Unsupervised Brain Tumor Segmentation using a Symmetric-driven Adversarial Network

Xinheng Wu[a], Lei Bi[a,*], Michael Fulham[a,b], David Dagan Feng[a,c], Luping Zhou[e], and Jinman Kim[a,*]

[a] *School of Computer Science, University of Sydney, NSW, Australia*

[b] *Department of Molecular Imaging, Royal Prince Alfred Hospital, NSW, Australia*

[c] *Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China*

[e] *School of Electrical and Information Engineering, University of Sydney, NSW, Australia*

** Corresponding authors: lei.bi@sydney.edu.au and jinman.kim@sydney.edu.au*

Abstract

The aim of this study was to computationally model, in an unsupervised manner, a manifold of symmetry variations in normal brains, such that the learned manifold can be used to segment brain tumors from magnetic resonance (MR) images that fail to exhibit symmetry. An unsupervised brain tumor segmentation method, named as symmetric driven generative adversarial network (SD-GAN), was proposed. SD-GAN model was trained to learn a non-linear mapping between the left and right brain images, and thus being able to present the variability of the (symmetry) normal brains. The trained SD-GAN was then used to reconstruct normal brains and to segment brain tumors based on higher reconstruction errors arising from their being unsymmetrical. SD-GAN was evaluated on two public benchmark datasets (Multi-modal Brain Tumor Image Segmentation (BRATS) 2012 and 2018). SD-GAN provided best performance with tumor segmentation accuracy superior to the state-of-the-art unsupervised segmentation methods and performed comparably (less than 3% lower in Dice score) to the supervised U-Net (the most widely used supervised method for medical images). This study demonstrated that symmetric features presenting variations (i.e., inherent anatomical variations) can be modelled using unannotated normal MR images and thus be used in segmenting tumors.

KEYWORDS: Generative Adversarial Network, Symmetry, Unsupervised Anomaly Detection, Brain MRI, Deep Learning.

1 INTRODUCTION

Brain tumor segmentation is fundamental for clinical decision support systems (CDSS) where the paradigm is that CDSSs can provide a second opinion to assist image interpretation. In clinical workflows the usual approach is to segment tumors manually. Manual segmentation is tedious, time-consuming and can be prone to intra- and inter-observer differences [1]. Many investigators have thus developed automated segmentation methods. Deep learning (DL) methods are the state-of-the-art for automated brain tumor segmentation. This is primarily attributed to the ability of DL to leverage large labelled datasets to derive feature representations with high-level semantics. There is a scarcity of annotated brain tumor training data, however, due to costs involved in labelling the multiple Magnetic Resonance (MR) imaging scans / sequences that are usually carried out. Further, various MRI scanning vendors / manufacturers employ different naming conventions for the data acquisitions and there are also differences in how scans are performed from site-to-site [2]. In addition, primary brain tumors differ in size, shape, location and degree of enhancement after intravenous contrast (see Figure 1). Primary brain tumors comprise approximately 2% of all malignancies in adults and 20% in children. They are usually separated into low-grade and high grade gliomas and in most cases they are unilateral. Low grade gliomas (LGGs) tend to slowly infiltrate normal brain tissue whereas high-grade gliomas (HGGs) grow rapidly, destroy normal brain, enhance with contrast, have associated vasogenic edema and may be hemorrhagic/necrotic. Hence, without training data that include all these variations, DL has difficulty in generating effective feature representations for these tumors.

We suggest that modeling variations in the normal brain, with constrained anatomical variability (i.e. bilateral symmetry), can be used to segment primary brain tumors and may remove the reliance of large annotated training data. Our hypothesis is that because the normal brain is generally symmetrical, a methodology that identifies asymmetry will be able to detect primary brain tumors. In this study, we propose an unsupervised DL method that models variations in symmetry. An unsupervised approach offers the advantage that it can exploit the abundant amount of unlabeled data generated during routine clinical imaging.

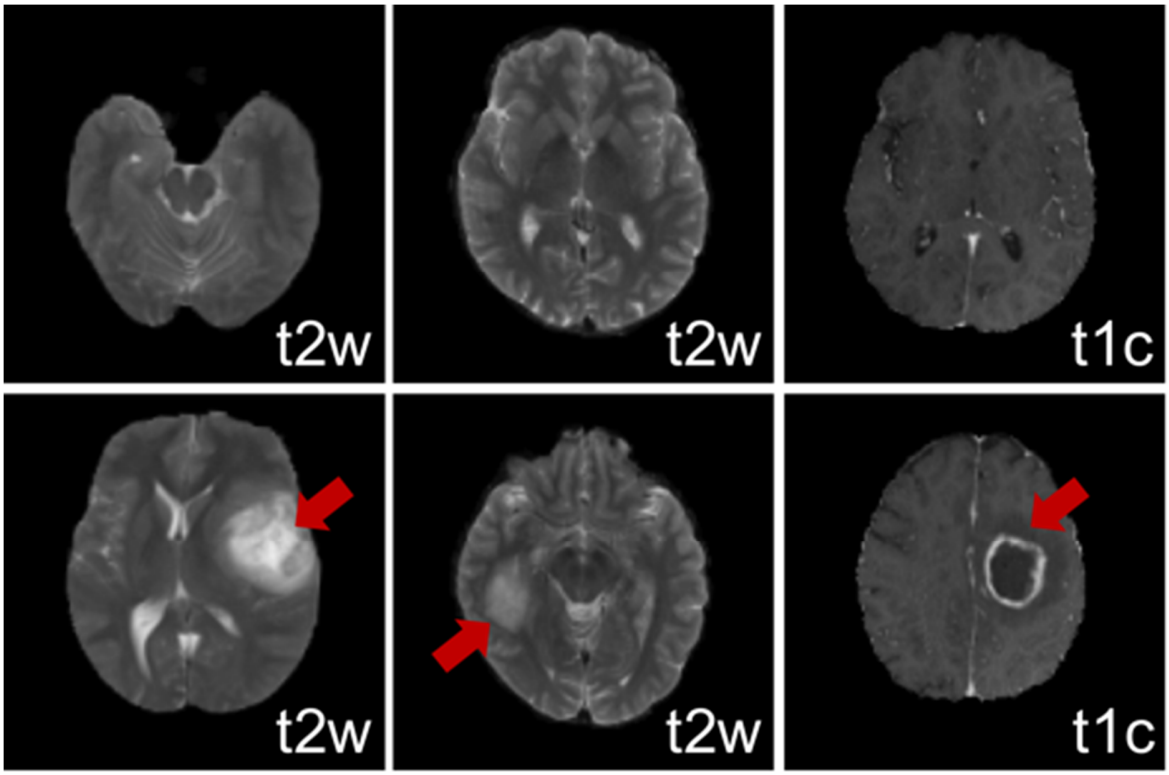

Figure 1. **Top row:** Three transaxial MR image slices at different levels – two are T2-weighted (T2W) and one is a T1-post contrast (T1c) from three brains. **Bottom row:** three transaxial images slices showing primary brain tumors (two T2W, one T1c) showing the differences in texture, size and shape.

1.1 Related Work

Our work is related to unsupervised brain tumor segmentation methods that can be separated into 2 main groups: a) Local based methods [11]–[22], where local features e.g., intensity values, are first calculated and then used for classification by an unsupervised classifier. Commonly used features and unsupervised classifiers include: thresholding [11], region growing [12], [13], K-means [14], Fuzzy C-means [15]–[18], Markov Random Field (MRF) methods [19], [20] and methods where multiple methods are combined [21]–[22]. Symmetry analysis [5][6][45][46] is a local based method that segments tumors based on their extracted symmetry features. b) Image based discriminative methods [23]–[27], where a common approach is unsupervised anomaly detection (UAD). UAD assumes that normal data have constrained variability whereas abnormal data (e.g. tumors) have diverse appearances and can be differentiated from normal. UAD constrains normal variations as a manifold and uses it to detect tumors that cannot be fitted into the 'learned' manifold. This parallels the human process of tumor detection which can be considered as searching for

anomalies compared to prior learned knowledge of what is normal. The aim of UAD is to find a lower dimensional embedding of the input data where the distance between anomalies and normal data are large [8].

The elements that directy relate to our method are anomaly detection and symmetry analysis. The UAD approach we used is based on the approach used by An and Cho [8]. The current UAD methods used in brain tumor segmentation are auto-encoder (AE) methods [23], [24], [38], [39], [40]. Baur et al. [23] investigated deep spatial auto-encoding (SAE) models on 2D whole brain MR images at an image level by comparing the input to the reconstructed image. Chen and Konukoglu [24] argued that the high variability across brain MR images, i.e., different slices of 3D brain volume, can cause a situation where the dissimilarity between two normal images could be larger than the dissimilarity between an abnormal brain and its 'normal' version. So, they enhanced the representative ability of an auto-encoder based model by imposing a consistency in the latent space to constrain the encoder to find a latent space where the projections of the input image and the reconstructed image are close to each other. Zimmerer et al. [43] used a variational auto-encoder with the Kullback-Leibler divergence to measure reconstruction errors. The AEs are able to simulate non-linear transformations from the latent space to input data, and then to detect anomalies as a deviation from the transforms by measuring the reconstruction error. The reliance on AEs means that these methods may not have sufficient feature representation to describe the brain regions. AEs also tend to reconstruct blurred images for the hand-engineering measurement of similarity between the reconstructions and the inputs, e.g., element-wise L1 or L2 distance [27]. UAD has also been applied to detect abnormalities with other medical imaging modalities using GANs [27], [9]. Schlegl et al. proposed AnoGAN [25] and f-AnoGAN [26] to model a latent space on normal 2D optical coherence tomography (OCT) samples. Unlike AEs, GANs can produce detailed images as, rather than predefining a similarity measurement, a GAN learns an objective function by itself via the adversarial training process.

