High Tissue Contrast MRI Synthesis Using Multi-Stage Attention-GAN for Glioma Segmentation

Mohammad Hamghalam,[1,2] **Baiying Lei,**[1*] **Tianfu Wang**[1]
[1]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, 518060,
[2]Faculty of Electrical, Biomedical and Mechatronics Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
m.hamghalam@gmail.com, {leiby, tfwang}@szu.edu.cn

Abstract

Magnetic resonance imaging (MRI) provides varying tissue contrast images of internal organs based on a strong magnetic field. Despite the non-invasive advantage of MRI in frequent imaging, the low contrast MR images in the target area make tissue segmentation a challenging problem. This paper demonstrates the potential benefits of image-to-image translation techniques to generate synthetic high tissue contrast (HTC) images. Notably, we adopt a new cycle generative adversarial network (CycleGAN) with an attention mechanism to increase the contrast within underlying tissues. The attention block, as well as training on HTC images, guides our model to converge on certain tissues. To increase the resolution of HTC images, we employ multi-stage architecture to focus on one particular tissue as a foreground and filter out the irrelevant background in each stage. This multi-stage structure also alleviates the common artifacts of the synthetic images by decreasing the gap between source and target domains. We show the application of our method for synthesizing HTC images on brain MR scans, including glioma tumor. We also employ HTC MR images in both the end-to-end and two-stage segmentation structure to confirm the effectiveness of these images. The experiments over three competitive segmentation baselines on BraTS 2018 dataset indicate that incorporating the synthetic HTC images in the multi-modal segmentation framework improves the average Dice scores 0.8%, 0.6%, and 0.5% on the whole tumor, tumor core, and enhancing tumor, respectively, while eliminating one real MRI sequence from the segmentation procedure.

Introduction

Among brain tumors, glioma is the most prevalent tumor that begins from the tissue of the brain and can affect the brain function (Elazab et al. 2018). In brain magnetic resonance (MR) images, the intensity distributions of pixels are largely overlapping in regions of interest (ROIs), therefore leading to low tissue contrast and creating the main challenge for tissue segmentation. In glioma, ROIs exhibit similar levels of intensity in MR images, making tissue segmentation quite challenging. Fig. 1(a) demonstrates a brain lesion in the FLAIR MR slice with three overlapping tissues: whole tumor (WT), tumor core (TC), and enhancing tumor (ET). We define target images as a high-contrast do-

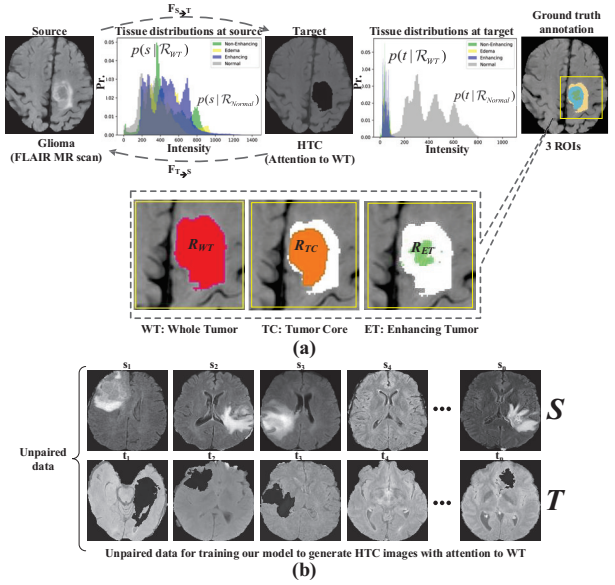

Figure 1: The MR images show low tissue contrast in the source domain, implying the most challenging for tissue segmentation. (a) A glioma lesion in the FLAIR MR image (left) with its intensity distributions (i.e., non-enhancing, edema, and enhancing) as well as normal tissue distribution. The corresponding HTC target image (middle) with attention to WT is achieved based on the manual labels (right). (b) Unpaired training data, consisting of a source set (first row) $s \sim p(s)$ and a target set (second row) $t \sim p(t)$, with no information provided as to which s matches which t.

*Corresponding author. This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, 61801305, and 81571758), National Natural Science Foundation of Guangdong Province (No. 2017A030313377), Guangdong Pearl River Talents Plan (2016ZT06S220), Shenzhen Peacock Plan (Nos. KQTD2016053112051497 and KQTD2015033016104926), and Shenzhen Key Basic Research Project (Nos. JCYJ20170413152804728, JCYJ20180507184647636, JCYJ20170818142347251, and JCYJ20170818094109846).

