This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3189735

A Dual Discriminator Fourier Acquisitive GAN for Generating Retinal Optical Coherence Tomography Images

Mahnoosh Tajmirriahi, Rahele Kafieh, Zahra Amini, and Vasudevan Lakshminarayanan

***Abstract*—Optical coherence tomography (OCT) images are widely used for clinical examination of the retina. Automatic deep learning-based methods have been developed to classify normal and pathological OCT images. However, lack of the big enough training data reduces the performance of these models. Synthesis of data using generative adversarial networks (GANs) is already known as an efficient alternative to increase the amount of the training data. However, the recent works show that despite high structural similarity between synthetic data and the real images, a considerable distortion is observed in frequency domain. Here, we propose a dual discriminator Fourier acquisitive GAN (DDFA-GAN) to generate more realistic OCT images with considering the Fourier domain similarity in structural design of the GAN. By applying two discriminators, the proposed DDFA-GAN is jointly trained with the Fourier and spatial details of the images and is proven to be feasible with a limited number of training data. Results are compared with popular GANs, namely, DCGAN, WGAN-GP, and LS-GAN. In comparison, Fréchet inception distance (FID) score of 51.30, and Multi Scale Structural Similarity Index Measure (MS-SSIM) of 0.19 indicate superiority of the proposed method in producing images resembling the same quality, discriminative features, and diversity, as the real normal and Diabetic Macular Edema (DME) OCT images. The statistical comparison illustrates this similarity in the spatial and frequency domains, as well. Overall, DDFA-GAN generates realistic OCT images to meet requirements of the training data in automatic deep learning-based methods, used for clinical examination of the retina, and to improve the accuracy of the subsequent measurements.**

***Index Terms*—Optical coherence tomography, Generative Adversarial Network, Dual discriminator, Multi-task learning, Diabetic macular edema, Fourier analysis.**

I. Introduction

Optical Coherence Tomography (OCT) is a noninvasive, high-resolution imaging technique which projects cross sectional images from the human retina. Since ocular pathologies generally cause changes in the retina, 2D OCT images (B-scans) have been widely used by ophthalmologists for early detection of the pathology. OCT images can be used in computer aided diagnosis (CAD) systems for automatic classification of various disorders. For instance, diabetic macular edema (DME) is diabetes-related complication which causes central vision loss [1] and can be detected through OCT B-scans.

Deep learning-based classification methods can automatically detect the discriminative features of pathological B-scans efficiently without any human supervision [2]. These methods need large, diverse, and well-balanced training datasets to be trained in order to work effectively. However, collecting and labeling of data can be time consuming and costly. In addition, sharing medical images is not allowed due to privacy regulations (E.G., the US Health Insurance Portability and Accountability Act.). As a result, availability of the training data is limited.

To overcome the limitation of the training data, several augmentation strategies have been developed [3]. Data augmentation can increase the generality of the classifier model, help resolve the class imbalance problem and inhibit the model from over-fitting. Since distribution of the shape of the retinal layers is relatively fixed (because of anatomical constraints), few classic image augmentation methods such as rotation, horizontal flip, shifting, adding white noise, adding multiplicative speckle noise, elastic deformation, and occluding patches could have been used to increase retinal data [4]. Recently, Generative adversarial networks (GANs) have been utilized for various applications [5], [6]. They learn patterns of a dataset and generate new examples resembling the input data. GANs have been widely used in various fields such as generating artificial fault signals in rolling bearing to improve classification results [7], or synthesis of magnetic resonant images (MRI) to improve segmentation results [8]. They can produce large and diverse retinal datasets with high quality and reduce the need for data acquisition, manual annotation, and

This paragraph of the first footnote will contain the date on which you submitted your paper for review.

Mahnoosh Tajmirriahi, Rahele Kafieh, and Zahra Amini are with the Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan 81746734641, Iran. (e-mail: mata.riahi@yahoo.com; rkafieh@gmail.com; zahraamini64@yahoo.com.au)

Rahele Kafieh is also with Department of Engineering, Durham University, South Road, Durham, UK

Vasudevan Lakshminarayanan is with the Theoretical and Experimental Epistemology Lab, School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario N2L3G1, Canada (e-mail: vengulak@uwaterloo.ca)

Corresponding author: Rahele Kafieh

data augmentation steps [9]. Realistic synthesized OCT images can be used for educational purposes in the training of retinal specialists and can be successfully used for training of accurate deep learning-based methods for data analytics and biomarker measurements.

However, despite the outstanding performance of GANs in synthesizing natural images, non-convergence, instability, and mode collapse are remained as common challenges. Several re-engineered network architectures have been proposed to tackle these problems. Moreover, to improve learning the probability distribution of the data, the objective functions were modified in networks such as seep convolutional GAN (DCGAN) [10], Wasserstein GAN (WGAN) [11], WGAN with gradient penalized (WGAN-GP) [12], and least square GAN (LS-GAN) [13]. To control the diversity of the generated images, conditional GAN [14] was proposed in which the class information is embedded in the training data. Info-GAN [15] and auxiliary classifier GAN (AC-GAN) [16] are other GAN architectures proposed to control the generated data. Progressive Growing GAN (PG-GAN) [17] is a new method for stable training of GANs and to generate large, high-quality images. The process starts with a very small image and blocks of layers are added to the generator to increase the output size. Meanwhile, these GANs often require large amount of data to learn the probability distribution of the real data and generate the realistic data.

