Abstract:
Currently there is strong interest in data-driven approaches to medical image classification. However, medical imaging data is scarce, expensive, and fraught with legal concerns regarding patient privacy. Typical patient data consent forms only allow images to be used in medical journals or for education purposes, meaning the majority of medical data is not available for general public research. Synthetic medical images promise a solution to these problems. We propose a novel, two-stage pipeline for generating synthetic medical images from a pair of generative adversarial networks (GANs), which we test in practice on retinal fundus images. The first stage generates synthetic segmentation masks and the second stage converts the masks to photorealistic images. The images from Stage-II, along with their corresponding segmentations from Stage-I, are used to train a u-net segmentation network. On the u-net, our synthetic data pipeline received an F1 score of 0.8877, in comparison to a score of 0.8988 when tested with ground truth data, showing a negligible difference between synthetic and real patient datasets.

I. INTRODUCTION

Computer aided medical diagnosis is a powerful tool that medical professionals use to assist in the interpretation of medical images. Recently, deep learning algorithms have shown the potential to perform as well or even better than humans in certain medical image understanding tasks, such as segmentation and classification [1]. Along with accuracy, deep learning improves the efficiency of data analysis tremendously, due to its automated and computational nature. Since most medical data is produced in large volumes, and is often 3-dimensional (MRIs, CTs, etc.), it can be cumbersome and inefficient to manually annotate.

There is strong interest in computer aided medical diagnosis systems that rely on machine learning techniques [2]. However, due to proprietary and privacy reasons limiting data access [3], the development and advancement of these systems cannot be accelerated by public contributions. Researchers are not allowed to make any type of medical image public without patient consent [4]. In addition, the publicly available datasets often lack size and expert annotations, rendering them useless for the training of data-hungry neural networks. The design of these systems is therefore done exclusively by researchers that have access to private data, limiting the growth and potential of this field of research.

In the last 10 years, many breakthroughs in artificial intelligence attribute success to extensive public datasets such as ImageNet. The annual ImageNet competition decreased image recognition error rates from 28.2% to 6.7% [5] in the span of 4 years from 2010 to 2014. This showcases that the presence of large and accurate datasets is extremely important for building accurate models. However, current research in the field of medical imaging has relied on hand-tuning models rather than addressing the underlying problem with data. We believe that a public dataset for medicine can spark exponential growth in imaging tasks.

We propose a novel pipeline for generating synthetic medical images, allowing for the production of a public and extensive dataset, free from privacy concerns.

II. RELATED WORKS

Researchers across a variety of disciplines have taken private data to the public domain using synthetic data. For example, the U.S. Census collects personally identifiable information (PII) such as occupation, education, income, and geographical data for the US population. Due to the natural specificity of the data, even if sources are de-identified and obfuscated, there is considerable risk of deanonymization [6]. This valuable data, which holds many potentially useful hidden statistical correlations, is publically unavailable because of privacy issues. Reiter, a researcher at Duke University, solved this privacy problem by generating synthetic business census data [7]. In 2011 their work was released in the form of a Synthetic Longitudinal Business Database [8], the first time a public record-level database on business establishments was made available.

As seen in Reiter’s research, previous uses of synthetic data in order to bring private data to the public domain had been purely with scalar quantities. With the growing power of data-driven computer vision techniques, we explore in this project the idea of synthetic data for images. Recent developments of neural networks, specifically the GAN (generative adversarial network) [11], promise the possibility for more realistic image generation. However,
images produced by a GAN may often still contain artifacts and noise, due to the instability when locating a saddle point in the energy landscape. We address this issue by creating a novel image generation pipeline using a pair of GANs to promote increased stability.

III. GENERAL PIPELINE

To generate a high quality synthetic dataset, we propose the use of two GANs, breaking down the generation problem into two parts:

Stage-I GAN: Produce segmentation masks that represent the variable geometries of the dataset.

Stage-II GAN: Translate the masks produced in Stage-I to photorealistic images.

IV. DATA

We used the DRIVE database [9] for Stage-I of our pipeline. It contains forty pairs of retinal fundus images and segmentation masks extracted manually by two experts. The segmentation images were cropped to 512x512 pixels. Stage-II was provided with segmentation masks, derived from a CNN segmentation network on the MESSIDOR database [10]. The CNN segmentation network was trained with the images from DRIVE to create an alternate dataset of corresponding ground truth segmentation masks and retinal fundus images. We also used DRIVE to train the single stage GAN to compare our results.