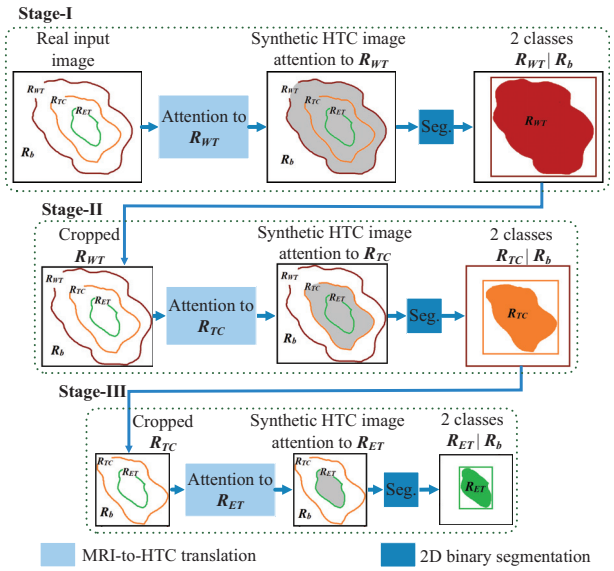

Figure 2: Design of the proposed multi-stage structure for segmentation of glioma within three stages. In each step, we first generate synthetic HTC images with the minimum overlapping area for the class-conditional densities between $(\mathcal{R}_f)$ and $(\mathcal{R}_b)$ through MRI-to-HTC block. Then, the synthetic HTC images are applied for segmentation in both end-to-end and two-stage training tactics. Synthetic HTC images deal with $\mathcal{R}_{WT}$, $\mathcal{R}_{TC}$, and $\mathcal{R}_{ET}$ in each stage, sequentially.

main such that tissues have limited overlapping area, while the source domain has overlapping tissue distributions. Our goal is to increase the intensity contrast between the underlying tissue region and others through image-to-image translation technique based on unpaired training data (Fig. 1(b)) to improve segmentation performance.

Image translation aims to learn the mapping between an input image following the source domain distribution to an output image with a defined distribution using a training set of paired (Isola et al. 2017) or unpaired images (Zhu et al. 2017). Despite the limitations of the synthetic images in the clinical application, these data have suggested promising results through generative adversarial networks (GANs) (Goodfellow et al. 2014), including data augmentation (Bowles et al. 2018), image reconstruction (Sharma and Hamarneh 2019), and segmentation (Huo, Xu, and Moon 2019; Zhang, Yang, and Zheng 2018; Chartsias et al. 2017; Nie et al. 2018; Wolterink et al. 2017; Zhao et al. 2017). The paired methods require training source images that aligned with the target ones to learn image generation through the forward adversarial loss, while the unpaired approaches frequently employ unaligned images through the CycleGAN structure.

The CycleGAN models have two main components, i.e., a source-to-target and a target-to-source block. Each part consists of a generator G and a discriminator D. G aims to generate a real image from a noise vector and an input image, while D is finding the difference between an actual image and the image produced by G. The key challenges in medical image synthesis either inter-modality (T1-to-T2, FLAIR-to-T1, and others) or cross-modality (MRI-to-CT, PET-to-CT, PET-to-MRI) translation are to predict the structure and fine-grained content of the target modality from the source one (Huo, Xu, and Moon 2019; Zhang, Yang, and Zheng 2018; Chartsias et al. 2017; Nie et al. 2018; Wolterink et al. 2017; Zhao et al. 2017). The CycleGAN provides effective supervision using cycle consistency between the source inputs and the reconstructed images as well as between the target images and corresponding reconstructed ones.

However, state-of-the-art medical image synthesis methods are restricted by the model's disability to attend a specific tissue. In this paper, we propose a multi-stage model to segment only one tissue through a segmentation block following an attention-guided synthesis block in each stage (Fig. 2). Specifically, the synthesis block generates high tissue contrast (HTC) images with attention to the relevant tissue for the segmentation task. In image synthesis block, we have two mappings: MRI-to-HTC and HTC-to-MRI. The former accepts 2D MR slices and generates HTC images, which further fed into the latter to reconstruct the input MR images. In the segmentation block, the HTC images are passed to the convolution layers to produce a binary segmentation map and a bounding box for the next stage. To provide attention to the specific tissue during synthesis process, two strategies have been used: (1) attachment of the attention block into the CycleGAN, (2) using high contrast image during training phase. The attention block guides G towards the expected region for translation via an attention map. This trainable map is further employed in D input to filter out irrelevant areas. Regarding the training, we use the ground-truth (GT) labels to form images with the minimum overlapping area between the tissue intensity distribution of the foreground and background in each stage (depicted tissue distribution in Fig. 1 (a)).

Furthermore, to produce a more detailed synthesis and consequently more accurate segmentation map, we explore the multi-stage architecture to deal with only one region in each stage. This structure alleviates the artifacts of the synthesized images by decreasing the gap between the source and target domain. The attention module effectively learns attention maps to guide the generator attentively select more important regions for generating an HTC image. The generated HTC image closely follows the distribution of the target domain and boosts the segmentation performance significantly. Besides, our model is based on the CycleGAN framework to leverage the vast quantities of unpaired data sets for training within the same modality. The experiments are conducted on multi-modal BraTS 2018 dataset (Menze et al. 2015) to segment internal parts of glioma. Specifically, we employ real modalities, i.e., FLAIR, T2, and T1c, to generate synthetic one with attention to the WT, TC, and ET in each stage, respectively. The contributions of this paper are summarized as:

- We design a novel framework to increase the contrast among sub-regions of glioma in MR images. Training on high contrast images as well as an unsupervised attention block inside the adversarial network guide our model to

pay attention to the particular regions.

- We propose a multi-stage structure that decreases the gap between the source and target domain to enhance the resolution of synthetic HTC images.
- We employ HTC MR images in both the end-to-end and two-stage segmentation structure on BraTS dataset to confirm the effectiveness of these images.

