Article

Multimodal CT Image Synthesis Using Unsupervised Deep Generative Adversarial Networks for Stroke Lesion Segmentation

Suzhe Wang * , **Xueying Zhang, Haisheng Hui, Fenglian Li and Zelin Wu**

College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China
* Correspondence: wangsuzhe@tyut.edu.cn

Abstract: Deep learning-based techniques can obtain high precision for multimodal stroke segmentation tasks. However, the performance often requires a large number of training examples. Additionally, existing data extension approaches for the segmentation are less efficient in creating much more realistic images. To overcome these limitations, an unsupervised adversarial data augmentation mechanism (UTC-GAN) is developed to synthesize multimodal computed tomography (CT) brain scans. In our approach, the CT samples generation and cross-modality translation differentiation are accomplished simultaneously by integrating a Siamesed auto-encoder architecture into the generative adversarial network. In addition, a Gaussian mixture translation module is further proposed, which incorporates a translation loss to learn an intrinsic mapping between the latent space and the multimodal translation function. Finally, qualitative and quantitative experiments show that UTC-GAN significantly improves the generation ability. The stroke dataset enriched by the proposed model also provides a superior improvement in segmentation accuracy, compared with the performance of current competing unsupervised models.

Keywords: stroke lesion segmentation; generative adversarial network; unsupervised data augmentation

1. Introduction

Stroke is the problem with blood supply blocking in cerebral vessels, and it is the most prevalent cause of mortality and acquired handicap [1,2]. Among various types of strokes, ischemic stroke is reported in a large proportion, and it mainly induces brain cell death and fatal paralysis. Hence, early diagnosis and quantification of the lesions could help stroke patients achieve effective recovery, which also benefits clinicians in optimizing the therapeutic schedules. Quantitative stroke lesion segmentation from medical imaging is a necessary procedure for the doctor to make decisions. Additionally, computed tomography (CT) is a typical effective non-incursion technique to evaluate the lesion regions of stroke patients [3,4]. It also has the merits of speediness, wide availability and inexpensiveness in detecting brain structure by ionizing radiation. Moreover, CT perfusion modalities including cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT) and time to peak of the residue function (Tmax) are also successfully used to assess stroke infarct core size [5]. However, precise segmentation based on these diagnostic means requires rich experience and a significant amount of time from physicians.

Recently, deep neural network-based methods have shown a remarkable impact on the segmentation accuracy for various medical images [6,7]. However, the scarcity of labelling multimodal pictures due to the enormous time cost and complex acquisition procedures often leads to low segmentation accuracy. Although previous image augmentation strategies such as rotation, flipping and elastic deformation have been applied widely to expand the volume of the dataset [8–10], it is impossible to produce a wide diversity of new features from the aspects of texture, shape and location. The disadvantages also limit the learning

capacity of the deep medical image segmentation model. Generative Adversarial Network (GAN) is a solution for verisimilar image generation and domain translation [11,12]. Nonetheless, this technique always relies on fully annotated paired images and supervised training, which is impractical to collect all modalities for each patient. Additionally, they always have an arduous process for the cross-modality translation. Several recent GAN variants try to tackle the problem by encouraging the extra encoders to capture the domain features [13–16]. However, the quality of the synthetic image may be negatively influenced when the difference between domains increases.

In this work, we proposed a GAN-based data enhancement architecture for CT ischemic stroke lesion segmentation. An unsupervised translation cycle generative adversarial network (UTC-GAN) is presented for the sake of the segmentation accuracy improvement. The main contribution includes the following:

- We develop an image augmentation architecture that is capable of synthesizing CT images and automatic learning translation from CT to its perfusion domains. It allows us to tackle the data scarcity problem for stroke segmentation.
- A Gaussian mixture–translation representation module (GM-TRM) is proposed to learn the transformations between CT modalities automatically, and its corresponding translation loss function is defined. The module guarantees the model could flexibly learn the various translations.
- Experiments demonstrate the proposed model offers a better quality of synthesis imaging and lesion segmentation accuracy on popular datasets than its counterparts.

2. Related work

2.1. Medical Image Segmentation

Numerous deep network designs have been exploited for medicine image segmentation in recent years [17–19]. Albert Clèrigues et al. [20] introduced symmetrical residual auto-encoding U-Net to perform lesion segmentation on CT images. Meanwhile, modality augmentation is utilized to provide more symmetric samples. Liu et al. [21] embeds an attention component in the deep CNN architecture to improve the predictive quality for white matter hypertension lesions. Furthermore, Zhang et al. [22] employed a 3D DenseNet model with dense block and multi-scale unit to localize the stroke lesions with harsh noise and low picture quality. Among those convolutional neural networks, U-Net is an influential architecture for biological image segmentation. For example, in [23], a U-Net model was applied to complete interwoven neurons and neurites segmentation tasks. Instead of only including contracting and expanding paths, Cui et al. [24] exploited a Bi-Directional ConvLSTM U-Net for blood vessel segmentation to fuse higher resolution features and semantic information.

Traditional works mainly focus on supervised learning. However, it requires sufficient training images with pixel-wise annotations. Therefore, some segmentation studies based on data augmentation are reported [25,26]. In these works, the supervised generative adversarial networks (GAN) are well used to expand the data amount by modelling proper data distribution in a two-player game framework. Jelmer et al. [27] employed DCGAN to complete brain CT image synthesis from MRI. A similar method has been employed in [28] to convert the T1 MRI scans into the T2 modality. Moreover, a multi-stage GAN method is designed to form image–mask pairs for segmentation task, by using the U-Net-like WGAN-GP as the central architecture [29]. To gain a higher augmentation quality and segmentation performance, a multi-scaled GAN framework, which also preserves the boundary of the tumor core, is composed to collaborate with the U-Net [12]. In addition, GAN is utilized for the unbalanced semantic segmentation task to balance data distribution [30]. Most GANs are hard to translate more than two domains or need to obtain the domain label for all training images.

2.2. Unsupervised Generative Adversarial Network

Several works use GANs to shift domain by unsupervised transferring of information between multiple modalities. Andrade et al. [31] trained Cycle-GAN to transform skin images from the macroscopic domain into the dermoscopic domains, which helps to acquire better segmentation capacity. Chen et al. [32] adapted a formation and appearance-detached augmentation GAN for unannotated cardiac CT image segmentation, synthesizing data from the annotated MRI slices. A unified generative adversarial network is established to execute 3D multimodal segmentation [33]. Such ideas above use other information to guide the generator in GAN to master the domain translation mapping. However, it is hard to capture semantic information across multimodal domains. To improve shape transformation and focus on the difference among the domains, U-GAT-IT [13] introduces an attentional component and a new layer instance-normalization technique in the least-squares GAN to complete the unsupervised domain transfer task. More recently, a tuple of concurrent GANs are designed to perform multi-class unsupervised image domain translation through conditional image generator and multi-task adversarial discriminator [14–16], where the generator is used for encoding content and class. In other words, the figure from a certain domain is composed of content and type information simultaneously.

