A Survey of Synthetic Medical Image Generation For Improving Disease Classification

Bradley Hu
Stanford University
bradleyh@stanford.edu

Michael Maffezzoli
Stanford University
maff2023@stanford.edu

Brendan Mclaughlin
Stanford University
bmc0407@stanford.edu

Abstract

Ocular diseases are a major cause of visual impairment and blindness, necessitating early detection and accurate diagnosis to improve treatment outcomes. However, the scarcity of comprehensive labeled datasets for various ocular conditions poses a significant challenge for training robust deep learning models. This study explores the potential of generative models to address this limitation by creating synthetic images to augment the limited training datasets available for ocular diseases. We experimented with four generative methods: Variational Autoencoders (VAEs), finetuning a pretrained StyleGAN, SmallGAN, and Stable Diffusion. Our results demonstrate that the StyleGAN2-ADA approach produced the highest quality synthetic images, and showing resemblance to real fundus photographs, despite challenges with vein detail and image blurriness. VAE and SmallGAN showed potential but were limited by dataset size and model complexity. Stable Diffusion required extensive compute resources and did not achieve the desired image detail. This study underscores the potential of synthetic data augmentation to overcome the limitations of small medical image datasets and improve automated disease classification systems.

1. Introduction

Ocular diseases are a leading cause of visual impairment and blindness worldwide. Early detection and accurate diagnosis are crucial for preventing disease progression and improving treatment outcomes. Traditional diagnostic methods rely heavily on the expertise of ophthalmologists and the quality of fundus photographs. However, the increasing prevalence of ocular diseases necessitates the development of automated and intelligent diagnostic systems to assist clinicians and improve patient outcomes.

While tasks such as skin melanoma classification benefit from abundant labeled training data, facilitating the development of robust CNN models, the classification of various ocular diseases suffers from a scarcity of comprehensive labeled data. This limitation necessitates the exploration of alternative methods to enhance training datasets and improve model performance.

In this study, we focus on generating synthetic images of ocular diseases to supplement the limited training dataset available for these conditions. Specifically, we experiment with four generative methods: Variational Autoencoders (VAEs), finetuning a pretrained StyleGAN, SmallGAN, and Stable Diffusion. By leveraging these generative approaches, our aim is to create a more robust and diverse dataset that can support effective training of deep learning models. The synthetic images generated through these methods are intended to supplement the existing dataset and future work will investigate whether these augmented datasets improve a classifier's ability to accurately classify various ocular diseases.

2. Related Works

2.1. Transfer Learning as a Baseline

In "Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets," Spolaor et al. explore the effectiveness of fine-tuning pre-trained CNNs for medical image classification with limited data. The study investigates eight fine-tuning strategies on VGG-based networks initially trained on ImageNet, using the ISIC and PH2 dermoscopic image datasets as test cases [13]. Baseline models without fine-tuning achieved accuracies of 78.5% and 80.3% on the ISIC and PH2 datasets, respectively, and fine-tuning improved these accuracies by around 10%. The findings suggest that with appropriate fine-tuning, pretrained CNNs can be effectively leveraged for medical image classification tasks.

2.2. Synthetic Data Improves Classifier

In "Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images," Yoo et al. address the challenge of detecting conjuncti-

val melanoma with limited labeled data (403 ocular surface images). To enhance the dataset, they employed CycleGAN and PGGAN to generate synthetic images, expanding the training data to 2000 images through linear transformations. The synthetic images were used to finetune several pre-trained deep learning architectures and yielded significant improvements in classification accuracy with the GAN-augmented dataset [15].

2.3. Data Augmentation Methods

2.3.1 Variational Auto-Encoder (VAE)

In the original VAE paper, "Auto-Encoding Variational Bayes" [9], Kingma and Welling introduce the variational auto-encoder, which allows efficient approximate posterior inference and model parameter learning without expensive iterative processes. This foundational work demonstrated that VAEs could effectively model the probability distribution of training data in a latent space, setting the stage for their use in image generation.