With respect to symmetry analysis, human brains are relatively symmetric. The brain is constrained by the bony skull so that any mass lesion in the brain, such as a primary brain tumor, results in asymmetry because it displaces normal brain tissue. Thus, many existing algorithms capture the asymmetry induced by brain tumors. These algorithms combine symmetry analysis with traditional supervised methods such as Support Vector Machine (SVM) [49][50], AdaBoost [48], Fuzzy [51], decision forest [3][7]. Anthony et al. [3] estimated the mid-sagittal plane (MSP) by locating symmetric interest points as proposed by Yu et al. [4]. Then, they calculated symmetric texture and symmetric intensity features to measure the difference between the regions reflected across the MSP. Consequently, they incorporated these hand-crafted symmetric features to improve the performance of a traditional decision forest classifier. In unsupervised brain tumor segmentation, investigators segment tumors based on an asymmetry score, or symmetry map

calculated from their estimation of the distance between hemispheres. Manu et al. [5] implemented a registration plus a simple intensity extraction between flipped brains to calculate asymmetrical areas as candidate tumor regions for further unsupervised region growing segmentation. Zhao et al., [53] conducted a 3D registration [54] on two hemispheres before detecting asymmetric areas. Saha et al. [45] used the Bhattacharya coefficient, which was calculated based on intensity values, to measure the difference between symmetrical brains to detect tumors. Erihov et al. [46] detected tumor regions as a salient object which were distinctive locally and across its symmetrical region. Hassan et al. [6] estimated the MSP to locate symmetrical areas based on intensity features and then used a deformation model and spatial relations between the tumor and other tissues to segment the brain tumor.

In summary, all the works referred to above first estimated the MSP or registered the brains to mitigate asymmetry introduced by motion. Asymmetrical candidate regions were detected based on hand-crafted symmetry features (e.g., intensity). Then, classical supervised or unsupervised classifiers were further applied to refine the segmentation. However, symmetry property in normal human brain is a high-level semantic concept measuring the visual similarity between two hesmihperes. Therefore, the reliance on using hand-crafted features means that these methods have limited representative capacity of capturing symmetry.

1.2 Contributions

We propose a symmetry driven GAN (SD-GAN) for the unsupervised segmentation of brain tumors. We model symmetry variations as normal brain patterns and then use them to differentiate tumors. We use a conditional GAN (cGAN) [42] to model the transformation between the normal left and right hemispheres, where the normal symmetry variations are embedded in our learning model. Our method, compared to existing methods, contributes the following: a) It tolerates large normal variations e.g., alignment and movement; consequently, it eliminates reliance on annotated training data. b) It provides detailed and realistic brain volumes that resemble the input images; the reconstructed images are conditioned by a latent space of symmetry variations that is learned through the symmetric transformation training. The capacity for generating detailed and realistic brain volumes enables segment brain tumors to show asymmetry. c) We leverage a cGAN to iteratively learn the normal brain appearance in an end-to-end manner by incorporating the symmetry analysis into our voxel-wise classifier. In addition, different from the existing symmetry-based methods [5][6][7][53] which rely on additional registration steps or estimating the MSP, our SD-GAN learns the symmetry only according to the middle vertical line. Thus, our method removes the reliance on applying additional registration steps and estimating the MSP.

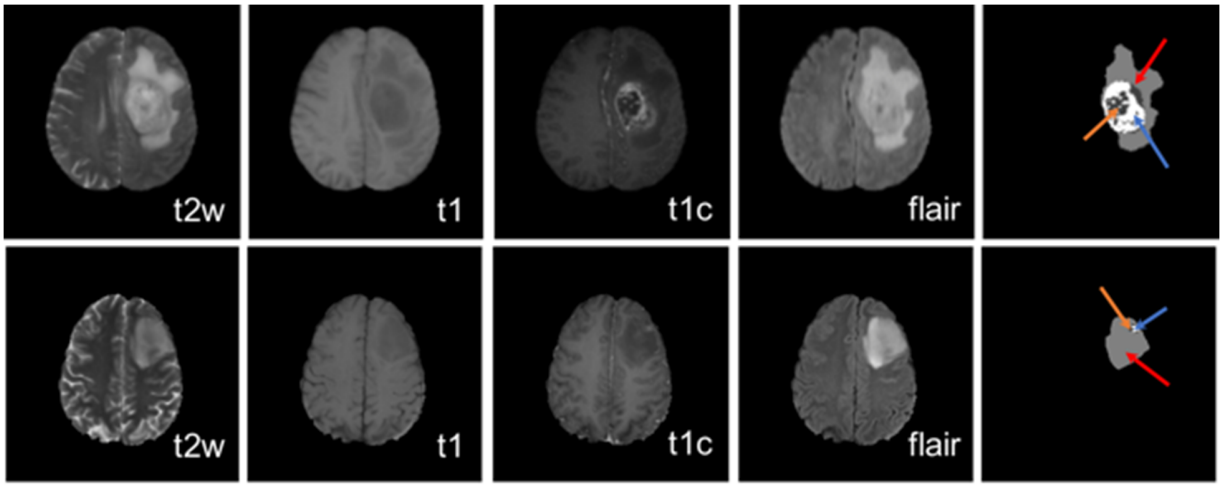

Figure 2. Two MR examples (HGG top row, LGG bottom row) from the 2018 BraTs dataset showing all the MR sequences. Two example tumors (top and bottom rows) with the annotated ground truth (right), where red arrows indicate edema, orange arrows indicate necrosis and non-enhancing core and, blue arrows indicate gadolinium-enhancing tumor.

2 METHODS

2.1 Materials

We used the public Brain Tumor Segmentation Challenge Datasets (BraTSs) from 2012 and 2018 that included LLGs and HGGs. The BraTS 2012 dataset had 20 HGGs, 10 LGGs and 50 synthesized brain images that were HGGs or LGGs [37]. The 2018 dataset [35], [36] had 285 patients with LGGs and HGGs from 19 institutions where different protocols were used to acquire the MR data. Each sample, however, had T1 (T1), T1W+contrast (T1c), T2-weighted (T2W) and FLAIR images. These two datasets were used because the brain tumors were segmented manually by neuroradiologists identifying the tumor, the contrast enhancement and any surrounding vasogenic edema. For each subject, four MR sequences (T1, T2w, T1c, flair) have been co-registered (using "VersorRigid3DTransform" [55] in ITK) to the T1c MRI. However, we note that these data were not registered (i.e., spatial normalization) to be symmetrical, as stated in [35], that no attempt was made to put the individual patients onto a common reference space. We used the T2W images which were annotated to include the entire tumor volume to evaluate our method [35]. We used two-fold cross-validation, that is, 143 were used for training and the rest 142 subjects were for testing, and vice versa. The testing dataset had images that were not included for training. In BraTS 2012 dataset, we

constructed training dataset from the synthesized images and part of the clinical images following [3], [6]. Six studies in the BraTS 2012 dataset were manually rotated to correct the movement.

2.2 Conditional Generative Adversarial Network (cGAN)

A traditional conditional GAN [28] architecture has two adversarial trained networks: a generator and a discriminator. The generator network learns a mapping from a noise vector z and an image x (in the context of an image translation task) to an output image similar to the ground truth y to 'fool' the discriminator network. The learned mapping can be formulated as $G:\{x,z\}\rightarrow y$. Such conditioning on the input image makes cGAN suitable for image-to-image translation tasks [9]. The generator network is typically a version of a fully convolutional network (FCN) [29] that has down- and up-sampling components. The down-sampling component has convolutional layers to extract high-level abstract information [30]. The up-sampling component contains deconvolutional layers [29] that decode the feature maps to the output image. Unlike the traditional FCN networks, used for image segmentation tasks that output masks, the generator network in the GAN generates an image similar to the ground truth image. To generate a realistic image, the generator network uses intensity based *L1* or *L2* loss [31], [32] rather than the SoftMax loss [33] used in traditional FCN architectures. The *L1* distance and *L2* distance are defined as:

$$\mathcal{L}_{L1}(G) = \frac{1}{n}[\|\boldsymbol{y} - G(\boldsymbol{x},\boldsymbol{z})\|_1] \tag{1}$$

$$\mathcal{L}_{L2}(G) = \frac{1}{n}[\|\boldsymbol{y} - G(\boldsymbol{x},\boldsymbol{z})\|_2] \tag{2}$$

where n is the number of voxels in an image, $\boldsymbol{y}$ is the ground truth image, G is the generator network. $\|\cdot\|_1$ is the sum of voxel-wise residuals and $\|\cdot\|_2$ is the sum of squared voxel-wise residuals of intensity value. *L1* and *L2* distance measure the similarity between the ground truth and the generated image $G(x,z)$.

The discriminator network is a classical CNN classifier that contains successive convolutional layers and a fully connected layer. The discriminator network takes the input image and the generated image or the ground truth image as input, and outputs a scaler as the posterior probability to distinguish the real and the generated images. These two networks are designed to play a 'min-max game' to obtain a powerful G to generate realistic images. The adversarial loss and the final objective functions are defined as:

$$\mathcal{L}_{GAN}(G,D) = E_{x,y}[\log(D(\boldsymbol{x},\boldsymbol{y})] + E_{x,z}[\log(1 - D(x,G(\boldsymbol{x},\boldsymbol{z}))] \quad (3)$$

$$G^* = argmin_G max_D \, \mathcal{L}_{GAN}(G,D) + \lambda \, \mathcal{L}_{L1}(G) \quad (4)$$

where D is the discriminator network that maximizes the posterior probability of allocating real examples, the "real" label and generated samples $G(\boldsymbol{x},\boldsymbol{z})$ the "fake" label. The generator network G is simultaneously trained to fool D via minimizing $[log(1 - D(x,G(\boldsymbol{x},\boldsymbol{z}))]$, which is equivalent to maximizing $D(x,G(\boldsymbol{x},\boldsymbol{z}))$. The first term in equation 4 (i.e., the GAN objective) represents the adversarial learning which learns itself an objective function that serves to force the generative model to produce a realistic image $G(\boldsymbol{x},\boldsymbol{z})$ at a high level of abstract, that is by asking whether the generated image $G(\boldsymbol{x},\boldsymbol{z})$ is real or fake.

