

RESEARCH

Open Access



Latent space autoencoder generative adversarial model for retinal image synthesis and vessel segmentation

K. Radha¹ and Yepuganti Karuna^{2*}

Abstract

Diabetes is a widespread condition that can lead to serious vision problems over time. Timely identification and treatment of diabetic retinopathy (DR) depend on accurately segmenting retinal vessels, which can be achieved through the invasive technique of fundus imaging. This methodology facilitates the systematic monitoring and assessment of the progression of DR. In recent years, deep learning has made significant steps in various fields, including medical image processing. Numerous algorithms have been developed for segmenting retinal vessels in fundus images, demonstrating excellent performance. However, it is widely recognized that large datasets are essential for training deep learning models to ensure they can generalize well. A major challenge in retinal vessel segmentation is the lack of ground truth samples to train these models. To overcome this, we aim to generate synthetic data. This work draws inspiration from recent advancements in generative adversarial networks (GANs). Our goal is to generate multiple realistic retinal fundus images based on tubular structured annotations while simultaneously creating binary masks from the retinal fundus images. We have integrated a latent space auto-encoder to maintain the vessel morphology when generating RGB fundus images and mask images. This approach can synthesize diverse images from a single tubular structured annotation and generate various tubular structures from a single fundus image. To test our method, we utilized three primary datasets, DRIVE, STARE, and CHASE_DB, to generate synthetic data. We then trained and tested a simple UNet model for segmentation using this synthetic data and compared its performance against the standard dataset. The results indicated that the synthetic data offered excellent segmentation performance, a crucial aspect in medical image analysis, where smaller datasets are often common. This demonstrates the potential of synthetic data as a valuable resource for training segmentation and classification models for disease diagnosis. Overall, we used the DRIVE, STARE, and CHASE_DB datasets to synthesize and evaluate the proposed image-to-image translation approach and its segmentation effectiveness.

Keywords Vessel segmentation, Diabetic retinopathy (DR), Early diagnosis, Generative adversarial models, Data generation

*Correspondence:

Yepuganti Karuna
karuna.y@vitap.ac.in

¹School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

²School of Electronics Engineering, VIT-AP University, Amaravati, India



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

The advancements in medical imaging technologies have facilitated the prompt and accurate detection of diseases. Furthermore, significant steps have been made in the realm of computer vision technologies, which have enabled the automatic processing of medical images without the need for human intervention. These developments have greatly enhanced the efficiency and precision of medical diagnosis and treatment and hold great promise for the future of healthcare [1].

Long-term diabetes can lead to vision loss, which is becoming increasingly prevalent due to the growing number of people with diabetes worldwide. This rise in the diabetic population has also resulted in a rise in eye diseases. To combat this issue, the World Health Organization began a global initiative in 2020 aimed at preventing visual impairment [2]. According to global statistics, the primary causes of blindness are cataracts, age-related macular degeneration, and DR. Fundus imaging is a non-invasive imaging technique for early diagnosis and monitoring of DR [3]. It allows physicians to inspect the geometrical [4] and physiological properties of the vessel structures, providing valuable clinical information for early diagnosis of DR [5]. However, it is time-consuming to perform manually since we have a large diabetes population. To overcome this, many semi-automated [6, 7] and automated retinal vessel segmentation techniques [8, 9] have been proposed to help ophthalmologists. However, these techniques need a huge number of annotated images, particularly for automated vessel segmentation techniques. Acquiring annotated images for these retinal fundus vessel images remains challenging. Typically, only a limited number of annotated images are available where domain experts precisely define the vessel structures through a protracted manual process. This lack of annotated images poses a significant challenge to the development of automated retinal vessel segmentation models. The progress in artificial intelligence has led to a surge in the utilization of synthetic data generation, particularly due to the emergence of GAN. This trend is indicative of the impactful role played by advanced AI techniques in facilitating synthetic data creation [10].

With the growing need for huge annotated datasets in the realm of medical image analysis, the generation of synthetic data has emerged as a viable solution to the challenges posed by manual annotation. Recent developments in GANs and Variational Autoencoders (VAEs) have facilitated the creation of high-quality synthetic images that significantly enhance the training of deep-learning models [11]. Research indicates that GAN-based architectures, including Pix2Pix, CycleGAN, and StyleGAN, are particularly effective in generating annotated data that captures the inherent variability of real-world patient populations. For instance, CycleGAN has proven

to be instrumental in producing cross-domain images, while StyleGAN allows for precise manipulation of specific attributes, which is advantageous for generating variations in the shapes, sizes, and intensities of vascular structures. Despite the progress made in the field, substantial challenges persist in synthesizing anatomically precise and clinically relevant medical images. A key difficulty lies in ensuring that synthetic images accurately represent realistic anatomical variations, which is crucial for tasks such as detecting retinal vessels. Furthermore, generating high-resolution images that exhibit intricate vessel structures closely resembling those in authentic fundus images remains a significant challenge. The limited diversity present in current synthetic datasets can result in overfitting, thereby reducing the model's ability to generalize when evaluated with novel images. In addition, the scarcity of specialized evaluation metrics aimed at medical image synthesis poses difficulty in quantitative assessment of the generated images' practicality and clinical applicability.

To address these challenges, our proposed GAN model with an integrated latent space autoencoder focuses on shape preservation, which is crucial for accurate vessel segmentation. By enhancing the diversity and anatomical correctness of generated data, we aim to provide a more robust solution for annotated data augmentation in retinal vessel segmentation. Even though many research activities were performed on retinal vessel segmentation [12, 13], only a few efforts have been taken to synthesize the data [14]. So, we proposed a GAN model with an integrated Latent space autoencoder to generate synthetic data. This could help achieve adequate training data with different intensities and vessel structures to create a fundus retinal vessel segmentation model using deep learning. This paper introduces the following contributions:

1. Our objective is to generate synthetic retinal images and their corresponding segmentation masks utilizing GAN. This approach seeks to address the challenge of limited data availability in developing deep learning models for retinal analysis.
2. This paper introduces a Latent space Autoencoder responsible for generating multiple images from a single image and ensuring the accurate preservation of anatomical shapes in generated masks, which is critical for precise DR diagnosis and analysis.
3. We conducted retinal vessel segmentation utilizing real and synthetic images through a simple UNet architecture. The results demonstrated competitive performance, indicating the potential application of GANs to synthesize the dataset for improved deep-learning segmentation.

Related works

Retinal image synthesis

Sagar et al. initiated the development of a retinal image synthesis application with the objective of establishing a lifelike virtual environment beneficial to eye surgery. This innovation involved using advanced visual and mechanical simulations to improve the accuracy and efficiency of surgical procedures. The retinal image synthesis application has been widely adopted in the medical field and continues to be a valuable tool for ophthalmologists [15]. Retinal image analysis faces a significant challenge due to the absence of annotated vessel structures. The task of annotating these structures is expensive and laborious. Furthermore, the subjectivity and annotation errors restricting from inter-observer and intra-observer variabilities of human observers further create the issue [16]. Fluorescein angiography (FA) is an essential medical procedure utilized to capture high-resolution images of the intricate blood vessels in the retina. It is, however, noteworthy that the dye necessary for this procedure is associated with potentially harmful side effects. At present, the only non-invasive alternative is Optical Coherence Tomography Angiography (OCTA), even though it is both costly and provides a limited field of view. In contrast, retinal fundus photography, a secure imaging modality, comprehensively captures the structural intricacies of the retina.

To visualize the retinal vasculature accurately, in a non-invasive and cost-effective manner, Tavakkoli et al. [17] proposed deep learning conditional GAN to produce FA-like images from fundus photographs. Schlegl et al. [18] employed GAN for the investigation of data distribution within the healthy retinal tissue region using patches. The GAN model was subsequently assessed for its efficacy in detecting anomalies in retinal images within regions of unseen and healthy image. The study's results show the potential of GANs in detecting anomalies in medical images with high precision and accuracy. This approach can potentially be used in clinical settings to detect anomalies in medical images early, thereby improving patient outcomes.

Beers et al. [19] have conducted a study aiming to develop a novel method for synthesizing medical images of fundus pictures that depict premature retinopathic vascular pathology. The researchers employed a progressively grown generative adversarial network (PGGAN). The PGGAN is a deep learning model that has been shown to produce high-quality synthetic images that are virtually indistinguishable from real ones. The network was trained on a large dataset of fundus images labeled with retinopathy grades. The generated images were evaluated by clinicians, who confirmed their accuracy and usefulness for educational and diagnostic purposes. This

method has the potential to facilitate early diagnosis and improve patient outcomes in the field of ophthalmology.

