

Generation of Data for Training Retinal Image Segmentation Models

Srinjoy Bhuiya¹[0009–0002–2583–9294],
Suchandra Chakraborty¹[0009–0001–9319–0127],
Subhopriyo Sadhukhan¹[0009–0004–7006–2167],
Deba Prasad Mandal²[0009–0007–7768–6679], and
Dinabandhu Bhandari³[0009–0001–4412–9848]

- ¹ Heritage Institute of Technology, Kolkata WB 700107, IND
{srinjoy.bhuiya, suchandra.chakraborty,
subhopriyo.sadhukhan}.cse23@heritageit.edu.in
² Indian Statistical Institute, Kolkata WB 700108, IND
dpmandal@isical.ac.in
³ Heritage Institute of Technology, Kolkata WB 700107, IND
dinabandhu.bhandari@heritageit.edu

Abstract. Biomedical image segmentation requires pixel-wise labelling which is extremely time-consuming and the availability of additional training data is highly beneficial for training Deep Learning models. In addition to using Classical Image augmentation, generative adversarial networks have been used to augment the training data. This work is an investigation of the usefulness of generated medical images for Deep Learning segmentation models. An attempt has been made to create a computer-generated retinal image segmentation dataset using various state-of-the-art image generative deep learning models and use that dataset to train a supervised image segmentation model. Our experiments demonstrate that the generated data can be used successfully for medical segmentation tasks and improves the model’s performance over using classical augmentations.

Keywords: Retinal Image Segmentation, Image Generation, Image Augmentation, Image Translation.

1 Introduction

Biomedical imaging is essential for practitioners in the field of medical study. However, clinically verified labelling of biomedical image data is extremely time-consuming with many legal barriers for their use. The scarcity of high-quality data causes overfitting and poor generalizability in many learning models. Data augmentation has been used to generate additional data points that can alleviate these problems to a certain degree. Classically, only simple affine and non-affine image transformations have been used in this data augmentation process. Recently, Generative Adversarial Networks (GANs) have emerged as a promising

solution for generating additional medical images, where data scarcity and imbalanced data categories are common challenges. Wasserstein GANs (WGAN) [1] are proposed to produce synthetic multi-sequence brain magnetic resonance images (MRI) to improve the accuracy and consistency of medical diagnoses. For high-resolution images, progressively grown GANs [2, 3] are utilized to generate synthetic medical image data for retinal fundus photographs and multi-modal MRI of glioma. Conditional GANs [4] are exploited for denoising, super-resolution, modality conversion, and reconstruction in MRI and for computing tomography images. CycleGan and Pix2pix [5] are utilized in detecting ischemic strokes. In MRI imaging, Swin-transformer-based GANs [6] for multi-modal medical image translation are used to perform the missing and corrupted modalities from the available existing ones using image-to-image translation architectures.

The objective of this investigation is to demonstrate the effectiveness of the generated retinal fundus images in training a segmentation model. Initially, we have incorporated the Perceptual path length (PPL) regularization [22] to get smooth and continuous changes in the latent space while generating images. Subsequently, an image-to-image GAN [7, 8] is applied to generate the corresponding vessel masks. A fully convolutional neural network based image segmentation model is developed to examine the usefulness of our generated data. In this model, a custom loss function [9, 10] is considered instead of the popularly used binary cross entropy. The proposed image generation framework diverges from traditional image augmentation methods and focuses on the benefits of using images generated using generative adversarial networks for deep learning-based model training. Through experiments, the results show that training with generated data leads to notable improvements compared to conventional augmentation methods.

Section 2 describes our proposed image generation model and the corresponding image segmentation model. The experimental results with the generated retina image dataset are presented in Section 3. Section 4 finds the conclusions and future scope of this investigation.

2 Proposed GAN based Image Generation Model

In this work, we are concerned with generating retinal image data and verifying its effectiveness. Here, three different tasks are performed: firstly, the retinal images are generated using a GAN; then the respective masks (pixel-wise labels) are generated corresponding to each generated fundus image using a conditional GAN; and finally, the utility of using generated datasets in retinal vasculature segmentation is verified using a CNN model. Moreover, some additional experiments are performed using a mixture of real, generated and classically-augmented datasets to verify the usefulness of using additional generative data in such tasks. The generated dataset aims to capture and mimic the distribution of disease markers present in the input dataset used for training. The focus is on replicating the overall distribution of the entire dataset, rather than generating images based on specific visual characteristics or disease markers. This approach

ensures that the generated data aligns with the distribution of disease markers in the original dataset.