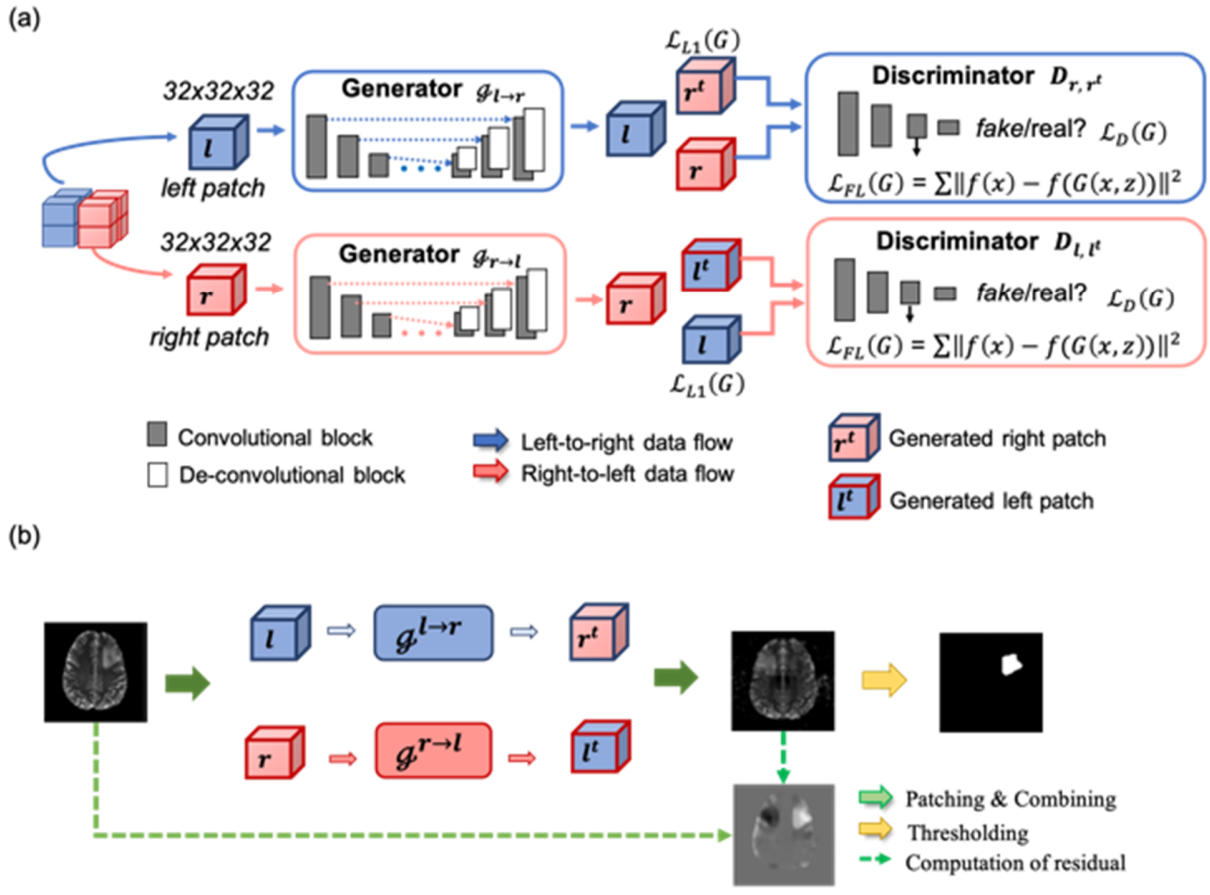

Figure 3. Overview of the SD-GAN framework. (a) shows a schematic of the data flow and the generator and discriminator networks; paired horizontal symmetrical patches train the SD-GAN separately. (b) the SD-GAN at testing phase, showing the source image (transaxial T2W image of a LGG) with generation of

the reconctructed image and the residual. The red and blue dash lines in the generator networks represent skip connections (i.e., the concatenation operations between the encoder and the decoder layers). The red and blue colors represent the right-to-left and left-to-right translation data streams.

2.3 Modeling Symmetry Variations

An overview of our proposed network is illustrated in Figure 3. We trained two cGAN [42] branches to learn non-linear mapping functions between the left and the right hemisphere images that present the variability of the (symmetry) training images only based on the set of unlabeled images. We denote the normal training dataset as $\boldsymbol{I_N} = \{x, y\}$, where N is the number of normal pairs, and $\{x_i, y_i\}$, with $i = 1, 2, \cdots, N$, denotes a pair of normal symmetrical brain patches. For each left-to-right and right-to-left branch, the generator network produces a synthesized corresponding symmetrical brain patch $\widehat{y_i}$ from the input volume x_i. The discriminator network subsequently takes a pair of images as input, i.e., $\{x_i, \widehat{y_i}\}$ and $\{x_i, y_i\}$, and aims to distinguish between the real and synthetic pairs. During this training processes, the generator G_1, G_2 and the discriminator D_1, D_2 are simultaneously optimized. Here, the generators learn to generate images of the training distribution that captures the symmetry variability. The discriminators estimate the fit of synthesized images to the distribution of training images $\boldsymbol{I_N}$. The trained generators are used with fixed weights for tumor segmentation in Section 2.4. For testing process, we have $\boldsymbol{I_M} = \{x, y, l\}$, where M is the number of unseen brain volumes during training, l_k with $i = 1, 2, \cdots, M$, is the annotation used for statistical evaluation. A description of the process is as follows.

1）Data preprocessing. Our method models the variations in symmetry in normal brains via learning a transformation mapping between symmetric brain patches in each hemipshere. We did not apply any additional pre-processing steps (e.g., spatial transformations) to register the brains so that they are prefectly symmetrical. The training dataset was generated as follows. The images were normalized to range from 0 to 1 and the brain objects were extracted from 3D volumes and then placed in the center of the images (see Figure 2). Paired patches, size of 32×32×32, were then generated from each 3D brain volume with a stride of 24. From these normalized and paired volumes, we then eliminated patches that contained < 20% non-zero voxels to stabilize the training process. After that, 2,281 paired healthy patches were used to compose the training dataset, denoted as $\boldsymbol{I_N} = \{x, y\}$ and N = 2281. No lesion voxels were selected as the input for the training process. During the training process, we did not flip the patches but instead, we let the model learn to flip the patches. The testing dataset comprised patches with normal and lesion-containing voxels from the subjects that were not included in the training set. The voxel-wise annotations of lesions from the

BraTS dataset were only used for statistical evaluation but were not fed into the network during training or test phases.

2）Generator network with skips. Our generator and discriminator architectures were motivated by the work of Isola et al. [9]. Unlike the traditional cGAN [28] where the random noise were sampled from a prior distribution, we incorporated the random noise *z* by applying dropout on every layer of our generator at training and testing stages (inspired by [9]). Modules of the form convolution-InstanceNorm-ReLu were used in the generator and discriminator. For the generator network, we followed the 5-layer 3D U-Net of Cicek et al. [34] with skip connections between encoder and decoder layers that enable the low-level information to pass across the network directly. We employed 3D convolutional kernels (i.e., 4×4×4) along with a stride of 2. The negative slope of leaky ReLu was set to 0.2. The connections and layers of the generator network are provided in Table 1.

3）Discriminator network with consistent feature representation. The discriminator network takes a pair of input images i.e., $\{x_i, \hat{y_i}\}$ and $\{x_i, y_i\}$, and aims to distinguish between the real and synthetic image pairs. By doing so both the discriminator network and the generator network are conditioning on the input original patch x_i. There is variability in normal brain MR images. The difference between two normal brain images might be larger than the distance between an image with lesion and its contralateral normal hemisphere (intra-subject asymmetry). That is to say, inter-subject variability could be larger than intra-subject asymmetry in human brain data. Chen et al. [24] minimized the inter-subject variability by adding a regularization term $\| z_h - z_{h'}\|^2$ to enforce consistency in the latent space, where z_h, $z_{h'}$ are the projections of the healthy image X_h and $X_{h'}$. Rather than modelling inter-subject variability, we trained our model to learn intra-subject variability in asymmetry. The discriminator network, in the original cGAN model, outputs a scalar to force the generator network to produce realistic images. Our approach is to generate realistic images and model the variations in normal asymmetry. Symmetry is a high-level semantic feature and it requires the discriminator network to focus on high-frequency structure. Hence, we regard the discriminator network as a feature extractor to produce the embedding of the input images and the synthesized images. Hence, we propose a feature level (FL) loss, which is defined as:

$$\mathcal{L}_{FL}(G) = \sum \| f(x) - f(G(x, \mathrm{z}))\|^2 \tag{5}$$

where *G* is generator network and $f(\cdot)$ is the third down-sampling layer of the discriminator network used as a feature extractor. The FL loss can be propagated back to *G* and encourages it to produce consistency in a lower dimensional feature space where the distance of the embedding of the input symmetrical normal brain patches are minimized. Consequently, the *G* is trained to 'fool' the discriminator, based on a scalar,

and to generate data that have similar feature representations to the input data at a high level of abstraction. We set the receptive field of this layer to 10, so that the FL loss evaluates 10×10×10 patches. The final objective loss function is thus:

$$G^* = arg \min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda_1 \mathcal{L}_{L1}(G) + \lambda_2 \mathcal{L}_{FL}(G) \quad (6)$$

where $\mathcal{L}_{\mathrm{cGAN}}$ is the traditional cGAN loss from scalar output, $\mathcal{L}_{\mathrm{L1}}(G)$ is the original L1 loss of G, and thus λ_1 serves to generating realistic images in a L1 sense; λ_2 serves to control the similarities of the images so that their projections are closer in a common feature space. The selection process of λ_1 and λ_2 is described in Section 5. The connections and layers of the discriminator network are provided in Table 1.