3. Network Implementation

In this section, the overall UTC-GAN framework is firstly given. After that, we then offer the details of the translation representation module (GM-TRM) and objective loss functions accordingly.

3.1. Model Architecture

The image synthesizing is realized by a UTC-GAN module, which allows translating the CT slices from one modality domain to another. This model inherits the benefits of Cycle-GAN [31], which consists of a forward and backward cycle. The forward cycle of UTC-GAN is illustrated in Figure 1 and vice versa. It is noted that generator G_A is designed to transfer CT images to its perfusion modalities on the dimension of latent space, and G_B provides the reversed mapping images. The discriminators D_B try to discriminate whether the synthesized images are fact or fiction. Inspired by the observation in the auto-encoding transformations approach [34], a paralleled auto-encoding structure is embedded in discriminator D_B to extract the representation of modality transformation automatically. Therefore, the discriminator is partitioned into diverse components: encoder E, decoder D_e and classifier C. Additionally, the architecture of each part is illustrated in Figure 2. Firstly, two encoders E_A, E_B are trained to extract the desired parameters of the training image from different domains, which uses a Siamese structure to share weight and obtain a co-training information. Each encoder introduces the AlexNet [17] as the backbone, and the Inception v1 block is embedded to enhance the network convergence. The decoder D_e, a network with one convolutional layer and one fully connected layer, is coupled with the encoders to estimate the modality translation from the fused features. Finally, one classifier C_{Adv} is added upon the encoder E_A from generated domain to decide whether the synthesized image is real or not, and another C_{Tran} is built upon the decoder D_e to distinguish which transformation is inputted. Each classifier contains two convolutional layers followed by two other fully connected layers.
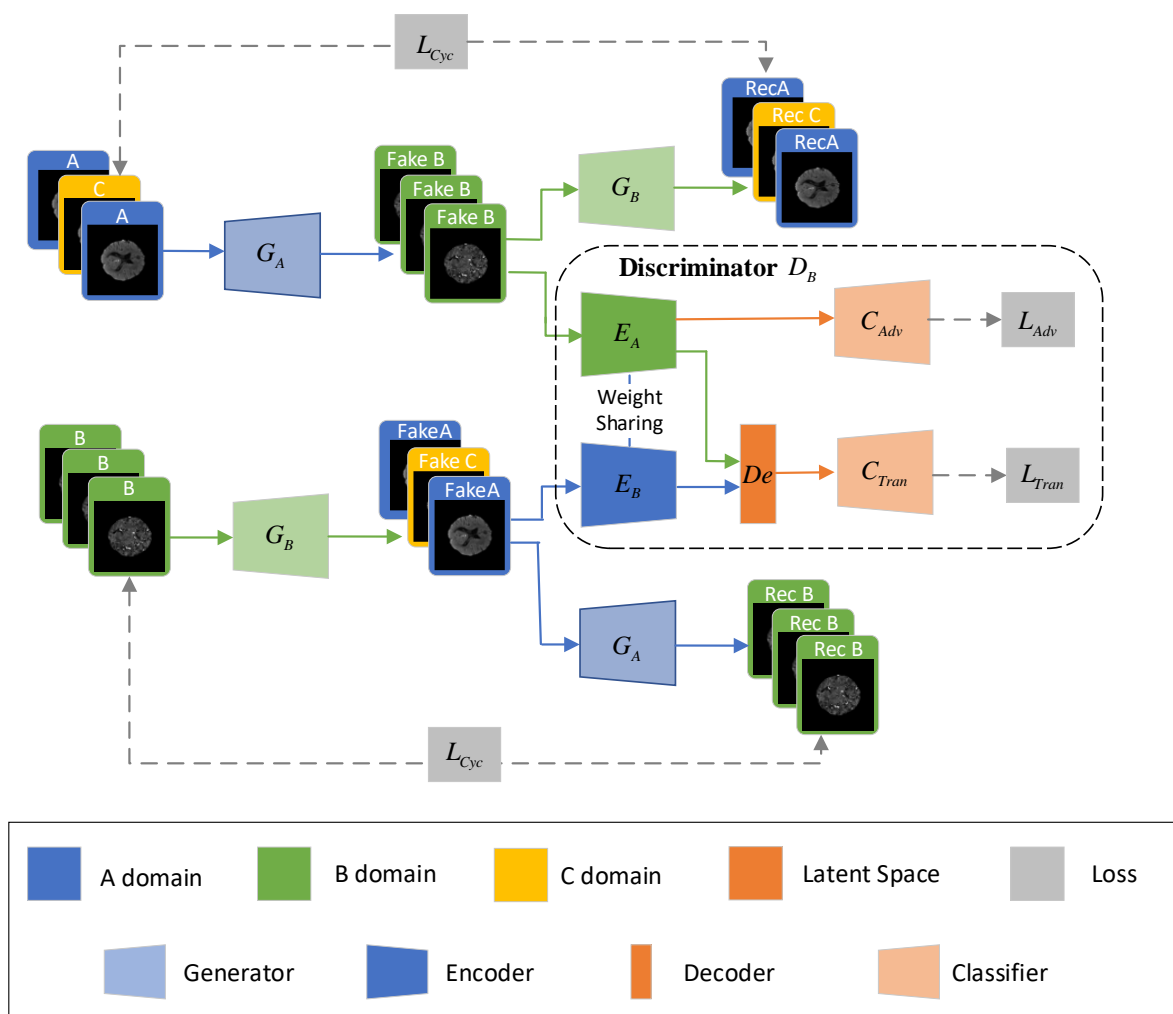
Figure 1. Illustration of the UTC-GAN architecture.

3.2. Gaussian Mixture–Translation Representation Module

As in the UTC-GAN, the auto-encoding structure, also called the Gaussian mixture–translation representation module (GM-TRM), is proposed for learning the modality transformation representations automatically. The GM-TRM module first encodes the CT images from different modalities into a latent space. Then, the translation representation is extracted (self-supervised) in the decoding part via estimating the mutual information across the latent features. The GM-TRM module would lead the generator to build highly entangled translation representations, and would enforce the discriminator to obtain more additional supervision on image generation. Here, we elaborate on the principle of the GM-TRM as shown in Figure 3.

