Towards Synthetic Generation of Clinical Rosacea Images with GAN Models

1st Anwesha Mohanty
School of Computing
Dublin City University
Dublin, Ireland
anwesha.mohanty2@mail.dcu.ie

2nd Alistair Sutherland
School of Computing
Dublin City University
Dublin, Ireland
alistair.sutherland@dcu.ie

3rd Marija Bezbradica
School of Computing
Dublin City University
Dublin, Ireland
marija.bezbradica@dcu.ie

4th Hossein Javidnia
School of Computing
Dublin City University
Dublin, Ireland
hossein.javidnia@dcu.ie

2022 33rd Irish Signals and Systems Conference (ISSC) | 978-1-6654-5227-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISSC55427.2022.9826207

***Abstract*—Computer-aided skin disease diagnosis has recently attracted much attention in the scientific and medical research community due to advances in computer vision and machine learning algorithms. These methodologies essentially rely on large datasets collected from hospitals and medical professionals. Data scarcity is a vital problem in the medical domain, especially facial skin conditions, due to privacy concerns. For instance, some facial skin conditions, e.g. Rosacea, require observation of the entire face, which reveals the patient's identity. Rosacea is a lamentably neglected skin condition in the computer-aided diagnosis research community, due to the limited availability of Rosacea datasets. Hence, there is a need for exploring alternative ways to deal with the limited available data for Rosacea. A common approach to expanding small datasets is to utilise augmentation techniques. One of the most powerful augmentation methods in machine learning is Generative Adversarial Networks (GANs). Recently, GANs, principally the variants of StyleGAN, have successfully generated synthetic facial images. In this paper, a small dataset of a particular skin disease, Rosacea, with 300 images is used to examine the potential of a variant of StyleGAN known as StyleGAN2-ADA. The preliminary experiments and evaluations show promising signs towards addressing the data scarcity for computer-aided Rosacea diagnosis.**

***Index Terms*—Skin Disease, Rosacea, Synthetic Face Generation, GANs, Data Augmentation.**

I. Introduction

Computer-aided diagnosis of skin diseases using Deep Learning (DL) models has become popular since the introduction of Inception v3 [1] which achieved an accuracy of 93.33% [2] in classifying various cancerous skin conditions. In the study by Esteva et al. [2] the training of Inception v3 was originally done on approx. 129,450 skin cancer images. However, gathering a dataset this large for every other type of skin disease, such as Rosacea, is not feasible. Many skin conditions including Rosacea can lead to fatal consequences in the long run. However, cancer has been taken most seriously of all, which has resulted in more data being gathered over time. Hence, the availability of skin cancer images in many teledermatology websites are substantial. On the other hand, there is a very limited amount of data for non-fatal chronic skin conditions such as Rosacea.

Rosacea is a common skin condition in native people from northern countries with fair skin or Celtic origins [3]. Facial flushing and redness, inflammatory papules and pustules, telangiectasias, and facial edema are the common signs and symptoms of Rosacea. It is a chronic facial skin condition that goes through a cycle of fading and relapse [4]. The prolonged condition of Rosacea causes cutaneous vascular disorder [5]. The diagnosis of Rosacea demands regular check-ups, medication, and laser treatment if necessary. As an act of caution, early diagnosis of Rosacea can be supported with the state-of-the-art methodologies of deep learning and computer vision.

Deep Convolutional Neural Networks (DCNNs) like Inception v3 classification models have proven to perform efficiently given a large dataset at the training stage. However, the performance of such models deteriorates in applications with scarce and limited data samples. Hence, it is essential to leverage synthetic data generation methodologies to expand the data volume for applications with scarce datasets. The synthetic images may later help in classifying skin conditions as a support to computer-aided disease diagnosis. This paper investigates the potentials of Generative Adversarial Networks (GANs) [6] and particularly a variant of StyleGAN [7] architecture called StyleGAN2-ADA [8] in synthetic facial image generation for the Rosacea skin condition using a limited dataset of 300 full-face images.

II. Related work on Synthetic face generation

The very first facial image generation using GANs was obtained by Goodfellow et al. [6] in 2014. The generated synthetic faces were very noisy and required significant further work. Later, in 2015, DCGAN [9] was presented, capable of generating facial images without any augmentation. DCGAN came with a few notable features such as strong topology of architectures and specific filters to improve image generation resulting in better synthetic faces. However, initial GAN models like the baseline GAN and DCGAN had some limitations such as (i) model instability, (ii) mode collapse, (iii) filter leaking after longer training time, (iv) small resolutions of generated images. Later, the Progressive Growing of GANs

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224.