Han et al. enhanced the HP-VAE-GAN model by incorporating the Convolutional Block Attention Module (CBAM) into the encoder of the Patch-VAE component, resulting in higher-quality synthetic images. Their study showed that combining real and generated images significantly improves classification accuracy compared to using only real images [5]. Similarly, Gur et al. demonstrated the effectiveness of HP-VAE-GAN in generating diverse, high-quality video samples from minimal training data, highlighting its potential in overcoming mode collapse and outperforming baseline models [4].

2.3.2 Stable Diffusion

In the paper, "Denoising Diffusion Probabilistic Models", Ho et al.introduce diffusion probabilistic models, a novel class of latent variable models that generate high-quality images by progressively adding and then removing Gaussian noise from data. This process, trained via variational inference, achieved state-of-the-art results, including an FID score of 3.17 on CIFAR-10 [6]. While the results of diffusion are promising, these models may require a larger amount of training data than other generative methods to produce detailed synthetic images.

Interestingly, Bansal et al. found that the type of noise does not matter at all for diffusion. In their paper, "Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise" [2], they introduce cold diffusion, which generalizes diffusion models to use arbitrary image transformations (i.e. blur, masking, downsampling) instead of Gaussian noise for both the forward degradation and reverse restoration processes. The success of cold diffusion without noise opens up new possibilities for building diffusion models with different properties by using various degradation processes.

2.3.3 General Adversarial Networks (GANs)

In "Generative Adversarial Nets" [3], Goodfellow et al.introduce the framework for GANs, where a generative model (G) and a discriminative model (D) are trained simultaneously in a minimax two-player game to produce realistic synthetic images without the need for Markov chains or inference during training. This paper set the foundation for many variants of GANs, including two explored in this paper, smallGAN and StyleGAN.

Noguchi and Harada's "Image Generation From Small Datasets via Batch Statistics Adaptation" presents a method for generating images from small datasets by adapting a pre-trained generator's scale and shift parameters while keeping the convolutional kernel parameters fixed. This technique significantly reduces the number of trainable parameters, enabling the generator to learn from small datasets without overfitting [11]. Experiments demonstrated that the proposed method achieves higher quality and more diverse image generation compared to fine-tuning the entire generator or only a subset of layers.

In another paper, "A Style-Based Generator Architecture for Generative Adversarial Networks", Tero Karras et al. propose an adaptive discriminator augmentation mechanism that stabilizes GAN training in cases of limited training data [8]. The authors introduce an adaptive augmentation strategy for the discriminator that prevents overfitting, a common issue when training GANs with small datasets that leads to the discriminator becoming too confident in distinguishing between real and fake images, which subsequently provides poor feedback to the generator. The adaptive discriminator augmentation has become a state-of-the-art method for limited data scenarios and demonstrated that high-quality GANs can be trained with much smaller datasets than previously thought possible.

3. Methods

3.1. Variational Autoencoder (VAE)

We implemented a VAE from scratch in PyTorch to address the challenge of limited labeled data in eye disease classification. In general, VAE consists of two main components: the encoder and the decoder. The encoder compresses the input images into a latent space, where it generates two sets of parameters: the mean and variance of a Gaussian distribution. The decoder then takes samples from the distribution to reconstruct the original images. During training, the model minimizes the reconstruction error between the original and the reconstructed images while also regularizing the latent space to follow a Gaussian distribution. In addition to reconstruction error, the loss function for VAE typically includes Kullback-Leible (KL) divergence loss as a regularization term. The KL divergence loss measures the difference between the learned probability distri-

bution and the predefined prior distribution.