Table 1. Network architecture used in the generator network and discriminator network.

Generator		
Layer	**Details**	**Size**
Input	Left brain patch	32×32×32×1
Conv3d_1	4×4×4, 2, 1 LeakyReLU	16×16×16×64
Conv3d_2	4×4×4, 2, 1, InstanceNorm, LeakyReLU	8×8×8×128
Conv3d_3	4×4×4, 2, 1, InstanceNorm, LeakyReLU	4×4×4×256
Conv3d_4	4×4×4, 2, 1, InstanceNorm, LeakyReLU	2×2×2×512
Conv3d_5	4×4×4, 2, 1, ReLU, BatchNorm	1×1×1×512
ConvTranspose3d_1	4×4×4, 2, 1, BatchNorm, ReLU	2×2×2×512
ConvTranspose3d_2	4×4×4, 2, 1, InstanceNorm, ReLU	4×4×4256
ConvTranspose3d_3	4×4×4, 2, 1, InstanceNorm, ReLU	8×8×8×128
ConvTranspose3d_4	4×4×4, 2, 1, InstanceNorm, ReLU	16×16×16×64
ConvTranspose3d_5	4×4×4, 2, 1, Than()	32×32×1

Discriminator		
Layer	**Details**	**Size**
Input	Left brain patch, right brain patch, synthesized right brain patch	32×32×32×2
Conv3d_1	4×4×4, 2, 1, LeakyReLU	16×16×16×64
Conv3d_2	4×4×4, 2, 1, InstanceNorm, LeakyReLU	8×8×8×128
Conv3d_3	4×4×4, 2, 1, InstanceNorm, LeakyReLU	4×4×4×256
Conv3d_4	4×4×4, 1, 0	1×1×1

2.4 Brain Tumor Segmentation

Normal brains are generally symmetrical when compared to brains with structural abnormalties (e.g. a tumor) and so a well reconstructed normal brain patch should have much smaller reconstruction errors when compared to a brain patch containing a tumor. Therefore, during brain tumor segmentation (Figure 3) a new 3D brain MR volume was dispatched to symmetrical patches and then passed to the trained generators to output reconstructed brain patches. These reconstructed patches were included in a whole 3D brain MR volume for comparison to the original image volumes to obtain residual volumes. The residual volumes were simply obtained by the reconstruction errors $\| y - G(x, z)\|_1$. We applied a simple 3×3×3 median filter to smooth the raw residual volume and then conducted tresholding to obtain a binary segmentation. No other post-processing steps were adopted. The threshold was empirically set at the value corresponding to the 1% voxels with the highest reconstruction errors.

2.5 Experimental Design

The generator network *G* and the discriminator network *D* are trained alternately. We iteratively trained one gradient step on *D* and then trained another on *G*. All networks were trained with Adam solver and a batch size of 1 for 100 epochs. The learning rate was set to 0.0002 for the first 50 epochs and then linearly decayed to 0 for the next 50 epochs. We implemented our method in PyTorch; all experiments were done on an NVIDIA GeForce GTX 2080 Ti GPU with 11 GB memory. We evaluated our method with the commonly accepted evaluation metrics for brain tumor segmentation - including Dice score (Dice), sensitivity (Sens.) and precision (Prec.). These terms are defined as follows:

$$Dice = \frac{2*|GT \cap PM|}{|GT+PM|} \tag{7}$$

$$Sens. = \frac{|TP|}{|TP+FN|} \tag{9}$$

$$Prec. = \frac{|TP|}{|TP+FP|} \tag{10}$$

The ground truth annotation of brain tumors provided in the BraTS datasets is *GT*; *PM* is the predicted mask of the tumor by our method and the comparison methods; *TPs* are the true positive tumor voxels; *FPs* the false positive tumor voxels; *FNs* are the false negative background voxels and *TNs* are the true negative background voxels.

For the BraTS 2018 dataset, we compared our method to the state-of-the-art UAD methods for brain tumor segmentation that included: (a) AAE [44] – auto encoder using adversarial training to find latent space; (b) VAE [44] – variational auto encoder; (c) AE-GAN [23] – autoencoder using GAN to reconstruct the image; (d) VAEGAN [23] – variational autoencoder using GAN to reconstruct normal brain images ; (e) f-AnoGAN [26] – WGAN [41] that uses an auxiliary encoder where the encoder encodes a latent space from the input images for further reconstruction via GAN. We reimplemented above UAD methods on BraTS datasets. The training dataset only consists of healthy brain slices without tumor annotations and the testing dataset consists of brain slices which were excluded from the training set. It is worth to note that unlike existing UAD methods which were globally trained on normal brain images, our SD-GAN was trained between the symmetrical brain hemispheres to learn the constrained intra-subject asymmetry variations.

We compared our method to a very simple heuristic method where we flipped the MR images and conducted the subtraction to examining the differences between the two hemispheres. We also compared our method to supervised symmetry analysis methods and DL methods: (a) Sym-ST – a supervised spatial tree method incorporating a context rich symmetry feature using the BraTS 2012 dataset [6], (b) Sym-T – a supervised tree method with symmetric intensity and texture features using the BraTS 2012 dataset [3], and (c) the supervised 3D U-Net [34], which is an established network optimized for volumetric medical images, using the BraTS 2018 dataset.

We evaluated the performance of our method at individual stages, where we investigated the utilization of the proposed FL losson BraTS 2012 and 2018 datasets. Further, we evaluated our methods on studies with high normal asymmetry variability. Among BraTS 2018 dataset, we identified the top 10% samples (28 patients) that had highest level of asymmetry score on normal slices based on the approach used by Zhao et al. [53]. We measured the local symmetry on normal brain slices by $Asym(p) = min_{q \in N(p)} \| I(p) - I_{mgr}(q) \|$, p is pixel(s) of original image I, $N(p)$ is the neighborhood of p, and I_{mgr} is the flipped image.

To analyse the effectiveness of the 3D cGAN model, we made comparison to 2D variant of our model using 2D convolutional layers. We noted that the 2D model used the normal brain slices as inputs while 3D model used image patches instead. For the 2D model, the training data consisted of 2D normal brain slices (7509 slices) without lesions. To study the proposed architecture of two networks, we conducted an additional experiment using a “lighter” architecture which only has a single network. Specifically, we trained a unique network to learn to reconstruct patches either from left-to-right or right-to-left at a probability of 0.5.

3 RESULTS

In Table 2 we present the evaluation metrics of our method when compared to the state-of-the-art UAD methods and the 3D U-Net. Our approach had the best in Dice at 61.9%, and Sensitivity at 61.3%. Our method was competitive when compared to the supervised 3D U-Net with a Dice of 61.9% versus 64.9%). Our result of AAE using BraTS dataset was 1.5% lower in Dice Score than those reported in the referenced paper [44] (our AAE: 39.5%, AAE [44]: 41%). Our result of VAE using BraTS dataset is 0.9% higher in Dice Score than those reported in the seminal paper [44] (our VAE: 42.9%, VAE [44]: 42%). Our reimplemented AE-GAN [23] and VAEGAN [23] got notably lower results in Dice Score on the BraTS dataset (our AE-GAN: 44.9%, AE-GAN [23]: 52.6%; our VAEGAN: 49.0%, VAEGAN [23]: 60%). The reproduced f-AnoGAN achieved a Dice Score of 52.4%. We also reimplemented AnoGAN but only generated unrealistic brain MR images (in Figure 5 Left) which is consistent with the paper [43] where they obtained a notably lower result of 37.4% in Dice Score. In Table 3 we show the comparison results of our method with supervised symmetry driven methods (i.e., Sym-T, Sym-ST). Our method was competitive to supervised symmetry methods with a Dice of 64.6% versus 68% and 69%.

In Table 4 we present the component analysis on using FL loss where it improved Dice (58.1% vs 61.9%, 61.3% vs 64.6%) on BraTS datasets. Further, when we assessed the FL loss on the identified top 10% asymmetry studies, our method achieved higher sensitivity of 62.8%, and a higher Dice of 59.2%, compared with not using the FL loss at 57.2% and 57.0%.

We evaluated the differences between the two brain hemispheres. The results show that our learned model outperformed the simple subtraction method (61.9% vs 35.5% for Dice, 61.3% vs 44.3% for Sensitivity). We further investigated the utilization of the proposed 3D cGAN model and the results show that our 3D model achieved higher performance when compared to 2D variant of our method (61.9% vs 40.5% for Dice, 61.3% vs 47.4% for Sensitivity). When we study the architecture of using one single network, such "lighter" architecture resulted in a 10.8% drop in Dice Score (51.1% vs 61.9%) when compared to our results of training two networks.