![Figure 1. Example vessel tree segmentation mask and retina fundus image from DRIVE.](image)

V. GENERATIVE ADVERSARIAL NETWORK

The Generative Adversarial Network (GAN), as proposed by Goodfellow et al. [11] in June 2014, involves the competition between two models: the discriminator D and the generator G. D is a binary classifier that classifies the data produced by G as either part of the training set (realistic) or not (unrealistic). G minimizes its loss function by producing data that D will classify as real, as modeled by:

$$\min_{D, G} \max_{D} \mathbb{E}_{x \sim p_{\text{data}}} \left[ \log D(x) \right] + \mathbb{E}_{z \sim p_z} \left[ \log (1 - D(G(z))) \right]$$

The discriminator is a standard convolutional neural network (CNN) that takes an input image and returns a scalar that represents how real the input image is. There are two convolutional layers that identify 5x5 pixel features and, as with most CNNs, there are fully connected layers at the end. The generator is initialized with a random noise vector while D is trained with a small set of ground truth data. The generator is a deeper neural network, having more convolutional layers and nonlinearities. The noise vector is upsampled and the weights of G are learned through backpropagation, eventually producing data that is classified as real by the discriminator.

We utilized this novel neural network model to create our pipeline for the generation of synthetic medical images.

VI. STAGE-I GAN

The purpose of the Stage-I GAN is to generate variable segmentation masks. It is based on the deep convolutional generative adversarial network (DCGAN) architecture [12], and built on the TensorFlow platform. This network has demonstrated competitive results while simultaneously improving training stability in comparison to the standard GAN. The distinctive feature of the DCGAN, compared to other generative models, is it being fully convolutional, meaning convolutional layers were used instead of pooling layers. Pooling layers reduce the spatial size of the representation, and although they improve computational efficiency, they also result in the loss of important features found in medical images. The generator is initialized with a noise vector, which is fed through multiple strided convolutions to generate a synthetic image.

We used the cross-entropy loss function to train the discriminator in the Stage-I GAN:

$$l_D = \frac{1}{m} \sum_{i=1}^{m} [\log(D(G(z^i))) + \log(1 - D(x^i))]$$

where D is the discriminator, G is the generator, m refers to mini-batch size, z is the corresponding input noise vector, x is the image, and i is the index of the image. The generator’s loss is described by:

$$l_G = \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(x^i))$$
As a result of these two connected loss functions, the generator and discriminator are constantly competing with each other to minimize their respective loss functions. We trained this GAN for 8 hours on an NVIDIA Tesla K80 GPU.

**VII. STAGE-II GAN**

The purpose of Stage-II GAN is to translate segmentation masks to corresponding photorealistic images. Stage-II is also built on the TensorFlow platform. Our model is based on an image-to-image translation network proposed by Isola et al. [13] in November 2016. Specifically, for the retinal images, we use a vessel-to-retina implementation built by Costa et al. [14].

This network is a special form of GAN known as a conditional generative adversarial network (CGAN). It aims to condition the two networks D and G to some vector y and input image X that represents the mapping between the segmentation mask and photorealistic image. Similar to the regular GAN, the CGAN can be modeled by this function (with the additional parameter y):

\[
\min_G \max_D V(D, G) =
\]

\[
E_{p_{\text{data}}} [\log D(x, y)] + E_{z \sim p_z} [\log(1 - D(G(z, y))), y]
\]

GAN II is trained with corresponding pairs of real fundi and segmentations masks in order to find a mapping between the two classes of images. Given a segmentation mask, the model will translate the given geometry to a photorealistic medical image.

**VIII. U-NET**

To evaluate the reliability of our synthetic data, we used it to train a u-net that creates a segmentation mask given a photorealistic medical image. The u-net architecture, specifically formulated for biomedical image segmentation [15], is derived from an autoencoder architecture that relies on unsupervised learning for dimensionality reduction. The u-net is especially useful for biomedical applications since it does not contain fully connected layers, imposing no restriction on input image size and allowing a significantly higher number of feature channels than a regular CNN. Receptive fields after convolution are also concatenated with receptive fields in the decoding process. This allows the network to use original features along with ones after the up-convolution.

