallel to each of the convolutional layers of the pre-trained network and update them using VERDICT-MRI. iv) MUNIT: we use MUNIT to map from source to target without segmentation supervision. v) CycleGAN + $\mathcal{L}_{Seg}^{Synth}$: we use CycleGAN and segmentation supervision to perform the translation, an approach similar to the one proposed in [28]. vi) CycleGAN + $\mathcal{L}_{Seg}^{Synth}$ + RAs: we use (v) for the translation and introduce RAs to the segmentation network. We

evaluate the performance based on the average recall, precision, dice similarity coefficient (DSC), and average precision (AP). We report the results in Table 1. The proposed approach yields substantial improvements and outperforms all baselines including CycleGAN, which indicates that accommodating the uncertainty in the cross-domain mapping allows us to learn better representations for the target domain. Compared to the naive MUNIT without segmentation supervision, $\mathcal{L}_{Seg}^{Synth}$, our approach performs better since it successfully preserves the lesions during the translation. Finally, introducing RAs in the segmentation networks further improves the performance of both CyclgeGAN + $\mathcal{L}_{Seg}^{Synth}$ and MUNIT + $\mathcal{L}_{Seg}^{Synth}$.

Impact of the ratio of synthesized to real data on the performance. Using synthesized data is motivated by the fact that annotating large datasets can be challenging in medical applications. We therefore evaluate the impact of the ratio of synthesized to real data. To this end, we first vary the percentage of real data while keeping fixed the amount of synthesized data (Fig. 3 (left)). We compare our approach to a segmentation network trained only on real data and to [28] where CycleGAN is used for the generation of synthesized data. Our approach outperforms both baselines. Figure 3 (middle) shows the performance when we vary the percentage of synthesized samples while fixing the percentage of real ones. The AP of our approach increases as we increase the amount of synthesized data. The baseline also improves but we systematically outperform it. Figure 3 (right) shows the performance of our approach when we vary the percentage of real data while fixing the percentage of synthesized. Here, we also vary the ratio of real to synthesized data in a mini-batch during training. Note that when the percentage of real data is small, a large ratio of synthesized to real data in the mini-batch delivers better results.

4 Conclusion

In this work we propose a domain adaptation approach for lesion segmentation. Our approach exploits the inherent uncertainty in the cross-domain mapping to generate multiple outputs conditioned on a single input allowing the extraction of richer representations for the task of interest in the target domain. We demonstrate the effectiveness of our approach in lesion segmentation on VERDICT-MRI, which is an advanced imaging modality for prostate cancer characterization. However, our approach is quite general can be applied in other application where the amount of labeled training data is limited.

Acknowledgments

This research is funded by EPSRC grand EP/N021967/1. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the GPU used for this research.

References

1. Ahmed, H.U., El-Shater Bosaily, A., et al.: Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. The Lancet (2017)
2. Cai, J., Zhang, Z., Cui, L., Zheng, Y., Yang, L.: Towards cross-modal organ translation and segmentation: A cycle and shape consistent generative adversarial network. MedIA (2019)
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV (2018)
4. Chiou, E., Giganti, F., Bonet-Carne, E., Punwani, S., Kokkinos, I., Panagiotaki, E.: Prostate cancer classification on VERDICT DW-MRI using convolutional neural networks. In: MLMI (2018)
5. Chiou, E., Giganti, F., Punwani, S., Kokkinos, I., Panagiotaki, E.: Domain adaptation for prostate lesion segmentation on VERDICT-MRI. In: ISMRM (2020)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NIPS (2014)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
8. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: CyCADA: Cycle-consistent adversarial domain adaptation. In: ICML (2018)
9. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: ICCV (2017)
10. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: ECCV (2018)
11. Jiang, J., Hu, Y.C., Tyagi, N., Zhang, P., Rimner, A., Mageras, G.S., Deasy, J.O., Veeraraghavan, H.: Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation. In: MICCAI (2018)
12. Johnston, E.W., Bonet-Carne, E., et al.: VERDICT-MRI for prostate cancer: Intracellular volume fraction versus apparent diffusion coefficient. Radiology (2019)
13. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., Glocker, B.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: IPMI (2017)
14. Lee, H.Y., Tseng, H.Y., Huang, J.B., Singh, M., Yang, M.H.: Diverse image-to-image translation via disentangled representations. In: ECCV (2018)
15. Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H.: Computer-aided detection of prostate cancer in MRI. TMI (2014)
16. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: NIPS (2017)
17. Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., Rueckert, D.: Data efficient unsupervised domain adaptation for cross-modality image segmentation. In: MICCAI (2019)
18. Panagiotaki, E., Chan, R.W., Dikaios, N., et al.: Microstructural characterization of normal and malignant human prostate tissue with vascular, extracellular, and restricted diffusion for cytometry in tumours magnetic resonance imaging. Investigate Radiology (2015)
19. Panagiotaki, E., Ianus, A., Johnston, E., et al.: Optimised VERDICT MRI protocol for prostate cancer characterisation. In: ISMRM (2015)
20. Panagiotaki, E., Walker-Samuel, S., Siow, B., et al.: Noninvasive quantification of solid tumor microstructure using VERDICT MRI. Cancer Research (2014)

21. Paszke, A., Gross, S., Chintala, S., Chanan, G., et al.: Automatic differentiation in pytorch. In: Autodiff Workshop, NIPS (2017)
22. Rebuffi, S., Vedaldi, A., Bilen, H.: Efficient parametrization of multi-domain deep neural networks. In: CVPR (2018)
23. Ren, J., Hacihaliloglu, I., Singer, E.A., Foran, D.J., Qi, X.: Adversarial domain adaptation for classification of prostate histopathology whole-slide images. In: MICCAI (2018)
24. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
25. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In: CVPR
26. Yang, J., Dvornek, N.C., Zhang, F., Chapiro, J., Lin, M., Duncan, J.S.: Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation. In: MICCAI (2019)
27. Zhang, Y., Miao, S., Mansi, T., Liao, R.: Task driven generative modeling for unsupervised domain adaptation: Application to X-ray image segmentation. In: MICCAI (2018)
28. Zhang, Z., Yang, L., Zheng, Y.: Translating and segmenting multimodal medical volumes with cycle and shape consistency generative adversarial network. In: CVPR (2018)
29. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017)
30. Zhu, J.Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward multimodal image-to-image translation. In: NIPS (2017)