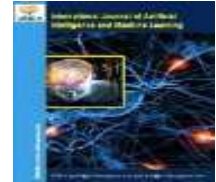




SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

An Improved Residual Pix2Pix Generative Adversarial Network for Finger Vein Segmentation

Anju Vincent¹, A. Anitha²

¹Research Scholar, Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kanyakumari, Tamil Nadu, India. E-mail: anju2511@gmail.com

²Associate Professor, Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kanyakumari, Tamil Nadu, India. E-mail: anithaa@niuniv.com

Abstract

Today, biometric technology has become a reliable solution for secure authentication. The uniqueness of finger vein patterns, contactless data collection, and low implementation costs make finger vein (FV) recognition stand out among other biometric methods. Researchers have developed many techniques to accurately recognize and classify vein patterns in fingers. In this work, Generative Adversarial Networks (GANs) are used to create a precise finger vein prediction model. The results highlight the effectiveness of the ResNet50 PIX2PIX Generative Adversarial Network and its potential applications in biometric recognition systems. Images are sourced from the THU-FVFD1 dataset and the Kaggle dataset. The process begins with standardizing the data and removing noise by using the Noise2Void model and pre-processing the images, followed by GAN-based segmentation to generate Segmented masks from the pre-processed images. This approach produced good results for finger vein recognition. An Akaze Detector is used for feature matching. The performance of the ResNet50 Pix2Pix GAN achieves an accuracy of 98.73%, which is greater than kaggle dataset. These findings demonstrate the effectiveness of ResNet50 Pix2Pix GAN in improving finger vein segmentation and its potential to enhance biometric authentication and identification industries.

Keywords: Matching, Biometrics, Segmentation Performance Analysis, Finger Vein Recognition, Noise2Void.

1. Introduction

In today's security-conscious society, biometrics-based verification frameworks are utilised in numerous significant applications. The majority of these applications require managing a large amount of complex biometric information simultaneously [1]. Therefore, conventional ordering systems are unsuitable for a biometric-based ID framework to recover biometric details faster. Biometrics is the study of utilising the physiological or conduct qualities of a human to check or distinguish the character of an individual. The finger vein recognition (FVR) technique has garnered more attention from researchers due to its exceptional security and accuracy [2]. Finger vein (FV)-based biometrics provide a significant advantage in live identification, as each person's venous network is concealed beneath the skin. FVR-based biometrics is still in its infancy and has many problems. Several factors can impact the performance of FV verification. These variables include uniform illumination, finger placement angle, and lighting conditions. Therefore, for feature extraction, recognition, and verification of images, robust and high-precision algorithms are needed [3]. Several researchers have proposed deep learning-based techniques for identifying finger veins. ML approaches can discover data-describing features that are used in discriminative training to distinguish between a large number of individuals [4].

This study proposes an approach for finger vein detection that uses a Pix-to-Pix GAN model to forecast the mask and further extract the traits' features. In finger vein recognition, feature extraction and segmentation enhance output quality and training stability. GAN can acquire more profound image features, and deep learning-based techniques can produce more consistent recognition outcomes compared to machine learning techniques. Using an Akaze keypoint detector, the feature matching can be performed and finds the matching performance.

Section 1 presents the introduction; Section 2 provides a detailed description of the related research associated with this work. The methodologies employed in this investigation are described in Section 3. The observations and outcomes are discussed in Section 4. Section 5 concludes the study.

2. Literature Review

Yang et al. [5] developed a restoration technique based on Gabor texture feature constraints; however, these methods struggle when the image is corrupted or missing vein patterns. Deep learning (DL) approaches have become a practical method for image inpainting. Still, they often fail to fully utilize texture information

during restoration, leading to poor matching authentication performance. Wang et al. [6] developed an FV restoration approach based on the neighbor binary Wasserstein GAN, but their focus was on restoration and did not prioritize textual properties for segmentation, a gap our feature-focused approach aims to fill. Lu et al. [7] used a transfer learning technique and developed a pre-trained Xception model with a modified depth-wise separable CNN.