The loss function for a Variational Autoencoder (VAE) can be written as:

$$\mathcal{L}(\theta, \phi; \mathbf{x}, \mathbf{z}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$

where:

$$\mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) = \int q_\phi(\mathbf{z}|\mathbf{x}) \log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p(\mathbf{z})} d\mathbf{z}$$

Here: - $\mathbf{x}$ is the input data. - $\mathbf{z}$ is the latent variable. - $q_\phi(\mathbf{z}|\mathbf{x})$ is the approximate posterior distribution. - $p_\theta(\mathbf{x}|\mathbf{z})$ is the likelihood of the data given the latent variable. - $p(\mathbf{z})$ is the prior distribution of the latent variable. - $\mathbb{E}$ denotes the expectation. - KL denotes the KL divergence.

To apply the VAE to our fundus dataset, we followed these steps:

1. **Data Preparation:** We preprocessed the images by resizing them to 64x64 pixels and normalizing the pixel values to a range of [-1, 1].

2. **Model Architecture:** The encoder includes two linear layers activated by LeakyReLU functions, producing the mean and variance of the latent space. Similarly, the decoder uses two linear layers and LeakyReLU activations to reconstruct the images, with a final sigmoid activation to produce pixel values in the range [0, 1].

3. **Training:** We trained the VAE by minimizing the combination of reconstruction loss and KL divergence loss. The Adam Optimizer, with a learning rate of 1×10^{-3}, was used to update the model parameters.

4. **Cross Validation:** We ran a hyper-parameter search to find the optimal learning rate, hidden dimension, and latent dimension which were 2048, 128, and 2×10^{-3} respectively.

3.2. Stable Diffusion

We also experimented with using diffusion models to generate synthetic data samples of fundus images. Diffusion models work by gradually adding noise to a training image and learning how to reverse that noise at each time step. The final model should then be able to create similar images from Gaussian noise by sampling the learned parameters. A diagram of this is shown below in Figure 1.

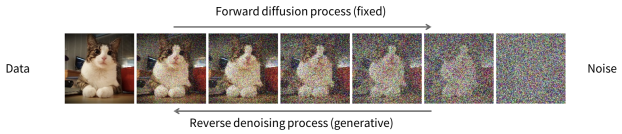

Figure 1. Simple diffusion image generation process diagram.

Algorithm 1 Training

repeat
  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
  $t \sim \mathrm{Uniform}(\{1, \dots, T\})$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Take gradient descent step on
    $\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$
until converged

Algorithm 2 Sampling

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, \dots, 1$ **do**
  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
  $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
end for
return $\mathbf{x}_0$

Figure 2. Diffusion training and sampling algorithms from original paper.

More technically, a diffusion probabilistic model is a parameterized Markov chain trained using variational inference to produce samples matching the data after finite time [6]. Ho et al. summarize the algorithm in the figure below.

For our experiment we used a Github respository [14] that created a package compatible with colab and pytorch, and was built based on Ho et al.'s original tensorflow implementation. The code for this can be found at https://github.com/lucidrains/denoising-diffusion-pytorch. The package includes a Trainer class that allowed us to pass in a data folder along with hyperparameters to then train and sample the diffusion model.

3.3. SmallGAN

Another approach we tried fine-tuning a pre-trained generator by introducing scale and shift parameters to each hidden activation of the generator and updating only these parameters, enabling the model to be transferred to a small dataset due to the low number of trainable parameters. This approach was adapted from Noguchi and Harada's "Image Generation From Small Datasets via Batch Statistics Adaptation" which was able to achieve promising results from datasets containing roughly 100 images [11].

Closely following the approach proposed by the paper, our implementation contained only a pre-trained generator with no discriminator [10]. We then introduced trainable scale and shift parameters into each of the hidden activations of the generator to transfer the model to our small dataset. By used a pre-trained model with a smaller number of trainable parameters overfitting, a common issue with GAN is reduced.

Instead of a discriminator, supervised learning is used in this approach by directly passing a trainable latent vector z through the generator such that the generated image is as close as possible to the true image.