Related works

Segmentation

Numerous machine learning (Hatami et al. 2019) and deep learning methods have been introduced to address segmentation problems, especially glioma subregions (Soleymanifard and Hamghalam 2019). Fully convolutional networks (FCNs) (Long, Shelhamer, and Darrell 2015; Ronneberger, Fischer, and Brox 2015; Drozdzal et al. 2016; Chen et al. 2018; Jégou et al. 2017) as an extension of convolutional neural networks (CNNs) (He et al. 2016; Huang et al. 2017) with down-sampling and up-sampling layers have been considered as a benchmark of segmentation. Replacement of fully connected layers with convolution layers facilitates FCNs to take the global features and provides localization in an end-to-end framework (Long, Shelhamer, and Darrell 2015). In U-Net (Ronneberger, Fischer, and Brox 2015), authors used U-shaped architecture of FCNs with the skip connection to combine features extracted in the encoder side to the decoder ones. In other work, Drozdzal *et al.* (Drozdzal et al. 2016) added the *residual blocks* (He et al. 2016) to the U-Net framework to improve the segmentation accuracy by reducing the effect of vanishing gradient (Res-U-Net). Chen *et al.* (Chen et al. 2018) also extended the fully convolutional version of residual networks (ResNets) (He et al. 2016) by incorporating the dilation to the main structure. Jegou *et al.* (Jégou et al. 2017) continued the DenseNet (Huang et al. 2017) to fully convolutional DenseNet (FC-DenseNet) without post-processing for segmentation. This architecture leads to implicit in-depth supervision and allows capturing contextual information.

Segmentation in Adversarial Framework

Adversarial methods have been successfully exploited in medical image analysis to address the shortage of large and diverse annotated databases (Bowles et al. 2018), missing/corrupted MR pulse sequences (Sharma and Hamarneh 2019), as well as boost the segmentation performance in typical applications. These latter approaches can be categorized as two-stage training techniques (Chartsias et al. 2017; Wolterink et al. 2017; Zhao et al. 2017; Nie et al. 2018; Hamghalam, Lei, and Wang 2019) and end-to-end methods (Huo, Xu, and Moon 2019; Zhang, Yang, and Zheng 2018). The former considers the synthesis and segmentation as two individual training stages, while the latter incorporates the segmentation loss into the adversarial loss during the training.

Chartsias *et al.* (Chartsias et al. 2017) produced synthetic cardiac data from unpaired images coming from different individuals (CT-to-MRI cardiac image) based on CycleGAN. They found that training on both real and synthetic images lead to a statistically significant improvement compared to training on real data. Wolterink *et al.* (Wolterink et al. 2017) proposed MRI-to-CT synthesis on pairwise aligned training images of the same patient in the treatment planning of brain tumors. They analyzed paired and unpaired image mapping from 2D brain MR image slices into 2D CT ones. Authors found that the synthetic CT images taken via the model trained with unpaired data seemed more realistic, contained fewer artifacts than those obtained through the model trained with paired data. Zhao *et al.* (Zhao et al. 2017) introduced a multi-atlas based hybrid approach to synthesize T1w MR images from CT and CT images from T1w MR images using random forest synthesis framework. This method used a set of random forest regressors within each label for synthesizing intensities on pairs of MR and CT images of the whole head. In other works, Nie *et al.* (Nie et al. 2018) first applied FCN Model to generate MR from CT image as well as 7T MR from 3T MR images based on the CycleGAN. Next step, they employed synthetic images for the task of semantic segmentation.

In the end-to-end framework, Huo *et al.* (Huo, Xu, and Moon 2019) integrated the CycleGAN and segmentation into an end-to-end structure to train a segmentation network for both MRI-to-CT and CT-to-MRI without having manual labels in the target modality. In their architecture, called SynSeg-Net, authors demonstrated that end-to-end training achieved better performances compared to the two-stage one for segmentation. Zhang *et al.* (Zhang, Yang, and Zheng 2018) presented 3D cross-modality synthesis approach (CT-to-MRI) to segment cardiovascular volumes by adding shape-consistency loss to the CycleGAN framework. They also validated that coupling the generator and segmentor module resulted in better segmentation accuracy than training them exclusively.

Method

The proposed framework is composed of K stages, where K denotes the number of labels in input images. Each step consists of two main modules: (1) image synthesis with attention, and (2) segmentation block. The former is learned in an adversarial framework to generate synthetic HTC images with attention to an individual foreground $\mathcal{R}_f^{(k)}$, while the latter performs supervised binary segmentation for the foreground and background $\mathcal{R}_b^{(k)}$ region. The bounding box which calculated from the segmentation map in stage k will be considered for the next step, $k+1$. Fig. 2 shows an overview of the proposed structure for segmentation of brain lesion with three regions ($K=3$), including $\mathcal{R}_{WT}$, $\mathcal{R}_{TC}$, and $\mathcal{R}_{ET}$. This section first describes how the image synthesis block transforms the tissue intensity distribution of the foreground from source to target domain, and then provides details of incorporating the synthetic images into the segmentation framework which is expected to produce more accurate results than using real MR images.

HTC Image Synthesis via Attention-GAN (MRI-to-HTC)

Let MR source image at stage k, $s^{(k)} \in S^{(k)}$, be the union of foreground, $s_f^{(k)}$, and background pixels, $s_b^{(k)}$, in the source domain as:

$$s = [s_f \sim p(s|\mathcal{R}_f)] \cup [s_b \sim p(s|\mathcal{R}_b)] \quad (1)$$

we omit the superscript k for simplicity. Similarly, in the target domain, we have HTC image, $t^{(k)} \in T^{(k)}$ as:

$$t = [t_f \sim p(t|\mathcal{R}_f)] \cup [t_b \sim p(t|\mathcal{R}_b)] \quad (2)$$

where $p(s|\mathcal{R})$ and $p(t|\mathcal{R})$ are the class-conditional distributions of tissue in the source and target domain, respectively. We also assume that the distribution of the foreground and background have a little overlap in the target space.