(ProGANs) [10] introduced a significant advantage in improving the resolution of the generated images with a stable and faster training process. ProGAN has successfully generated images with resolutions from 4×4 to 1024×1024. It utilized the CelebA-HQ dataset [10] with 30,000 real images to generate 1024×1024 synthetic face images. Although ProGAN successfully generated facial images with large resolution, it did not perform adequately in terms of generating realistic features and microstructures.

The introduction of StyleGAN [7] came with further improvements in synthetic image generation, which mainly focused on generating facial images. It paid attention to specific characteristics and phases in facial image generation. StyleGAN introduced a new dataset called Flickr Faces HQ (FFHQ) [7]. FFHQ has 70,000 images with a diverse range of details at 1024×1024 resolution. Sequentially, another variant of StyleGAN was introduced by Karras et al. called StyleGAN2 [11], in which the key focus was on analysis of the latent space. It was effective in rectifying some undesirable artifacts in the generated images. Another set of GAN architectures called BigGAN and BigGAN-deep [12] improved the span of variety and fidelity of the generated images. Although a face image dataset was not used in the evaluation of BigGAN architecture, it brought a lot of freedom in expanding and controlling the sample variety and fidelity during the image generation stage.

The amalgamation of all the ongoing developments in the GAN architectures has brought facial image generation to an excellent state for further applications. Nevertheless, the amount of data used in these GAN models is substantial; as a result, the generated images are high fidelity and realistic. Except for real-world objects and a few other domains where data is easily accessible, it is impossible to acquire a large amount of data for some specific fields/domains. Due to privacy and ethical constraints, it is challenging to acquire images in the healthcare domain, such as medical or clinical images. To improve the handling of limited data in image generation, StyleGAN2-ADA [8] came up with a wide range of flexibility designed to function satisfactorily when only a handful of data is available. StyleGAN2-ADA introduced a dataset that contains 1336 paintings of human portraits called METFACES and another dataset that has 1994 breast cancer histopathology images. StyleGAN2-ADA works reasonably well in generating synthetic images with only 1000 images from FFHQ dataset [7].

Until now, only a few studies have been carried out on Rosacea images using computer vision and deep learning, starting from 2019 onwards. There are a few notable works conducted on Rosacea classification by Thomsen et al. [13], Binol et al [14], Xie et al. [15], Zhao et al. [16], and Zhu et al. [17] with a significant quantity of data collected from dermatology departments of the affiliated hospitals. However, the datasets used in these studies were entirely confidential. In these studies, the early detection problem of Rosacea is addressed by performing image classification among different subtypes of Rosacea and other common skin conditions. Except for classification on partial face images of Rosacea, no studies have been conducted to generate synthetic images for this skin condition. In this paper we use StyleGAN2-ADA [8] because it is able to handle small amounts of data. A small dataset of Rosacea (with 300 images) is discussed in the further sections with experiments and evaluations.

III. Methodology

A. StyleGAN2 with Adaptive Discriminator Augmentation

The hypothesis of StyleGAN2-ADA [8] is that decent outcomes are achievable with a limited number of images, typically a few thousand training images. This ability is supported by the concept of "Adaptive Discriminator Augmentation (ADA)". This concept of ADA supports the Discriminator network and Generator network to perform with stability without causing divergence during training. Unlike other GAN models, StyleGAN2-ADA helps achieve relatively good results with limited datasets due to the following significant developments:

1) The Stochastic Discriminator Augmentation is a flexible type of augmentation that, by showing all the applied augmentation to the discriminator, prevents the discriminator from becoming overly confident. This practice assists the generator in generating desired ideal outcome. The flow-diagram of Stochastic Discriminator Augmentation is presented in Fig. 1. The term 'Stochastic' refers to the use of a probability value 'p' that determines the strength of the augmentation. In this approach, the discriminator is only evaluated on augmented images. Both real and generated images are augmented using the same augmentation techniques.
2) Invertible transformations are applied to ensure the generated images follow the same distribution as the input images. The proposed augmentation pipeline contains 18 types of augmentations grouped in 6 categories: pixel blitting, more general geometric transformations, colour transforms, image-space filtering, additive noise, and cutout. Invertible transformations are helpful when a wide range of augmentations are applied during the training process.
3) Addition of Adaptive Discriminator Augmentation (ADA) can adjust the strength of augmentation 'p' in every interval of 4 mini-batches 'N'. This technique helps prevent the augmentation from leaking at the early stage of training and achieving convergence without overfitting, irrespective of the volume of the input dataset.