This is an intuitive approach for small datasets because in a typical GAN, a separate discriminator is being trained to determine given image x, is real or fake. This works well

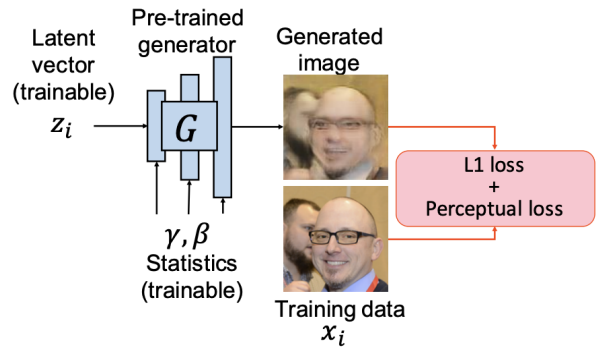

Figure 3. SmallGAN training pipeline

for large datasets that densely fill the distribution but is a very hard task with a small dataset that is better suited for a distance-based supervised loss approach.

Our implementation forks Noguchi and Harada's implementation, modifying the data preprocessing and hyperparameters to suit our dataset. Because we had roughly 300 images per disease class, which is close to the 100-image dataset for which SmallGAN is optimized, we trained on a single class (the Glaucoma class) instead of the whole dataset.

3.4. Finetuned StyleGAN

We employed the StyleGAN2-ADA (Adaptive Discriminator Augmentation) [7] implementation to generate synthetic images of ocular diseases. The finetuning process was performed on the pretrained *ffhq.pkl* model, which was trained on the FFHQ dataset [12] at a resolution of 1024x1024 using the original StyleGAN2.

The ODIR dataset is used for finetuning. The images in the dataset are preprocessed to be standardized 1024x1024 pixels and RGB format. The pretrained *ffhq.pkl* model is then loaded as the base model for finetuning. To address the issue of limited data, StyleGAN2-ADA implements adaptive discriminator augmentation (ADA) on top of the original StyleGAN2 implementation, and involves adjusting the augmentation probabilities based on the discriminator's performance during training. This technique helps prevent overfitting and stabilizes the training process, allowing for effective learning even with a small dataset. We set up the generator and discriminator to require gradients for backpropogation, and both models are set in training mode to allow for proper gradient updates. We then feed the preprocessed images into the training pipeline with an initial batch size of 8, learning rate of 1e-4, and trained for 10 epochs. After training, the finetuned generator and discriminator models are saved for image generation.

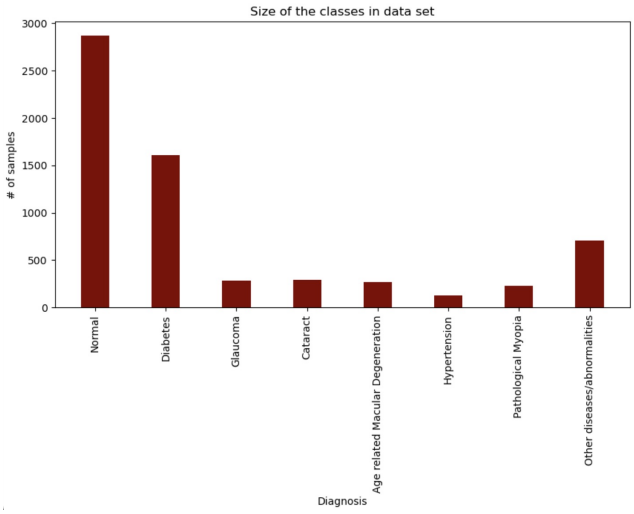

Figure 4. Dataset Class Sizes

4. Dataset and Features

For this study, we leverage the Ocular Disease Intelligent Recognition (ODIR) dataset, a structured ophthalmic database comprising data from 5,000 patients, collected by Shanggong Medical Technology Co., Ltd. from various hospitals and medical centers in China [1]. The dataset includes color fundus photographs of both left and right eyes, along with diagnostic keywords provided by trained human readers.