In Figure 4 we show the segmentation results from four brain tumor examples with our reconstructed images, the intermediary segmentation results, the residual image and, the GT. There is strong correspondence between our results and the GT; in row 1 and row 4, our tumor segments are visually slightly smaller than the GT at the edge of the tumor regions; in row 2 our result has an inconsistent surface in tumor edema where shows relative smaller differences to normal regions. We notice that the residual volumes (i.e., differences between the input volumes and the generated volumes) would indicate the tumor regions with higher positive errors (and higher negative errors in the contralateral regions). This is because,

SD-GAN was trained on normal paired patches which are symmetrical. During the testing process, the trained "normal" model assumes a symmetrical pair of inputs and generates a symmetrical contralateral patch for an input patch (regardless of the presence of tumor regions). As a result, the reconstruction errors would only present smaller absolute values (i.e., close to zero) when the input pairs do not contain tumors (i.e., showing symmetry).

In Figure 5 (left) we show an example of a normal brain and the reconstructed synthesized images from our method and comparison methods. Our produced images are pixelated but have the strongest correspondence to the source image whereas other methods produced blurry (AAE and f-AnoGAN) or unrealistic (f-AnoGAN-part) images. In Figure 5 (right) we outline the comparison results between our method using FL loss and without. In row 1, the absence of the FL loss resulted in false detections of tumor regions (indicated by the red arrow); in row 2 the removal of the FL loss caused an undersegmented results with missing tumor regions (indicated by the blue arrow). In Figure 6 (a) the distribution of the reconstruction errors of normal patches compared to patches that contain tumors is shown. The distribution of the reconstruction errors over the normal patches (blue lines) had smaller means and variances compared to abnormal patches (red lines). This is consistent with the fact that brains have restricted asymmetry variations and constrained anatomy compared with brain patches containing tumors which show asymmetry and vary in its anatomical appearance.

In Figure 6 (b) the probability distribution function (PDF) and the cumulative distribution function (CDF) of the voxel-wise reconstruction errors on normal brains are shown. The errors were normally distributed and achieved over 98% of the errors after 0.10. The segmentation results with varying thresholding values from 0.10 to 0.15 are stable, as shown in Figure 6 (c). Additional segmentation results on asymmetry cases are included in the supplementary materials.

In Figure 7, we present two failed cases from using our method.

Table 2. Metrics of our SD-GAN compared to UAD methods and supervised 3D u-Net on BraTS 2018 dataset.

Models	**Metrics**		
	Dice	**Sens.**	**Prec.**
AAE [44]	0.395	0.492	0.305
VAE [44]	0.429	0.491	0.347
AE-GAN [23]	0.449	0.493	0.401
VAEGAN [23]	0.490	0.517	0.503
f-AnoGAN [26]	0.524	0.548	0.560
Ours	0.619	0.613	0.735
3D U-Net (supervised)	0.649	0.771	0.791

Table 3. Comparison with symmetry driven methods on BraTS 2012 dataset.

Methods	**Metrics**		
	Dice	**Sens.**	**Prec.**
Simple-Subtraction	0.355	0.443	0.298
Sym-ST [6] (supervised)	0.68	-	-
Sym-T [3] (supervised)	0.69	-	-
Ours	0.646	0.802	0.701

Table 4. Component analysis on FL.

Methods		**Metrics**		
		Dice	**Sens.**	**Prec.**
BraTS 2018	Without FL	0.581	0.657	0.719
	Ours	0.619	0.613	0.735
BraTS 2012	Without FL	0.613	0.713	0.683
	Ours	0.646	0.802	0.701

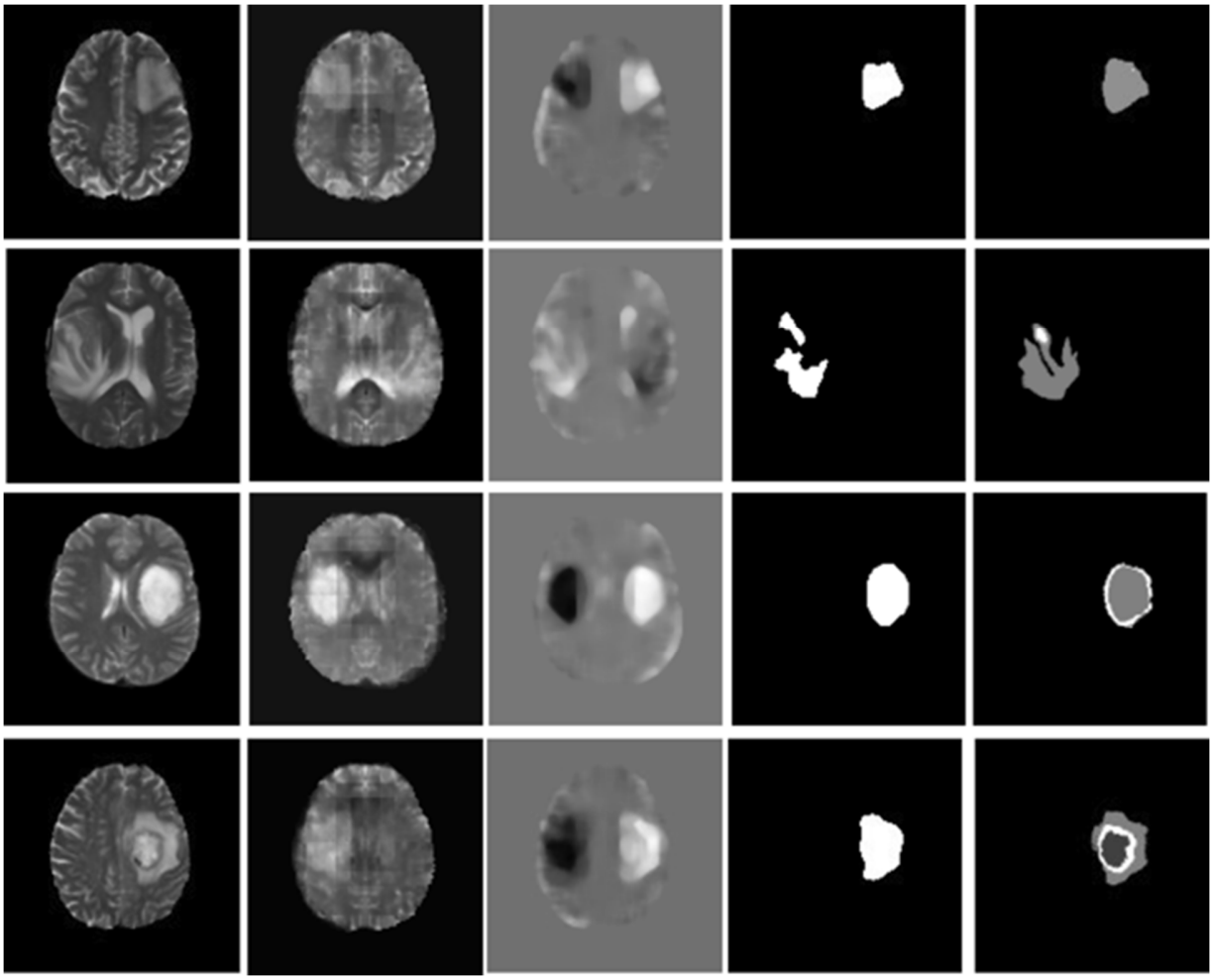

Figure 4. Four brain tumor examples in T2W images with LGG in the top two rows and HGG in the bottom two rows. The columns show the segmentation results and the intermediary stages. Column 2 shows the reconstructed brain volumes with our method; column 3 shows the residual images obtained from the reconstruction errors; column 4 shows the segmentation results with our method and, column 5 shows the ground truth from the BraTS datasets.

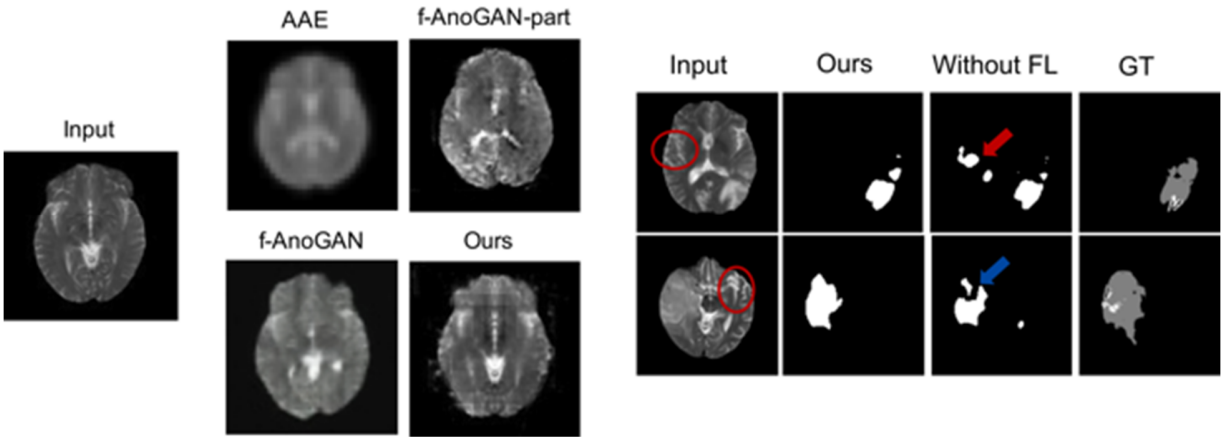

Figure 5. Left: a visual comparison of reconstructed images from a normal brain example between our method and other UAD methods. Right: comparison of segmentation results from our method with and without the FL loss . The red arrow indicates the falsely positives while the blue arrow indicates the false negatives.

