Unpaired Data based Cross-domain Synthesis and Segmentation Using Attention Neural Network

Xiaoming Liu LXMSPACE@GMAIL.COM
Xiangkai Wei AITEWEI@FOXMAIL.COM
Aihui Yu METAPHORLOVE@FOXMAIL.COM
College of Computer Science and Technology, Wuhan University of Science and technology, Wuhan, 430065, China
Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, 430065, China

Zhifang Pan PANZHIFANG@WMU.EDU.CN
Information Technology Center, Wenzhou Medical University, 325035, China

Editors: Wee Sun Lee and Taiji Suzuki

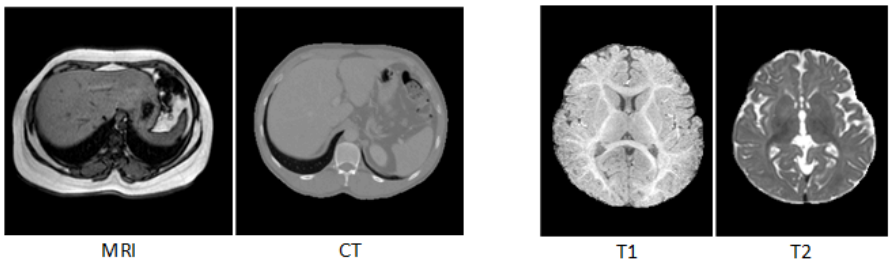
Abstract

Medical images from different modalities (e.g. MRI, CT) or contrasts (e.g. T1, T2) are usually used to extract abundant information for medical image analysis. Some modalities or contrasts may be degraded or missing, caused by artifacts or strict timing during acquisition. Thus synthesizing realistic medical images in the required domain is meaningful and helpful for clinical application. Meanwhile, due to the time-consuming of manual annotation, automatic medical image segmentation has attracted much attention. In this paper, we propose an end-to-end cross-domain synthesis and segmentation framework SSA-Net. It is based on cycle generative adversarial network (CycleGAN) for unpaired data. We introduce a gradient consistent term to refine the boundaries in synthesized images. Besides, we design a special shape consistent term to constrain the anatomical structure in synthesized images and to guide segmentation without target domian labels. In order to make the synthesis subnet focusing on some hard-to-learn regions automatically, we also introduce the attention block into the generator. On two challenging validation datasets (CHAOS and iSeg-2017), the proposed method achieves superior synthesis performance and comparable segmentation performance.

Keywords: medical image, synthesis and segmentation, unpaired data, CycleGAN

1. Introduction

Different feature information are helpful to improve medical image analysis (Frangi et al. (2018)) and clinical diagnosis (Wenqi Lu and Qiang Su (2016)), which are extracted from multiple modalities (e.g. MRI, CT) or multiple contrast (e.g. T1, T2 in MRI). However, some sequences often be missing or destroyed during collection. Cross-modality or cross-contrast (they are referred as cross-domain in the paper) medical image synthesis has attracted much attention and can be used as an intermediate for medical image analysis tasks (Dar et al. (2018)). But it still faces a significant challenge that the appearance difference between different domains are complex (see Fig. 1).



(a) (b)

Figure 1: Appearance differences between unpaired sequences of different modalities or contrast. (a) CHAOS dataset (CHAOS (2019)), the real slice size is 256×256. (b) iSeg-2017 dataset (Wang et al. (2019)), the real slice size is 144×192.

Unlike intra-modality generation (Frid-Adar et al. (2018); Isola et al. (2016)), cross-domain synthesis is based on different domains (Huo et al. (2019)). Previous cross-domain synthesis methods (Burgos et al. (2012); Huang et al. (2017); Jog et al. (2017)) usually based on image patches and with a lot of computational overhead. Several deep-learning based methods are developed recently, such as Generative adversarial network (GAN) (Goodfellow et al. (2014); Dar et al. (2018)) based methods, which often require the aligned data. Recently, CycleGAN-based methods (Zhu et al. (2017); Hiasa et al. (2018)) are proposed to handle unpaired data. However, basci CycleGAN is not enough to supervise the result of synthesis, such as boundary (Hiasa et al. (2018)).

Segmentation labels are important in clinical applications and deep-learning based segmentation (Pham et al. (2019); Liu et al. (2019)) is also prevalent. Recently, several synthesis-based segmentation methods (Chartsias et al. (2017); Zhang et al. (2018); Huo et al. (2019)) has shown the potential to use realistic synthesized data to improve segmentation, and they are generally based on the CycleGAN framework. Among them, (Chartsias et al. (2017)) is a two-stage method (see Fig. 2(a)), which performs synthesis before segmentation.

Inspired by the above works, we propose a CycleGAN-based one-stage medical image synthesis and segmentation framework (Fig. 2(b)), which is end-to-end trained by unpaired data. In order to constrain the consistency of anatomical structure and boundaries, we design a gradient consistent (Hiasa et al. (2018)) and a special shape consistent (Zhang et al. (2018)) loss as additional loss terms. Only labels in the source domain are required in the proposed method, which means higher utilization of existing data and better adaptability. Moreover, an modified Attention U-net (Oktay et al. (2018)) is used as generator to focus on hard-to-synthesis image regions automatically.

We summarize the contributions of our work as follows:

(1) We propose a novel CycleGAN-based end-to-end cross-domain synthesis and segmentation network, with using unpaired data and without target domain labels.

