IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 June 2022, accepted 10 July 2022, date of publication 25 July 2022, date of current version 5 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3193668

RESEARCH ARTICLE

DEEPFAKE Image Synthesis for Data Augmentation

NAWAF WAQAS[1], SAIRUL IZWAN SAFIE[1], KUSHSAIRY ABDUL KADIR[2], (Senior Member, IEEE), SHEROZ KHAN[3], AND MUHAMMAD HARIS KAKA KHEL[2]

[1]Department of Instrumentation and Control Engineering, Universiti Kuala Lumpur Malaysian Institute of Industrial Technology, Kuala Lumpur 81750, Malaysia
[2]Electronic Section, British Malaysian Institute, Universiti Kuala Lumpur, Jalan Sungai Pusu, Kuala Lumpur, Selangor 53100, Malaysia
[3]Department of Electrical Engineering, Onaizah College of Engineering and Information Technology, Qassim 2053, Saudi Arabia

Corresponding authors: Nawaf Waqas (nawaf.waqas@s.unikl.edu.my) and Sairul Izwan Safie (sairulizwan@unikl.edu.my)

ABSTRACT Field of medical imaging is scarce in terms of a dataset that is reliable and extensive enough to train distinct supervised deep learning models. One way to tackle this problem is to use a Generative Adversarial Network to synthesize DEEPFAKE images to augment the data. DEEPFAKE refers to the transfer of important features from the source image (or video) to the target image (or video), such that the target modality appears to animate the source almost close to reality. In the past decade, medical image processing has made significant advances using the latest state-of-art-methods of deep learning techniques. Supervised deep learning models produce super-human results with the help of huge amount of dataset in a variety of medical image processing and deep learning applications. DEEPFAKE images can be a useful in various applications like translating to different useful and sometimes malicious modalities, unbalanced datasets or increasing the amount of datasets. In this paper the data scarcity has been addressed by using Progressive Growing Generative Adversarial Networks (PGGAN). However, PGGAN consists of convolution layer that suffers from the training-related issues. PGGAN requires a large number of convolution layers in order to obtain high-resolution image training, which makes training a difficult task. In this work, a subjective self-attention layer has been added before 256×256 convolution layer for efficient feature learning and the use of spectral normalization in the discriminator and pixel normalization in the generator for training stabilization - the two tasks resulting into what is referred to as Enhanced-GAN. The performance of Enhanced-GAN is compared to PGGAN performance using the parameters of AM Score and Mode Score. In addition, the strength of Enhanced-GAN and PGGAN synthesized data is evaluated using the U-net supervised deep learning model for segmentation tasks. Dice Coefficient metrics show that U-net trained on Enhanced-GAN DEEPFAKE data optimized with real data performs better than PGGAN DEEPFAKE data with real data.

INDEX TERMS DEEPFAKE, PGGAN, self-attention layer, spectral normalization, unbalanced dataset.

I. INTRODUCTION

Recently we have seen a rise in DEEPFAKE data in every domain. DEEPFAKE data are sometimes used for good but can also be used for bad purposes that mostly impacts the social aspect of life. What makes DEEPFAKE so important today is their low barrier to entry, which means that readily available tools and models can be used by people with moderate programming skills to create highly realistic fake data. When this is taken into account in the context of image processing domains like Segmentation, Detection or Reconstruction, the impact of DEEPFAKE data could be quite significant.

Existing supervised learning methods for image segmentation rely heavily on a large amount of high-quality training data. The problem becomes apparent with the resurgence

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

of deep learning, whose training requires huge volume of labelled data. To build-up large scale training datasets is a daunting task for most image analysis researchers due to the enormous financial costs and expert label time involved. Meanwhile, traditional augmentation methods such as scaling, rotating, flipping and elastic deformation, especially in the case of medical image synthesis, fail to account for capturing minute featured differences resulting from size, shape, location and appearance of specific pathology [1].

With updates of the General Data Protection Regulation (GDPR) regulations in the EU, the free flow of data has been restricted to ensure patient privacy and anonymity [2]. Even anonymised or de-identified data must not be shared between research groups in different countries, because of the combination of some variables in an anonymized dataset may allow for individual identification [3]. For example, knowing the zip code, birthday and sex is enough to identify 87% of US citizens [4]. The European GDPR rules are more strict than those in the US Health Insurance Portability and Accountability Act (HIPAA) rules for health data exchange [5]. EU demands that health data protection in a third country is essentially equivalent to that in the EU, which is not the case with the US HIPAA system [6]. All health data transfers require to ensure that informed consent is received from each individual, which makes most transatlantic collaboration impossible, if not planned in advance.

Generating realistic synthetic DEEPFAKE data is an alternative solution to the privacy issue. Synthetic DEEPFAKE data should contain all the desired characteristics of a given dataset, but without any sensitive content, which makes it impossible to identify individuals. Therefore, properly manufactured synthetic DEEPFAKE data is a solution to the privacy problem that allows data to be shared between research groups.

The Generative Adversarial Network (GAN) [7] is a robust and unsupervised training approach. GANs have made remarkable progress in the domain of DEEPFAKE images [8] and DEEPFAKE voice syntheses [9]. They learn the pattern of the input samples and generate new DEEPFAKE based on the basic structural information in the training data. As a result, GANs are very useful for deep superimposing fake images. Prospectively, DEEPFAKE synthetic image-based augmentation provides a solution to the lack of manually annotated data and the inflexibility of traditional augmentation.

In the past decades, the field of medical imaging has seen improved in performance with a small dataset. This is made possible due to the prevailing prior knowledge in Deep Neural Network [10]. The U-net architecture is appreciated for bio-medical segmentation images that has shown how powerful the use of DEEPFAKE synthesized data is to overcome the deficit of the small amount of quantity training data available to train deep neural networks [11]. The authors in [12] have used DCGAN as DEEPFAKE synthesized model making 64×64 liver lesion (or ulcer) Region of Interest (ROI) to improve the classification performance of the model into three categories [13]. In [14], the authors have used DCGAN to augment the data for segmentation purposes to synthesize DEEPFAKE lung field data of cardiac images and their corresponding masks. The problem with using DCGAN is that it can only synthesize low-resolution images, which makes defining the region of interest very challenging. The authors have proposed GAN for synthesizing DEEPFAKE high resolution images of retinal fungi with their corresponding mask. The authors have compared the performance of these DEEPFAKE synthesized dataset trained on segmentation model dataset with that of real dataset trained on segmentation model [15]. The authors have proposed in [16] the pix2pix architecture which is the GAN architecture used for image-to-image translation. They have used pix2pix for translating simple brain MRI images into images of DEEPFAKE brain tumor images for augmentation purposes [17]. The authors have proposed PGGAN in [18] an architecture to synthesize DEEPFAKE 256×256 brain MR images with and without tumor for improving the detection task [19]. Beers *et al.* have proposed PGGAN architecture for the synthesis of DEEPFAKE MR images of reticulocytes and their vessel maps [20]. Unlike Han *et al.* and Beers *et al.*, in this paper we have proposed adding the self-attention layer to the PGGAN architecture along with spectral normalization [30] in the PGGAN's discriminator alongside with pixel normalization in PGGAN's generator to synthesize the 256×256 realistic DEEPFAKE knee MR images. Zhang *et al.* have proposed GAN architecture to which self-attention layer with spectral normalization has been applied and has achieved state-of-the-art results. This model aggregates images with a resolution of 128×128 which is one of the limitations [21], and is therefore not suitable to support the ROI. BIGGAN proposed by Brock *et al.* is another model in which the self-attention layer is imposed with convolution layers and have achieved the best results using high-resolution images, but the only problem with BIGGAN is that it is computationally very expensive [22].