In the field of ophthalmology, vessel segmentation is a vital step for disease finding and screening. However, the intricate morphological vessel structure can be challenging when designing segmentation models. Researchers have been using GAN for segmentation. The GAN has two parts - the generator and the discriminator. The generator uses a U-shaped structure with a skip connection for gradient diffusion, while the discriminator adopts deep convolution blocks to evaluate the generator's segmentation response effectively [20]. The authors reported that an effective loss function could be helpful for the further development of the model. Researchers also have experimented with GAN models for vessel segmentation. One such technique is an attention-augmented Wasserstein GAN. The generator is designed with a combination of attention-augmented convolution and squeeze and excitation modules incorporated into the basic UNet. A complex vessel structure makes the segmentation of tiny vessels complex. To overcome that, an attention-augmented convolution and squeeze-excitation module are utilized to handle pixel dependency and channel feature maps. The authors reported that the complexity of the model can be reduced [21].

Recent studies have found that the structure of blood vessels in the retina can be difficult to analyze in low-contrast regions due to their complex and indefinable nature. Researchers have proposed a symmetric equilibrium generative adversarial network to address this issue. This network utilizes a generator to produce realistic images that include local details, which can aid in extracting the morphological parameters of the vessels [22]. A GAN variant featuring a single generator and three discriminators is utilized. The generator is a U-shaped network that executes the segmentation process. Three discriminators with varying receptive fields are integrated to assess the generator's performance. This effectively directs the generator to generate more intricate segmentation outcomes [23].

A conditional generative adversarial network has been developed. It utilizes a fully stacked convolutional neural network to balance the loss. The generator has two fully convolutional networks and a multi-kernel block that helps to balance the scale invariance. The discriminator uses deep convolutional networks to evaluate the generator's response by comparing the generated image to the corresponding ground truth [24]. Kar et al. [25] proposed a GAN for fundus retinal vessel segmentation. Multi-scale feature extraction is facilitated by incorporating the inception module. A discriminator with multiple self-attention modules stacked with a point-wise fully connected feed-forward network acts as a binary classifier to discriminate vessel and non-vessel pixels.

Liang et al. [26] presented an end-to-end conditional GAN incorporating class feature loss and enhanced retinal detail loss to tackle the scarcity of annotated retinal images in academic research. The network generates high-quality retinal images and evaluates them using objective metrics: subjective effect, FID, and SWD. Kim et al. [27] aims to use deep learning to create realistic retinal images for computer-aided diagnosis systems. The efficacy of these synthetic images in augmenting imbalanced datasets with specific diseases is being evaluated through various tests and analyses, including deep learning-based classification performance.

Saeed et al. [28] aims to create realistic fundus images with morphologically changed vessel structures using a technique called Sharpening and Varying Vessels (SVV). SVV sharpens and varies vessel structures, reducing the need for expensive annotated medical data. The approach also preserves critical fundus characteristics using recent advancements in GANs and Variational Autoencoder. Ahn et al. [29] proposed a high-performance model, to generate fundus disease images for DR and Age-Related Macular Degeneration (AMD) using a semi-supervised learning approach using GAN. The model addresses the lack of labeled data and data imbalance issues by utilizing a guidance mask and a disease-feature matching loss to accurately generate initial symptoms for the two diseases.

Researchers have been exploring various GAN models for image-to-image translation. However, only very little effort has been taken to synthesize retinal images to expand the training data and accurately segment blood vessels. The synthesis of retinal images has the potential to significantly enhance the segmentation of retinal vessels using deep learning techniques. Automatic fundus retinal image analysis has been developed to address inconsistencies in human observers' assessments and to support ophthalmologists in managing the increasing population of individuals with diabetes. Deep learning techniques have become increasingly popular for segmenting and classifying medical images for disease diagnosis. It is essential to note that vessel segmentation plays a pivotal role in the early diagnosis of DR. Several vessel segmentation algorithms based on deep learning have been proposed. Since deep learning algorithms are data-hungry, a large dataset is necessary to develop a generalized prediction model or segmentation algorithm. Obtaining ground truth from ophthalmologists for this specific problem is a critical and long process due to the large amount of data, ethical considerations, and patient consent. We used a GAN model to synthesize data and perform vessel segmentation. The proposed methodology is explained in further sections. Section 3 presents the proposed methodology for synthetic data generation; Sect. 4 discusses the results of the proposed GAN model;

and Sects. 5 and 6 describe the discussion and conclusion on the proposed model.

Methodology

Conditional Generative Adversarial Networks (cGANs) are one variant of the generative adversarial network model. In cGANs, the generation of an image depends not only on random noise but also on some specific input image, which allows for the generation of targeted and controlled image outputs. cGANs comprise two main neural networks: the generator and the discriminator. These networks participate in a competitive association, constantly improving and evolving their capabilities. The generator is responsible for creating synthetic data, while the discriminator's role is to differentiate between real and generated data. The generator strives to produce data visually indistinguishable from real data, while the discriminator aims to become increasingly effective at detecting any differences. By effectively training the generator and discriminator in this adversarial manner, cGANs can produce outputs that closely resemble real data, yielding more realistic and reliable results.

The overall concept of the proposed architecture is shown in Fig. 1, which consists of a generator and a discriminator. The generator follows a U-shaped network design, incorporating two downsampling encoders and one upsampling decoder. The downsampling encoders consist of a latent space autoencoder aimed at preserving vessel shape and a second encoder featuring a sequence of convolution and downsampling operations. On the other hand, the decoder network entails a series of upsampling operations. The discriminator's training data is sourced from two different channels. The first source comprises real data instances, such as actual fundus images, which the discriminator uses as positive examples during training. The second source consists of data instances created by the generator, which the discriminator uses as negative examples during training. The two input images depict these data sources feeding into the discriminator. It is important to note that the generator does not undergo training during discriminator training. Instead, its weights remain constant while it produces examples for the discriminator to train on. Detailed working operation of a proposed model is described in the following sections.

Generator

The proposed generator utilizes a UNet architecture, which is well-known for its ability to perform effectively with minimal data. This generator comprises two parallel encoders: a latent space autoencoder, a downsampling encoder, and a decoder with an upsampling block and skip connections.

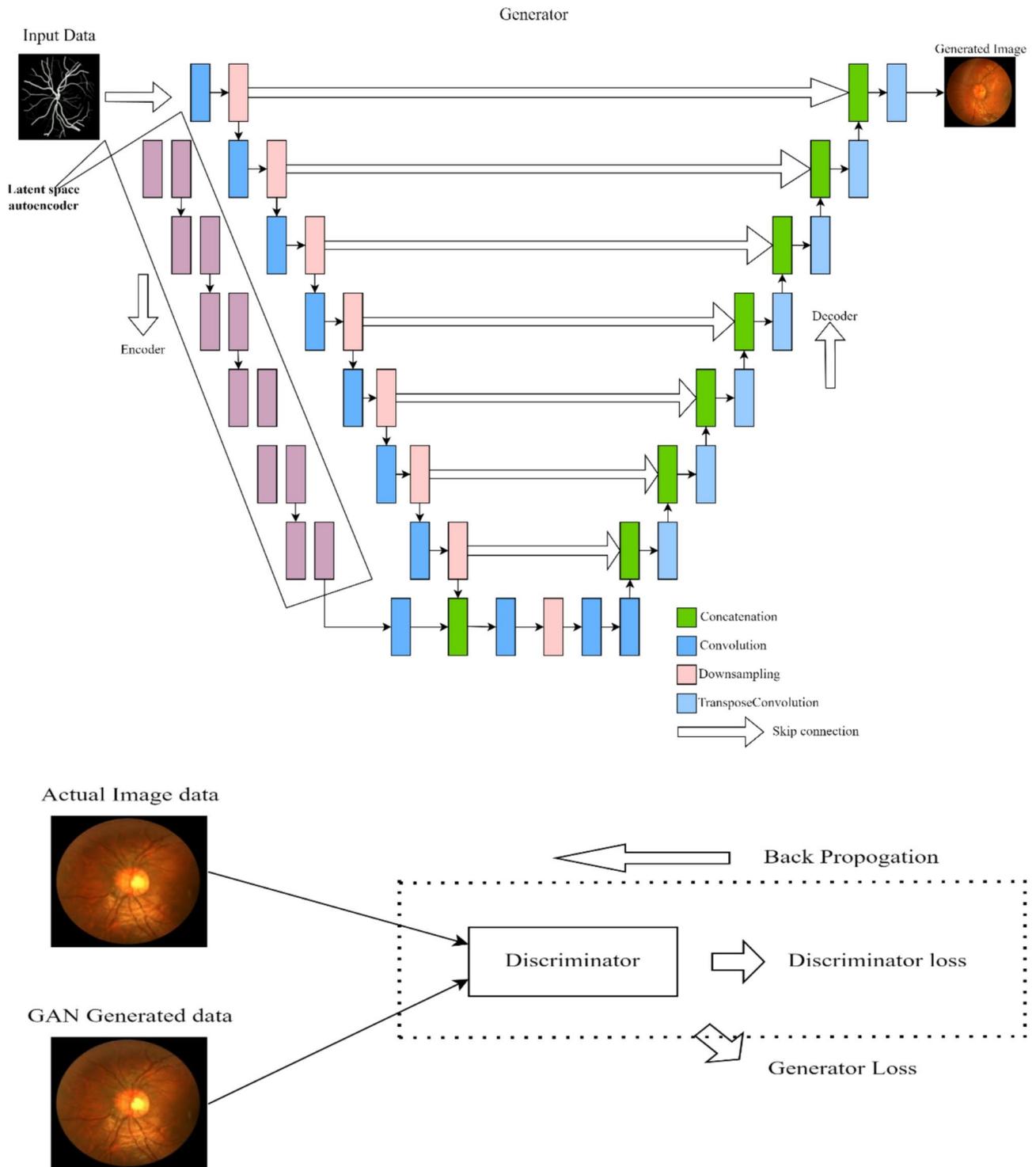


Fig. 1 Overview of the proposed architecture

Latent space autoencoder for vessel mask generation

Ideally, an image-to-image translation algorithm should generate a realistic vessel network from a fundus retinal image. The model should be able to absorb original data and produce a wide range of vessel networks as needed while maintaining anatomical precision. In this work, we

propose achieving this through the use of a Latent Space auto-encoder [30].