(2) We integrate the gradient- and shape-consistency terms as additional loss terms to supervise the boundary and the anatomical structure separately. We introduce an adjusted Attention U-Net into synthesis subnet to refine the hard-to-learn regions during synthesis.

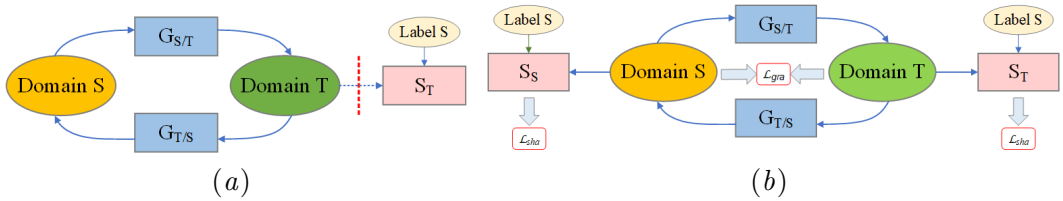


Figure 2: The diagram of two-stage method (Chartsias et al. (2017)) and our proposed two-stage method, they are based on CycleGAN. "Domain S" and "Domain T" represent the source and the target domain, respectively. $G_{S/T}$ is a source-to-target generator, $G_{T/S}$ is a target-to-source generator. S_S is a segmentor for the source domain and S_T is a segmentor for the target domain. "Label S" is the segmentation label in the source domain. Blue arrows indicate the flow of data information. $\mathcal{L}$ denotes the loss term. (a) Cycle+Seg (Chartsias et al. (2017)). (b) Our proposed SSA-Net.

(3) Our proposed method has shown good performance on two challenging datasets. As far as we konw, this is also the first end-to-end cross-domain synthesis and segmentation on iSeg-2017.

2. Related Work

2.1. Cross-domain Synthesis for Medical Image

For medical images, cross-domain synthesis can predict the target modal or contrast of a given object, by learn the pixel-level mapping from source to target domain (Chartsias et al. (2017)). Technically, they can be divided into three categories: 1) registration-based methods, 2) sparse-coding-based methods, and 3) deep-learning-based methods. Registration-based methods (Burgos et al. (2012)) are to register the atlas to the target and are sensitive to the registration accuracy. Sparse-coding-based methods (Huang et al. (2017)) are mainly based on encoding the non-linear mapping between different domain patches. However, both classes of methods are time-consuming and ignore spatial information. Several learning-based schemes are also patch-based and to learn the pixel-wise mapping between different domain, such as (Jog et al. (2017)). Various deep-learning-based approaches have brought new potential for cross-domain synthesis, such as GAN (Goodfellow et al. (2014)). Several GAN-based method (Dar et al. (2018)) has been introduced to enhance the fidelity of synthesis by using game theory. However, there is a common limitation that they are mainly based on paired data in different domains. In order to solve it, CycleGAN (Zhu et al. (2017)) combines two GANs with a cycle consistency to receive unpaired data. In (Hiasa et al. (2018)), a gradient consistency is integrated into CycleGAN to refine the boundaries.

2.2. Segmentation with Synthesis

Cross-domain synthesiszed images can be used to improve the medical image processing performance, such as segmentation (Iglesias et al. (2012); Liu et al. (2018)). These synthesis-based segmentation methods are grouped into two classes: 1) two-stage methods (Chartsias et al. (2017)), they first synthesize and then segment separately. 2) one-stage methods

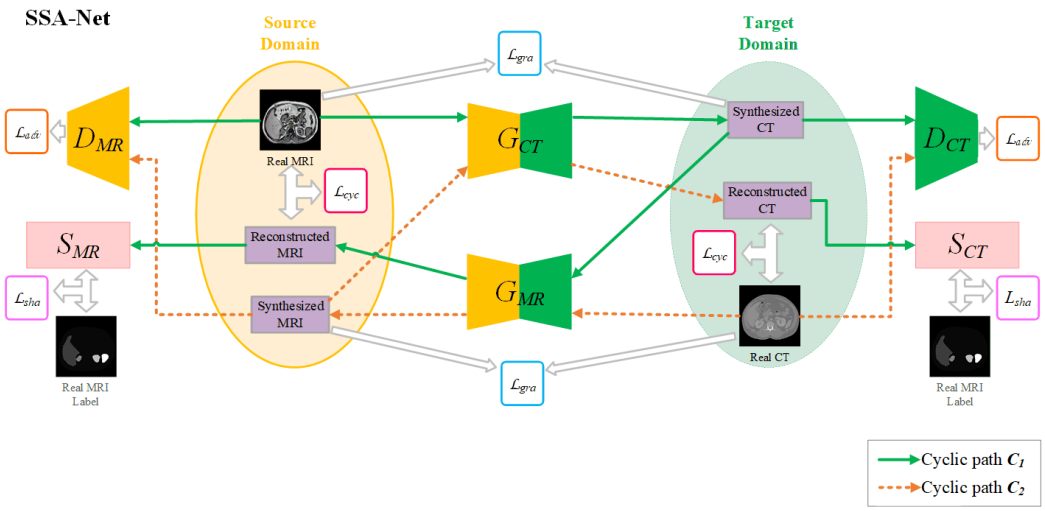


Figure 3: The proposed SSA-Net is based on a CycleGAN with four loss terms.

