Multimodal MR domain conversion using cyclic residual architecture and evolutionary algorithm

Erfan Darzidehkalani[1], Hossein Vahabie[1]

Control and Intelligence Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

Erfandarzi@ut.ac.ir, h.vahabie@ut.ac.ir

Abstract

Cross modality image transformations have a variety of applications in the diagnosis and treatment. Existing methods for training require a lot of data, and labeled in two different modalities that are not generally available in the world of medical imaging. Also generative adversarial networks have the problem of mode collapse due to the low amplitude of the density function. In this study, a combination of representations of different brain image scales was used to process the images, and the representations were combined with shortcuts. In modalities T1 and T2, a transformation was performed in which the partial structures of the brain images were preserved and the network was able to perform the modality transformation using multi-step conversion without paired data and in an unsupervised manner. The mean absolute error for T1 data is 6.843 and and for T2 data is 7.849, and the peak signal-to-noise ratio is 26.72 in T2 data and 26.24 in T1 data. Which has improved over existing methods despite unsupervised learning.

Keywords*: MR image synthesis, Generative Adversarial Networks, Evolutionary optimization, Residual learning, Unsupervised Learning*

I. INTRODUCTION

In many domains, such as radiation therapy planning to segment tumor volume and at-risk organs, T2-MR image is required as well as a T1-MR image. Acquiring these volumes separately is time consuming, and costly for the patient. In addition, the voxel spatial alignment between T2 and T1 images may be compromised, which requires accurate recording of T1 and T2 volumes. Therefore, to circumvent the acquisition of a separate image, a range of methods have been proposed in which an alternative or artificial image is taken from the existing image.

There is a standard solution to this problem, and that is to augment the input data artificially. For grayscale images, Affine transformations are commonly used,

which include rotation, zooming, translation, and flipping. These transformations do not take into account changes that may result from different imaging methods, or different signal sequencing protocols. It may also have its own specific pathological phenomenon, shape, appearance, color and variety, that these traditional methods do not provide any additional information about the above features.

Researchers are looking for an answer to this problem using machine learning algorithms. Autonecoders were one of the primary methods based on artificial intelligence. Several articles also examined the production of pseudo-realistic images and the addition of data using Convolutional Neural Networks (CNNs). However, such articles could not produce images with high resemblance to real images, since CNNs were generally not designed to learn the distribution domain.

One of the recent topics in computer science is Generative Adversarial Networks or GANs. These networks use two encoder-decoder networks, with competition between them. The generator network learns how to produce images very similar to the real image so that the detection network is not able to distinguish the fake from the real. The distance of this

data distribution with the original sample is evaluated by the discriminator network.

One application of GANs is to produce pseudo-realistic images. This can provide researchers with a tool to distinguish between healthy and sick people. Although the above network has shown a high ability in various fields, but for a specific application, the transfer of a particular image from one domain to another has the some problems. The aim of this research is to provide an optimal structure for producing pseudo-realistic images. In this issue, GANs are used as a basis and the goal is to fundamentally change and improve its structure to transfer images from one modality to another.

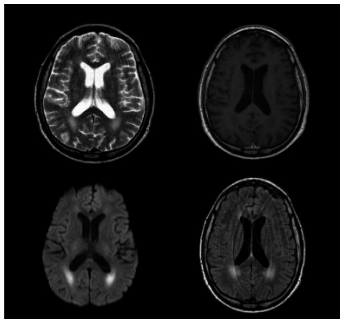

Figure 1 - MRI of brain. T2 Up-Left, DWI Low-left, which are brighter with higher intensity values, and FLAIR Low-right and T1 up-right .

II. RELATED RESEARCH

Physics based images synthesis

In the production of images of physical methods in magnetic resonance images, multiple images of a person's body are obtained using different pulse sequences, and by using mathematical functions based on the physical properties of MRI, images from different amplitudes can be compared to each other. [1]

With physical parameters in hand, these imaging equations can be used to generate new magnetic resonance images, as if they were taken with the same pulse frequency. A major advantage of this type of imaging is that the images are produced in a deterministic way based on mathematics, and therefore those images can be obtained without taking images in the desired modality. A view of this can be seen in the images on the left-right knee.

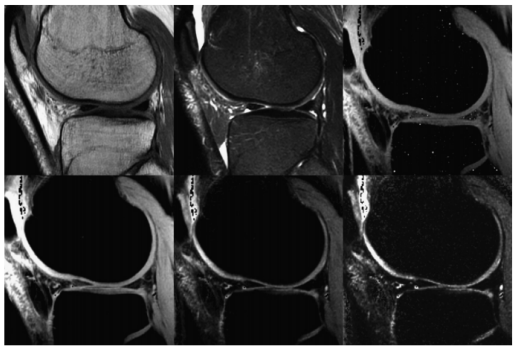

Figure 2 Synthesis of physics-based images

In the Figure 2, in the upper left, the image of rapid magnetic resonance is taken. In the middle and top, a T2 image of saturated fat can be seen. The four additional images are actually generated images taken at different echo times.

Classification based image synthesis

Producing physics-based images requires knowing the texture parameters in each voxel. Generating images based on classification is a facilitated method based on the previous method. In the classification-based method, the only parameter required for conversion is the type of texture. In this case, the nominal values corresponding to that type of texture are used in imaging methods and a new image with a new contrast is produced.

For example, if an image contains only a brain image that contains three types of tissue, including gray matter, white matter, and cerebrospinal fluid, the intensity of the voxel counts in T1 and T2 modalities has already been calculated and can be found by having the type.

Voxel calculated the image in the new modality from the previous image. This method of converting images from different modalities is valuable in that it uses both physical functions and mathematical equations related to imaging. This method is also the core of the Brain Web project [1].Brain Web is a magnetic resonance imaging and processing tool that is widely used to evaluate the performance of image processing methods.

Registration based image synthesis

This method was first proposed by Miller [2].The registration-based approach requires a booklet or atlas, which is a complete set of many images of one or more subjects. When the atlas is transformed to match the anatomy of the subject, all the images in that atlas can be

transferred with the same conversion function and adapted to the subject. Since atlas images are structurally similar to the individual anatomy after conversion, the adapted images can be considered as individual images.

The original method was actually based on a huge database of various images of magnetic resonance, CT and PET. For example, in the paper [3] researchers examined the method of producing a registration-based image using reference books, or a pair of image atlases. Using the above method, they were able to generate CT images from MRI. They compared the atlas of magnetic resonance images with a magnetic resonance image of a subset and applied the resulting conversion function to the corresponding CT image in that atlas to obtain a pseudo-image in CT modality.

Example based image synthesis

Example-based methods use atlases or training data to estimate the output image. However, unlike registration-based methods, they are not based on the spatial arrangement of the subject image and the atlas image. One of the first examples of this method is provided by [3] This method is inspired by the topic of image similarity in machine vision field. This topic actually tries to answer the question of what image B' is similar to image B that bears the same resemblance as image A' to image A.