The latent space auto-encoder is a powerful tool that shrinks the original input image (I) into a lower-dimensional representation (Z) in a smooth and abstract manner. This reduced representation captures a wide range

of input image features, such as shape, texture, and other vital characteristics. In our work, we have connected the latent space auto-encoder alongside the UNet encoder and decoder to produce vessel masks from the input image. By compressing the image into latent space, the auto-encoder aims to retain the essential aspects of the fundus data while filtering out the noise, which is crucial for generating accurate masks.

The auto-encoder outputs a mean μ and the logarithm of the variance $\log(\sigma^2)$ of the latent variables mentioned in Eqs. (1),

$$\mu, \log(\sigma^2) = E(I) \tag{1}$$

The latent variable Z is sampled from a normal distribution parameterized by μ and σ , sampling the latent variable mentioned in Eqs. (2),

$$Z = \mu + \sigma \odot \epsilon \tag{2}$$

Where $\epsilon \sim \mathcal{N}(0, I)$, here the ϵ represents noise when sampled from a normal distribution. The Latent auto-encoder loss calculation is given as follows: first, the reconstruction loss ensures the reconstructed image is close to the original image as given in Eq. (3), \hat{I} is the reconstructed image.

$$\downarrow_{rec} = E_I \|I - \hat{I}\|_1 \tag{3}$$

The KL divergence is used to regularize the latent space to follow a standard normal distribution given in Eqs. (4),

$$\downarrow_{KL} = E_I [D_{KL}(q(Z|I)|p(z))] \tag{4}$$

Where $q(Z|I)$ is the approximate posterior, and $P(Z)$ is the prior (standard normal distribution). Total latent space autoencoder loss is given in Eq. (5), where β is the weighting factor.

$$\downarrow_{KL} = \downarrow_{rec} + \beta \downarrow_{kl} \tag{5}$$

Downsampling encoder

In our implementation, the UNet generator model is designed to transform input images into target images by gradually decreasing the spatial dimensions and increasing the number of feature channels through a series of downsampling operations shown in Fig. 2. This process allows for the extraction of high-level features and representations from the input image. The downsampling path is composed of eight sequential blocks, each exactly carrying out convolution operations followed by batch normalization and activation functions. This deliberate design is aimed at ensuring that the model can proficiently capture a rich hierarchy of features at various levels of granularity. The fundamental components of each block encompass the following key aspects: **Convolution Layer:** Every block employs a kernel size of 4×4 and strides of 2×2 , effectively reducing the spatial dimensions of the feature maps with each step while extracting features. **Batch Normalization:** This step stabilizes and accelerates training by normalizing the inputs of each layer, ensuring that the outputs have zero mean and unit variance. Notably, batch normalization is purposefully omitted in the first downsampling block to enable the network to learn an initial set of features without normalization constraints. **Leaky Rectified Linear Unit (LeakyReLU) Activation Function,** this non-linearity is strategically applied to facilitate the network in learning more intricate representations. LeakyReLU’s ability to allow a small gradient when the unit is not active assists in mitigating the issue of dying neurons, thereby promoting improved learning outcomes mentioned in Eqs. (6),

$$D_i = LeakyReLU (BatchNorm (Conv2D (D_{i-1}))) \tag{6}$$

The downsampling process involves reducing the spatial dimensions of the input image from 256×256 to 1×1 while concurrently increasing the number of feature

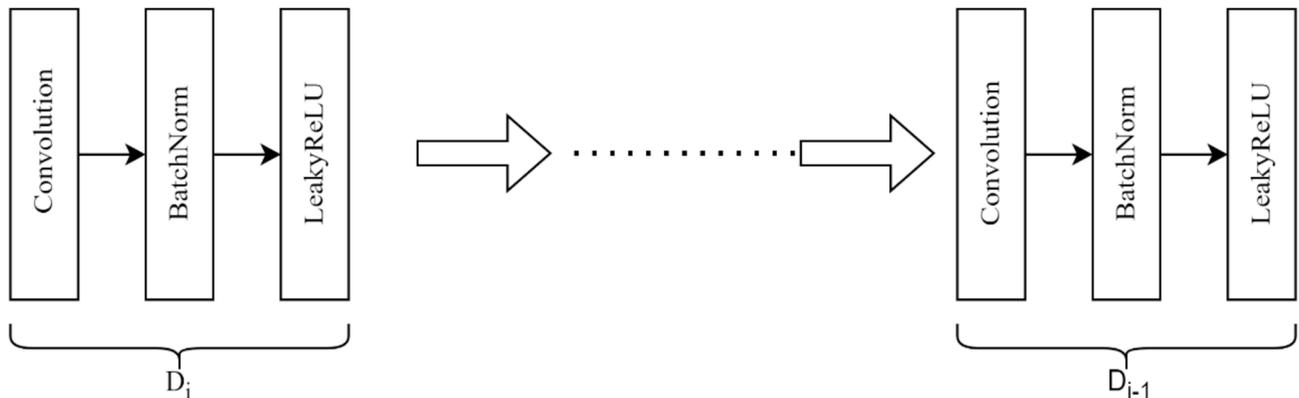


Fig. 2 Downsampling encoder

channels from 3 (RGB) to 512. In Eq. (6) D_i represents the number of down-sampling layers. This transformation enables the neural network to abstract high-level features by aggregating spatial information across progressively larger image regions. The outputs of each block are maintained as skip connections, which are later utilized during the upsampling phase to incorporate fine-grained spatial detail. These skip connections allow the network to merge high-level features with detailed spatial information, thus enhancing the overall accuracy and quality of the generated images.

Upsampling decoder

Upsampling serves as an important technique to convert low-resolution feature maps into higher-resolution ones within a neural network. This transformation enables the model to capture and reconstruct intricate details present in the original image. Specifically, in the UNet architecture, upsampling is strategically used to progressively elevate the resolution of the feature maps until they reach the size of the original input. The model can generate output images with precise and fine-grained details through this method. In the proposed architecture, the upsampling block includes transposed convolution, batch normalization, ReLU activation, and optional dropout mentioned in Eq. (7),

$$U_i = \text{ReLU}(\text{BatchNorm}(\text{Conv2DTranspose}(U_{i-1}))) \quad (7)$$

Incorporating optional dropout [31] in the upsampling blocks of the generator functions as an efficient regularization technique to avoid overfitting during the training process. Dropout is an effective method of regularization used to protect neural networks from overfitting. During training, dropout randomly deactivates a proportion of the input units in each iteration, preventing the network from relying excessively on any individual neuron. This

approach encourages the network to acquire more intricate features, improving new data generalization. In the proposed model, dropout is selectively applied within the upsampling blocks of the generator, which helps refine lower-dimensional feature maps to produce the final output image.

Discriminator

The discriminator classifies the real and generated synthetic images by concatenating them and passing through a series of downsampling blocks and convolutional layers, shown in Fig. 3. The discriminator plays a critical role in classifying image pairs as either real or fake. By combining the generated image with the target image, the discriminator gains valuable context about the relationship between the input image and the generated output. This understanding is particularly important for tasks like image translation. The discriminator's ability to detect differences goes beyond the appearance of the generated image alone, as it also considers how well it aligns with the target image. This joint evaluation significantly enhances the discriminator's capacity to guide the generator in creating more contextually accurate images. Loss calculation helps the generator to produce more realistic images.

