Abstract—We developed a new joint probabilistic segmentation and image distribution matching generative adversarial network (PSIGAN) for unsupervised domain adaptation (UDA) and multi-organ segmentation from magnetic resonance (MRI) images. Our UDA approach models the co-dependency between images and their segmentation as a joint probability distribution using a new structure discriminator. The structure discriminator computes structure of interest focused adversarial loss by combining the generated pseudo MRI with probabilistic segmentations produced by a simultaneously trained segmentation sub-network. The segmentation sub-network is trained using the pseudo MRI produced by the generator sub-network. This leads to a cyclical optimization of both the generator and segmentation sub-networks that are jointly trained as part of an end-to-end network. Extensive experiments and comparisons against multiple state-of-the-art methods were done on four different MRI sequences totalling 257 scans for generating multi-organ and tumor segmentation. The experiments included, (a) 20 T1-weighted (T1w) in-phase mduiton and (b) 20 T2-weighted (T2w) abdominal MRI for segmenting liver, spleen, left and right kidneys, (c) 162 T2-weighted fat suppressed head and neck MRI (T2wFS) for parotid gland segmentation, and (d) 75 T2w MRI for lung tumor segmentation. Our method achieved an overall average DSC of 0.87 on T1w and 0.90 on T2w for the abdominal organs, 0.82 on T2wFS for the parotid glands, and 0.77 on T2w MRI for lung tumors. We have made our code available at: https://github.com/harveerar/PSIGAN.

Index Terms—Unsupervised domain adaptation, generative adversarial network, MRI segmentation, lung tumor, parotid glands, abdominal organs.

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is rapidly emerging as the modality for image-guided adaptive radiation therapy treatments due to its better soft tissue contrast compared with computed tomography (CT) scans. However, a critical requirement for MR-guided radiotherapy is fast, accurate, and consistent segmentation of target and surrounding normal organs at risk (OAR). [2]

Deep learning-based methods have shown remarkable success in diverse image analysis tasks when they can be trained using large annotated datasets. However, acquiring large expert-segmented medical image datasets is difficult. This is because, slice-by-slice delineations of several organs is highly time consuming, and such delineations require a domain expert, such as a radiologist or a radiation oncologist.

Domain adaptation is a commonly used approach to overcome the issue of learning from limited and unlabeled target modality datasets, where a model for the target domain is trained by using an existing labeled dataset from a different modality, called the source domain. In unsupervised domain adaptation (UDA) based segmentation, the focus of this work, no target domain labeled data is available for training.

UDA segmentation has been accomplished by either using feature-level or pixel-level adaptation. In feature-level adaptation, the encoder networks are adversarially trained to extract domain-invariant features, such that a single segmentation model trained on source data is applicable to both domains [4]. [5]. [6]. [7]. In pixel-level adaptation, generative adversarial networks (GAN) model the complex inter-modality anatomical relationships and compute image to image (I2I) translations [6]. [10]. The target modality model is then learned by using the source to target transformed images. Cyclical consistency and feature disentanglement losses are often used in unpaired I2I translation methods to circumvent the lack of corresponding source and target modality images. [3]. [13]. [7]. Hybrid methods combine feature and pixel-level adaptation in order to ensure good pixel-level I2I transformations and the preservation of low-level edge and mid-level textural characteristics of the target modality images.

A major issue while performing UDA is mode collapse [18].

Fig. 1. Difference between traditional UDA and PSIGAN segmentation. In both (a) and (b), the translation network T parameterized by θ_T produces labeled pseudo target data \{
X^m_c, Y_c\} from unpaired source X_c and target X_m data and trains the segmentation (S) network, parameterized by θ_S. PSIGAN also uses a structure discriminator \mathcal{D}_{st} to match the joint distribution of image-segmentation probability maps \{x_m, \psi_m\}, \{x^m_c, \psi^m_c\} to further optimize \theta_T.

This work was supported by the MSK Cancer Center core grant P30 CA08748.

Jue Jiang, Neelam Tyagi, Yu Chi-Hu Hu, Joseph O. Deasy, Sean Berry, and Harini Veeraraghavan are with the department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, USA (email: veerarah@mskcc.org)

Nancy Lee and Andreas Rimner are with the department of Radiation oncology, Memorial Sloan Kettering Cancer Center, New York, USA.
wherein multiple distinct inputs are mapped to a same output. Modality hallucination is a related manifestation of this issue that commonly occurs in medical image I2I translations, wherein distinct organ characteristics like geometry (or overall shape) and appearance (or intensity distribution) are ill-preserved or are removed \[10\] in translated images. This is because the commonly used losses based on matching global or marginal intensity statistics of whole images cannot sufficiently constrain the generator to model the local organ/tumor geometry and appearance statistics. Feature disentanglement methods have been proposed to reduce the aforementioned issues by using shared content and domain-specific attribute encoders \[12\], with demonstrated success in medical image applications \[13\]. Nevertheless, the lack of explicit conditioning of losses with respect to the geometry and appearance of the various structures of interest (SOI), may not yield the desired results for the output task.

Prior works have used output segmentation as an adversarial loss to constrain UDA and improve segmentations \[14\], \[19\], \[20\], \[21\], \[10\]. As shown in \[13\], task-specific losses can improve training stability and reduce chances of mode collapse. However, adversarial losses computed only using the segmentation output can constrain the geometry but not necessarily the appearance of SOIs in I2I translation.

A key difference of our approach compared to prior works is the use of joint distribution matching adversarial loss. The joint distribution is represented as a channel-wise concatenation of images and their voxel-wise segmentation probability maps. Prior works have employed similar joint-distribution matching to constrain bidirectional matching between images and a low dimensional latent distribution vector \[22\], \[23\] or images and scalar output categories \[24\]. Ours on the other hand constrains pixel-level relationships between I2I translations and segmentations. To our best knowledge, this is the first approach to compute joint distribution matching of images and segmentation probabilities for UDA segmentation (Fig. \[1\]). This approach also leads to a cyclical optimization where both segmentation and generator outputs constrain each other. In prior approaches like \[21\], \[8\], segmentation outputs do not constrain generator network gradient computation.

Our contributions are:

- **Organ geometry and appearance constrained unpaired cross-modality adaptation.** We introduced an UDA approach that constrains organ geometry and appearance in I2I translations by computing adversarial losses to minimize the mismatch in the joint distribution of images and their segmentation probability maps.

- **A cyclic feedback based UDA segmentation.** In our approach, the generator outputs are used to train the segmentation network, while the segmentation outputs are used to compute losses for the generator.

- **Comprehensive performance comparisons** were done against multiple state-of-the-art methods using three datasets for multiple organs and tumor segmentation. Extensive ablation experiments were done to measure the impact of joint distribution matching on both segmentation and I2I translation.