As illustrated in Figure 1, the structure of human knee is composed of multiple types of musculoskeletal tissue, consisting of three cartilage and corresponding bone components that make up the overall structure of the knee. Their anatomical geometry changes significantly across the image slices [23]. An effective medical frame image synthesis that can maintain training stability and capture the minute details of irregular knee structure poses a major challenge that has never been encountered before. Hence, a novel DEEPFAKE image synthesis of the knee profile via hierarchical framework has been designed and proposed. In summary, the main contributions of this paper include:

1. To propose an Enhanced-GAN, which is capable to generate DEEPFAKE knee images at 256×256 resolution of realistic DEEPFAKE knee images. The use of self-attention layer enables Enhanced-GAN to learn the finest features and patterns from given data. Moreover, spectral normalization is also used to further improve

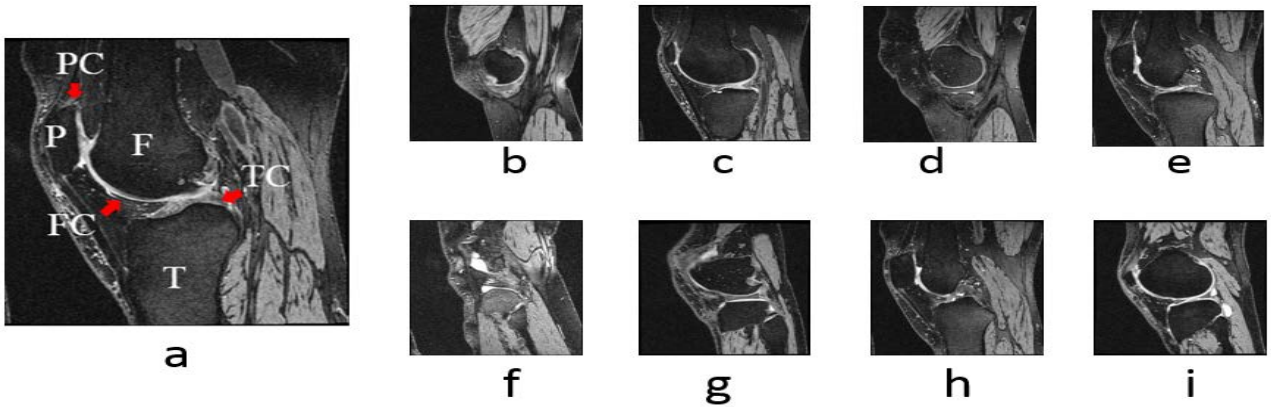

FIGURE 1. Knee structure with Patella (P), Femur (F), Tibia (T), Patellar cartilage (PC), Femoral cartilage (FC) and Tibial cartilage (TC) (a). Knee structure components such as knee bone, cartilage, muscles and ligaments have changing anatomical geometry given in (b) to (i).

stability during training of high-resolution images with Enhanced-GAN.

2. Instead of the Inception Score and Frechet Inception Distance applied to most assessment of natural images, probability-based Mode Score and AM Score are adopted to assess the performance of the proposed Enhanced-GAN framework.
3. The DEEPFAKE data from Enhanced-GAN is used as augmentation data for the knee cartilage segmentation task using U-net and to compare its performance with PGGAN DEEPFAKE synthesized data using the dice-coefficient [24].

II. MATERIALS AND METHODS

A. IMAGE DATASETS

The study has comprised of 75 normal knee image datasets. MR image data has been acquired by using 3.0 Tesla (T) MRI Scanner (Siemens Magnetom Trio, Erlangen, Germany) with quadrature transmit-receive knee coil (USA Instruments, Aurora, OH). Dual Echo Steady State (DESS) with Water Excitation (WE) imaging sequence was selected. All knee image datasets have been chosen randomly from the Osteoarthritis Initiative (OAI) database. The images have section thickness of 0.7 mm and an in-plane resolution of $0.365 \times 0.365\ mm^2$ (field of view $= 140 \times 140$ mm, flip angle $= 25°$, TR/TE $= 16.3/4.7$ msec, matrix size $= 384 \times 384$mm, bandwidth $= 185$ Hz/pixel). Additionally these datasets were also annotated using Slicer Software [25] in which FC, TC and PC were labeled.

B. ENHANCED-GAN

In order to improve the DEEPFAKE data synthesis task, this paper suggests optimizing it as another variant of PGGAN. The PGGAN is enhanced by adding a self-attention layer and spectral normalization to improve stability during training by capturing the finest and detailed features and patterns in high resolution images. It defines to identify class preserving variations to generate valid and representative samples from knee images with its respective segmentation masks. These samples can be used by both academics and industrial researchers to augment data from prospective medical imaging of the knee. Thus, this improvement of PGGAN proves to be the solution to the medical image data augmentation problem.

Initially, the discriminator of the PGGAN model is trained at a relatively low resolution of 8×8 layers, which is then gradually evolved with 2^n which means resolution of each layers will grow according to this formula 2^n from 8×8 to 16×16 onward to 32×32 and so on, until it settles reaching the 256×256 layer. The input data with fixed size low resolution are centrally cropped to get the above mentioned resolutions. This gradual expansion of the model makes it easier for the layers to learn different variants, styles and classes of an image. The model does not require the layers to learn how to draw a linear vector of different size of the images, either 256×256 or 512×512, and instead starts learning gradually with lower resolution images starting at 8×8, 16×16, 32×32 to reach a resolution of 256×256.

The fadeout block in the PGGAN helps to smooth out the process of up-scaling across the image dimensions during training at each resolution while using the WGAN loss function [26]. After scaling, a new layer is created by merging the weights of the previous layer as input and grouping them together using the weighted sum [27]. Then a self-attention layer is added to capture the long-range dependencies of the image to generate an accurate image with a large number of categories. In addition, the training process is normalized by adding spectral normalization in the discriminator to improve the training stability and image generation quality of PGGAN. An overview of PGGAN and Enhanced-GAN is as shown in Figure 2 and Figure 3 respectively.

C. LEARNING FINEST FEATURE USING ATTENTION

In most of high resolution images research, PGGAN has been used in medical image synthesis tasks [28]. They have used convolution layers to create PGGAN model. Unfortunately, there is a problem with the convolution layer architecture.