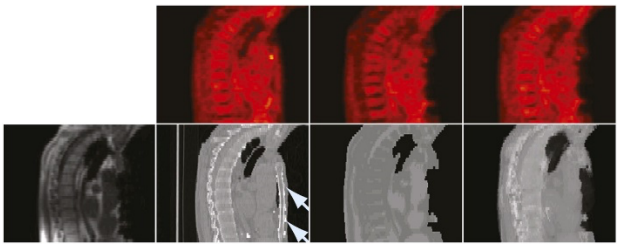

Figure 3. From left to right: MR image; CT image, with arrows indicating location of metal implant; pseudo-CT from 5-class MR image segmentation; and AT&PR based pseudo-CT. Top row: corresponding PET images reconstructed using pseudo-CT images for AC, generated with Example based method [3]

There is actually a collection of images A and A' images in the atlas. But instead of using a spatial conversion to transformation, it uses the properties and intensities of the images in the atlas set to create a mapping function. It then applies the generated mapping function to image B to produce image B '. Here, the properties in the atlas are examples of how we learn the conversion function. This is why this method is called example-based image production.

Applications of GANs in Medical Images

GANs are generally used in two aspects in medical imaging, the first way emphasizing aspects of image production that help us learn the hidden structure of data and reproduce similar images. This work is done using the generator part of the GAN.

The second characteristic uses the network discriminator. Which can be trained by recognizing a set of correct images, such as normal images of a category, and acts as a discriminator if it sees unrealistic images. In the following, examples of different applications of GANs in medical images are reported and explained.

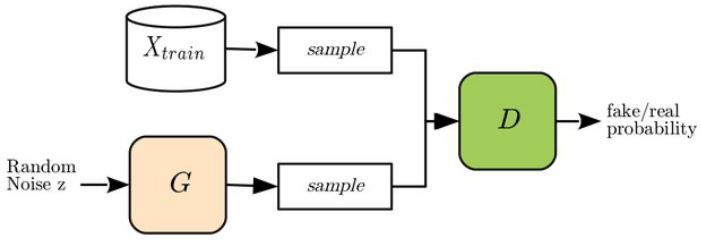

Figure 4- A schema of Generative Adversarial Network (GAN) structure

Reconstruction

Due to various problems in clinical conditions, such as radiation dose or patient conditions, images may be present due to noise and interfere with the diagnosis and treatment process. In the last decade, articles have generally shifted from the reconstruction of patients with computational methods to deep learning based methods. Also note the machine learning methods that have attracted attention in recent years. [4] These methods generally produce the final images either from the raw data of the imaging device or, as a post-processing step, eliminate new images produced by conventional methods.

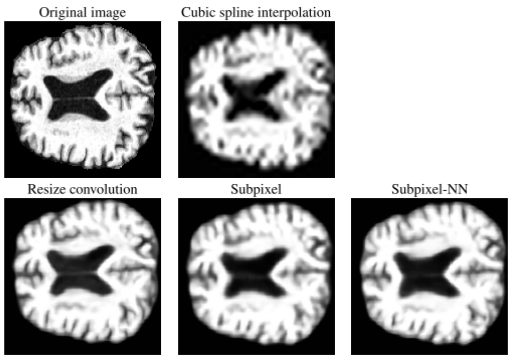

Figure 5 - Illustration of High resolution MR generation results. The first row shows the original high resolution image and the result of cubic spline interpolation. The next row presents the results of applying GAN , proposed in [5]

The converted images are used for noise reduction of CT images, reconstruction of magnetic resonance images, and noise reduction of PET images. . [6]] used a segmentation map of vessel images to increase the resolution of corneal images. The same idea was used by [7]in the reconstruction of computed tomography CT images. Other types of error functions have also been used to reset image structures for reconstruction. For example, changes were made to the error function by , [8] which weighs separately with each image voxel.

Domain conversion

The production of multimodal images (for example, the production of CT images from magnetic resonance imaging) is useful for a variety of reasons, including less imaging time and lower cost. Medical images can also be generated by applying conditions to segmentation maps, on the generated images. [9]Since GANs can receive inputs of different types and produce the desired output by appropriate structure change, recent research has significantly focused on the application of various inputs and the use of various medical information in image synthesis.

[10] Using special conditions, he attempted to produce synthesized images of lung nodules that connect to the lung border. Also, segmentation maps can be generated using arbitrary conditions. This is done using GANs by [11]

There are also some works done in mammogram image synthesis using GANs. [12]

Classification

Classification is probably one of the most successful areas in which deep learning has entered. Image features can be extracted hierarchically from a deep neural network and the image can be accurately classified by those features. GANs have also been used for classification purposes, with structures designed using both the generating network and the discriminator network for classification. [13]

Other structures have also been designed that use only the discriminator network for classification. This is done by adding output layers to the discriminator network to the desired number of classes, and applying an activation function to the end-layer neurons. [14]used both WGAN and InfoGAN networks to extract cell features in pathology images. [15] also is a good example of using GANs in pathology image classification.

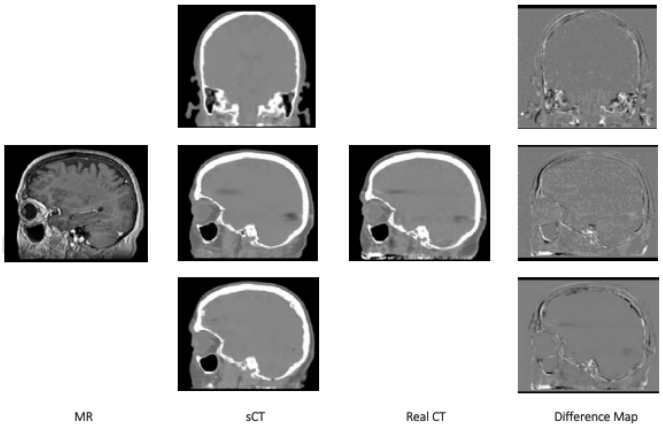

Figure 6- First column: MR; second column: syntheized CTs , third column real CTs, fourth column differences. CT images from MRI were synthesized using DCGAN, in article [16]

Segmentation

In general, researchers have considered cost functions based on pixels or voxels to classify medical images. Since a class must be assigned to each pixel or voxel in image segmentation, cost functions can be used to calculate classification accuracy. [17] However, these functions alone are not enough because, the output images under these networks are usually very rough and completely different from the actual images. [18] Introduced the popular U-net network, which used both low-level and high-level features to split images. However, this network does not guarantee compatibility in different parts of the production image. Also, the adversarial cost functions introduced in the discriminator network can also consider higher-order potentials. [19]

III. LIMITATIONS

Cost functions in most of researches mentioned above do not ensure the preservation of the output image structures. Therefore, if the original cost function, which is, for example, the RMSE between real and fake images, is minimized but the main structures of the image are shifted (e.g., borders and angles) so that the output image is visually very different from the target image, the network will not notice.

Another problem with GANs in medical domains is the lack of search in the global parameter space.

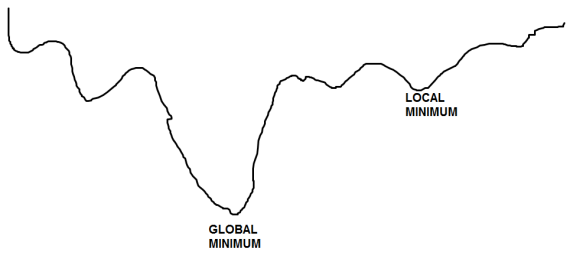

Figure 7 The problem of local minimum entrapment in neural networks

It is possible that with a series of parameters specific to the generator network, GAN converges in a direction that no longer gets better and its performance increase will be stopped. At this point, it is said that the network has converged to the local minimum point. However, reaching this point may not be sufficient for the intended purpose, and to achieve higher quality images and more accurate mapping, it may be more beneficial to find the global minimum point.