B. Rosacea Dataset- 'rff-300'

There are a few online teledermatology resources/datasets to support computer-aided diagnosis research and development. To identify the nature of Rosacea, it is crucial to have high quality images of the region of interest on the skin, which can be frequently hindered by the modality of the images and by watermarks appearing in images available in the teledermatology web sources. In order to generate synthetic images

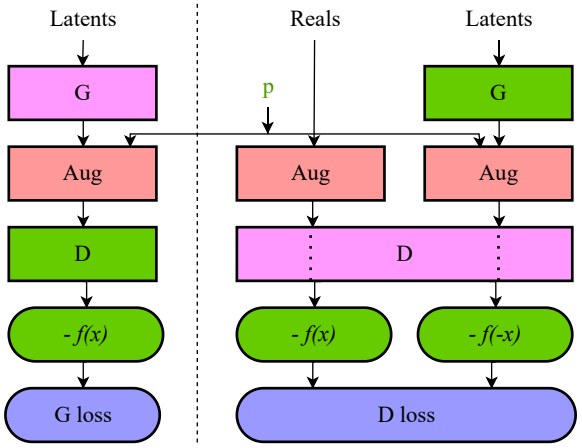

Fig. 1. The flow diagram of Stochastic Discriminator Augmentation, where G is the generator and D is the discriminator. The red boxes represent the 18 augmentation operations. The set of selected augmentation are controlled by the augmentation probability p and these augmentation can be visible to the discriminator D in the green. The pink boxes represent the networks that get trained during the training process and violet boxes represent the loss calculated after the training. In this set up, the non-saturating logistic loss is accommodated to calculate the final probability of the images being predicted as fake.

using GANs, it is essential to select a dataset with a uniform distribution and minimum artifacts such as watermarks.

Hence acquiring full-face images of Rosacea from web resources has proven to be a difficult task. However, for the purpose of this study we rely on a small dataset provided to us by the Powell Lab, University College Dublin, Ireland [18]. This dataset is referred to as the "Irish Dataset". The Irish Dataset contains 70 full-face Rosacea images, among which 67 images are included in this study. Another 67 images are taken from SD-260 [19]. A few more images were obtained from Google search results and teledermatology websites, making 300 images in total. For ease in understanding and usage, the entire dataset used in the experiment is given the name "**r**osacea-**f**ull-**f**ace-300" or rff-300. The resolution of the images is resized to 512×512 to keep the optimum details of the disease.

IV. Experiments and Evaluation

In this section the performance of StyleGAN2-ADA [8] is examined using the rff-300 dataset in multiple scenarios: (i) training the model from scratch (ii) training using transfer learning (on a pre-trained facial dataset) with/without layer freezing. A system equipped with a Nvidia Geforce RTX 3090 (24GB), an AMD Ryzen 9 5900X 12 core CPU, and 32 GB RAM is used to carry out the experiments. The complete implementation was carried out on Pytorch 1.7.1. with CUDA version 11.1 on Linux. In this experiment, keeping the dataset size in mind, all the images in rff-300 have been x-flipped, making a total of 600 images, that are fed into the model for training with a learning rate of 0.0025. The same learning rate, chosen experimentally, is used for all the experiments presented in this section in order to find the performance of different training set ups. However, selection of the augmentation differs in each experiment, which is explained in the next sections. In all the experiments, the augmentation strength 'p' is controlled by setting the initial value to 0 and adjusted in every interval of 4 mini-batches. The increment/decrement of 'p' is determined based on the measure of overfitting during training.

