

# Improve Image-based Skin Cancer Diagnosis with Generative Self-Supervised Learning

Zhihang Ren Yunhui Guo Stella X. Yu David Whitney

*UC Berkeley / International Computer Science Institute*

*Berkeley, CA, USA*

*{peter.zhren,yunhui,stellayu,dwhitney}@berkeley.edu*

**Abstract**—Skin cancer is the most common malignancy in developing countries. This is largely due to the lack of early detection. The best method for early detection of skin cancer is to track the changes in skin lesions. But, it is hard to implement in developing countries due to the scarcity of experts and their availability in remote areas. Teledermatology provides a promising technology for monitoring skin cancer. Currently, with the involvement of deep learning, teledermatology has become more efficient. However, deep learning, and in particular supervised learning, requires a large amount of data, while collecting and labeling skin lesion images is tedious and requires a high degree of expertise. It is thus expensive to collect enough labeled data to train deep neural networks for skin cancer analysis. Recently, self-supervised learning has proven itself useful for learning representations directly from unlabeled images. Yet, for some rare diseases, e.g. Actinic Keratosis, it is also infeasible to collect enough unlabeled images. In this paper, we utilize Generative Adversarial Network (GAN) to generate synthetic unlabeled images which have high semantic similarity with existing unlabeled medical data. In particular, we evaluate the use of StyleGAN for the data augmentation of skin cancer image self-supervised learning. We utilized StyleGAN to generate new training samples which have the same semantics as the original unlabeled training images. We then combined the new GAN-generated samples with the original unlabeled images as the new training dataset for self-supervised learning. The self-supervised pre-trained network is used as a fixed feature backbone for supervised classification with a limited number of labeled skin cancer images. Quantitative results confirm that our GAN-based data augmentation can boost the accuracy of self-supervised skin cancer image classification by 11.17% on BCN20000 and 3.07% on HAM10000.

**Keywords**-skin cancer; self-supervised learning; generative adversarial networks; data augmentation

## I. INTRODUCTION

Skin cancer is increasingly becoming a public health concern. It is the most common cancer in many countries including the United States [1], [2]. In certain developing countries, e.g. Brazil, the situation is even worse [3]. Currently, the best method for early detection of skin cancer is to track the changes in skin lesions. But it is hard to implement in developing countries due to the scarcity of experts and their availability in remote areas. Thus, in Brazil, certain types of skin cancer, such as basal cell cancer (BCC) and squamous cell cancer (SCC), are usually diagnosed at advanced stages [4].

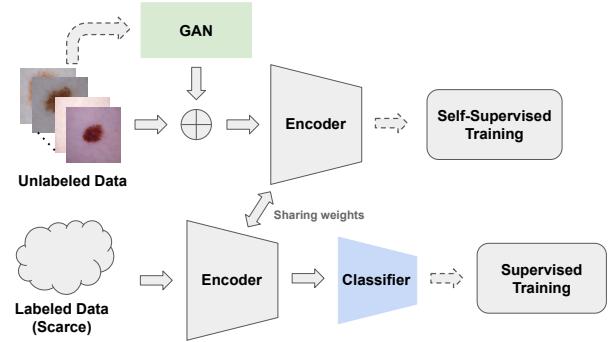


Figure 1: Proposed method. We first train StyleGAN [9] on unlabeled data to generate high quality skin cancer images which are semantically similar to the unlabeled training dataset. Then, we train a feature encoder via self-supervised learning. At last, a linear classifier is attached to the feature encoder to test the performance of skin cancer classification on the scarce labeled data.

Teledermatology provides a promising technology for monitoring skin cancer [5]–[8]. This is mainly attributed to the accessibility and ubiquity of smartphones. App users can be diagnosed remotely by a group of dermatology experts without meeting the dermatologist in-person. Dermatologists can therefore serve not only their local patients but also patients far from their working site. Most available mobile health apps for skin cancer detection utilize machine learning algorithms which heavily rely on handcrafted features [8]. Currently, deep learning has been widely utilized in medical image analysis, where features are learned automatically by neural networks.

Deep learning has achieved great success in general image recognition tasks [10]–[12], and researchers have also applied deep learning methods to medical image analysis, in particular; For example, in lesion classification [13], [14], lesion detection [15], [16], lesion segmentation [17], [18], and so on. While the models become deeper and deeper, a data scarcity issue emerges. Supervised models need ground truth labels along with image data. For traditional computer vision tasks, researchers have created publicly labeled datasets, such as ImageNet [19], Microsoft COCO [20], and CIFAR-10 [21]. However, collecting and labeling

those data is tedious and expensive. In medical image arenas, researchers usually collect data from hospitals. But the requisite data processing procedures, such as file categorization, annotation, and data de-identification, are time-consuming. Moreover, it requires experts (e.g., extensively trained and highly paid radiologists or pathologists) to perform meticulous annotations that is even harder or impossible [22]. For teledermatology and other mobile health technologies, the data scarcity issue may be even worse due to the lack of users at the early stage.

With limited labeled data, few-shot learning [23]–[25] has become popular. In few-shot learning, we are given some categories where each category only has a limited number of images. Few-shot learning methods [23]–[25] utilize the knowledge learned from some base categories which are different from the given categories. Hence, it is possible to learn a novel category by showing one or several images [26]. Few-shot learning has already been successful in handwritten characters, birds, dogs, and other natural images [23], [27]. However, it is still difficult to apply the few-shot learning techniques to medical images because of the lack of base category data.

Unsupervised learning is a promising approach to reduce the labelling cost for training deep neural networks [28]–[32]. Unsupervised learning methods [28]–[31] aim to learn a useful representation directly from unlabeled data. Then, the learned representation can be reused for supervised learning with limited labeled data [30], thus reducing the cost of data labeling. For traditional computer vision tasks, there are usually sufficient unlabeled data for self-supervised learning. However, in clinical practice, medical images with lesions are anomalies, and are therefore rare and naturally hard to find and collect. In particular, certain skin cancers, such as Merkel cell carcinoma, are so rare that it is impossible to find sufficient data [33]. Therefore, it is still difficult for medical image tasks to obtain adequate data for self-supervised learning. Notably, the effectiveness of the learned representation with self-supervised learning depends on the size of the unlabeled dataset [31]. It is thus critical to augment the unlabeled data when the amount of unlabeled data itself is limited in medical image analysis.

To improve the performance of self-supervised learning on skin cancer images, we propose to use a Generative Adversarial Network (GAN) to augment the unlabeled dataset. Generative Adversarial Networks have been utilized to create a range of authentic images [9], [34]–[36], including faces, cars, landscapes, and so on. Trained on real image datasets, a GAN can learn to estimate the manifold that represents the training images. Through training, the learned manifold and the real image manifold can be practically aligned. In doing this, GANs can learn both local and global statistics of the real images from the training dataset, and the generated images can have similar semantic content to that of real images. However, it is unclear whether GAN generated

images can be utilized to boost the performance of self-supervised learning on skin cancer images.