The property of the global minimum point is that it can give the best possible approximation of the mapping from the input distribution to the target distribution, taking into account the number of network parameters, which in fact can be said to guarantee the learning the objective function.

Since the production of medical images is related to treatment and is very sensitive, it is better to make the network more complex by forcing it to achieve higher accuracy.

This motivates the proposed optimization structure to be designed in such a way that, in this research paper, in addition to the local point guaranteed by the conventional Stochastic Gradient Descent (SGD) algorithm, the search for a global minimum point is considered to be an objective for the proposed network.

Given the above, it seems that there is a need for comprehensive research to address all of the above problems in medical image conversion. There may be a variety of ideas for solving the problems in the above network, but it will be very difficult to apply the right ideas so that they can work well together and be tested in practice.

Another problem commonly faced in training GANs is mode collapse, which, is a case when the distribution learned by G focuses on a few limited modes of the data distribution. Hence instead of producing diverse images, the model produces similar images over and over.

IV. PROPOSED SCHEME

First the data is collected, then the desired category is attributed to images (T1 or T2) and preprocessing is done. Generator and discriminator networks are then trained by receiving the corresponding images, data output, and corresponding cost functions. The network is then

optimized and the parameters of the generator and discriminator networks are updated using the methods described above. Then the next image is selected and the process is repeated until the final image. However, the cost functions presented in the figure are not selected from the same set of images, but multiple images are compared with each other and each has its own cost function, each function will be discussed in detail below.

Addressed questions

In this work, two questions have been answered in general:

Model design: Which architecture of GANs is more suitable for producing pseudo-realistic images? In fact, it is possible to use different architectures of GANs that can perform the mapping operation in the best way and also do not cause any problems in terms of stability.

Medical data augmentation: Can images of different MRI images with inherent differences in measurement methods be converted into each other? Images may arise from two different MR sequences. For example, an image may inherently represent a specific structure of brain images that may not be available at all in other imaging modalities. Can AI-based methods find hidden patterns in images from one domain so that another domain can be produced?

Results obtained

The results obtained in this research are as follows:

MR image synthesis: This study shows that GANs, especially GANs that use the information in different layers and scales of the image, can produce pseudo-realistic magnetic resonance images, which have valuable clinical applications. Clinical applications discussed earlier include increasing the accuracy of neural networks for analyzing medical images and adding disease-related data to physicians.

Medical image synthesis: In this research, the differences in different modalities of medical images are trained by the network and a neural network is designed to produce pseudo-real images by considering the inherent differences between different modalities. Although the proposed network uses structures that are specific to brain images and magnetic resonance imaging, it can be used to produce medical images by removing structures specific to brain images and providing a more general framework.

Network Architecture

After normalization, T1 and T2 images go to generator networks A and B and become the corresponding image in the other modality. Discriminator networks are also trained using the generated image and the real image. Also, the generated images are then returned to the original domain. The network consists of two generators and two detectors that are trained one after the other. The network has used special functions in each section for proper training, which will be mentioned below.

Generator network

We show the distribution of images in two different modalities as $\mathbf{x} \sim \mathbf{p_{data}}(\mathbf{x})$ and $\mathbf{y} \sim \mathbf{p_{data}}(\mathbf{y})$. As mentioned, the proposed model performs two types of mapping: $\mathbf{G}: \mathbf{X} \rightarrow \mathbf{Y}$ and $\mathbf{F}: \mathbf{Y} \rightarrow \mathbf{X}$. And the converted images are from another domain, namely $\mathbf{F}((\mathbf{Y}))$ Accordingly, the goal of $\mathbf{D_Y}$ will be to discriminate between $\mathbf{Y}$ and $\mathbf{G}((\mathbf{X}))$. In the generator network, convolutional layers are used in encoder-decoder networks. Which will be discussed below. That is a fundamental change compared to the original GANs offered by Goodfellow et al.

Residual connections

Shortcuts are placed between the convolution layers, as shown in the image below. In fact, after each convolution layer and its corresponding activation layer, a shortcut from the previous layer is added to bring the image representation at different scales directly together. In addition to improving network accuracy, this is also intuitively meaningful. Because instead of using image representation on a small scale and after applying various convolutional layers and activation operators, various

image representations in different scales are used and in fact the features of different scales produce the corresponding final image. One of the crucial benefits of this architecture is the use of shortcut blocks.

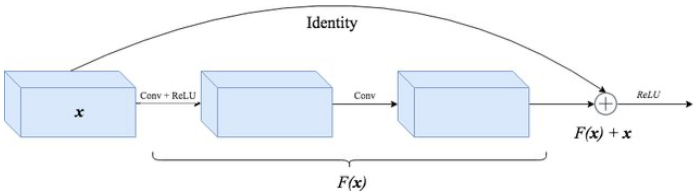

Figure 8 A schema of residual connections

Convolution layers

To achieve the highest accuracy in the production of images, the use of convolution layers is one of the methods in these networks.

The integration of convolution layers in GANs has been created with the aim of changing the architecture of the generator and discriminator networks and with the aim of solving the problem of network stability or improving the quality of images. Deep Convolutional Generative Adversarial Network (DCGAN) with the aim of improving network stability and developing training was introduced in the article [20]

In this architecture, convolutional neural network (CNN) is used for the generator and discriminator network. Due to the high stability and significant success of the DCGAN, as well as the proven advantage of convolutional structure in image processing, convolutional layers were used in the network.

Our model includes the convolutional layers introduced in the deep convolutional network, but there are some differences with it. Including that in the generator network, in each block, the ReLU activation function is used. Also in our network we have an initial convolution layer, and then two convolutional layers that come with the batch normalization layers and the ReLU activation function are added to the model, which will form a block of our model.

The other two blocks are designed to up-sample the image. Then the other two blocks are made in the same way but with decreasing dimensions so that the image returns to its original dimensions (down-sampling). This causes the output image of each generator, which is actually an image in one modality, to be the same size in terms of the input image as we actually are in another modality, after combining different layers of convolution.

