ScienceDirect
www.sciencedirect.com

ELSEVIER

Computers in Biology and Medicine Volume , 2021, 00 (2021) 1–27

Paired-Unpaired Unsupervised Attention Guided GAN with Transfer Learning for Bidirectional Brain MR-CT Synthesis

Alaa Abu-Srhan[a], Israa Almallahi[b], Mohammad A. M. Abushariah[c], Waleed Mahafza[b], Omar S. Al-Kadi[c]

[a]*Department of Basic Science, The Hashemite University, Zarqa, Jordan.*

[b]*Department of Diagnostic Radiology, Jordan University Hospital, Amman 11942, Jordan.*

[c]*King Abdullah II School of Information Technology, The University of Jordan, Amman 11942, Jordan.*

Abstract

Medical image acquisition plays a significant role in the diagnosis and management of diseases. Magnetic Resonance (MR) and Computed Tomography (CT) are considered two of the most popular modalities for medical image acquisition. Some considerations, such as cost and radiation dose, may limit the acquisition of certain image modalities. Therefore, medical image synthesis can be used to generate required medical images without actual acquisition. In this paper, we propose a paired–unpaired Unsupervised Attention Guided Generative Adversarial Network (uagGAN) model to translate MR images to CT images and vice versa. The uagGAN model is pre-trained with a paired dataset for initialization and then retrained on an unpaired dataset using a cascading process. In the paired pre-training stage, we enhance the loss function of our model by combining the Wasserstein GAN adversarial loss function with a new combination of non-adversarial losses (content loss and L1) to generate fine structure images. This will ensure global consistency, and better capture of the high and low frequency details of the generated images. The uagGAN model is employed as it generates more accurate and sharper images through the production of attention masks. Knowledge from a non-medical pre-trained model is also transferred to the uagGAN model for improved learning and better image translation performance. Quantitative evaluation and qualitative perceptual analysis by radiologists indicate that employing transfer learning with the proposed paired-unpaired uagGAN model can achieve better performance as compared to other rival image-to-image translation models.

INDEX TERMS:
Generative Adversarial Networks, Unsupervised Image Translation, Paired and Unpaired Data, MR-to-CT Image Synthesis, Transfer Learning.

Corresponding author: o.alkadi@ju.edu.jo (Omar S. Al-Kadi)

1. Introduction

Medical image analysis can assist in revealing the internal structure of body organs. It is widely used in many applications, such as classification [1], detection [2], segmentation [3], and registration [4]. Image acquisition is considered as a critical step for subsequent image analysis. There are many medical image acquisition modalities including Magnetic Resonance (MR), Computed Tomography (CT), Positron Emission Tomography (PET), and Single Photon Emission Computed Tomography (SPECT). The most widely used modalities for brain tumor imaging are MR and CT, where MR is specifically used for tumor volume segmentation and the diagnosis of neuronal pathology (i.e., Parkinson's disease), while CT is used for radiotherapy treatment planning [5].

Medical image acquisition faces many considerations, such as cost, radiation dose, patient age, and limitation of images for certain types of brain disease. Therefore, we could benefit from medical image synthesis by generating required medical images without any physical scan [6]. Radiation is emitted during CT acquisition, thereby limiting the number of scans a patient could undergo without exceeding the permitted radiation levels. However, MR does not involve any radiations. Therefore, various methods have been proposed for CT images estimation from the available MR images. On the other hand, MR imaging is considered costly, and it takes more time than CT scans. If the latter was the major motivation, then MR images could be generated from its corresponding CT images [7]. In some cases, MR and CT imaging are needed. For example, MR scan could show neural elements (spinal cord and nerve roots) exquisitely, but accurately differentiating between bone and soft tissue using MR is difficult. In this case, a CT scan that could accurately delineate bony boundaries, is required [5]. In this paper, the term CT-MR image synthesis is used to indicate bidirectional translation between CT and MR images.

CT-MR image synthesis is useful in a variety of applications, including data augmentation, in which the generated images can be used to improve generalizability in classification and segmentation tasks [8]. CT-MR image synthesis is also used for CT-MR registration, which is necessary for accurate delineation of the tumor and other structures [4]. It is also used as a first step for medical image segmentation [9]. Therefore, accurate CT-MR synthesis model is needed.

A Generative Adversarial Network (GAN) is considered an interesting model for CT-MR image synthesis. It is the next-generation artificial intelligent approach that has shown promising results in image generation and image synthesis[10] [11] [12]. For image synthesis, the GAN model does not generate images from a random noise vector as input. In this case, the input is an image required to be mapped to another image domain. The GAN model needs to be fed with training datasets, which could either be paired or unpaired. Dealing with unpaired training datasets is difficult, because mapping between input and output does not exist. Treating the problem as unsupervised learning makes the task of developing a GAN model harder than when paired trained datasets exist. However, paired training datasets are not available for many tasks in practice. In addition, obtaining paired training datasets for several applications is relatively expensive and difficult [13]. Regarding CT-MR image synthesis, the acquisition of CT and MR images separately is time-consuming, costly, and a burden to the patient. Therefore, the available training datasets are mostly unpaired.

Another issue that faces CT-MR image synthesis using GAN model is the size of the training datasets, where majority of the available datasets are small. Given that the number of training images is critical in obtaining realistic images, transfer learning and data augmentation are

usually used to overcome the problem of small training datasets [14]. The data augmentation algorithm generates data by performing a series of transformations on the original data, including elastic transformations, pixel-level transformations, and other transformations [15]. GAN is a powerful model that has recently been used as a novel data augmentation technique to increase the size of the training dataset [8]. GAN-based data augmentation methods are more appropriate for medical image generation than traditional methods because color adjustment or rotation, for example, may alter the model's ability to distinguish medical images. Transfer learning is expressed through the use of pre-trained models, which are trained on a large dataset to solve a problem that is similar to the one that we aim to solve in the present study. The pre-trained model can be used as the starting point for the model on the second task of interest, where the model continues to train with the new dataset. The learned weights and biases will be copied from the pre-trained network to the target network [16].

Specifying the location and extent of tumors is critical for medical diagnosis, prognosis and surgical planning. Automatic tumor segmentation helps specialists in treatment planning and tumor measurements. Segmentation of the bony structure from MR and segmentation of soft tissues from CT are quite challenging. As a solution for bone structure segmentation and MR image availability, GAN is used to generate a realistic CT image from MR image and perform segmentation on the generated CT image.

Most previous work lack proper synthesis of fine tissue details with large variations of brain anatomies across subjects. This is especially evident with CT-MR image synthesis which still remains a challenge, and is usually hindered by the problem of small-size paired datasets (e.g. pix2pix model [17] and Conditional Adversarial Networks (cGAN) [18]). On the other hand, the state-of-the-art unpaired models, such as cycle generative adversarial networks (cycleGAN), have their limitations. CycleGAN models utilize the cycle consistency loss function to handle the small-size paired MR-CT datasets problem, but it may lead to mismatch of anatomical structures in the generated images [19][20]. Therefore, there is no guarantee that the input and generated images are structurally consistent. Figure 1 shows that the generated images are quite different from the ground truth images, especially in skull region in case of CT and in soft tissue area in case of MR.

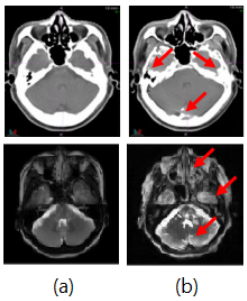

Figure 1: CycleGAN image synthesis showing both (a) ground truth, and (b) generated image, for CT (first row) and MR (second row), respectively. (Red arrows indicate regions with discrepancies)

In this paper, we develop a new Unsupervised Attention Guided Generative Adversarial Network (uagGAN) model for bidirectional MR-CT image synthesis. The proposed architecture takes into consideration both paired and unpaired image data. Pre-training is performed for fine-

tuning the network parameters with a paired dataset. Then unsupervised training follows with unpaired image data with an optimized loss function. Loss functions are known to give variable errors for the same prediction, and thus could have a considerable effect on the performance of the model. For this purpose, a Wasserstein GAN (WGAN) adversarial loss function is integrated with a set of non-adversarial loss functions to generate more realistic high-quality clinical images. WGAN employs the Earth Mover's distance that overcomes the mode collapse problem [21], and it is chosen after a comparison with the state-of-the-art loss functions. The main contributions of this research are as follows:

1. The paired-unpaired uagGAN model is initialized by supervised pre-training and then subsequently followed by unsupervised training for fine-tuning the medical image translation task.
2. Both high and low frequency components of the output images are captured by enhancing the adversarial loss function with an optimized combination of non-adversarial loss functions.
3. The proposed model is applied on real cases for bidirectional MR-to-CT and CT-to-MR translations, where knowledge transfer from a non-medical pre-trained model is used to tackle the problem of limited-size of paired MR-CT image data.
4. Performance is quantitatively evaluated by four different well-known image quality assessment metrics, and qualitatively through a perceptual study by three experienced radiologists.

The rest of the paper is organized as follows: Section 2 presents a review of related work of image synthesis in the medical imaging domain. Section 3 provides coverage of image-to-image translation using GAN models. Section 4 explains the proposed paired-unpaired unsupervised learning model with transfer learning. Section 5 discusses experimental results and performance. Section 6 contains the discussion. Section 7 is the conclusion and future work.

2. Related Works

Deep learning models have been developed for image-to-image translation with state-of-the-art results. Convolutional neural network (CNN) is a popular deep learning model for computer vision and medical imaging fields [22]. Xiang et al. [23] proposed a model for MR-to-CT synthesis using deep embedding CNN (DECNN). Li et al. [24] applied the recursive CNN with improved super-resolution algorithm to estimate the PET image from the MR image, which could increase the performance of the generated images without increasing the parameterized complexity.