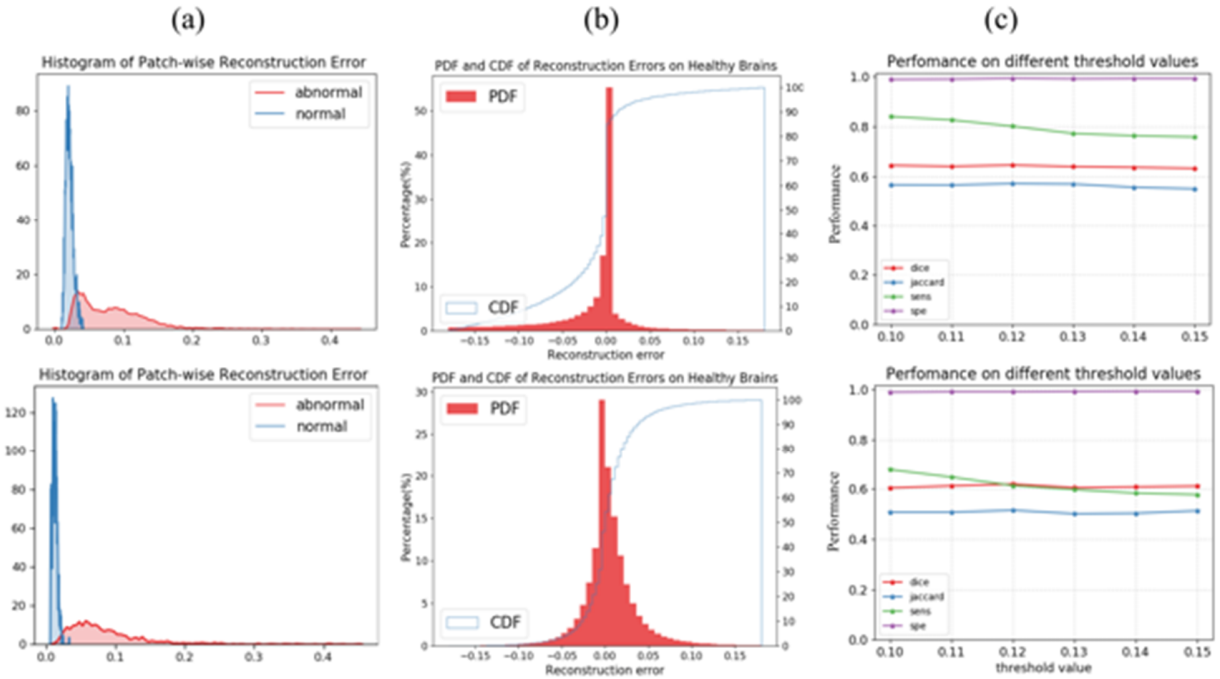

Figure 6. (a) Patch-wise reconstruction error distribution in normal and abnormal brains; (b) Voxel wise reconstruction error on normal brains, (c) Ablation study on different thresholding values on BraTS 2018 dataset (top) and 2012 dataset (bottom).

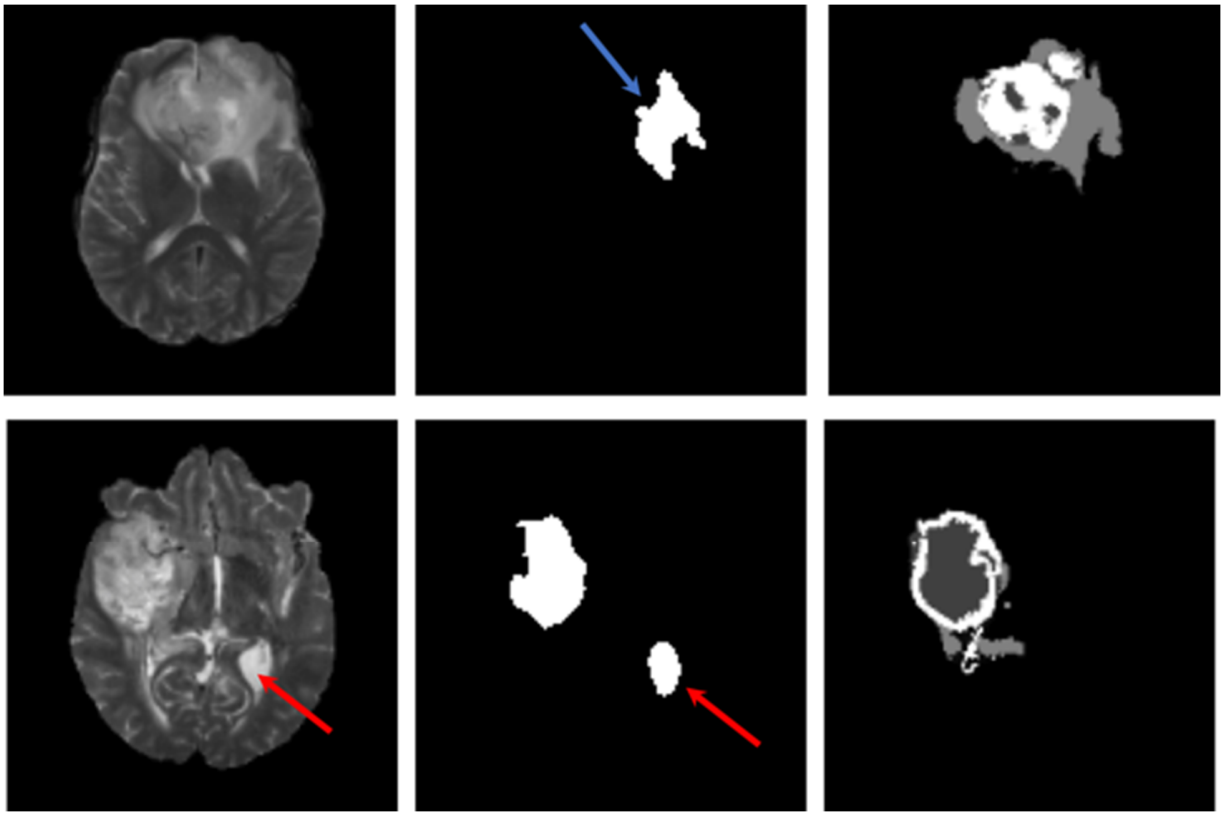

Figure 7. Examples of two failed cases. Row 1: a failed case where the tumor was located adjacent to the midline. Row 2: an example of HGG when the tumor involves and deforms the left lateral ventrical. The red arrow indicates that normal asymmetrical lateral ventrical was falsely detected as tumor regions.

4 DISCUSSION

Our main findings are that our SD-GAN: a) outperformed comparative deep UAD brain tumor segmentation methods; b) perform competitively to supervised DL method and symmetry analysis methods; c) tolerated most normal aymmetry variations thus improving segmentation.

4.1 Comparison with brain tumor segmentation methods

Our method achieved the best result compared with other UAD methods on brain tumor segmentation. When compared to GAN based methods, we found that AE based methods (e.g., AAE, VAE) had limited generative capacity and tended to produce blurry images. As a consequence, variations in normal brains were more likely to be detected as anomalies, resulting in lower performances on Dice scores (AAE: 39.5%, VAE: 42.9%). However, AE based methods with GAN for training, e.g., the re-implemented AE-GAN,

and VAEGAN improved performances, 44.9% and 49.0%, respectively, on Dice scores. This is as expected since GAN training adds improved reconstruction capacity compared to simple AEs. We note that, when compared to the reference paper [23], our reimplemented AE-GAN and VAEGAN had notably lower results (our AE-GAN: 44.9%, AE-GAN [23]: 52.6%; our VAEGAN: 49.0%, VAEGAN [23]: 60%) in Dice Score. The lower performance on BraTS dataset is likely attributed to the fact that their methods were optimized to the private dataset e.g., hyper-parameters and heavily post-processing steps. Consequently, these methods have limited generalizability to the BraTS dataset, which has a different data distribution (i.e., image characteristics) and data acquisition protocols. Apart from AE based methods, we investigated f-AnoGAN [26]. f-AnoGAN improved over AE based methods and performed competitively with AE-GAN [23]. We also investigated f-AnoGAN-part by taking out of the auxiliary encoder used in f-AnoGAN. The experimental results showed that the generated images were less realistic. This indicated that the auxiliary encoder used in f-AnoGAN was trained to simulate a conditional probability distribution on the input brains and hence improve the reconstruction. Without using the auxiliary encoder, our method further improved upon the f-AnoGAN (Dice Scores:61.9% vs 52.4%, Sens.: 61.3% vs 54.8%) by imposing symmetry information via one-stage conditional GAN [42]. This is because, due to the large normal variability in human brains, existing UAD methods (e.g., f-AnoGAN) cannot identify tumors when they have mildest differences to normal vairiations thus resulting a lower sensitivity of 54.8%. Our method also outperformed another unsupervised AE method using KL proposed by Zimmerer et al. [43] where they trained ther model on additional 1000 healthy brains and obtained a 44% of Dice Score on the BraTS 2017 dataset. We attribute the superiority of our symmetric driven GAN model to segment abnormalities (i.e., tumors) by learning a restricted intra-subject asymmetry variability between brain hemispheres.

We note that another unsupervised asymmetry-based methods proposed by Erihov et al. [46] achieved 75% Dice score on 20 patients from BraTS 2014. This method, however, was designed by measuring the distinctness and smoothness of brain patches on four modalities. Although we designed our algorithm to work with a single modality, i.e., T2w, we anticipate that other modalities can contribute complementary information to further improve the segmentation results. We discuss the potential in using additional modalities as part of our limitation and future work in Section 4.4. In addition, Erihov used L1 and L2 distance measures which have difficuilties in the detection of tumors with lower contrast to the background, i.e., segmentation results of < 20% Dice score for patients with larger edema regions. In contrast, our method is not limited by the contrast differences and is consistent in segmentation performance to tumors with various appearances.