Recently, GAN-based methods have been proposed for denoising [18] and super resolution of the OCT images [19]; but to the best of our knowledge, only a few studies have been conducted on the production of synthetic retinal data using GAN networks. Odaibo [20] augmented OCT B-scans up to 500,000 images and used them to train a DCGAN to generate synthetic OCT B-scans. Liu et al. [9] utilized DCGAN and WGAN to synthesize new retinal fundus images. They utilized large amount of training data from Retina Image Bank (RIB) [21] to train the GANs. Yanagihara et al. [22] utilized 6875 paired OCT B-scans to train and validate a conditional GAN to generate synthetic data. Burlina et al. [23] used 133,821 age-related retinal fundus images to train a PG-GAN and provided a deep learning based classification method to show that synthetic images lead a classifier which performs as well as the real data. Zheng et al. [24] utilized 108,312 OCT B-scans to train a PG-GAN to generate high-resolution OCT images with urgent and non-urgent referrals. Sengupta et al. [25] applied a deep residual variational auto-encoder (RSVAE) to generate blood vessel annotation of fundus images. They used these annotations to train a pix2pix GAN [26] to generate a complete fundus image dataset. Using this, they could generate fundus images reducing the number of real training images to as low as twenty.

However, despite structural similarity between synthetic data and the real images, a considerable distortion is observed in frequency information of the synthetic data produced by previous methods. Singh et al. [27] showed that Fourier components of real and synthetic fundus images are quite different especially at higher frequencies and this can be used to discriminate real and synthesized fundus images.

On the other hand, there are studies to investigate the relationship between the image resolution and its Fourier transform details. Mizutani et al. [28] determined the spatial resolution of real X-ray micro tomography images using a logarithmic plot of the squared norm of their Fourier transforms. Wang [29] illustrated that quality of spatial image and its corresponding frequency domain are related to each other. He showed that in a blurred image the magnitude of high frequencies is much smaller than a sharp image which have more high frequency components spread along with the two central stripes.

Motivated by these investigations and in order to fill in the existing gap of dissimilarity between real and synthetic images in the frequency domain, in this study we extend the GAN model with dual discriminators and propose a dual-adversarial GAN which can jointly learn both spatial and frequency domain characteristics of the OCT B-scans. To the best of our knowledge, this is the first GAN model which comprises frequency information of the OCT images in training the network. Since preserving the frequency components of images plays an important role in their quality, the ultimate goal of this study is to investigate the effect of contributing Fourier information in training GANs and to prove the feasibility of generating high resolution, more realistic DME and normal B-scans with a limited number of training data. Thus, the main contributions of this study are: 1) A novel generative model, called dual-discriminator Fourier acquisitive GAN (DDFA-GAN) is proposed that is jointly trained in an end-to-end manner with multi-task learning of both spatial and Fourier information of the OCT B-scans. 2) The advantages of variational auto-encoders (VAE), and auto-encoders (AE) are used by transfer learning to increase the stability of the GAN, to provide faster convergence, and to prevent mode collapse. 3) An evaluation method is proposed for the generated B-scans to measure their quality and validate their usefulness. Quantitative evaluation, examination of synthetic data for data augmentation, and ablation study confirmed that the proposed DDFA-GAN can resolve the distortions in frequency information and increase the diversity and quality of the generated normal and DME target images despite utilizing limited number of the training data.

The rest of the paper is organized as follows. In section II, a brief background for GAN networks is presented. The details of the proposed network are expressed in section III. In section IV, the experiments and evaluation results of the proposed model are provided and discussed. In section V, some concluding remarks and future works are expressed.

II. Theoretical Background

The original formulation of a GAN is a min-max game between the discriminator $D(x)$ and generator $G(z)$ which try to outwit each other. The generator $G(z): z \rightarrow x$, maps the noise vector z to the data space. The discriminator, $D(x) \rightarrow [0,1]$, takes a point x in data space and computes the probability that x is either sampled from real data, $D(x)$, or generated by the generator, $D(G(z))$. This adversarial training process can be

formulated as min-max problem as depicted in (1) where $P_{data(x)}$ is the distribution of real data [30].

$$min_G max_D V(D,G) = \mathbb{E}_{x\sim P_{data(x)}}[\log(D(x))] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (1)$$

$p_z(z)$ is usually a Gaussian distribution $\mathcal{N}(0,1)$ that is used to draw the samples and $\mathbb{E}$ is the expectation operation. The GAN training is performed through two iterative trainings of discriminator and generator. The goal of discriminator training is to maximize the loss function expressed in the following [30].

$$\mathcal{L}_D = \mathbb{E}_{x\sim P_{data(x)}}[\log(D(x_r))] + \mathbb{E}_{x_G\sim p_G(x)}[\log(1-D(x_G))] \quad (2)$$

where x_r is a batch of real data randomly selected from the training set, and x_G are the randomly selected batches from the generated images. The loss function is maximized through the gradient descent algorithm during the training process.

The generator training contains updating the weights of the generator by minimizing the following loss function:

$$\mathcal{L}_G = \mathbb{E}_{x_G\sim p_G(x)}[\log(1-D(x_G))] \quad (3)$$

Several challenges such as mode collapse, vanishing gradient, and convergence failure may occur during training of GAN. One solution to overcome such ordinary defect of GANs, in particular mode collapse, is to contribute multiple adversarial losses in training a single generator through multiple discriminators [31]. Moreover, multi-task learning (MTL) algorithms can extract structure and similarities across different learning problems and force the GAN network to learn multiple tasks jointly [32], [33].

As we have mentioned before, studies investigated the relationship between the image resolution and its Fourier transform details. By contributing Fourier domain information of OCT images in training of the GAN, Fourier components of generated data especially at higher frequencies, as well as, the details of the synthetic data greatly resemble the real data. Here, this is provided by employing dual task-specific discriminators to learn spatial and frequency tasks. In this way, the generator is forced to learn both spatial and Fourier domain information of data to minimize the prediction error of realistic data through dual discriminators. In the training process, after updating the weights of first discriminator, the weights of the generator are modified by second discriminator to improve the generated images. Since generator employs losses of two parallel complementing-task discriminator for updating its parameters at each iteration, it ultimately produces samples with minimum error and higher quality images.