Generator loss

The generator aims to fool the discriminator, which is measured using the mean Square Error (MSE) and Mean Absolute error (MAE) mentioned in Eqs. (8), (9), and (10).

$$\mathcal{L}_{GAN} = \mathbb{E}_{X,Y} [(D(X, G(X)) - 1)^2] \quad (8)$$

$$\mathcal{L}_{L1} = \mathbb{E}_{X,Y} [Y - G(X)_1] \quad (9)$$

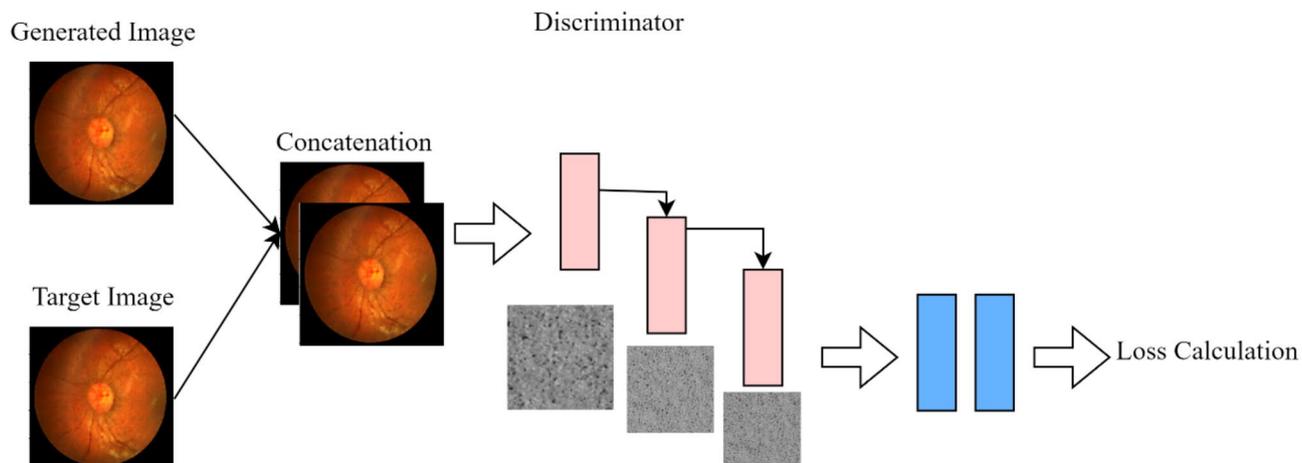


Fig. 3 Proposed discriminator

$$\mathcal{L}_{GEN} = \mathcal{L}_{GAN} + \lambda \mathcal{L}_{L1} \quad (10)$$

Here, G is the generator, D is the discriminator, X is the input image, and Y is the target image. λ is a hyperparameter that balances GAN loss and L1 loss.

Discriminator loss

The discriminator loss is a combination of the MSE and MAE for real and generated images mentioned in Eqs. (11), (12), and (13).

$$\mathcal{L}_{disc-real} = \mathbb{E}_{X,Y} \left[(D(X, Y) - 1)^2 \right] \quad (11)$$

$$\mathcal{L}_{disc-gen} = \mathbb{E}_{X,Y} \left[(D(X, G(X)))^2 \right] \quad (12)$$

$$\mathcal{L}_{DISC} = \mathcal{L}_{disc-real} + \mathcal{L}_{disc-gen} + MAE \text{ loss terms} \quad (13)$$

Different optimization functions are used in the discriminator and generator, RMS prop for convolution layers, and Adam for remaining layers. Where $\mathcal{L}_{disc-real}$ (Eq. 11) and $\mathcal{L}_{disc-gen}$ (Eq. 12) are based on MSE loss, and the MAE loss term is intended to capture the absolute deviation between the discriminator's output for real and generated images. $MAE = E(X, Y) [|D(X, Y) - D(X, G(X))|]$. $E(X, Y)$ represents the expected value over the joint distribution of input image X and the target image Y. $D(X, Y)$ represents the discriminator function D takes an input image X and its real target image Y and outputs a probability indicating how real the image pair is. $G(X)$ is the generator function G takes an input image X and generates a synthetic output $G(X)$, which is intended to be as close as possible to the real target image Y. The model is trained by altering between optimizing the generator and discriminator. During each training step, gradients are clipped to $-1.0 < gradient < 1.0$ to stabilize the training.

During training, the generator and discriminator gradient descent are adjusted using Eqs. (14) and (15).

$$\theta_D \leftarrow \theta_D - \eta \Delta_{\theta_D} \mathcal{L}_{DISC} \quad (14)$$

$$\theta_G \leftarrow \theta_G - \eta \Delta_{\theta_G} \mathcal{L}_{GEN} \quad (15)$$

Here θ_D and θ_G are the parameters of the discriminator and generator, respectively, and η is the learning rate.

Results of the proposed GAN model

Dataset details and experimental setup

We have utilized the DRIVE [32], STARE [33], and CHASE_DB [34] datasets, each containing 20, 10, and 14 training images. Employing the proposed GAN model, we generated 200, 100, and 140 images from these datasets, respectively. As these images are intended for vessel

segmentation, we used Mask data as input to ensure accurate extraction of vessel information and subsequently generated RGB images. The Mask data was created using the RGB image as input, a process that will be essential in assessing the performance of the GAN model. We generated 440 synthetic RGB images and corresponding masks derived from these three standard datasets. The same GAN model depicted in Fig. 1 was employed to generate both Mask and RGB images. The generator produces a synthetic RGB image for the mask input, as illustrated in Fig. 4. Similarly, synthetic mask data is generated from the RGB image shown in Fig. 5. The model yields pairs of images with a resolution of 256×256 . Each actual image was downsampled to the synthetic data size for a meaningful comparison.

To balance the GAN Loss and L1 loss, we used a hyperparameter λ . In the generator loss equation (Eq. 10), λ is utilized to control the weight of the L1 loss in relation to the GAN loss. This enables fine-tuning of the balance between generating realistic images (represented by the GAN loss) and ensuring similarity to the target images (represented by the L1 loss). Our model's λ value is set to 100, effectively scaling the L1 loss alongside the GAN loss. By prioritizing the minimization of the L1 loss, the generator is encouraged to produce images that closely match the target images in terms of pixel-wise similarity. The learning rate is established at 0.0002 for both the generator and the discriminator, facilitating an optimal balance between training stability and convergence speed. This configuration is particularly effective when employed with the Adam optimizer, which enhances overall model performance. The weighting factor is set to $\lambda = 100$, balancing the contributions of the L1 loss and the GAN loss. Emphasizing the L1 loss encourages the generator to produce images that closely align with the target in pixel space. This enhances the generation process's stability and promotes improved structural fidelity in the resulting images.

The peak signal-to-noise ratio (PSNR) and structural Similarity index (SSIM) are calculated between the original input image and the Generated Synthetic data mentioned in Table 1 to evaluate the proposed GAN model's performance. The PSNR is a critical metric for evaluating the fidelity of a reconstructed image in relation to its original counterpart. It quantifies the disparity between the original and generated images and is typically expressed in decibels (dB). High image similarity is typically represented by PSNR values of 30 dB and above [35]. The SSIM is a metric used to assess the similarity between two images. In contrast to traditional metrics like MSE and PSNR, SSIM considers changes in structural information, luminance, and contrast, offering a more comprehensive evaluation of image quality. A range of $0.9 < SSIM < 1$ signifies a very high degree of similarity,