Let d(x) presents the data distribution of an image x from the original modality A. When it is translated to another modality B, a translation function t is assumed to turn $d(x)$ into $t(d(x))$. The encoder E_A is considered to produce a representation from the image $t(x)$ with network parameters θ, which also maps the low-dimensional input data $t(x)$ to a high-level latent variable r. The classical reflection of the latent variable r from encoder is a statistical formulation, which is specified by the mean m_θ and variance σ_θ^2 of a normal distribution $N(\varepsilon|0, I)$, such that:

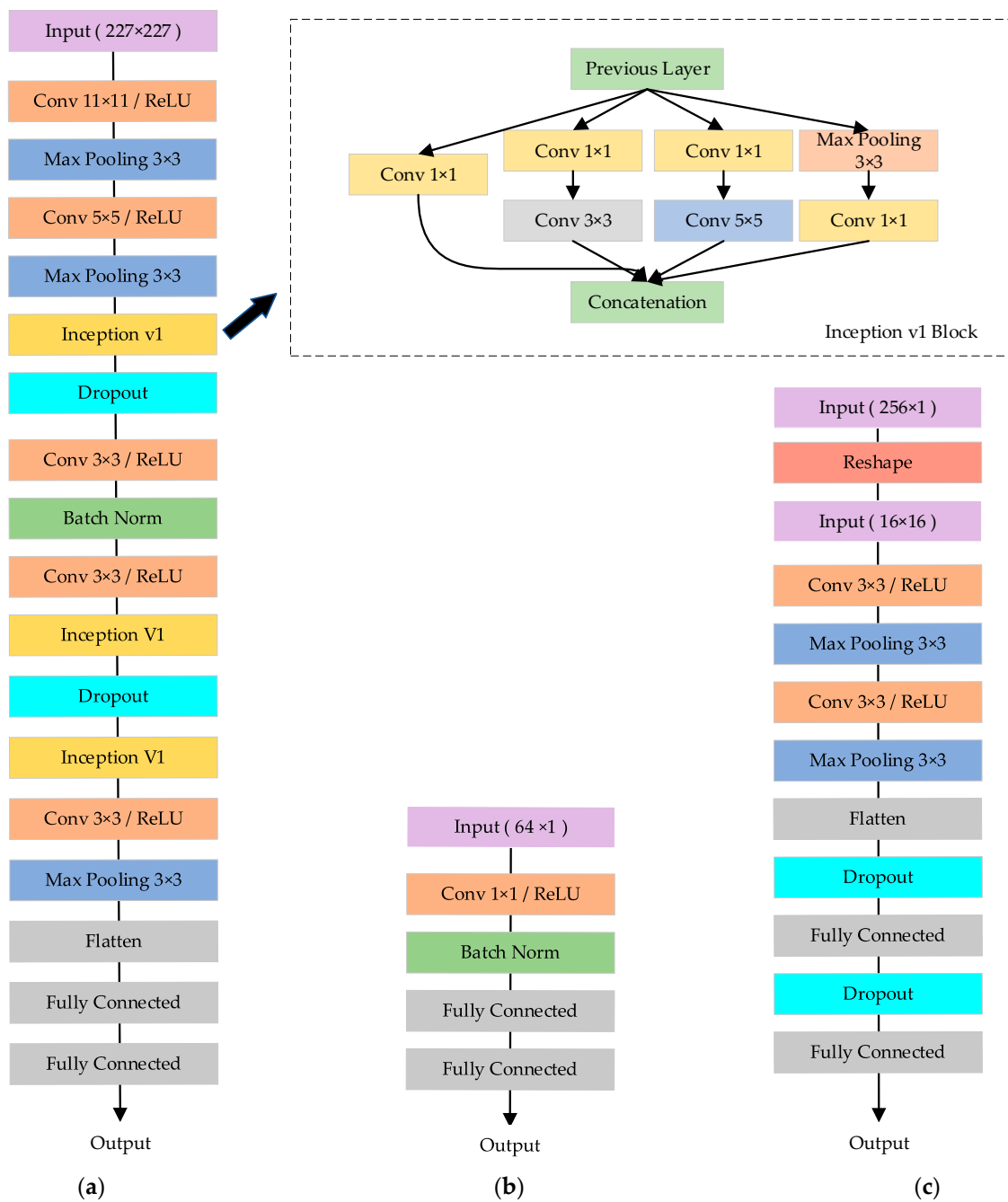$$r = m_\theta(t(x)) + v_\theta(t(x)) \tag{1}$$

Figure 2. Illustration of the component architecture in discriminator for (**a**) Encoder; (**b**) Decoder; (**c**) Classifier.

To better approximate the transformation in latent space, the probabilistic density of encoding images can be assumed to follow the Gaussian mixture model instead of a single Gaussian. The resultant probabilistic representation of true posterior can be formulated as:

$$d_\theta(r|t,x) = \sum_{i=1}^{k} \phi_i N\left(r_i \Big| m_{i\theta}(t(x)), \sigma_{i\theta}^2 t(x)\right) \sum_{i=1}^{k} \phi_i = 1 \qquad (2)$$

where N denotes a standard normal distribution, k is the total count of transformations, ϕ_i is a weight vector, and $m_{i\theta}$ and $\sigma_{i\theta}^2$ are defined as the mean and variance of the ith Gaussian component, respectively.
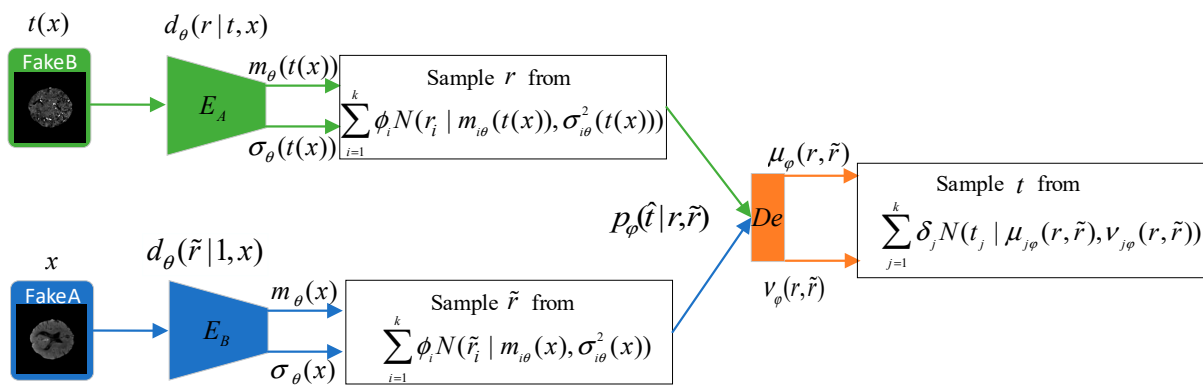
Figure 3. The workflow of Gaussian mixture translation representation module (GM-TRM).

Meanwhile, we train a decoder D_e to estimate the applied transformation parameters t by comparing the representations from encoded image features of original and target modality. The probabilistic density of decoder is defined as $p_\varphi(t|r,\tilde{r})$ with the parameter φ, and $\tilde{r}$ represents the probabilistic density from sample of original domain.