CNN has introduced several models for image generation and translation to improve the modeling of nonlinear mapping from input to output and produce more realistic images [25]. The most prominent models among them are the GAN models. These models have achieved promising results in the field of image generation as they produce more realistic images even in unsupervised settings. In 2014, Goodfellow et al. [26] proposed the GAN model with an aim to generate new images from scratch. Consequently, the idea expanded to address the problem of image-to-image translation with more promising results. Several research works have been performed to improve the performance of the GAN model by modifying either the GAN architecture or its adversarial loss function, which enhanced the training process and generated more

realistic images. Emami et al. [27] proposed a GAN model to generate CT images from T1 MR images. They used a residual network for the generator network and CNN with five convolutional layers for the discriminator networks. They performed five-fold cross-validation to evaluate their model. They also compared their model with the CNN model. The results showed that their model outperformed the CNN model, the generated CT images preserved details better than CNN, and the abnormal regions in the generated MR images were well represented.

The GAN models have been recently used to perform medical image-to-image translation tasks. These models are classified as supervised and unsupervised models. For supervised models, pix2pix is a general-purpose model for image-to-image translation tasks. Nie et al. [6] utilized the pix2pix model with the gradient loss function added to the generator architecture for MR-to-CT translation. Due to patch-wise training, this proposed architecture has a limited modeling capacity, which rendered the invisibility of the training. Therefore, the auto-context model (ACM) could be used to train multiple GAN models one after another and enhance the results. The pix2pix model has been also utilized by Wolterink et al. [28] for the translation of low-dose CT images into their high-dose counterpart. It also has been used with Wasserstein distance and perceptual loss by Yang et al. [29] for CT denoising and 2T-to-1T MR translation [30]. Han et al. [31] incrementally incorporated a high-roughness bounding box into progressive growing GAN (PGGAN) and proposed a conditional PGGAN (CPGGAN) to place regions of interest (in this case, brain metastases) at desired positions/sizes on MR images. This model has been used to generate additional training data to address the small-sized training dataset issue. The results showed that the proposed model improved training robustness and increased 10% sensitivity in diagnosis. Cao et al. [32] proposed a framework for medical image generation called self-supervised collaborative learning. The authors presented an auto-encoder network that was used to obtain information about the target modality to generate any missing image modality. The authors also created a mask vector for the target modality to be used as a label. Chen et al. [33] proposed the Target-aware Generative Adversarial Network, a paired GAN model (TarGAN). TarGAN is a multi-modality image-to-image translation model that uses target area labels to improve target area generation quality. The TarGAN generator uses a proposed crossing loss function to translate the entire image and target area. Tang et al. [34] developed a GAN model for T1 MR-to-CT translation. For its generator, this GAN model employs a U-net network. The training dataset contains 27 rigidly registered brain cancer MR-CT pairs, whilst the testing dataset contains 10 pairs to evaluate the performance of the proposed model using mean absolute error. The clinical volumetric modulated arc therapy protocol was also used, followed by gamma analysis and a dose–volume histogram comparison on both generated and real CT images.

Several unsupervised models that were trained with unpaired input-target images have been developed, including CycleGAN [19]. Chartsias et al. [35] utilized cycleGAN to synthesize cardiac MR images from CT images with an unpaired dataset. Consequently, they applied the generated MR images in a segmentation task. The findings indicated that the generated MR images could be accurately used further in medical image segmentation, and these images improved the accuracy of the used segmentation algorithm by 16%. The authors recommended following the same approach of CT-to-MR synthesis for other body organs and with more training examples to improve the results. Hiasa et al. [36] also applied cycleGAN to unpaired head CT and MR images to synthesize CT images from MR images. They extended cycleGAN by adding gradient consistency loss to improve the accuracy. In addition, Jiang et al. [37] developed a model that started with unsupervised CT-to-MR tumor synthesis and then with semi-supervised tumor segmentation. They utilized cycleGAN with a new introduced loss called tumor-aware

loss. Zhang et al. [9] performed cycleGAN for synthesis of realistic-looking 3D images. The synthesized images were utilized to improve the volume segmentation algorithm. The generator of this GAN was trained with a shape consistency loss in addition to the cycleGAN loss functions, and more accurate results were generated. Furthermore, Wei et al. [38] used the cycleGAN model to generate a CT image from an MR image. The original MR and CT images were registered using traditional mono-modal image registration of the synthesis CT image and the original CT image. This fast MR-CT image registration method guides the thermal ablation of liver tumours. Experimental results from a real clinical dataset confirmed that the proposed method outperforms state-of-the-art methods with high registration accuracy and fast computing. Han et al.[39] combined noise-to-image GAN and image-to-image GAN to improve the efficacy of data augmentation for tumour detection. Their proposed model is a two-step GAN for generating brain MR images to be used in the training stage to handle a small-sized dataset. The results showed that the proposed two-step GAN-based data augmentation outperforms classic data augmentation.

Many other GAN models have been developed to translate medical images from one form to another. For instance, Calimeri et al. [40] proposed a new model of LapGAN to generate MR images of the human brain. They used quantitative and human-based evaluations to assess the effectiveness of the proposed method. Nie et al. [7] applied 3D Context-Aware GAN (CGAN) with ACM and used brain and pelvic datasets to test their proposed method. The 3D CGAN generated CT images from MR images, while the ACM refined the CT images. The results indicated that the performance of GAN improved when using the ACM. The authors considered the task of CT prediction only and claimed that their proposed model could be applied to other related tasks that include generative process. Zhao et al. [13] suggested Tub-GAN for multiple realistic-looking retinal image synthesis. Although they utilized small-sized samples of 10–20 images, they indicated that their model works very well. Furthermore, the authors claimed that the model has the ability to handle image synthesis from the same tubular structure. Dar et al. [41] offered conditional generative adversarial networks (cGAN) for multi-contrast MRI. Their proposed model for T1-to-T2 MR synthesis showed enhanced performance by using two types of losses: pixel-wise loss for registered images and cycle-consistency loss for unregistered images. The main idea behind cGAN was that the input data fed not only the generator but also the discriminator.

The GAN models have been employed for other medical applications. For example, Zhao et al. [42] proposed a cascaded GAN model for bony structure segmentation with deep supervision discriminator (Deep-supGAN). They generated CT images from MR images and then segmented the bony structures from both generated images. The combination of the MR and CT images provided a complete bony structure information needed for the segmentation task. Their model exhibited two blocks. The first block was for generating CT images from MR images, while the second block was for bony structure segmentation of the MR images and the generated CT images. The results showed that the generated CT images had clear structural details and the bony structure segmentation had more accurate results than the state-of-the-art models. Nema et al. [43] designed a residual cyclic unpaired encoder–decoder network (RescueNet) to segment an entire tumour in a brain MR image, followed by the core and enhanced region. RescueNet trains with an unpaired training dataset to eliminate the need for a paired dataset because preparing a large paired dataset is difficult. The proposed network was tested on BraTS 2015 and BraTS 2017 datasets, and its performance was evaluated through DICE and sensitivity measurements. The results showed that the proposed network outperforms existing methods of brain tumour

segmentation.

In contrast to the current medical image-to-image-translation models, which have been trained either on paired images or unpaired images, our model has been trained with both paired and unpaired datasets to handle the registration problem of paired dataset and to tackle the misalignment problem of the unpaired dataset. Moreover, our work focuses on addressing the size limitation of the training dataset by transferring knowledge of the non-medical pre-trained model to our medical model. Furthermore, we selected the appropriate combination of loss functions to capture the high and low frequency details of the generated images and generate fine structure images.

3. Unsupervised Image-to-Image Translation

The GAN models are a class of unsupervised machine learning models. The original GAN model consists of generative and discriminator neural networks that work with one against the other and become trained by using the minimax game theory. A generative model attempts to generate a new image from a random or a very-low-resolution image. Although the generative model is trained to fool the discriminator, the latter reviews the generated data to decide whether it belongs to the actual training dataset or not (0 for a fake image and 1 for a real image) [26]. The GAN model architecture is shown in Figure 2.

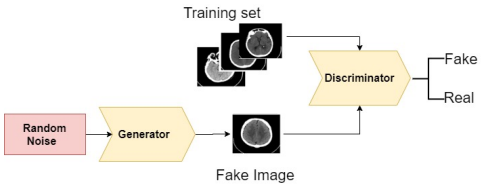

Figure 2: A typical generative adversarial network architecture.

The GAN model uses adversarial process to estimate the generative model by training its two models: generative model G and a discriminative model D. These models are fed with training dataset, which could be paired or unpaired. Paired training dataset consists of training examples x_i (data from first domain) and y_i (data from second domain) $i = [1, n]$, where the correspondence between x_i and y_i exists, unpaired training dataset consists of a source set $\{x_i\}$ $i = 1 \ldots n$ ($x_i \in$ X, {X: Source domain}) and a target set $\{y_j\}$ $j = 1 \ldots n$ ($y_j \in$ Y, {Y: Target Domain}), with no information provided as to which x_i matches which y_j, as shown in Figure 3.

The Conditional Adversarial Networks (cGAN) is an example of paired GAN models. The cGAN model is a general approach for many image-to-image translation tasks, because it could be used in a wide range of image domains, and it is considered the first paired GAN model used for image synthesis. This cGAN model is an extension of the original GAN, but with a change made to the discriminator input; the generator's input image is also provided to a discriminator [18]. The pix2pix model is an extension of cGAN' work. This model is a general solution to supervised image-to-image translation problems. It improves image translation output by changing the loss function of the generator network by adding L1 to the adversarial loss function. The generator of the pix2pix model translates the input images into target images by minimizing the adversarial loss and L1 loss functions. This model is very effective in image synthesis, has the potential to achieve reasonable results, and is widely applicable and easy to adopt [17].

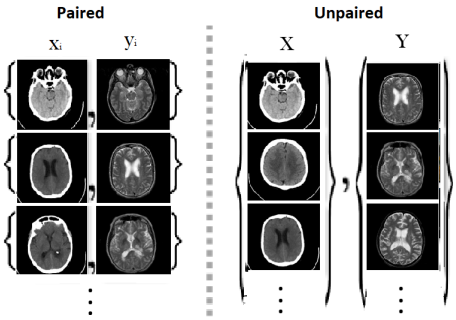

Figure 3: Paired and unpaired training datasets.