2.1 Retinal Image Generation

The image generation model described in [11] is utilized here for retinal image generation. The weight modulation and demodulation in the generator structure helped eliminate any artefacts by normalizing them with estimated statistics. A hierarchical generator with skip connections, similar to MSG-GANs, has been exploited to reduce the phase artefacts. Additionally, Adaptive Discriminator Augmentation (ADA) regularization has been used to improve the training stability and sample quality by dynamically adjusting the complexity of the discriminator during training. Path length regularization [11] was employed in the generator loss to encourage smooth and continuous changes in the latent space, increasing the training stability of the generator. The real and generated images are processed using a residual connection network as a discriminator. An R1 regularization penalty was used in the nonsaturating discriminator loss to stabilize the training of the discriminator, thereby improving the quality of the generated images.

2.2 Vasculature Retinal Masks Generation

The GAN framework introduced in [7] by Phillip Isola et al. is exploited for an image-to-mask translation system. This model includes a Conditional GAN [8] that creates images based on an input retinal image, in this instance the retinal image, and the output is a vascular structure segmentation mask. The generating network is a U-style autoencoder connected by skip connections, similar to the architecture provided in [7]. The encoder extracts the features of the input retinal image into a latent feature vector and the decoder uses the latent feature vector to generate the translated image. Finally, the output from the decoder network is passed through a *Tanh* activation function. A PatchGAN discriminator [7] is being used to downsample the input using convolutions and strided-convolutions to produce a single-channel prediction map. 70x70 image patches are used as the discriminator input. The model uses the Leaky ReLU as the activation function, batch normalization for regularization and the Wasserstein loss function to maintain the Lipschitz continuity of the critic function.

2.3 Image Segmentation Model

To compare the performance of the generated retinal image segmentation dataset with that of a real retinal dataset. Here a lightweight self-attention-based medical image segmentation model showing state-of-the-art performance retinal vasculature segmentation available in [23] is chosen. As shown in Figure 1, the model uses a simple U-style autoencoder CNN and incorporates the concept of a convolutional self-attention mechanism [24] as the bottleneck. In the proposed model,

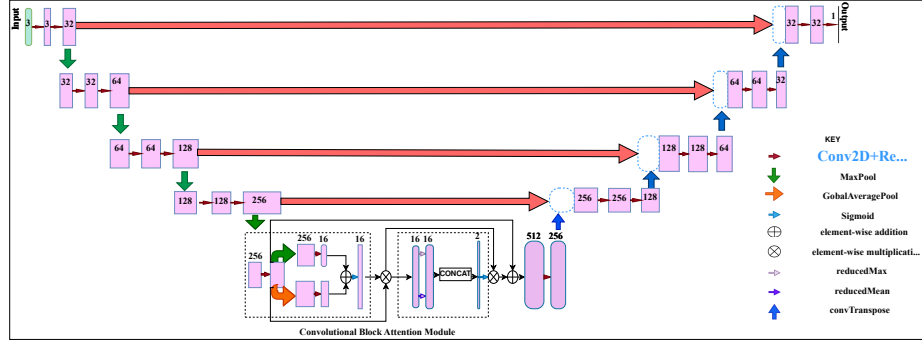


Fig. 1. Architecture of the segmentation model used

a down-sampling block to the decoder part of the autoencoder is added resulting a feature vector width of 256 channels. The final segmentation model used in this work is fully convolutional and has approximately 7 million learnable parameters. The model is then trained on two different datasets. The first was the clinician-labelled data of 4000 training image-mask pairs, the other was a generated dataset of the same size. Additional image-mask pairs are generated for use as the validation dataset. The performance of both models is tested using real data. To deal with the high levels of the background to vessel pixels imbalance, a custom loss function [9, 10] based on a combination of dice and focal loss is incorporated. The aim of the proposed work is to evaluate the effectiveness of generated retinal image rather than proposing a novel architecture or method for retinal image generation. The methodology involves leveraging pre-existing image generation models and adapting them through parameter adjustments and modifications to suit retinal image generation. The fine-tuned models are subsequently trained on a carefully curated dataset.