From an information-theoretic viewpoint [35], the transformation t can be equivalent to the joint mutual information $I(t,(r,\tilde{r}))$ between itself and the encoding latent space $(r,\tilde{r})$, and the maximum $I(t,(r,\tilde{r}))$ can be considered the optimized representation. However, the posterior $p_\varphi(t|r)$ cannot be calculated directly. So, a moment matching approximation approach is introduced to compute $p_\varphi(t|r)$ conveniently, which can be derived to the following formulation:

$$
\begin{aligned}
I(t,r) &= H(t) - H(t|r) \\
&= H(t) + \sum_{i=1}^{k} \phi_i E \log d_{i\theta}(t_i|r_i) \\
&= H(t) + \sum_{j=1}^{k} \delta_j E \log p_{j\varphi}(t_j|r_j) + \sum_{i=1}^{k}\sum_{j=1}^{k} E(KL(d_{i\theta}(t_j|r_i)||p_{j\varphi}(t_j|r_i))) \\
&\geq H(t) + \sum_{j=1}^{k} \delta_j E_{d_\theta(t,r|x)} \log p_{j\varphi}(t_j|r_i)
\end{aligned}
\tag{3}
$$

where $KL(d_{i\theta}(t_j|r_i)||p_{j\varphi}(t_j|r_i)$ represents a non-negative Kullback–Leibler divergence between density $d_{i\theta}$ and $p_{j\varphi}$, δ_j is the mixture weight. From this representation, the variational posterior distribution $d_\theta(r|t,x)$ can be tractable and replaced approximated by an upper-bounded parameterized model $p_\varphi(t|r)$. In addition, the entropy $H(t)$ is independent of parameters θ and φ with respect to GAN model. So, we can maximize the lower variational bound $I(t,r)$ through only calculating $\log p_\varphi(t|r,x)$.

In the meantime, the decoded transformation vector can be associated with the input images pairs, which enforces the generator to utilize classified transformation information as well. Hence, the corresponding generated image will learn different attributes from the other domains.

3.3. Loss Function

The UTC-GAN not only relies on the model architecture mentioned above but also the appropriate loss function to perfect the model performance. Our loss includes three components, of which the translation representation loss is newly proposed to learn the style transferring from the CT slices pair automatically.

3.3.1. Adversarial and Cycle-Consistency Loss

The adversarial and cycle-consistency loss from Cycle-GAN [30] are employed both in the generation cycle and discriminator. For the images translation $A \rightarrow B$, the adversarial loss can be expressed as:

$$L_{Adv}(G_A, D_B) = E_{v \sim p_B(v)}[log D_B(v_B)] + E_{u \sim p_A(u)}[log(1 - D_B(G_A(u_A)))] \qquad (4)$$

where u and v are the training slices come from source and target modality. Similarly, the adversarial loss of the translation $B \rightarrow A$ is denoted as:

$$L_{Adv}(G_B, D_A) = E_{u \sim p_A(u)}[log D_A(u_A)] + E_{v \sim p_B(v)}[log(1 - D_A(G_B(v_B)))] \qquad (5)$$

In order to ensure the generated slices can be reconstructed to their previous modality simultaneously, a cycle-consistency loss $L_{Cyc}(G_A, G_B)$ is utilized into architecture to associate the reconstructed image $G_A(G_B(u_A))$ with the input image u. Thus, the loss function with forward–backward consistency is defined as:

$$
\begin{aligned}
L(G_A, G_B, D_A, D_B) \quad &= L_{Adv}(G_A, D_B) + L_{Adv}(G_B, D_A) + \lambda_1\left(L_{Cyc}(G_A, G_B)\right. \\
&\left. +L_{Cyc}(G_B, G_A)\right)
\end{aligned}
\qquad (6)
$$

where λ_1 is the relative importance of GAN loss concering cycle loss.

3.3.2. Translation Representation Loss

Since the adversarial loss cannot detect the transformation directly from encoders, we have used the joint mutual information $I(t, v|u)$ as the decoder to predict domain translation. Additionally, the lower variational bound of $I(t, v|u)$ can be maximized by learning the expectation over posterior distribution $p_\varphi(t|v)$ according to (2). Thus, the translation forecasting loss L_{Tran} can be described as:

$$
\begin{aligned}
L_{Tran} &= max I(t, v|u) = max_{\theta, \varphi} E log p_\varphi(t|r) \\
&= max_{\theta, \varphi} \delta_j \sum_{j=1}^{k} log N(t|u_j(r), v_j^2(r))
\end{aligned}
\qquad (7)
$$

where the mean u_j and variance v_j are derived from the encoder, respectively, and δ_j is a weight vector. This loss function, L_{Tran}, is added to the discriminator D_B to learn the transformation functions between different domains. Then, the whole augmented objective function is given by:

$$L_{Total} = L_{Adv} + \lambda_1 L_{Cyc} + \lambda_2 L_{Tran} \qquad (8)$$

where λ_2 is a hyper-parameter applied for affecting the significance proportion in total loss.

4. Experiments

4.1. Experimental Settings

4.1.1. Dataset

The Ischemic Stroke Lesion Segmentation Challenge (ISLES) 2018 dataset is used to execute the training and assessment of our augmenting-based segmentation. The dataset contains multiple modalities, including CT and four derived perfusion maps, i.e., mean transit time (MTT), time to peak of the residue function (Tmax), cerebral blood flow (CBF) and cerebral blood volume (CBV) [36,37]. In our experiments, the 94 labelled cases with the CT and its perfusion modalities serve as input for the UTC-GAN network.

Additionally, all images in the dataset are preformed augmentation via skull stripping and traditional operations such as flipping, scaling and rotating. Finally, eighty percent of all the image scans are treated as the training and validation set, and the remainder is the testing set. That is, we uses 10,980 slices for training, 3660 slices for validating and 3660 slices for testing, respectively.

4.1.2. Baseline Model and Evaluation Measures

For comparison in augmentation, we use the ISLES2018 dataset to compare UTC-GAN to five existing unsupervised baseline models: Cycle-GAN [30], DRIT++ [14], EGSC-IT [15], FUNIT [16] and U-GAT-IT [13]. The Cycle-GAN and U-GAT-IT model have a similar mechanism and effectiveness to the proposed method. DRIT++ introduces the disentangled representation model to learn the mapping from the CT to other perfusion modalities. EGSC-IT realizes unsupervised image-to-image translation by adopting weight-sharing architecture and feature masks. FUNIT is chosen as another comparable baseline model to perform image translation in that it has a multi-task synthesis structure with reconstruction and feature-matching loss function. At the segmenting phase, a multiple-scale minus network (MSNet) [38] is trained to evaluate the improvement by the generative model. To input the multi-modalities CT slices, the encoder path of the MSNet is duplicated five times and concatenated to the decoding part. Due to computational constraints, every slice image is resized to 256×256.

We verify the synthesis methods by adopting four widely used metrics: Peak Signal to Noise Ratio (PSNR), Normalized Mean Squared Error (NMSE) and Structural Similarity Index Measurement (SSIM) [39].

Six performance metrics are chosen to analyze the improvement in the segmentation accuracy for the testing set, including Dice Coefficient (DC), Intersection-over-Union (IoU) score, Precision, Accuracy, Recall and Hausdorff Distance (HD) [40].