The uagGAN model is an example of unpaired GAN models [44]. The uagGAN model utilizes the attention unsupervised mechanism inspired by the significant role of human perception. It differs from other unsupervised image-to-image translation models. It focuses its attention on multiple objects within the image and alters the background in the case of a single object, leading to a more realistic translation compared with that in other recent relevant approaches. The uagGAN model follows the same architecture of the cycleGAN model. It adds two attention networks, AS and AT (source and target), to the cycleGAN model architecture. The uagGAN model's two generator is built with a special built-in attention mechanism. These two generators can generate attention masks by utilizing the attention mechanisms, M_x and M_y of image x and y, respectively. In addition to attention masks, they generate the content masks, R_x and R_y, of images x and y, respectively. It also employs element-wise products to apply the learned mask to the generated images and then uses the inverse mask to add the background. In other words, the uagGAN model locates the area that needed to be translated inside the image and then applies the appropriate translation to that location. In addition, a new loss function called pixel loss was introduced and used in addition to the cycle consistency loss and the GAN adversarial loss for enhanced model optimization. The pixel loss is described on Eq. (1).

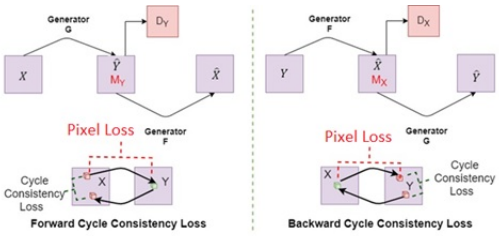

Figure 4: UagGAN model. M_x and M_y are the attention masks of x and y images, respectively.

$$L_{pixel}(G_{X\to Y}, G_{Y\to X}) = \|G_{X\to Y}(x) - x\|_1 + \|G_{Y\to X}(y) - y\|_1 \quad (1)$$

Where x represents the input image when translating from x to y, and y represents the input

image when translating from y to x. The generated image from x and y is represented by $G_{X\to Y}(x)$ and $G_{Y\to X}(y)$, respectively. Pixel loss is used with paired translation models such as pix2pix, but it is used for unpaired translation in the case of the uagGAN model.

4. Method

The uagGAN model performs bidirectional MR-CT image synthesis using paired–unpaired data with transfer learning. The first step in establishing the model is to identify the best model that deals with unpaired training images. We compare state-of-the-art GAN models, namely, cycleGAN, dualGAN, discoGAN, comboGAN,UNIT, and uagGAN. These models have two generators. Thus, they can be trained with an unpaired dataset. We follow the model architecture of Tripathy et al. [45] but we replace cycleGAN with the best unpaired model. The best model is then trained with paired and unpaired datasets. The paired and unpaired datasets are used to enhance the translation performance and solve the problem of having a limited paired dataset. During training with the paired dataset, we modify our model loss function. In this case, the model functions as a pix2pix model with modification to its loss function to improve model performance. The knowledge from non-medical pre-trained model has been transferred to our paired-unpaired model. Algorithm 1 presents the methodology of our paired-unpaired uagGAN model with transfer learning. Figure 5 illustrate the training steps of our proposed model.

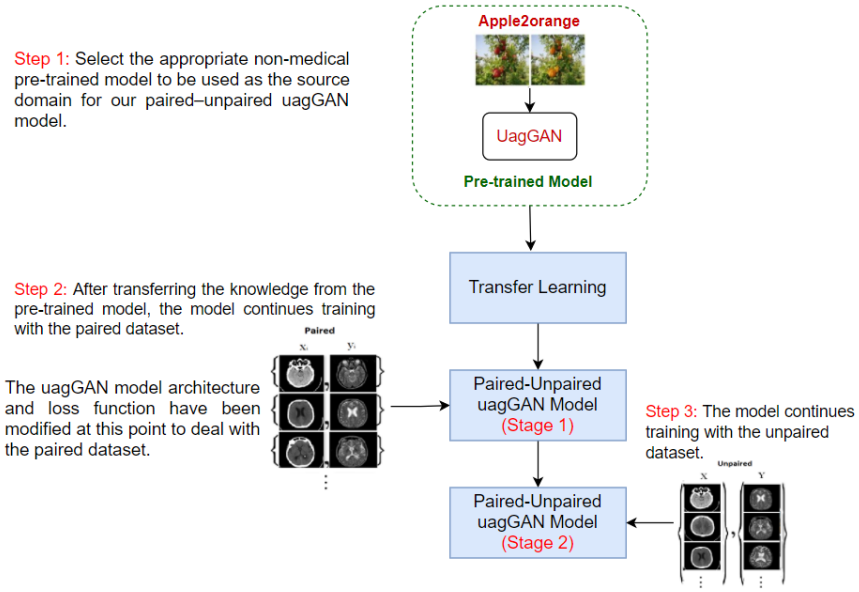

Figure 5: The training process of our proposed paired–unpaired uagGAN with transfer learning.

4.1. Multi-Loss Functions Combination

A well selected and optimized loss function may impact the stability of the GAN model training stability and performance. In a GAN model, the loss function is considered as an adversarial

loss that estimates the distance between the distribution of generated data and the distribution of real data. Thus, combining the GAN model adversarial loss function with traditional loss functions could be beneficial. For instance, the loss function of the pix2pix model, which is an extension of the conditional GAN model framework, is modified by adding L1 loss functions, leading to powerful results. In the case of paired training, we investigate the effect of the loss function by replacing the adversarial loss function with state-of-the-art adversarial losses. Additionally, we incorporate a new combination of non-adversarial loss functions. Several comparisons are performed in order to select the suitable combination of L1 and one of the following loss functions: structure, gradient, content-based, Kullback–Leibler divergence, and softmax.

4.2. Comparison with State-of-the-art Unpaired Models

Studying the GAN models that have been trained with unpaired datasets is important because paired training data are not available in certain tasks. One of the most popular GAN models dealing with the existence of unpaired training datasets is the CycleGAN model. This model is built on the basis of the pix2pix model, but it removes the paired input dataset and translates the trained images two times. The dualGAN, discoGAN, and Unsupervised Attention Guided GAN (uagGAN) models follow the same concept but use different loss functions.

The uagGAN model follows the same architecture of the cycleGAN model but with further improvement. It integrates an attention mechanism into unsupervised image synthesis that significantly improves the generated image quality and produces sharper images than the other methods. Additionally, this model utilizes the attention-guided discriminator to learn accurate maps and to focus on the attended content, reflecting where the discriminator looks before evaluating whether an image is real or fake.

According to Mejjati et al. [44], the uagGAN model has some limitations in robustness to shape changes between domains. However, this limitation occurs in some translation tasks but not all cases. For instance, this limitation will occur when ‘mapping zebras to lions’ because lions and zebras have differing shapes, whereas this problem will not occur when translating horse to zebra because the horse and zebra are similar in shape. In our case, the shapes of the CT and MR images are similar, and the distinction between MR and CT images is in the details within the image (e.g. MR images have more tissue details than CT). Therefore, the translation between MR and CT domains with no change in shape will not affect the robustness of the translation images.

In order to show that the uagGAN model produces sharper images than other unpaired methods, an image sharpness assessment test was carried out, where it is inspected using the gradient magnitude. This method is a non-reference metric that is based on a statistical analysis of local edge gradients. The gradient magnitude is appropriate for evaluating image sharpness, because the sharpness of an image is linked to the sharpness of its edge. The gradient of an image is computed by selecting an image region called the focus window with a size of m *times* n and then performing Eq. (2) [46].

Algorithm 1: Paired-unpaired augGAN model with transfer learning

Input: Paired training dataset P$x_n, y_n, n = 1 \ldots N$, Unpaired training dataset U$x_m, y_m, m = 1 \ldots M$, pre-trained model

Initialization: Number of training epochs N_{epoch}; Number of adversarial loss A; adv-lossf[A]=[WGAN-loss,WGANgp-loss, lsGAN-loss]; nonadv-loss[6]=[L1,$L_{content}$,L_{KLD},$L_{gradient}$,$L_{structure}$,$L_{softmax}$]; best-advloss[A]; uagGAN model generator uagGen=[G_y,G_x]; train_data

Output: Generated MR and CT images.

Select an appropriate non-medical pre-trained model to be used as the source domain for the paired–unpaired uagGAN model.

uagGen=[G_x]

Modify the adversarial loss function:

for $i = 1 \rightarrow A$ **do**

  adv-loss= adv-lossf[i]

  for $k = 1 \rightarrow N_{epoch}$ **do**

    for $n = 1 \rightarrow N$ **do**

      $\mathcal{L}_{L_1}(x_n, y_n) = \|y_n - G(x_n)\|_1$

      nonadv-loss1 +=adv-loss+$\mathcal{L}_{L_1}(x_n, y_n)$

      $\mathcal{L}_{content}(x_n, y_n) = \frac{1}{2} \sum_{i,j}(F_{i,j}(G(x_n)) - F_{i,j}(x_n))^2$

      nonadv-loss2 +=adv-loss+$\mathcal{L}_{content}(x_n, y_n)$+$\mathcal{L}_{L_1}(x_n, y_n)$

      $\mathcal{L}_{KLD}(x_n, y_n) = y_n * log(y_n/G(x_n))$

      nonadv-loss3 +=adv-loss+$\mathcal{L}_{KLD}(x_n, y_n)$+$\mathcal{L}_{L_1}(x_n, y_n)$

      $\mathcal{L}_{gradient}(x_n, y_n) = \sum_{i,j} \|y_{n_{i,j}} - y_{n_{i-1,j}}| - |G(x_n)_{i,j} - G(x_n)_{i-1,j}| + |y_{n_{i,j}} - y_{n_{i,j-1}}| - |G(x_n)_{i,j} - G(x_n)_{i,j-1}\|$

      nonadv-loss4 +=adv-loss+$\mathcal{L}_{gradient}(x_n, y_n)$+$\mathcal{L}_{L_1}(x_n, y_n)$