Our goal in each stage is to estimate a mapping function, $F_{S\to T}$: MRI-to-HTC, from a source domain S (MRI image) to the target domain T (HTC image) based on independently sampled data instance, such that the distribution of the mapped samples, s', matches the probability distribution $p(t)$ of the target. For the cycle consistency, a domain inverse mapping, $F_{T\to S}$: HTC-to-MRI, also generates the reconstructed images, s'', to match closely to the input image $s \approx s''$.

Attention Block In our mapping, we need to generate HTC images that provide maximum segmentation accuracy in $\mathcal{R}_f$. To this end, we first need to locate the $\mathcal{R}_f$ to translate in each image and then apply the translation to that region. Specifically, we achieve this by adding two attention networks $\mathcal{A}_S$ and $\mathcal{A}_T$, which select areas to translate by maximizing the probability that the discriminator makes a mistake in the source and target domain, respectively. The attention block is an FCN network consists of convolution, deconvolution, and the ResNet (He et al. 2016) unit, followed by the soft-max layer. For each input image, it produces a per-pixel attention map with the same size of the input image indicating the importance of the spatial information. Mainly, after feeding the input image to the generator, we employ the attention mask to the generated image using an element-wise product ($\odot$), and then add the background using the inverse of the mask applied to the input image.

As shown in Fig. 3, s is split into two parts: the first part is fed to the source attention block, $\mathcal{A}_S$, to create the attention map, $s_a = \mathcal{A}_S(s)$, while the second part is considered as an input of the generator $G_{S\to T}$ to highlight the foreground region. To eliminate the background region, s_a is element-wisely multiplied by $G_{S\to T}(s)$ to make masked image as: $s_f = s_a \odot G_{S\to T}(s)$. Finally, the synthetic HTC image can be calculated as:

$$s' = s_a \odot G_{S\to T}(s) + (1 - s_a) \odot s \quad (3)$$

where s' is passed to the segmentation block to segments the $\mathcal{R}_f$ and fed to the domain inverse mapping for the reconstruction. Likewise, we have:

$$s'' = t_a \odot G_{T\to S}(s') + (1 - t_a) \odot s' \quad (4)$$

where $t_a = \mathcal{A}_T(s')$ is the attention map in target domain.

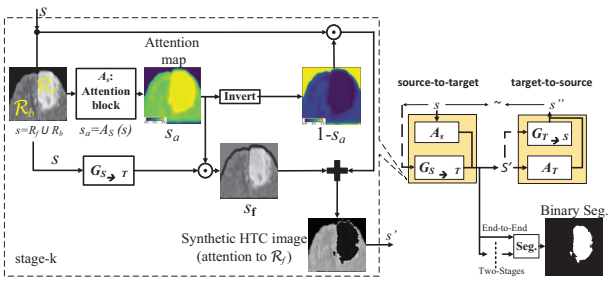

Figure 3: Image synthesis with attention to WT at stage-I. MRI and HTC images are considered as source and target.

Training Procedure The training of MRI-to-HTC network requires a discriminator D_T to discern the translated outputs from the real HTC images t. Likewise, the discriminator at the source domain, D_S, encourages HTC-to-MRI network to translate t into source domain indistinguishable from the source domain. We train both discriminators such that they only rate attended regions. Particularly, instead of employing an entire image as the input, we first filter both generated and real image via an element-wise multiplication with the attention map at source and target domain. Then the filtered images are fed into the discriminator for evaluation. According to (Mejjati et al. 2018), to avoid the mode collapse, we train the network on whole images (s and t) for 25 epochs and then switch to masked ones ($s \odot s_a$ and $t \odot t_a$), when the attention blocks $\mathcal{A}_S$ and $\mathcal{A}_T$ have trained moderately.

According to Equations 3 and 4, as long as $\mathcal{A}_S$ and $\mathcal{A}_T$ attend to the background regions, the generated images will preserve their input domain classes. Thus, the discriminators can simply detect the images as fake ones. To be thriving in two-player minimax game, $\mathcal{A}_S$ and $\mathcal{A}_T$ have to concentrate on the objects or regions that the corresponding discriminator thinks are the most descriptive within its domain (i.e., the foreground). Finally, the network finds an equilibrium between the generator, attention map, and discriminator to produce realistic images.