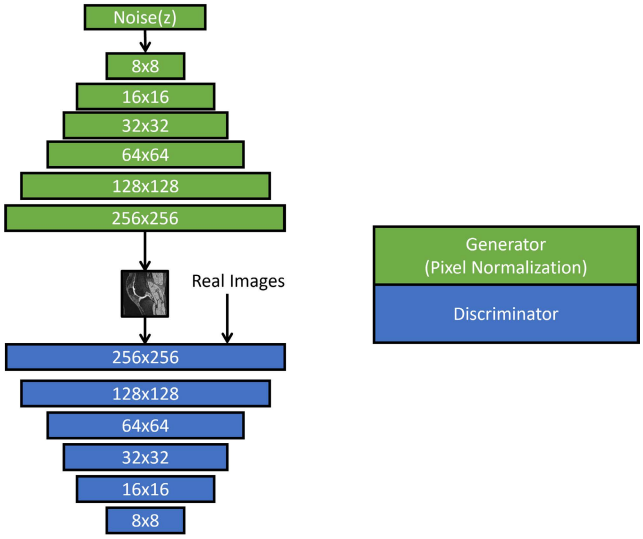

FIGURE 2. Overview of PGGAN.

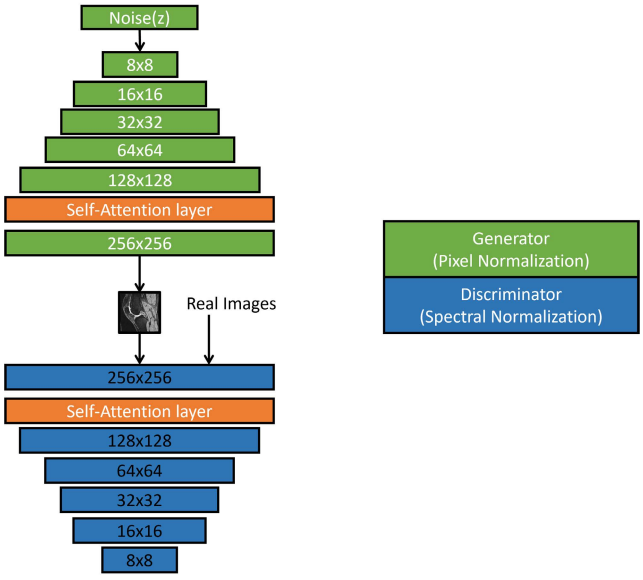

FIGURE 3. Overview of Enhanced PGGAN.

Convolution layers process data in a local neighbourhood which is a big drawback of convolution layers because it makes the model learning difficult and computationally expensive with images that contain long-range of classes.

Accordingly, we rely on adding the self-attention block by the authors of [29] in PGGAN framework which enables the generator to synthesize an image in which specific details at different locations are carefully coordinated with similar features in the far away portion of the image. Furthermore, it allows the discriminator to distinguish those regions of the image, and to filter the response of the feature to maintain only the activation relevant to the specific task. Therefore, in our Enhanced-GAN, details can be generated from the features in the image. Besides this, the discriminator is capable to identify the highly detailed features in distinct and those areas of the image that are consistent. Figure 4 illustrates the architecture of self-attention layer or block.

Where we define $\mathbf{x} = \{x\}_{\mathbf{i=1}}^{N_l}$ as the feature maps obtained in the previous convolution layer. The feature maps feed as input is translated into two feature spaces **f**, **g** to compute the attention, where $f(x) = W_f x$, $g(x) = W_g x$

$$\beta_{j,i} = \frac{exps_{ij}}{\sum_{i=1}^{N} exp\left(s_{ij}\right)} \quad \{where\ s_{ij} = f(x_i)^T g(x_j) \tag{1}$$

where $\beta(j, i)$ is an attention map that indicates the extent to which the model attends to the **$_i$th** location when synthesizing the **$_j$th** region. The output of the self-attention layer is defined as:

$$o_j = \sum_{i=1}^{N} H\left(x_i\right) \beta_{j,i} \quad \{where\ H(x_i) = W_h x_i \tag{2}$$

In the above equations, W_f, W_g, and W_h are the weight matrices of the 1×1 convolutional layer. To enable the generator to learn the local dependency of the image as well as the global long-range dependence, we multiply the output of the self-attention layer o_j, which is the weight coefficient γ and add it to the input feature map x_i to obtain the final output of the module of the self-attention y_i as given in Equation (3).

$$y_i = \gamma o_j + x_i \tag{3}$$

D. TRAINING STABILITY VIA SPECTRAL NORMALIZATION

The Spectral Normalization has been adopted to stabilize GAN training embedded in the discriminator block of GAN [30]. Spectral Normalization takes advantage of the spectral criterion on the discriminator block parameter matrix. Thus, the network satisfies the Lipshitz constraints, therefore smoothing the Discrimator block parameter to achieve stabilized training.

Spectral Normalization initializes random vector of real values in the beginning. Then for each update and each layer, it performs three tasks: first, it normalizes weights by applying the power iteration method; secondly, it computes the spectral norm, and third, it updates the weight using stochastic gradient descent on mini batch dataset. This enables to train the model to be stable and smooth until it converges.

III. RESULTS

A. EXPERIMENTAL SETTINGS

In this experiment, the Enhanced-GAN framework is compared with state-of-art PGGAN. A total of 30 dataset or 4800 training slices or images have been used. Enhanced-GAN performance is evaluated with two performance metrics as:

1) AM Score.
2) Mode Score

The AM Score is based on the idea to evaluate any sort of GAN synthetic data on the classifier other than pre-trained on Imagenet as proposed in [31]. The pre-trained Imagenet

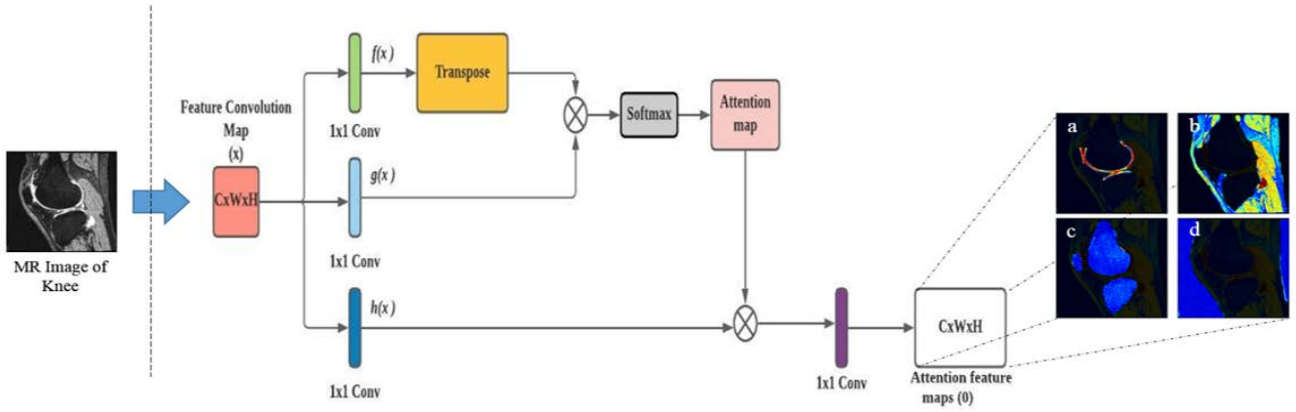

FIGURE 4. Diagram of attention mechanism within the Enhanced-GAN architecture. The attention feature convolution map is a result of matrix multiplication between the attention map, $\beta(j, i)$ and third feature space, h(x) and 1 x 1 convolution filter.

classifiers used by Inception Score for evaluation. Inception Score works fine for natural images but fails on the evaluation of medical images. Thus, we have trained the classifier on 30 knee images dataset in which each dataset containing 160 slices, and each slice represents a single class. Mathematically, the AM score can be represented as given in Equation (4):

$$KL(p(y^*)||p(y)) + E_x H(y/x) \tag{4}$$

where $p(y^*)$ is the empirical label distribution derived from the training data, and p(y) represents marginal distribution, and H(y/x) represents the entropy of the predicted class label for sample image x. The AM score primarily calculates the image quality value, and a low AM score value indicates that the image is of good quality [32].