      $\mathcal{L}_{structure}(x_n, y_n) = \frac{1}{n} \sum_{x_n, y_n}[1 - SSIM(x_n, y_n)]$

      nonadv-loss5 +=adv-loss+$\mathcal{L}_{structure}(x_n, y_n)$+$\mathcal{L}_{L_1}(x_n, y_n)$

      $\mathcal{L}_{softmax}(x_n, y_n) = log(exp(-y_n) + exp(-x_n) + exp(-5))$

      nonadv-loss6 +=adv-loss+$\mathcal{L}_{softmax}(x_n, y_n)$+$\mathcal{L}_{L_1}(x_n, y_n)$

      θ_D[i] + = $logD(x_n, y_n) + log(1 - D(y_n, G(y_n)))$

  θ_D= max θ_D[i]

  final-advloss[i] = argmin ($nonadv - loss_f$), f=1...6

θ_G= min final-advloss[i]

train_data= Px_n, y_n

uagGen=[G_y,G_x]

train_data= Ux_m, y_m

Apply to unseen MR and CT images

return synthesized MR and CT images

$$SharpnessValue = Maximum\left[\left|(1/m)\sum_{j=1}^{m} a(i,j) - (1/m)\sum_{j=1}^{m} a(i+1,j)\right|\right]_{i=1....n-1} \quad (2)$$

4.3. Paired-Unpaired Training with Transfer learning

An unsupervised GAN model is used after changing the network architecture to account for paired data pre-training. Our model consists of cascade training, for which paired and unpaired image data are used. The model starts with the paired dataset by disconnecting one generator to act as a pix2pix model. At this stage, the adversarial loss function will be optimized. This leads to a more stable training and produces higher quality images.

To realise accurate bidirectional MR-CT image synthesis and tackle the problem of small-sized paired datasets, transfer learning is used from non-medical data. To select the appropriate non-medical pre-trained model for our paired–unpaired GAN model, we compare apple2orange, horse2zebra and lion2 tiger pre-trained models. The horse2zebra and lion2tiger pre-trained models are used for comparison because the conversion happens by adding lines to the translated image, similar to adding tissue details to CT images when translated from CT-to-MR. Apple2orange is also used for comparison because MR-CT images have a similar structure with apple-oranges images, that is, an oval or round shape structure. The selected pre-trained model has been used as the source domain for our paired–unpaired uagGAN model. After transferring the required knowledge from the non-medical pre-trained model, the uagGAN model continues to the next stage for training with unpaired data. As illustrated in Figure 6, the paired-unpaired uagGAN model employs a built-in attention mechanism that uses M_x for the generator G_{AB} and M_y for the generator G_{BA}. In the first stage, the model uses only one attention mask (M_x), while in the second stage, it uses both M_x and M_y.

We also use data augmentation as another method to deal with small datasets, and compare the results to those obtained using the transfer learning concept. WGAN has been used for data augmentation by generating new realistic medical images. The role of the GAN model in this case is to generate additional images. We trained WGAN twice: once with real MR images to generate new MR images and again with real CT images to generate new CT images. As a result, the training dataset size was increased by the WGAN-generated images. As a result, our paired-unpaired model will be trained on a larger dataset. The convergence time will be reduced as a result of transfer learning. Therefore, we compare the amount of time required to train the model using transfer learning and data augmentation.

5. Experimental Results

5.1. Dataset

We used two real MR-CT datasets during the experiments to demonstrate the capability of our model. For the paired dataset, we used the dataset provided by Han et al. [47], which includes

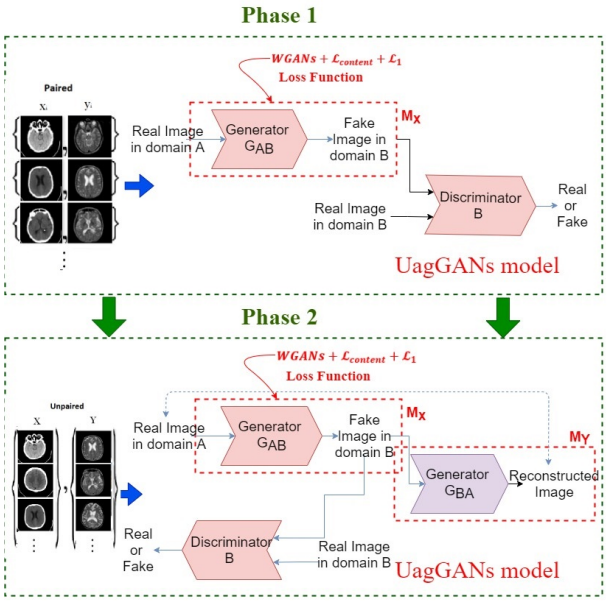

Figure 6: Our proposed paired–unpaired uagGAN model. In the first stage, one of the uagGAN model generators is used, and generator G_{BA} is disconnected. The uagGAN model acts as a pix2pix model that deals with paired dataset.

367 paired MR-CT brain images from 18 patients. To align each MR/CT pair, the authors used a mutual information rigid registration algorithm. They re-sampled the CT images to match the resolution of the MR images, and then corrected the MR images with the N3 bias field correction algorithm [48]. To standardize image intensities for all patients, the histogram-matched method [49] was applied to MR images.

For the unpaired dataset, we use our collected MR-CT brain images from the Radiology Department of the Jordan University Hospital (JUH) that has been collected between the period of April 2016 and December 2019. Both datasets contain normal and abnormal (i.e. contains tumors) MR and CT images[1].

The JUH dataset has been collected after receiving Institutional Review Board (IRB) approval of the hospital, and the consent of the patients to take their data. All procedures followed are consistent with the ethics of handling patients' data. Our dataset consists of the brain CT and MR images of 20 patients scanned for radiotherapy treatment planning for brain tumors. The dataset contains T2-MR and CT images for 20 patients aged between 26-71 years with mean-std equal to 47-14.07. The MR images of each patient were acquired with a 5.00mm T Siemens Verio 3T using a T2-weighted without contrast agent, 3 Fat sat pulses (FS), 2500-4000 TR, 20-30 TE, and 90/180 flip angle. The CT images were acquired with Siemens Somatom scanner with 2.46mGY.cm dose length, 130KV voltage, 113-327 mAs tube current, topogram acquisition protocol, 64 dual source, one projection, and slice thickness of 7.0mm. Smooth and sharp filters have been applied to the CT images. The MR scans have a resolution of 0.7×0.6×5 mm^3, while the CT scans have a resolution of 0.6×0.6×7 mm^3. There are a total of 840 2D axial image slices

[1]The JUH dataset can be downloaded from the IEEE DataPort at `https://dx.doi.org/10.21227/fe9x-qg64`.

in our final unpaired dataset (420 MR and 420 CT 2D axial image slices).

The final training dataset consists of 267 paired images from 13 patients and 770 unpaired images (385 MR images and 385 CT images) related to 17 patients (JUH dataset), while testing is carried out on a separate dataset of 135 images from 8 patients (5 from the paired dataset and 3 from the JUH dataset). The experiments are performed using 2D image slices that are extracted using the RadiAnt DICOM viewer software. The extracted images are transformed to png image data format with a resolution of 256×256 pixels. The model is trained using Google Colab cloud service, Tensorflow 2.0, python3 framework. All models are trained for 200 epochs with a batch size of 1. Peak-signal-to-noise-ratio (PSNR), Structural Similarity Index (SSIM), Universal Quality Index (UQI), and Visual Information Fidelity (VIF) have been used to quantitatively evaluate our model.

5.2. Comparison with State-of-the-art Unpaired Models

In the case of image synthesis, these metrics calculate the amount of distortion in the generated images. The simplest way to assess image quality is to calculate PSNR, but PSNR does not always correlate with human visual perception and image quality. Additional parameters were recommended to resolve the constraint of PSNR metrics, that is, SSIM. The performance of our model was evaluated on the above-mentioned datasets using PSNR[50] [51], SSIM[52], UQI [53], and VIF [54].

For the first set of experiments, we compare the state-of-the-art unpaired GAN models including cycleGAN, discoGAN, dualGAN, comboGAN, UNIT, and uagGAN models. Figure 7 shows the MR-to-CT and CT-to-MR translations. Table 1 shows the PSNR, SSIM, UQI, and VIF results of the bidirectional MR-CT synthesis (MR-to-CT and CT-to-MR translation) for the four selected unpaired models. All of these models have been trained by an end-to-end fashion for 200 epochs. For both MR-to-CT and CT-to-MR translations, the uagGAN model outperforms the other approaches. It results in the best score across the large majority of the chosen metrics: higher PSNR, SSIM, UQI, and VIF in the case of MR-to-CT translation; higher PSNR, SSIM and VIF in the case of CT-to-MR translation. The values of std are considered low for the majority of the metrics, indicating the stability of the results.

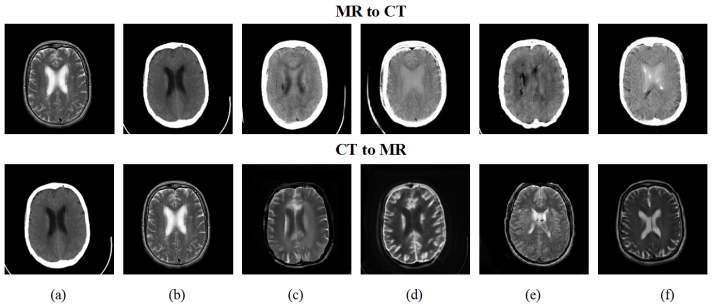

Figure 7: Bidirectional MR-CT translation results of unpaired models (a) input, (b) ground truth, (c) cycleGAN, (d) dualGAN, (e) discoGAN, and (f) uagGAN, respectively.