4.1.3. Implementation Details

In the data augmentation stage, we adopt the ResNet architectures from Miyato et al. [20] as the backbone of the UTC-GAN generator. For UTC-GAN discriminator, encoder E_A and E_B in each branch consists of InceptionV1 block, then the output features from decoder D_e are concatenated to a convolutional classifier C_{Tran}. Classifier C_{Adv} is framed as same as C_{Tran}. During the segmentation stage, a standard U-Net framework is trained for segmenting stroke lesion areas. The back-propagation of both networks were completed by adopting ADAM optimization algorithm [41] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate was positioned at 2×10^{-4} for all networks. The loss-balancing weight parameters λ_1 and λ_2 were determined as 15 and 10, respectively. The exponential decay rate is used at a rate of 0.001 every 30 epochs, training the model for a total of 200 epochs with batch size of 24. Each paired input contains the original CT slice and their perfusion modality counterpart.

The synthesis and segmentation networks were implemented by Pytorch on the NVIDIA Titan XP GPU device. Furthermore, the overall experiments are conducted five times with varying random seeds, and the average value is reported.

4.2. The Impact of GM-TRM Module

First, we compare the generation quality to assess the effectiveness of components in our UTC-GAN model. The main component including the GM-TRM module and the classifier C_{Tran} concatenated to decoder D_e are replaced in the architecture and compared with other existing schemes in sequence. As the competitors for comparison, the AET module also employs the auto-encoding architecture to learn the unsupervised transformation representation. Additionally, the AVT module creates the translation information by applying a constrained variational approach to the similar auto-encoders network. Table 1 firstly presents the quantitative synthesis performance with GM-TRM and other modules. We observe that the GM-TRM enhances the synthesized performance by leading to 1% improvement on average over PSNR, and 0.3% over NMSE. The evaluation result demonstrates that the mixture density can derive better transformation representation. Moreover, we also quantitatively compare the results by varying different classifiers upon the GM-TRM. The traditional non-linear classifier adopts fully connected layers to discrete transformations, whereas the convolutional type takes advantage of the convolution kernel to narrow the range of the transformation prediction. From the results, we can see that

the convolutional classifier is consistently better than the non-linear classifier, and our synthesis model can almost achieve the best PSNR no matter which classifier is used.

Table 1. Synthesizing performance of the UTC-GAN composed of different components.

Method	PSNR	NMSE	SSIM
AET + non-linear [42]	24.12	0.125	0.8654
AET + conv [40]	24.86	0.125	0.8675
AVT + non-linear [34]	24.47	0.122	0.8590
AVT + conv [34]	25.19	0.119	0.8678
GM-TRM + non-linear	25.97	0.109	0.8816
GM-TRM + conv	26.40	0.096	0.918

4.3. Comparison with Other Unsupervised Data Augmentation Methods

Moreover, we next evaluate the UTC-GAN compared with various corresponding synthesis models.

Table 2 reports the quantitative synthesis results for all baselines. From the results underlined in Table, the UTC-GAN surpasses the other five unsupervised models by raising the PSNR from 23.13 to 26.40, and the NMSE approximates 0.096 when the SSIM achieves 0.918. This indicates our model attains the transformation characteristic across different modalities, and that it contributes to achieving superior synthesis effectiveness.

Table 2. Synthesizing performance comparison with different baselines.

Method	PSNR	NMSE	SSIM
Cycle-GAN	23.13	0.153	0.8594
U-GAT-IT	23.25	0.149	0.8673
FUNIT	23.09	0.233	0.8516
EGSC-IT	23.86	0.146	0.8624
DRIT++	24.72	0.115	0.8946
UTC-GAN	26.40	0.096	0.918

Figure 4 shows a visualization comparison under different types of modalities between the proposed UTC-GAN model and other unsupervised baselines. As we can see, the UTC-GAN generates much more realistic synthesis images, while samples from Cycle-GAN and U-GAT-IT generate some unclear regions or fail to create a detailed feature. In contrast, samples from FUNIT, EGSC-IT and DRIT++ show similar results, which supply more complex attributes of the brain to a certain extent but also have some unwanted artifacts that can be found in the image. Overall, the proposed method yields higher visual realism results for all CT modalities than the others, as indicated by qualitative and quantitative measures.