III. Proposed DDFA-GAN

The overall architecture of the DDFA-GAN is depicted in Fig. 1. The proposed model consists of single-generator and dual-discriminator variants that attempt to better approximate max $V(G,D_k)$ where $k=S,F$ indicates the spatial and Fourier domain discriminations. Details of the structure of generator and discriminators are provided in the following. In the training process, the generator takes a random vector z of size 128 and generates the fake OCT B-scan. Then the fake $G(z)$ and the real OCT images (x) are applied to D_S which minimizes the error in predicting fake images produced by the generator through the binary cross-entropy loss.

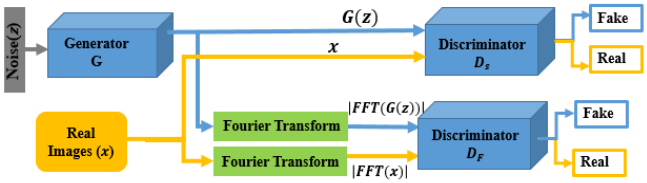

Fig. 1. Block diagram of the proposed DDFA-GAN model. See text for details.

In each iteration of the training process, the absolute value of the Fast Fourier transform (FFT) for both fake and real images are calculated and applied to the second discriminator (D_F) to minimize the binary cross-entropy loss. After several iterations, the DDFA-GAN is trained to generate realistic normal and DME OCT images. The training process is formulated in Algorithm 1.

A. Architecture of Generator

As mentioned above, ordinary GAN networks suffer from instability and non-convergence challenges. For most cases, the bad learning performance originates from the non-stationary characteristics of the training data. In this study, normal distribution is considered for prior distribution of the latent-space and this assumption is rather enforced due to the pre-training step of the generator which is done by taking the advantages of VAEs. That is, the same structure of the decoder part of a pre-trained VAE is used for the generator. Then, the weights of the decoder were transferred to the generator for initialization. The VAE consists of an encoder and a decoder network [34]. The encoder network assumes that the distribution of the data x is normal $\mathcal{N}(0,1)$ and samples the distribution z of the data. The decoder network then reconstructs the data x from the sampled distribution z. The loss of VAE consists of the prior regularization parameter and the reconstruction loss. The regularization parameter controls the distribution of the encoder $\mathcal{N}(0,1)$ through the KullbackLeibler (KL) divergence, and the pixel-wise binary Cross-entropy loss between real images x and reconstructed images $\hat{x}$ guarantees the similarity of x and $\hat{x}$. The VAE is trained using variational inference to approximate the prior distribution. Because the prior distribution is a normal distribution, the approximated posterior distribution would also be normally distributed.

In an ordinary GAN model, the generator tries to minimize errors between distributions of real and generated data. However, the generator does not know the distribution of the data from the beginning of the training and can be trained to approximate any distribution in particular for limited data. The variational decoder make the generative distribution be initialized with approximately the normal distribution. Therefore, the generator is inhibited from converging to a different distribution which prevents mode collapse. The VAE networks are stable and the pre-trained VAE can provide this stability for the GAN network as well [35]. The architecture of the utilized VAE is depicted in Fig. 2. The encoder part consists of two Conv2D layers of size 32×3×3, 64×3×3 and a dense layer of size 16.

Algorithm 1: DDFA-GAN for OCT image synthesis

Input: Training samples $\boldsymbol{S_{train}} := \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N}\}$
Output: Normal and DME synthesized images
For number of iterations do:

Sample mini-batch from Gaussian noise $\boldsymbol{z} := \{z_1, z_2, \ldots, z_N\}, z_i \sim p_z(z)$
Generate data from noise samples $\boldsymbol{x_G} := \{x_{G1}, x_{G2}, \ldots, x_{GN}\}, x_{Gi} \sim p_{G(z)}(G(z))$
Train $\boldsymbol{D_S}$
Sample mini-batch from real data distribution $x := \{x_1, x_2, \ldots, x_N\}, x_i \sim p_{data}(x)$
Sample mini-batch x_G from generated data distribution $x_G := \{x_{1G}, x_{2G}, \ldots, x_{NG}\}, x_{iG} \sim p_{G(z)}(x_G)$
Update the D_S as follows:

$$maximize\ \mathbb{E}_{x\sim p_{data(x)}}[\log(D_S(x))] + \mathbb{E}_{x_G\sim p_{G(z)(G(z))}}[\log(1 - D_S(x_G))]$$

Calculate absolute FFT of each mini-batch x to obtain samples $X := \{X_1, X_2, \ldots, X_N\}, X_i \sim p_{data_{FFT}}(X)$
Calculate absolute FFT of each mini-batch of generated images to obtain samples $X_G := \{X_{1G}, X_{2G}, \ldots, X_{NG}\}$, $X_{iG} \sim p_{G(z)_{FFT}}(X_G)$
Train $\boldsymbol{D_F}$
Sample mini-batch from distribution of Fourier transform of real data $p_{data_{FFT}}(X)$
Sample mini-batch from distribution of Fourier transform of generated data $p_{G(z)_{FFT}}(X_G)$
Update the $D_F$as follows:

$$maximize\ \mathbb{E}_{X\sim p_{data_{FFT}}(X)}[\log(D_F(X))] + \mathbb{E}_{X_G\sim p_{G(z)FFT(X_G)}}[\log(1 - D_F(X_G))]$$