Table 2 shows the sharpness of the generated images of these unpaired models. The uagGAN framework results in an increased sharpness of the translated images as well as a notable improvement of the quantitative metrics compared to the other models.

Table 1: Image quality evaluation metrics for unpaired GAN models

GAN model	MR-to-CT							
	PSNR		SSIM		UQI		VIF	
	Mean	(Std)	Mean	(Std)	Mean	(Std)	Mean	(Std)
CycleGAN	25.404	1.648	0.554	0.052	0.779	0.021	0.361	0.070
DualGAN	21.880	2.607	0.541	0.053	0.688	0.038	0.234	0.103
DiscoGAN	24.675	2.136	0.553	0.042	0.743	0.034	0.067	0.017
ComboGAN	21.219	2.143	0.453	0.056	0.676	0.074	0.221	0.058
UNIT	25.572	1.821	0.573	0.040	0.718	0.046	0.370	0.069
UagGAN	**27.599**	**1.769**	**0.595**	**0.043**	**0.781**	**0.042**	**0.387**	**0.073**
	CT-to-MR							
CycleGAN	30.529	2.318	0.529	0.058	**0.607**	**0.083**	0.049	0.012
DualGAN	27.292	1.715	0.360	0.044	0.360	0.046	0.121	0.052
DiscoGAN	28.316	3.494	0.393	0.083	0.612	0.074	0.040	0.032
ComboGAN	30.320	2.301	0.382	0.060	0.462	0.072	0.042	0.039
UNIT	30.701	1.435	0.539	0.021	0.602	0.097	0.072	0.033
UagGAN	**31.049**	**1.306**	**0.542**	**0.051**	0.543	0.084	**0.178**	**0.029**

Table 2: The sharpness of the generated images of unpaired GAN models

	UagGAN	CycleGAN	DualGAN	DiscoGAN	ComboGAN	UNIT
MR	**0.925**	0.915	0.907	0.918	0.902	0.911
CT	**1.087**	1.077	1.068	1.015	1.017	1.063

5.3. Loss Functions Performance Analysis

The second set of experiments determines the best loss function to be used in the cGAN paired model. The modified model will then be used as the first phase of our paired–unpaired uagGAN model. We start with replacing the cGAN and pix2pix (cGAN $\| \mathcal{L}_{L1}$) adversarial loss with the adversarial loss of the state-of-the-art GAN models, namely, WGAN, WGAN-GP and lsGAN. We want to show which adversarial loss is the most suitable for cGAN and pix2pix architecture. Table 3 shows the PSNR, SSIM, UQI, and VIF results of a different cGAN model evaluation on MR-CT dataset. We find that adding L1 loss function to the cGAN model produces better performance, regardless of the used adversarial loss. The results show that {WGAN $\| \mathcal{L}_{L1}$} and {lsGAN $\| \mathcal{L}_{L1}$} are among the best adversarial loss functions, as they have the highest mean value and lowest std value for all evaluation metrics.

We are going further in our experiments by adding non adversarial loss functions to the cGAN and the pix2pix models. Table 4 shows the image evaluation metrics (PSNR, SSIM, UQI, and VIF) results after adding non-adversiral loss functions to the cGAN and pix2pix adversarial loss. The results show that adding L_1 to any of the used non-adversarial loss functions gives better results. For instance, SSIM loss preserves contrast in high-frequency regions. On the other hand,

Table 3: Image quality evaluation metrics for paired GAN models

GAN model	PSNR		SSIM		UQI		VIF	
	Mean	(Std)	Mean	(Std)	Mean	(Std)	Mean	(Std)
cGAN	30.234	3.948	0.668	0.067	0.854	0.045	0.573	0.123
WGAN	29.047	2.364	0.626	0.035	0.759	0.020	0.508	0.088
WGAN-GP	27.054	5.522	0.619	0.085	0.841	0.039	0.523	0.119
lsGAN	28.096	3.447	0.651	0.064	0.838	0.046	0.514	0.119
	adding $\mathcal{L}_{L1}$ (pix2pix)							
cGAN || $\mathcal{L}_{L1}$ (pix2pix)	31.159	2.929	0.697	0.056	0.877	0.037	0.603	0.073
WGAN || $\mathcal{L}_{L1}$	**31.560**	**2.629**	**0.724**	**0.047**	**0.887**	**0.028**	**0.633**	**0.077**
WGAN-GP || $\mathcal{L}_{L1}$	26.776	5.225	0.627	0.081	0.838	0.041	0.535	0.105
lsGAN || $\mathcal{L}_{L1}$	**32.913**	**2.571**	**0.739**	**0.049**	**0.896**	**0.025**	**0.687**	**0.072**

Bold indicates the highest two metric scores.

L_1 maintains low-frequency. This indicates that the combination of SSIM and L_1 loss functions produces better results than using SSIM alone. In addition, adding a non-adversarial loss function to the pix2pix model produces better results than using L_1 alone.

Table 4: Image quality evaluation metrics for cGAN and pix2pix after adding non-adversarial loss function to the generator network.

GAN model	PSNR		SSIM		UQI		VIF	
	Mean	(Std)	Mean	(Std)	Mean	(Std)	Mean	(Std)
cGAN	30.234	3.948	0.668	0.067	0.854	0.045	0.573	0.123
$cGAN + \mathcal{L}_{gradient}$	29.676	3.351	0.668	0.079	00.851	0.049	0.567	0.143
$cGAN + \mathcal{L}_{KLD}$	27.809	2.726	0.640	0.065	0.836	0.042	0.497	0.085
$cGAN + \mathcal{L}_{softmax}$	28.815	3.337	0.648	0.070	0.833	0.055	0.536	0.137
$cGAN + \mathcal{L}_{content}$	30.904	2.872	0.641	0.063	0.893	0.054	0.657	0.093
$cGAN + \mathcal{L}_{structural}$	29.007	2.991	0.642	0.061	0.871	0.055	0.588	0.097
	cGAN+ $\mathcal{L}_{L1}$ (pix2pix)							
pix2pix	31.159	2.929	0.697	0.056	0.877	0.037	0.603	0.073
pix2pix+$\mathcal{L}_{gradient}$	31.591	3.774	0.697	0.068	0.879	0.044	0.637	0.095
pix2pix+$\mathcal{L}_{KLD}$	31.047	2.708	0.700	0.056	0.874	0.032	0.627	0.098
pix2pix+ $\mathcal{L}_{softmax}$	30.745	3.407	0.689	0.059	0.866	0.044	0.671	0.625
pix2pix+$\mathcal{L}_{content}$	31.332	2.859	0.789	0.061	0.891	0.036	0.623	0.093
pix2pix+$\mathcal{L}_{structural}$	30.497	2.842	0.715	0.059	0.882	0.029	0.612	0.874

We proceed by adding the traditional loss functions apart from the L1 loss function to the lsGAN and WGAN models. One loss function is added at a time. Then, we compare the results to determine the best one. Table 5 shows the PSNR, SSIM, UQI, and VIF results for these models. The results show that the best loss function for the paired GAN model is {WGAN || $\mathcal{L}_{content}$ || $\mathcal{L}_1$}. It has the best values for all evaluation metrics. Output samples are blurred and

lack a high-frequency structure that uses L1 loss function on its own, while content loss offers the training stability required for convergence. Content loss is used to detect the features in images, which allows the loss function to know what features are in the target ground truth image rather than merely comparing pixel differences. This process allows the model being trained with this loss function to produce a much finer detail of the generated features and outputs.

Table 5: Image quality evaluation metrics after multi-loss functions combination

GAN model	lsGAN							
	PSNR		SSIM		UQI		VIF	
	Mean	(Std)	Mean	(Std)	Mean	(Std)	Mean	(Std)
lsGAN $\|\|$ $\mathcal{L}_1$	32.913	2.571	0.739	0.049	0.896	0.025	0.687	0.072
lsGAN $\|\|$ $\mathcal{L}_{gradient}$ $\|\|$ $\mathcal{L}_1$	33.005	2.388	0.739	0.046	0.892	0.022	0.649	0.076
lsGAN $\|\|$ $\mathcal{L}_{KLD}$ $\|\|$ $\mathcal{L}_1$	33.005	3.357	0.727	0.049	0.885	0.036	0.668	0.092
lsGAN $\|\|$ $\mathcal{L}_{softmax}$ $\|\|$ $\mathcal{L}_1$	31.596	3.115	0.732	0.049	0.888	0.034	0.642	0.084
lsGAN $\|\|$ $\mathcal{L}_{content}$ $\|\|$ $\mathcal{L}_1$	33.014	2.532	0.742	0.039	0.901	0.029	0.696	0.079
lsGAN $\|\|\mathcal{L}_{structural}$ $\|\|$ $\mathcal{L}_1$	32.024	3.235	0.733	0.050	0.887	0.033	0.667	0.101
	WGAN							
WGAN $\|\|\mathcal{L}_1$	31.560	2.629	0.724	0.047	0.887	0.028	0.633	0.077
WGAN $\|\|\mathcal{L}_{gradient}$ $\|\|\mathcal{L}_1$	32.429	2.281	0.736	0.038	0.894	0.023	0.671	0.066
WGAN $\|\|$ $\mathcal{L}_{KLD}$ $\|\|$ $\mathcal{L}_1$	32.114	2.989	0.734	0.046	0.889	0.028	0.654	0.091
WGAN $\|\|$ $\mathcal{L}_{softmax}$ $\|\|$ $\mathcal{L}_1$	31.376	2.969	0.726	0.048	0.882	0.032	0.635	0.085
WGAN $\|\|$ $\mathcal{L}_{content}$ $\|\|$ $\mathcal{L}_1$	**33.225**	**2.280**	**0.756**	**0.038**	**0.904**	**0.022**	**0.700**	**0.089**
WGAN $\|\|$ $\mathcal{L}_{structural}$ $\|\|$ $\mathcal{L}_1$	32.233	3.496	0.725	0.040	0.885	0.037	0.672	0.067

5.4. Paired-Unpaired Training

The uagGAN model is the best unpaired model according to the comparison with the state-of-the-art unpaired model. To add supervision behavior to the uagGAN model, we modify its architecture to train it with both paired and unpaid datasets. This model starts training with paired dataset, consequently inducing the uagGAN model to act as pix2pix in this stage, but the loss function is changed to {WGAN $\|$ $\mathcal{L}_{content}$ $\|$ $\mathcal{L}_1$}. The model continues training with unpaired dataset. Therefore, we obtain a new uagGAN model which deals with both paired and unpaired datasets. We compare the paired–unpaired cycleGAN to the paired–unpaired uagGAN model to ensure that using the uagGAN is appropriate as a paired–unpaired model. Both models have been trained in our dataset for 200 epochs. The results of MR-to-CT and CT-to-MR are shown in Figure 8. Table 6 shows the PSNR, SSIM, UQI, and VIF results of uagGAN and cycleGAN models trained on paired–unpaired dataset.