(Huo et al. (2019); Zhang et al. (2018)), they integrate synthesis and segmentation in end-to-end. As shown in Fig. 2(a), (Chartsias et al. (2017)) introduced a two-stage method via CycleGAN, which first obtained synthesized data for the next segmentation with using target labels. For the one-stage methods, (Huo et al. (2019)) integrated a CycleGAN and a segmentation network with using only source domain label. TranSeg-Net (Zhang et al. (2018)) added a segmentor at each domain of the CycleGAN, and used a shape consistent loss with using the both domains label to made the synthesis and segmentation work together. All these methods used CycleGAN to handle unpaired data, and training segmentation with synthesized data. We integrate the advantages of both of them (Fig. 2(b)) and propose a shape consistency loss with only requires source domain labels, which means better adaptability and higher utilization of data.

3. The Proposed Method

Taking synthesis of MRI-to-CT as an example (MRI is the source and CT is the target domain), the workflow of the proposed method is depicted in in Fig. 3. The proposed SSA-Net consists of a synthesis subnet and a segmentation subnet (they are trained in end-to-end), and the inputs of it include real MRI, real CT and labels on MRI. The model can be divided into two cyclic paths C_1 (the blue arrows) and C_2 (the red arrows) according to the data flow. In path C_1, the generator G_{CT} (for MRI-to-CT) generates synthesized CT from real MRI, and then the generator G_{MR} reconstructs the synthesized CT back to the source domain as the reconstructed MRI. In path C_2, fake MRI is synthesized from real CT by generator G_{MR} (for CT-to-MRI), and then it is input to G_{CT} to be reconstructed back to the target domain as the reconstructed CT. Here, two segmentors only worked in cyclic path C_1, segmentor S_{CT} segments synthesized CT and segmentor S_{MR} segments reconstructed MRI separately. It should be noted that the labels only in MRI are used as segmentation ground truth for both segmentors, while working with loss $\mathcal{L}_{sha}$ to supervise the shape consistency in C_1. We also design a gradient consistent term $\mathcal{L}_{gra}$ to constraint

the boundary in the synthesis subnet. In both C_1 and C_2, adversarial learning is achieved by two discriminator (D_{MR} or D_{CT}) with the adversarial loss $\mathcal{L}_{adv}$, and an original cycle consistent loss $\mathcal{L}_{cyc}$ in CycleGAN (Zhu et al. (2017)) is also performed. During testing, real MRI is input to G_{CT} (the parameters in the model have been fixed) to obtain the realistic synthesized CT, and the segmentation results are also output by S_{CT}.

3.1. Synthesis and Segmentation Network

In this paper, the source domain MRI and target domain CT are marked as X_{MR} and X_{CT}. The proposed network consists of two subnets: synthesis subnet and segmentation subnet. They are introduced as follows.

3.1.1. CycleGAN-based Synthesis Subnet with Attention

We use the Attention U-Net structure (see Fig. 4) as the generators G_{MR} and G_{CT}. This U-Net consists of a contraction and an expansion path, including three down-sampling blocks (convolution with max pooling), three upsampling blocks (convolution with deconvolution) and several convolutions with stride 1. We also employ Instance normalization (IN) and rectification linear unit (ReLU). Note that the original skip connections in U-Net are replaced by the attention blocks, which connect the contraction and the expansion layers. As shown in Fig. 4, the attention block can calculate a grid-wise attention matrix from the feature maps to enhance or suppress (by element-wise multiplication) the propagation of some local features. It allows the generators focus on the hard-to-synthesis region during optimizing. We use the basic basic PatchGAN (Isola et al. (2016)) as the discriminators to identify synthesized local region with the adversarial loss $\mathcal{L}_{adv}$.

These two generators and the two discriminators form a CycleGAN-based synthesis subnet with a cycle consistent loss $\mathcal{L}_{cyc}$ (Zhu et al. (2017)), which simultaneously learns two cross-domain mappings (MRI-to-CT and CT-to-MRI). Besides, a gradient consistent loss $\mathcal{L}_{gra}$ is introduced to constraint the texture in synthesized result, and a segmentation subnet with the shape consistent loss $\mathcal{L}_{sha}$ is also used to constrain the anatomical structure during synthesis, they are described in Section 3.2.

3.1.2. Segmentation Subnet via U-Net

We employ a simplified U-Net (Huo et al. (2019)) as the structure of our segmentors to obtain reliable segmentation results on the synthesized images, as shown in Fig. 5. The segmentor consists of two symmetrical contraction (down-sampling block with max pooling) and expansion (upsampling block with unpooling) paths. Here are several skip connections (as shortcuts) between different layers to achieve deeper supervision.

The segmentation subnet works with a special shape consistent loss $\mathcal{L}_{sha}$, which can constrain the structural consistency during synthesis and guide segmentation. $\mathcal{L}_{sha}$ is driven by two segmentors, S_{CT} (in target domain) segments the synthesized target domain images and S_{MR} (in source domain) segments the reconstructed image returned to the source domain. In order to better deal with the limitation of data acquisition in real situations, we make a compromise between performance and generalization. Here only source domain labels are used as the ground truth for segmentation, thus $\mathcal{L}_{sha}$ mainly supervise the cyclic path C_1. However, because the whole framework is trained in end-to-end and the object