When compared to supervised DL method and supervised symmetry analysis methods, the results showed that our method obtained competitive performance (within 3% and 4% lower in Dice score). The higher

performance from the supervised methods are expected since the manual expert annotations can guide the model to learn the discriminative features between tumors and normal regions in a supervised way. In our method, the models were learnt in an unsupervised end-to-end manner and outperformed other unsupervised deep learning methods.

As an unsupervised DL method, our method can be trained to better model the normal anatomical variabilities with large datasets and hence discriminating tumors. To exemplify our method's ability to improve from additional unannotated data, we added 100 human connectome project (HCP) [52] normal brain data into the BraTS2018 training set. The additional data resulted in 1.2% improvement in Dice (63.1%) by improving the modelling of the larger normal variations between the brain images.

4.2 Tumor segmentation based on modelling normal variations

Our method showed its ability to reconstruct normal brains with the cGAN models, and the regions with higher reconstruction errors were detected as brain tumors because they were not symmetrical. The small overlapped regions between the two distributions in Figure 6 (a) indicates that our method was able to separate tumors from normal variations. Our method reconstructed normal brain images with realistic appearance and low reconstruction errors. In the normal brain cases, the generated images resembled the input images based on the symmetry of the normal brains (See Figure 4 Left). We attribute this to the reconstructed images which were conditionally obtained from a learned latent space which represents the symmetry of the human brains. In the cases of brain patches containing tumors, when compared to the input images, the normal regions resembled the input images because of the symmetry, whereas the tumor regions were reconstructed ('flipped') in the contralateral hemisphere (See Figure 5 first two columns). This is because the cGAN models were trained on normal brain hemispheres, and normal brain hemispheres are symmetrical. Therefore, tumor regions have higher reconstruction errors in their residual images when compare to normal tissues.

The results obtained by the simple subtraction method (Dice: 35.5%, Sens. 44.3%) are notably lower than our learning model. We attribute the large differences to the fact that the human brains are not perfectly symmetrical and there are large variations e.g., symmetry variations, exist between the two hemispheres. Consequently, it's difficult to use simple subtraction to separate tumor regions from the normal tissues. In contrast, our method leveraged UAD to model the symmetry variations as a high level of abstraction. Therefore, the learned residual maps (i.e., differences between the reconstructed MR image and the input MR image) were constrained to highlighting the tumor regions while tolerating the reconstruction errors in normal regions showing asymmetry.

The results from 2D experiments (Dice:40.5%, Sens.47.4%) are notably lower than our 3D method (61.9%, 61.3%). This shows that the 2D model failed to learn sufficient normal symmetrical variations, as it can falsely detect asymmetrical normal structure as tumors. Moreover, the 2D model can get incoherent masks between slices, which increased false detections. This is because brain and brain tumors are in 3D by nature and our 3D method segments the tumors by modelling normal symmetrical variations that are related to the 3D structures in MRIs. We attribute our better performance to the proposed 3D GAN model which can capture the spatially structural 'symmetry' via minimizing the feature disparity between paired 3D patches.

4.3 Ablation study

We evaluated the segmentation results with varying thresholding values from 0.10 to 0.15 with the results depicted in Figure 6 (c). The evaluation performance measures were stable with only the specificity being slightly decreased along with the ascending thresholding values. This was expected, where a lowered thresholding value means that the model would always classify more voxels as the tumor. However, the stable Dice score indicates that the higher specificity was aligned with higher false positives. Thus, for the segmentation task, Dice score was found to be a more indicative metric. In this situation, our method was more consistent with various thresholding values.

We studied our architecture at individual stages. We found that the FL loss ensured that our GAN model was able to produce consistent segmentation results on asymmetrical cases. In the 28 highly asymmetrical samples, our method achieved higher Sens. (62.8%) and Dice score (59.2%) using FL loss compared with not using FL (57.2% and 57.0%). Figure 5 (right) indicates that the use of FL loss can improve the modelling of normal asymmetry variations (indicated by red circles), and thus reducing false detections. We attribute our superiority of using FL loss to measure symmetry in a high level of abstract, thus intensity differences caused by normal asymmetry variations in human brains will not hurt the performance. We included the example segmentation results on asymmetry cases in Figure 1 in the supplementary materials.

Further, the results obtained by using an unique network caused a 10.8% drop in Dice Score, when compared to our architecture of training two networks. We attribute the better performance to the fact that training two networks allows to separately model the left-to-right and right-to-left translations, which enables to learn subtle variations across the left and right hemispheres within the dataset. In addition, the training of two networks can be considered as a form of ensemble mechanism, which allows to produce more stable segmentation results when compared with a single network architecture. Furthermore, we note that even with such a lighter network, we outperformed the other UAD methods in Dice Score (single

network: 51.1%, VAEGAN: 49.0%, AE-GAN: 44.9%). These results further indicate the superiority of learning the constrained symmetry variations for unsupervised brain tumor segmentation.

4.4 Limitations and Future Work

Figure 7 shows examples of two failed results. In row 1, the tumor was located in the midline, involving the corpus callosum with virtually identical involvement of both medical frontal lobes. The blue arrow indicates a detected asymmetrical tumor region but missing the symmetrical regions. This failure was expected because the undetected tumor region is symmetrical in texture, and therefore the tumor was missed by our method. The inability to work on tumors located in the midline is a known limitation among all symmetry driven methods [3][6]. To overcome this limitation, as part of our future work, we will explore post-processing with e.g., unsupervised region growing algorithm [12], where we will use the falsely detected region as an initialization to refine the segmentation. Moreover, the red arrow in Figure 7 row 2 marks a false detected tumor region which is an unsymmetrical region between the bodies of lateral ventricle. This asymmetry is caused by a tumor involving and deforming the left lateral ventricle in another hemisphere. In such a case, we suggest using a different MR sequence, e.g., T1c which could be used to identify the tumor and separate it from the normal ventricular system.

5 Model Variants and Fine-tunining

5.1 Model variants

We conducted additional experiments on three smaller model variants. The detailed model configurations and the segmentation results are provided in Table 5. The “small” and “medium” models are variants in which fewer channels were adopted in the generator networks (with using the same discriminator network of a size of 2.9M). The other medium network (“medium-2”) is a variant when we only use a 4-layer U-Net in the generator network. The medium sized network obtained a slightly decrease in dice score (61.3% and 60.7% vs 61.9%) whereas the small network had a 5.1% reduction (56.8% vs 61.9%) when compared to the current network. To evaluate the risk of overfitting, we also evaluate the overall mean and variance on a 5-fold and 10-fold cross validation setting. As expected, segmentation performance across different models was consistently improved with a larger number of folds and the variance was decreased. For example, our large network obtained a 1.1% increase in dice score on a 10-fold setting over a 2-fold setting and the variance had been largely reduced. These results suggest that our method can further benefit from a larger training dataset. Although a large network tends to overfit to a relatively small dataset, through

these experiments, we identified that a relatively large model is still essential in obtaining the optimal brain tumor segmentation performance over the risk of overfitting.

Table 5. Performance (Dice.) of model variants using k-fold cross validation.

Model Variants	Channel	Params.	Layers	2-fold		5-fold		10-fold	
				Mean	Var.	Mean	Var.	Mean	Var.
Ours-Small	16	4M	5	0.568	6.5E-04	0.570	5.57E-05	0.571	5.28E-05
Ours-medium	32	19M	5	0.613	4.3E-04	0.618	5.70E-05	0.619	**1.53E-05**
Ours-medium-2	64	16M	4	0.607	4.0E-04	0.610	5.78E-05	0.612	1.60E-05
Ours-Large	64	66M	5	**0.619**	4.2E-04	**0.628**	**5.65E-05**	**0.630**	1.55E-05

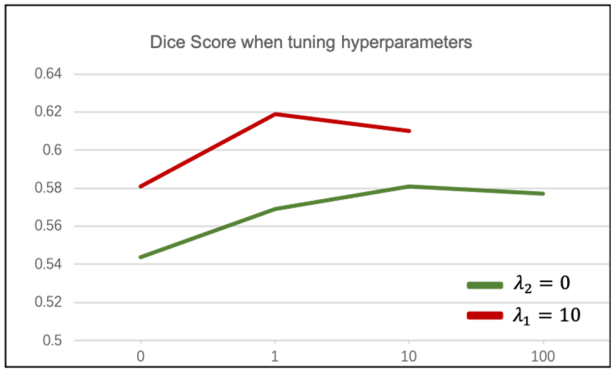

Figure 8. Performance on BraTS 2018 trained with different hyperparameters. The green line presents the results of tuning λ_1 on a range of {0, 1, 10, 100}. The red line is the results of tuning λ_2 over {0, 1, 10}.

5.2 Fine-tuning

To select the optimal hyperparameters, we firstly tuned λ_1 on a range of {0, 1, 10, 100} while keeping λ_2 equal to zero. The optimal result was obtained when λ_1 = 10. We then started tuning λ_2 on a same basis of tenfold, (e.g., 0, 1, 10, …) while keeping λ_1 equal to 10. The optimal performance was found when λ_2 = 1. The experimental results are shown in Figure 8. We notice that the model only using the cGAN loss (λ_1=0 and λ_2=0) had the worst performance. This is as expected as using cGAN loss alone is known to produce artifacts in the image reconstruction tasks such as in image generation and image translation [9] [57]. The optimal performance was achieved when both L_1 loss and feature level loss were added (λ_1=10, λ_2 = 1). This is because using L_1 and cGAN loss (λ_1=10), which help the model to produce realistic brain images. Moreover, an additional constrain on the feature level disparity (when set λ_2 = 1) further improved the model's capacity of learning symmetry of human brains.