4.4. Segmentation Using Data Augmentation

To investigate the influence of UTC-GAN on the segmentation improvement, we first use the real and synthetic data to provide for the MSNet segmentation model, the performance evaluation is described in Figure 5. The synthetic images have a positive effect on the segmentation accurateness where the synthetic data ratio is less than 60%. The increasing proportion of the synthetic images leads to 7% and 8% improvement on the dice score and precision than the results obtained only by the real images. Moreover, we found that too many synthetic images could degrade the performance.
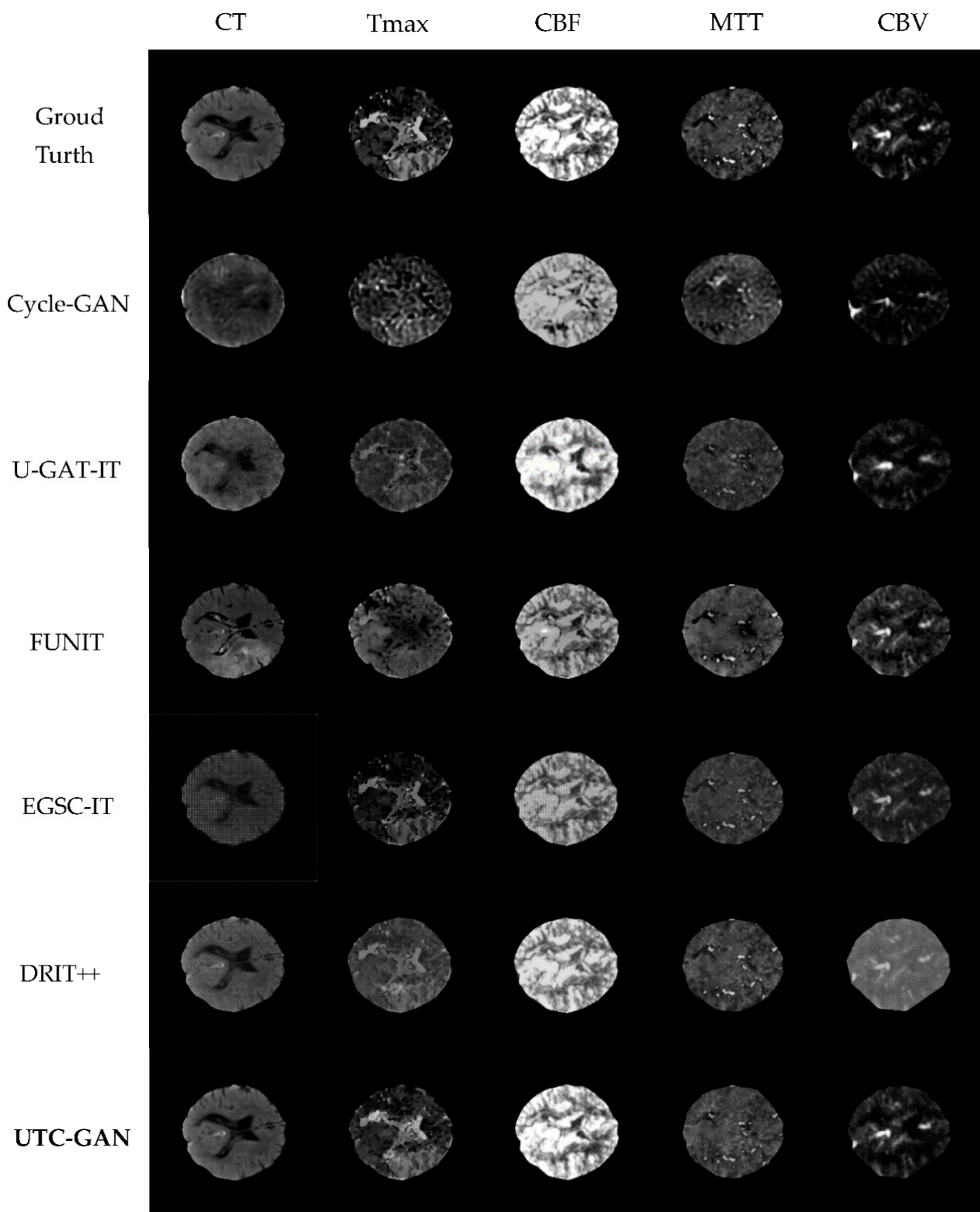
Figure 4. The visualizations of real and generated CT slices under different modalities. From left to right: CT, Tmax, CBF, MTT, CBV modality. The top-down results: Input ground truth, Cycle-GAN, U-GAT-IT, FUNIT, EGSC_IT, DRIT++ and our augmentation method.
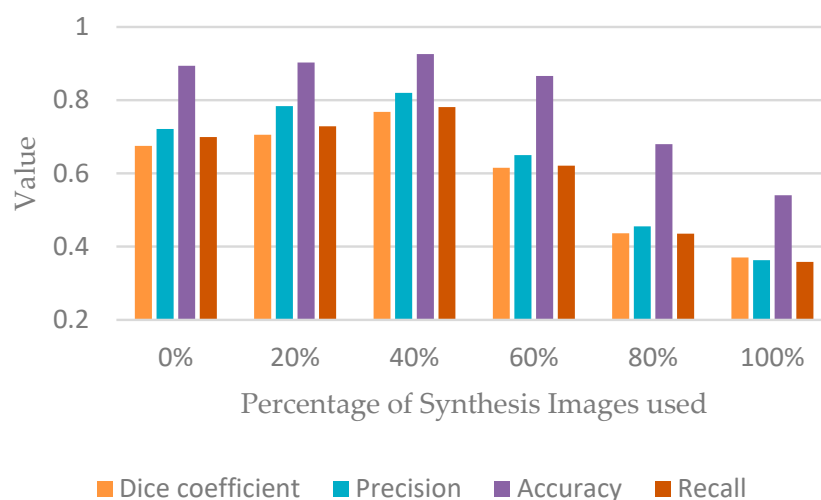
Figure 5. The performance of stroke segmentation by changing the percentage of synthetic images.

The second experiment is conducted on the task of segmentation by applying both all the data augmentation methods mentioned earlier and the MSNet model on ISLES2018 data. The quantitative results of segmentation can be viewed in Table 3. According to the demonstration in the table, all of the synthesis models provide more significant improvement on all evaluation metrics than only using the MSNet segmentation method with the randomness added by data pre-processing. Additionally, our process appears to produce the most significant improvement when generating transformation detectable images. For example, our proposed UTC-GAN allows the Dice score to improve from 0.675 to 0.768, increasing the precision by and 10% and reducing the Hausdorff distance by 9.3 mm, respectively. The IoU score and accuracy also achieved the highest value. This improvement may benefit the most due to the high contrast and variation from the generated images. Furthermore, the boxplot of the Dice coefficient is also used to analyze the segmentation robustness in Figure 6. Overall, the mean value and the median line of the UTC-GAN expresses relatively higher improvement than the other six competitors.

Table 3. Segmentation performance comparison with different strategies.

Method	Dice Coefficient	IoU	Precision	Accuracy	Recall	Hausdorff Distance (mm)
MSNet	0.675	0.578	0.721	0.894	0.699	27.03
Cycle-GAN + MSNet	0.709	0.633	0.773	0.903	0.725	25.00
U-GAT-IT + MSNet	0.684	0.602	0.745	0.907	0.687	24.21
FUNIT + MSNet	0.718	0.645	0.786	0.912	0.732	20.34
EGSC-IT + MSNet	0.751	0.683	0.809	0.925	0.773	34.11
DRIT++ + MSNet	0.735	0.669	0.781	0.915	0.758	19.99
UTC-GAN + MSNet	0.768	0.685	0.820	0.926	0.781	17.73