Train Generator
Sample mini-batch from $p_z(z)$
Update the generator as follows:

$$minimize\ \mathbb{E}_{x_G\sim p_{G(z)}(x_G)}[1 - \log(D_S(x_G))]$$
$$minimize\ \mathbb{E}_{X_G\sim p_{G(z)_{FFT}}(X_G))}[1 - \log(D_F(X_G))]$$

End

The architecture of the proposed generator is exactly the same as the decoder part. It consists of a dense layer of size 128, Conv2D of size 32×5×5, 64×5×5 each followed by the Leaky-ReLU (α=0.1) activation functions. These layers are followed by two residual blocks. Residual blocks directly propagate the forward and backward signals from one block to any other block, using the skip connections and after-addition activation. [36]. It has been shown that these blocks accelerate the speed of training, reduce vanishing gradients and increase the performance of the network by learning more effective features of the data while keeping the network structure shallow. This is a very effective advantage in this study where we want to train the lightweight network with limited amount of data. Accordingly, two residual blocks are used each composed of a Conv2D layer of size 64×3×3 and a ReLU activation function following by an element-wise summation.

B. Architecture of D_S

The first discriminator (D_S) compares real and generated OCT images in spatial domain. The architecture of D_S is depicted in Fig. 3. It consists of four blocks of plain convolutional layers with size 128× 5×5, and the last block with size 256×5×5. The first four blocks include a Leaky-ReLU (α=0.2) activation function. A Drop-Out layers with rate 0.3 are utilized in all plain convolutional blocks.

C. Architecture of D_F

The second discriminator (D_F) compares the absolute value of Fourier transform between real and generated OCT images. To increase the speed of learning and to prevent instability, an AE network is first trained utilizing the absolute value of real images' Fourier transforms. During the training, the encoder part of AE extracts low dimensional features and, accordingly, the encoder part of the pre-trained AE entails essential constructive Fourier features of data. The encoder can be frozen and be applied to D_F. The architecture of the pre-trained AE, and D_F are illustrated in Fig. 4. The encoder network of the AE consists of Conv2D with sizes 16×3×3, and 8×3×3 and the decoder part has three Conv2D layers with sizes 8×3×3, 16×3×3, and 1×3×3. Eventually, D_F is composed of the frozen pre-trained encoder followed by two Conv2D layers with size 64×3×3, and 32×3×3, a dense layer with size one and a sigmoid activation function.

IV. Experiments

A. Dataset

The training dataset consisted of OCT retinal images from 50 normal and 50 DME subjects obtained from Heidelberg SD-OCT imaging system. These images were obtained from the Noor Eye Hospital in Tehran, Iran [37], [38]. In this dataset, the lateral and azimuthal resolutions vary in the subjects, but the axial resolution is 3.5μm with a dimension of 8.9×7.4 mm². Therefore, the width of the B-scans varies among 512 or 768; and 19, 25, 31, or 61 B-scans were acquired per volume for different subjects.

B. Implementation of DDFA-GAN

To generate OCT B-scans with the proposed DDFA-GAN, three networks are implemented and trained utilizing the dataset of normal and DME images including 2200 B-scans of size 128 × 128. First, VAE network (Fig. 2) is trained utilizing this dataset. During the training, both binary cross-entropy and KL loss are optimized. The VAE is trained for 200 epochs using batches of size 128.

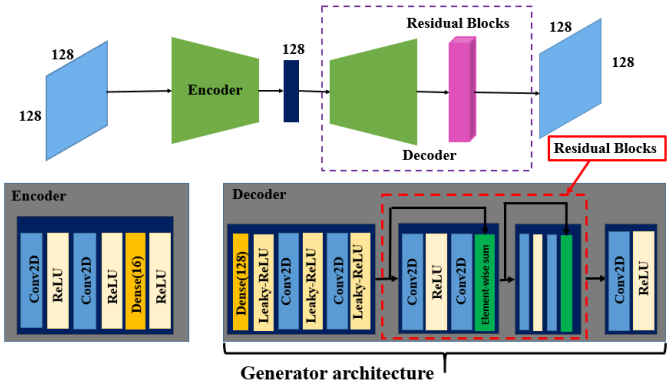

Fig. 2. The architecture of the generator which is the same as the decoder part of the pre-trained VAE.

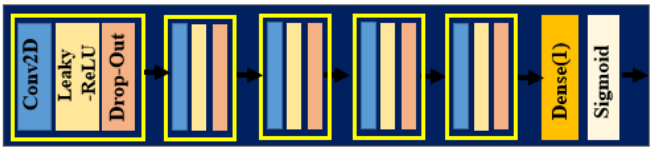

Fig. 3. The architecture of the spatial domain discriminator (D_S).

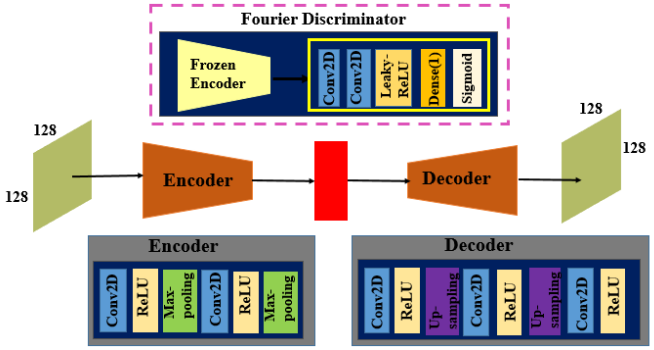

Fig. 4. The architecture of the Fourier discriminator (D_F). The frozen encoder of the pre-trained auto-encoder is fine tuned in the D_F.