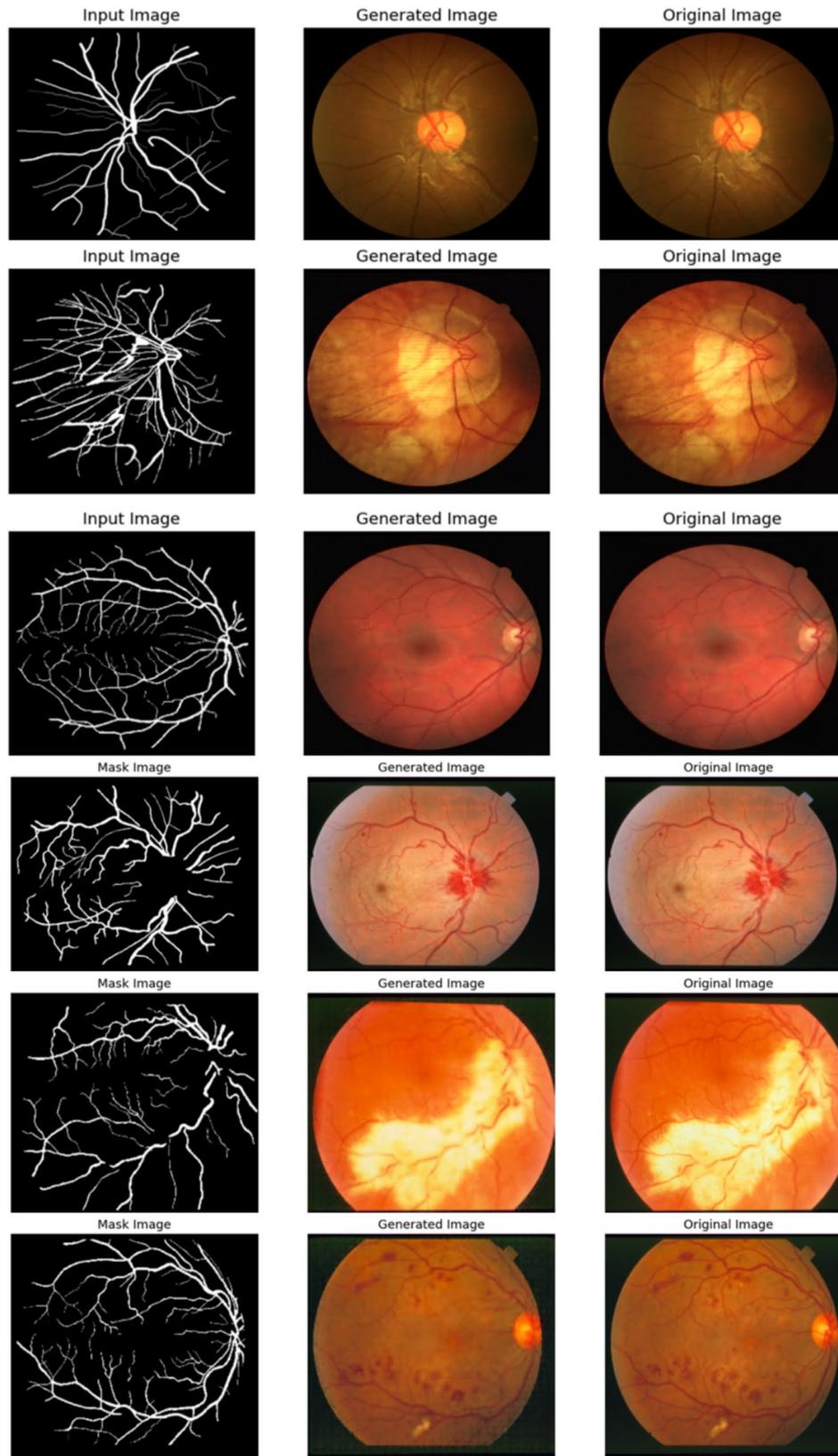


Fig. 4 Generated synthetic data for Mask input (Columns 1, 2, and 3 represent the input to the network, the generated synthetic data, and the ground truth data used to evaluate the generated image, respectively)

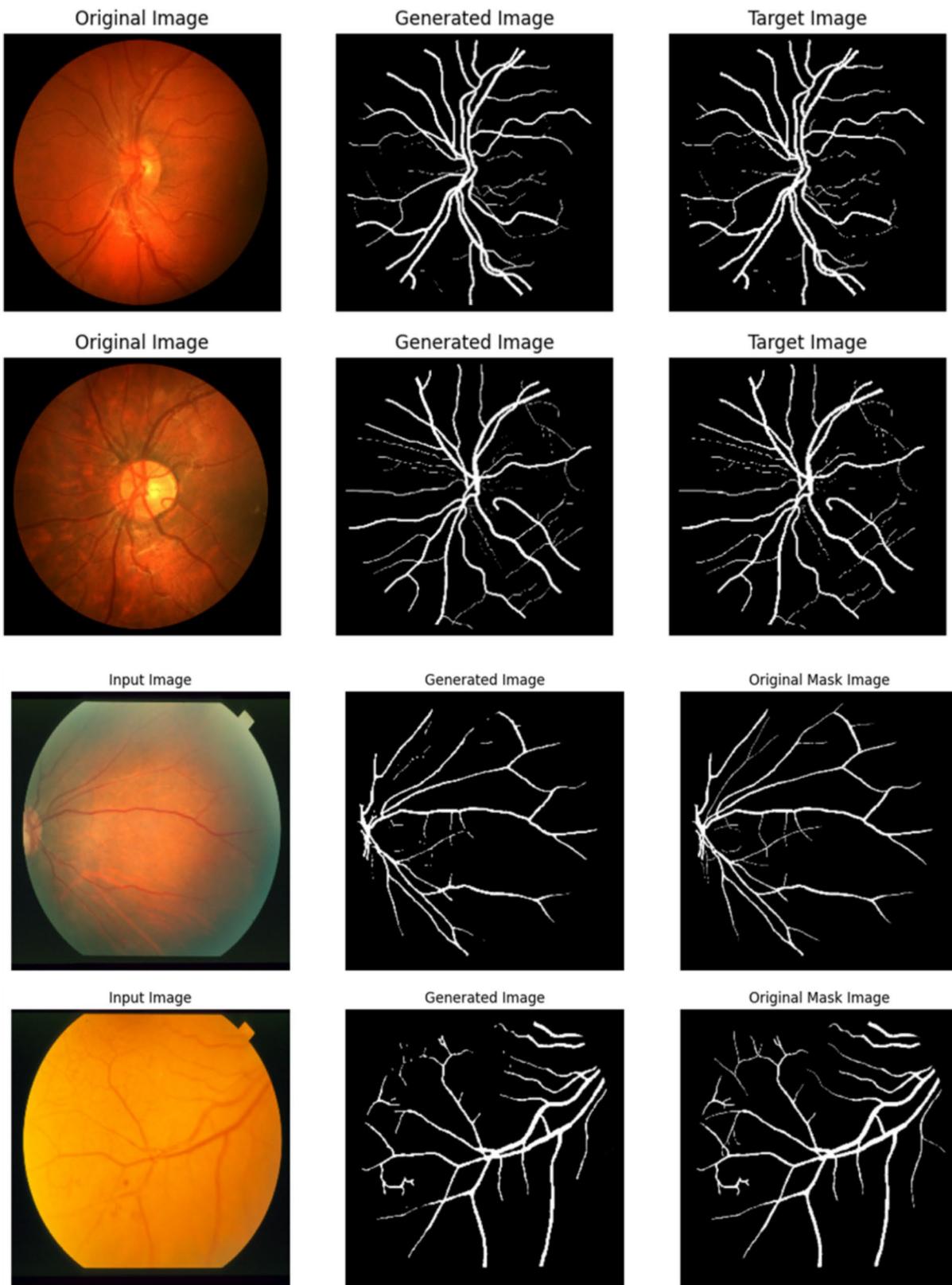


Fig. 5 Generated Synthetic Mask data for RGB input (Columns 1, 2, and 3 represent the input, the synthetic data generated, and the ground truth data used for evaluating the generated image, respectively.)

Table 1 Performance analysis of the proposed GAN model

Dataset	SSIM	PSNR (dB)
DRIVE Dataset		
Image 1	0.9826	33.54
Image 2	0.8542	32.34
Image 3	0.9755	32.35
Image 4	0.9822	33.67
Image 5	0.9900	32.88
Image 6	0.9767	33.51
Image 7	0.8955	32.91
Image 8	0.9855	32.87
Image 9	0.8654	32.98
Image 10	0.9823	33.12
Image 11	0.9703	34.98
Image12	0.8688	33.16
Image 13	0.8527	33.67
Image 14	0.8955	32.89
Image 15	0.9932	32.97
Image 16	0.9862	33.65
Image 17	0.8878	31.97
Image 18	0.9998	31.45
Image 19	0.9829	32.67
Image 20	0.8907	31.52
STARE Dataset		
Image 1	0.9684	33.55
Image 2	0.9692	30.98
Image 3	0.9899	32.88
Image 4	0.9732	31.99
Image 5	0.9741	32.16
Image 6	0.9907	32.78
Image 7	0.9890	33.67
Image 8	0.9910	34.54
Image 9	0.9967	32.97
Image 10	0.9901	32.25
CHASE_DB1 Dataset		
Image 1	0.9599	32.11
Image 2	0.9867	32.29
Image 3	0.9874	37.36
Image 4	0.9901	31.56
Image 5	0.9654	34.46
Image 6	0.9891	32.23
Image 7	0.9865	38.17
Image 8	0.9962	32.24
Image 9	0.9879	32.02
Image 10	0.9655	33.51
Image 11	0.9891	32.05
Image 12	0.9761	32.20
Image 13	0.9845	32.71
Image 14	0.9899	31.05

with the images appearing almost indistinguishable to the human eye [35]. We achieved strong performance metrics for the generated synthetic data from the proposed GAN model and selected only 440 images based on the PSNR and SSIM values from these generated images.

Using artificial data for retinal vessel segmentation

The primary reason for this work is the necessity of a large training dataset for utilizing deep learning techniques in automatic segmentation. The paper introduces a method that generates synthetic images and their respective vessel trees. The importance of this task is to evaluate how well the generated images can be used to separate vessel trees from eye fundus images and assess how well the synthetically generated data performs in this task; we employed the recommended generator as a segmentation network, excluding the use of a latent space autoencoder shown in Fig. 6. The segmentation model initially underwent training using real data, followed by training using only synthetic data for comparison.

UNet is a specialized, fully convolutional neural network architecture tailored for the task of biomedical image segmentation, particularly for identifying retinal vessels. Its design features an encoder-decoder framework paired with skip connections, enabling the model to effectively capture both fine details and broader contextual information crucial for accurate segmentation. The training data consists of both real and synthetic fundus images, as detailed in Table 2. The datasets employed include DRIVE, STARE, and CHASE_DB, which are recognized benchmark collections for retinal vessel segmentation. These images come in various resolutions (565×584 , 700×605 , and 999×960). To address the limited real data, a GAN model generated 440 synthetic images, all sized at 256×256 . These synthetic images add diversity, enhancing the model's ability to generalize. Both types of images are divided into smaller patches of 64×64 pixels, which helps lower computational demands and increases the volume of training samples. This patch-based approach enables a focus on localized details, making it easier for the model to learn the structures and characteristics of the vessels effectively. Additionally, using patches accommodates variations in image sizes by maintaining a uniform input size for the model.