The results show that our proposed paired–unpaired uagGAN model outperforms the paired–unpaired cycleGAN model because it has the highest and stable evaluation values for all the evaluation metrics. The results also show that synthesised images have higher quality with paired and unpaired data than when using a single type of data on its own. Owing to the limited size of the training dataset, the results remain unreliable. Several details do not appear in translated images in both MR-to-CT and CT-to-MR. The generated CT and MR images lose anatomical information in areas of soft brain tissue and contain artefacts in areas with bony structures.

Table 6: Image quality evaluation metrics for paired–unpaired cycleGAN and paired–unpaired uagGAN models

	MR-to-CT							
Model	PSNR		SSIM		UQI		VIF	
	Mean	(Std)	Mean	(Std)	Mean	(Std)	Mean	(Std)
	Unpaired Results							
cycleGAN	25.404	1.648	0.554	0.052	0.779	0.021	0.361	0.070
uagGAN	27.599	1.769	0.595	0.043	0.781	0.042	0.387	0.073
	Paired-Unpaired Results							
cycleGAN	32.928	4.023	0.692	0.068	0.812	0.091	0.411	0.091
uagGAN	**34.786**	**2.362**	**0.739**	**0.060**	**0.813**	**0.053**	**0.482**	**0.097**
	CT-to-MR							
	Unpaired Results							
cycleGAN	30.529	2.318	0.529	0.058	0.607	0.083	0.049	0.012
uagGAN	31.049	1.306	0.542	0.051	0.543	0.084	0.178	0.029
	Paired-Unpaired Results							
cycleGAN	30.121	1.769	0.512	0.049	0.535	0.052	0.148	0.051
uagGAN	**31.821**	**2.451**	**0.603**	**0.065**	**0.631**	**0.106**	**0.187**	**0.094**

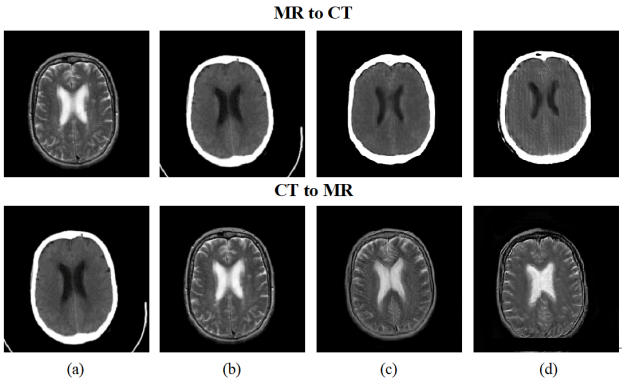

Figure 8: MR-to-CT translation results of paired–unpaired cycleGAN and uagGAN models (a) input, (b) ground truth, (c) cycleGAN, and (d) uagGAN, respectively.

Any GAN model must be trained with a large number of images to generate realistic images, as the training stability in terms of the generative loss increases when using more training data. We perform an experiment to show how the number of images affects the quality of the generated images (Figure 9). The results show that the PSNR value increases with the increase in the number of images. In addition, using 600 training images or more will produce good generated images but still need improvement. Therefore, in the next set of experiments, transfer learning is used to deal with the problem of a small-sized dataset."

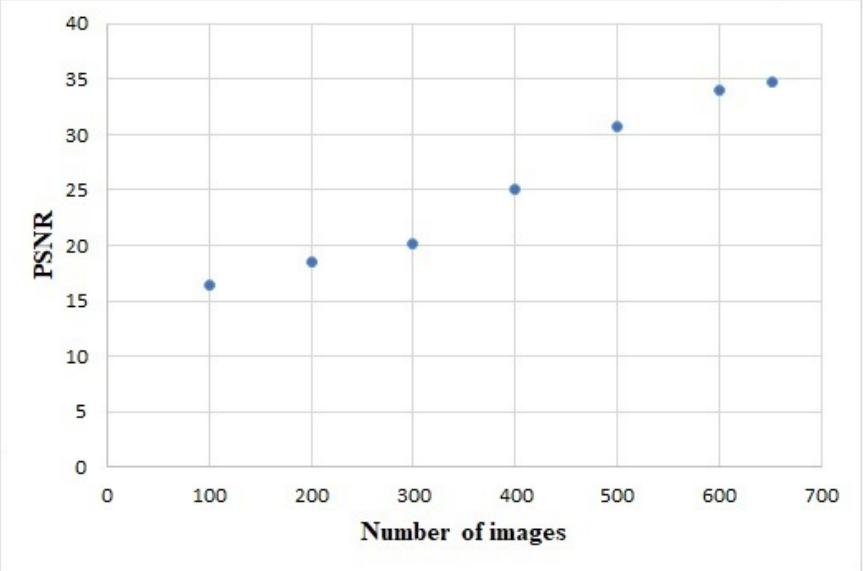

Figure 9: The impact of the number of images on image generation quality.

5.5. *Effect of Transfer Learning*

The size of paired and unpaired datasets remains small. Therefore, to enhance the quality of this paired–unpaired model, we proceed by using the transfer learning concept. We use three uagGAN pre-trained models that were trained with different datasets, as a source model for our proposed model, which are: Apple2Orange containing 996 apple images and 1020 orange images, Horse2Zebra with 939 horse and 1177 zebra images, and Lion2Tiger dataset with 916 lion images and 854 tiger images. Figure 10 shows the final results of MR-to-CT and CT-to-MR syntheses for our paired–unpaired uagGAN model with transfer learning. Table 7 shows the PSNR, SSIM, UQI, and VIF for our final model with transfer learning. The results show that the performance of our proposed model has improved for all evaluation metrics when transfer learning has been used, regardless of the type of pre-trained model used. It also shows that the best pre-trained model is the apple2orange pre-trained model for both MR-to-CT and CT-to-MR translation.

To demonstrate the robustness of the proposed model with abnormal cases, Figure 11 shows the result of image synthesis for four different cases with tumors of varying sizes. The results are

Table 7: Image quality evaluation metrics for paired-unpaired uagGAN model with transfer learning

	MR-to-CT							
Model	PSNR		SSIM		UQI		VIF	
	Mean	(Std)	Mean	(Std)	Mean	(Std)	Mean	(Std)
	Training from Scratch							
Paired-Unpaired UagGAN	34.786	2.362	0.739	0.060	0.813	0.053	0.482	0.097
	Transfer Learning							
Pre-trained Dataset								
Apple2orange	**68.105**	**2.413**	**0.953**	**0.023**	**0.892**	**0.008**	**0.899**	**0.047**
Horse2zebra	57.681	2.968	0.888	0.085	0.817	0.082	0.826	0.098
Lion2tiger	38.438	1.079	0.856	0.008	0.756	0.040	0.358	0.015
	CT-to-MR							
	Training from Scratch							
Paired-Unpaired uagGAN	31.821	2.451	0.603	0.065	0.631	0.106	0.187	0.094
	Transfer Learning							
Pre-trained Dataset								
Apple2orange	**57.969**	**4.803**	**0.943**	**0.028**	**0.958**	**0.035**	**0.699**	**0.083**
Horse2zebra	46.890	2.504	0.855	0.024	0.861	0.023	0.556	0.044
Lion2tiger	41.301	1.449	0.858	0.015	0.898	0.013	0.351	0.026

significant even with the existence of tumors in both MR and CT images. The proposed model can capture the tumor area inside 2D brain slices in both MR and CT images. The tumors' shape and size appears spatially correct in translated MR and CT images. Many details appear correctly despite the appearance of a small amount of artefacts in the generated MR images. The generated images have a similar global structure as the corresponding ground truth images. Compared to other rival image-to-image models, the proposed model can be viewed as robust method when translating images with fine details and complex structures such as MR images.

5.6. Data Augmentation

We also used data augmentation to increase the size of the dataset in order to improve the quality of the images generated by our paired-unpaired augGAN model. Table 8 shows the comparison between transfer learning and data augmentation. According to the results, both methods improved accuracy. It also demonstrates that transfer learning is best suited for dealing with small-sized problems with the highest evaluation metrics. GAN training takes a significant amount of time, especially for unpaired GAN models. The time required by each unpaired GAN model is shown in the Table 9. The results show that the time required to train these models is close, though cycleGAN requires less time. The model requires 14 hours to train with data augmentation and 8 hours to train with transfer learning. As a result, using transfer learning leads to better results in less time.

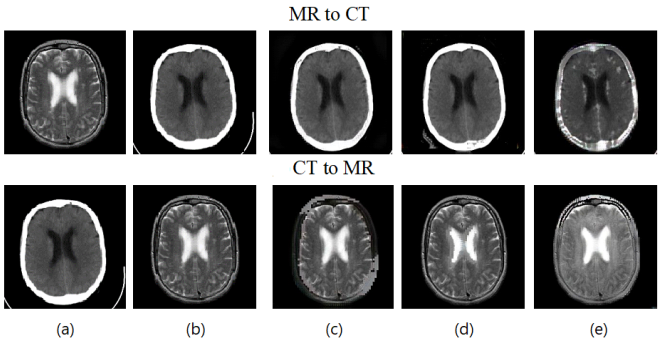

Figure 10: Bidirectional MR-CT translation results of paired–unpaired uagGAN model with transfer learning (a) input, (b) ground truth, (c) apple2orange, (d) horse2zebra, and (e) lion2tiger, respectively.