We have evaluated the same classifier features with mode score computation. Mode score can quantify the difference between real data and generated data using term Kullback–Leibler (KL) [33]. Mode score calculates two aspects of samples, quality and diversified variety. Mathematically Mode score can be expressed as given in Equation (5):

$$Exp(E_x KL(p(y|x)||p(y)) - KL(p(y^*)||p(y))) \tag{5}$$

where p(y|x) is a classifier output trained on knee real images. High Mode score indicates high quality of images and more versatility in synthetic images.

B. EVALUATION OF SYNTHETIC IMAGE QUALITY

The assessment of realism between real knee images and DEEPFAKE is presented in Table 1. The results suggested both PGGAN and Enhanced-GAN have been improved via training from 8 × 8 up to 256 × 256. Table 1 shows a comparison after training for each layer with best mode and AM scores recorded of every layer. In the 8 × 8 layer, a low AM score of 1.4632 is recorded which reached to AM score of 3.0349 for 128 × 128 layer image samples. The deployment of self-attention layer after the 128 × 128 layer reduced the AM score for the 256 × 256 layer to decrease instead of increasing it because with increasing resolution, the evaluation metrics evaluating synthetic quality also become worse [34]. Thus, the self attention layer has made it possible to obtain additional information in image that the convolution layer in the PGGAN has missed.

The images AM score and Mode Score generated during different iterations of training the 256 × 256 and 128 × 128 layers are given in Table 2 and Table 3. In these iterations, the AM and Mode Score are recorded to help determine the valuable information acquired by these images samples as visible to human eye. In addition, the self-attention layer is implemented between the 128 × 128 and 256 × 256 layers. The use of self-attention layer in this position is due to its better performance on the high resolution images. Using Self-attention layer in low resolution layers gives almost equal results like convolution layer. These AM and Mode Score of the synthetic images are compared with PGGAN in Table 2.

It can be noticed from the values in Table 3 that our model performs better than PGGAN. For good synthetic images, the AM score should be low and the Mode score should be high. Our model has an AM score of 3.0100, and the Mode score is 0.9950 while PGGAN has AM score of 3.2654 and Mode score is 0.9843 for images with a resolution of 128 × 128 at 190K iteration, outperforming the PGGAN model. At 114K our model tends towards mode collapse as its Mode Score worsened compared to PGGAN. This means that the varieties in generated samples are lower those at 114K iteration for Enhanced-GAN than for the PGGAN model. However, at the level of 152K iterations the Mode Score is recovered.

Table 3 shows the training results from our Enhanced-GAN when compared with those of the PGGAN model using the AM score and Mode Score for the 256 × 256 resolution layer, and our proposed model clearly outperforms PGGAN at different iteration levels. We have compared the AM Score and Mode scores of the 256 × 256 layer of our model with PGGAN at different iteration levels. To get good synthetic images, the AM score should be low and the Mode score

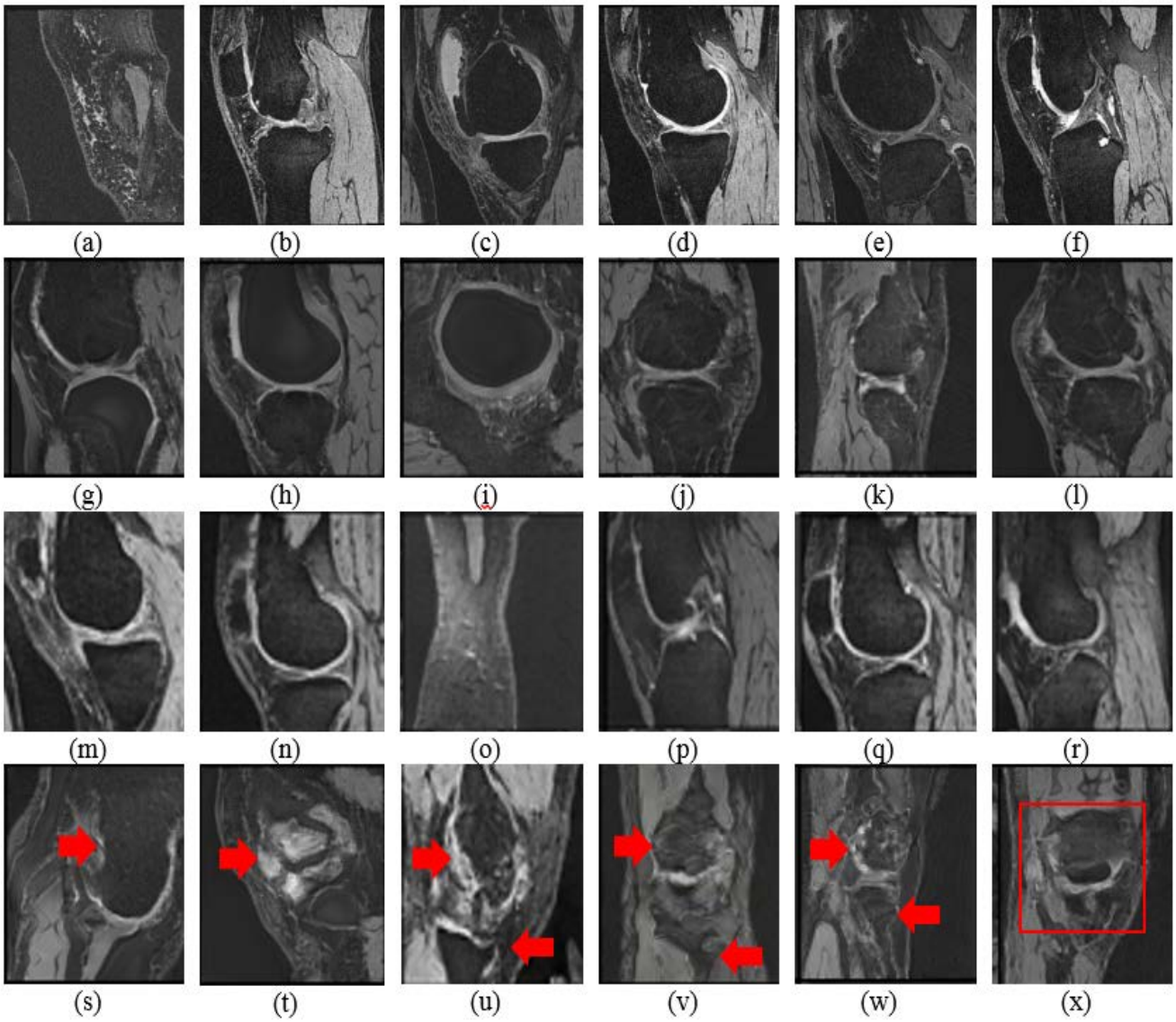

FIGURE 5. Comparison of real MR image of knee (a) to (f) against synthetic knee image at 128 × 128 scale (g) to (l) and 256 × 256 scale (m) to (r). Failure cases were indicated by red arrows in PGGAN (s) to (w) and by red box in Enhanced-GAN (x).

should be high, for our model at every iteration it outperforms PGGAN, for example, the proposed model AM score is 3.0667 and Mode score is 0.9641 with 228K iterations which is better than PGGAN results for 256 × 256 resolution images at iteration of 228K.