### II. RELATED WORKS

#### A. Feature-level UDA segmentation

Feature-level UDA segmentation methods extract a domain-invariant feature encoding, such that a model trained on source data is applicable to both source and target domains. This method is often used for performing domain adaptation between related image sequences. Example applications of this include prostate gland segmentation from MRIs acquired from different scanners \[5\], and brain tumor segmentation from similar MRI contrasts \[4\]. Joint adversarial training strategy combining a domain critic network with a segmentation network has been used for the more challenging cross-modality adaptation between CT and MRI in \[25\]. However, as shown in \[25\], these methods may require different number of feature layers to be adapted for segmenting various organs. Hence, network feature-adaptation depth is an important hyper-parameter. Such a depth tuning can require computationally intensive training when generating segmentation of large number of organs for radiation therapy applications. This issue has been avoided by using joint latent space learning with variational autoencoders \[7\] and disentangled feature representations \[26\]. The work in \[27\] employed output feature matching for producing scanner invariant estimation of cardiothoracic ratio (or heart size) from chest X-rays. Segmentation probability maps produced from softMax layer \[28\] and entropy maps indicative of pixel-wise classification uncertainties \[27\] have been used for domain invariant (synthetic and camera-acquired) natural image segmentation.

#### B. Pixel-level UDA segmentation

Pixel-level domain adaptation and segmentation methods model the inter-modality relationships and compute I2I image translations. The translated images are used to train a segmentation model for the target domain \[29\], \[30\], \[3\], \[14\]. Two-step training consisting of I2I translation (e.g., CT to MRI) followed by segmentation network training has been used for MRI lung tumor \[10\] and fundus image segmentation \[31\]. I2I translation and segmentation were combined into one network to segment cardiac structures from CT \[14\], \[17\], abdominal organs from CT and MRI \[3\], and knee structures from MRI \[50\]. Multiple works \[32\], \[30\], \[3\] have combined cyclical consistency losses of the CycleGAN \[11\] with a segmentation network to train with unpaired source and target modalities. Cyclical consistency loss shrinks the space of possible mappings in GANs by computing global intensity distribution mismatches. But this loss alone is insufficient to preserve the SOI geometry and appearance in I2I translation of medical images \[33\]. Inclusion of style and perceptual losses have shown improvements in I2I translations \[34\]. Geometry preserving losses implemented by backpropagating segmentation losses \[14\] to the generator and high-level segmentation features matching losses have also shown improvements for both tumor \[10\] and semantic segmentation of real-world images \[16\]. However, none of the aforementioned losses provide constraints to sufficiently control both geometry and appearance of the SOIs in I2I translation. We address this prob-
Fig. 2. Approach overview. Generator $G_M^C$ converts CT image $x_c$ into pseudo MR image $x^m_c$, which is used to train the segmentor $S$. $S$ uses split encoders $E_M, E^C_M$ and shared decoder $D_E$ to separate the gradient flows for $D_{struct}$ and $G_M^C$ through sub-networks $S_M$ and $S_M^C$. $D_{struct}$ computes joint distribution matching $(\{x^m_c, \psi^m_c\}, \{x_m, \psi_m\})$ adversarial loss $L_{struct}$, where $\psi^m_c$ and $\psi_m$ are produced by $S_M$. The corresponding adversarial loss $L_{struct}$ for $G_M^C$ uses $(x^m_c, \psi^m_c)$, where $\psi^m_c$ is produced by $S_M^C$. $D_M$ and $D_C$ are global intensity discriminators for MRI and CT domain; $G_M^C$ converts MRI to CT images for enforcing cyclically consistent I2I translations.

le by combining images and their segmentation probability maps as a joint density for adversarial learning.

III. METHOD

Goal: Learn MRI multi-organ segmentation models by using unpaired expert-segmented CT and unlabeled MRI images.

Notation: CT $(X_C, Y_C)$ is the source domain that consists of images $x_c \in X_C$ and expert-segmentations $y_c \in Y_C$ for training. MRI is the target domain that is provided with only MRI images $x_m \in X_M$ for training. As the learning optimization involves a finite set of examples, the probability distribution of CT and MRI images are represented as $p(x_c)$ and $p(x_m)$. The probability distribution of pseudo MRI $x^m_c$ and pseudo CT $x^m_c$ resulting from I2I translations of CT to MRI and MRI to CT are represented as $p(x^m_c)$ and $p(x^m_m)$, respectively. The joint probability distribution of the real MR and pseudo MR images and their probabilistic segmentations are represented as $p(x_m, \psi_m)$ and $p(x^m_c, \psi^m_c)$, respectively.

A. Background

In supervised learning with finite set of training examples $N$, given a joint probability distribution of inputs and outputs $p(x, y)$ and a model parameterized by $\theta$, a chosen loss function $L(x, \cdot)$ is used to compute the empirical risk in predicting the outputs $y$ from inputs $x$:

$$\mathbb{E}[L(x, y, \theta)] = \arg\min_{\theta \in \Theta} \sum_{i=1}^{N} p(x_i, y_i) L_c(x_i, y_i, \theta).$$  \hspace{1cm} (1)

In pixel-level UDA segmentation, we seek to minimize the empirical risk of training with pseudo target examples, obtained through a model $\phi : x_c \rightarrow x^m_c$, by assuming that $p(x^m_c) \approx p(x_m)$:

$$\mathbb{E}[L(x^m_c, y, \theta)] = \arg\min_{\theta \in \Theta} \sum_{i=1}^{N} p(x^m_m, y_i) L_c(x^m_m, y_i, \theta).$$  \hspace{1cm} (2)

where $p(x^m_m, y)$ is the joint distribution on an intermediate representation or pseudo domain. However, $\phi$ does not produce a perfect mapping, that is $p(x^m_m) \neq p(x_m)$. Therefore, an additional domain translation loss $L_1$ needs to be added to (2). With $p(x^m_m, x_m)$ as the joint probability distribution over pseudo and real target samples, the risk is computed as:

$$\mathbb{E}[L(x^m_m, x_m, y, \phi)] = \arg\min_{\phi \in \Phi} \sum_{j=1}^{M} p(x^m_m, x_m_j) L_1(x^m_m, x_m_j, \phi) + \sum_{i=1}^{N} p(x^m_m, y_i) L_c(x^m_m, y_i, \theta).$$  \hspace{1cm} (3)

Equation (3) can be optimized by training the domain translation and segmentation networks either sequentially [10], [31] or jointly [8], [13], [16]. However, this optimization ignores the co-dependency of domain translation and segmentation. For instance, no explicit constraint exists to preserve any inherent target modality appearance or geometric characteristics of the SOIs that distinguishes them from the background in the I2I translated images, which is a cause of sub-optimal performance. We model this co-dependency by computing adversarial losses using the joint distribution of images and
their segmentation probability maps as a pair obtained from the translated \((x^m_c, \psi^m_c)\) and real target images \((x_m, \psi_m)\). In other words, we require \(p(x^m_c, \psi^m_c) \approx p(x_m, \psi_m)\).