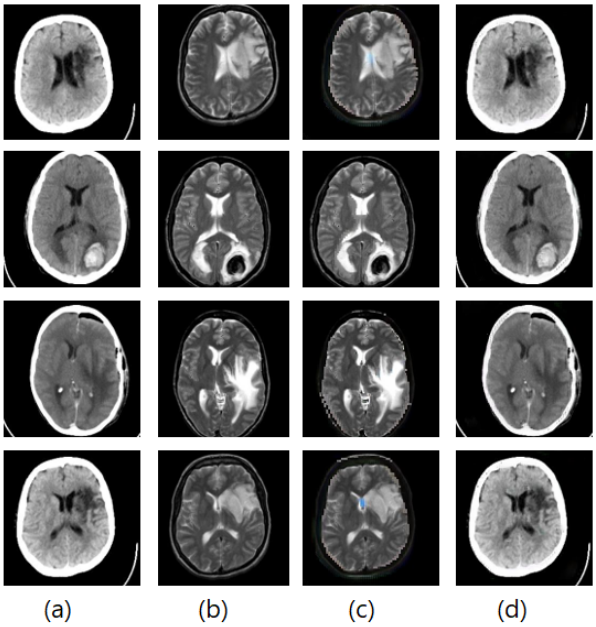

Figure 11: MR-CT bidirectional translation results of paired–unpaired uagGAN model with transfer learning for abnormal cases, where (a) and (c) are real images, and (b) and (d) are synthesized images.

Table 8: Comparison between transfer learning and data augmentation

	MR-to-CT							
Method	PSNR		SSIM		UQI		VIF	
	Mean	(Std)	Mean	(Std)	Mean	(Std)	Mean	(Std)
Method								
Paired-Unpaired UagGAN	34.786	2.362	0.739	0.060	0.813	0.053	0.482	0.097
Transfer learning	**68.105**	**2.413**	**0.953**	**0.023**	**0.892**	**0.008**	**0.899**	**0.047**
Data Augmentation	53.062	3.063	0.882	0.009	0.827	0.010	0.673	0.059
	CT-to-MR							
Paired-Unpaired uagGAN	31.821	2.451	0.603	0.065	0.631	0.106	0.187	0.094
Transfer earning	**57.969**	**4.803**	**0.943**	**0.028**	**0.958**	**0.035**	**0.699**	**0.083**
Data Augmentation	47.816	2.784	0.799	0.025	0.753	0.102	0.537	0.087

Table 9: Training time for unpaired GAN models

Model	Time (minutes)
UagGAN	487.44
CycleGAN	484.19
DualGAN	506.16
DiscoGAN	506.24
ComboGAN	491.36
UNIT	489.27

5.7. Perceptual Study and Validation

To evaluate the accuracy of our translated MR and CT images, we present three experienced radiologists from JUH a group of images containing the translated and ground truth images, which appear in a randomized order. This study adopts our final model, the uagGAN model, and the paired–unpaired uagGAN model for MR and CT generated images. Each radiologist was presented 60 translated images from each model (The above-mentioned three models) in addition to the 60 ground truth images. The final number of images presented to the radiologists is 240 MR images and 240 CT images with a resolution of 256×256 pixels. The main purpose of this study is to demonstrate the importance of working with paired and unpaired datasets to train the GAN model instead of training the model with an unpaired dataset only. Another purpose is to evaluate the images generated by our paired–unpaired uagGAN with the transfer learning model, and compare with ground truth images. Radiologists are asked to identify the ground truth images and use a four-point scoring method (with '4' denoting the most realistic image) to rate the accuracy of each image.

Table 10 presents the results of the perceptual study evaluated by radiologists for MR-to-CT and CT-to-MR translations. The final column of this table shows the percentage of images classified as real by the radiologists over the total number of images. In both MR-to-CT trans-

lations, using paired and unpaired datasets to train the uagGAN model outperforms the use of unpaired dataset alone rated with a mean score of 0.91 for the paired–unpaired uagGAN model than 0.25 achieved by the unpaired uagGAN model in the case of MR-to-CT translation. The performance of paired–unpaired uagGAN model is also reflected in the CT-to-MR translation, where this model achieves a mean score of 1.08 compared with 0.39. Additionally, 93.89% of the generated CT images and 68.89% of the generated MR images by our model successfully convince the radiologists that they are ground truth images from a real scanner.

Table 10: Results of perceptual study

	MR-to-CT		
Model	Mean	Std	*Real%*
Unpaired uagGAN	0.25	0.14	18.89%
Paired-unpaired uagGAN	0.91	0.30	43.75%
Our model*	3.08	0.19	93.89%
Ground truth	3.24	0.21	95.56%
	CT-to-MR		
Model	Mean	Std	*Real%*
Unpaired uagGAN	0.39	0.08	25.29%
Paired-unpaired uagGAN	1.08	0.09	40.11%
Our model*	1.64	0.86	68.89%
Ground truth	2.49	0.89	80.95%

* Paired-unpaired uagGAN with transfer learning

We compute the Lin's concordance correlation coefficients (CCC) for the results of perceptual study, which quantifies the agreement between the ground truth images and the generated images. Table 11 presents the Pearson's correlation coefficient (ρ_c) and accuracy ($C\beta$) for both MR-to-CT and CT-to-MR translation. The results indicate high or medium correlation between ground truth and the generated images for all radiologists.

Table 11: The CCC results of perceptual study

GAN model	MR-to-CT		CT-to-MR	
	(ρ_c)	($C\beta$)	(ρ_c)	($C\beta$)
Radiologist 1	0.479	0.993	0.663	0.933
Radiologist 2	0.453	0.933	0.636	0.929
Radiologist 3	0.758	0.986	0.592	0.946

6. Discussion

This work proposes a paired–unpaired uagGAN model that has been trained with small-sized paired and unpaired datasets. A limited number of paired dataset images were integrated

in a semi-supervised cascaded procedure with unpaired dataset images to train our model. This procedure overcomes the unpaired training context misalignment problem and alleviates the generation of blurred images during paired data training.

For this purpose, the best combination of loss functions is optimized on a paired dataset for fine-tuning network parameters to generate realistic MR and CT brain images. This combination is used to handle the low and high frequency components of target images. In the supervised phase of the GAN model, the adversarial loss function is chosen based on procedure that automatically selects the best loss function. The weight parameters of both generator and discriminator θ_G, θ_D are assessed and then the best combination of loss functions are returned. The WGAN adversarial loss gave best performance among other loss functions. Additionally, since the L1 loss function can handle low-frequency components of the image, and the content loss function deals with the high-frequency image components, thus the combination of these two functions can further enhance the WGAN adversarial loss performance. Instead of comparing pixel differences, the content loss function can give focus to shape and structure features inside the ground truth image, and works on generating more realistic context details. Also, using a pre-trained network knowledge decreases the convergence time and improves image quality significantly, especially when target data is limited.

The definition of the adversarial loss function is a critical aspect in the design of GAN models because it affects the GAN's performance and the quality of the results. The WGAN adversarial loss function is selected according to comparison with the state-of-the-art adversarial loss functions (cGAN, WGAN, WGAN-GP, lsGAN adversarial loss functions). One of the most important characteristics of WGAN is its ability to continuously estimate the Earth Mover's distance by training the discriminator network to achieve the best status. Therefore, using the WGAN adversarial loss function results in stable training, avoiding the mode collapse problem, and improving the GAN model's performance [55] [21].

Preparing a large-size paired dataset is difficult and expensive in the medical domain, therefore an unpaired datasets can assist in network training. In the unsupervised phase of our GAN model, a uagGAN model is used for training the unpaired dataset. The proposed uagGAN method utilizes the capability of the cycleGAN model in generating images from unpaired data, in addition to the use of pixel and attention losses. By utilizing a built-in attention mechanism, the model attention-guided generators produce attention masks for high-quality target images. Image sharpness is also inspected using the gradient magnitude to measure the capability of generated fine tissue details of the unpaired models. Results show sharper images as compared to other methods.

GAN model needs to be trained with a large dataset to generate realistic images. Therefore, transfer learning from a non-medical data source is used to handle the small-sized paired and unpaired datasets. This would shorten the convergence time and minimize the effort required for collecting a suitable number of medical images to train the GAN model. The optimization of the uagGAN loss function and the use of transfer learning leads to homogeneous and realistic global structures and fine texture details. The results indicate that our synthetic model can estimate structures efficiently inside complicated 2D brain slices, such as soft brain vessels, and bones.

Clinical significance was assessed by a perceptual study. Three experienced radiologists performed qualitative evaluation and assessed the reliability of the generated MR and CT images. The quality of the MR and CT images generated by the augGAN model was rated near to ground

truth. A subset of the generated images were identified as real (source) images by the participating radiologists in the study. The reported qualitative results demonstrate the fidelity of the generated images by our model. Moreover, the results were statistically significant to clinicians, and may serve as a useful application tool in real-time actual MR and CT scanning scenarios (e.g. confirming the location of a tumor before surgery when using a single imaging modality).

Despite qualitative and quantitative experiments, the quality of the MR translated images can be improved as some of the generated images may contain artifacts in the skull region. This probably happens due to the model attempting to provide much more soft tissue details to the input CT images to generate the corresponding MR images, our model may incorrectly add more details to the bone area too. Adding a feature map layer in the MR image generation may assist with the poorly translated or mistranslated patterns. Additionally, this work is constrained to a single type of MR images, and performing the bidirectional image translation in 3D volumes can provide radiologists with better structural information and richer anatomical details.