The weights of the decoder part of VAE is then applied for initialization of the generator. The second network (the AE network (Fig. 4)) is trained utilizing the absolute value of FFT of the images in the dataset. It is trained for 200 epochs utilizing mean square error (MSE) loss, and batches of size 16. Then, the encoder part of the trained AE is frozen and applied in the D_F. For both VAE and AE, the weights are updated utilizing the Adam optimizer with $lr = 0.001$ and $\beta_1 = 0.9$. Finally, the 3rd network, DDFA-GAN, is trained after initialization of the weights of its generator with the pre-trained decoder of the VAE model. The binary cross-entropy loss is optimized by the Adam optimizer with $lr = 0.002$ and $\beta_1 = 0.5$ for the generator, and both discriminators. The model is trained for 800 epochs, with batches of size 16. In addition, label smoothing is applied to improve the performance of the network by preventing the discriminators from enforcing large gradients to the generator. All the models are implemented in Python, TensorFlow backend utilizing 12 GB NVIDIA K80 GPU, 324 MHz memory clock and it took 13 hours to completely train the DDFA-GAN (the source code of this method will be publicly available at https://github.com/ OCT-synthesis/DDFA-GAN)

C. Results

To evaluate the performance of the DDFA-GAN, it is implemented to generate synthetic data. To perform robust visualization, principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) scatter plots of real and synthetic data generated by DDFA-GAN are illustrated in Fig. 5. The generated data is following the pattern of original data closely. This indicates the ability of proposed method in generating realistic data despite using limited training data. Furthermore, by inspecting the t-SNE grid, it is observed that the distribution of generated data is not concentrated in a single region. This indicates that the mode collapse issue does not affect DDFA-GAN and its generator produces diverse and non-repeated outputs. Some samples of normal and DME images generated by DDFA-GAN, compared with results of the DCGAN [10], WGAN-GP [12], and LS-GAN [13] are illustrated in Fig. 6 for visual inspection. It can be seen that the proposed DDFA-GAN generates more realistic images with higher resolutions than the comparing GAN networks.

In order to evaluate whether the generated images of the DDFA-GAN have the same discriminative features as the real data, a simple AE network proposed in [39] is implemented to extract the features. The Euclidean distance between two feature vectors are then calculated [40]. For this purpose, utilizing the DDFA-GAN and mentioned comparing methods, four datasets with 2200 synthetic images are generated. Next, the AE is separately trained for each of the synthesized datasets and the real images. For each of the datasets, the feature vectors of the latent space of the AE with size of 32 ×32×4 are extracted. The Euclidean distances between normalized feature vectors of real images ($\boldsymbol{x}$) and generated images ($\boldsymbol{y}$) are calculated from equation (4) where N denotes the number of the samples in the vector.

$$d(\boldsymbol{x},\boldsymbol{y}) = \left(\sum_{k=1}^{N} |x_k - y_k|^2\right)^{\frac{1}{2}} \quad (4)$$

The calculated Euclidean distances are reported in Table I. According to this table, the DDFA-GAN produces more realistic images than compared method in terms of discriminative features due to lower Euclidean distance. Frechet Inception Distance (FID) is another widely used metric to evaluate the generative models [41]. It highly correlates with the visual quality of images and low FID-scores indicate the superiority of the generative model. The FID score is calculated for real and four mentioned synthetic datasets (Table I). The lower FID-scores of DDFA-GAN indicates the higher quality of its generated images compared to other methods.

In addition, multi-scale structural similarity metric (MS-SSIM) scores of 100 randomly chosen pairs of images within each class (normal, DME) are also computed to assess the diversity of the images in the real data and generated data, as proposed in [16]. Higher (lower) diversity within a class, corresponds to the lower (higher) mean MS-SSIM score. Calculated results are reported in Table I. According to Table I, the DDFA-GAN generated more diverse images than other comparing methods due to lower value of the average MS-SSIM. In addition, the synthesized images have diversity comparable to the training data which is useful in augmentation applications. More visula details related to the layers and textures of generated OCT images by DDFA-GAN and DCGAN are presented in Fig. 7. It can be seen that the DDFA-GAN can produce more information of the layers and their textures.

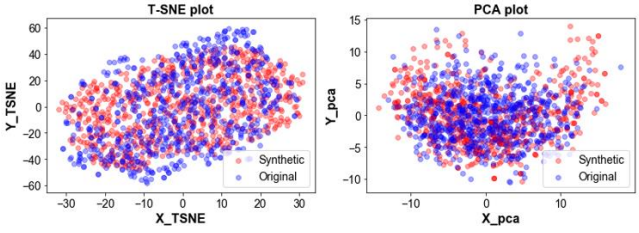

Fig. 5. Two-dimensional grid illustrating the distribution of 1000 real and generated synthetic images after applying t-SNE and PCA.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3189735

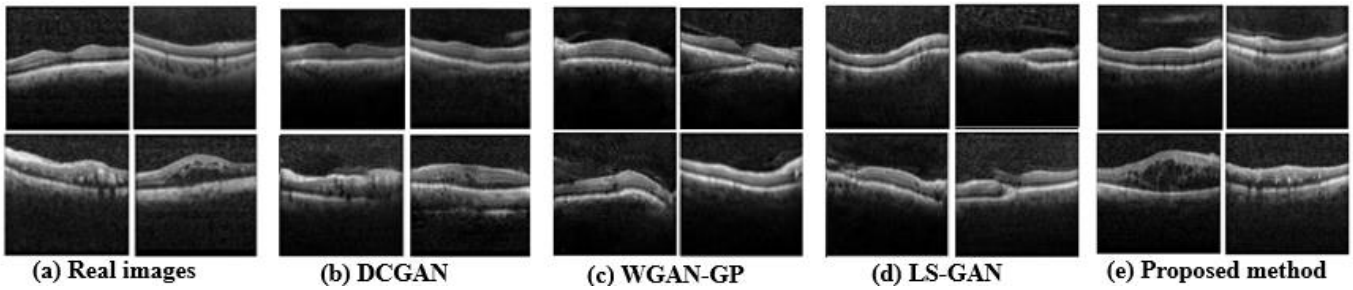

Fig. 6. Illustration of sample normal and DME B-scans, (a) the applied dataset and generated images by (b) DCGAN [10], (c) WGAN-GP [12], (d) LSGAN [13], and (e) the proposed DDFA-GAN. The first row contains normal samples and the second row shows DME samples.