Preparing Target HTC Images The assumption of non-overlapping tissue distribution in the target domain can be achieved through the GT labels. We change the class-conditional distributions of $p(t|\mathcal{R}_f)$ and $p(t|\mathcal{R}_b)$ according to the manual labels as depicted in Fig. 4. We minimize the inter-class variance while maximizing the intra-class distance between $p(t|\mathcal{R}_f)$ and $p(t|\mathcal{R}_b)$ in the target space. Value of the mean and variance in the target domain are considered as a hyperparameter. Choosing an appropriate amount will produce maximum segmentation accuracy. However, there is a trade-off between class overlap at large values and visual artifacts at small ones. Particularly, low values for variance will generate much sharper results but introduces visual artifacts, which leads to a decline in segmentation performance. Fig. 4 (left column) demonstrates the considerable class overlap between the distributions of the foreground and background tissue in the source domain on the BraTS dataset. We use FLAIR, T2, and T1c sequence

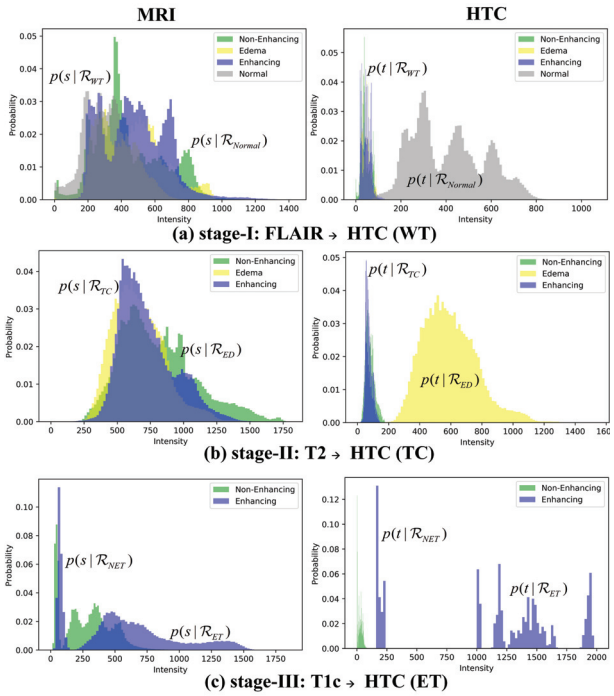

Figure 4: Tissue distributions of glioma at the source (left) and target (right) domain. (a) FLAIR MR images are used in the first stage to produce synthetic HTC images with attention to $\mathcal{R}_{WT}$. (b) At the second stage, T2 images are cropped according to the bounding box at the first stage and employed to increase tissue contrast between $\mathcal{R}_{TC}$ and Edema $\mathcal{R}_{ED}$. (c) The third stage is dedicated to enhancing $\mathcal{R}_{ET}$ and non-enhancing tumor $\mathcal{R}_{NET}$ from T1c MR images.

to segment WT, TC, and ET in each stage, respectively. Fig. 4 (right column) shows distributions of the corresponding tissues in the defined target domain.

Segmentation Block

Segmentation block provides feedback for the image synthesis one during the training in the case of end-to-end strategy. We apply the 2D binary segmentation structure with the weighted cross-entropy loss, $\mathcal{L}_{seg}$, to handle the class imbalance, especially in the first stage. Specifically, FC-DenseNet comprises the Dense blocks (batch normalization (BN), ReLU, 3×3 convolution, and Dropout), the Transition down blocks (BN, ReLU, 1×1 convolution, Dropout, and 2×2 Max Pooling), and the Transition up block (3×3 Transposed convolution with stride of 2). We also consider non-overlapping max pooling and Dropout with $p = 0.2$. Each Dense block contains four layers of convolution which each layer calculates 12 feature maps. These features are sequentially concatenated to build 48 feature maps at the output of Dense block. In the training phase, the bounding boxes are automatically generated based on the GT, whereas, in the testing phase, the bounding boxes are obtained based on the binary segmentation results of the preceding stage.

Loss Functions

In addition to the segmentation loss, $\mathcal{L}_{Seg.}$, there are four loss functions to generate HTC images in each stage. The adversarial loss to take advantage of GAN networks at the source, $\mathcal{L}^s_{adv}$, and target domain, $\mathcal{L}^t_{adv}$, as:

$$\mathcal{L}^s_{adv}(F_{S\to T}, A_S, D_T) = \mathbb{E}_{t\sim p(t)}[\log(D_T(t))] + \mathbb{E}_{s\sim p(s)}[\log(1 - D_T(s'))] \quad (5)$$

$$\mathcal{L}^t_{adv}(F_{T\to S}, A_T, D_S) = \mathbb{E}_{s\sim p(s)}[\log(D_S(s))] + \mathbb{E}_{t\sim p(t)}[\log(1 - D_S(t'))] \quad (6)$$

Meanwhile, and similarly to CycleGAN, we add a cycle-consistency loss to the adversarial ones by enforcing a one-to-one mapping between true image, s, and cycle reconstructed ones, s'', as a forward cycle consistency loss:

$$\mathcal{L}^s_{cyc}(s, s'') = \|s - s''\|_1 \quad (7)$$

where $s'' = F_{T\to S}(F_{S\to T}(s))$. In the backward path, we also have the backward cycle consistency loss as:

$$\mathcal{L}^t_{cyc}(t, t'') = \|t - t''\|_1 \quad (8)$$

where $t'' = F_{S\to T}(F_{T\to S}(t))$. Finally, we combine the defined loss functions with different weights. The final objective for the image synthesis, $\mathcal{L}_{synth.}$, is:

$$\begin{aligned}\mathcal{L}_{synth.} = \lambda_1.\mathcal{L}^s_{adv}(F_{S\to T}, A_s, D_T) + \\ \lambda_2.\mathcal{L}^s_{cyc}(F_{S\to T}, F_{T\to S}, S) + \\ \lambda_3.\mathcal{L}^t_{adv}(F_{T\to S}, A_T, D_S) + \\ \lambda_4.\mathcal{L}^t_{cyc}(F_{T\to S}, F_{S\to T}, T)\end{aligned} \quad (9)$$

where λ_1, λ_2, λ_3, and λ_4 are the scalar hyper-parameters to regularize the loss functions.