In this paper, we investigate how to leverage GAN generated skin cancer images to improve the self-supervised learning performance on skin cancer classification task for teledermatology (Fig. 1). We first train StyleGAN [9] on unlabeled data to generate high quality skin cancer images which are semantically similar to the unlabeled training dataset. Then, we train a feature encoder via self-supervised learning using the augmented training dataset which includes the StyleGAN generated images and the labeled training images. At last, a linear classifier is attached to the feature encoder to test the performance of skin cancer classification on the scarce labeled data.

#### A. Contributions

In this paper, we propose to use a Generative Adversarial Network (GAN) to augment training data for self-supervised learning on skin cancer images. The contributions of this work can be summarised as follows:

- We propose to use StyleGAN [9] for data augmentation to boost the self-supervised skin cancer classification accuracy. To the best of our knowledge, it is the first time that GAN-based data augmentation is applied to self-supervised learning algorithms for skin cancer image classification tasks.
- The self-supervised skin cancer classification accuracy can be boosted by 11.17% on BCN20000 [37] and 3.07% on HAM10000 [38] after StyleGAN-based data augmentation.

## II. RELATED WORK

**Generative Adversarial Networks:** The GAN is a promising image synthesis model. The model consists of two networks, a generator network and a discriminator network. Inspired by game theory, those two networks are trained in an adversarial process where the generator generates fake but authentic images to fool the discriminator and the discriminator discriminates between the real and fake images repeatedly [39]. Conceptually, the training process can be described as a minmax game, which is formulated as follows:

$$\min_G \max_D E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

where  $G$  represents the generator,  $D$  represents the discriminator,  $p_{data}(x)$  indicates the real data distribution, and  $p_z(z)$  indicates the noise vector distribution. While generating new images, the generator takes in noise vectors  $z$  sampled from distribution  $p_z(z)$  and maps onto the estimated image manifold. The training process guarantees that the estimated image manifold is aligned with the training image manifold by optimizing this minmax loss, i.e. the adversarial

loss. Ideally, this minmax game has a global optimum at  $p_g = p_{data}(x)$ , where  $p_g$  is implicitly defined by the generator  $G$  while  $G(z)$  is the sample when  $z \sim p_z(z)$ .

Originally, the training process of Generative Adversarial Networks (GAN) is highly unstable. This makes the optimum point of the training hard to reach. Hence, the generated images from this pioneering work are blurry and hard to recognize. Later work [40]–[42] focuses on improving the loss metrics and training strategies, which improves the generated image quality. A modified approach, the PGGAN [35] proposed to train the GAN in a coarse to fine manner. Starting with low resolution, high resolution layers will be added and trained after the lower layers. Upon the same training strategy, StyleGAN [9] added another mapping from original latent space  $\mathcal{Z}$  into the  $\mathcal{W}$  space through a non-linear mapping network and then merged into the synthesis network via adaptive instance normalization (AdaIN) at each convolutional layer [43], [44]. This potentially improves the representational ability of StyleGAN and allows it to generate stunningly high resolution images.

In medical image applications, Frid-Adar et al. [45] utilized DCGAN [46] and ACGAN [47] to generate CT liver lesion patches and boosted the liver lesion classification performance. Han et al. [48] proposed to use WGAN [40] to generate MR images for data augmentation and physician training. Nie et al. [49] used GAN to predict CT images from MR images. And, Cao et al. [50] proposed an Auto-GAN to synthesize missing modality for medical images. In particular, GAN has been widely used for skin cancer image generation and purification [51]–[54].

**Unsupervised Learning:** Unsupervised learning aims at learning useful representations from unlabeled data. In [28], Wu et al. try to learn an embedding function by enforcing the features to be discriminative among individual instances. In unsupervised contrastive learning, the goal is to learn a good representation by pulling together positive sample pairs and pushing apart negative sample pairs. The idea of unsupervised contrastive learning is instantiated via different self-supervised learning methods [29]–[32], [55], [56] in which the positive sample pairs are crafted by applying different data augmentations on the same image. In self-supervised learning, the augmentations of the same image are attracted and the augmentations of different images are repulsed in the embedding space. In particular, SimCLR [30] leverages the composition of data augmentations and large batch sizes to improve the effectiveness of the representation. MoCo [31] uses a momentum encoder to improve the consistency of the queue of negative samples. From an augmented view of an image, BYOL [32] trains an online network to predict a target network representation of the same image under a different augmented view.

In addition to supervised learning approaches, several groups have applied unsupervised learning to medical image registration and classification tasks [57]–[59]. For example,

Armanious et al. [60] proposed an unsupervised translation framework for PET-CT translation and MR motion correction. Li et al. [61] utilized multi-modal data for retinal disease diagnosis via self-supervised learning. In particular for skin cancer images, [62]–[64] used self-supervised learning for skin cancer classification tasks.

**Traditional Data Augmentation:** Data augmentation is a traditional approach to improve model generality. Common methods include cropping, rotation, occlusion, flipping, shearing, zooming in/out, image blurring, and changing brightness or contrast. In supervised learning, traditional augmentation methods have been widely utilized [65]. But the performance improvement is limited since those elementary image operations do not introduce much variety to the training data. Recently, GAN-based data augmentation methods have been widely utilized. Shin et al. [66] used GAN-based data augmentation to improve the performance of tumor segmentation in brain MRI. Lim et al. [67] proposed an adversarial autoencoder to augment the data for unsupervised anomaly detection. Waheed et al. [68] proposed CovidGAN to enhance the performance of CNN for COVID-19 detection.

Our proposed method aims to utilize GAN generated skin cancer images to augment the training data for self-supervised learning. Unlike [51]–[54], which mainly aim at skin cancer image generation, and [62]–[64], which mainly focus on the unsupervised learning for skin cancer images, our proposed method leverages the advantages of both methods and improves the performance of the self-supervised learning.

### III. GAN AUGMENTATION FOR SELF-SUPERVISED LEARNING ON SKIN CANCER IMAGES

In this paper, we propose to utilize Generative Adversarial Networks (GANs) to generate synthetic unlabeled data, which is then used for self-supervised learning of skin cancer images. For traditional computer vision tasks, unlabeled data is easy to collect. However, for medical image analysis, even unlabeled data is scarce, particularly for some rare diseases. Our proposed approach allows self-supervised learning on a limited number of unlabeled data. The self-supervised pre-trained model can be further utilized to boost performance on skin cancer image classification.

#### A. Self-supervised learning on skin cancer images

It is infeasible to train deep neural networks with a limited number of labeled skin cancer images, so we employed self-supervised learning to pretrain the model on unlabeled images. In self-supervised learning, the goal is to learn a useful representation directly from unlabeled data. Several factors influence the success of self-supervised learning: the amount of unlabeled data [31], the training batch size [30], and the composition of data augmentation operations [30]. In

contrast to natural images, unlabeled medical images are also expensive to collect. Therefore, we augment the unlabeled data with GAN generated synthetic images to increase the size of the unlabeled dataset. We employ two recently proposed self-supervised learning methods, SimCLR [30] and BYOL [32], to pretrain the model on unlabeled images.