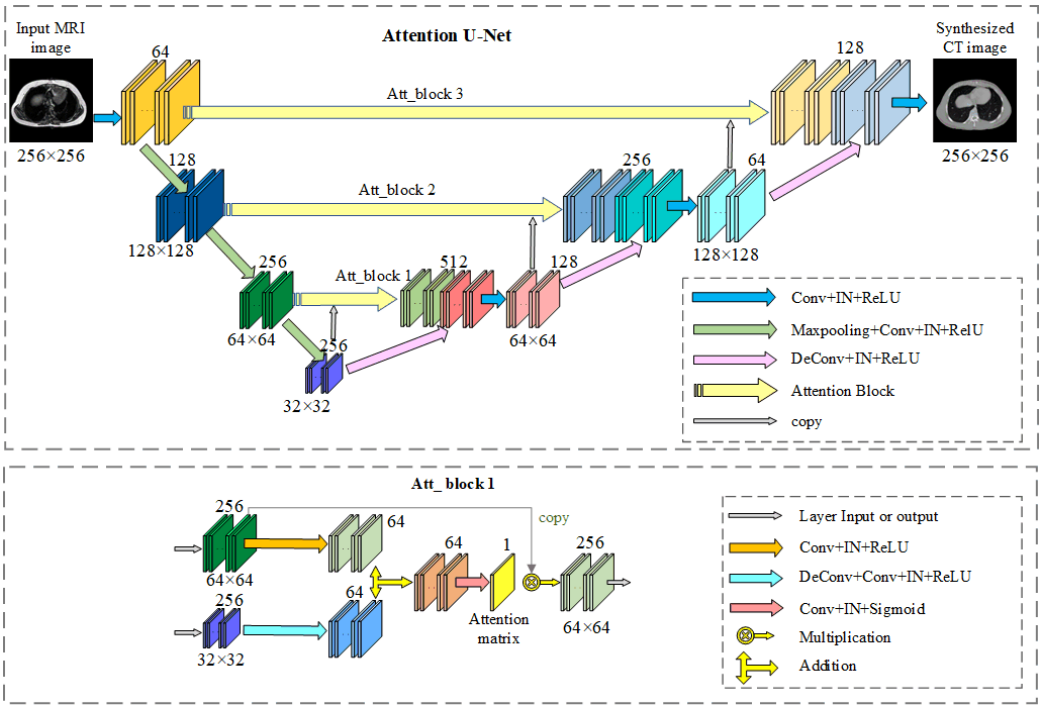


Figure 4: The structure of the proposed SSA-Net. Each group of overlapped color cubes represents a group of feature maps. The channel number of the feature maps is marked on the top of the cube, the size is marked on the bottom.

supervised by $\mathcal{L}_{sha}$ belongs to both domains, shape consistent loss can also indirectly work in the C_2.

3.2. Objective Function

During the training, the total objective loss $\mathcal{L}_{total}$ (Eq. 1) is a combination of four loss terms: an adversarial loss $\mathcal{L}_{adv}$, a cycle consistent loss $\mathcal{L}_{cyc}$, a gradient consistent loss $\mathcal{L}_{gra}$ and a shape consistent loss $\mathcal{L}_{sha}$. Here α, β, γ and δ are the weight coefficients. The adversarial loss $\mathcal{L}_{adv}$ is the least square error (Qi (2017)) between the synthesized and real images. The cycle consistent loss $\mathcal{L}_{cyc}$ refers to (Zhu et al. (2017)), which can force two cyclic paths to capture the features from their respective domains and allows CycleGAN to handle unpaired data. The gradient consistent and shape consistent loss are described in detail bellow.

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{adv} + \beta\mathcal{L}_{cyc} + \gamma\mathcal{L}_{gra} + \delta\mathcal{L}_{sha} \tag{1}$$

3.2.1. The Gradient Consistent Loss

In order to supervise the contours in synthesized images, we use a gradient consistent loss $\mathcal{L}_{gra}$ (see Eq. 2) to maximize the gradient correlation (GC, between 0 and 1) (Penney et al. (1998)) between synthesized and real images. Here x_{mr}, x_{ct} are two unpaired images in both domains, G_{CT} is the generator for CT domain (MRI-to-CT), and G_{MR} is the generator for

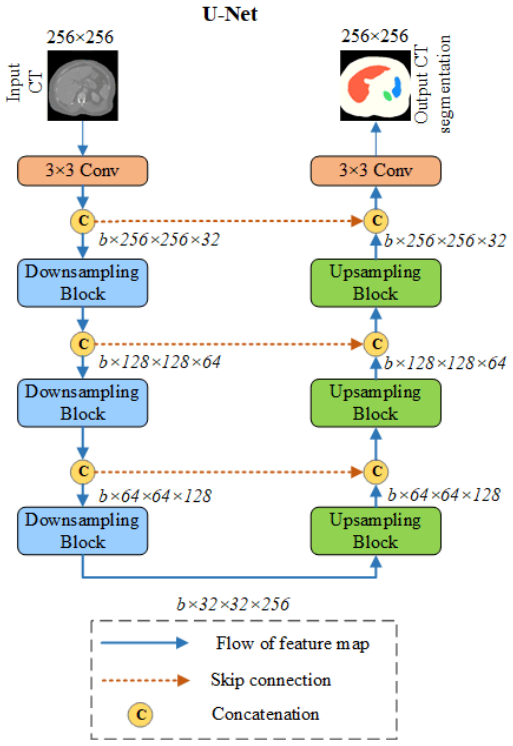


Figure 5: The U-Net used as the segmentors in the proposed SSA-Net, includes Down-sampling block and Up-sampling block. The size $batchsize \times width \times heigth \times channel$ of the feature maps is marked around the arrow.