In the two-stage training strategy, we first minimize $\mathcal{L}_{synth.}$ to generate HTC images with attention to the specific region. Then, we optimize $\mathcal{L}_{Seg.}$ with fixed synthesis loss, as two independent training steps. While, in the case of end-to-end, we optimize $\mathcal{L}_{total}$ which incorporates the segmentation loss into the adversarial one during training as:

$$\mathcal{L}_{total} = \lambda_5\mathcal{L}_{Seg.} + \mathcal{L}_{Synth.} \quad (10)$$

where λ_5 balances the effect of $\mathcal{L}_{Seg.}$ to equip our HTC synthesis model with the segmentation feedback.

Experiments and Results

We conduct several experiments on BraTS 2018 dataset to demonstrate the effectiveness of the proposed methods on the synthesizing HTC images and the segmentation task. Each sequence has been normalized separately by subtracting the mean and dividing by the standard deviation of the brain region, and the non-brain area is set to zero. Furthermore, all networks are trained for 180 epochs with Adam learning rate of 0.0001. Our implementation is developed employing TensorFlow on an NVIDIA TITAN X GPU with 12G of RAM.

Table 1: K-S test results between the synthetic and target HTC images applying different loss weights.

Loss weights		Brain lesions			
λ_1, λ_3	λ_2, λ_4	WT	TC	ET	Normal
1	10	**0.31**	**0.30**	**0.34**	**0.14**
10	1	0.28	0.26	0.22	0.10
1	100	0.30	0.27	0.32	0.06
1	1	0.12	0.18	0.12	0.13

Dataset The performance of the proposed method is evaluated on publicly available BraTS 2018 dataset (Menze et al. 2015; Bakas et al. 2017; 2018), gathered from various scanners with an in-plane matrix size of $240 \times 240 \times 155$. Four MR sequences are available for each patient consist of FLAIR, T1, T1c, and T2. Evaluation is performed for three ROIs, including the WT (all internal parts of tumor), TC (enhancing and non-enhancing), and ET. Around 2K axial slices are randomly selected and center-cropped to 128×128 for training, such that each slice has non-zero value on at least half of the pixels. Specifically, the first stage, 2D FLAIR MR images are used to generate synthetic HTC one in our cyclic framework with attention to WT as: FLAIR$\leftrightarrow$FLAIR$^{'}$. Then, we segment FLAIR$^{'}$ with the end-to-end (FLAIR$\leftrightarrow$FLAIR$^{'}\rightarrow\mathcal{R}_{WT}$) as well as the two-stage (FLAIR$\leftrightarrow$FLAIR$^{'}$, FLAIR$^{'}\rightarrow\mathcal{R}_{WT}$) approach. Accordingly, for the segmentation of TC, we extract T2 patches to 96×96 from the corresponding slices. Thus, we have: T2$\leftrightarrow$T2$^{'}\rightarrow\mathcal{R}_{TC}$ and T2$\leftrightarrow$T2$^{'}$, T2$^{'}\rightarrow\mathcal{R}_{TC}$ for the end-to-end and two-stage, respectively. In the last stage, segmentation of ET, we apply T1c patches with a size of 64×64 to generate the synthetic HTC images and predict pixel labels of ET (T1c$\leftrightarrow$T1c$^{'}\rightarrow\mathcal{R}_{ET}$ and T1c$\leftrightarrow$T1c$^{'}$, T1c$^{'}\rightarrow\mathcal{R}_{ET}$).

Evaluation of the Synthetic HTC MR Images To evaluate the synthetic HTC MR images, we calculate the Kolmogorov-Smirnov (K-S) statistic on the target domain to estimate the goodness-of-fit between the intensity distribution of the synthetic HTC and real HTC images for each class label. Table 1 lists the consequences of the K-S test for the WT, TC, ET, and Normal of the brain tumor for various loss weight values. Note that the segmentation block is bypassed ($\lambda_5 = 0$) to assess the quality of synthetic HTC images. We further appraise the quality of synthetic HTC images on each stage using peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM) in Table 2. In these experiments, the loss weights are considered as $\lambda_1, \lambda_3 = 1$ and $\lambda_2, \lambda_4 = 10$. Moreover, Fig. 5 shows examples of synthetic HTC images with attention to ET at stage III. The first column presents the real T1c MR patches in the source domain, the second column displays the attention maps, and the third one shows the corresponding synthetic HTC patches, and the last column depicts the real HTC images in the target domain.

Table 2: Quality evaluation of the synthetic HTC images in our multi-stage framework.