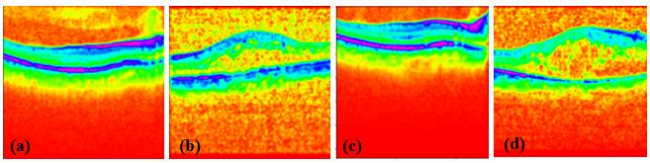

Fig. 7. Examples of two B-scans in Hue, Saturation, Value (HSV) model. (a),(b) real images, (c), (d) DDFA-GAN generated.

Furthermore, the autocorrelation (AC) values of the real dataset and synthetic datasets, generated by DDFA-GAN and comparing methods, were calculated and their cumulative distribution functions (CDF) are depicted in Fig.8 (a). It is evident that the CDFs are identical for real images and those generated by DDFA-GAN, while this is not true for images generated by the other comparing methods. The Kolmogorov Smirnov (K-S) test indicated that the AC of data generated by DDFA-GAN has the same distribution as real data (values of test statistics was less than critical value = 0.014 at the 99% confidence level).

In addition, two individual DDFA-GANs were implemented for normal (DDFA-GAN-N) and DME (DDFA-GAN-DME) images. Evaluation metrics are then calculated to show how separated training dataset yield different generated images. According to this Table I, diversity of generated data in each group is decreased compared to diversity in real images (higher MS-SSIM). However, Euclidean distance for generated images using DDFA-GAN-N has lower value than DDFA-GAN. This implies that DDFA-GAN-N can generate more realistic normal B-scans than DDFA-GAN in terms of discriminative features. Furthermore, the quality of images generated by two separate DDFA-GANs is better than single DDFA-GAN, according to lower FID-scores. This analysis indicates the capability of DDFA-GAN in generating various type of images by being trained individually.

D. Fourier analysis

To analyze the differences between real and generated OCT images in the Fourier transform domain, the Fourier spectrum of the real and generated images and their AC values were calculated. CDFs of AC functions of spectrums of real and generated images are depicted in Fig. 8 (b). In real and generated data by DDFA-GAN, the CDFs are significantly close to each other (values of K-S test statistics are less than critical value = 0.0089 at the 99% confidence level). This underscores the ability of the proposed DDFA-GAN in acquisition of frequency information from real data. Figure 9 shows the sample Fourier spectrum of the real images and generated data using DDFA-GAN and DCGAN. It can be inspected that the periodic structure of the Fourier spectrum in real images is learnt by the DDFA-GAN and preserved during the generation of new images, while this not true for DCGAN.

In Fig. 10, illustration of the horizontal and vertical cuts of the Fourier spectrum (averaged over each dataset of real and generated images) is provided. The averaged components for both cuts overlay in real and generated images by DDFA-GAN in low and high frequencies (Analysis with t-test, P-value >0.05). In addition, the strength of these components decays exponentially (linear in log scale) as the frequency increases for both images. However, the vertical and horizontal cuts of the images generated by other comparing methods are further away from those of real images.

TABLE I
RESULTS OF EVALUATION METRICS COMPUTED FOR GENERATED IMAGES

	MS-SSIM		Euclidean distance	FID score
	Normal	DME		
Real Images	0.17	0.24	0	0
DCGAN [10]	0.33	0.46	0.58	73.20
WGAN-GP [12]	0.58	0.73	0.87	87.31
LS-GAN [13]	0.41	0.55	0.64	82.65
DDFA-GAN-N	0.21	-----	0.32	50.08
DDFA-GAN-DME	-------	0.31	0.41	51.00
DDFA-GAN	0.19	0.27	0.34	51.30

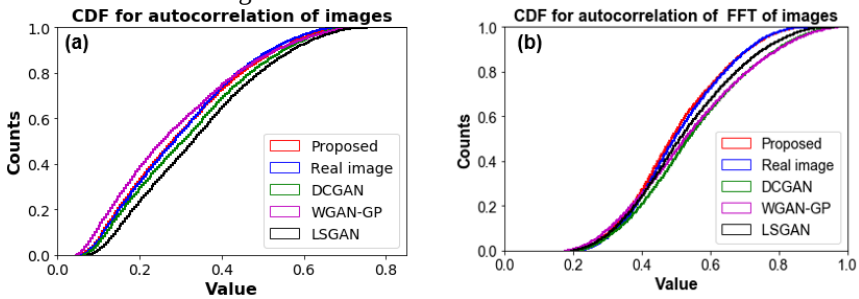

Fig. 8. (a) CDF of AC functions of real and generated images by utilizing DCGAN [10], WGAN-GP, LSGAN, and DDFA-GAN. (b) CDF of AC functions of spectrum of real and generated images by utilizing DCGAN [10], WGAN-GP [12], LSGAN [13], and proposed DDFA-GAN.