Proposed cost functions

The purpose of this study includes four separate cost functions:

- **Adversarial cost function**: Wasserstein least squares function, to match the distribution of images produced with the target images
- **Cyclic cost function**: to prevent the mapping learned by the generator networks from interfering with each other
- **Identity cost function**: The distance of the input image to the generator that generates itself with the output, to extract / significantly combine features in different modalities
- **Cost function for structure stability**: use of feature-independent feature to maintain structural similarity of images in different modalities

Adversarial cost function

The adversarial cost function applies to both maps. For the mapping function $\boldsymbol{G}: \boldsymbol{X} \rightarrow \boldsymbol{Y}$ and its discriminator network $\boldsymbol{D_X}$, the cost function is defined as follows: While the generating network $\boldsymbol{G}$ tries to produce images $\boldsymbol{G}$**(x)** similar to the images of domain $\boldsymbol{Y}$, $\boldsymbol{D_Y}$ tries to distinguish between the images produced by $\boldsymbol{G}$**(x)** and the input images,which were fed to it namely $\boldsymbol{Y}$**.** $\boldsymbol{G}$ has a goal and it is to minimize its cost function, and the discriminator $\boldsymbol{D}$ has the goal to maximize the loss function.

$$\min_{G.F} \max_{D_X.D_Y} \boldsymbol{L_{GAN}}(\boldsymbol{G}.\boldsymbol{D_Y}) + \boldsymbol{L_{GAN}}(\boldsymbol{F}.\boldsymbol{D_X}) \quad . 1$$

There is also a similar relationship for the mapping function $\boldsymbol{F}: \boldsymbol{Y} \rightarrow \boldsymbol{X}$, and its associated detection network, $\boldsymbol{D_X}$:

$$\min_{F.G} \max_{D_Y.D_X} \boldsymbol{L_{GAN}}(\boldsymbol{G}.\boldsymbol{D_X}) + \boldsymbol{L_{GAN}}(\boldsymbol{F}.\boldsymbol{D_Y}) \quad . 2$$

Where**:**

$$\hat{\boldsymbol{y}} = \boldsymbol{G_\theta(x)} \quad . 3$$

$$\hat{\boldsymbol{\theta}} = \mathbf{argmin} \sum_i \boldsymbol{L}(\, \hat{\boldsymbol{y}}_i(\boldsymbol{\theta}). \mathbf{y(i)}) \quad . 4$$

Where $\mathbf{y(i)}$ represents a particular image in the target domain and $\hat{\boldsymbol{y}}\,(\boldsymbol{i})$represents the output of the model trying to approach $\mathbf{y(i)}$

Also for the cost function or L, the Wasserstein distance introduced in the WGAN paper is used. However, with the difference that the input values to the WGAN arguments, instead of the output network output values, their distance L2 was to the target values. WGAN network is presented in the article [21]in order to improve the learning stability of the network and solve the problem of failure to produce pseudo-real images. It was utilized in [22] for medical image classification, as a copy-paste of basic WGAN model, which performed very well. But for our specific use of comparing two converted images in two different modalities, it has been rewritten. Also, since there are two detector discriminator, this cost function is used twice.

Cyclic stability cost function

Adversarial training can, in theory, learn different mappings from the G and F generator networks that can produce outputs that have an exact distribution equal to the distribution of the X and Y domains, respectively. In fact, we can say that G and F must be random functions.

However, with enough data and due to the high capacity of GAN, this network can map a set of inputs to completely random instances of the output image distribution. Each of these learned mappings can have a distribution equal to the distribution of the target images. Therefore, the adversarial cost function alone cannot guarantee that the network-learned function can map an arbitrary input to its corresponding output.

In order to reduce the amount of possible parallel mappings from the input domain to the target domain, the condition can be applied to the function that the mapping functions must also be cyclically stable. That is, for every x image in the input range X, the image conversion cycle must be able to return the final image x to the original image. i.e.

$$\boldsymbol{x} \rightarrow \boldsymbol{G_\theta(x)} \rightarrow \boldsymbol{F_\theta(G_\theta(x))} \approx \boldsymbol{x} \quad . 5$$

This equation is called the forward cyclic stability equation. For each y image of the Y domain, the generator networks G and F must also be able to provide cyclic stability in the reverse direction. that's mean:

$$\boldsymbol{y} \rightarrow \boldsymbol{G_\theta(y)} \rightarrow \boldsymbol{F_\theta(G_\theta(y))} \approx \boldsymbol{y} \quad . 6$$

Using the cyclic stability cost function, the network goes in the direction of forcing the above two operations.

An attempt was also made to define a new cost function, based on the L1 distance, between the output images of $\boldsymbol{F_\theta(G_\theta(x))}$ and the target images $\boldsymbol{x}$, as well as between the output images of the network $\boldsymbol{G_\theta(F_\theta(x))} \approx \boldsymbol{x}$ and the target images of $\boldsymbol{y}$, Which did not show a positive effect.

The proposed model can be considered in another way, and with a little negligence, the training of two autoencoder networks. The first $\boldsymbol{F}\circ\boldsymbol{G} : \boldsymbol{X} \rightarrow \boldsymbol{X}$ encoder network and the second $\boldsymbol{G}\circ\boldsymbol{F} : \boldsymbol{Y} \rightarrow \boldsymbol{Y}$ decoder network are updated simultaneously based on the outputs of each other.

But in the intended application, autoencoder networks each have a series of specific features that differentiate the proposed model from the sum of two autoencoder networks.

I also evaluated my method with only cycle loss in one direction and show that a single cycle is not sufficient to regularize the training for this under constrained problem.

Cost function to maintain structural stability

Since the cyclic cost function does not guarantee structural stability, the proposed method adds an additional cost function between the pseudo-real images produced and the input images. However, since the two images are in two different domains, they are first measured using a new feature that can compare the features of images in different domains, and then the corresponding cost function is minimized. In this respect, this function is very suitable for comparing two images in two different modalities. The equations related to this functionality are discussed below.

In this research, "Modality Independent Neighborhood Descriptor" or MIND has been used as a common feature. MIND actually divides the image into smaller parts and measures the similarity of the image to itself. Since there are two generator networks in the proposed structure, each has its own structure cost stability function. This cost function is also defined between the intermediate output image in one modality and the output / reconstructed image in another modality.

$$\mathbf{L_{structure}}(\mathbf{x}.\mathbf{y}) = \boldsymbol{MIND}(\boldsymbol{x}.\boldsymbol{y}) \quad .7$$

$$\mathbf{y} \rightarrow \mathbf{G_\theta}(\mathbf{y}) \approx \mathbf{x} \quad .8$$

The output of this function is the similarity of the different parts of the image to itself, which can replace the values of the voxel intensity and therefore acts independently of the imaging method.

In some studies, MIND has been used to match CT and MR images as a measure of similarity. In the following, the MIND feature and the structural stability cost function are presented in detail.

The MIND feature extracts different structural parts of the image and compares each part with all neighboring parts. For the voxel or pixel x in Figure I, the MIND attribute, denoted $\boldsymbol{F_x}$, is actually a vector of length $|\boldsymbol{R_{nl}}|$, where $\boldsymbol{R_{nl}}$ represents the areas around the voxel $\boldsymbol{x}$. And each component of the image $[\![\boldsymbol{F}_{\boldsymbol{x}}^{(\boldsymbol{\alpha})}$ is defined for the $\boldsymbol{x}+\boldsymbol{\alpha} \in \boldsymbol{R_{nl}}$ voxels belonging to $\boldsymbol{R_{nl}}$, as follows:

$$\boldsymbol{F}_{\boldsymbol{x}}^{(\boldsymbol{\alpha})}(\boldsymbol{I}) = \frac{\mathbf{1}}{\boldsymbol{Z}} \boldsymbol{exp}\left(-\frac{\boldsymbol{D_p}(\boldsymbol{I}.\boldsymbol{x}.\boldsymbol{x}+\boldsymbol{\alpha})}{\boldsymbol{V}(\boldsymbol{I}.\boldsymbol{x})}\right) \quad .9$$

where Z is a normalization constant so that the maximal component of $\boldsymbol{F_x}$ is 1. $\boldsymbol{D_p}(\boldsymbol{I}.\boldsymbol{x}.\boldsymbol{x}+\boldsymbol{\alpha})$ denotes the L2 distance between two image patches **P** respectively centered at voxel $\boldsymbol{x}$ and voxel $\boldsymbol{x}+\boldsymbol{\alpha}$ in image $\boldsymbol{I}$, and $\boldsymbol{V}(\boldsymbol{I}.\boldsymbol{x})$ is an estimation of local variance at voxel x, which can be written as