The ODIR dataset is designed to represent a real-life set of patient information, with fundus images captured using various cameras, such as Canon, Zeiss, and Kowa, leading to varied image resolutions. This diversity in imaging conditions presents a realistic challenge for developing robust diagnostic models.

Patients in the ODIR dataset are classified into eight categories based on their ocular conditions:

1. **Normal (N)** - 2873 images
2. **Diabetes (D)** - 1608 images
3. **Glaucoma (G)** - 284 images
4. **Cataract (C)** - 293 images
5. **Age-related Macular Degeneration (A)** - 266 images
6. **Hypertension (H)** - 128 images
7. **Pathological Myopia (M)** - 232 images
8. **Other diseases/abnormalities (O)** - 708 images

The images in the dataset vary in size, angle, and zoom, but most are 512x512. Below are examples from each class (all left eyes):

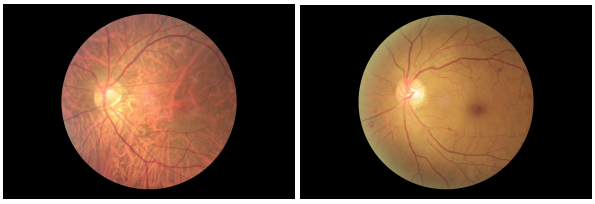

Figure 5. N (left) and D (right)

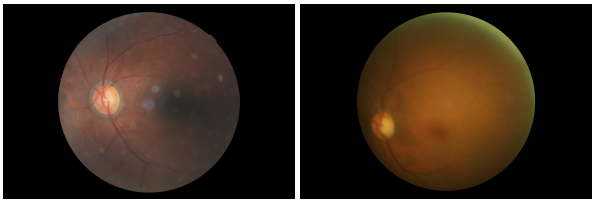

Figure 6. G (left) and C (right)

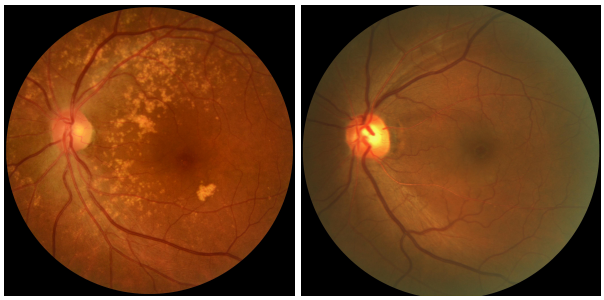

Figure 7. A (left) and H (right)

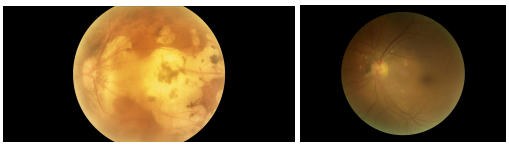

Figure 8. M (left) and O (right)

5. Experiments, Results, and Discussion

5.1. Preliminary VAE

Through implementing two versions of our VAE in which we experimented with activation functions and hyperparameters, we were able to drastically improve the synthetic images we generated. On our initial dataset, which was a set of 411 photographs of eyes with different diseases, our VAE was not able to produce the quality of synthetic images we needed to actually use them in training. Below are the initial images that the two versions of our VAE produced.

We later learned that VAEs suffer when trained on small datasets. This was one of the contributing factors that caused us to switch our dataset to the medium-sized fundus dataset described above. With the larger dataset, we re-trained our VAE after conducting a new hyperparameter search. We were able to get much higher quality results but

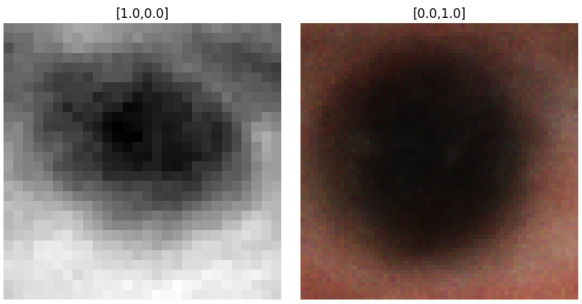

Figure 9. VAE version 1 (left) and 2 (right) synthetic eye images

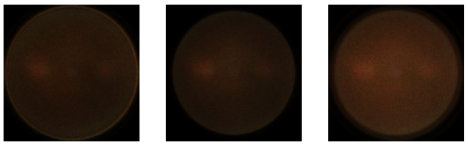

Figure 10. VAE synthetic fundus photographs.