Table 2 and Table 3 are utilizing AM and Mode Score for different layers to compare with similar results for the Enhanced-GAN and PGGAN during training while Table 1 shows the comparison after training. It can be seen clearly from Table 1, that AM score increases horizontally layer by layer for the 8 × 8 layer images sample. In this layer, a low AM score of 1.4632 is recorded which reaches an AM score of 3.0349 for the 128 × 128 layer image samples. The deployment of self-attention layer after 128 × 128 layer reduced the AM score of 256 × 256 layer instead of increasing it. Thus, the self attention layer has made it possible to obtain additional information in the image that were missed by the convolution layer in PGGAN [34].

On the other hand, the Mode Score depends not only on the quality of the image but also dependent on its diversity of variety. The score fluctuates strangely in Table 1 at different layers. But the model at the desired layer of 256 × 256 layer achieves higher value producing quality results with more variety.

C. EFFECT OF DEEPFAKE ON SEGMENTATION MODEL

The U-net [35] is tuned to serve as a segmentation model to simulate the diagnostic decision-making process on segmentation of different cartilage boundaries of the selected knee images. The U-net model is trained on our real dataset and DEEPFAKE synthesized dataset of Enhanced-GAN and

TABLE 1. Best AM scores and Mode Scores For every layer of Enhanced-GAN and PGGAN of synthetic images.

Models	Layers	8x8	16x16	32x32	64x64	128x128	256x256
PGGAN	AM	1.8005	1.8901	2.5456	2.9053	3.0357	3.0490
	Mode Score	0.9545	0.9762	0.8125	0.8382	0.9220	1.1009
Enhanced-GAN	AM	1.4632	1.7624	2.1915	2.6440	3.0349	3.0250
	Mode Score	0.9769	1.0623	0.8566	0.8680	0.9540	1.1256

TABLE 2. Assessment of image quality between Enhanced-GAN and PGGAN synthetic DEEPFAKE knee images at 128 × 128 layer.

Models	Iteration	38K	76K	114K	152K	190K	228K
PGGAN	AM	3.319062	3.3208	3.2891	3.3513	3.2654	2.6109
	Mode Score	0.8832	0.9815	0.9678	0.8909	0.9843	1.0090
Enhanced-GAN	AM	3.1794	3.0085	3.0741	3.0250	3.0100	2.4190
	Mode Score	1.01123	1.0257	0.9904	1.1256	0.9950	1.3833

TABLE 3. Assessment of image quality between Enhanced-GAN and PGGAN synthetic DEEPFAKE knee images at 256 × 256 layer.

Models	Iteration	38K	76K	114K	152K	190K	228K
PGGAN	AM	3.2190	3.1998	3.1349	3.2047	3.2245	3.2041
	Mode Score	0.7854	0.8417	0.8671	0.8086	0.7754	0.8234
Enhanced-GAN	AM	3.1263	3.0391	3.1248	3.0937	3.0731	3.0667
	Mode Score	0.8054	1.0188	0.8031	0.8266	0.9613	0.9641

TABLE 4. Average dice co-efficient Score of U-net model.

Schemas	Dice-Coefficient
A	0.8357
B	0.8249
C	0.8469

PGGAN, while Dice-Coefficient(DC) [24] is adopted as performance metric.

The masks corresponding to the DEEPFAKE synthesized knee images are generated by PGGAN and by Enhanced-GAN. The DEEPFAKE synthesized knee image is concatenated with its mask channel by channel, resulting in 384 × 384 × 2 instead of 384 × 384 × 1 images. The generated knee images and their corresponding mask samples are as shown in Figure 6. The dice coefficient is applied as assessment metric to evaluate performance of segmentation model as listed in Table 4.

The U-net model has been trained three times, namely: firstly it is trained on the original data denoted as Schema-A, secondly, it is trained on the original data with PGGAN DEEPFAKE synthesized data denoted as Schema-B, and thirdly, it is trained on the original data with DEEPFAKE Enhanced-GAN synthesized data denoted as Schema-C in Table 4. It can be observed that the average dice-coefficient score of Schema-C is better when compared to other schemas. The average dice-coefficient scores have been recorded for unseen test 10 slices or images of knee datasets.

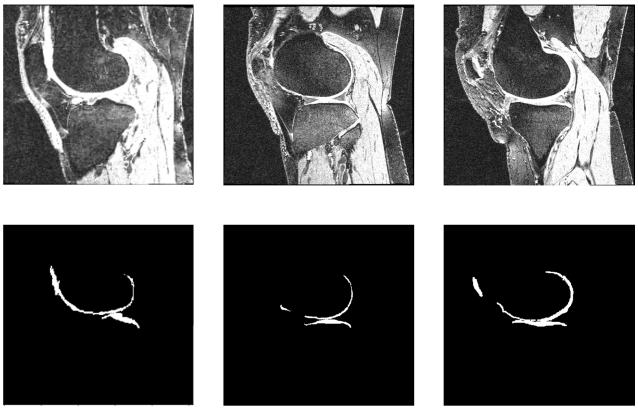

FIGURE 6. Synthesized knee image with its corresponding masks using Enhanced-GAN. Upper Row consist of synthesized Knee image and lower row consists of their Corresponding mask.

IV. DISCUSSION

This is a novel work on GAN-based DEEPFAKE image synthesis framework as it helps in augmentation of data for

supervised deep learning algorithms. The training performance of Enhanced-GAN has been compared with state-of-art PGGAN using relevant scales. Next, we have evaluated the difference in data distribution between Enhanced-GAN and PGGAN knee images using AM and Mode scores. Finally, we have validated the efficacy of enhanced data using a supervised segmentation model and an image of the synthetic knee with annotation at a resolution of 256 × 256. The knee image synthesis is a challenging task due to its complex structure and diverse anatomical geometry [36], [37]. The proposed framework has succeeded in producing realistic DEEPFAKE images through the use of spectral normalization technique. Below, we describe the main lessons learned from this work.