$$\boldsymbol{D_p}(\boldsymbol{I}.\boldsymbol{x}.\boldsymbol{x}+\boldsymbol{\alpha}) = \sum_{\boldsymbol{p}\in\boldsymbol{P}}\left(\boldsymbol{I}(\boldsymbol{x}+\boldsymbol{p}) - \boldsymbol{I}(\boldsymbol{x}+\boldsymbol{\alpha}+\boldsymbol{p})\right)^2 \quad .10$$

$$\boldsymbol{V}(\boldsymbol{I}.\boldsymbol{x}) = \frac{\mathbf{1}}{\mathbf{4}}\sum_{\boldsymbol{n}\in\boldsymbol{N}} \boldsymbol{D_p}(\boldsymbol{I}.\boldsymbol{x}.\boldsymbol{x}+\boldsymbol{\alpha}) \quad . \quad .11$$

Since Z is a constant for normalizing the output, this constant is considered to be such that the maximum of the $\boldsymbol{F_x}$function is equal to 1. $\mathbf{D_p}(\mathbf{I}.\mathbf{x}.\mathbf{x}+\boldsymbol{\alpha})$ is actually the same distance L2 between two different P parts of image I, which are located in the voxel $\mathbf{x}+\boldsymbol{\alpha}$ and the voxel $\mathbf{x}$, respectively.

$\mathbf{V}(\mathbf{I}.\mathbf{x})$also represents the local variance in voxel $\mathbf{x}$, which can be written as follows:

$$\mathbf{V}(\mathbf{I}.\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{4}}\sum_{\mathbf{n}\in\mathbf{N}} \mathbf{D_p}(\mathbf{I}.\mathbf{x}.\mathbf{x}+\boldsymbol{\alpha}) \quad .12$$

The set of N voxels represents the 4 neighboring voxels of the x voxel. It is difficult to calculate the operator $\mathbf{D_p}$ and its gradient in a deep neural network. Instead, as mentioned in the article; It can be expressed in another way, which is based on the convolution or operator.

$$\mathbf{D_p}(\mathbf{I}.\mathbf{x}.\mathbf{x}+\boldsymbol{\alpha}) = \mathbf{C} * (\mathbf{I} - \mathbf{I}'(\boldsymbol{\alpha}))^2 \quad .13$$

In the above equation, C represents a kernel with values of 1 and a size equal to part **P**, and $\mathbf{I}'(\boldsymbol{\alpha})$ represents the same I but with a transition of size $\boldsymbol{\alpha}$. By doing this, the structural property can be done with a few simple operations and the gradients of these operators can be easily calculated.

The cost function, which is defined based on the MIND attribute, actually maintains the stability of the image structures. And this function is applied in two parts. The above can be summarized as follows:

$$\mathbf{L_{structure}}(\mathbf{x}.\mathbf{y}) = \frac{\mathbf{1}}{\mathbf{N_{MR}}|\mathbf{R_{nl}}|}\sum_{\mathbf{x}} ||\mathbf{F_F}(\hat{\mathbf{y}}_\mathbf{i}(\boldsymbol{\theta}).\boldsymbol{x}(\mathbf{i}))|| + \frac{\mathbf{1}}{\mathbf{N_{MR}}|\mathbf{R_{nl}}|}\sum_{\mathbf{x}} ||\mathbf{F_G}(\hat{\mathbf{x}}_\mathbf{i}(\boldsymbol{\theta}).\mathbf{y}(\mathbf{i}))|| \quad .14$$

Where $\mathbf{N_{MR}}$ represents the number of voxels of the input images, and $||\circ||$ Expresses the L1 norm

Training cost function:

Based on the definitions of adversarial cost functions, cyclic stability and structural stability introduced above,

the cost function of the proposed method is defined as follows:

$$\mathbf{L}(\mathbf{G_a}\,.\,\mathbf{G_b}\,.\,\mathbf{D_a}\,.\,\mathbf{D_b}) = \mathbf{L_{GAN}}(\mathbf{G_a}\,.\,\mathbf{D_a}) + \mathbf{L_{GAN}}(\mathbf{G_b}\,.\,\mathbf{D_b}) + \boldsymbol{\lambda_1}\,\mathbf{L_{cycle}}(\mathbf{G_a}\,.\,\mathbf{G_b}) + \mathbf{L_{structure}}(\mathbf{G_a}\,.\,\mathbf{G_b}) + \boldsymbol{\lambda_2}\mathbf{L_{Identity}} \quad .15$$

Where the parameters λ1 and λ2 determine the relative importance of the cost function sentences, λ1 is set to 10 and λ2 to 5 when training. To optimize L, instead of the conventional method of updating $\mathbf{D_a}$ (with fixed $\mathbf{G_a}$) and updating $\mathbf{D_b}$ (with fixed $\mathbf{G_b}$) an evolutionary algorithm-based method is used, the details of which will be discussed below

Network optimization

Since the introduction of the GANs, extensive efforts have been made to resolve its problems. Studies have focused on improving the efficiency of GANs. In this work, using adversarial methods based on evolutionary algorithms, action will be taken to optimize GANs.

Evolutionary Algorithm

Unlike conventional GANs, which update the generator network and the detector network, respectively, and based on the evolutionary algorithm, an environment is first defined, where the parameters of the detector network are $\mathbf{D}$. A population is then generated from the set of $\mathbf{G}$ generator network parameters. Evolutionarily, each member of the population, which is actually a set of $\mathbf{G_\theta}$generator network parameters, represents a possible response in the generator network parameter space. In the evolutionary process, it is expected that the set of parameters of the generating network will gradually adapt to its environment. This means that Evolutionary GANs can produce more realistic examples and eventually learn to distribute real data. During the evolutionary process, each stage consists of three sub-stages as follows:

- **Variation**: Having a set of generator network parameters in the form of $\mathbf{G_\theta}$, the change method is used to generate its child parameters, ie $\{\mathbf{G_{\theta_1}}.\mathbf{G_{\theta_2}}.\cdots\}$. Specifically, multiple copies of each initial parameter or parent parameter are generated. Each of which is changed by random and different mutations. Each modified sample is then considered a child.

- **Assessment**: For each child, its performance is assessed using the $\mathbf{F(\cdot)}$ compatibility function. In this particular case, the investigator function is the same as the generating network.

- **Selection**: All children are selected based on the degree of adaptation to the environment, the worse parts are eliminated and the better parts remain and advance to the next stage.

After each evolutionary stage, the discriminator network (environment) is updated to be able to distinguish between true x-sample images and y-unreal images. That is, the following equation must be solved:

$$\max_{\mathbf{w}\in\mathbf{W}} \mathbb{E}_{\mathbf{x}\sim\mathbf{P_x}}[\mathbf{f_w}(\mathbf{x})] - \mathbb{E}_{\mathbf{y}\sim\mathbf{P_{(y)}}}[\mathbf{f_w}(\mathbf{g_\theta}(\boldsymbol{y}))] \quad .16$$

Therefore, the discriminator network (or environment) can continuously define different cost functions to select the best set of generator network parameters. Finally, the set answers are optimized.