MRI domain (CT-to-MRI).

$$\mathcal{L}_{gra} = \sum_{x_{mr} \in X_{MR}} (1 - GC(x_{mr}, G_{CT}(x_{mr}))) + \sum_{x_{ct} \in X_{CT}} (1 - GC(x_{ct}, G_{MR}(x_{ct}))) \tag{2}$$

For two given images p and q, their gradients in both horizontal and vertical directions are marked as p_v, q_v, p_h and q_h. Then their GC is calculated according to Eq. 3, where $NCC(\cdot)$ is the normalization cross-correlation (NCC) (Hiasa et al. (2018)).

$$GC(p, q) = \frac{1}{2} \left[NCC(p_v, q_v) + NCC(p_h, q_h) \right] \tag{3}$$

3.2.2. The Shape Consistent Loss

In order to alleviate the geometric distortion (Zhang et al. (2018)) caused by insufficient constraint in basic CycleGAN and obtain reliable segmentation in target domains, we design a shape consistent loss $\mathcal{L}_{sha}$ (see Eq. 4) with two segmentors. S_{CT} is the segmentor for CT domain, and S_{MR} is the segmentor for MRI domain.

$$\mathcal{L}_{sha} = \mathcal{L}_{sha}(G_{CT}, S_{CT}) + \mathcal{L}_{sha}(G_{MR}, G_{CT}, S_{MR}) \tag{4}$$

Here $\mathcal{L}_{sha}(G_{CT}, S_{CT})$ is the shape consistency error between the synthesized target image and the source image, and $\mathcal{L}_{sha}(G_{MR}, G_{CT}, S_{MR})$ is between the reconstructed image

of the synthesized target image and the source image. They are defined as two multi-class cross-entropy:

$$\mathcal{L}_{sha}(G_{CT}, S_{CT}) = -\sum_{\substack{x_{mr}\in X_{MR}\\ y_{mr}\in Y_{MR}}} (y_{mr}\log(S_{CT}(G_{CT}(x_{mr})))) \tag{5}$$

$$\mathcal{L}_{sha}(G_{MR}, G_{CT}, S_{MR}) = -\sum_{\substack{x_{mr}\in X_{MR}\\ y_{mr}\in Y_{MR}}} (y_{mr}\log(S_{MR}(G_{MR}(G_{CT}(x_{mr}))))) \tag{6}$$

Here $y_{mr} \in Y_{MR}$ represents the segmentation labels of MRI inputs and used as the shape ground truth for shape consistency. Unlike (Zhang et al. (2018)), here only source domain (MRI) labels are required to directly constrain C_1, which also indirectly affect C_2 due to the features can cyclically propagate among different domains in CycleGAN.

4. Experimentation and Results

4.1. Dataset

To investigate the performance of cross-domain synthesis and segmentation of the methods, an abdominal dataset CHAOS (CHAOS (2019)) is used for cross-modality experiment, and an infant brain MRI dataset iSeg-2017 (Wang et al. (2019)) is used for cross-contrast experiment.

CHAOS (CHAOS (2019)) includes unpaired CT and MRI images as well as their labels. Note that the labeled organs in both modalities are different, CT label marks liver and abdomen, MRI label marks abdomen, liver, left kidney, right kidney, and spleen. It allows our method to try to segment organs unmarked in CT (target domain) label (see Section 4.4). We extracted a total of 1120 unpaired MRI/CT images (256×256) and their labels, 80% (896) of them are used for training, and the remaining 20% (224) are used for testing.

For iSeg-2017 (Wang et al. (2019)), there are more fuzzy boundaries and irregular tissues in the images, beacause the brain of 6-month infant is undeveloped. It means greater unpredictability for image processing task. As far as we know, there is no cross-domain synthesis and segmentation method has been explored on this dataset before. iSeg-2017 includes T1 and T2 images as well as their labels (label marks cerebrospinal fluid (CSF), white matter (WM) and gray matter (GM)). To make these two contrast sequences unaligned, we disrupt the order of them. A total of 1200 T1/T2 images (144×192) and their labels are extracted. 80% (960) of them are training samples, and the remaining 20% (240) are testing samples.

4.2. Implementation Details and Compared Methods

All method are carried on a Linux workstation with a NVIDIA Tesla V100-SXM2 GPU (16 GB memory). During training, the coefficients of four loss terms are set to $\alpha = 1$, $\beta = 10$, $\gamma = 0.5$ and $\delta = 0.5$ respectively.

We select two advanced synthesis and segmentation methods to compare with the proposed method. They are (1) Cycle+Seg (Chartsias et al. (2017)) (Fig. 2(a)) and (2) TranSeg-Net (Zhang et al. (2018)). All methods are evaluated with the same dataset partitioning, and noting that the target domain labels are used to train TranSeg-Net (Zhang