First, synthetic DEEPFAKE knee images are useful for segmentation tasks. One potential application involves diversifying real training data with synthetic data to improve robustness of a supervised deep learning model. Accordingly, three training configurations of real data alone, synthetic DEEPFAKE data only, and real-synthetic DEEPFAKE data combined, have been used to augment the training data of the U-net segmentation model [38]. The same configuration is also present in our study but with slight change. In a single configuration we have used PGGAN Real-synthetic DEEPFAKE images instead of only synthetic DEEPFAKE images and it proves that Enhanced-GAN Real-synthetic DEEPFAKE training images configuration had reported superior performance compared to previous two configurations. Given the growing number of research works on medical image segmentation using supervised deep learning models [39], further investigations based on their findings will benefit the supervised deep learning models.

Second, the optimization of GAN training remains an active research topic with its attractive potential applications. The choice of normalization techniques has a profound effect on the quality of image synthesis. For instance, the use of standalone spectral (or pixel) in Discriminator and spectral (or pixel) in Generator normalization have produced low quality knee images with background blur or minimal contrast as shown in Figure 7, which cannot be adopted into subsequent deep learning segmentation models. Based on the training data during different iterations recorded in Table 2 and Table 3 in this study, it is evident that PGGAN suffers from training instability as traditional GANs tend to experience mode collapse and fading gradient issues during training process. After deploying the improvements to model the training stability and salient feature of Enhanced-GAN manages to successfully avoid mode collapse even at 128 × 128 layer between 114k and 152k iterations.

Third, StyleGAN [40] is an extension of PGGAN. It has generated high-resolution features in natural images. Recently, it has been extended to synthesize CT and MR images [41]. However, the implementation of StyleGAN is limited by its extremely heavy computation. It is useless to apply it to common medical image synthesis. Attention layer has been employed in our Enhanced-GAN framework as alternative. Our quantitative results have suggested that the

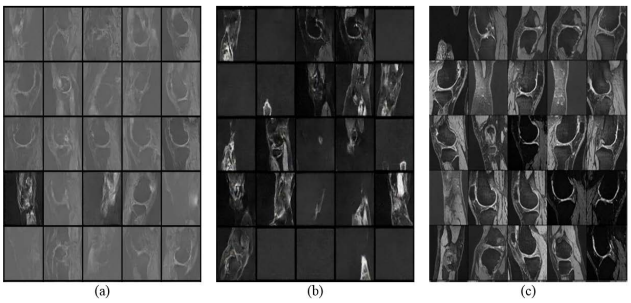

FIGURE 7. Synthetic knee images generated by HieGAN by using different normalization technique configuration in generator (G) and discriminator (D). (a) G: Spectral; D: Spectral, (b) G: Pixelwise; D: Pixelwise, (c) G: Pixelwise; D: Spectral.

images have achieved a high degree of realism, especially at 256 × 256 resolution. The attention layer has successfully guided the discriminator to pay more attention to the various features of knee images in order to compel the generator to create high resolution images with detailed information. Specifically, the overall image brightness is preserved, the boundaries of cartilage-bone interface are well-preserved, the contrast between bone, cartilage and background is made clear, and the anatomical shape and size of cartilage and bone are conserved.

On the other hand, we have detected failure cases from samples generated by PGGAN. As such, PGGAN have produced seriously deformed knee structures wherein the features of the femur and tibia have been altered. Moreover, the boundary between femur and surrounding musculoskeletal tissues is excessively diffused in several samples. The failure samples with severe deformation can potentially mislead the learning of deep learning models. Nonetheless, we also have observed minor irregularity in one sample produced by Enhanced-GAN. The proposed model failed to distinguish between shrinking femur and tibia from the background musculoskeletal tissues. The boundary between knee bones and background is considered blur. These failure cases provide us with valuable insights to improve the model in the future.

CelebA dataset which is the dataset consisting of various celebrity faces as shown in Figure 8, is used as a second dataset. It is also synthesized for DEEPFAKE images of human faces at a resolution of 64 × 64 as shown in Figure 9, and it has been observed that proposed model has synthesized such images which are so real but deepfaked enough that they do not belong to any body in the world and these images can be used to abuse usage [42]. However, on the other hand in medical image analysis these DEEPFAKE images help in applications such as data augmentation [43], reconstruction [44], inter-modalities translation [45]. These deepfaked images have been synthesized by Enhanced-GAN in such a way that it learns the feature from training data or Celebrity dataset and synthesize DEEPFAKE data accordingly. Besides, we have decided to generate the DEEPFAKE human faces image up to 64 × 64 scale in order to better

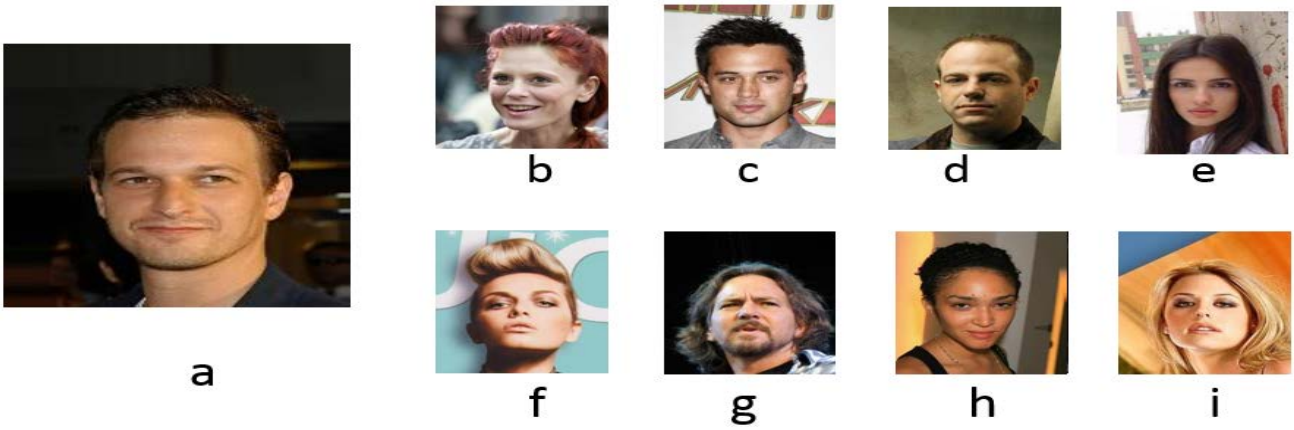

FIGURE 8. Various Celebrities Images from CELEBA dataset.