Additionally, three representative cases of visual segmentation scan for variant combinations are displayed in Figure 7. Compared with other data augmentation means, the UTC-GAN produces considerable improvements in the segmentation mask, allowing segmentation regions to be closer to their manual annotation counterpart. This intuitively demonstrates that the stroke CT images augmented by the UTC-GAN model help obtain a more accurate lesion region.
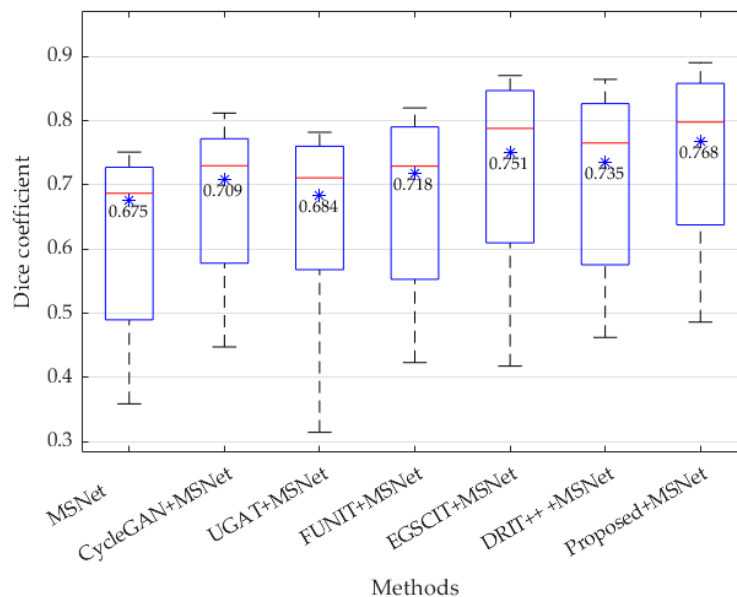
Figure 6. The Dice coefficient in boxplot by different methods, where the asterisk * and number in black represent the respective mean value.
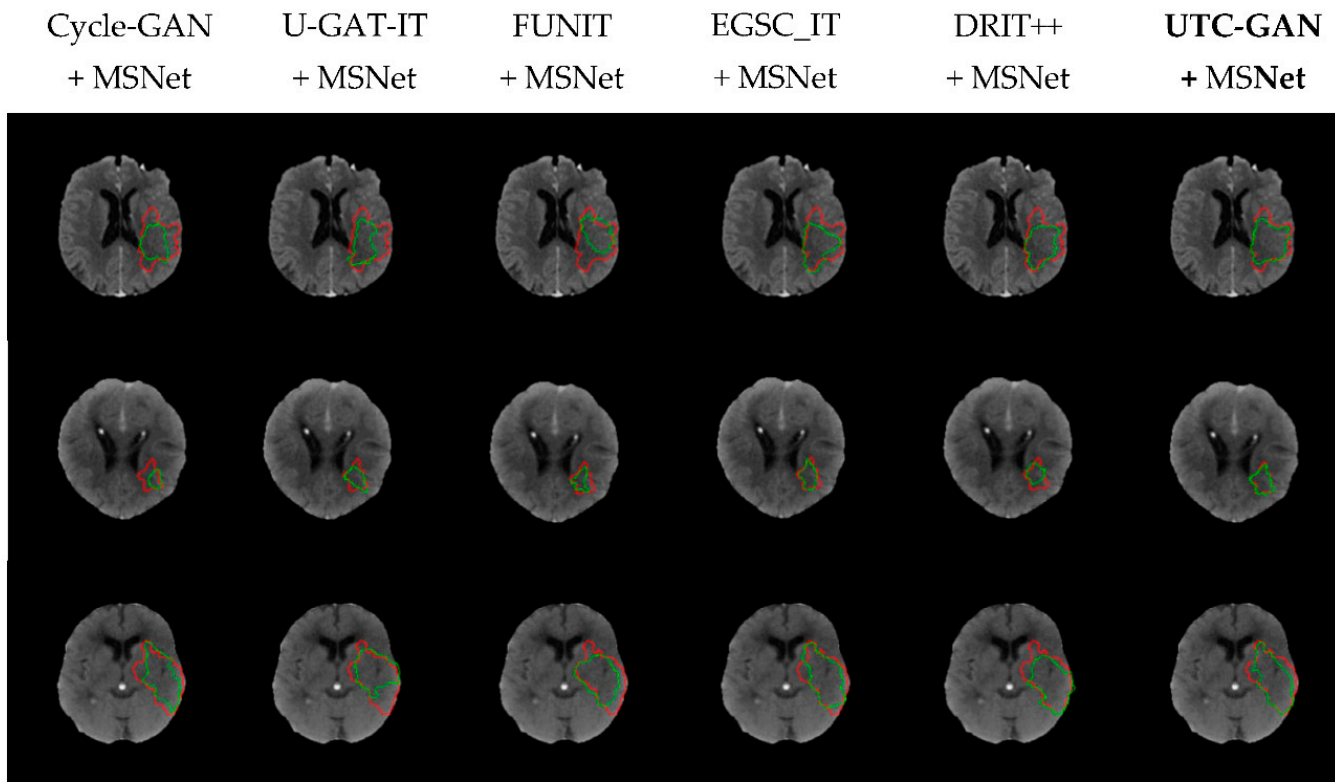

Figure 7. Visual results of lesion segmentation processed via the MSNet and data augmentation models. The boundaries of expertise-based and unsupervised learning-based segmentation are painted with red and green curves, respectively.

5. Conclusions

In this paper, a GAN-based data augmentation paradigm is presented to promote the exactness of ischemic stroke segmentation. By integrating Siamesed auto-encoders and information-theoretic loss into a Cycle-generative adversarial framework, the architecture can learn sufficient representations about transformation from the original to the target

domain in an unsupervised fashion. The sampling problem on latent data space is further solved by introducing a Gaussian mixture probability distribution to better approximate the characteristics of the transformation. Based on the experimental evaluation and comparison, it demonstrates that the proposed method outperforms alternative structures and provides better high-quality generated stroke images. Meanwhile, the augmentation method could yield higher segmentation quality improvement by cooperating with traditional segmentation methods.

Author Contributions: Conceptualization and methodology, S.W. and X.Z.; formal analysis, S.W., X.Z., H.H., F.L. and Z.W.; software, validation and writing—original draft, S.W., X.Z. and H.H.; writing—review and editing and data curation, S.W., X.Z., H.H. and F.L.; supervision and funding acquisition, S.W., X.Z. and F.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The ISLES 2018 dataset is publicly available at http://www.isles-challenge.org/ (accessed on 15 July 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Ethical Approval: Dataset was obtained from ISLES challenge, there is no live interaction with subjects and the dataset is anonymous. Hence this article does not contain any studies with live human participants or animals performed by any of the authors.

References

1. Soltanpour, M.; Greiner, R.; Boulanger, P.; Buck, B. Improvement of automatic ischemic stroke lesion segmentation in CT perfusion maps using a learned deep neural network. *Comput. Biol. Med.* **2021**, *137*, 104849. [CrossRef] [PubMed]
2. Gautam, A.; Raman, B. Segmentation of ischemic stroke lesion from 3D MR images using random forest. *Multimed. Tools Appl.* **2019**, *78*, 737–738. [CrossRef]
3. Ostman, C.; Garcia, C.; Lillicrap, T.; Tomari, S. Multimodal Computed Tomography Increases the Detection of Posterior Fossa Strokes Compared to Brain Non-contrast Computed Tomography. *Front. Neurol.* **2020**, *11*, 588064. [CrossRef] [PubMed]
4. Maegawa, T.; Sasahara, A.; Ohbuchi, H.; Chernov, M. Cerebral vasospasm and hypoperfusion after traumatic brain injury: Combined CT angiography and CT perfusion imaging study. *Surg. Neurol. Int.* **2021**, *12*, 361. [CrossRef] [PubMed]
5. Honda, M.; Ichibayashi, R.; Yokomuro, H.; Yoshihara, K.; Masuda, H.; Haga, D.; Seiki, Y.; Kudoh, C.; Kishi, T. Early cerebral circulation disturbance in patients suffering from severe traumatic brain injury (TBI): A xenon CT and perfusion CT study. *Neurol. Med. Chir.* **2016**, *56*, 501–509. [CrossRef] [PubMed]
6. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
7. Chilamkurthy, S.; Ghosh, R.; Tanamala, S.; Biviji, M.; Campeau, N.; Venugopal, V.; Mahajan, V.; Rao, P.; Warier, P. Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* **2018**, *392*, 2388–2396. [CrossRef]
8. Oliveira, A.; Pereira, S.; Silva, C.A. Augmenting data when training a CNN for retinal vessel segmentation: How to warp. In Proceedings of the 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG), Coimbra, Portugal, 16–18 February 2017; pp. 1–4.
9. Noguchi, S.; Nishio, M.; Yakami, M.; Nakagomi, K.; Togashi, K. Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques. *Comput. Biol. Med.* **2020**, *121*, 103767. [CrossRef]
10. Koh, Y.J.; Kim, C.S. Primary object segmentation in videos based on region augmentation and reduction. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–27 July 2017; pp. 7417–7425.
11. Majurski, M.; Manescu, P.; Padi, S.; Schaub, N.; Hotaling, N.; Simon, C., Jr.; Bajcsy, P. Cell image segmentation using generative adversarial networks, transfer learning, and augmentations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1–9.
12. Mok, T.C.; Chung, A. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16–17 September 2018; pp. 70–80.
13. Kim, J.; Kim, M.; Kang, H.; Lee, K. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv* **2019**, arXiv:1907.10830.
14. Lee, H.Y.; Tseng, H.Y.; Mao, Q.; Huang, J.B.; Lu, Y.D.; Singh, M.; Yang, M.H. Drit++: Diverse image-to-image translation via disentangled representations. *Int. J. Comput. Vis.* **2020**, *128*, 2402–2417. [CrossRef]