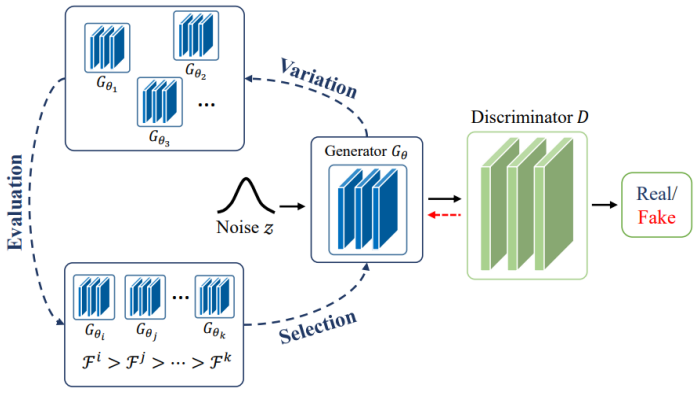

Figure 9- E-GAN framework, the evolutionary optimization for GANs was first proposed in E-GAN paper, which a version of this framework was used in our method [23]

Mutation

Asexual reproduction is used to create mutations. That is, there is only one primary factor (father) and children are created from a variety of father parameters. In particular, these mutations occur randomly, and the ultimate goal of

the mutation process is to bring the parameters of the generator network closer to the ideal parameters of the discriminator network.

V. EXPERIMENT

Data

The Human Connectome Project dataset was used to examine T1 and T2 data. The Human Connectome Project is a project that aims to identify and create a comprehensive map of human brain connections, with a structural and functional approach. These data are part of the data used in this project, including brain images of people. The above data were obtained from a Siemens device with a power of 3 T. All images were adapted to the same brain shape, as if in the same position and size.

In this research, only sagittal axis images have been used. Also, a two-dimensional image of 120 slice is used. Data were divided into training and evaluation. 900 data from each imaging method were used for training and the remaining 213 data were used for evaluation. Also, the pixel intensity range of the images is [255, 0].

Implementing the network using the Nvidia graphics processing unit provided by Google, with 12 GB of memory, the training time is 3180 seconds per round and the model is trained 200 rounds. The production time of each image is 0.014 seconds for each image.

The optimizer function was Adam, and the learning rate fixed at 0.0002 the first 100 rounds and zero linearly in the next 100 rounds. Three sample images are displayed in T1 modality and three sample images are displayed in T2 modality.

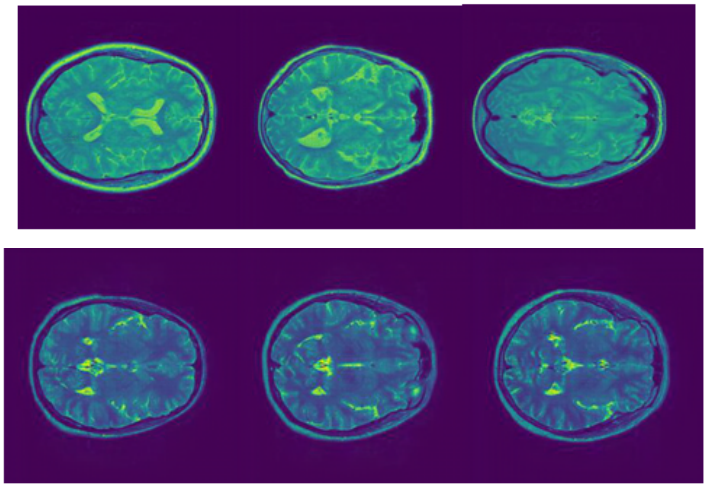

Figure 10 – Sample images in T1 and T2 modalities, images are scaled between red and green for better visualization

The total cost function, as formulated in the previous section, is the sum of the structural, adversarial, identity, and cyclic stability cost functions for the generator network. The cost function per 200 epoch training is shown in the figure below.

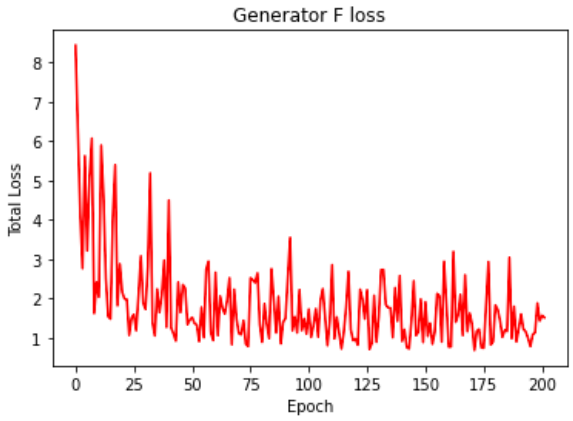

Figure 11 -The Loss plot of the generator network F

This function is shown for the generator network F, which converts images T1 to T2 is shown in the figure below.

The loss plot in the figure below is the generator network function for conversion from modality T2 to modality T1. Which is actually the G generator network.

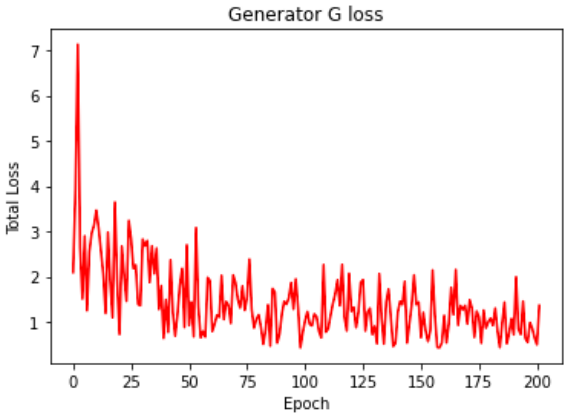

Figure 12 The Loss plot of the generator network G

The discriminator network corresponding to the generator network F is shown in the figure below. The grid actually has two T-mode inputs that are averaged between their corresponding cost functions.

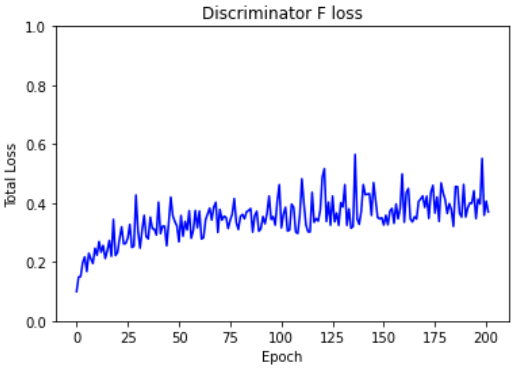

Figure 13 Discriminator network F loss plot

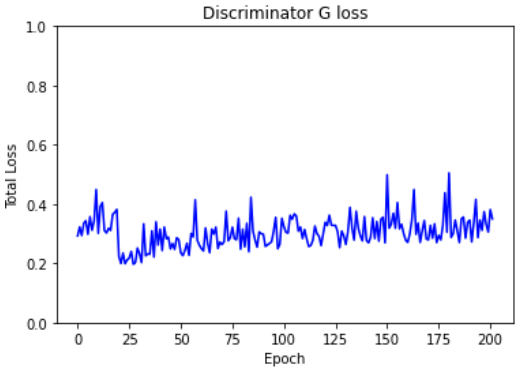

Figure 14 Discriminator network G loss plot

VI. RESULTS

Evaluation Criteria

Images generated by the network in the target imaging method, and the corresponding real images in that domain, can be compared as follows using the absolute mean error value criterion:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |F_\theta(G_\theta(x_i)) - x_i| \qquad .17$$

The variable i counts the corresponding voxels in the actual and generated images. In addition, the similarity between the output images and the actual images is measured by the maximum signal-to-noise ratio criterion.

$$\mathrm{PSNR} = `20\log_{`10}\left(\frac{\mathrm{Max_i}}{\sqrt{\mathrm{MSE}}}\right) \qquad .18$$

$\mathrm{Max_i}$ is the maximum brightness of the image and MSE is the mean square error.