For consistent comparison and to account for the low resolution of the produced synthetic images, we down-sampled the original DRIVE, CHASE_DB, and STARE datasets to a size of 256×256 . We developed our segmentation model by training it on two sets of images: real and synthetic. The dataset details for the segmentation task involving these images can be found in Table 3. In both cases, we used a test set comprising 20, 10, and 14 images for testing cases.

We also calculated the Dice Similarity Coefficient (DSC), Jaccard Index (IoU), and Hausdorff Distance as quantitative metrics to further evaluate the proposed GAN model, as shown in Table 2. This represents the average performance across all images. We have compared our model with existing state-of-the-art GAN models. While only a few studies have utilized GANs for

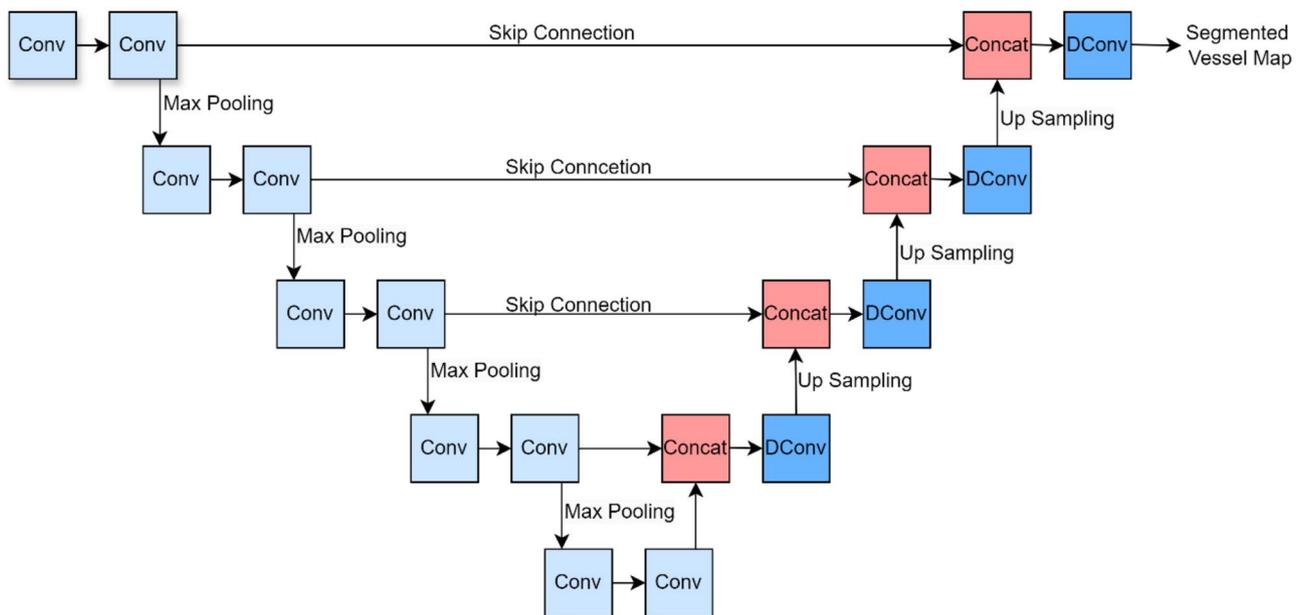


Fig. 6 Simple UNet for vessel segmentation to analyse the efficiency of synthesized retinal images

Table 2 Shape similarity metrics for the proposed GAN model with existing models

	Dice similarity coefficient (DSC)	Jaccard index (IOU)	Hausdorff distance	F1 score	Accuracy	Sensitivity	Specificity
CGAN [26]	0.8112	-	-	0.8106	-	-	-
Style-Based_GAN [27]	-	0.989	-	-	0.970	0.957	0.982
Fundus GAN [29]	-	-	-	-	0.8605	0.8654	0.8689
Proposed Model	0.9853	0.9900	1.20	0.9299	0.9804	0.9745	0.9867

Table 3 Dataset details for training the segmentation model

Dataset	Number of available images	Preprocessed data with a size of 64 × 64 for training	Testing accuracy		
			Accuracy	AUC	IOU
DRIVE, STARE, and CHASE_DB	20, 10, and 14 images with sizes of 565 × 584, 700 × 605, and 999 × 960 respectively.	5936 image patches for training	0.9647	0.9857	0.7023
Synthetic data generated from the GAN model	440 Images with a size of 256 × 256	7040 image patches for training	0.9956	0.9899	0.9825

synthesizing retinal images, many have focused on segmentation. Therefore, our comparisons are limited to the existing methodologies in this specific context.

The DSC quantifies how much the generated image overlaps with the target image (input data). It is especially responsive to small variations in overlap, which makes it an effective metric for measuring the quality of the generated image. In comparison, the IoU directly penalizes false positives and false negatives more rigorously, offering a stricter assessment of overlap accuracy. Meanwhile, the Hausdorff Distance differs from both DSC and IoU by concentrating on the spatial alignment of boundaries, showing how closely the edges of the generated image correspond to those of the target image.

To assess the effectiveness of the synthetic data, we analyzed accuracy, AUC, and IoU for both real and synthetic data in segmentation tasks. Accuracy measures how well

the model identifies pixels as either vessel or background; however, in retinal vessel segmentation, where there are typically fewer vessel pixels than background pixels, relying solely on accuracy may not provide a complete picture of the model's performance. To address this, we calculated the AUC, which represents the area under the Receiver Operating Characteristic (ROC) curve, illustrating the relationship between the true positive rate (sensitivity) and the false positive rate (1-specificity). A higher AUC score indicates better differentiation between vessel and non-vessel pixels. A high AUC in segmentation indicates the model's capability to effectively identify vessel structures amidst the background across various thresholds. As shown in Table 3, the segmentation model utilizing synthetic data achieved a notable AUC of 0.9899, highlighting the advantages of synthetic data. IoU, or Intersection over Union, quantifies the overlap between

Table 4 Performance analysis comparison of the simple UNet with existing state-of-the-art methods when trained on Synthetic data

Dataset	Methods	Accuracy	Specificity	Sensitivity	IOU	AUC
DRIVE	SUD-GAN [20]	0.9560	0.9820	0.8340	-	0.9786
	AA-WGAN [21]	0.9651	0.9903	0.7923	-	0.9788
	Refined Equilibrium GAN [22]	0.9563	0.9812	0.8294	-	0.9830
	M-GAN [24]	0.9706	0.9836	-	0.7129	0.9868
Real Images	Ours (UNet)	0.9508	0.8915	0.9144	0.7984	0.8814
Synthetic Data	Ours (UNet)	0.9915	0.9950	0.9890	0.9891	0.9897
STARE	SUD-GAN [20]	0.9663	0.9897	0.8334	-	0.9734
	AA-WGAN [21]	0.9719	0.9923	0.8234	-	0.9873
	Refined Equilibrium GAN [22]	0.9671	0.9781	0.8812	-	0.9863
	M-GAN [24]	0.9876	0.9938	-	0.7198	0.9873
Real Images	Ours (UNet)	0.9513	0.8814	0.9159	0.7990	0.8820
Synthetic Data	Ours (UNet)	0.9900	0.9895	0.9897	0.9893	0.9898
CHASE_DB1	AA-WGAN [21]	0.9644	0.9914	0.8335	-	0.9781
	M-GAN [24]	0.9736	-	-	-	0.9859
Real Images	Ours (UNet)	0.9451	0.8993	0.9017	0.7877	0.8859
Synthetic Data	Ours (UNet)	0.9917	0.9883	0.9891	0.9890	0.9889

Table 5 SSIM for Figs. 4 and 5

Image	SSIM
Figure 4	(SSIM between Generated image and original image in Fig. 4)
Row 1	0.9602
Row 2	0.9002
Row 3	0.9054
Row 4	0.9100
Row 5	0.9856
Row 6	0.9214
Figure 5	(SSIM between Generated image and Target Image in Fig. 5)
Row 1	0.9051
Row 2	0.9756
Row 3	0.9815
Row 4	0.9204

the predicted and actual segmentation, calculated by dividing the intersection of predicted and true vessel regions by their union. IoU is particularly useful in segmentation tasks as it evaluates how closely the model's predicted vessel map aligns with the ground truth. The higher IoU score of 0.9825 for synthetic data compared to real data suggests superior accuracy and precision in vessel segmentation. In summary, while accuracy offers a general sense of correctness in pixel classification, AUC reflects the model's ability to distinguish between vessels and non-vessels, and IoU assesses the quality of the segmentation by measuring the alignment between predicted and actual vessel regions.