1) *SimCLR*: SimCLR [30] is a recently proposed contrastive self-supervised learning method. The goal of SimCLR is to learn representations by attracting differently augmented views of the same data example in the latent space. For each image  $x$  in a given set of  $N$  images, SimCLR generates two augmented views of  $x$  via a stochastic data augmentation module, resulting in a total of  $2N$  images. The two differently augmented views of the same image form a positive pair and the other  $2(N - 1)$  images are negative samples. SimCLR applies a neural network base encoder and a projection head to embed each image in a latent space. The embedded vector is denoted as  $\mathbf{z}$ . For a positive pair  $\{\mathbf{z}_i, \mathbf{z}_j\}$ , the contrastive loss is written as,

$$c_{i,j} = -\log \frac{\exp(\mathbf{z}_i^T \mathbf{z}_j / \tau)}{\sum_{k=1, k \neq i}^{2N} \exp(\mathbf{z}_i^T \mathbf{z}_k / \tau)} \quad (2)$$

where  $\tau$  is a temperature parameter. Positive pairs will be attracted in the latent space by minimizing Equation 2.

2) *BYOL*: More recently, BYOL [32] was proposed as a self-supervised learning method which does not rely on negative samples. BYOL learns the representation by iteratively predicting one augmented view of a given image via a differently augmented view of the same image.

Formally, given an image  $x$ , BOYL applies stochastic data augmentation to generate two augmented views  $x'$  and  $x''$ . The online network with parameter  $\theta$  generates a representation  $z'_\theta$  based on  $x'$  and the target network with parameter  $\xi$  generates a representation  $z''_\xi$  based on  $x''$ . Then the target network outputs a prediction  $c_\theta(z'_\theta)$  of  $z''_\xi$  with a classifier  $c_\theta$ . The prediction  $c_\theta(z'_\theta)$  and  $z''_\xi$  are both L2-normalized and are optimized to be close via a mean squared error  $\mathcal{L}'_{\theta, \xi}$ . The loss function is further symmetrized by feeding  $x'$  to the target network and  $x''$  to the online network to obtain  $\mathcal{L}''_{\theta, \xi}$ . The final loss function is written as  $\mathcal{L}_{\theta, \xi} = \mathcal{L}'_{\theta, \xi} + \mathcal{L}''_{\theta, \xi}$ .

The parameter  $\theta$  of the online network is optimized via stochastic gradient descent and the parameter  $\xi$  of the target network is updated with a moving average,

$$\theta \leftarrow OPT(\theta, \nabla \mathcal{L}_{\theta, \xi}) \quad \xi \leftarrow \gamma \xi + (1 - \gamma) \theta \quad (3)$$

where  $OPT$  is an optimizer and  $\nabla \mathcal{L}_{\theta, \xi}$  is the gradient of the loss function.  $\gamma$  is a hyperparameter which controls the smoothness of the moving average.

## B. Method

In this section, we describe the proposed pipelines of data augmentation with GAN for self-supervised learning on

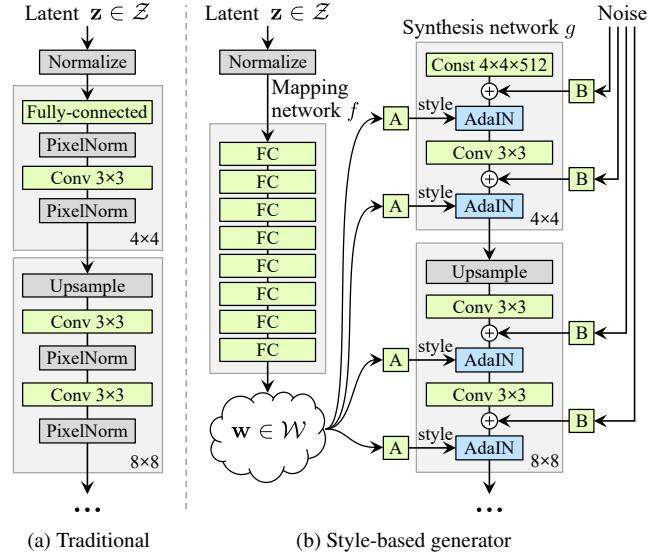


Figure 2: StyleGAN Architecture. Compared to traditional GAN models, whose generator directly takes in the latent code only from the input layer, the generator of StyleGAN first maps the latent space to an intermediate latent space  $\mathcal{W}$  using a 8-layer Multilayer Perceptron (MLP). Then it will be merged into each convolutional layer via adaptive instance normalization (AdaIN). Gaussian noise will be added after each convolution before the activation layer. "A" represents a learned affine transform and "B" represents learned per-channel scaling factors to the noise input. (Figure is reprinted from [9])

skin cancer images. The pipeline of our proposed method is shown in Figure 3. First, StyleGAN is trained on the unlabeled skin cancer images and generates authentic samples for self-supervised learning. Then we train the feature encoder on the augmented dataset including the scarce labeled images and the generated samples from StyleGAN. At last, we leverage the self-supervised learned feature encoder on the skin cancer image classification task on scarce labeled data.

1) *StyleGAN-based Data Augmentation*: StyleGAN is the state-of-the-art high resolution image synthesis model. The architecture is shown in Figure 2. Unlike a traditional generator, the latent code  $z$  will first be mapped to  $w$  in an intermediate latent space through a non-linear mapping network, i.e. a 8-layer Multilayer Perceptron (MLP). Then the learned affine transformations specialize  $w$  to styles  $y = (y_s, y_b)$  that control adaptive instance normalization (AdaIN) [43], [44] operations after each convolution layer of the synthesis network. The AdaIN operation is defined as

$$AdaIN(x_i, y) = y_{s,i} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i} \quad (4)$$