The proposed method is compared with UNIT network, UNIT network with paired data, Pix2Pix network and DCGAN network. As mentioned, the presented network, like the UNIT network, performs the image conversion operation in an unsupervised manner, i.e. without having paired data, so in cases where the data is presented to a network in pairs, the provided networks will perform better. Of course, in real cases, this is often impossible due to the unavailability of paired data. The average absolute error for T1 data is 6.843 and for T2 data is 7.849 which is higher than similar and existing methods. The results of similar articles can also be evaluated in the image below.

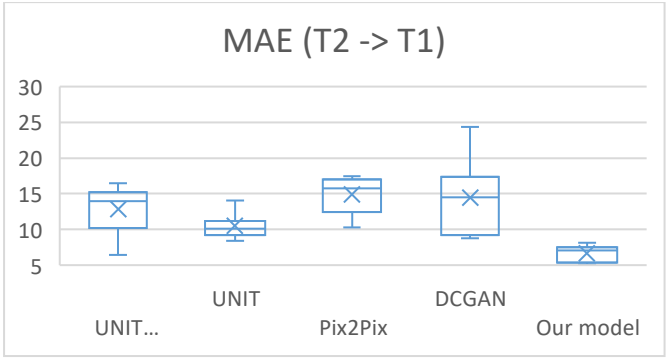

Figure 15 Quantitative error measurement- MAE for the GAN models

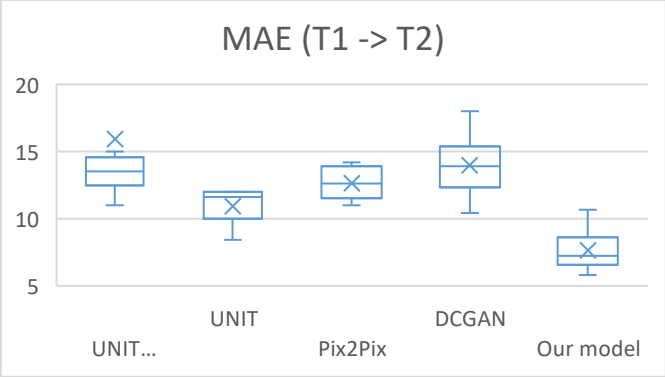

Figure 16 - Quantitative error measurement- MAE for the GAN models

Also, based on the PSNR criterion in the test data, the proposed method with a value of 26.72 in T2 data and .26.24 in T1 data has gained higher accuracy than similar methods.

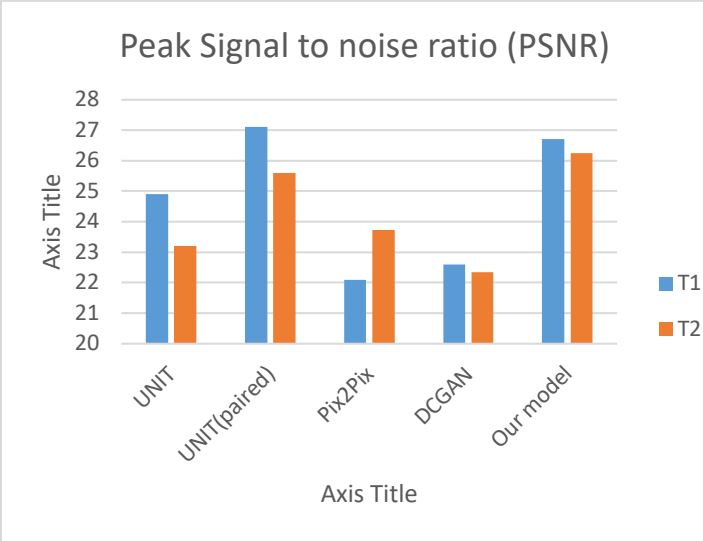

Figure 17 - Quantitative error measurement- PSNR for the GAN models.

The left column represents the input images, the middle column represents the generated images and the right column of the images correspond to the original image. The upper row is the input image in T1 modality and the corresponding image is generated in T2 modality. The bottom row is the input image in T2 modality and the corresponding image is generated in T2 mode.

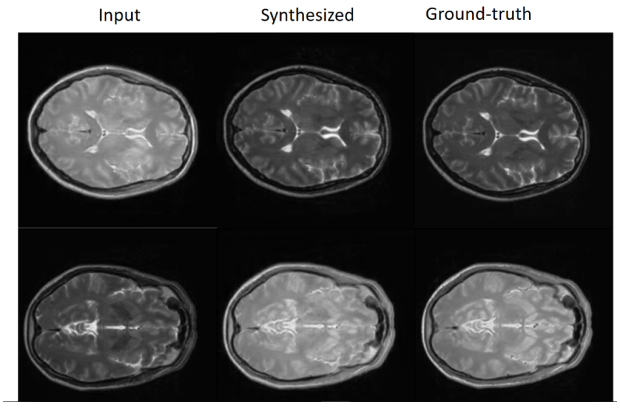

Figure 18 - From left to right Input MR image, synthesized MR image in another domain, reference real MR image

Also in the images below, to better understand and compare the preservation of structures in the images produced, part of the MR image is isolated and then magnified. In the figure below, the top row of production images and the bottom row are real images.

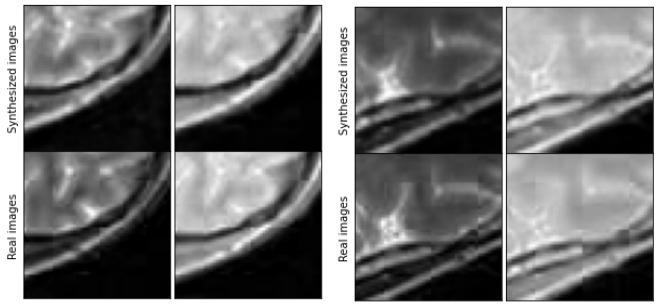

Figure 19 - Real and synthesized images, a comparison between structures

The left row of images is T2 and the right row of images are T1. Noise patterns may appear on the image due to magnification. But what is clear is that even the smallest structures are well preserved. Of course, since there is always noise and the gradient changes of the magnetic field are different from what is in the theory of MRI devices, even the corresponding T1 and T2 images may not be able to reflect all the details of each other.

VII. DISCUSSION

In this part of the research, the results obtained from the studies have been summarized. In the following, the

results of the research is described and main limitations of this research are stated. Finally, suggestions for future studies are provided.