Table 4 Compares the segmentation performance of the UNet trained on synthetic data with existing state-of-the-art methods. Our goal is to generate synthetic data and assess the effectiveness of synthetic data in segmentation tasks. The results indicate that synthetic data significantly enhances segmentation performance across

all three datasets using a simple UNet architecture. This demonstrates the efficacy of synthetic data in the design of segmentation and classification models for medical images. It is particularly beneficial when there are challenges in collecting real-time medical images and their annotations, thereby aiding researchers in developing more advanced deep-learning algorithms.

We need to generate visually indistinguishable images obtained in Figs. 4 and 5. We have calculated the SSIM for Figs. 4 and 5 to differentiate the image quantitatively, as shown in Table 5. For the RGB image (Fig. 5), we calculated the SSIM for each channel and then averaged these values to obtain the final metric. For the mask data (Fig. 5), we directly computed the SSIM by comparing the generated image with the target image. In Fig. 4, we referred to the original image, while in Fig. 5, we referred to the Target image; both terms denote the same object, namely the original image from the dataset.

Discussion

The effectiveness of generative models depends on the amount of training data and its ability to capture real-world data variability. In this work, we utilized a Latent Space Autoencoder GAN trained on a modest dataset of 44 images from the DRIVE, STARE, and CHASE_DB datasets. Despite the limited training set, our Latent Space Autoencoder GAN successfully produced synthetic images that differed notably from those used during training. Images generated from the tubular structure and RGB images are shown in Figs. 7 and 8. After generating 250 images from each original image and selecting those with SSIM and PSNR values exceeding 0.9 and 30dB, respectively, we obtained a dataset of 440 high-quality images. Using these images, we trained a segmentation model to assess the performance of synthetic

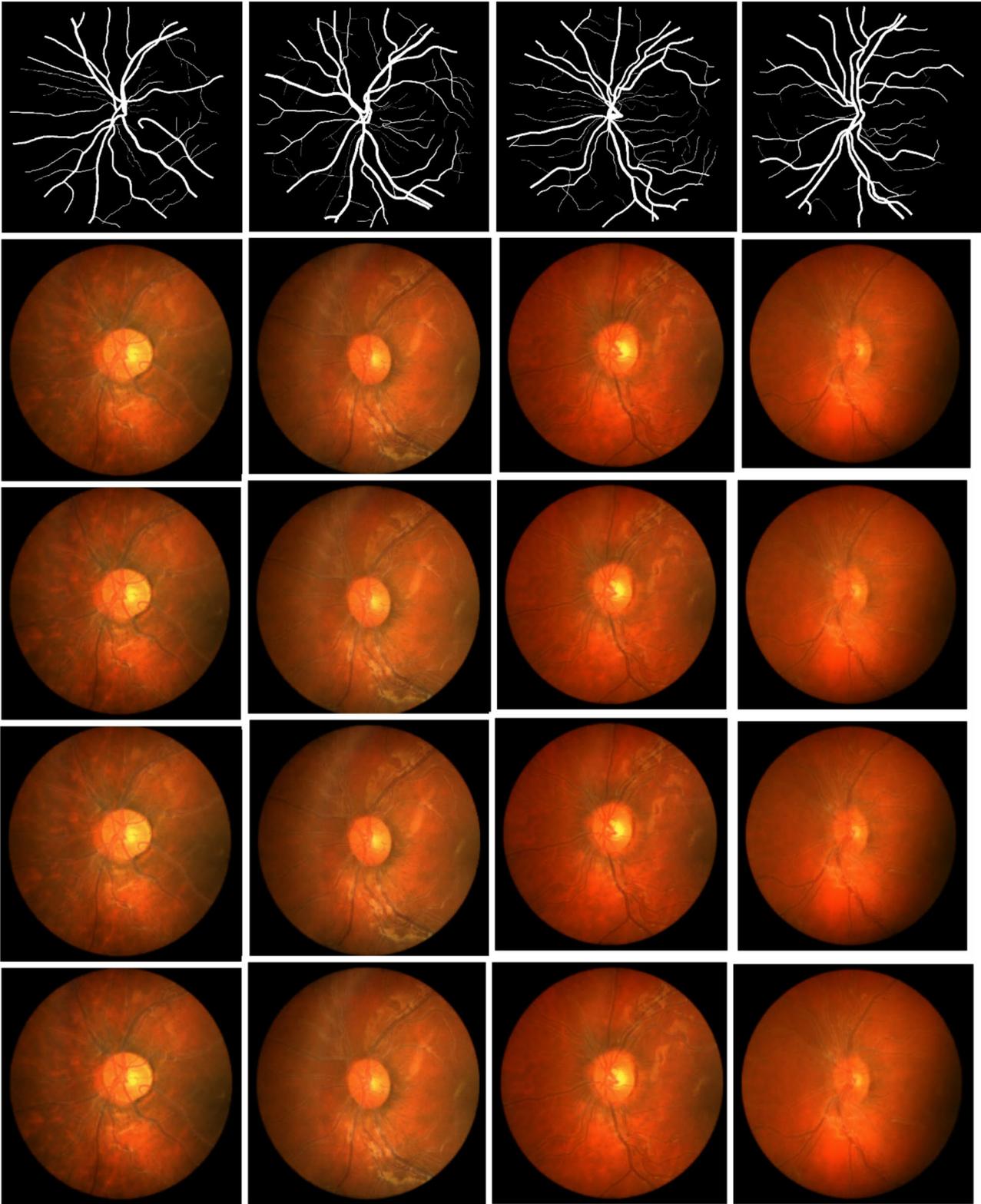


Fig. 7 RGB images generated from the mask input

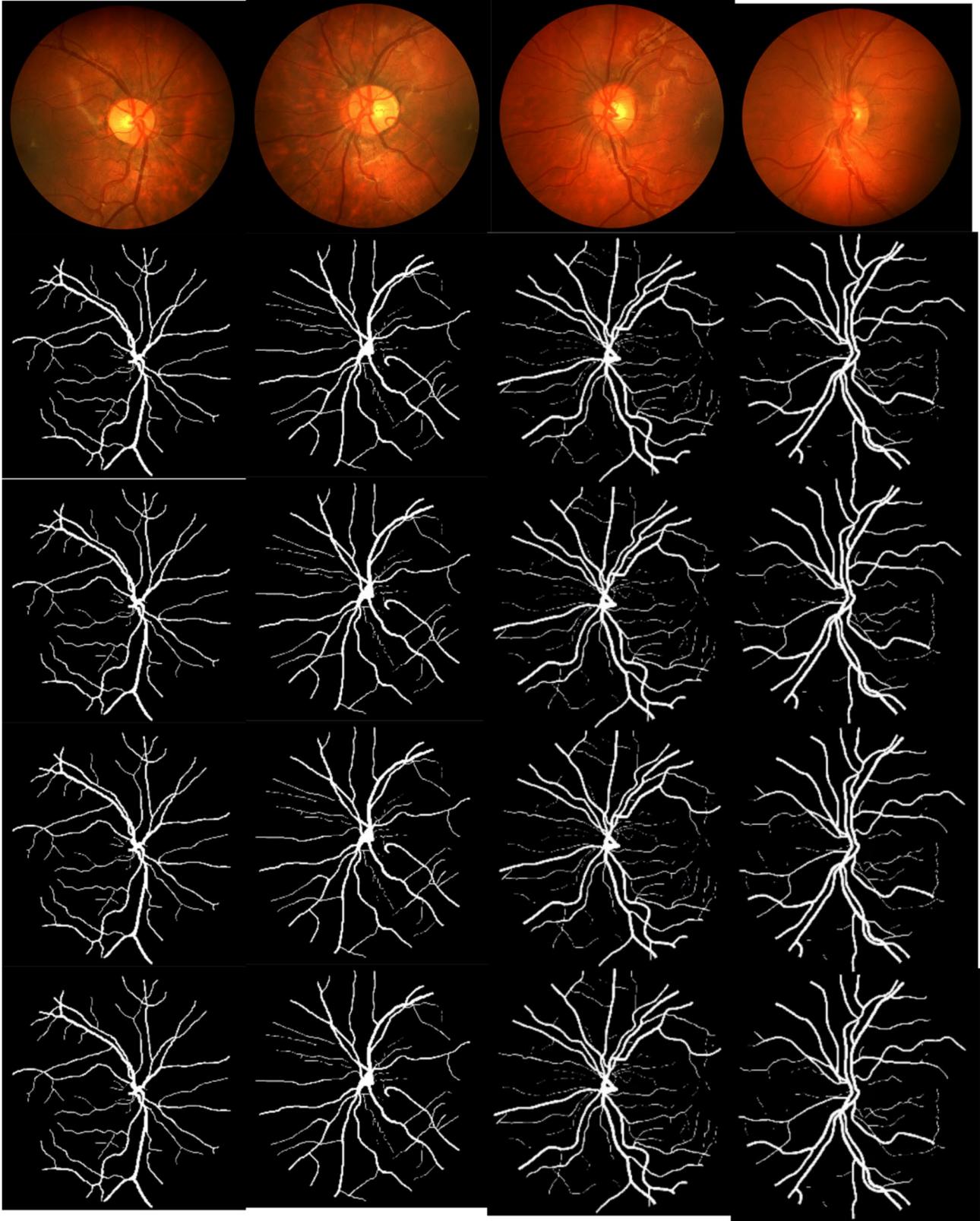


Fig. 8 Mask images generated from RGB input

data when tested with unseen images, resulting in a segmentation accuracy of 0.9956, surpassing that achieved with real images. This demonstrates the potential of GAN-generated synthetic data to address data scarcity issues. However, it is important to note the limitations of this approach, such as the smaller image size (256×256) compared to current fundus image databases and the significant memory requirements, which necessitate advanced hardware when it is implemented for real-time applications.