**IX. EVALUATION METRICS**

Our pipeline produced synthetic segmentation masks along with corresponding fundi images. We used this data to train a u-net segmentation network. We evaluated the u-net on test images from the DRIVE database and compared them with the ground truth to calculate an F1 score. We also calculated the variance between the
synthetic and real datasets through a Kullback-Leibler (KL) divergence score.

When considering GANs, we must analyze the adversarial divergence to calculate the statistical correlation between the generated and original data. The KL divergence score has been the standard to measure this for generative models, calculated by:

$$KL(P, Q) = \sum_i P_i \ln \frac{P_i}{Q_i}$$

We also used the universal F1-score, calculated by taking the harmonic mean of precision and recall. This score is a simple way of displaying the precision and accuracy of our model.

X. QUANTITATIVE RESULTS

In Table I, accuracy scores for segmentation and corresponding fundi images are displayed from DRIVE and our synthesized dataset. We received an F1 accuracy rating of 0.8877 for synthetic data and an F1 accuracy of 0.8988 on the DRIVE dataset. When testing variance, we received a KL-divergence score of 4.752.

| Type | Synthetic | Real |
|------|-----------|------|
| F1   | 0.8877    | 0.8988 |
| KL Divergence | 4.759 | 4.212 x 10^{-4} |

XI. QUALITATIVE RESULTS

Figure 3. Graphic displaying examples from DRIVE and our synthesized dataset.

Figure 4. Graphic displaying examples from BUBIL and a corresponding synthesized dataset.

To confirm the flexibility of our pipeline, we tested it on a second dataset. Using the BUBIL database [16] of 35 rat smooth muscle cell images and segmentations as the training data for our pipeline, we were able to produce a synthetic version of the data.

We chose this database due to its complex structure and variation. Each cellular image contains a large amount of noise, making it difficult for the GAN to learn which features are relevant. However, through our dual GAN pipeline, we were able to successfully produce realistic smooth muscle cell images as well as corresponding segmentations.

As described by our pipeline, we first generated segmentation masks of the smooth muscle cells from rats using Stage-I GAN. We then transferred the segmentations to Stage-II GAN where they were translated into photorealistic smooth muscle cells.

It is important to note this was done on an extremely small dataset of 35 images to test the limits of our pipeline. The results show that it was able to learn the correspondence between the segmentation mask and the photorealistic image, but a greater variety of data would be helpful to develop the natural background found in the original images.

XIII. DISCUSSION

Due to the extreme variation of medical imaging data (various illuminations, noise, patterns, etc.), a single GAN is unable to produce a convincing image (see Figure 5). The GAN is unable to determine complex structures, as seen with the poorly defined vessel tree structure and dark spots. However, it is able to identify simple features such as general color, shape, and lighting.

This lack of detail is unacceptable for medical image generation, as the body has many intricacies that must be accurately captured for the images to be usable. Our
dual GAN architecture improves the quality of our synthetic images by breaking down the challenging task of generating medical images. This process allows the unstable nature of GANs to be controlled by providing each GAN with a relatively elementary task. Stage-I GAN focuses only on a much lower dimensional problem: generating unique segmentation geometries, while ignoring photorealism. This allows Stage-II GAN to only generate the colors, lighting, and textures of the medical image from the given geometry. Because the geometry is generated in a lower dimensional image by a separate GAN, an unrealistic vessel geometry causes a larger loss compared to a single GAN that produces unrealistic vessel geometries in its high dimensional fundus images. This system allows both GANs to perform at a high level and reach convergence faster, creating images with more realistic geometries and textures than an ordinary single GAN system [18].

In addition, the nature of our pipeline produces a wider variety of images than the original dataset. This is because our pipeline generates images that are between the original data and the data that formed the distribution. As measured by our KL score, our synthetic dataset keeps the statistical variation of the original dataset while producing de-identified images. Therefore, the synthetic data generated by our pipeline can be effectively used for data-driven machine learning tasks while avoiding legal concerns regarding patient privacy.

To our knowledge, our pipeline is the first to synthesize synthetic medical images and segmentation pairs for the training of a segmentation network.

XIV. CONCLUSION

We have proposed a pipeline that is able to generate medical images for a segmentation task end-to-end, using a pair of generative adversarial networks. Our method decomposes the image generation process into two parts: Stage-I GAN which focuses on creating variable geometries of the segmentation mask and Stage-II GAN which transforms the geometry into a photorealistic image. Given a dataset of real images, it can produce large amounts of synthetic data that is not an image of any real patient, meaning that data produced by our pipeline can be distributed in the public domain. This is a significant step towards the creation of a public and synthetic medical image dataset, analogous to ImageNet. To further this purpose, we have created an online synthetic medical imaging database known as SynthMed\(^1\). We plan to populate this database with synthetic data from private research.