7. Conclusions and Future Works

The paired–unpaired uagGAN model is a new approach for medical image translation tasks. The model alleviates the rigid registration task of training using small-size paired data and handles the context misalignment problem of unpaired datasets. In addition, a new combination of non-adversarial loss functions is incorporated to enhance model consistency and generate sharper images with higher soft-tissue contrast. Furthermore, knowledge transfer from a non-medical pre-trained model improved the proposed uagGAN model and achieved better image translation performance. The experimental results show efficient model performance in MR-to-CT translation. However, further image enhancement might be required in the case of CT-to-MR translation. Future works will be deal with 3D multi-channel volumes. We plan to generalize the model to accommodate other types of MR contrast images, such as T1 and proton density. Currently, the proposed model is being assessed for automatic tumour segmentation from bidirectional MR-CT generated images.

Acknowledgment

The authors would like to thank Ashjan Alshakkah, Hanin Kayed and Hamsa Abdulmunem from the Department of Radiology – Jordan University Hospital for assisting with the perceptual study.

References

[1] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE J BIOMED HEALTH*, vol. 21, no. 1, pp. 31–40, 2016.

[2] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *ISBI*. IEEE, 2015, pp. 294–297.

[3] M. Forouzanfar, N. Forghani, and M. Teshnehlab, "Parameter optimization of improved fuzzy c-means clustering algorithm for brain mr image segmentation," *ENG APPL ARTIF INTEL*, vol. 23, no. 2, pp. 160–168, 2010.

[4] S. Roy, A. Carass, A. Jog, J. L. Prince, and J. Lee, "Mr to ct registration of brains using image synthesis," in *Medical Imaging 2014: Image Processing*, vol. 9034. SPIE, 2014, p. 903419.

[5] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Deep mr to ct synthesis using unpaired data," in *SASHIMI*. Springer, 2017, pp. 14–23.
[6] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed Eng*, vol. 65, no. 12, pp. 2720–2730, 2018.
[7] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *Med Image Comput Comput Assist Interv*. Springer, 2017, pp. 417–425.
[8] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.
[9] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network," in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, 2018, pp. 9242–9251.
[10] B. Yu, Y. Wang, L. Wang, D. Shen, and L. Zhou, "Medical image synthesis via deep learning," *Deep Learning in Medical Image Analysis*, pp. 23–44, 2020.
[11] B. Xin, Y. Hu, Y. Zheng, and H. Liao, "Multi-modality generative adversarial networks with tumor consistency loss for brain mr image synthesis," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1803–1807.
[12] C. Wang, G. Yang, G. Papanastasiou, S. A. Tsaftaris, D. E. Newby, C. Gray, G. Macnaught, and T. J. MacGillivray, "Dicyc: Gan-based deformation invariant cross-domain information fusion for medical image synthesis," *Information Fusion*, vol. 67, pp. 147–160, 2020.
[13] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Med. Image Anal.*, vol. 49, pp. 14–26, 2018.
[14] M. Milicevic, I. Obradovic, K. Zubrinic, and T. Sjekavica, "Data augmentation and transfer learning for limited dataset ship classification," *Wseas Trans. Syst. Control*, vol. 13, pp. 460–465, 2018.
[15] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data augmentation for brain-tumor segmentation: a review," *Frontiers in computational neuroscience*, vol. 13, p. 83, 2019.
[16] Y. Wang, C. Wu, L. Herranz, J. van de Weijer, A. Gonzalez-Garcia, and B. Raducanu, "Transferring gans: generating images from limited data," in *Comput Vis ECCV*, 2018, pp. 218–234.
[17] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, 2017, pp. 1125–1134.
[18] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
[19] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
[20] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, J. L. Prince, and Z. Xu, "Unsupervised mr-to-ct synthesis using structure-constrained cyclegan," *IEEE transactions on medical imaging*, vol. 39, no. 12, pp. 4249–4261, 2020.
[21] A. Alotaibi, "Deep generative adversarial networks for image-to-image translation: A review," *Symmetry*, vol. 12, no. 10, p. 1705, 2020.
[22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3DV 2016*. IEEE, 2016, pp. 565–571.
[23] L. Xiang, Q. Wang, D. Nie, L. Zhang, X. Jin, Y. Qiao, and D. Shen, "Deep embedding convolutional neural network for synthesizing ct image from t1-weighted mr image," *Med. Image Anal*, vol. 47, pp. 31–44, 2018.
[24] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, and S. Ji, "Deep learning based imaging data completion for improved brain disease diagnosis," in *Med Image Comput Comput Assist Interv*. Springer, 2014, pp. 305–312.
[25] S. Barua, S. M. Erfani, and J. Bailey, "Fcc-gan: A fully connected and convolutional net architecture for gans," *arXiv preprint arXiv:1905.02417*, 2019.
[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Adv Neural Inf Process Syst*, 2014, pp. 2672–2680.
[27] H. Emami, M. Dong, S. P. Nejad-Davarani, and C. K. Glide-Hurst, "Generating synthetic cts from magnetic resonance images using generative adversarial networks," *Med Phys*, vol. 45, no. 8, pp. 3627–3636, 2018.
[28] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose ct," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, 2017.
[29] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, 2018.
[30] Q. Yang, N. Li, Z. Zhao, X. Fan, E.-C. Chang, Y. Xu *et al.*, "Mri image-to-image translation for cross-modality image registration and segmentation," *arXiv preprint arXiv:1801.06940*, 2018.
[31] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, H. Nakayama, and S. Satoh, "Learning more with less: Conditional pggan-based data augmentation for brain metastases detection using highly-rough annotation on mr images," in *CIKM '19*, 2019, pp. 119–127.
[32] B. Cao, H. Zhang, N. Wang, X. Gao, and D. Shen, "Auto-gan: self-supervised collaborative learning for medical

image synthesis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 10 486–10 493.
[33] J. Chen, J. Wei, and R. Li, "Targan: Target-aware generative adversarial networks for multi-modality medical image translation," *arXiv preprint arXiv:2105.08993*, 2021.
[34] B. Tang, F. Wu, Y. Fu, X. Wang, P. Wang, L. C. Orlandini, J. Li, and Q. Hou, "Dosimetric evaluation of synthetic ct image generated using a neural network for mr-only brain radiotherapy," *Journal of Applied Clinical Medical Physics*, vol. 22, no. 3, pp. 55–62, 2021.
[35] A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, "Adversarial image synthesis for unpaired multi-modal cardiac data," in *SASHIMI*. Springer, 2017, pp. 3–13.
[36] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J. L. Prince, N. Sugano, and Y. Sato, "Cross-modality image synthesis from unpaired data using cyclegan," in *SASHIMI*. Springer, 2018, pp. 31–41.
[37] J. Jiang, Y.-C. Hu, N. Tyagi, P. Zhang, A. Rimner, G. S. Mageras, J. O. Deasy, and H. Veeraraghavan, "Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation," in *Med Image Comput Comput Assist Interv*. Springer, 2018, pp. 777–785.
[38] D. Wei, S. Ahmad, J. Huo, P. Huang, P.-T. Yap, Z. Xue, J. Sun, W. Li, D. Shen, and Q. Wang, "Slir: Synthesis, localization, inpainting, and registration for image-guided thermal ablation of liver tumors," *Med. Image Anal.*, vol. 65, p. 101763, 2020.
[39] C. Han, L. Rundo, R. Araki, Y. Nagano, Y. Furukawa, G. Mauri, H. Nakayama, and H. Hayashi, "Combining noise-to-image and image-to-image gans: Brain mr image augmentation for tumor detection," *IEEE Access.*, vol. 7, pp. 156 966–156 977, 2019.
[40] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *ICANN*. Springer, 2017, pp. 626–634.
[41] S. U. Hassan Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *arXiv preprint arXiv:1802.01221*, 2018.
[42] M. Zhao, L. Wang, J. Chen, D. Nie, Y. Cong, S. Ahmad, A. Ho, P. Yuan, S. H. Fung, H. H. Deng *et al.*, "Craniomaxillofacial bony structures segmentation from mri with deep-supervision adversarial learning," in *Med Image Comput Comput Assist Interv*. Springer, 2018, pp. 720–727.
[43] S. Nema, A. Dudhane, S. Murala, and S. Naidu, "Rescuenet: An unpaired gan for brain tumor segmentation," *BIOMED SIGNAL PROCES*, vol. 55, p. 101641, 2020.
[44] Y. A. Mejjati, C. Richardt, J. Tompkin, D. Cosker, and K. I. Kim, "Unsupervised attention-guided image-to-image translation," in *Adv Neural Inf Process Syst*, 2018, pp. 3693–3703.
[45] S. Tripathy, J. Kannala, and E. Rahtu, "Learning image-to-image translation using paired and unpaired training samples," in *ACCV*. Springer, 2018, pp. 51–66.
[46] A. K. Srivastava and N. Kandpal, "Cumulative gradient based image sharpness evaluation algorithm for auto focus control of thermal imagers," 2015.
[47] X. Han, "Mr-based synthetic ct generation using a deep convolutional neural network method," *Med Phys*, vol. 44, no. 4, pp. 1408–1419, 2017.
[48] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *IEEE transactions on medical imaging*, vol. 17, no. 1, pp. 87–97, 1998.
[49] I. J. Cox, S. Roy, and S. L. Hingorani, "Dynamic histogram warping of image pairs for constant image brightness," in *Proceedings., International Conference on Image Processing*, vol. 2. IEEE, 1995, pp. 366–369.
[50] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, 2010.
[51] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.
[52] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," *arXiv preprint arXiv:1511.08861*, 2015.
[53] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.
[54] A. Saha and Q. J. Wu, "Full-reference image quality assessment by combining global and local distortion measures," *SIGNAL PROCESS*, vol. 128, pp. 186–197, 2016.
[55] J. Adler and S. Lunz, "Banach wasserstein gan," *arXiv preprint arXiv:1806.06621*, 2018.