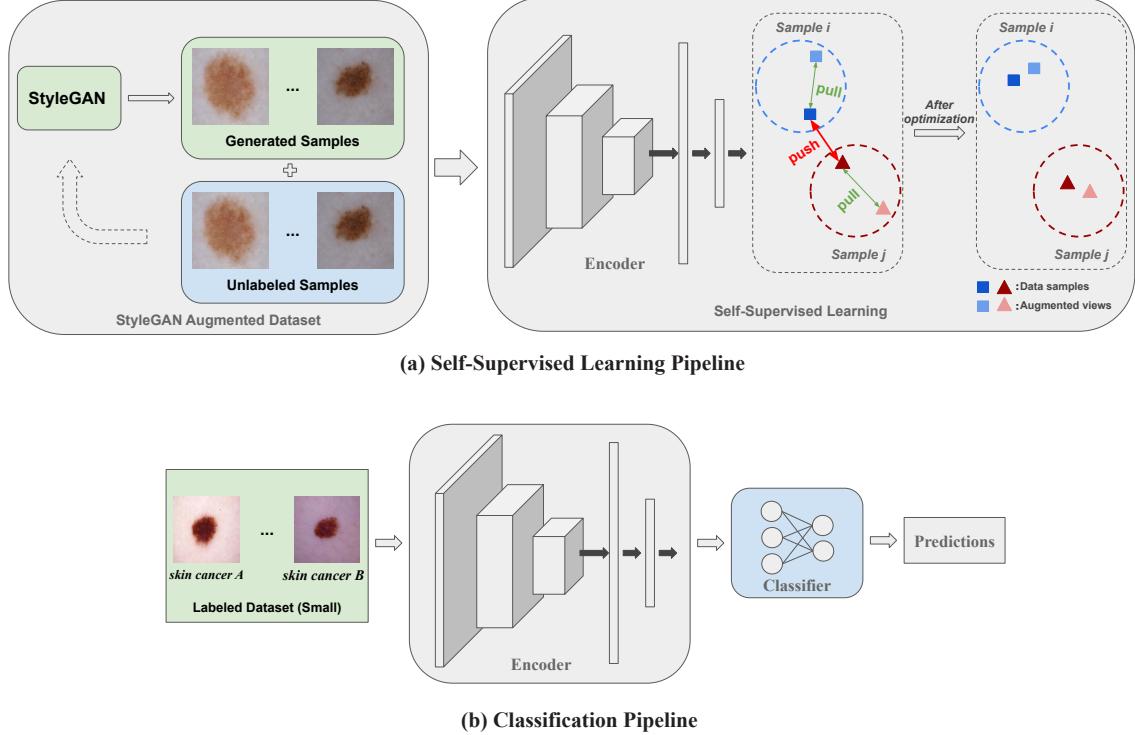


Figure 3: Proposed Pipeline. (a) Self-supervised learning pipeline: StyleGAN is first trained using the unlabeled samples and generates authentic skin cancer samples to augment the original training dataset. Then we use self-supervised learning to train a feature encoder. We generate augmented views for each sample in the augmented dataset. The augmented views are treated as positive pairs that are trained to pull towards each other. The augmented views from other samples form negative pairs that are pushed away from each other. (b) Classification pipeline: we leverage the self-supervised trained feature encoder on the skin cancer image classification with limited labeled data. During training, we attach a fully connected layer as the classifier. Only the parameters of the classifier are updated.

where  $\mathbf{x}_i$  is the feature map at each layer. It will be normalized separately, then scaled and biased according to the scalar components from styles  $\mathbf{y}$ .

StyleGAN is trained in a progressive manner similar to PGGAN [35]. The training starts from  $4 \times 4$  resolution. Then after previous resolution layers finish training, layers for the next resolution will be attached for training. In this paper, the generator network consist of 14 layers – two for each resolution ( $4^2 - 256^2$ ). The final resolution for the generated image is  $256 \times 256$ .

Using the same training dataset as the one for self-supervised learning, we train a StyleGAN. Then, we sample vectors  $z$  in the latent space and pass them into the StyleGAN generator to generate extra skin cancer images for data augmentation. Finally, the GAN-generated images and original training data are combined together for self-supervised learning. In total, 20,000 skin cancer images are generated for data augmentation, augmenting the training dataset size to 25,000.

2) *Self-supervised learning on skin cancer images:* For SimCLR [30], we use a Resnet18 [10] backbone for feature encoding. During training, we generate 2 augmented views for each image via random cropping, random horizontal flipping, random color jittering, and random grayscaling. Augmented views from the same image are treated as positive pairs. In SimCLR, the positive pairs are attracted in the latent space. While for augmented views from different images, they are negative pairs, which will be repelled from each other. Positive pairs augmented from an example skin cancer image are shown in Figure 4. For BYOL [32], we generate 2 augmented views by using the same image operations as SimCLR. And the online network and target network are optimized iteratively. After self-supervised learning, the feature encoder is fixed.

For both methods, we use 5,000 images randomly subsampled from the training dataset (BCN20000 [37] or HAM10000 [38]) to train the feature encoder. We do not use all the image from the training dataset because we aim to simulate the data scarcity problem which widely exists in

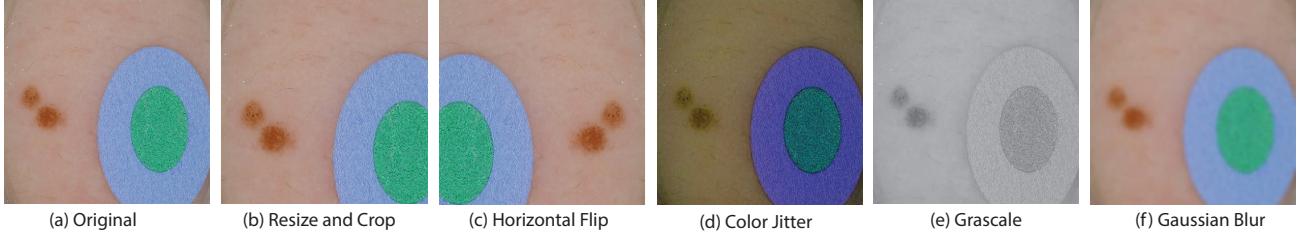


Figure 4: Illustration of the operations for SimCLR augmented views. Here, we show all elementary operations. During training, each augmented view is generated by randomly combining those operations. In this paper, we generated two augmented views for self-supervised training.

medical image arenas. Moreover, in this data scarcity setting, we can test whether the GAN-based data augmentation is able to boost the self-supervised learning performance. While using StyleGAN-based data augmentation, we add the aforementioned 20,000 generated samples together with the original training images for self-supervised learning.

*3) Classification via StyleGAN-boosted feature encoder* : For the baseline model, we use the Resnet18 [10] feature encoder trained by self-supervised learning methods as mentioned in the previous section using subsampled 5,000 skin cancer images. We add one fully connected layer attached to the feature encoder to classify the skin cancer images. Then, we finetune this classifier. Samples from this dataset are shown in Figure 5.

During training, the feature encoder parameters are fixed, and only the parameters of the last fully connected layer will be optimized. The training utilizes 80% of the scarce labeled dataset (5000 images subsampled from BCN20000 [37] or HAM10000 [38]). The remaining 20% of the dataset is used for testing. For all the experiments, we repeatedly train the last fully connected layer for 5 times with different random seeds and record the mean test accuracy and its standard deviation.

#### IV. EXPERIMENTS

In the experiments, we investigate the following questions:

- Does the self-supervised pretraining improve the accuracy of the skin cancer classification?
- Does the StyleGAN-based data augmentation improve the quality of the representation learned by self-supervised learning?
- Does the quantity of augmented images influence the improvement of skin cancer classification performance?

##### A. Evaluation datasets

BCN20000 [37] is the dataset from the International Skin Imaging Collaboration (ISIC) 2019 Challenge. It contains 25,331 labeled but unbalanced skin cancer images. 8 skin cancer types are included: nevus, melanoma, basal cell carcinoma, seborrheic keratosis, actinic keratosis, squamous cell carcinoma, dermatofibroma, vascular lesion.

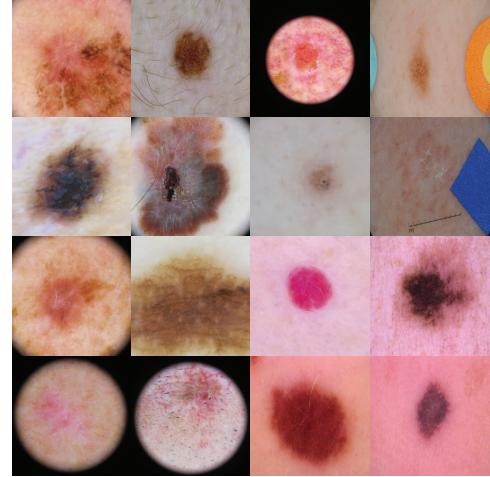


Figure 5: Training samples extracted from BCN20000 [37]. It is clear that the variety of the dataset is large. The images have various skin tones, dark corners, hairs, and color patches, which makes the classification extremely hard without a good feature encoder.