### B. Probabilistic segmentation and image matching GAN (PSIGAN)

The overview of PSIGAN is shown in Fig. 2. PSIGAN consists of a CT to MRI generator \(G^M_C: x_c \rightarrow x^m_c\), global intensity discriminator \(D_M\), a structure discriminator \(D_{struct}\), and a target domain segmentor \(S: x_m \rightarrow \{\psi_m, y_m\}\), where \(\psi_m\) is the predicted map of voxel-wise segmentation probabilities for \(x_m\) and \(y_m\) is the K-organ segmentation for \(x_m\). We also include a MR to CT generator \(G^C_M: x_m \rightarrow x^m_c\) and global intensity discriminator \(D_C\) to implement cyclical consistency losses for unpaired I2I translation. \(D_M\) computes a global adversarial loss to penalize mismatches in the marginal intensity distribution of pseudo and real MRI images in order that \(p(x^m_c) \approx p(x_m)\):

\[
\max_{D^G_M} \min_{G^C_M} \mathcal{L}_{D \rightarrow M}(G^M_C, D^M) = \mathbb{E}_{x_m \sim p(x_m)}[\log(D^M(x_m))] + \mathbb{E}_{x_c \sim p(x_c)}[\log(1 - (D^M(G^M_C(x_c)))]
\]

\[(4)\]

\(D_{struct}\) computes a joint distribution adversarial loss to reduce SOI appearance (intensity distribution) and geometry (overall shape) mismatches between pseudo \(x^m_c\) and real MRIs \(x_m\). It accomplishes this by matching the joint distributions \(p(x_m, \psi_m)\) and \(p(x^m_c, \psi^m_c)\), implemented by concatenating the images and their segmentation probability maps. The segmentor \(S\) (Fig. 2) produces the segmentation probability maps \(\psi_m, \psi^m_c\) from real \(x_m\) and pseudo MRI \(x^m_c\). \(D_{struct}\) loss is computed by using both real and pseudo MRI as:

\[
\max_{D^{struct}} \min_{G^M_C} \mathcal{L}_{D \rightarrow struct}(G^M_C) = \mathbb{E}_{(x_m, \psi_m) \sim p(x_m, S(x_m))}[\log(D_{struct}(x_m, \psi_m))] + \mathbb{E}_{(x^m_c, \psi^m_c) \sim p(x^m_c, S(x^m_c))}[\log(1 - D_{struct}(x^m_c, \psi^m_c))].
\]

\[(5)\]

The inclusion of segmentation probability maps in this loss constrains the SOI geometry similar to prior works [3], [14], [31], [17]. The inclusion of images in this loss additionally constrains the appearance. The generator \(G^M_C\) is optimized by adversarial loss, which is computed using the pseudo MR images \(x^m_c \approx G^M_C(x_c)\) as:

\[
\mathbb{E}_{(x^m_c, \psi^m_c) \sim p(x^m_c, S(x^m_c))}[\log(1 - D_{struct}(x^m_c, \psi^m_c))]. \quad \psi^m_c, \psi^m_c \text{ are produced by aggregating label assignment probabilities } e_{i,j} \text{ at location } i,j \text{ for } K \text{ SOIs from } K \text{ channels as generated by the softMax layer of the network } S \text{ as:}
\]

\[
\psi_{i,j} = \sum_{n=2}^{K} e_{i,j}^n. \quad \text{(7)}
\]

The first channel corresponds to background and is thus ignored in the aggregation shown in (7). The map \(\psi\) has continuous values in the range [0,1], where higher values indicate higher likelihood of a voxel corresponding to a SOI, while lower values are indicative of background.

The segmentor \(S\) can be trained using the generated pseudo MRI and the associated expert-segmentations available on the source modality \(x^m_c, y_c\) and optimized using cross-entropy losses. The output segmentation is computed using an argmax function.

Given a fixed \(G^M_C\) and \(S\), the optimal \(D_{struct}\) at any point in the optimization of (5) is given by \(\mathbb{P}(p(x_m, \psi_m)) = \mathbb{P}(p(x^m_c, \psi^m_c))\). The global equilibrium is achieved if and only if the joint distribution of \(p(x^m_c, \psi^m_c)\) and \(p(x_m, \psi_m)\) are matched, i.e., \(p(x^m_c, \psi^m_c) = p(x_m, \psi_m)\). Thus, unlike methods in [3], [14], [10], [31], [17] that only constrain geometry, ours constrains both organ geometry and appearance.

The joint distribution formulation leads to a cyclical optimization of the generator \(G^M_C\) and the segmentation \(S\) networks. Concretely, \(S\) requires the output \(x^m_c\) of \(G^M_C\) for its training, while \(G^M_C\) requires the output of \(S\), namely \(\psi^m_c\) for gradient computation. The gradient flow relation between \(D_{struct}\), \(G^M_C\) and \(S\) is illustrated in Fig. 3(a) and 3(b).

![Fig. 3: Gradient flow between \(G^M_C\), \(D_{struct}\), and segmentors using (a) single segmentor \(S\), and (b) split segmentors \(S^C_M\) and \(S_M\).](image-url)

The gradient flow resulting from a network using a single segmentor \(S\) is shown in Fig. 3(a), where the outputs of \(G^M_C\) and \(S\) are used to update each other. Concretely, \(x^m_c\) is used to train \(S\), and \(\{x^m_c, \psi^m_c\}\) is used to compute the joint distribution adversarial loss for \(G^M_C\). As a result, the outputs of \(G^M_C\) and \(S\) can co-adapt to facilitate easy segmentation, but without \(x^m_c\) matching the target distribution. In this case, \(D_{struct}\) can easily distinguish \(p(x^m_c, \psi^m_c)\) from \(p(x_m, \psi_m)\), allowing it to reach a stable local minima well before \(G^M_C\) reaches its local minima, and possibly prevent GAN convergence. This is because, GAN optimization hinges on achieving Nash equilibrium, whereby all players (generator and discriminator) achieve equal payoffs (reach local minima at similar times). This would also lower the generalization accuracy of \(S\) on real MRI. We address this potential issue by separating the networks used for producing the segmentation probability maps for \(G^M_C\) and \(D_{struct}\). Thus, segmentation probability maps from network \(S_M\) is used to compute \(D_{struct}\) loss, and segmentation probability maps from \(S^C_M\) is used for to compute \(G^M_C\) loss. The modified configuration of

1 We only show the components related to our contribution for simpler explanation. Other parts like \(D_C, G^C_M, D_M\) are used as done in CycleGAN.
gradient flow is shown in Fig. 3(b).

C. Split segmentor network

The split segmentation network consists of two encoder sub-networks, called $E_G$ and $E_M^C$ with a shared decoder network (DE) (Fig. 2), from which two segmentations are generated via $S_M = E_M^S \circ DE$ and $S_C^M = E_M^S \circ DE$. Shared decoder is used because the high-level contextual features for segmentation should be same between pseudo and real MRI. It also reduces the number of parameters required for training $S_M$ and $S_C^M$.