after a qualitative evaluation, we determined that the image classes were not clearly distinguishable from one another. We began training with a learning rate of 0.0001 and a batch size of 32. We noticed consistent spikes in the loss curve as it decreased meaning that our learning rate was likely too high and causing the model to overfit to each batch. Additionally, we noticed that our GPU utilization was low. In response, we tripled the batch size and doubled the learning rate to manifest a proportionally lower learned rate with respect to the batch size. Figure 9 depicts the results produced by our VAE implementation trained on 6,484 fundus photographs with a final learning rate of 0.0002 a latent dimension of 128, a hidden dimension of 2048, and a batch size of 96. We trained for 3 epochs which took just over an hour on an L4 GPU.

The images are blurry which is a common characteristic of VAEs when undertrained. With more training data and more compute and time for hyperparameter searching, we likely could improve our VAE experiment results. While VAE was worth trying, we now better understand why GAN-based architectures are typically the strongest choice for medical synthetic image generation.

5.2. Stable Diffusion

We executed the stable diffusion package for 1000 training steps where the Gaussian diffusion had 1000 time steps and a sample time step of 250. We ran the training with a batch size of 32, a learning rate of 8×10^{-5} and accumulated the gradient every 2 steps. The output images were set at 128x128. Below are four samples from this model.

This experiment took just over an hour to run and had a final loss of 0.015, which was promising. As the images show, however, the diffusion process was unable to learn the precise details of fundus images like the veins

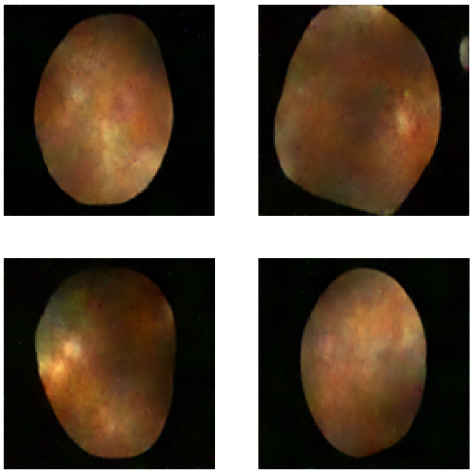

Figure 11. Diffusion experiment samples.

and optic nerve head. We hypothesize that the lack of color variety and the very subtle differences between images and classes necessitate many more training iterations and higher-resolution output. We attempted these parameters but quickly ran out of compute, and even at lower resolutions, with many iterations the training timeline became days to weeks for a single GPU.

5.3. SmallGAN

After adapting the data preprocessing to suit our needs, we started by making an initial selection of hyperparameters based on the goal of leveraging our 300 image class datasets even though the model was architected for under 100 image datasets. To accommodate this, we slightly increased the batch size effectively lowering the effect of the learning rate and increasing parallelization. The default implementation began training the pre-trained GAN on top of another checkpoint. We decided to drop this checkpoint and extend our training to allow the model to learn a little more complex provided by our 3x larger dataset.

As seen in Figure 12, our results show some promise, clearly displaying distinct coloration of the fundus which is a key variable in disease classification. Similar to the results shown in Noguchi and Harada's paper, the images are rather blurry suggesting poor edge detection and undertraining.

5.4. StyleGAN

Through iterative experimentation with the StyleGAN2-ADA model, we attempted to generate high-quality synthetic images of ocular diseases. Our initial training setup was characterized by a batch size of 8 and learning rate of 1e-4, running for 10 epochs. This configuration, however, resulted in extremely large losses for both the generator and discriminator, which was indicative of exploding gradients. To address the issue, gradient clipping was applied, which helped stabilize training.