3 Results

To generate data in training the segmentation model, two datasets were created to train two separate deep-learning models viz., the retinal image generation and the retinal mask generation models. The first dataset for retinal image generation was compiled from six publicly available benchmark datasets - DRIVE [12], CHASE_DB1 [13], RIDB [14], DIARETDB1 V2.1 [15], HRF [16], and IDRiD [17], totalling 5000 retinal images. The second dataset for retinal vessel mask generation was created by collecting image and mask pairs from DRIVE, CHASE_DB1, ARIA [18], FIVES [19] and STARE [20], totalling 1500 image pairs. The images were resized to a resolution of 512 X 512 while maintaining their aspect ratio. Centre cropping was also performed on all images.

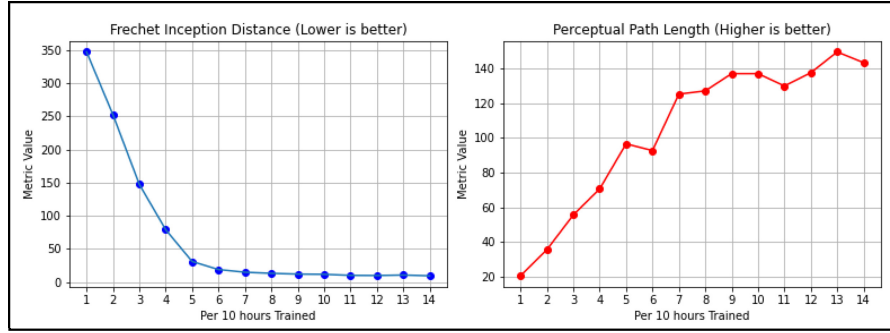


Fig. 2. FID and PPL Metric

3.1 Retinal Image Generation

The proposed image generation system has achieved satisfactory results in generating high-quality retinal fundus images with a fully generated vasculature structure. The Frechet Inception Distance (FID) [21] and Perceptual Path Length Metric [22] are used to evaluate the image generation model. After 50 epochs the model resulted in an FID score of 9.55 and a Perceptual Path Length value of 143.54. As illustrated by Figure 2, the performance of the model has increased steadily with the increase in the number of epochs trained. Figure 3 presents some examples of the retinal images generated from the model, centred on the macula with proper focus. A good visualization of the parafoveal vessels was achieved over the two disk diameters around the fovea. To further demonstrate the model's superiority, it is compared to another image generation model discussed in [22] that was unable to generate connected vasculature and contained noisy artefacts; its FID score was only 25.46, indicating significantly inferior performance than the proposed approach. The average computation time required to generate the retinal images was 0.2446 seconds per image on an Nvidia P100 Graphics Processor.

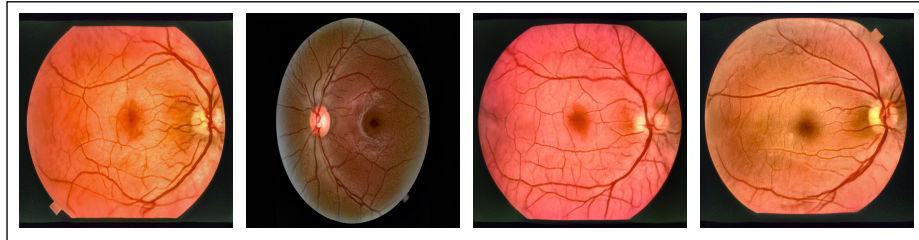


Fig. 3. Samples of the Retinal Fundus images Generated

3.2 Retinal Mask Generation

The proposed image-to-Mask generation model has also shown good results in generating retinal vasculature masks for both generative and real retinal images. Accuracy, F1-Score, Specificity, and Sensitivity metrics are used to evaluate its performance. The proposed model achieves an Accuracy of 94.456%, an F1-Score of 0.8232, a Specificity value of 0.954, and a Sensitivity value of 0.875. Comparisons with some state-of-the-art supervised binary segmentation models such as [23, 25] indicates that the model has achieved comparable performance to even the above-referenced supervised deep convolutional segmentation model. Figure 4 depicts some of the retinal masks generated, most of which show all of the major retinal vessels with only some of the finest vessels being either missing or deviating from the ground truth. The average computation time required to generate the vessel masks was 0.455 seconds per image on an Nvidia P100 Graphics Processor.