MRI-to-HTC (attention)	SSIM	PSNR
FLAIR $\leftrightarrow$ FLAIR$^{'}(WT)$	0.6132	17.37
T2 $\leftrightarrow$ T2$^{'}(TC)$	0.6284	18.32
T1 $\leftrightarrow$ T1c$^{'}(ET)$	0.6449	19.87

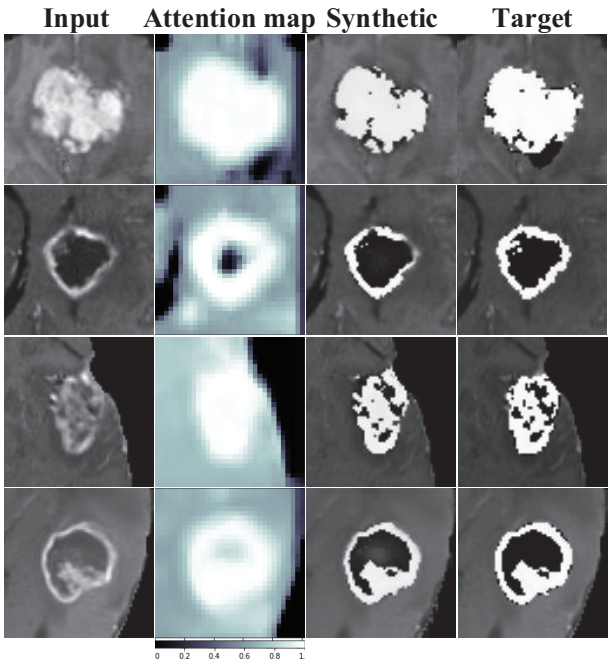

Figure 5: Examples of MRI-to-HTC translation with attention to ET on BraTS dataset. From left to right: input image, attention map, synthetic HTC image, and target image.

Table 3: Ablation study of multi-stage MRI-to-HTC model.

Model	SSIM	PSNR
CycleGAN	0.6072	17.64
CycleGAN+$\mathcal{A}_S$	0.6384	18.21
CycleGAN+$\mathcal{A}_t$	0.6387	18.29
CycleGAN+$\mathcal{A}_S$+$\mathcal{A}_t$	0.6647	18.86
CycleGAN+$\mathcal{A}_S$+$\mathcal{A}_t$+multi-stage	**0.6849**	**19.87**

Ablation Analysis We measure PSNR and SSIM as similarity metrics between the synthetic HTC and target images. In Table 3, we first employ the plain CycleGAN (Zhu et al. 2017) to generate HTC images with attention to four regions. Then, we judge the model with only one attention block in either the source (CycleGAN+$\mathcal{A}_S$) or the target domain (CycleGAN+$\mathcal{A}_T$). Finally, we repeat the experiment to consider only one region (ET) to assess our multi-stage MRI-to-HTC structure with attention blocks.

Table 4: Segmentation accuracy for WT, TC, and ET of brain lesion in MR images via cross-validation.

Dice Mean (±Std.(%))	CycleGAN + Segmentation		Proposed	
	End-to-End	Two-Stage	End-to-End	Two-Stage
WT	0.9231 (0.24)	0.8804 (0.37)	**0.9508 (0.58)**	0.9304 (0.42)
TC	0.9016 (0.11)	0.8652 (0.13)	**0.9304 (0.51)**	0.9061 (0.17)
ET	0.8699 (0.14)	0.8370 (0.18)	**0.8891 (0.13)**	0.8612 (0.11)

Table 5: Dice scores and HD95 with and without the proposed synthetic HTC images on BraTS'18 Validation datasets. FLAIR* indicates the synthetic HTC MR images by attention to non-enhancing, edema, enhancing, and normal regions.

Method	Modality concatenation	Dice			HD95 (mm)		
		EN	WT	TC	EN	WT	TC
U-Net	FLAIR, T1, T1c, T2	0.7874	0.8913	0.8402	4.15	5.61	**7.71**
	FLAIR, FLAIR↔FLAIR*,T1c,T2	**0.7921**	**0.8998**	**0.8462**	**3.94**	**5.24**	8.02
Res-U-Net	FLAIR,T1,T1c,T2	0.7891	0.8951	0.8413	4.03	5.5l	8.66
	FLAIR, FLAIR↔FLAIR*,T1c,T2	**0.7944**	**0.9033**	**0.8475**	4.03	**5.02**	**6.31**
FC-DenseNet	FLAIR,T1,T1c,T2	0.7890	0.8965	0.8439	**4.05**	5.41	7.95
	FLAIR, FLAIR↔FLAIR*,T1c,T2	**0.7943**	**0.9041**	**0.8498**	4.33	**4.95**	**7.75**

Comparisons with Other Synthetic Segmentation Methods We compare our method with recently proposed approaches (Huo, Xu, and Moon 2019) and (Chartsias et al. 2017), which employed synthetic images for segmentation in the end-to-end and two-stage manner, respectively. The former combines the segmentation loss with the adversarial one during training, while the latter individually trains the image synthesis and segmentation block. In Table 4, we measure the segmentation accuracy for WT, TC, and ET via the 4-fold cross-validation and observe that the proposed end-to-end method with the attention block achieves the highest accuracy compared to others. Table 4 also demonstrates the advantage of end-to-end training over the two-stage one in terms of accuracy. Note that we need roughly 27 ms to generate synthetic HTC image (2D) in the two-stage framework during the inference time.

Synthetic HTC Volumes in 3D Multi-Modal Segmentation Framework We evaluate the effect of synthetic HTC images in the 3D multi-modal segmentation framework based on the two-stage training approach. To this end, we substitute T1 MR volume for the corresponding FLAIR↔FLAIR* sequence, while increasing contrast among the non-enhancing, edema, enhancing, and normal regions. We experiment with three state-of-the-art segmentation models, including U-Net (Ronneberger, Fischer, and Brox 2015), Res-U-Net (Drozdzal et al. 2016), and FC-DenseNet (Jégou et al. 2017). To have a fair comparison, we perform experiments using four sequences in both cases, i.e., FLAIR, T1, T1c, and T2 for the real segmentation as well as FLAIR, FLAIR*, T1c, T2 for the synthetic one. Since T1 modality has less information regarding glioma compared to other sequences, we eliminate T1 in our experiments. Table 5 presents Dice and modified Hausdorff distance (HD95) on BraTS'18 validation set (Leaderboard), reported by the CBICA image processing online portal. Segmentation with HTC sequences improves Dice scores in three clinically important sub-regions, including WT (0.8%), TC (0.6%), and ET (0.5%). We also achieve averagely 0.4 mm improvement in WT in terms of HD95. However, we need approximately 4.2s to generate each FLAIR* volume form real FLAIR.