A. Training From Scratch

In this experiment, StyleGAN2-ADA network was trained from scratch for 3000 steps/iterations and all the augmentation techniques such as pixel blitting, geometric augmentation, colour augmentation (contrast, brightness etc), filter (band-pass), RGB noise and cutout were employed in this process. Fig. 2 and Fig. 3 illustrate the progress of the evaluation metric value KID [20] throughout the training process at each step and a few generated images obtained from the best model, KID = 6.8, respectively. For most of the images, it is visible that the features are pixellated and blurred, which means the model fails to generate the fine details of Rosacea. However, it captures the general redness of the Rosacea skin condition. Zooming in further on these images reveals a more severe blur around the edges. Such a condition could be addressed using Transfer learning on a model that is pre-trained on generic facial images. This may help in capturing more facial features.

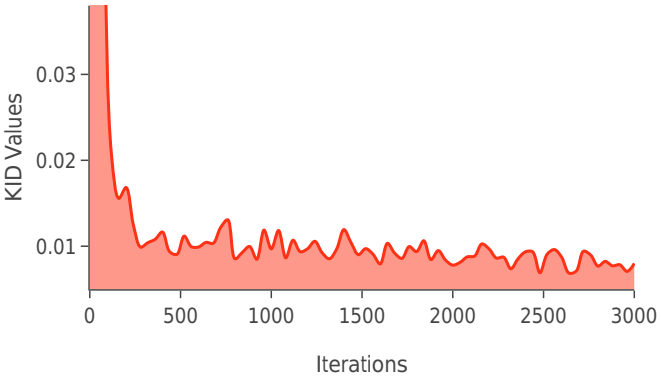

Fig. 2. The progress of KID value throughout the training from scratch.

B. Transfer Learning from FFHQ Dataset

To see better results with rff-300, the next experiment was performed with transfer learning using the FFHQ [7] dataset. As the FFHQ dataset is fundamentally a facial dataset, it was expected to have a wide range of facial features in the resulting generated images. In this experiment, only a few augmentations such as pixel blitting and geometric augmentations were utilised. Fig. 4 represents the trend of KID [20] values during the training process. The best KID value achieved from the model evaluation is 3.6 at training step 120. It can be observed that the KID value improves at a very early stage of the training. However, there is a significant climb of the KID value at step 800. This indicates that the model learns to generate the synthetic images quickly by step 120 and is likely to overfit at later steps. It is observed in this experiment that training

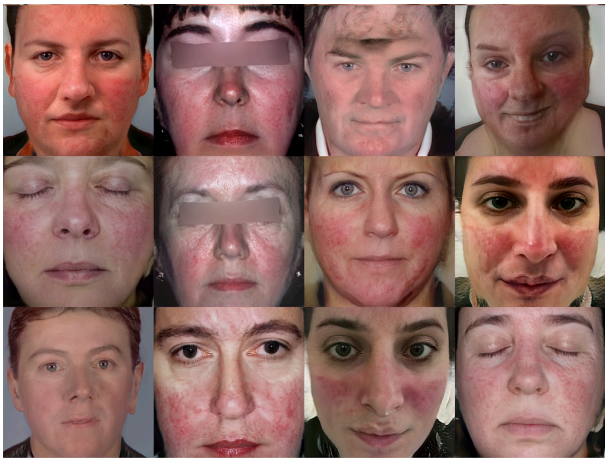

Fig. 3. Generated faces with Rosacea from the best KID value when training from scratch.

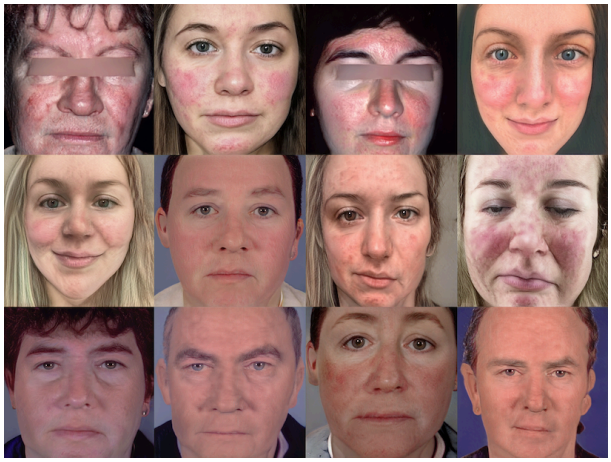

Fig. 5. Generated Rosacea faces with the best KID value in Transfer Learning.

the model through transfer learning generates nearly twice as better results than training from scratch, as presented in Table I.