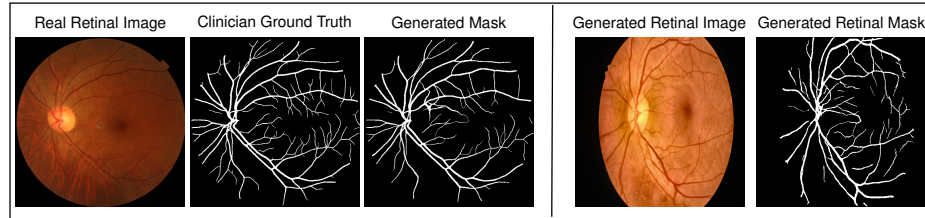


Fig. 4. A sample of the masks generated from the model

3.3 Real-data versus Generated-data

To demonstrate the benefits of generated (fake) data, the segmentation model is trained on four different datasets. The first contains real images and clinically verified vessel segmentation masks, the next has generated images and vessel masks, and the third has only the real images but was augmented with some affine and non-affine transformation techniques. Finally, a segmentation dataset containing both real images and generated images together is chosen; this dataset is used to compare the utility of using classical image augmentation techniques with the GANs-based image generation process mentioned in earlier an section.

Table 1 shows the performance of all the different models on a fixed testing dataset comprising real test images and clinically verified segmentation masks. Figure 5 depicts samples of generated masks in comparison to the ground truth masks. Metrics such as Accuracy, DICE Score, Precision, Recall, Specificity, and Sensitivity are used for assessing the performance between real data and generated data for training image segmentation models. It has been found that the generated data leads to better results than merely using the real data- this

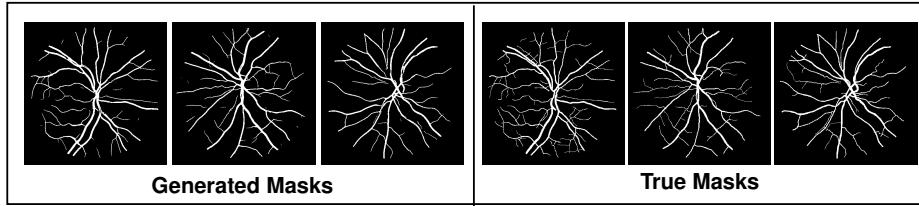


Fig. 5. Generated and True Masks

Table 1. Performance metrics of segmentation model based on different data-sets

Dataset	Accuracy	IOU	DICE	F1	Precision	Recall	Specificity	MCC
Real Data	0.978	0.728	0.842	0.842	0.864	0.822	0.990	0.832
Generated Data	0.981	0.762	0.864	0.864	0.895	0.836	0.992	0.855
Real + Gen	0.992	0.871	0.931	0.932	0.925	0.937	0.993	0.905
Real + Augmented	0.983	0.892	0.887	0.888	0.905	0.894	0.994	0.903

is primarily due to a low number of real training image-mask pairs. The model trained with generated data yield a high IoU score of 0.871, a Precision value of 0.925, and a recall value of 0.937. These metrics indicate that generated data can be used to train medical image segmentation models for real-life applications without any serious complications. The use of combined datasets of the real and generated datasets together is found to perform better than when either real and augmented data together or only generated data is used. This may suggest that using medical images created through GANs leads to better training outcomes than only applying affine and non-affine transformations on real datasets.

4 Conclusions

This paper presents a high-resolution retinal fundus image generation pipeline that generates retinal images and vasculature masks. The pipeline was rigorously tested on the medical image generation task and was able to reproduce high-quality medical images that can enhance the training of deep-learning models in medical imaging. The use of additional generated training samples was able to significantly improve the performance of the segmentation model. It is also shown that augmenting medical datasets with generative data results in better performance than using simple transformations for data augmentation. In future work, the generative model architecture used to create a medical-specific image generator will be modified, compare and contrast the use of various types of regularization techniques in the loss function and explore the use of diffusion models for medical image generation.