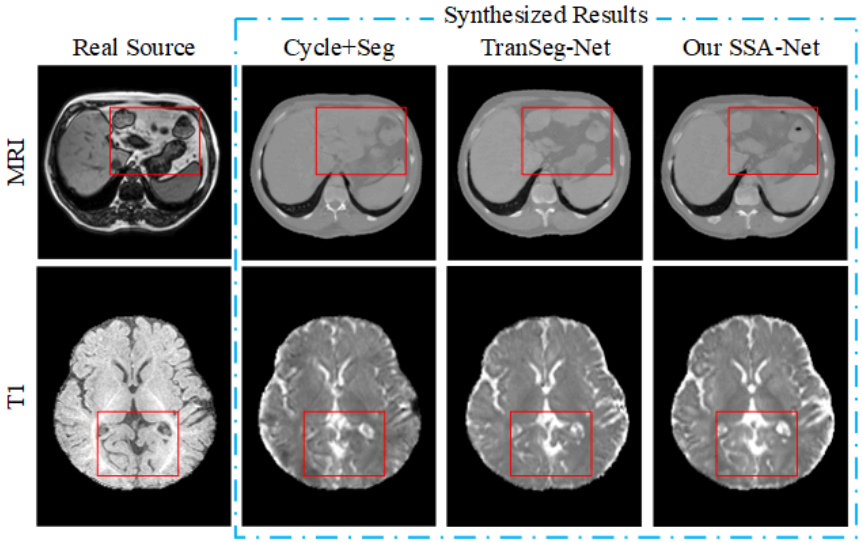


Figure 6: The visual results of synthesized images in target domain of all methods. The first row shows the results on CHAOS, and the second row shows the results on iSeg-2017.

et al. (2018)). Because of the lack of the corresponding ground truth in target domain (this paper are based on unpaired data), we chose an indirect evalution criteria S-score (Zhang et al. (2018)) (see Eq. 7) for synthesis, which is worked with an additional independent segmentation network S_{test}. S_{test} (using a basic U-net in this paper) is independently trained with the real images and manual labels in target domain. The S-score is calculated by Dice similarity coefficient (DSC). Here x_{mr} and y_{mr} represent the source MRI image and its segmentation label, N denotes the number of testing images.

$$S\text{-score} = \frac{1}{N}\sum_{i \in N} DSC_i(S_{test}(G_{CT}(x_{mr})), y_{mr}) \tag{7}$$

4.3. Results of Cross-Domain Synthesis

To investigate the cross-domain synthesis performance of our method, all methods are evaluated on two datasets (CHAOS (CHAOS (2019)) and iSeg-2017 (Wang et al. (2019))). For quantitative evaluation with S-score, S_{test} is trained by real images and labels in target domain (CT or T2). Here, the results of Cycle+Seg (Chartsias et al. (2017)) are as the baseline.

4.3.1. Visual Comparison and Qualitative Analysis

The synthesized results of our method and other comparative methods on both datasets (CHAOS and iSeg-2017) are displayed in Fig. 6. Among the synthesized results (synthesized CT and synthesized T2) on both datasets, the proposed SSA-Net have achieved best visual performance than other two compared methods. The result of the baseline Cycle+Seg (Chartsias et al. (2017)) has rough edges of tissue and low contrast between different organs. TranSeg-Net (Zhang et al. (2018)) and our SSA-Net obviously outperform

Table 1: The S-scores of our method and other comparative methods on two datasets for cross-domain synthesis. The largest is marked in blod and the second on italics

Methods	**CHAOS**		**iSeg-2017**	
	CT	MRI	T2	T1
Cycle+Seg (baseline)	0.796±0.052	0.764±0.043	0.684±0.049	0.647±0.036
TranSeg-Net	*0.863±0.035*	*0.825±0.032*	0.716±0.031	0.677±0.021
SSA-Net (ours)	**0.877±0.028**	**0.846±0.034**	**0.734±0.033**	**0.692±0.023**

the baseline, due to the advantage of additional semantic segmentation information (labels) used to strengthen supervision. Furthermore, the result of our proposed SSA-Net has clearer contours and higher contrast than TranSeg-Net (Zhang et al. (2018)). As shown in the red boxes, our model can obtain better details and clearer boundaries by integrating gradient consistent loss, which guarantees the contrast between different tissues. Besides, some small details are also retained in the result of SSA-Net, because of the attention component can make our generators attend to detailed texture.

4.3.2. Quantitative Comparison and Analysis

The S-scores of cross-domain synthesis on two datasets are presented in Tab. 1, the proposed SSA-Net outperforms other methods and achieves the highest S-score. Similar to visual results, TranSeg-Net (Zhang et al. (2018)) and our method exceed the Cycle+Seg (Chartsias et al. (2017)) (baseline) to a certain extent. This reveals that for the medical images with complex textures and specificity, a single CycleGAN is not sufficient to map the synthesis transition between different domains. This also means that additional semantics segmentation information can provide more constraints to synthesis and promote it. The S-score of TranSeg-Net (Zhang et al. (2018)) is higher than Cycle+Seg (Chartsias et al. (2017)), because all semantic labels in both domains are used in TranSeg-Net (Zhang et al. (2018)) to ensure the shape consistency. Although only MRI labels are used in our method to supervise the shape consistency to C_1, the attention component and the gradient consistent term are mainly conducive to improve synthesis (see the ablation experiment in Section 4.5) and can compensate for this weakness.

4.4. Results of Cross-domain Synthetic Segmentation

A comparison is carried out on all methods to verify the performance of segmentation based on cross-domain synthesized images. Taking the cross-domain synthetic segmentation of source-to-target as the example, the visual results and quantitative analysis are described below.