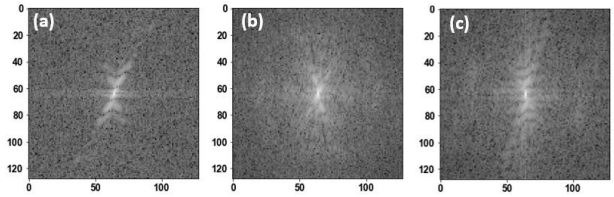

Fig. 9. sample Fourier spectrum of the (a) real image, and generated images by (b) DCGAN [10], (c) proposed DDFA-GAN

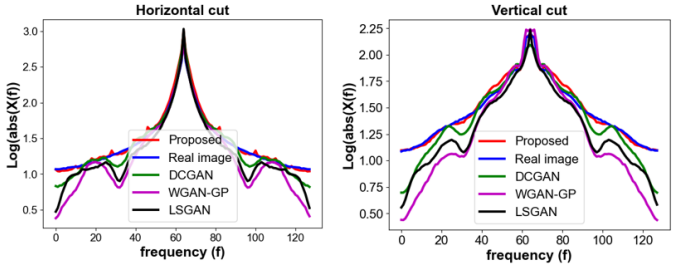

Fig. 10. Averaged log transformed Fourier components of both horizontal and vertical cuts of real image and generated images utilizing DCGAN [10], WGAN-GP [12], LSGAN [13], and DDFA-GAN.

This indicates that by training the frequency information through the proposed DDFA-GAN, distortions in the frequency domain, which is prevalent in the existing methods, are resolved. DDFA-GAN can generate images with similar frequency information, in particular for high frequency components, as real data.

E. Data Augmentation

To confirm the advantages of the synthetic data for augmentation, data generated from the DDFA-GAN is used to augment the new dataset of normal and DME images in [42].Support vector machine (SVM) classifier is trained on 1000 normal and DME images. Grid search is done to find the best parameters of SVM and RBF kernel is selected to achieve the best performance. Before training, PCA is used to remove redundant features and reduce the dimension of each image from 16384 to 100. Generated images are used for data augmentation and several augmentation steps are considered (namely, N_{250}, N_{500}, N_{750}, and N_{1000}). In each step 250 synthetic images are added to the training dataset. For example, N_{250} stands for adding 250 synthetic data to the real dataset. The accuracy and F1-score of the classifier is recorded in each step and reported in Table II. It can be observed from this table that as the number of the synthetic dataset is increased, the performance of the classifier is improved, indicating the effectiveness of generate images for data augmentation.

TABLE II
RESULTS OF EVALUATION METRICS COMPUTED FOR ABLATION STUDY

	Real	**N_{250}**	**N_{500}**	**N_{750}**	**N_{1000}**
Accuracy	86.33	87.73	88.81	90.28	90.16
F1-score	86.50	88.72	89.17	90.13	91.23

F. Ablation study

The key components of the proposed method are: 1) The dual discriminator architecture which provided joint learning in spatial and Fourier domain, 2) The pre-trained VAE which helped with increasing the stability and preventing the mode collapse, 3) The residual blocks which provided more efficient performance of the generator in spite of limited training data. In order to verify the contribution of the mentioned key components, the proposed DDFA-GAN is evaluated with ablation experiments. To this end, three different modes are picked to train the model with 300 epochs. There were 'NO-FFT' which denotes that the model does not contain the FFT module, 'NO-RE' which denotes that the residual blocks are removed, and 'NO-VAE', which denotes that the model does not contain pre-training with VAE. Table III shows the quantitative results of these modes. It can be clearly seen that these three modes obtain relatively higher MS-SSIMs, and FID-scores than DDFA-GAN, demonstrating the effectiveness of key components in design of the DDFA-GAN as a powerful method for OCT image synthesis.

TABLE III
RESULTS OF EVALUATION METRICS COMPUTED FOR AUGMENTATION STUDY

	MS-SSIM	**FID score**
Real image	0.15	0
NO-FFT	0.41	70.81
NO-RE	0.27	63.27
NO-VAE	0.36	61.55
DDFA-GAN	0.18	51.30

V. CONCLUSION

In this study, a dual discriminator GAN network, DDFA-GAN, is proposed which can jointly learn spatial and Fourier domain information of the retinal OCT images. The close relationship between the high frequency components of the synthesized images and their resolution and quality, demonstrate that the proposed DDFA-GAN can generate more realistic OCT images by learning to preserve Fourier domain information. The stability advantage of the pre-trained VAE is used in the generator of the proposed network to increase the stability of the model and to prevent mode collapse. In addition, the residual blocks utilized in the generator provide more efficient performance of the generator in spite of limited training data and a relatively shallow generator structure. The ablation experiments verified the contribution of employed FFT, VAE, and residual blocks in the proposed method.

The PCA, and t-SNE visualization of the generated B-scans demonstrate that that the mode collapse issue does not happen in DDFA-GAN and the generator produced diverse and non-repeated outputs despite the limited amount of the training data. In addition, statistical analysis indicated the significant similarity between the distribution of ACs in generated images and real images. This significant similarity is confirmed in Fourier domain, as well. Results of the qualitative and quantitative evaluation proves that DDFA-GAN can generate images with higher quality, more reality and diversity than existing approaches and is feasible with limited amount of data.

Applying the synthesized images for data augmentation in a classification experiment, it is shown that generated images can be used in clinical applications to improve the performance of machine learning-based CAD systems. In addition, by training DDFA-GAN with images obtained from other imaging modalities or different pathological disorders, the proposed generative model can be extended to generate various synthetic OCT images.

REFERENCES

[1] N. Bhagat, R. A. Grigorian, A. Tutela, and M. A. Zarbin, "Diabetic macular edema: pathogenesis and treatment," *Surv. Ophthalmol.*, vol. 54, no. 1, pp. 1–32, 2009.