HAM10000 [38] is the dataset from the ISIC 2018 Challenge. It contains 10,000 skin cancer images, including actinic keratosis, basal cell carcinoma, benign keratosis, dermatofibroma, melanocytic nevi, melanoma, and vascular lesion.

Both datasets are highly imbalanced. The quantity of each skin cancer category varies a lot. Compared to HAM10000 [38], BCN20000 [37] is a more challenging dataset. BCN20000 contains lesions found in hard to diagnose locations (nails and mucosa) [37]. Most of the images would be considered hard-to-diagnose [37].

##### B. Model training and implementation details

The latent vector for StyleGAN generator has the dimension of 512. The generator consists of 14 layers – two for each resolution ( $4^2$ – $256^2$ ). The discriminator has the mirrored structure of the generator – also two for each resolution ( $256^2$ – $4^2$ ). We use the Adam optimizer [69] with  $\beta_1 = 0.0$  and  $\beta_2 = 0.99$ . The learning rate is set to 0.002.

During training, the images are reshaped to  $256 \times 256$ .

While training via SimCLR [30], we augment 2 views for each training image. We train the Resnet18 [10] backbone for 200 epochs with a batchsize of 256. The learning rate is set to 0.0003. During training, the images are reshaped to  $96 \times 96$ .

For the fine tuning, we attach 1 fully connected layer to the Resnet feature encoder for skin cancer image classification. We fix the parameters for the feature encoder and train the attached fully connected layer 200 epochs. The Adam optimizer [69] is used with default parameters. The learning rate is 0.0001.

### C. Quantitative Results

*1) With vs. Without Self-supervised Pretraining:* In order to investigate whether self-supervised learning would improve the accuracy of the skin cancer classification, we compare the classification results using the feature encoder with and without self-supervised pretraining on both BCN20000 [37] and HAM10000 [38]. For the result without self-supervised pretraining, we randomly initialize the Resnet18 feature encoder parameters. While for self-supervised pretraining, we train a Resnet18 feature encoder using SimCLR and BYOL. During testing, we attach a fully connect layer as the classifier and train its parameters with the feature encoder parameters fixed. Here, both SimCLR and BYOL are trained using the 5,000 images subsampled from BCN20000 [37] or HAM10000 [38]. The comparison of the skin cancer classification accuracy is shown in Table.I.

	BCN20000	HAM10000
w/o pretraining	$26.05 \pm 1.24\%$	$67.87 \pm 0.38\%$
SimCLR	$34.73 \pm 1.07\%$	$71.84 \pm 0.23\%$
BYOL	$35.71 \pm 2.04\%$	$71.37 \pm 0.36\%$

Table I: Classification accuracy w/o vs. w/ self-supervised pretraining on BCN20000 [37] and HAM10000 [38]

On BCN20000 [37], the classification accuracy is  $26.05 \pm 1.24\%$  without self-supervised pretraining. SimCLR and BYOL achieve  $34.73 \pm 1.07\%$  and  $35.71 \pm 2.04\%$  respectively. On HAM10000 [38], the classification accuracy is  $67.87 \pm 0.38\%$  without self-supervised pretraining. SimCLR and BYOL achieve  $71.84 \pm 0.23\%$  and  $71.37 \pm 0.36\%$  respectively.

Clearly, with SimCLR and BYOL self-supervised pre-training, we can improve the skin cancer classification accuracy compared to a random feature encoder (without self-supervised pretraining). This indicates that self-supervised learning methods can learn useful representations directly from unlabeled skin cancer images. It further reveals that it is possible to utilize the knowledge from unlabeled images to improve the medical image classification.

*2) With vs. Without StyleGAN-based Data Augmentation:* We then investigate whether the StyleGAN-based data augmentation would improve the self-supervised learning per-

formance on the skin cancer classification. First, we train a Resnet18 network as the feature encoder via self-supervised learning using 5,000 skin cancer images subsampled from BCN20000 [37] or HAM10000 [38]. We utilize both SimCLR and BYOL. For the StyleGAN-based data augmentation, 20,000 generated skin cancer images are added into the training dataset for self-supervised learning, augmenting the training dataset size to 25,000. The comparison of the skin cancer classification accuracy is shown in Table.II.

	BCN20000	HAM10000
SimCLR w/o DA	$34.73 \pm 1.07\%$	$71.84 \pm 0.23\%$
SimCLR w/ DA	$38.55 \pm 0.44\%$	$72.52 \pm 0.25\%$
BYOL w/o DA	$35.71 \pm 2.04\%$	$71.37 \pm 0.36\%$
BYOL w/ DA	$46.88 \pm 0.48\%$	$74.44 \pm 0.28\%$

Table II: Classification accuracy w/o vs. w/ GAN-based Data Augmentation (DA) on BCN20000 [37] and HAM10000 [38]

For SimCLR, the classification performance is boosted from  $34.73 \pm 1.07\%$  to  $38.55 \pm 0.44\%$  on BCN20000 [37], and the classification performance is boosted from  $71.84 \pm 0.23\%$  to  $72.52 \pm 0.25\%$  on HAM10000 [38]. For BYOL, the classification performance is boosted from  $35.71 \pm 2.04\%$  to  $46.88 \pm 0.48\%$  on BCN20000 [37], and the classification performance is boosted from  $71.37 \pm 0.36\%$  to  $74.44 \pm 0.28\%$  on HAM10000 [38].

Clearly, the StyleGAN-based data augmentation can improve the self-supervised learning performance on the skin cancer classification. Since we are using the feature encoder trained via self-supervised learning methods, it further indicates that the StyleGAN-based data augmentation can help self-supervised learning methods learn more useful representation from unlabeled skin cancer images.

*3) Influence of the StyleGAN Augmented Sample Quantity:* In this experiment, we vary the quantity of both raw unlabeled training images and StyleGAN augmented samples to train self-supervised classification via SimCLR. The experiment is conducted on BCN20000 [37]. The quantity of raw unlabeled images is set at  $1k$ ,  $3k$ ,  $5k$  and  $7k$ . The quantity of StyleGAN augmented samples is set at  $0$ ,  $10k$  and  $20k$ . We investigate the classification accuracy under different combinations of those two parameters, i.e. at different augmentation ratio. The augmentation ratio is defined as follow.

$$ratio = \frac{Q_{raw}}{Q_{augmentation}} \quad (5)$$

where  $Q_{raw}$  is the quantity of raw unlabeled images and  $Q_{augmentation}$  is the quantity of StyleGAN augmented samples. The skin cancer classification accuracy at different augmentation ratios is shown in Fig.7.