References

1. Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Furukawa, Y., Mauri, G., Nakayama, H.: Gan-based Synthetic Brain Mr Image Generation. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). (2018). <https://doi.org/10.1109/ISBI.2018.8363678>
2. Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D.A., Hernández, M.V., Wardlaw, J., Rueckert, D.: Gan augmentation: Augmenting training data using generative Adversarial Networks, <https://arxiv.org/abs/1810.10863>. (2018) <https://doi.org/10.48550/arXiv.1810.10863>
3. Beers, A., Brown, J., Chang, K., Campbell, J.P., Ostmo, S., Chiang, M.F., Kalpathy-Cramer, J.: High-resolution medical image synthesis using progressively grown generative adversarial networks, <https://arxiv.org/abs/1805.03144>. (2018) <https://doi.org/10.48550/arXiv.1805.03144>
4. Kaji, S., Kida, S.: Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in Medical Imaging - Radiological Physics and Technology, <https://link.springer.com/article/10.1007/s12194-019-00520-y>. (2019) <https://doi.org/10.1007/s12194-019-00520-y>
5. Platscher, M., Zopes, J., Federau, C.: Image translation for medical image generation: Ischemic stroke lesion segmentation. Biomedical Signal Processing and Control. 72, 103283 (2022). <https://doi.org/10.1016/j.bspc.2021.103283>
6. Yan, S., Wang, C., Chen, W., Lyu, J.: Swin transformer-based gan for multi-modal medical image translation. Frontiers in Oncology. 12, (2022). <https://doi.org/10.3389/fonc.2022.942511>
7. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks, <https://arxiv.org/abs/1611.07004>. (2018). <https://doi.org/10.48550/arXiv.1611.07004>
8. Mirza, M., Osindero, S.: Conditional generative adversarial nets, <https://arxiv.org/abs/1411.1784>. (2014). <https://doi.org/10.48550/arXiv.1411.1784>
9. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 42, 318–327 (2020). <https://doi.org/10.1109/TPAMI.2018.2858826>
10. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. 240–248 (2017). https://doi.org/10.1007/978-3-319-67558-9_28
11. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data, <https://arxiv.org/abs/2006.06676>. (2020). <https://doi.org/10.48550/arXiv.2006.06676>
12. Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging. 23, 501–509 (2004). <https://doi.org/10.1109/tmi.2004.825627>
13. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Transactions on Biomedical Engineering. 59, 2538–2548 (2012). <https://doi.org/10.1109/tbme.2012.2205687>

14. Akram, M.U., Abdul Salam, A., Khawaja, S.G., Naqvi, S.G., Khan, S.A.: RIDB: A dataset of fundus images for retina based person identification. *Data in Brief*. 33, 106433 (2020). <https://doi.org/10.1016/j.dib.2020.106433>
15. Kauppi, T., Kalesnykiene, V., Kamarainen, J.-K., Lensu, L., Sorri, I., Raninen, A., Voutilainen, R., Uusitalo, H., Kalviainen, H., Pietila, J.: The diaretdb1 diabetic retinopathy database and evaluation protocol. *Proceedings of the British Machine Vision Conference 2007*. (2007). <https://doi.org/10.5244/c.21.15>
16. Budai, A., Bock, R., Maier, A., Hornegger, J., Michelson, G.: Robust vessel segmentation in fundus images. *International Journal of Biomedical Imaging*. 2013, 1–11 (2013). <https://doi.org/10.1155/2013/154860>
17. Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudhe, V., Meriaudeau, F.: Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research. *Data*. 3, 25 (2018). <https://doi.org/10.3390/data3030025>
18. Farnell, D.J.J., Hatfield, F.N., Knox, P., Reakes, M., Spencer, S., Parry, D., Harding, S.P.: Enhancement of blood vessels in digital fundus photographs via the application of Multiscale Line Operators, <https://research.manchester.ac.uk/en/publications/enhancement-of-blood-vessels-in-digital-fundus-photographs-via-th>.
19. Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., Zhang, Q., Wang, Y., Ye, J.: Fives: A fundus image dataset for artificial intelligence based vessel segmentation. *Scientific Data*. 9, (2022). <https://doi.org/10.1038/s41597-022-01564-3>
20. Hoover, A.D., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*. 19, 203–210 (2000). <https://doi.org/10.1109/42.845178>
21. Heusel, Martin, et al. “Gans Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.” *ArXiv.org*, 12 Jan. 2018, <https://arxiv.org/abs/1706.08500> (2019). <https://doi.org/10.48550/arXiv.1706.08500>
22. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (2019).
23. Bhuiya, S., Choudhury, S.R., Aich, G., Maurya, M., Sen, A.: Retinal blood vessel segmentation and analysis using lightweight spatial attention based CNN and data augmentation. 2022 *IEEE Calcutta Conference (CALCON)*. (2022). <https://doi.org/10.1109/CALCON56258.2022.10060189>
24. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: Convolutional Block Attention Module. *Computer Vision – ECCV 2018*. 3–19 (2018). https://doi.org/10.1007/978-3-030-01234-2_1
25. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on U-Net (R2U-net) for medical image segmentation. <https://doi.org/10.48550/arXiv.1802.06955>