The networks $S_M$, $S_C^M$ are trained using pseudo MRI data with label $(x_m^p, y_c)$ and optimized with cross-entropy losses $L_{seg}^M$ (first part of summation in (8)) for $S_M$ and $L_{seg}^C$ (second part of summation in (8)) for $S_C^M$, respectively. The overall loss $L_{seg} = L_{seg}^M + L_{seg}^C$ is computed as:

$$L_{seg} = \frac{1}{\lambda} \left[ \mathbb{E}_{x_p^m \sim p(x_p^m), y_c \sim Y_c} \left[ \log P(y_c | S_M(x_p^m)) \right] + \mathbb{E}_{x_p^m \sim p(x_p^m), y_c \sim Y_c} \left[ \log P(y_c | S_C^M(x_p^m)) \right] \right].$$

The network $S_M$ is used to produce $\psi_m$ and $\psi_c^m$ from $x_m$ (solid blue arrow Fig. 2) and $x_c^m$ (solid orange arrow Fig. 2), respectively, to compute the gradients for $D_{struct}$. Gradient flow with respect to $G_C^M$, $S_M$, and $D_{struct}$ is shown in Fig. 3(b). The modified loss $L_{D_{struct}}^D$ to optimize $D_{struct}$, obtained by replacing $S$ in (5) with $S_M$ is computed as:

$$\max_{D_{struct}} L_{D_{struct}}^D = \mathbb{E}_{(x_m, \psi_m) \sim p(x_m, S_M(x_m))} \left[ \log (D_{struct}(x_m, \psi_m)) \right] + \mathbb{E}_{(x_p^m, \psi_c^m) \sim p(x_p^m, S_M(x_p^m))} \left[ \log (1 - (D_{struct}(x_p^m, \psi_c^m))) \right].$$

The network $S_C^M$ is used to produce $\psi_c^m$ from $x_m$ (dotted orange arrow Fig. 2) in order to compute the gradients for $G_C$. We use $\psi_c^m$ to differentiate from $\psi_m$ that is generated using $S_M$ in (8). The loss $L_{G_C^M}^D$ to optimize $G_C^M$ is now computed by replacing $S$ in (5) with $S_C^M$ and $x_c^m = G_C^M(x_m)$:

$$\min_{G_C^M} L_{G_C^M}^D = \mathbb{E}_{(x_p^m, \psi_c^m) \sim p(x_p^m, S_C^M(x_p^m))} \left[ \log (1 - (D_{struct}(x_p^m, \psi_c^m))) \right].$$

The network $S_M$, which is never used in generator update is used to segment target datasets at test time. We found that this network produces more accurate MRI segmentations than the $S_C^M$ as shown in results.

D. Additional losses for unpaired mapping

In order to train with unpaired CT and MR datasets, we enforce consistent reverse mapping by computing adversarial penalties from target to source generator $G_C^M$ and global intensity discriminator $D_C$. This loss is expressed as:

$$\max_{D_C} \min_{G_C^M} L_{adv}^{D_C}(G_C^M, D_C) = \mathbb{E}_{x_c \sim p(x_c)} \left[ \log (D_C(x_c)) \right] + \mathbb{E}_{x_m \sim p(x_m)} \left[ \log (1 - (D_C(G_C^M(x_m)))) \right].$$

The global adversarial loss is then computed as $L_{adv} = L_{adv}^{D_C} + L_{adv}^{D_{struct}}$. Cyclical consistency loss (11) is computed to account for lack of any pixel-level correspondence between the source and the target domain images in the I2I translation $(G_{C \circ M} = G_C(G_M(x_c)); G_{M \circ C} = G_M(G_C(x_m)))$ as:

$$L_{cyc}(G_C^M, G_M^C) = \mathbb{E}_{x_c \sim p(x_c)} \left[ \| G_{C \circ M}(x_c) - x_c \|_1 \right] + \mathbb{E}_{x_m \sim p(x_m)} \left[ \| G_{M \circ C}(x_m) - x_m \|_1 \right].$$

The total loss is expressed as:

$$L_{total} = L_{adv} + \lambda_{cyc} L_{cyc} + \lambda_{struct} L_{struct} + \lambda_{seg} L_{seg}.$$ (13)

The algorithm for the proposed method is shown in Algorithm 1.

Algorithm 1: PSIGAN

input : Source (CT) domain: $(X_c, y_c)$, Target (MRI) domain: $(X_m)$
output: Segmentation model $G_M$ for MRI
1 $\theta_M, \theta_D, \theta_{G_M}^I, \theta_{S_M}^I, \theta_{G_C^M}^I \leftarrow$ initialize ;
2 for $b \leq$ Maximum Iter do
3 $X_c, X_m \leftarrow$ sample mini-batch from CT, MRI domains;
4 $L_{adv}, L_{cyc}, L_s \leftarrow$ calculated using (11) ;
5 Compute segmentation probability map $\psi_c \leftarrow S_C^M$;
6 $L_{struct} \leftarrow$ calculated using (10) ;
7 $\theta_G \leftarrow -\Delta \theta_G (L_{adv} + \lambda_{cyc} L_{cyc} + \lambda_{struct} L_{struct} + \lambda_{seg} L_{seg})$;
8 Compute segmentation probability maps $\psi_m, \psi_c^m \leftarrow S_M$;
9 $L_{D_{struct}} \leftarrow$ calculated using (9) ;
10 $\Delta \theta_{D_{struct}} \leftarrow -\Delta \theta_{D_{struct}} (L_{adv} + \lambda_{struct} L_{struct})$;
11 $L_{seg} \leftarrow$ calculated using (8) ;
12 $\theta_{S_M}, \Delta \theta_{S_M} \leftarrow -\Delta \theta_{S_M} (L_{seg})$;
end

The generators $(G_M^C, G_C^M)$, discriminators $(D_M, D_C)$, and segmentors $(S_M, S_C^M)$ are updated with the following gradients, $-\Delta \theta_{G} (L_{adv} + \lambda_{cyc} L_{cyc} + \lambda_{struct} L_{struct} + \lambda_{seg} L_{seg})$, $-\Delta \theta_{D_{struct}} (L_{adv} + \lambda_{struct} L_{struct})$, and $-\Delta \theta_{S_M} (L_{seg})$, respectively.

The algorithm for the proposed method is shown in Algorithm 1.

E. Implementation and network structure

All networks were implemented using the Pytorch library and were trained on Nvidia GTX V100 with 16 GB memory. Training was done using ADAM algorithm [55] with an initial learning rate of 1e-4 and batch size of 2. We set $\lambda_{cyc}=10$, $\lambda_{struct}=0.5$ and $\lambda_{seg}=5$ in the training. The appropriate values for these hyper-parameters were selected empirically from the T2w MR parotid dataset set (see Supplementary document Sec. I). The learning rate was kept constant for the first 30 epochs and decayed to zero in the next 30 epochs.