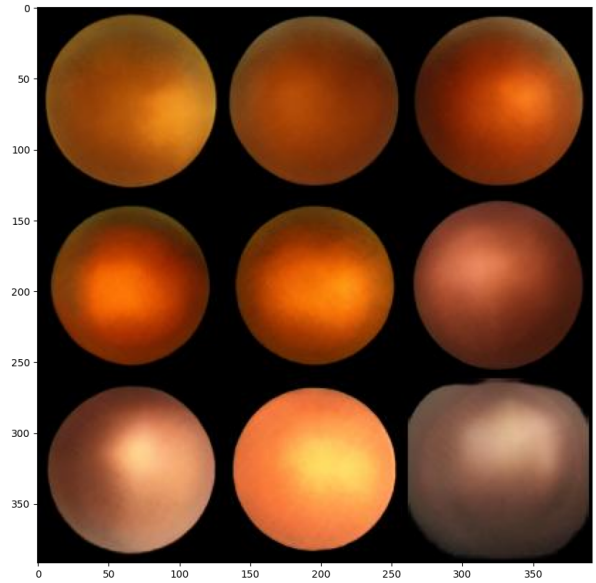

Figure 12. SmallGAN trained on Glaucoma class dataset (284 images)

Furthermore, the discriminator loss initially dropped rapidly while the generator loss remained large, which indicated that the discriminator was learning too well. This imbalance suggested that the generator struggled to produce realistic images while the discriminator easily distinguished between real and synthetic images. To combat this, we adjusted our training strategy by training the discriminator more frequently than the generator (every other image), to improve the discriminator's ability to challenge the generator.

The adjusted training strategy improved the stability of the training process and quality of the generated images. Figure X shows the the progression of synthetic images produced over epochs. Training the StyleGAN model for 10 epochs took approximately 3.5 hours on an NVIDIA T4 GPU. Among the four generative methods explored, StyleGAN2-ADA consistently produced synthetic images that most closely resembled the real fundus images in terms of structure and detail. This can be attributed to the adaptive discriminator augmentation architecture, which stabililizes GAN training even with limited data.

However, despite this, there was still a noticeable loss in fidelity of certain details, such as vein structures, color accuracy, and overall blurriness in the generated images. We believe that further hyperparameter tuning and optimizations such as learning rate scheduling and Batch Size Adjustments could improve upon these shortcomings. Additionally, given more time, we believe extended training duration could further improve results.

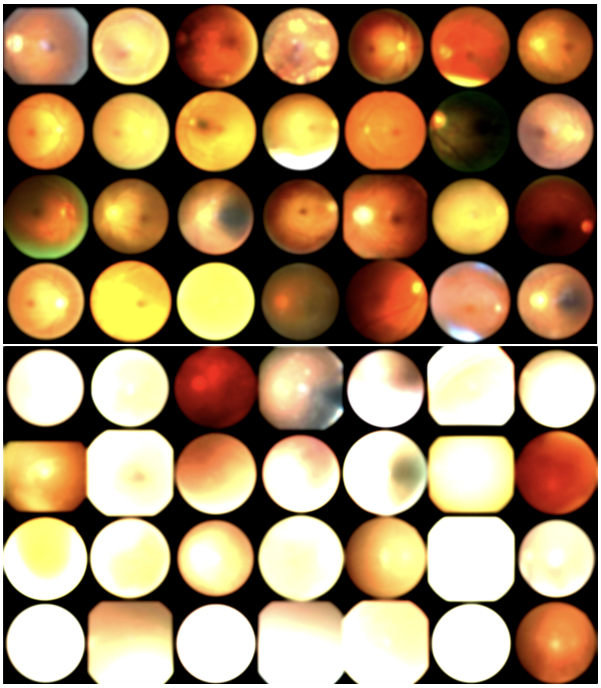

Figure 13. StyleGAN trained on full ODIR dataset (top) and StyleGAN trained on Glaucoma (G) class only (bottom)