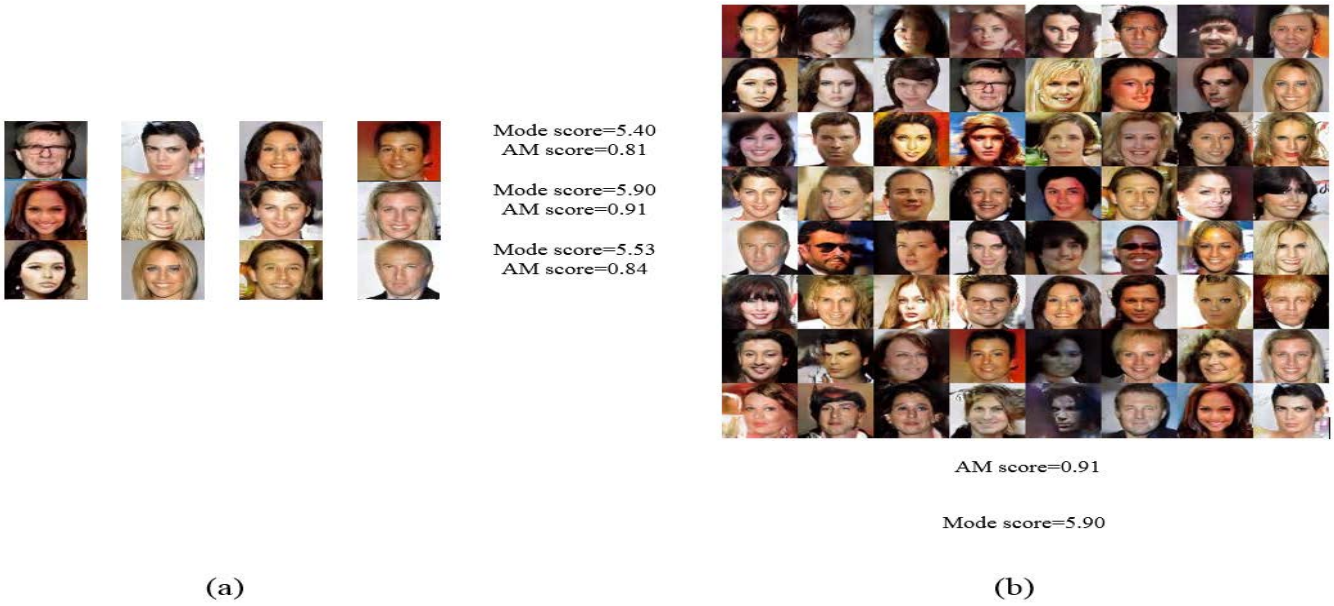

FIGURE 9. (a) Synthetic DEEPFAKE Face images generated by Enhanced-GAN from Different grids having different scores. (b) Each grid Synthetic DEEPFAKE face images from Enhanced-GAN consists of 64 images with 64 × 64 resolution.

understand the balanced results considering the recognition of salient features and acceptable image resolution. In future, we will propose a deep learning model detection algorithm which will help in detecting fake images that can be used for misleading objectives in social life.

V. CONCLUSION

In this paper, a novel method Enhanced-GAN has been capable to generate real-looking and high resolution DEEPFAKE images with perfect class recognition details (cartilage, bone, background, etc) compared to the already available and a widely used PGGAN architecture to generate high resolution images. Enhanced-GAN is developed by incorporating self-attention layer with convolution layer alongside spectral normalization in the discriminator and pixel normalization in the generator. We then evaluate Enhanced-GAN by means of two parameters AM and Mode scores, at 128 and 256 resolution images of Enhanced-GAN have shown to be higher than PGGAN. With a resolution of 128 × 128 during 114K iterations, Enhanced-GAN tends towards mode collapse as its mode score has worsened compared to PGGAN which affects training to some extent but later at 152K iteration it is recovered. Lastly, DEEPFAKE synthesized data from proposed Enhanced-GAN and PGGAN is then used as data augmentation with real data for U-net segmentation model, to prove its performance and evaluate U-net segmentation model using Dice-Coefficient. The score at Schema-C in which real and Enhanced-GAN DEEPFAKE synthesized data has been mixed equally higher compared to Schema-B and Schema-A. In future, generation of synthetic human face images at higher resolution of 256 × 256 and 512 × 512 will be attempted, we have achieved 64 × 64 resolution synthetic

human face images as shown in figure 8 and figure 9 to see how well it works on natural images with three channels. We have already started working on generating high resolution human face images and we will analyze these synthetic DEEPFAKE natural image (Human face) and synthetic DEEPFAKE medical images(Knee MR images) results using DEEPFAKE detecting algorithms based on deep learning and non-deep learning.

ACKNOWLEDGMENT

The authors would like to thank the administrative and technical support they received from UniKL and the use of laboratory facilities and equipment in the experiments.

REFERENCES

[1] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," 2018, *arXiv:1809.07294*.

[2] P. Voigt and A. Von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*, vol. 10, no. 3152676, 1st ed. Cham, Switzerland: Springer, 2017, p. 5555.

[3] Y.-A. de Montjoye, L. Radaelli, V. K. Singh, and A. Pentland, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, 2015.

[4] K. El Emam, E. Jonker, L. Arbuckle, and B. Malin, "A systematic review of re-identification attacks on health data," *PLoS ONE*, vol. 6, no. 12, Dec. 2011, Art. no. e28071.

[5] L. Bradford, M. Aboy, and K. Liddell, "International transfers of health data between the EU and USA: A sector-specific approach for the USA to ensure an 'adequate' level of protection," *J. Law Biosci.*, vol. 7, no. 1, Jul. 2020, Art. no. lsaa055.

[6] D. Hallinan, A. Bernier, A. Cambon-Thomsen, F. P. Crawley, D. Dimitrova, C. B. Medeiros, G. Nilsonne, S. Parker, B. Pickering, and S. Rennes, "International transfers of personal data for health research following Schrems II: A problem in need of a solution," *Eur. J. Human Genet.*, vol. 29, no. 10, pp. 1502–1509, Oct. 2021.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[8] H. Zunair and A. B. Hamza, "Synthesis of COVID-19 chest X-rays using unpaired image-to-image translation," *Social Netw. Anal. Mining*, vol. 11, no. 1, pp. 1–12, Dec. 2021.

[9] S. Liu, S. Li, and H. Cheng, "Towards an end-to-end visual-to-raw-audio generation with GAN," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1299–1312, Mar. 2022.

[10] N. K. Singh and K. Raza, "Medical image generation using generative adversarial networks: A review," in *Health Informatics: A Computational Perspective in Healthcare*. 2021, pp. 77–96.

[11] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2337–2346.

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

[13] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[14] W. Dai, N. Dong, Z. Wang, X. Liang, H. Zhang, and E. P. Xing, "Scan: Structure correcting adversarial network for organ segmentation in chest X-rays," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 263–273.

[15] D. Mahapatra, B. Bozorgtabar, S. Hewavitharanage, and R. Garnavi, "Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Sep. 2017, pp. 382–390.

[16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[17] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *Proc. Int. Workshop Simulation Synth. Med. Imag.* Cham, Switzerland: Springer, Sep. 2017, pp. 1–11.

[18] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*.

[19] C. Han, L. Rundo, R. Araki, Y. Nagano, Y. Furukawa, G. Mauri, H. Nakayama, and H. Hayashi, "Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection," *IEEE Access*, vol. 7, pp. 156966–156977, 2019.

[20] A. Beers, J. Brown, K. Chang, J. P. Campbell, S. Ostmo, M. F. Chiang, and J. Kalpathy-Cramer, "High-resolution medical image synthesis using progressively grown generative adversarial networks," 2018, *arXiv:1805.03144*.

[21] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, May 2019, pp. 7354–7363.

[22] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, *arXiv:1809.11096*.

[23] H.-S. Gan, T.-S. Tan, K. A. Sayuti, A. H. A. Karim, and M. R. A. Kadir, "Multilabel graph based approach for knee cartilage segmentation: Data from the osteoarthritis initiative," in *Proc. IEEE Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2014, pp. 210–213.