We hope that future researchers will apply similar synthetic data techniques to provide public access to their private data for the further advancement and development of computer aided medical diagnosis.

\(^1\) synthmed.github.io

XV. FUTURE WORK

We believe that our pipeline of dual generative adversarial networks can also be applied to fields outside of medical imaging. Specifically, scene generation has been a challenging topic in Computer Vision, due to the complexity and variance of the images. Our two-stage pipeline may be used to simplify the problem where simple features of the scene can be generated using Stage-I GAN and details can be learned through Stage-II. Researchers have shown in the past that GANs are able to translate manually done photo segmentations to realistic scenes, as seen with Isola et al. in facade generation [13].

Using our pipeline for different datasets may require the tuning of hyperparameters for increased effectiveness, as is the case with most neural networks. In addition, the development of deeper and more advanced architectures could be implemented to replace certain networks in our pipeline.

Our pipeline relies on a set of accurate data with high variance. For our pipeline to be executed on a variety of medical images, we must have access to private research data. Access to these private collections of images to generate synthetic data is the key to opening up public collaboration for more advanced automated medical image interpretation.
XVI. REFERENCES

1. Baris Kayalibay, Grady Jensen, and Patrick van der Smagt. Cnn-based segmentation of medical imaging data. CoRR, abs/1701.03056, 2017.

2. Ishida, T, and S Katsuragawa. “[Overview of computer-Aided diagnosis].” Nihon Igaku Hoshasen Gakkai zasshi. Nippon acta radiologica., U.S. National Library of Medicine, July 2002, www.ncbi.nlm.nih.gov/pubmed/12187835

3. Secretary, HHS Office of the, and Office for Civil Rights (OCR). “Your Rights Under HIPAA.” HHS.gov, US Department of Health and Human Services, 1 Feb. 2017.

4. Christopher Cunniff, Janice L.B. Bryne Louanne M. Hudgins, John B. Moeschler, Ann Haskins Olney, Richard M. Pauli, Lauri H. Seaver, Cathy A. Stevens, Christopher Figone. Informed consent for medical photographs, Dysmorphology Subcommittee of the Clinical Practice Committee, American College of Medical Genetics.

5. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Fei-Fei Li.
Imagenet large scale visual recognition challenge. CoRR, abs/1409.0575, 2014.

6. Aditi Ramachandran, Lisa Singh, Edward Porter, Frank Nagle. Exploring Re-identification Risks in Public Domains, Georgetown University, Harvard University.

7. Jarmin, R. and Louis, T. (2014). [ebook] Washington: U.S. Census Bureau, Center for Economic Studies.

8. Satkatar K. Kinney, Jerome P. Reiter, Arnold P. Reznek, Javier Miranda, Ron S. Jarmin, and John M. Abowd. Towards Unrestricted Public Use Business Microdata: The Synthetic Longitudinal Business Database. International Statistical Review, 79(3):362–384, December 2011.

9. J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 23(4):501–509, 2004.

10. Decenci`ere et al.. Feedback on a publicly distributed database: the Messidor database. Image Analysis and Stereology, v. 33, n. 3 p. 231-234, aug. 2014. ISSN 1854-5165.

11. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. ArXiv e-prints, June 2014.

12. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR, abs/1511.06434, 2015.

13. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. ArXiv e-prints, November 2016

14. P. Costa, A. Galdran, M. Ines Meyer, M. D. Abramoff, M. Niemeijer, A. M. Mendonca, and A. Campilho. Towards Adversarial Retinal Image Synthesis. ArXiv e-prints, January 2017

15. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. CoRR, abs/1505.04597, 2015.

16. D. Gurari, D. Theriault, M. Sameki, B. Isenberg, T. A. Pham, A. Purwada, P. Solski, M. Walker, C. Zhang, J. Y. Wong, and M. Betke. "How to Collect Segmentations for Biomedical Images? A Benchmark Evaluating the Performance of Experts, Crowdsourced Non-Experts, and Algorithms." Winter conference on Applications in Computer Vision (WACV), 8 pp, 2015. [In Press].