From the bar chart, it is clear that without StyleGAN-based data augmentation, increasing the raw unlabeled

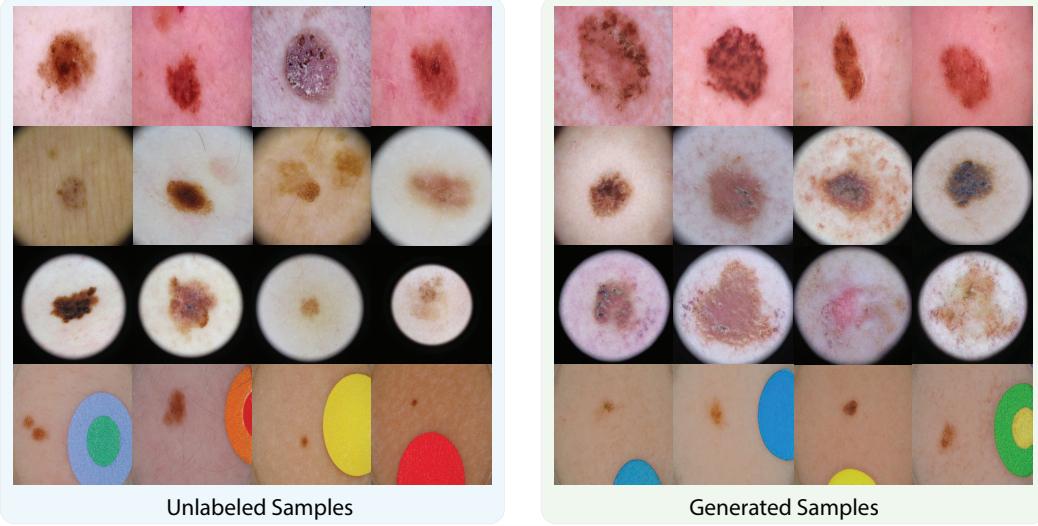


Figure 6: Uncurated set of novel images produced by StyleGAN on BCN20000 [37]. Compared to the images from unlabeled training dataset, the generated samples well maintained the semantic statistics, such as the skin tone, the dark corner of the image, and some color patches. The generated skin cancer image resolution is  $256 \times 256$ .

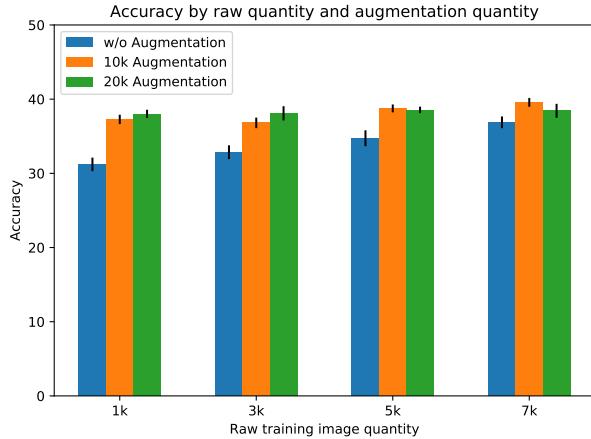


Figure 7: Classification accuracy on BCN20000 [37] at different StyleGAN augmented sample quantities.

image quantity can help to improve the self-supervised classification result. While applying StyleGAN-based data augmentation, for small raw unlabeled images quantities where the augmentation ratio is large, such as  $1k$  and  $3k$ , the skin cancer classification accuracy can gain a lot. However, for larger raw unlabeled images quantities where the augmentation ratio is small, such as  $5k$  and  $7k$ , the accuracy boost is reduced.

#### D. Qualitative Results

1) *StyleGAN generated results:* After StyleGAN has been trained, we randomly sample latent codes  $z$ , then pass them to the generator. The generated images are shown in

Figure.6 and Figure.8 for BCN20000 [37] and HAM10000 [38] respectively. From the generated results, it is clear that the StyleGAN generator has learned the semantic statistics of the training dataset, such as the skin tone, the dark corner of the image, and some color patches. Moreover, the generator can utilize the learned statistics to produce novel images which do not exist in the real world.

Additionally, it is clear that BCN20000 [37] is a more challenging dataset since it has more diverse image texture compared to HAM10000 [38]. Intuitively, this indicates that compared to HAM10000, images in BCN20000 scatter in a sparser way on the image manifold such that StyleGAN based data augmentation can efficiently interpolate between the image samples. On the contrary, HAM10000 is less diverse, i.e., the image samples are locally denser on the image manifold. Therefore, the performance boost from StyleGAN based data augmentation is limited on HAM10000.

2) *Comparison between PGGAN and StyleGAN:* We also train PGGAN to perform the skin cancer image generation on BCN20000 [37]. We compare PGGAN generation quality with StyleGAN generation quality because they both share the same progressive training manner and have similar network structures. During training, we use the same number of epochs with the same optimizer setting and learning rate. The generation results are randomly picked for both models and are arranged based on certain semantics. The comparison is shown in Figure.9.

In general, StyleGAN generation is better than the PGGAN generation visually. As indicated by the red arrows, StyleGAN can generate sharper details for hair, lesion texture, surrounding skin texture and color patches, while those details are blurry and unreasonable in PGGAN generated

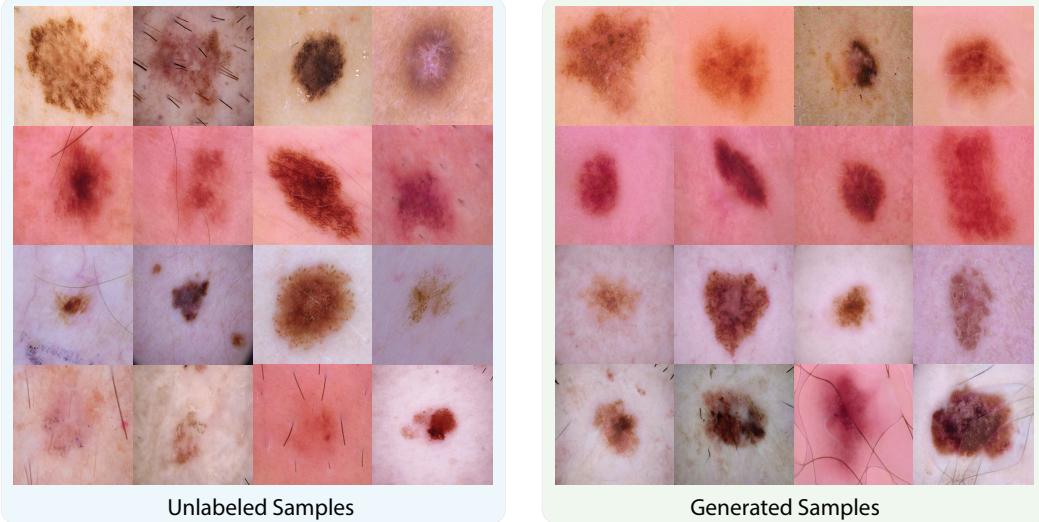


Figure 8: Uncurated set of novel images produced by StyleGAN on HAM10000 [38]. The generated skin cancer images are semantically similar to the unlabeled training samples. It is clear that compared to the images in BCN20000 [37], HAM10000 [38] has less diverse image texture.