6. Conclusion and Future Work

In this study, we explored the potential of four generative models - VAEs, finetuned StyleGAN, SmallGAN, and Stable Diffusion - for synthesizing fundus images to augment the limited ODIR dataset for ocular disease classification. While the VAE and diffusion models struggled to capture fine diagnostic details, the StyleGAN and SmallGAN approaches showed promise in generating realistic and diverse fundus images. The results of this study demonstrate the potential of generative models, particularly GANs, for synthesizing realistic fundus images to augment limited datasets. However, further work is needed to generate higher resolution images with finer diagnostic details. Potential improvements include modifying architectures, loss functions, and training schemes to encourage the models to capture subtle features of different ocular diseases. Experimenting with additional techniques such as transfer learning, attention mechanisms, and class conditioning could also yield benefits.

Once sufficiently realistic synthetic images can be produced, the next key step is to evaluate their impact on downstream disease classification performance. Controlled experiments should assess whether augmenting real datasets with increasing proportions of synthetic data boosts accuracy, sensitivity and specificity of ocular disease classifiers. It will also be important to examine the failure modes and biases of classifiers trained on hybrid real/synthetic data.

Finally, to maximize clinical utility, future research should seek to generate synthetic images spanning the full spectrum of disease presentations and severities observed in clinical practice. Rarer conditions and imaging artifacts are particularly critical to capture. Techniques for amplifying the representation of uncommon pathologies and features in the synthetic data could help develop more robust diagnostic models for real-world deployment. Collaboration with ophthalmologists will be essential to guide medically relevant data augmentation.

Synthetic data augmentation is a promising approach to alleviating the scarcity of labeled data, which currently limits deep learning for ocular disease diagnosis. With further research to refine image generation and validate clinical utility, this technique could accelerate the development of effective AI screening and diagnostic systems to preserve vision and improve access to ophthalmic care worldwide.

7. Contributions and Acknowledgements

1. Bradley Hu: StyleGAN model and generation, background research, report writing

2. Brendan McLaughlin: HP-VAE-GAN investigation, VAE model and generation, SmallGAN model and generation, report writing

3. Michael Maffezzoli: Milestone classifier, dataset preparation, diffusion model and generation, report writing

References

[1] Ocular disease recognition - odir5k. https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k/data. Accessed: 2024-06-05. 4

[2] A. Bansal, E. Borgnia, H.-M. Chu, J. S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise, 2022. 2

[3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014. 2

[4] S. Gur, S. Benaim, and L. Wolf. Hierarchical patch VAE-GAN: generating diverse videos from a single sample. *CoRR*, abs/2006.12226, 2020. 2

[5] Y. Han, Y. Liu, and Q. Chen. Data augmentation in material images using the improved hp-vae-gan. *Computational Materials Science*, 226:112250, 2023. 2

[6] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models, 2020. 2, 3

[7] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training generative adversarial networks with limited data, 2020. 4

[8] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks, 2019. 2

[9] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2022. 2

[10] A. Noguchi. Small dataset image generation, 2020. Accessed: 2024-06-05. 3

[11] A. Noguchi and T. Harada. Image generation from small datasets via batch statistics adaptation, 2019. 2, 3

[12] NVlabs. Flickr-faces-hq dataset (ffhq), 2019. Accessed: 2024-06-05. 4

[13] N. Spolaôr, H. D. Lee, A. I. Mendes, C. V. Nogueira, A. R. S. Parmezan, and W. S. R. Takaki. Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets. *Multimedia Tools and Applications*, 83:27305–27329, 2024. 1

[14] P. Wang. Denoising diffusion probabilistic models in pytorch, 2021. Accessed: 2024-06-05. 3

[15] T. Yoo. Conjunctival melanoma detection using deep learning in smartphone images. 2020. 2