Conclusion

This paper presented an advanced generative model that employed a Latent space auto-encoder to create synthetic vessel mask images and their corresponding eye fundus images. This algorithm learned the intricate structure of original retinal images from vessel networks and fundus images, enabling it to generate synthetic vessel masks and retinal images with diverse visual characteristics not present in the training examples. The method's ability to produce realistic vessel geometries and retinal image textures while maintaining a consistent global structure was especially noteworthy. Furthermore, the generated synthetic data was leveraged to train the segmentation model, resulting in high accuracy, AUC, and IOU scores of 0.9956, 0.9899, and 0.9825, respectively. These results represented a significant enhancement over the existing dataset used for training the segmentation network. This demonstrates the potential of synthetic data as a valuable resource for training segmentation and classification models, thereby addressing the challenges associated with the availability of annotated data.

Acknowledgements

Not applicable.

Author contributions

Problem formulation, manuscript writing, simulation -Radha K Problem formulation, Supervision, Manuscript correction-Dr. Yepuganti Karuna. Authors have made a substantial contribution to the design of the article. Revised it critically for important logical content and approved the version to be published. Problem Formulation, supervision, results validation - Yepuganti Karuna*.

Funding

Not applicable.

Data availability

The authors have used publicly available data in this manuscript. The dataset links are DRIVE, STARE - Retinal Image Databases - medicmind, (CHASE_DB1 retinal vessel reference dataset - Kingston University Research Data Repository).

Declarations

Ethical approval

This research did not involve any real-time human subjects. All datasets used in this study are publicly available and anonymized, and therefore, ethics approval and consent to participate were not required.

Consent for publication

Not applicable. This study does not include any individual person's data in any form, and no consent for publication is required. Problem Formulation, Implementation, Manuscript writing - Radha k.

Competing interests

The authors declare no competing interests.

Received: 12 September 2024 / Accepted: 25 April 2025

Published online: 05 May 2025

References

1. Yousef R, Gupta G, Yousef N, Khari M. A holistic overview of deep learning approach in medical imaging. *Multimedia Syst.* 2022;28(3):881–914.
2. Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T, Simó R. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol.* 2020;8(4):337–47.
3. Radha K, Karuna Y. Retinal vessel segmentation to diagnose diabetic retinopathy using fundus images: A survey. *Int J Imaging Syst Technol.* 2024;34(1):e22945.
4. Krestanova A, Kubicek J, Penhaker M. Recent techniques and trends for retinal blood vessel extraction and tortuosity evaluation: a comprehensive review. *Ieee Access.* 2020;8:197787–816.
5. Abramoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. *IEEE Rev Biomed Eng.* 2010;3:169–208.
6. Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans Med Imaging.* 1989;8(3):263–9.
7. Mendonca AM, Campilho A. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans Med Imaging.* 2006;25(9):1200–13.
8. Khan TM, Alhusein M, Aurangzeb K, Arsalan M, Naqvi SS, Nawaz SJ. Residual connection-based encoder decoder network (RCED-Net) for retinal vessel segmentation. *IEEE Access.* 2020;8:131257–72.
9. Tu Z, Zhou Q, Zou H, Zhang X. A multi-task dense network with self-supervised learning for retinal vessel segmentation. *Electronics.* 2022;11(21):3538.
10. Creswell A, White T, Dumoulin V, Arulkumar K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. *IEEE Signal Process Mag.* 2018;35(1):53–65.
11. Wang L, Chen W, Yang W, Bi F, Yu FR. A state-of-the-art review on image synthesis with generative adversarial networks. *Ieee Access.* 2020;8:63514–37.
12. Radha K, Karuna Y. Modified depthwise parallel attention UNet for retinal vessel segmentation. *IEEE Access.* 2023;11:102572–88. <https://doi.org/10.1109/ACCESS.2023.3317176>
13. Radha K, Yepuganti K, Saritha S, Kamireddy C, Bavirisetti DP. Unfolded deep kernel estimation-attention UNet-based retinal image segmentation. *Sci Rep.* 2023;13(1):20712.
14. Zhao H, Li H, Maurer-Stroh S, Cheng L. Synthesizing retinal and neuronal images with generative adversarial Nets. *Med Image Anal.* 2018;49:14–26.
15. Sagar MA, Bullivant D, Mallinson GD, Hunter PJ. A virtual environment and model of the eye for surgical simulation. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques 1994 Jul 24 (pp. 205–212)*.
16. Trucco E, Ruggeri A, Karnowski T, Giancardo L, Chaum E, Hubschman JP, Al-Diri B, Cheung CY, Wong D, Abramoff M, Lim G. Validating retinal fundus image analysis algorithms: issues and a proposal. *Investig Ophthalmol Vis Sci.* 2013;54(5):3546–59.
17. Tavakkoli A, Kamran SA, Hossain KF, Zuckerbrod SL. A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs. *Sci Rep.* 2020;10(1):21580.
18. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal.* 2019;54:30–44.
19. Yurt M, Dar SU, Erdem A, Erdem E, Oguz KK, Çukur T. mustGAN: multi-stream generative adversarial networks for MR image synthesis. *Med Image Anal.* 2021;70:101944. <https://doi.org/10.1016/j.media.2020.101944>
20. Yang T, Wu T, Li L, Zhu C. SUD-GAN: deep Convolution generative adversarial network combined with short connection and dense block for retinal vessel segmentation. *J Digit Imaging.* 2020;33:946–57.

21. Liu M, Wang Z, Li H, Wu P, Alsaadi FE, Zeng N. AA-WGAN: attention augmented Wasserstein generative adversarial network with application to fundus retinal vessel segmentation. *Comput Biol Med.* 2023;158:106874.
22. Zhou Y, Chen Z, Shen H, Zheng X, Zhao R, Duan X. A refined equilibrium generative adversarial network for retinal vessel segmentation. *Neurocomputing.* 2021;437:118–30.
23. Chen Z, Jin W, Zeng X, Xu L. Retinal vessel segmentation based on task-driven generative adversarial network. *IET Image Proc.* 2020;14(17):4599–605.
24. Park KB, Choi SH, Lee JY. M-GAN: retinal blood vessel segmentation by balancing losses through stacked deep fully convolutional networks. *IEEE Access.* 2020;8:146308–22.
25. Kar MK, Neog DR, Nath MK. Retinal vessel segmentation using multi-scale residual convolutional neural network (MSR-Net) combined with generative adversarial networks. *Circuits Syst Signal Process.* 2023;42(2):1206–35.
26. Liang N, Yuan L, Wen X, Xu H, Wang J. End-to-end retina image synthesis based on CGAN using class feature loss and improved retinal detail loss. *IEEE Access.* 2022;10:83125–37.
27. Kim M, Kim YN, Jang M, Hwang J, Kim HK, Yoon SC, Kim YJ, Kim N. Synthesizing realistic high-resolution retina image by style-based generative adversarial network and its utilization. *Sci Rep.* 2022;12(1):17307.
28. Saeed AQ, Sheikh Abdullah SN, Che-Hamzah J, Abdul Ghani AT, Abu-Ain WA. Synthesizing retinal images using end-to-end vaes-gan pipeline-based sharpening and varying layer. *Multimedia Tools Appl.* 2024;83(1):1283–307.
29. Ahn S, Song SJ, Shin J. FundusGAN: fundus image synthesis based on semi-supervised learning. *Biomed Signal Process Control.* 2023;86:105289.
30. Pidhorskyi S, Adjero DA, Doretto G. Adversarial latent autoencoders. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020* (pp. 14104–14113).
31. Chen Y, Yi Z. Adaptive sparse dropout: Learning the certainty and uncertainty in deep neural networks. *Neural Comput.* 2021;25:450:354–61. <https://doi.org/10.1109/TNNLS.2022.3141665>
32. Staal J, Abramoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging.* 2004;23(4):501–9.
33. Hoover AD, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans Med Imaging.* 2000;19(3):203–10.
34. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, Barman SA. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng.* 2012;59(9):2538–48.
35. Vespe M, Jones G, Baker CJ. Lessons for radar. *IEEE Signal Process Mag.* 2009;26(1):65–75.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.