[2] D. Wang and L. Wang, "On OCT image classification via deep learning," *IEEE Photonics J.*, vol. 11, no. 5, pp. 1–14, 2019.

[3] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: when to warp?," in *2016 international conference on digital image computing: techniques and applications (DICTA)*, 2016, pp. 1–6.

[4] S. Kuwayama *et al.*, "Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images," *J. Ophthalmol.*, vol. 2019, 2019.

[5] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," *arXiv Prepr. arXiv1711.04340*, 2017.

[6] J. Ma, H. Zhang, Z. Shao, P. Liang, and H. Xu, "GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2020.

[7] T. Zheng, L. Song, J. Wang, W. Teng, X. Xu, and C. Ma, "Data synthesis using dual discriminator conditional generative adversarial networks for imbalanced fault diagnosis of rolling bearings," *Measurement*, vol. 158, p. 107741, 2020.

[8] B. Ma *et al.*, "MRI image synthesis with dual discriminator adversarial learning and difficulty-aware attention mechanism for hippocampal subfields segmentation," *Comput. Med. Imaging Graph.*, vol. 86, p. 101800, 2020.

[9] Y.-C. Liu *et al.*, "Synthesizing new retinal symptom images by multiple generative models," in *Asian Conference on Computer Vision*, 2018, pp. 235–250.

[10] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv Prepr. arXiv1511.06434*, 2015.

[11] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, 2017, pp. 214–223.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," *arXiv Prepr. arXiv1704.00028*, 2017.

[13] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[14] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv Prepr. arXiv1411.1784*, 2014.

[15] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 2180–2188.

[16] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International conference on machine learning*, 2017, pp. 2642–2651.

[17] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv Prepr. arXiv1710.10196*, 2017.

[18] A. Guo, L. Fang, M. Qi, and S. Li, "Unsupervised Denoising of Optical Coherence Tomography Images with Nonlocal-Generative Adversarial Network," *IEEE Trans. Instrum. Meas.*, 2020.

[19] V. Das, S. Dandapat, and P. K. Bora, "Unsupervised super-resolution of OCT images using generative adversarial network for improved age-related macular degeneration diagnosis," *IEEE Sens. J.*, vol. 20, no. 15, pp. 8746–8756, 2020.

[20] S. G. Odaibo, "Generative adversarial networks synthesize realistic OCT images of the retina," *arXiv Prepr. arXiv1902.06676*, 2019.

[21] R. I. Bank, "A project from the American Society of Retina Specialists." 2018.

[22] R. T. Yanagihara, C. S. Lee, D. Shu, W. Ting, and A. Y. Lee, "Methodological Challenges of Deep Learning in Optical Coherence Tomography for Retinal Diseases : A Review," *Assoc. Res. Vis. Ophthalmol.*, vol. 9, no. 2, p. 11, 2020.

[23] P. M. Burlina, N. Joshi, K. D. Pacheco, T. Y. A. Liu, and N. M. Bressler, "Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration," *JAMA Ophthalmol.*, vol. 137, no. 3, pp. 258–264, 2019.

[24] C. Zheng *et al.*, "Assessment of generative adversarial networks model for synthetic optical coherence tomography images of retinal disorders," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 29, 2020.

[25] S. Sengupta, A. Athwale, T. Gulati, J. Zelek, and V. Lakshminarayanan, "FunSyn-Net: enhanced residual variational auto-encoder and image-to-image translation network for fundus image synthesis," in *Medical Imaging 2020: Image Processing*, 2020, vol. 11313, p. 113132M.

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

[27] H. Singh, S. S. Saini, and V. Lakshminarayanan, "Real or fake? Fourier analysis of generative adversarial network fundus images," in *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*, 2021, vol. 11601, p. 116010H.

[28] R. Mizutani *et al.*, "A method for estimating spatial resolution of real image in the Fourier domain," *J. Microsc.*, vol. 261, no. 1, pp. 57–66, 2016.

[29] H. Wang, "Understanding low resolution facial image from image formation model." University of Twente, 2018.

[30] I. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[31] C. Hardy, E. Le Merrer, and B. Sericola, "Md-gan: Multi-discriminator generative adversarial networks for distributed datasets," in *2019 IEEE international parallel and distributed processing symposium (IPDPS)*, 2019, pp. 866–877.

[32] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, 2021.

[33] J. Oh and M. Kim, "PeaceGAN: A GAN-based Multi-Task Learning Method for SAR Target Image Generation with a Pose Estimator and an Auxiliary Classifier," *arXiv Prepr. arXiv2103.15469*, 2021.

[34] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv Prepr. arXiv1312.6114*, 2013.

[35] J.-Y. Lee and S.-I. Choi, "Improvement of Learning Stability of Generative Adversarial Network Using Variational Learning," *Appl. Sci.*, vol. 10, no. 13, p. 4528, 2020.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, 2016, pp. 630–645.

[37] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT classification using a multi-scale convolutional neural network ensemble," *IEEE Trans. Med. Imaging*, vol. 37, no. 4, pp. 1024–1034, 2017.

[38] "DATASET." [Online]. Available: https://misp.mui.ac.ir/en/dataset-oct-classification-50-normal-48-amd-50-dme-0.

[39] M. Tajmirriahi, R. Kafieh, Z. Amini, and H. Rabbani, "A Lightweight Mimic Convolutional Auto-encoder for Denoising Retinal Optical Coherence Tomography Images," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[40] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi, "The similarity metric," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3250–3264, 2004.

[41] A. Borji, "Pros and cons of gan evaluation measures," *Comput. Vis. Image Underst.*, vol. 179, pp. 41–65, 2019.

[42] P. P. Srinivasan *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express*, vol. 5, no. 10, pp. 3568–3577, 2014.