Qualitative analysis. The results of synthetic segmentation of all methods are visualized in Fig. 7. The segmentation result on CHAOS of two-stages Cycle+Seg (Chartsias et al. (2017)) has obvious deficiencies, such as the contours of segmented organ are irregular. The segmentation results of TranSeg-Net (Zhang et al. (2018)) and our model are close to the ground truth. For the challenging iSeg-2017 with Six-month infant brain, all methods can segment CSF reliably, but there is under-segmentation in other two tissues (especially

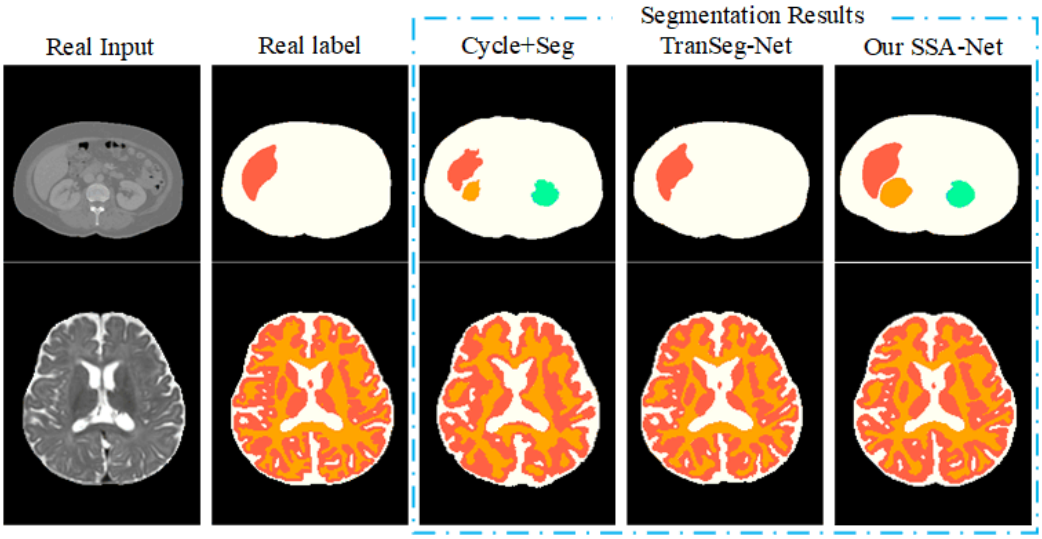


Figure 7: Visual results of cross-domain synthetic segmentation for all methods on two datasets. The first row is the results on CHAOS, white marks abdomen, orange marks liver, yellow marks right kidney, green marks left kidney, and blue marks spleen. The second row is the results on iSeg-2017, white marks cerebrospinal fluid, orange marks gray matter, and yellow marks white matter.

Table 2: The synthetic segmentation DSC scores of all methods for different tissues on two datasets. The largest is marked in blod and the second on italics

Methods	CHAOS		iSeg-2017		
	Abdomen	Liver	CSF	GM	WM
Cycle+Seg (baseline)	0.854±0.059	0.788±0.046	0.816±0.045	0.697±0.053	0.684±0.048
TranSeg-Net	**0.916±0.033**	**0.824±0.038**	**0.869±0.041**	**0.764±0.041**	*0.745±0.042*
SSA-Net (ours)	*0.910±0.031*	*0.805±0.033*	*0.855±0.034*	*0.757±0.037*	**0.751±0.033**

in WM). Compared with Cycle+Seg (Chartsias et al. (2017)), other two one-stage models (TranSeg-Net and our SSA-Net) have better segmentation performance. This confirms that one-stage framework optimizes the synthesis and segmentation networks jointly, so that synthesis and segmentation can promote each other. Note that the results of TranSeg-Net (Zhang et al. (2018)) on both datasets outperform the proposed SSA-Net. This is because TranSeg-Net (Zhang et al. (2018)) uses both domain labels to optimize segmentors and achieve the shape consistency. Although our proposed SSA-Net only use source domain labels, this also brings more generalization and flexibility to our framework, such as we can segment the organs (kidney and spleen on CHAOS) that do not exist in the target domain label. Besides, the segmentation contour of our method is smooth, which may be the effect of gradient consistency.

Quantitative analysis. The segmentation DSC scores on two datasets of all methods are listed in Tab. 2. It can be seen that all the methods get the lower DSC scores on

Table 3: The S-scores and DSC scores of our method and two ablation versions on CHAOS dataset. The largest is marked in blod and the second on italics

Methods	S-**score(synthesis)**	**DSC(segmentation)**	
	CT	Abdomen	Liver
SSA-Net NA	0.846±0.024	0.876±0.034	0.784±0.032
SSA-Net NG	*0.860±0.033*	*0.892±0.030*	*0.791±0.037*
SSA-Net (ours)	**0.877±0.028**	**0.910±0.031**	**0.805±0.033**

iSeg-2017, which indicates that segmentation of 6-month infant brain is also a challenge. Compared to Cycle+Seg (Chartsias et al. (2017)), our SSA-Net and TranSeg-Net (Zhang et al. (2018)) achieve better segmentation performance based on synthesis, which benefits from end-to-end optimized synthesis and segmentation subnets. Our SSA-Net gets higher scores than the Cycle+Seg (Chartsias et al. (2017)) on all tissues, which is due to the additional consistent loss terms can strengthen the supervision of the synthesis and indirectly improve the segmentation. Note that our model is second only (even the best result on WM) to TranSeg-Net (Zhang et al. (2018)), which is due to additional target domain labels used in TranSeg-Net (Zhang et al. (2018)) can directly improve the segmentation, but the best result on WM segmentation also reveals the potential of our method. Although our SSA-Net sacrifices little performance for segmentation with less labels, it also achieves better generalization and practicability.