The generator architectures were adopted from DC-GAN [36]. The generators $(G_M^C)$ and $G_C^M$ consisted of two stride-2 convolutions, 9 residual blocks and two fractionally strided convolutions with half strides and used tanh activation following the last convolutional layer to produce output images. The discriminator networks $(D_M, D_C$ and $D_{struct})$ were composed of 5 convolutional layers with a kernel size of $4 \times 4$ pixels that resulted in feature maps of size $64,128,256,512,1$ in each layer of these networks. Discriminators were implemented as $70 \times 70$ pixels patchGAN [37].
The segmentation network was based on the U-net [38] with batch normalization added after each convolution filter. Feature encoders, $E_M$ and $E_C^f$ of the segmentation network were implemented using 2 successive operations of convolution using kernels of size $(3 \times 3)$, batch normalization and $ReLU$ activation followed by max pooling. Four max-pooling operations were used in the encoder structure for subsampling the image feature maps leading to the feature sizes of 64, 128, 256 and 512. Skip connections from both the encoder networks layers are used to combine with the decoder layer features to prevent segmentation blurring. The decoder network (DE) was implemented using four unpooling operations to upsample the features back to the original image resolution.

F. Evaluation Metrics

Per-organ segmentation accuracies were measured using Dice similarity coefficient (DSC), computed from voxel-wise true positives (TP), false positives (FP) and false negatives (FN) ($DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$), and Hausdorff Distance at 95th percentile (HD95) as suggested in [39]. We also computed an overall DSC that is an average DSC over all the structures segmented in a given scan.

G. Compared methods

We compared our method against multiple UDA segmentation methods including the CycadaGAN [16], segmentation structure matching (SA) [28], segmentation entropy matching (ADVENT) [21], SynSeg-Net [3], and SIFA [17]. Also, we compared against the CycleGAN [11] with a U-net [38] and the UNIT [40] with a U-net. SynSeg [3] and SIFA [17] are methods that have been developed and applied to medical image segmentation, while all other compared methods were developed for analyzing natural images. Both SA and ADVENT methods employed additional segmentation matching based adversarial losses for domain adapted segmentation training. We also computed the performance on a model trained with only CT (source) images without domain adaptation. Moreover, we compared performance against supervised MRI segmentation model to ascertain the accuracy upper limit. The supervised MRI segmentation training was done by using the MRI validation set as the training set.

Default implementation of the various networks as available from the authors’ were used. All networks, including ours, were trained and tested using identical datasets with the same hyperparameter settings (learning rates and batch size).

IV. EXPERIMENTS AND RESULTS

We evaluated multiple organ and tumor segmentation from four different MRI sequences arising from two external institution (with two MR sequences) and two private and internal institution datasets. All networks were trained using $256 \times 256$ CT and MRI image patches enclosing the SOIs. All MR scans were standardized to remove patient-dependent signal intensity variations [41] and normalized to a range [-1,1] for meaningful computation of global and joint distribution adversarial losses.

Ablation experiments were done to assess the utility of the proposed structure discriminator ($L_{struct}$), the global intensity discriminator ($L_{adv}$) and cycle losses ($L_{cyc}$). We also evaluated alternative network design choices for computing the structure adversarial GAN losses by using (a) only segmentation probability maps, (b) joint distribution representation using multi-channel segmentation probability maps, (c) a single encoder-decoder segmentation network, and (d) SOI-specific discriminators.

A. MRI abdomen organs dataset

1) Data: Twenty MRIs (T1-DUAL in-phase and T2w spectral pre-saturation inversion recovery or SPIR) acquired on a 1.5T Philips machine from the Combined Healthy Abdominal Organ Segmentation (CHAOS) challenge data [42] were used to generate segmentation of liver, left and right kidneys, and spleen. Ten MRIs were used for training (without labels) and validation and the remaining 10 scans were used for testing. Both MRI sequences were acquired to perform fat suppression. These data sets were acquired using a 1.5T Philips MRI, with an image resolution of $256 \times 256$ pixels, slice thickness that ranged between 5.5mm to 9mm (average of 7.84mm). Additional details of the MR sequences are in supplemental document Sec.I.A.

Thirty contrast-enhanced portal venous phase CT scans with expert segmentations were obtained from a completely different dataset [43] for UDA training. The CT images had a resolution of $512 \times 512$ pixels, an in-plane resolution ranging between 0.54mm $\times$ 0.54mm to 0.98mm $\times$ 0.98mm, and slice thickness ranging between 2.5mm to 5.0mm. CT images were cropped to contain only body region and then resampled to $256 \times 256$ to have the same resolution as the MRI images. Following histogram standardization, T1w and T2w MRI were clipped to the range of [0,1136] and [0,1814] using the 95th percentile of the reference MRI intensity values, respectively. Separate segmentation networks were trained for the T1w and T2w MRI using $256 \times 256$ pixels image patches. These patches were extracted from 8000 T1w, 7872 T2w MRI slices, and 14038 CT slices.

We also evaluated the feasibility of our method to segment T2w MRIs acquired on different scanners with various repetition times (TR), echo times (TE), with and without fat suppression from external institution TCIA-LIHC [44] dataset. Six patients were downselected from a total of 97 patients; exclusion criteria were only CT scans (N=57), absence of T2w FSE/TSE MRI (N=22), motion artifacts (N=6), and large liver tumors(N=6). Table [I] shows the sequence details from the CHAOS and the six patients from TCIA-LIHC dataset.

2) Volumetric segmentation accuracies: The volumetric DSC segmentation accuracies for the multiple abdominal organs generated from the T2w and T1w MRI sequences on the testset are shown in Table [II]. A model trained with only CT images (w/o adaptation) was unable to produce clinically usable segmentations on MRI. PSIGAN achieved better overall average (computed over all organs) DSC of 0.90 on T2w and 0.87 on T1w MRI and a HD95 of 7.80mm on T2w and 6.90mm on T1w MRI than all other methods except supervised MRI segmentation. PSIGAN accuracy was only slightly lower than the fully supervised MRI segmentation.
Fig. 4 shows example segmentations generated by the various methods from T1w and T2w MRI. Segmentations produced without adaptation and supervised method are shown in Supplementary Fig. 2. As shown, the PSIGAN segmentations are nearly indistinguishable from the expert’s segmentations.

3) Evaluation on TCIA-LIHC dataset: PSIGAN produced an average DSC accuracy of 0.87 and an average HD95 accuracy of 8.59mm on this dataset, which is close to that achieved on the CHAOS dataset (Table II). PSIGAN segmentations were highly similar to expert delineations on all six cases with highly varying MR tissue contrasts (Supplementary Fig. 5).