15. Ma, L.; Jia, X.; Georgoulis, S.; Tuytelaars, T.; Van Gool, L. Exemplar guided unsupervised image-to-image translation with semantic consistency. *arXiv* **2018**, arXiv:1805.11145.

16. Liu, M.Y.; Huang, X.; Mallya, A.; Karras, T.; Aila, T.; Lehtinen, J.; Kautz, J. Few-shot unsupervised image-to-image translation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 10551–10560.

17. Dvořák, P.; Menze, B. Local structure prediction with convolutional neural networks for multimodal brain tumor segmentation. In Proceedings of the International MICCAI Workshop on Medical Computer Vision, Munich, Germany, 9 October 2015; pp. 59–71.

18. Rodríguez Colmeiro, R.G.; Verrastro, C.A.; Grosges, T. Multimodal brain tumor segmentation using 3D convolutional networks. In Proceedings of the International MICCAI Brainlesion Workshop, Quebec City, QC, Canada, 14 October 2017; pp. 226–240.

19. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Van Leemput, K. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef] [PubMed]

20. Clèrigues, A.; Valverde, S.; Bernal, J.; Freixenet, J.; Oliver, A.; Lladó, X. Acute and sub-acute stroke lesion segmentation from multimodal MRI. *Comput. Methods Programs Biomed.* **2020**, *194*, 105521. [CrossRef] [PubMed]

21. Liu, L.; Kurgan, L.; Wu, F.X.; Wang, J. Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. *Med. Image Anal.* **2020**, *65*, 101791. [CrossRef] [PubMed]

22. Zhang, R.; Zhao, L.; Lou, W.; Abrigo, J.M.; Mok, V.C.; Chu, W.C.; Shi, L. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. *IEEE Trans. Med. Imaging* **2018**, *37*, 2149–2160. [CrossRef]

23. Wang, W.; Yu, K.; Hugonot, J.; Fua, P.; Salzmann, M. Recurrent U-Net for resource-constrained segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 2142–2151.

24. Azad, R.; Asadi-Aghbolaghi, M.; Fathy, M.; Escalera, S. Bi-directional ConvLSTM U-Net with densely connected convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, 27–28 October 2019; pp. 1–10.

25. Long, M.; Zhu, H.; Wang, J.; Jordan, M.I. Unsupervised domain adaptation with residual transfer networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 136–144.

26. Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; Krishnan, D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3722–3731.

27. Wolterink, J.M.; Dinkla, A.M.; Savenije, M.H.; Seevinck, P.R.; van den Berg, C.A.; Išgum, I. Deep MR to CT synthesis using unpaired data. In Proceedings of the International Workshop on Simulation and Synthesis in Medical Imaging, Strasbourg, France, 27 September 2017; pp. 14–23.

28. Welander, P.; Karlsson, S.; Eklund, A. Generative adversarial networks for image-to-image translation on multi-contrast mr images-a comparison of cyclegan and unit. *arXiv* **2018**, arXiv:1806.07777.

29. Pandey, S.; Singh, P.R.; Tian, J. An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation. *Biomed. Signal Process. Control* **2020**, *57*, 101782. [CrossRef]

30. Liu, S.; Zhang, J.; Chen, Y.; Liu, Y.; Qin, Z.; Wan, T. Pixel level data augmentation for semantic image segmentation using generative adversarial networks. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1902–1906.

31. Andrade, C.; Teixeira, L.F.; Vasconcelos, M.J.M.; Rosado, L. Data Augmentation Using Adversarial Image-to-Image Translation for the Segmentation of Mobile-Acquired Dermatological Images. *J. Imaging* **2020**, *7*, 2. [CrossRef]

32. Chen, X.; Lian, C.; Wang, L.; Deng, H.; Kuang, T.; Fung, S.H.; Yap, P.T. Diverse data augmentation for learning image segmentation with cross-modality annotations. *Med. Image Anal.* **2021**, *71*, 102060. [CrossRef]

33. Yuan, W.; Wei, J.; Wang, J.; Ma, Q.; Tasdizen, T. Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images. *Med. Image Anal.* **2020**, *64*, 101731. [CrossRef] [PubMed]

34. Qi, G.J.; Zhang, L.; Chen, C.W.; Tian, Q. Avt: Unsupervised learning of transformation equivariant representations by autoencoding variational transformations. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 8130–8139.

35. Agakov, D.B.F. The IM algorithm: A variational approach to information maximization. *Adv. Neural Inf. Process. Syst.* **2004**, *16*, 201–210.

36. Cereda, C.W.; Christensen, S.; Campbell, B.C.; Mishra, N.K.; Mlynash, M.; Levi, C.; Lansberg, M.G. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. *J. Cereb. Blood Flow Metab.* **2016**, *36*, 1780–1789. [CrossRef] [PubMed]

37. Hakim, A.; Christensen, S.; Winzeck, S.; Lansberg, M.G.; Parsons, M.W.; Lucas, C.; Zaharchuk, G. Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: Lessons from the ISLES challenge. *Stroke* **2021**, *52*, 2328–2337. [CrossRef] [PubMed]

38. Zhao, X.; Zhang, L.; Lu, H. Automatic polyp segmentation via multi-scale subtraction network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2021; pp. 120–130.

39. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

40. Nielsen, A.; Hansen, M.B.; Tietze, A.; Mouridsen, K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. *Stroke* **2018**, *49*, 1394–1401. [CrossRef]

41. Sage, A.; Agustsson, E.; Timofte, R.; Van Gool, L. Logo synthesis and manipulation with clustered generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5879–5888.

42. Zhang, L.; Qi, G.J.; Wang, L.; Luo, J. Aet vs. aed: Unsupervised representation learning by auto-encoding transformations rather than data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2547–2555.