4.5. Ablation Experiment

The ablation experiment is carried out on the CHAOS datasets with two abridged versions (SSA-Net NA and SSA-Net NG) of our proposed method. SSA-Net NA replaces the attention block with the ordinary skip connection in U-Net. SSA-Net NG removes the gradient consistent loss term $\mathcal{L}_{gra}$ from the objective function. The quantitative results (S-score and DSC) are listed in Tab. 3. The complete version (SSA-Net) surpasses SSA-Net NA and SSA-Net NA in both metircs, which illustrates that these two components are positive for the proposed method in synthesis and segmentation. Besides, we can see that the attention block improves the synthesis scores (S-score) more than the gradient consistent loss term. It implies that the attention is more conducive to improve synthesis performance.

5. Conclusion

In this paper, we propose a novel one-stage cross-domain synthesis and segmentation method using unpaired data, which is CycleGAN-based and only requires the labels in source domain. We use the gradient prior information in synthesis and integrate the segmentation network to supervise the shape consistency. Besides, we introduce an attention U-Net as generator to automatically attend to the hard-to-synthesis regions. Experiments on two datasets (CHAOS and iSeg-2017) demonstrate the effectiveness of the proposed SSA-Net, which achieves good performance on cross-domain synthesis and shows the potential of segmentation.

Acknowledgments

This work is partially supported in part by the National Natural Science Foundation of China under Grants 61403287, 61472293, and 61572381, and in part by the Natural Science Foundation of Hubei Province under Grant 2014CFB288.

References

Ninon Burgos, M. Jorge Cardoso, Kris Thielemans, Marc Modat, Stefano Pedemonte, John Dickson, Anna Barnes, Rebekah Ahmed, Colin J. Mahoney, and Jonathan M. Schott. Attenuation correction synthesis for hybrid pet-mr scanners: Application to brain studies. *IEEE Transactions on Medical Imaging*, 33(12):2332–2341, 2012.

CHAOS. Isbi chaos challenge: Combined (ct-mr) healthy abdominal organ segmentation. https://chaos.grand-challenge.org/, 2019.

Agisilaos Chartsias, Thomas Joyce, Rohan Dharmakumar, and Sotirios A. Tsaftaris. Adversarial image synthesis for unpaired multi-modal cardiac data. 2017.

Salman Ul Hassan Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Çukur. Image synthesis in multi-contrast mri with conditional generative adversarial networks. *IEEE Transactions on Medical Imaging*, PP(99):1–1, 2018.

Alejandro F. Frangi, Sotirios A. Tsaftaris, and Jerry L. Prince. Simulation and synthesis in medical imaging. *IEEE Transactions on Medical Imaging*, 37(3):673–679, 2018.

Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 2018.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Xu Bing, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. 2014.

Yuta Hiasa, Yoshito Otake, Masaki Takao, Takumi Matsuoka, Kazuma Takashima, Jerry L. Prince, Nobuhiko Sugano, and Yoshinobu Sato. Cross-modality image synthesis from unpaired data using cyclegan: Effects of gradient consistency loss and training data size. 2018.

Yawen Huang, Ling Shao, and Alejandro F. Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. 2017.

Y. Huo, Z. Xu, H. Moon, S. Bao, A. Assad, T. K. Moyo, M. R. Savona, R. G. Abramson, and B. A. Landman. Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE Transactions on Medical Imaging*, 38(4):1016–1025, April 2019. ISSN 0278-0062. doi: 10.1109/TMI.2018.2876633.

Juan Eugenio Iglesias, Ender Konukoglu, Darko Zikic, Ben Glocker, Koen Van Leemput, and Bruce Fischl. Is synthesizing mri contrast useful for inter-modality analysis? 2012.

Phillip Isola, Jun Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. 2016.

Amod Jog, Aaron Carass, Snehashis Roy, Dzung L. Pham, and Jerry L. Prince. Random forest regression for magnetic resonance image synthesis. *Medical Image Analysis*, 35: 475–488, 2017.

X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu. Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning. *IEEE Access*, 7:3046–3061, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2889321.

Xiaoming Liu, Dong Liu, Tianyu Fu, Zhifang Pan, Wei Hu, and Kai Zhang. Shortest path with backtracking based automatic layer segmentation in pathological retinal optical coherence tomography images. *Multimedia Tools and Applications*, (2), 2018.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven Mcdonagh, Nils Y Hammerla, and Bernhard Kainz. Attention u-net: Learning where to look for the pancreas. 2018.

G.P. Penney, J. Weese, J.A. Little, P. Desmedt, D.L.G. Hill, and D.J. hawkes. A comparison of similarity measures for use in 2-d-3-d medical image registration. *IEEE Transactions on Medical Imaging*, 17(4):586–95, 1998.

D. D. Pham, G. Dovletov, S. Warwas, S. Landgraeber, M. Jäger, and J. Pauli. Deep learning with anatomical priors: Imitating enhanced autoencoders in latent space for improved pelvic bone segmentation in mri. pages 1166–1169, April 2019. doi: 10.1109/ISBI.2019.8759221.

Guo Jun Qi. Loss-sensitive generative adversarial networks on lipschitz densities. 2017.

Li Wang, Dong Nie, Guannan Li, Elodie Puybareau, Jose Dolz, Qian Zhang, Fan Wang, Jing Xia, Zhengwang Wu, and Jiawei and Chen. Benchmark on automatic 6-month-old infant brain segmentation algorithms: The iseg-2017 challenge. *IEEE Transactions on Medical Imaging*, PP(99):1–1, 2019.

Wenqi Lu and Qiang Su. Impact of clinical information quality on healthcare diagnosis quality: An empirical study. pages 1–5, June 2016. doi: 10.1109/ICSSSM.2016.7538608.

Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. 2018.

Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017.