[24] A. W. Setiawan, "Image segmentation metrics in skin lesion: Accuracy, sensitivity, specificity, dice coefficient, Jaccard index, and Matthews correlation coefficient," in *Proc. Int. Conf. Comput. Eng., Netw., Intell. Multimedia (CENIM)*, Nov. 2020, pp. 97–102.

[25] S. Pieper, M. Halle, and R. Kikinis, "3D slicer," in *Proc. 2nd IEEE Int. Symp. Biomed. Imag., Macro Nano*, Apr. 2004, pp. 632–635.

[26] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2017, pp. 214–223.

[27] J. R. S. C. Mateo, "Weighted sum method and weighted product method," in *Multi Criteria Analysis in the Renewable Energy Industry*. London, U.K.: Springe, 2012, pp. 19–22.

[28] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.

[29] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[30] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.

[32] Z. Zhou, H. Cai, S. Rong, Y. Song, K. Ren, W. Zhang, Y. Yu, and J. Wang, "Activation maximization generative adversarial nets," 2017, *arXiv:1703.02000*.

[33] Q. Xu, G. Huang, Y. Yuan, C. Guo, Y. Sun, F. Wu, and K. Weinberger, "An empirical study on evaluation metrics of generative adversarial networks," 2018, *arXiv:1806.07755*.

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

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[36] M. S. Harkey, N. Michel, C. Kuenze, R. Fajardo, M. Salzler, J. B. Driban, and I. Hacihaliloglu, "Validating a semi-automated technique for segmenting femoral articular cartilage on ultrasound images," *Cartilage*, vol. 13, no. 2, Apr. 2022, Art. no. 194760352210930.

[37] T. Xu, D. An, Y. Jia, J. Chen, H. Zhong, Y. Ji, Y. Wang, Z. Wang, Q. Wang, Z. Pan, and Y. Yue, "3D joints estimation of human body using part segmentation," *Inf. Sci.*, vol. 603, pp. 1–15, Jul. 2022.

[38] T. Russ, S. Goerttler, A.-K. Schnurr, D. F. Bauer, S. Hatamikia, L. R. Schad, F. G. Zöllner, and K. Chung, "Synthesis of CT images from digital body phantoms using CycleGAN," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 10, pp. 1741–1750, Oct. 2019.

[39] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *J. Digit. Imag.*, vol. 32, no. 4, pp. 582–596, 2019.
[40] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.
[41] L. Fetty, M. Bylund, P. Kuess, G. Heilemann, T. Nyholm, D. Georg, and T. Löfstedt, "Latent space manipulation for high-resolution medical image synthesis via the StyleGAN," *Zeitschrift Medizinische Physik*, vol. 30, no. 4, pp. 305–314, Nov. 2020.
[42] S. Singh, R. Sharma, and A. F. Smeaton, "Using GANs to synthesise minimum training data for deepfake generation," 2020, *arXiv:2011.05421*.
[43] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, "Data augmentation using learned transformations for one-shot medical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8543–8553.
[44] X. Pan, B. Dai, Z. Liu, C. Change Loy, and P. Luo, "Do 2D GANs know 3D shape? Unsupervised 3D shape reconstruction from 2D image GANs," 2020, *arXiv:2011.00844*.
[45] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, "MedGAN: Medical image translation using GANs," *Comput. Med. Imag. Graph.*, vol. 79, Jan. 2020, Art. no. 101684.

NAWAF WAQAS was born in Riyadh, Saudi Arabia, in March 1997. He received the B.Sc. degree in electrical engineering from the National University of Computer and Emerging Sciences, Pakistan, in 2018, and the M.Sc. degree by Research from the British Malaysian Institute (BMI), Gombak, Universiti Kuala Lumpur, Malaysia, in 2021. He is currently pursuing the Ph.D. degree with the Universiti Kuala Lumpur Malaysian Institute of Industrial Technology, Johar Baru, Malaysia. He is a Founder of Nawai-Wadana (Hisara), Charsadda, Khyber Pakhtunkhwa, Pakistan. His research interests include machine learning, deep learning, computer vision, medical image processing, segmentation and image and signal processing, and digital signal processing.

SAIRUL IZWAN SAFIE received the Bachelor of Engineering degree in electrical and the master's degree in electrical engineering (power) from Universiti Teknologi Malaysia, and the Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K. He is currently an Associate Professor with the Universiti Kuala Lumpur Malaysian Institute of Industrial Technology. His research interests include artificial intelligence, big data, signal, and image processing with application to biometric, psychological, and physiological signal, image, and video processing.

KUSHSAIRY ABDUL KADIR (Senior Member, IEEE) received the Bachelor of Science degree in electrical engineering from Case Western University, U.K., the Master of Science degree in information technology from the University of Loughborough, U.K., the M.Sc. degree in mechatronic in information system from International Islamic University Malaysia (IIUM), and the Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, U.K. He is currently the Head of Campus and the Dean of the British Malaysian Institute, Universiti Kuala Lumpur. He has been teaching within the Electronic Technology Department, British Malaysian Institute (BMI), since 2004. His research interest includes image and speech processing.

SHEROZ KHAN (Senior Member, IEEE) was born in Nawai-Wadana, Charsadda, Khyber Pakhtunkhwa, Pakistan. He received the B.Sc. degree in electrical engineering from the NWFP University of Engineering and Technology (UET), Pakistan, the M.Sc. degree in microelectronic & computer engineering from Surrey University, U.K., in 1990, and the Ph.D. degree from Strathclyde University, in 1994, Glasgow, U.K. He worked as a Principal Lecture at UNITEN, from 2000 to 2002, and as an Associate Professor and a Professor with the Department of ECE, International Islamic University Malaysia (IIUM), from 2002 to 2019. He has produced 22 M.Sc. and ten Ph.D.'s, two postdoctorate under his direct supervision while producing eight Ph.D.'s under co-supervision. He has been the PG Coordinator of ECE Department, IIUM, and the Founding Coordinator of the Wireless Communication and Signal Processing Research Group, since 2006. He has been the Co-Founder of ICSIMA, ICISE, ICIRD, and ICETAS. He is the Founder Coordinator of the IIUM-Limoges, France, and IIUM-Schmalkalden USA, Germany programs. Since December 2019, he has been working as a Professor and a Research Coordinator of the Department of Electrical Engineering, Onaizah College of Engineering and Information Technology. He is with QU-IIUM-UnikL Research Team of the KSA MoE RDO grant worth of 1.277M SAR. He is a member of IET and a Chartered Engineer (C.Eng.).

MUHAMMAD HARIS KAKA KHEL was born in Charsadda, Khyber Pakhtunkhwa, Pakistan, in January 1997. He received the B.S. degree in electrical engineering (major in communication) from the University of Engineering and Technology (UET) Peshawar, Pakistan, in 2019. He is currently pursuing the master's degree under the supervision of Dr. Kushsairy Abdul Kadir with Universiti Kuala Lumpur British Malaysian Institute, Malaysia. He is also on research attachment funded by Erasmus funded program by UniKL. His research interests include artificial intelligence (AI), machine learning, deep learning, computer vision, medical image processing, segmentation and image and signal processing, and digital signal processing.