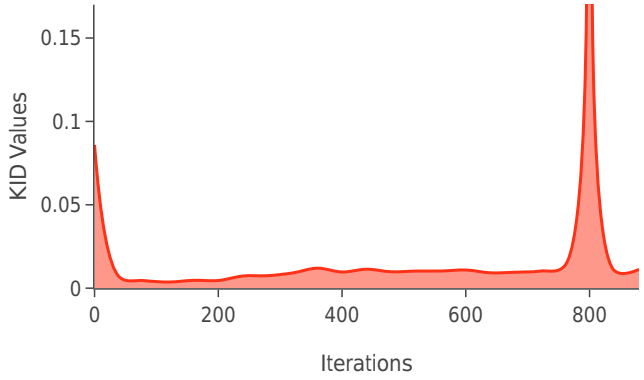

Fig. 4. Progress of KID value throughout the training from FFHQ.

As presented in Fig. 5, the generated images are sharper at the edges compared to the results obtained while training from scratch. However, the details of the skin condition are very smooth indicating that despite the image resolution of 512×512, there is a need for improvements in the image quality i.e. fine grain details such as sharpness of the skin artifacts in the generated images. Issues like this are commonly addressed by freezing the top layers of the discriminator to preserve the small features of the disease, which is discussed in the next section.

1) Transfer Learning from FFHQ Dataset with freeze-D: The final experiment relies on the Freeze-Discriminator (Freeze-D) [21] technique to improve the fine grained details of the images. In this experiment, the 4 top layers of the Discriminator are frozen, which improved the result faster than previous experiments. The augmentation choice was kept similar to the previous experiment i.e pixel blitting and geometric transformations. Fig. 6 represents the obtained KID values during the training process, in which the best value of

TABLE I
RESULTS WITH VARIOUS TRAINING SET UPS

Training set up	Best KID$\times 10^3$	Step no.
From scratch	6.8	2640
Transfer Learning from FFHQ	3.6	120
Transfer Learning with Freeze-D	3.5	80

KID = 3.5 is achieved at step 80. Hence, it is observed that the training process improves relatively faster when the top layers of the discriminator are frozen. As the transfer learning was performed based on FFHQ, the layers with resolution 512×512 were frozen, resulting in enhanced fine-grain details of the face and reduced blur. Zooming in on the images presented in Fig. 7 shows that the generated images of Rosacea contain improved details and characteristics. The sharpness at the edges of the face and overall details of the face are also visible. It is noticeable that the edges in a few images are still blurred. However, one thousand of high fidelity images are obtained from this particular experiment that can be useful for training a classification model in future.

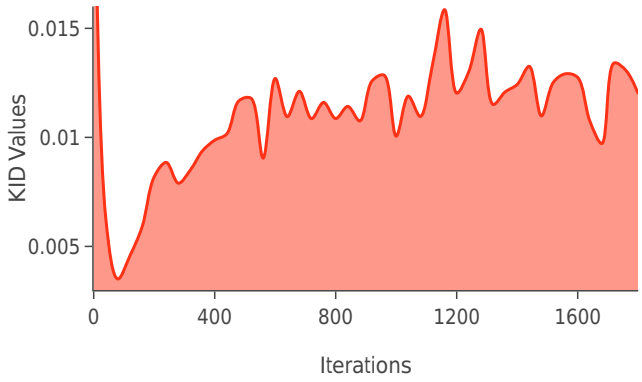

Fig. 6. Progress of KID value throughout the training from FFHQ with Freeze-D.

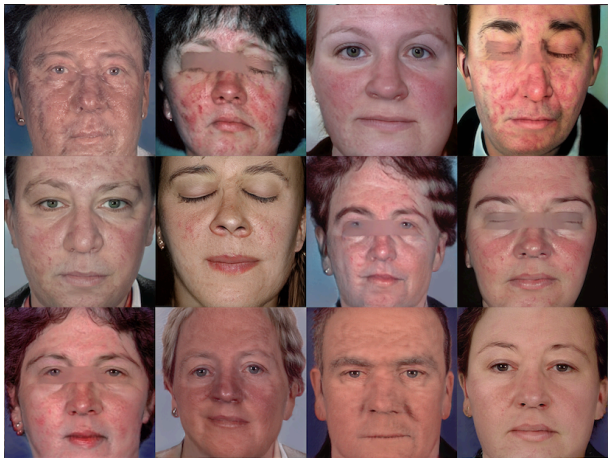

Fig. 7. Generated Rosacea faces with the best KID value in Transfer Learning with Freeze-D