Results: The results shown in the previous section are based on test data, and from images of the Human Connectome project. The above images are structurally very similar to the original image as well as the partial structures in the images, which is a direct result of applying a structural stability function to the presented network. Experiments show that this network is probably capable of receiving complex information and patterns in both origin and destination modalities. and in fact the network is able to compare images of different modalities with each other as well as intermediate images that are somehow impossible to compare.

Efficiency: The method of producing the presented images can help in the production of pseudo-real images in the segmentation of different parts of brain images, and this issue has been studied in various articles and its accuracy has been proven [24]. For example, the production of images in the traditional way, to a limited extent, can help increase the accuracy of classification and increase the performance of classification networks. Produced images can also be used to improve the classification and reconstruction of brain images. Its obvious application is in the study of brain tumors using deep neural networks. However, considering the high diversity of input data and also the successful performance of the network in the test data, it can be concluded that the network is robust to different data types.

Effectiveness: The main function of the network is to represent the image in a space independent of modalities and independent of image size. It is possible to train the network in general on the desired slices of these images, or to consider a three-dimensional network consisting of several or all of the slices that work separately for each brain slice. In addition, the work done considers the registered images in the same way / with the same algorithm. This means that the network is unable to perform registration operations.

VIII. CONCLUSION

The observed results generally prove the capability of the presented network in converting images from different domains. Also, the proposed network is very successful in terms of convergence. In terms of appearance, as seen in the resulting images, the images produced have a completely similar appearance to real images, which is sometimes difficult to distinguish from real images.

The generated T2 images are more similar to the real images. Also, the mean squared error is higher in the T1 images, which is 255 due to its brighter nature. Also qualitatively, in T2 images it is more difficult to detect whether the noise is real or not because it is darker. As expected, the greatest error and ambiguity in the images produced were in the parts that had the edge and the changes in light intensity were greater. Although attempts have been made to preserve the structures in the image (and of course their edges) with functions that try to preserve the structure, in some cases the above network has not been successful.

This research has had several innovations. First, solving the problem of convergence of GANs was on the agenda so that a network could be designed without the network having the problem of network disruption. The presented network converges well and gives acceptable results. Also, the improvement of the general search of the parameter space was presented using an evolutionary algorithm, which prevents the network from getting stuck in the optimal local points. Also, a coherent structure was applied to be able to reproduce existing image conversion methods in an unsupervised manner, which is especially useful in medical applications where there is a data constraint problem.

In addition, the structural information contained in the images in different modalities is used to solve the problem of comparison between the produced images in different modalities. In fact, T1 and T2 modalities are generally not directly compared, because the difference between the mean squares of their error increases due to one being light and the other dark, and the network may make a mistake that cannot work well. And large errors prevent the network from convergence.

Also, the lack of direct comparison of these images has led to the loss of valuable information, because in fact the ultimate goal of the network is the similarity of images in different modalities, which should have been done in a direct way. This has been done in this research.

The problems of this research are, first of all, a lot of time for network training, so that practically examining various network problems and improving network performance by changing its parameters is a time consuming task and user choices are very limited. The above method may also require a fundamental change for images of patients. Because more information about the location and boundaries of the tumor should be available so that the exact equivalent can be found in other imaging modalities.

IX. REFERENCES

[1] C. Cocosco, "Brainweb: Online interface to a 3D MRI simulated," *NeuroImage,* vol. 5, 1997.

[2] I. Miller, "Mathematical textbook of deformable neuroanatomies," *Proceedings of National Academy of Sciences,* vol. 20, pp. 11944-11948, 1993.

[3] M. Hofmann, "MRIbased attenuation correction for whole-body PET/MRI: quantitative evaluation of segmentation- and atlas-based methods," *Journal of Nuclear Medicine,* vol. 9, pp. 1392-1399, 2011.

[4] S. L.-. m. Bermudez, "Learning implicit brain mri manifolds with deep learning,," *Medical Imaging,* p. p. 105741L, 2018.

[5] V. V. Irina Sánchez, "Brain MRI super-resolution using 3D generative adversarial networks," Barcelona, 2017.

[6] . D. and Fulham, "Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs)," *Molecular Imaging, Reconstruction and Analysis of Moving Body Organ,* pp. 43-51, 2017.

[7] S. Kida, S. Kaji, K. Nawa, T. Imae, T. Nakamoto, S. Ozaki, T. Ohta, Y Nozawa and K. Nakagawa, "Cone-beam CT to plan- ning CT synthesis using generative adversarial networks," *arXiv:1901.05773,* vol. 54, no. 9, pp. 239-259, 2019.

[8] D. Jin, Z. Xu, Y. Tang, A. Harrison and D. Mollura, "Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation," *arXiv preprint ,* p. arXiv:1806.04051.

[9] C. Chen, D. Q. C. H. and H. P. , "Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation," *arXiv preprint ,* p. arXiv:1806.00600, 2018a.

[10] M. Chuquicusma, S. Hussein, J. R. Burt and U. Bagci, "How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis," in *IEEE 15th International Symposium on Biomedical Imaging*, Berlin, 2018.

[11] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Medical Physics,* vol. 43, no. 32, pp. 1408-1419, 2017.

[12] D. Korkinof, T. Rijken, M. O'Neill, J. Yearsley, H. H.-. vey and B. Glocker, "High-resolution mammogram synthesis using progressive generative adversarial networks," in *arXiv:1807.03401*, Irvine, CA, 2018.

[13] E. Wu, K. Wu, D. Cox and W. Lotter, "Conditional infilling GANs for data augmentation in mammogram classification," *Lecture Notes in Computer Science,* vol. 1758, no. 85, pp. 98-106, 2018.

[14] H. B, T. Y, Chang, E.I, Fan, L. Y., M., Xu and Y, "Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks," *arXiv preprint ,* p. arXiv:1711.11317.

[15] J. Ren, I. Hacihaliloglu, E. Singer, D. Foran and X. Qi, "Adversarial domain adaptation for classification of prostate histopathology whole-slide images," *arXiv preprint ,* p. arXiv:1806.01357.

[16] . N. Network, "Xiao Han," Maryland Heights, 2017.

[17] Z. Zhang, L. Yang and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Boulder. Colorado, United States, 2018.

[18] T. Brox, "U-net: Convolutional networks for biomedical image segmentation,," *International Conference on Medical image computing and computer-assisted intervention,* vol. 23, p. 234– 241, 2015.

[19] A. Srivastava, L. Valkov, C. Russell, M. U. Gutmann and C. Sutton, "VEEGAN: Reducing mode collapse in GANs using implicit variational learning," *Adv. Neural Inf. Process. Syst,* p. 3309–3319, 2017-Decem.

[20] A. Radford, L. Metz and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," Newyork, 2016.

[21] K. E. Baumgartner, "Visual feature attribution using wasserstein gans.," *arXiv preprint,* p. arXiv:1711.08998, 2017.

[22] X. Y, E. Walia and B. P, "Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by wasserstein distance for dermoscopy image classification," *arXiv,* p. arXiv:1804.03700.

[23] D. T. Chaoyue Wang, "Evolutionary Generative Adversarial Networks," in *Neural and evolutionary computing*, 2018, 2018.

[24] T. C. Mok and A. C. Chung, "Learning data augmentation for brain tumor segmentation with coarse-to-fine generative ad- versarial networks," *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries,* vol. 11383, no. 24, pp. 70-80, 2019.