6 CONCLUSIONS

We present our findings using an unsupervised symmetric-driven adversarial network method for segmentation of brain tumors. The novelty was in the modelling of a manifold of symmetry variations on normal brains that can be used to detect tumors based on their fitness to the modelled manifold. Our results on public brain tumor datasets show that our method achieved the best unsupervised segmentation performance and performed competitively to the supervised methods.

REFERENCES

[1] V. Martin, D. Müller, N. Verburg, R. Eijgelaar, M. Witte, F. Barkhof, P. de Witt Hamer, and J. de Munck *et al*., "Inter-observer variation in segmenting glioma on MRI before and after resection," In EMBEC & NBC, Springer, Singapore, 2017, pp. 161-164.

[2] G. Nelly, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," Magnetic resonance imaging, vol. 31, no. 8, pp. 1426-1438, 2013.

[3] B. Anthony, J. V. Miller, E. T. Tan, and A. Montillo, "Brain tumor segmentation with symmetric texture and symmetric intensity-based decision forests," In IEEE International Symposium on Biomedical Imaging, 2013, pp. 748-751.

[4] S. Yu, and B. Bhanu, "Reflection symmetry-integrated image segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 9, pp. 1827-1841, 2011.

[5] G. Manu, K. S. Gayatri, K. Harika, B. P. Rao, V. Rajagopalan, A. Das and C. Kesavadas, "Brain tumor segmentation by integrating symmetric property with region growing approach," In Annual IEEE India Conference, 2015, pp. 1-5.

[6] K. Hassan, O. Colliot, and I. Bloch, "Automatic brain tumor segmentation using symmetry analysis and deformable models," In Advances in Pattern Recognition, pp. 198-202. 2007.

[7] G. Ezequiel, B. H. Menze, and N. Ayache, "Spatial decision forests for glioma segmentation in multi-channel MR images," MICCAI Challenge on Multimodal Brain Tumor Segmentation, 34, 2012.

[8] J. An, and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," Special Lecture on IE, vol. 2, no. 1, 2015.

[9] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125-1134.

[10] M. Angulakshmi, and G. G. Lakshmi Priya, "Automated brain tumour segmentation techniques—A review," International Journal of Imaging Systems and Technology, vol. 27, no. 1, pp. 66-77, 2017.

[11] N. M. Saad, S. A. R. Abu-Bakar, S. Muda, and M. Mokji, "Segmentation of brain lesions in diffusion-weighted MRI using thresholding technique," In IEEE International Conference on Signal and Image Processing Applications, 2011, pp. 249-254.

[12] G. C. Lin, W. J. Wang, C. C. Kang, and C. M. Wang, "Multispectral MR images segmentation based on fuzzy knowledge and modified seeded region growing," Magnetic resonance imaging, vol. 30, no. 2, pp. 230-246, 2012.

[13] K. S. A. Viji, and J. JayaKumari, "Modified texture-based region growing segmentation of MR brain images," In IEEE Conference on Information & Communication Technologies, 2013, pp. 691-695.

[14] K. M. Nimeesha, and R. M. Gowda, "Brain tumour segmentation using K-means and fuzzy c-means clustering algorithm," Int J Comput Sci Inf Technol Res Excell, vol. 3, pp. 60-65, 2013.

[15] Z. Ji, Y. Xia, Q. Sun, Q. Chen, and D. Feng, "Adaptive scale fuzzy local Gaussian mixture model for brain MR image segmentation," Neurocomputing, vol. 134, pp. 60-69, 2014.

[16] Q. Ain, M. A. Jaffar, and T. S. Choi, "Fuzzy anisotropic diffusion-based segmentation and texture-based ensemble classification of brain tumor," Applied Soft Computing, vol. 21, pp. 330-340, 2014.

[17] H. Verma, R. K. Agrawal, and A. Sharan, "An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation," Applied Soft Computing, vol. 46, pp. 543-557, 2016.

[18] C. Militello, L. Rundo, S. Vitabile, G. Russo, P. Pisciotta, F. Marletta, M. Ippolito, C. D'arrigo, M. Midiri, M.C. Gilardi, "Gamma Knife treatment planning: MR brain tumor segmentation and volume measurement based on unsupervised Fuzzy C‐Means clustering," International Journal of Imaging Systems and Technology. 2015 Sep; 25(3):213-25.

[19] N. Subbanna, D. Precup, and T. Arbel, "Iterative multilevel MRF leveraging context and voxel information for brain tumour segmentation in MRI," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 400-405.

[20] N. Subbanna, and T. Arbel, "Probabilistic gabor and markov random fields segmentation of brain tumours in mri volumes," Proc MICCAI Brain Tumor Segmentation Challenge, 2012, pp. 28-31.

[21] A. Demirhan, and İ. Güler, "Combining stationary wavelet transform and self-organizing maps for brain MR image segmentation," Engineering Applications of Artificial Intelligence, vol. 24, no. 2, pp. 358-367, 2011.

[22] A. Moran, O. Aizenstein, T. Jonas‐Kimchi, N. Bornstein, L. Shopin, H. Hallevi, D. B. Bashat, and STIR and VISTA Imaging collaborators. "Classification of lesion area in stroke patients during the

subacute phase: A multiparametric MRI study." Magnetic resonance in medicine 72, no. 5 (2014): 1381-1388.

[23] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, "Deep autoencoding models for unsupervised anomaly segmentation in brain mr images," In International MICCAI Brainlesion Workshop, 2018, pp. 161-169.

[24] X. Chen, and E. Konukoglu, "Unsupervised Detection of Lesions in Brain MRI using constrained adversarial auto-encoders," In MIDL Conference book, 2018.

[25] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," In International Conference on Information Processing in Medical Imaging, 2017, pp. 146-157.

[26] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," Medical image analysis, vol. 54, pp. 30-44, 2019.

[27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[28] M. Mirza, S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," In Advances in neural information processing systems, 2012, pp. 1097-1105.

[31] S. Shalev-Shwartz, and A. Tewari, "Stochastic methods for l1-regularized loss minimization," Journal of Machine Learning Research, vol. 12, pp. 1865-1892, 2012

[32] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang, "Multi-class generative adversarial networks with the L2 loss function," arXiv preprint arXiv:1611.04076, 2016

[33] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," In ICML, 2016, vol. 2, no. 3, p. 7.

[34] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," In International conference on medical image computing and computer-assisted intervention, 2016, pp. 424-432.

[35] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, "The multimodal brain tumor image segmentation benchmark (BRATS)," IEEE transactions on medical imaging, vol. 34, no. 10, pp. 1993-2024, 2014.

[36] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, M. Prastawa, "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge," 2018, arXiv preprint arXiv:1811.02629

[37] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, "The virtual skeleton database: an open access repository for biomedical research and collaboration," Journal of medical Internet research, vol. 15, no. 11, pp. e245, 2013.

[38] D. P. Kingma, and M. Welling, "Auto-Encoding Variational Bayes," stat, vol. 1050, pp. 1, 2014.

[39] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," arXiv preprint arXiv:1511.05644, 2015

[40] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," In International Conference on Machine Learning, 2016, pp. 1558-1566.

[41] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," stat, vol. 1050, p. 9, 2017.

[42] M. Mirza, and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[43] D. Zimmerer, F. Isensee, J. Petersen, S. Kohl, K. Maier-Hein, "Unsupervised Anomaly Localization Using Variational Auto-Encoders," In MICCAI, 2019, Lecture Notes in Computer Science, vol. 11767.

[44] X. Chen, N. Pawlowski, M. Rajchl, B. Glocker, E. Konukoglu, "Deep generative models in the real-world: an open challenge from medical imaging," arXiv preprint arXiv:1806.05452.

[45] B. N. Saha, N. Ray, R. Greiner, A. Murtha, H. Zhang, "Quick detection of brain tumors and edemas: A bounding box method using symmetry," Computerized medical imaging and graphics, vol. 36, no. 2, pp. 95-107, 2012.

[46] M. Erihov, S. Alpert, P. Kisilev, S. Hashoul, "A cross saliency approach to asymmetry-based tumor detection," In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 636-643.

[47] S. You, K. C. Tezcan, X. Chen, E. Konukoglu, "Unsupervised Lesion Detection via Image Restoration with a Normative Prior," In International Conference on Medical Imaging with Deep Learning, 2019, pp. 540-556.

[48] X. Xuan, Q. Liao, "Statistical structure analysis in MRI brain tumor segmentation," In *Fourth International Conference on Image and Graphics*, 2007, pp. 421-426.

[49] H. B. Nandpuru, S. S. Salankar, V. R. Bora, "MRI brain cancer classification using support vector machine," In *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*, 2014, pp. 1-6.

[50] K. Machhale, H. B. Nandpuru, V. Kapur, L. Kosta, "MRI brain cancer classification using hybrid classifier (SVM-KNN)," In *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 60-65.

[51] H. Khotanlou, O. Colliot, J. Atif, I. Bloch, "3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models," *Fuzzy sets and systems*, 2009, pp. 1457-1473.

[52] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, K. Ugurbil, "for the WU-Minn HCP Consortium (2013). The WU-Minn Human Connectome Project: An overview," NeuroImage, 2013, vol. 80, pp:62-79.

[53] L. Zhao, W. Wu, J. J. Corso, "Semi-automatic brain tumor segmentation by constrained MRFs using structural trajectories," In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2013, pp. 567-575.

[54] S.Ourselin, R. Stefanescu, X. Pennec: Robust registration of multi-modal images: Towards real-time clinical applications. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002, Part II. LNCS, vol. 2489, pp. 140–147. Springer, Heidelberg (2002)

[55] L. Ibanez et al., The ITK Software Guide. Clifton Park, NY: Kitware, 2003.

[56] S. Bauer, T. Fejes, and M. Reyes, "A skull-stripping filter for ITK," Insight J., pp. 70–78, 2012

[57] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.