4) MR to CT UDA segmentation: We also evaluated our and other approaches to segment CT images by using either T1w or T2w MRI as the source modality. Validation was done using 15 CTs and testing was done on remaining 15 cases. Fig. 5 shows segmentations computed on a representative test case using the segmentation accuracies for the left and right parotid glands using the various methods including, supervised MRI training and MRI segmentation obtained using a CT network trained without domain adaption. The CT model trained without any domain adaption was unable to generate segmentation on MRI. PSIGAN produced an average DSC of 0.82 and the lowest HD95 of 3.06mm, which was slightly lower than supervised method’s accuracy with DSC of 0.84 and HD95 of 2.26mm. The next best method SIFA reached an average DSC of 0.72 and HD95 of 4.99mm. Fig. 6 shows segmentations generated on a representative test case using the compared methods; the corresponding segmentations produced without adaptation and with the supervised method are in Supplementary Fig. 4. As seen, PSIGAN segmentations were nearly indistinguishable from the expert contours.

C. MRI lung tumor dataset

1) Data: A private institution dataset with 75 T2-weighted turbo spin-echo (T2wTSE) MRIs obtained from 27 non-small
cell lung cancer (NSCLC) cancer patients scanned before and every week during conventional fractionated external beam radiotherapy of 60 Gy was analyzed. Motion-robust two-dimensional axial images were acquired by using respiratory triggering on a 3T Philips Ingenia scanner (Medical Systems, Best, Netherlands). This is the same dataset as used in our prior work [10]. MRI images were clipped to the range of [0,1198] using the 95th percentile of the reference MRI intensity values following image standardization. Training was done on 9696 unlabeled 2D image patches containing lung tumor extracted from 35 longitudinal MRI scans of 5 patients, while independent testing was done on the remaining 40 MRI scans from 22 patients. The CT source domain data was obtained from 377 expert-segmented CT scans of NSCLC patients and available from the Cancer Imaging Archive (TCIA) [46]. Training was done using 32,000 image patches of size 256×256 pixels containing lung tumor.

2) 

Volumetric segmentation accuracies: Table IV shows the lung tumor segmentation accuracies achieved by the various methods. The CT segmentation model without domain adapt-
delineations. The corresponding segmentations without adaptation and with supervised segmentation are in Supplementary Fig. 3.

D. Differences in feature maps extracted using global and joint distribution discriminator

Fig. 8 shows four randomly chosen feature maps of the first convolutional layer from the global intensity discriminator that used only images (Fig. 8(c)), a discriminator that used only segmentation probability maps (Fig. 8(d)), and structure discriminator that used a concatenation of image and aggregated segmentation probability map (Fig. 8(e)). We visualized the first layer features due to their proximity to the input images and to ascertain what low-level features were relevant for the discriminator. As seen, the image intensity matching global intensity discriminator extracts features without a clear focus on any part of the image (Fig. 8(c)), while the segmentation probability matching discriminator amplified features at the SOI boundaries (Fig. 8(d)), thereby, emphasizing organ geometry. On the other hand, feature responses are higher both within and at SOI boundaries, with a slight emphasis on some background structures (e.g. bottom row Fig. 8(e)) when using the structure discriminator. As a result, our method allows the structure discriminator to focus on both geometry and the content within the SOIs, while also preserving some background features.

E. PSIGAN network design experiments

1) Structure discriminator: We evaluated segmentation performance under the following structure discriminator network designs: (i) multi-channel segmentation probability only, (ii) aggregated segmentation probability map only, (iii) channel-wise concatenation of image and multi-channel segmentation probability, (iv) SOI specific structure discriminators that used channel-wise concatenation of image and segmentation probability map for each SOI, and (v) default PSIGAN, which used channel-wise concatenation of image and aggregated segmentation probability maps. These tests were done on both the MRI parotid and the T1w SPIR MRI abdomen dataset (CHAOS). Separate networks were trained in each one of these settings from scratch. The main difference between settings (i) and (ii) vs. the rest was the use of segmentation probability only in (i) and (ii) vs. joint-distribution matching of images and segmentation probabilities in (iii), (iv), and (v). Both (iii) and (v) used a single structure discriminator as opposed to $K$ structure discriminators in (iv). Finally, whereas $K$ channels for each organ was used to represent segmentation probability map in (iii), a single aggregated segmentation probability map was used in (v). All other losses including the global intensity discriminator and cycle consistency losses as used in PSIGAN were used all experiments.

As shown in Table V, the settings (iii), (iv), and (v) produced more accurate segmentations than (i) and (ii). Furthermore, the default setting of PSIGAN that aggregates the segmentation probabilities into a single map produced more accurate segmentation than all other methods. SOI-specific discriminator setting (iv) was similarly accurate as the setting using multi-channel segmentation probability maps in (iii). Segmentations produced on a representative case from the CHAOS dataset by all these methods is shown in Fig. 9. Segmentations produced on a case from the MR parotid dataset is in Supplementary Fig. 9.

2) Segmentation network: We evaluated performance when using single or split segmentor networks (Fig. 3). The structure discriminator and generator losses were computed using (5) and (6) for single segmentor, as opposed to using (9), (10) for the split segmentor case. In addition, we also measured the accuracy of segmentor $S_M^M$ for generating MRI segmentations instead of the default $S_M$ network used in PSIGAN. Both split
and single-segmentor networks were trained from scratch till convergence with identical training, validation, and testing sets for both parotid and T1w SPIR abdomen organs segmentation.

The single-segmentor configuration was less accurate than the split-segmentor (Table. II and Table. III) with an average DSC of 0.85 on the T1w SPIR abdomen segmentation, and an average DSC of 0.80 on the parotid segmentation, respectively. The network parameters were slightly higher for the split-segmentor (42.97M) compared with single segmentor (38.99M). Adversarial losses for $G^M$ and $D_{struct}$ during training are shown for split and single network configurations in Supplementary Fig. 10. As shown, the gap in losses for $G^M$ and $D_{struct}$ stabilized faster for the split than the single segmentor configuration.

The $S^M$ network produced a much lower accuracy than both $S_M$ and single-segmentor networks with an average DSC of 0.84 on the T1w SPIR abdomen segmentation, and an average DSC of 0.77 on the parotid segmentation, respectively. Organ-specific accuracies are in Supplementary Table II. Segmentations on representative examples using these three networks are shown in Supplementary Fig. 7 and Fig. 8 for both datasets.

I. Ablation experiments

1) Contribution of structure discriminator, global intensity discriminator, and cycle losses on accuracy: The goal of this experiment was to evaluate the contribution of each loss on segmentation performance. Both structure and global intensity discriminator compute an adversarial loss and can train the generator independent of each other. Separate networks were trained from scratch until convergence using identical training, validation, and testing sets from the T2wSPIR MRI parotid datasets with the following loss settings:

1) CT to MR global intensity discriminator loss $L_{adv}^{C \rightarrow M}$ (4): Only a global adversarial loss was computed using the discriminator $D_M$ for CT to MRI I2I translation.

2) CT to MR, MR to CT global intensity discriminators, and cycle losses $L_{adv}^{C \rightarrow M} + L_{adv}^{M \rightarrow C} + L_{cycle}$ (4, 11, 12): Cycle consistency, global adversarial losses computed using $D_M$ and $D_C$ (for MR to CT translation) were used.