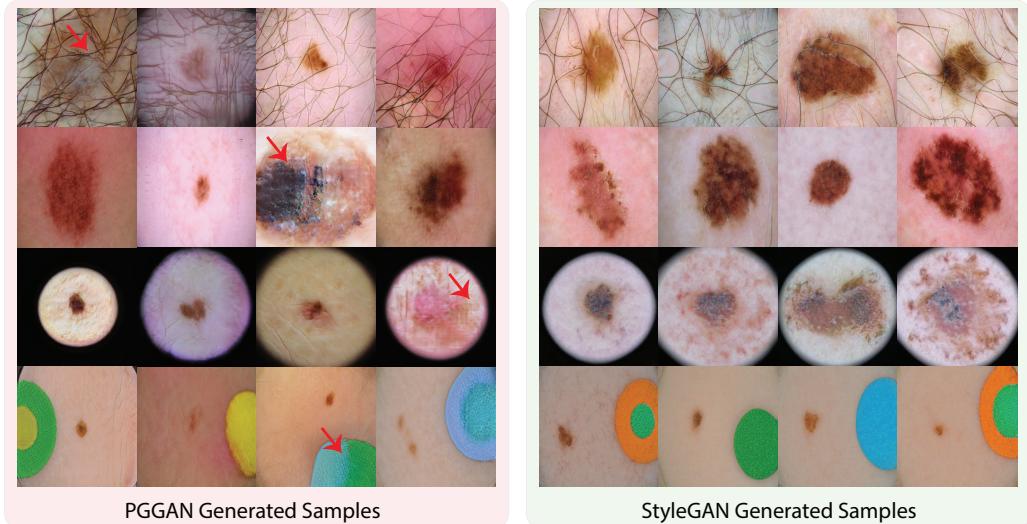


Figure 9: PGGAN and StyleGAN skin cancer image generation quality comparison. It is clear that overall StyleGAN generated skin cancer images have higher visual quality compared to those generated by PGGAN. As indicated by the red arrows, the skin cancer image details, such as hair, lesion texture, surrounding skin texture and color patches, are maintained sharper and more meaningful in StyleGAN generated samples.

skin cancer images. This is attributed to the non-linear mapping network from original latent space  $\mathcal{Z}$  into the  $\mathcal{W}$  space and the merging branch via adaptive instance normalization (AdaIN) at each convolutional layer [43], [44] in StyleGAN.

## V. DISCUSSION

In this paper, we showed that StyleGAN is capable of synthesising authentic skin cancer images. This is valuable because the ability to generate images like those

presented here helps ameliorate the significant problem of data scarcity. In particular, for rare skin cancer cases, the data augmentation benefit is even larger. Thus, the proposed approach can reduce the cost and human effort required for teledermatology.

Moreover, other mobile health apps will also suffer data scarcity issue at the early stage. We can apply the proposed method to other medical modalities as well. The generated images can also be used in other domains, such as medical image perception and medical image analysis.

A second goal of this paper was to test whether StyleGAN generated samples can be utilized for data augmentation for self-supervised learning. We found that StyleGAN-based data augmentation significantly boosted the performance of self-supervised skin cancer classification. Essentially, using the generated images helped the classifier better discriminate skin lesions. In a followup experiment, we found that the classification performance improvement was most significant in cases when there were fewer labeled training data. That is, the benefit of augmenting data is most pronounced when labeled data are scarce.

Compared to supervised learning, self-supervised learning only requires a small quantity of labeled data at the final training stage. Thus, with the gradually growing unlabeled training data from users, self-supervised learning system is easier to scale. Moreover, it only requires experts to label a small amount of key data. Therefore, it is more suitable for mobile health app systems. With our proposed generative self-supervised learning, the performance of mobile health apps could be much improved.

## VI. CONCLUSION

In this paper, we trained StyleGAN to augment the training dataset for self-supervised learning of skin cancer images for teledermatology. Our model was able to generate authentic skin cancer images, and those images were effective as a source of augmentation for self-supervised learning. The benefit of augmenting real datasets with StyleGAN-based generated data was most prominent when the original dataset was limited in size. Therefore, when real data are scarce (for example, in several types of skin cancer including Merkel cell carcinoma), the augmentation approach presented here could be highly beneficial. This, in turn, could be very helpful for mobile health applications.

## ACKNOWLEDGMENT

This work has been supported, in part, by National Institutes of Health (NIH) under Grant No. R01CA236793 to Z.R. and D.W., by National Science Foundation under Grant No. 1612843 and by Berkeley Deep Drive to Y.G. and S.Y.

## REFERENCES

- [1] G. P. Guy Jr, C. C. Thomas, T. Thompson, M. Watson, G. M. Massetti, and L. C. Richardson, “Vital signs: melanoma incidence and mortality trends and projections—united states, 1982–2030,” *MMWR. Morbidity and mortality weekly report*, vol. 64, no. 21, p. 591, 2015.
- [2] G. P. Guy Jr, S. R. Machlin, D. U. Ekwueme, and K. R. Yabroff, “Prevalence and costs of skin cancer treatment in the us, 2002- 2006 and 2007- 2011,” *American journal of preventive medicine*, vol. 48, no. 2, pp. 183–187, 2015.
- [3] R. Siegel, E. Ward, O. Brawley, and A. Jemal, “Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths.” *CA: a cancer journal for clinicians*, vol. 61, no. 4, pp. 212–236, 2011.
- [4] C. E. G. Silveira, T. B. Silva, J. H. G. T. Fregnani, R. A. da Costa Vieira, R. L. Haikel, K. Syrjänen, A. L. Carvalho, and E. C. Mauad, “Digital photography in skin cancer screening by mobile units in remote areas of brazil,” *BMC dermatology*, vol. 14, no. 1, pp. 1–5, 2014.
- [5] S. Markun, N. Scherz, T. Rosemann, R. Tandjung, and R. P. Braun, “Mobile teledermatology for skin cancer screening: a diagnostic accuracy study,” *Medicine*, vol. 96, no. 10, 2017.
- [6] A. Finnane, K. Dallest, M. Janda, and H. P. Soyer, “Teledermatology for the diagnosis and management of skin cancer: a systematic review,” *JAMA dermatology*, vol. 153, no. 3, pp. 319–327, 2017.
- [7] N. Chuchu, J. Dinnis, Y. Takwoingi, R. N. Matin, S. E. Bayliss, C. Davenport, J. F. Moreau, O. Bassett, K. Godfrey, C. O’Sullivan *et al.*, “Teledermatology for diagnosing skin cancer in adults,” *Cochrane Database of Systematic Reviews*, no. 12, 2018.
- [8] T. M. de Carvalho, E. Noels, M. Wakkee, A. Udrea, and T. Nijsten, “Development of smartphone apps for skin cancer risk assessment: progress and promise,” *JMIR Dermatology*, vol. 2, no. 1, p. e13376, 2019.
- [9] 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.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [12] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114.
- [13] J. Yap, W. Yolland, and P. Tschandl, “Multimodal skin lesion classification using deep learning,” *Experimental dermatology*, vol. 27, no. 11, pp. 1261–1267, 2018.
- [14] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, “Skin lesion classification using hybrid deep neural networks,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1229–1233.
- [15] C. Lam, C. Yu, L. Huang, and D. Rubin, “Retinal lesion detection with deep learning using image patches,” *Investigative ophthalmology & visual science*, vol. 59, no. 1, pp. 590–596, 2018.
- [16] T. Nair, D. Precup, D. L. Arnold, and T. Arbel, “Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation,” *Medical image analysis*, vol. 59, p. 101557, 2020.