V. Conclusion and Future work

This paper examined the potential of StyleGAN2-ADA in generating synthetic images of a particular skin condition known as Rosacea. The preliminary experiments and evaluations showed promising signs towards addressing the data scarcity problem in many applications of AI and particularly in skin disease diagnosis. From the results presented in Section IV, it is observed that improvements in training and generation of synthetic faces significantly increase through the adaption of transfer learning followed by the adaption of the freeze-Discriminator feature in StyleGAN2-ADA. The observable improvements noted in Table I offer motivation for further experiments with various hyperparameters that may help improve the quality of synthetic images. Therefore, a few more sets of experiments shall be considered in our future work. The selection of hyperparameters will be an important part of our future research and development with the same rff-300 dataset. The generated synthetic faces of Rosacea are currently being validated by dermatologists/non-specialist participants, and the results will be presented at ISSC 2022. Those synthetic images can be utilised to expand the existing Rosacea datasets for further usage in deep learning and computer vision applications for skin disease diagnosis. This phenomenon can help solve the data scarcity problem for Rosacea and other skin conditions or similar domains where the input datasets are short in supply.

References

[1] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[3] F. C. Powell, "Rosacea," *New England Journal of Medicine*, vol. 352, no. 8, pp. 793–803, 2005.

[4] J. Q. Del Rosso, R. L. Gallo, L. Kircik, D. Thiboutot, H. E. Baldwin, and D. Cohen, "Why is rosacea considered to be an inflammatory disorder? the primary role, clinical relevance, and therapeutic correlations of abnormal innate immune response in rosacea-prone skin." *Journal of drugs in dermatology: JDD*, vol. 11, no. 6, pp. 694–700, 2012.

[5] M. Steinhoff, J. Schauber, and J. J. Leyden, "New insights into rosacea pathophysiology: a review of recent findings," *Journal of the American Academy of Dermatology*, vol. 69, no. 6, pp. S15–S26, 2013.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[7] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[8] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 104–12 114, 2020.

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

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

[11] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8110–8119.

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

[13] K. Thomsen, A. L. Christensen, L. Iversen, H. B. Lomholt, and O. Winther, "Deep learning for diagnostic binary classification of multiple-lesion skin diseases," *Frontiers in medicine*, p. 604, 2020.

[14] H. Binol, A. Plotner, J. Sopkovich, B. Kaffenberger, M. K. K. Niazi, and M. N. Gurcan, "Ros-net: A deep convolutional neural network for automatic identification of rosacea lesions," *Skin Research and Technology*, vol. 26, no. 3, pp. 413–421, 2020.

[15] B. Xie, X. He, S. Zhao, Y. Li, J. Su, X. Zhao, Y. Kuang, Y. Wang, and X. Chen, "Xiangyaderm: a clinical image dataset of asian race for skin disease aided diagnosis," in *Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention*. Springer, 2019, pp. 22–31.

[16] Z. Zhao, C.-M. Wu, S. Zhang, F. He, F. Liu, B. Wang, Y. Huang, W. Shi, D. Jian, H. Xie *et al.*, "A novel convolutional neural network for the diagnosis and classification of rosacea: usability study," *JMIR medical informatics*, vol. 9, no. 3, p. e23415, 2021.

[17] C.-Y. Zhu, Y.-K. Wang, H.-P. Chen, K.-L. Gao, C. Shu, J.-C. Wang, L.-F. Yan, Y.-G. Yang, F.-Y. Xie, and J. Liu, "A deep learning based framework for diagnosing multiple skin diseases in a clinical environment," *Frontiers in medicine*, vol. 8, 2021.

[18] F. Powell. Powell Lab. [Online]. Available: https://www.ucd.ie/charles/research/researchgroups/thepowelllab/

[19] X. Sun, J. Yang, M. Sun, and K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *European Conference on Computer Vision*. Springer, 2016, pp. 206–222.

[20] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying mmd gans," *arXiv preprint arXiv:1801.01401*, 2018.

[21] S. Mo, M. Cho, and J. Shin, "Freeze the discriminator: a simple baseline for fine-tuning gans," *arXiv preprint arXiv:2002.10964*, 2020.