3) Structure discriminator loss $L_{struct}$ (9, 10): Joint distribution (image and aggregated segmentation prob-
ability map) matching adversarial loss was computed from $D_{struct}$.

4) Structure and CT to MR global intensity discriminator losses $L_{struct} + L_{adv}^{C\rightarrow M}$ (9), (10), (4): Losses from setting 1 and 3 were combined.

5) Structure, MR to CT global intensity discriminator, and cycle losses $L_{struct} + L_{adv}^{M\rightarrow C} + L_{cyce}$ (9), (10), (11), (12): Cycle loss and loss from setting 3 were combined.

6) Structure, CT to MR, MR to CT global intensity discriminators, and cycle losses $L_{struct} + L_{adv}^{C\rightarrow M} + L_{adv}^{M\rightarrow C} + L_{cyce}$: Default PSIGAN.

As shown in Table VI, $L_{struct}$ loss alone leads to a clear performance improvement compared with the combination of global adversarial and cycle losses ($L_{adv}^{C\rightarrow M} + L_{adv}^{M\rightarrow C} + L_{cyce}$). Addition of either the CT to MR global adversarial (setting 4) or cycle losses (setting 5) to the $L_{struct}$ loss resulted in equivalent performance improvements. PSIGAN, which combines all the losses produced the most accurate segmentation. Segmentations produced on a representative case by the various networks trained in the aforementioned settings are shown in Supplementary Fig. 11.

2) Impact of structure discriminator on I2I translation:

Fig. 10 shows I2I translations produced by networks trained without and with structure discriminator on two example images, first one from the T2w SPIR MRI abdomen and the second from the T2wFS MRI parotid dataset. The source CT image is shown for reference. As shown, the addition of structure discriminator improved the contrast of SOIs with respect to background Fig. 10(c) and more accurately modeled the internal characteristics like the regularity in the organization of blood vessels in the liver, as appearing on real MRI. Additional I2I translation results are in Supplementary Fig. 12.

Quantitative comparison of the distribution of MR signal intensities within the SOIs between the pseudo and real MRIs were computed using Kullback-Leibler (KL) divergence metric.  The method trained without the structure discriminator produced a higher KL-divergence of 1.5 within the parotid glands and 0.14 within liver. Whereas the method trained using structure discriminator produced a K-L divergence of 0.05 for parotid glands and 0.018 for liver.

G. Evolution of segmentation probability maps during training

Fig. 11 shows the evolution of segmentation probability maps produced on representative examples from the three analyzed datasets during early epochs in training. As seen, the various organs and the tumor are correctly detected and the segmentation probabilities improve, becoming sharper and more focused with training. The higher probabilities are indicated by red color while low probabilities correspond to blue color. These maps clearly indicate their usefulness to constrain I2I translation after training only for a few epochs.
Fig. 11. Example PSIGAN segmentation probability maps produced during the early training epochs. First column shows expert segmentations on MRI. The structure segmentation probabilities (blue - low probability, and red - high probability) steadily increase with training.

V. DISCUSSION

We introduced PSIGAN, a joint distribution matching method for unsupervised domain adaptation-based multiple organ segmentation. Our method produced highly similar accuracy as supervised method on three different datasets, indicating its ability to learn without requiring target modality labeled image sets. Our method also showed feasibility to segment CT scans when performing T1w to CT and T2w to CT UDA segmentation.

Joint distribution matching has previously been used to constrain the space of GAN mappings by learning the bidirectional mapping from image to a scalar latent variable [23], [22] or from image to vector of class categories [23]. To our knowledge, ours is the first to perform joint distribution matching using pairs of images for UDA segmentation.

We conducted ablation and network design experiments to determine the utility of the joint distribution matching structure discriminator for both I2I translation and segmentation. Our results show clear performance improvements when using joint distribution matching adversarial losses. Consistent with prior work that computed adversarial losses using segmentation [21], [8] or with a joint translation-segmentation network [3], [14], we also found that features extracted using adversarial discriminators using only the segmentation maps showed a strong preference for SOI geometry, by emphasizing features on SOI boundary. On the other hand, the joint-distribution matching discriminator heavily weighted features corresponding to both the geometry and the appearance features within the organs.

We also found that joint distribution formulations that used aggregated segmentation probability maps yielded more accurate segmentations than formulations using either multi-channel segmentation maps or SOI-specific structure discriminators. Performance improvement in the aggregated case could have resulted from increased contextual information from the other organs that was available to the structure discriminator.

Finally, the split segmentor showed a small improvement in accuracy over the single segmentor. However, the choice of the split segmentation network for MRI segmentation clearly impacted accuracy. More specifically, the network that was used for computing discriminator gradients was more accurate than the one used for computing the generator gradients.

Our idea of using segmentation probability maps to guide the translation is similar in principle to attention-guided translation methods [47], [48], which iteratively focus the domain translation network on to regions of interest and produce the desired translation. The main difference is that our method handles simultaneous translation of multiple target and background structures while attention-guided methods are typically restricted to transfiguring a single foreground. Also, as the optimization of the segmentation network is done in a supervised manner, pre-specified image to target relationships can be easily extracted through a fast converging network to constrain translation. Deriving unsupervised attention information on the other hand, would require pre-training of the self-attention network for several epochs before it can be combined with the generator as shown in [48].

A deficiency of our method as is common to most UDA methods is the inability to handle expert delineation variabilities across modalities that may arise as visibility of structures can vary across modalities. Preliminary evaluation of our method on an external dataset with a variety of MR contrasts and scanning parameters indicates that it is possibly robust to MR contrast variations. However, extensive validation and potential extension to handle large MR contrast variations on much bigger cohorts are needed and is work for future. Nevertheless, our method outperformed multiple state-of-the-art methods.

VI. CONCLUSION

We developed and evaluated a new unpaired domain adaptation segmentation approach using joint distribution matching structure discriminator for multiple organ segmentation on MRI datasets. Our approach outperformed multiple state-of-the art methods and demonstrated the value of structure discriminator in improving I2I translation and segmentation.

REFERENCES

[1] P. Kupelian and J. Sonke, “Magnetic-resonance guided adaptive radiotherapy: a solution to the future,” Semin Radiat Oncol, vol. 24, no. 3, pp. 227–32, 2014.
[2] H. Bainbridge, A. Salem, R. Tijssen, M. Dube, A. Wetscher, E. C. Van et al., “Magnetic resonance imaging in precision radiation therapy for lung cancer,” Trans Lung Cancer Research, vol. 6, no. 6, pp. 689–707, 2017.
[3] Y. Huo, Z. Xu, H. Moon, S. Bao, A. Assad, T. K. Moyo et al., “Synseg-net: Synthetic segmentation without target modality ground truth,” IEEE Trans. on Med. Imaging, vol. 34, no. 4, pp. 1016–1025, 2018.
[4] K. Kamnitas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane et al., “Unsupervised domain adaptation in brain lesion segmentation with adversarial networks,” in Information Processing in Medical Imaging, 2017, pp. 597–609.
[5] Z. Zhu, B. Du, and P. Yan, “Boundary weighted domain adaptive neural network for prostate MR image segmentation,” IEEE Trans. Med Imaging, no. 3, pp. 753–763, 2019.
H. Y. Lee, H. Y. Tseng, J. B. Huang, M. Singh, and M.-H. Yang, “Diverse application to cross-modality liver segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent, 2018, pp. 669–677.