Discussion and Conclusion

We have shown that a deep neural network can be trained on the unpaired dataset to synthesize an HTC image from an MR image. Our proposed supervised model modify the class-conditional distributions of ROIs for the segmentation task in each stage based on the GAN model, which is equipped with attention mechanisms to alter only relevant regions in the input image. We validate our approach on the sub-regions of glioma in multi-modal MR scans of BraTS 2018 dataset. The results of the K-S test confirm that proposed MRI-to-HTC can modify the distributions of WT, TC, and ET in the FLAIR, T2, and T1c MR images, respectively. The experiments over three segmentation baselines indicate that incorporating the synthetic HTC images with other modalities, i.e., FLAIR, T1c, and T2, improves Dice score and HD95 on BraTS 2018 Leaderboard while eliminating the T1 MR sequence from the segmentation procedure. Although the proposed MRI-to-HTC can achieve promising results, it still has a limitation on defining the mean and standard deviation of the class-conditional distribution in the HTC target images. Small standard deviation values generate much sharper results but introduce visual artifacts in the synthetic images, which reduce the segmentation accuracy. As a direction for future works, one can develop a framework to tackle corrupted or missing MR vol-

ume, that appears during scanning in the acquisition setting. Towards this end, the synthetic HTC volume can be replaced with the corrupted one to complement the information presented by the missing sequence for automated systems.

References

Bakas, S.; Akbari, H.; Sotiras, A.; et al. 2017. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data* 4:170117.

Bakas, S.; Reyes, M.; Jakab, A.; et al. 2018. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*.

Bowles, C.; Gunn, R.; Hammers, A.; et al. 2018. GANsfer learning: Combining labelled and unlabelled data for GAN based data augmentation. *arXiv preprint arXiv:1811.10669*.

Chartsias, A.; Joyce, T.; Dharmakumar, R.; et al. 2017. Adversarial image synthesis for unpaired multi-modal cardiac data. In *International Workshop on Simulation and Synthesis in Medical Imaging*, 3–13.

Chen, L.; Papandreou, G.; Kokkinos, I.; et al. 2018. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(4):834–848.

Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; and Pal, C. 2016. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*. Springer. 179–187.

Elazab, A.; Abdulazeem, Y. M.; Anter, A. M.; Hu, Q.; Wang, T.; and Lei, B. 2018. Macroscopic cerebral tumor growth modeling from medical images: A review. *IEEE Access* 6:30663–30679.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; et al. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, 2672–2680.

Hamghalam, M.; Lei, B.; and Wang, T. 2019. Brain tumor synthetic segmentation in 3D multimodal MRI scans. *arXiv preprint arXiv:1909.13640*.

Hatami, T.; Hamghalam, M.; Reyhani-Galangashi, O.; and Mirzakuchaki, S. 2019. A machine learning approach to brain tumors segmentation using adaptive random forest algorithm. In *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, 076–082.

He, K.; Zhang, X.; Ren, S.; et al. 2016. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Huang, G.; Liu, Z.; v. d. Maaten, L.; et al. 2017. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2261–2269.

Huo, Y.; Xu, Z.; and Moon, H. 2019. Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE Transactions on Medical Imaging* 38(4):1016–1025.

Isola, P.; Zhu, J.; Zhou, T.; et al. 2017. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 5967–5976.

Jégou, S.; Drozdzal, M.; Vazquez, D.; et al. 2017. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1175–1183.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Mejjati, Y. A.; Richardt, C.; Tompkin, J.; et al. 2018. Unsupervised attention-guided image-to-image translation. In *Advances in Neural Information Processing Systems*, 3693–3703.

Menze, B. H.; Jakab, A.; Bauer, S.; et al. 2015. The multimodal brain tumor image segmentation benchmark. *IEEE Transactions on Medical Imaging* 34(10):1993–2024.

Nie, D.; Trullo, R.; Lian, J.; et al. 2018. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering* 65(12):2720–2730.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Sharma, A., and Hamarneh, G. 2019. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. *arXiv preprint arXiv:1904.12200*.

Soleymanifard, M., and Hamghalam, M. 2019. Segmentation of whole tumor using localized active contour and trained neural network in boundaries. In *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, 739–744.

Wolterink, J. M.; Dinkla, A. M.; Savenije, M. H.; et al. 2017. Deep MR to CT synthesis using unpaired data. In *Simulation and Synthesis in Medical Imaging*, 14–23.

Zhang, Z.; Yang, L.; and Zheng, Y. 2018. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9242–9251.

Zhao, C.; Carass, A.; Lee, J.; et al. 2017. A supervoxel based random forest synthesis framework for bidirectional mr/ct synthesis. In *International Workshop on Simulation and Synthesis in Medical Imaging*, 33–40.

Zhu, J.; Park, T.; Isola, P.; et al. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision*, 2242–2251.