- [17] Y. Yuan, M. Chao, and Y.-C. Lo, “Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance,” *IEEE transactions on medical imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.
- [18] M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, “Skin lesion segmentation in clinical images using deep learning,” in *2016 23rd International conference on pattern recognition (ICPR)*. IEEE, 2016, pp. 337–342.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [20] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *European conference on computer vision*. Springer, 2014, pp. 740–755.
- [21] A. Krizhevsky, G. Hinton *et al.*, “Learning multiple layers of features from tiny images,” 2009.
- [22] M. J. Willemink, W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R. M. Summers, D. L. Rubin, and M. P. Lungren, “Preparing medical imaging data for machine learning,” *Radiology*, vol. 295, no. 1, pp. 4–15, 2020.
- [23] J. Snell, K. Swersky, and R. S. Zemel, “Prototypical networks for few-shot learning,” *arXiv preprint arXiv:1703.05175*, 2017.
- [24] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, “Learning to compare: Relation network for few-shot learning,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1199–1208.
- [25] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, “Meta-transfer learning for few-shot learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 403–412.
- [26] L. Fei-Fei, R. Fergus, and P. Perona, “One-shot learning of object categories,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 594–611, 2006.
- [27] N. Hilliard, L. Phillips, S. Howland, A. Yankov, C. D. Corley, and N. O. Hodas, “Few-shot learning with metric-agnostic conditional embeddings,” *arXiv preprint arXiv:1802.04376*, 2018.
- [28] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, “Unsupervised feature learning via non-parametric instance discrimination,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3733–3742.
- [29] A. v. d. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *arXiv preprint arXiv:1807.03748*, 2018.
- [30] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.
- [31] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738.
- [32] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. G. Azar *et al.*, “Bootstrap your own latent: A new approach to self-supervised learning,” *arXiv preprint arXiv:2006.07733*, 2020.
- [33] U. D. of Health, H. Services *et al.*, “The surgeon general’s call to action to prevent skin cancer,” 2014.
- [34] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, “Semantic image synthesis with spatially-adaptive normalization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2337–2346.
- [35] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” in *International Conference on Learning Representations*, 2018.
- [36] 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.
- [37] M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig *et al.*, “Bcn20000: Dermoscopic lesions in the wild,” *arXiv preprint arXiv:1908.02288*, 2019.
- [38] P. Tschandl, C. Rosendahl, and H. Kittler, “The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Nature Scientific Data*, vol. 5, p. 180161, 2018.
- [39] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *arXiv preprint arXiv:1406.2661*, 2014.
- [40] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved training of wasserstein gans,” *arXiv preprint arXiv:1704.00028*, 2017.
- [41] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *International conference on machine learning*. PMLR, 2017, pp. 214–223.
- [42] A. Brock, J. Donahue, and K. Simonyan, “Large scale gan training for high fidelity natural image synthesis,” *arXiv preprint arXiv:1809.11096*, 2018.
- [43] V. Dumoulin, J. Shlens, and M. Kudlur, “A learned representation for artistic style,” *arXiv preprint arXiv:1610.07629*, 2016.
- [44] X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1501–1510.

- [45] 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, 2018.
- [46] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint arXiv:1511.06434*, 2015.
- [47] A. Odena, C. Olah, and J. Shlens, “Conditional image synthesis with auxiliary classifier gans,” in *International conference on machine learning*. PMLR, 2017, pp. 2642–2651.
- [48] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, “Gan-based synthetic brain mr image generation,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 734–738.
- [49] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, “Medical image synthesis with context-aware generative adversarial networks,” in *International conference on medical image computing and computer-assisted intervention*. Springer, 2017, pp. 417–425.
- [50] B. Cao, H. Zhang, N. Wang, X. Gao, and D. Shen, “Auto-gan: self-supervised collaborative learning for medical image synthesis,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 10486–10493.
- [51] A. Bissoto, F. Perez, E. Valle, and S. Avila, “Skin lesion synthesis with generative adversarial networks,” in *OR 2.0 context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis*. Springer, 2018, pp. 294–302.
- [52] D. Bisla, A. Choromanska, R. S. Berman, J. A. Stein, and D. Polksy, “Towards automated melanoma detection with deep learning: Data purification and augmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.
- [53] A. Ghorbani, V. Natarajan, D. Coz, and Y. Liu, “Dermgan: Synthetic generation of clinical skin images with pathology,” in *Machine Learning for Health Workshop*. PMLR, 2020, pp. 155–170.
- [54] A. Bissoto, E. Valle, and S. Avila, “Gan-based data augmentation and anonymization for skin-lesion analysis: A critical review,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1847–1856.
- [55] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, “Learning deep representations by mutual information estimation and maximization,” *arXiv preprint arXiv:1808.06670*, 2018.
- [56] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, “Unsupervised learning of visual features by contrasting cluster assignments,” *arXiv preprint arXiv:2006.09882*, 2020.
- [57] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “An unsupervised learning model for deformable medical image registration,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9252–9260.
- [58] S. Zhao, T. Lau, J. Luo, I. Eric, C. Chang, and Y. Xu, “Unsupervised 3d end-to-end medical image registration with volume tweening network,” *IEEE journal of biomedical and health informatics*, vol. 24, no. 5, pp. 1394–1404, 2019.
- [59] D. Kuang and T. Schmah, “Faim—a convnet method for unsupervised 3d medical image registration,” in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2019, pp. 646–654.
- [60] K. Armanious, C. Jiang, S. Abdulatif, T. Küstner, S. Gatidis, and B. Yang, “Unsupervised medical image translation using cycle-medgan,” in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.
- [61] X. Li, M. Jia, M. T. Islam, L. Yu, and L. Xing, “Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4023–4033, 2020.
- [62] S. S. Boushehri, A. B. Qasim, D. Waibel, F. Schmich, and C. Marr, “Annotation-efficient classification combining active learning, pre-training and semi-supervised learning for biomedical images,” *bioRxiv*, 2020.
- [63] D. Wang, N. Pang, Y. Wang, and H. Zhao, “Unlabeled skin lesion classification by self-supervised topology clustering network,” *Biomedical Signal Processing and Control*, vol. 66, p. 102428, 2021.
- [64] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen *et al.*, “Big self-supervised models advance medical image classification,” *arXiv preprint arXiv:2101.05224*, 2021.
- [65] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [66] 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 *International workshop on simulation and synthesis in medical imaging*. Springer, 2018, pp. 1–11.
- [67] S. K. Lim, Y. Loo, N.-T. Tran, N.-M. Cheung, G. Roig, and Y. Elovici, “Doping: Generative data augmentation for unsupervised anomaly detection with gan,” in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 1122–1127.
- [68] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, “Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection,” *Ieee Access*, vol. 8, pp. 91916–91923, 2020.
- [69] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.