Y. Li, L. Yuan, and N. Vasconcelos, “Bidirectional learning for domain adaptation of semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2019, pp. 6936–6945.

Q. Dou, C. Ouyang, C. Chen, H. Chen, B. Glocker, X. Zhuang et al., “Pop-AdaNet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation,” IEEE Access, vol. 7, pp. 99065–99076, 2019.

J. Yu, T. Park, P. Isola, and A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in Int. Conf. Computer Vision. ICCV, 2017, pp. 2223–2232.

H. Y. Lee, H. Y. Tseng, J. B. Huang, M. Singh, and M.-H. Yang, “Unsupervised domain adaptation via disentangled representations,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2018, pp. 35–51.

J. Yang, N. C. Dvornek, F. Zhang, J. Chapiro, M. Lin, and J. S. Duncan, “Unsupervised domain adaptation via disentangled representations: Application to cross-modality liver segmentation,” Proc. Int. Conf. Med. Image Comput. Assist. Intervent, pp. 255–263, 2019.

Z. Zhang, L. Yang, and Y. Zheng, “Translating and segmenting multimodal medical volumes with cycle- and shape consistency generative adversarial network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2018, pp. 9242–9251.

K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, “Unsupervised pixel-level domain adaptation with generative adversarial networks,” in Proc. IEEE Conf. Computer Vision and Pattern Recognit, 2017, vol. 1, no. 2, pp. 3722–3731.

J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko et al., “CyCADA: Cycle-consistent adversarial domain adaptation,” in Proc. Int. Conf. on Machine Learning, vol. 2018, pp. 1989–1998.

C. Chen, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, “Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation,” arXiv preprint arXiv:1901.08211, 2019.

T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” in Adv. Neural Inf. Process. Syst, 2016, pp. 2234–2242.

C. Chen, Q. Dou, H. Chen, and P.-A. Heng, “Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest X-Ray segmentation,” in Machine Learning in Medical Imaging, 2018, pp. 143–153.

Y.-H. Tsai, K. Sohn, S. Schultzer, and M. Chandraker, “Domain adaptation for structured output via discriminative patch representations,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2019, pp. 1456–1465.

T. H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, “ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2019, pp. 2517–2526.

C. Li, H. Liu, C. Chen, Y. Pu, L. Chen, R. Henao et al., “ALICE: Towards understanding adversarial learning for joint distribution matching,” Proc. Adv. Neural Inf. Process. Syst, pp. 5495–5503, 2017.

J. Donahue, P. Krähenbühl, and T. Darrell, “Adversarial feature learning,” arXiv preprint arXiv:1605.06782, 2016.

L. Chongxuan, T. Xu, J. Zhu, and B. Zhang, “Triple generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst, 2017, pp. 4088–4098.

Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, “Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss,” arXiv preprint arXiv:1804.10916, 2018.

T. Joyce, A. Chartisars, and S. A. Tsafaris, “Deep multi-class segmentation without ground-truth labels,” in Proc. Int. Conf. Medi. Imag. with Deep Learning, 2018.

N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. Xing, “Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent, 2018, pp. 544–552.

Y.-H. Tsai, W.-C. Hung, S. Schultzer, K. Sohn, M.-H. Yang, and M. Chandraker, “Learning to adapt structured output space for semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2018, pp. 7472–7481.

Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim, “Image to image translation for domain adaptation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2018, pp. 4500–4509.

F. Liu, “SUSAN: segment unannotated image structure using adversarial network,” Magnetic resonance in medicine, vol. 81, no. 5, pp. 3330–3345, 2019.

H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, and L. Cheng, “Supervised segmentation of un-annotated retinal fundus images by synthesis,” IEEE Trans. on Med. Imaging, vol. 38, no. 1, pp. 46–56, 2018.

J. Cai, Z. Zhang, L. Cui, Y. Zheng, and L. Yang, “Towards cross-modal organ translation and segmentation: A cycle-and shape-consistent generative adversarial network,” Med. Im. Ana., vol. 52, pp. 174–184, 2018.

J. P. Cohen, L. Margaux, and H. Sina, “Distribution matching losses can hallucinate features in medical image translation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent, 2018, pp. 529–536.

K. Armannious, C. Yang, M. Fischer, T. Küstner, K. Nikolau, S. Gaditidis et al., “MedGan: Medical Image Translation using GANs,” vol. abs/1806.06397, 2018.

D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” Proceedings of the 3rd Int. Conf. on Learning Representations, 2014.

A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv preprint arXiv:1511.06434, 2015.

P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit, pp. 1125–1134, 2017.

O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent, 2015, pp. 234–241.

B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby et al., “The multimodal brain tumor image segmentation benchmark (BRATS),” IEEE Trans. on Med. Imaging, vol. 34, no. 10, pp. 1993, 2015.

M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” in Proc. Adv. Neural Inf. Process. Syst, 2017, pp. 700–708.

L. G. Nyul and J. K. Udupa, “On standardizing the MR image intensity scale,” Magnetic Resonance in Medicine, vol. 42, no. 6, pp. 1072–1081, 1999.

A. E. Kavour, M. A. Selver, O. Dicle, M. Barıs, and N. S. Gezer, “CHAOSS - Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data,” Apr. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.3362844

B. Landman, Z. Xu, J. Iglesias, M. Stynier, T. Langerak, and A. Klein, “MICCAI multi-atlas labeling beyond the cranial vault-workshop and challenge,” 2015.

B. Erickson, S. Kirk, Y. Lee, O. Bathe, M. Kearns, C. Gerdes et al., “Radiology data from the cancer genome atlas liver hepatocellular carcinoma [TCGA-LIHC] collection the,” Cancer Imaging Archive, 2016.

P. F. Raudaschl, P. Zaffino, G. C. Sharp, M. F. Spadea, A. Chen, B. M. Dawant et al., “Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015,” Medical physics, vol. 44, no. 5, pp. 2020–2036, 2017.

H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho et al., “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach,” Nature communications, vol. 5, pp. 4066, 2014.

Y. A. Meijati, C. Richardt, J. Tompkin, D. Cosker, and K. I. Kim, “Image to image translation using GANs,” in Proc. Adv. Neural Inf. Process. Syst, 2018, pp. 3693–3703.

H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, “Self-attention generative adversarial networks,” in Proc. of Int. Conf. on Machine Learning, vol. 97, 2019, pp. 7354–7363.