Jain et al. [8] introduced Fisher vector computation for improved matching. Driver Auth, a risk-based multi-model biometric recognition system, achieved a recognition rate of 99.13. Den et al. [9] proposed a GAN model that is not meant for denoising and is conceptually similar to the Noise to Void model. The model is trained to predict unknown pixel values based on their surroundings, and the network is later used to generate synthetic images. GAN was first introduced in FVR in 2019 [10]. They addressed serious issues with CNN for feature representation, such as (1) CNN's fully connected layer is limited to the number of FV images, which increases processing time, and (2) CNN's feature extraction is poor due to low-quality finger vein images. Self-supervised training methods, like Noise2Void [11], offer a promising alternative because they only work with single noisy images. This is achieved by blocking or masking the network's receptive fields in the center (blind spot).

Cai et al. [12] present a segmentation framework based on Generative Adversarial Networks (GANs) that uses image translation capabilities to convert original CT images into accurate segmented masks. Several GAN architectures, including Pix2Pix and Residual GANs, were tested on the LUNA16 lung CT dataset. The Residual GAN outperformed both traditional segmentation methods and standard GAN models, scoring 98.3% accuracy, a Dice coefficient of 0.96, and a Jaccard index of 0.92. These findings demonstrate the effectiveness of GANs in capturing complex lung structures, suggesting that such models can significantly improve computer-aided diagnosis systems and clinical decision-making in pulmonary care. Wang et al. [13] proposed ResGAN, a network that enhances image generation from degraded inputs by using a ResNet-like direct path and circular adversarial training. Tested on multiple datasets, . A novel GAN was introduced by Kim et al. [14] to normalise the illumination of FV images. This GAN generates a one-channel image using texture information using a residual image generating block. It improves recognition performance by restoring FV texture information that was distorted by extreme light nonuniformity. Cho et al. [15] introduced a modified GAN for restoring blurred images, followed by recognition of the restored images through a deep CNN. Experiments on two public databases, the SDUMLA-HMT FV database and the HKPU-FV database, showed that this method outperforms existing techniques. Gao et al. [16] proposed a framework combining the Vision Transformer architecture and Capsule Networks to enhance FVR by exchanging local information while considering global dependencies and spatial variations.

Huda F. et al. [17] present an advanced Pix-to-Pix Generative Adversarial Network (GAN) model designed to improve the segmentation of thyroid nodules in ultrasound images. This model utilizes an encoder-decoder framework inspired by the U-Net and CNN architecture for segmentation tasks. The discriminator is built using a multilayered CNN that distinguishes between real and generated masks. Experiment results show that this improved model achieves a detection accuracy of 97% for malignant areas in thyroid ultrasound images, surpassing the effectiveness of semi-supervised segmentation methods. This progress emphasizes the potential of combining supervised loss functions with GANs to improve the reliability and accuracy of medical image segmentation.

This study aims to develop a robust and reliable system that can effectively capture, preprocess, and segment finger vein images. It addresses issues such as low contrast, noise, and occlusions. Despite these challenges, the goal is to ensure high accuracy and security across various applications, including access control, border security, and financial transactions. Using deep learning methods, specifically Pix2Pix GANs, this research seeks to assist in creating a dependable and precise finger vein detection system. The results demonstrate effectiveness and potential applications in biometric recognition systems.

The objectives of the study are:

1. Enhance finger vein image quality using the Noise2Void technique to handle blurred and noisy images without clean ground-truth data.
2. Develop a hybrid ResNet50-U-Net based Pix2Pix GAN model for accurate finger vein segmentation.
3. Examine the effectiveness of ResNet 50 Pix-to-Pix GAN in finger veins and evaluate its performance with a ResNet model for segmentation.
4. Extract and match discriminative vein features using Akaze keypoint detection. Using the Akaze key point detector, and the matching percentage is calculated.
5. Compare the proposed deep learning-based method's results to those of conventional machine learning techniques. Our method not only improved FV segmentation accuracy but also removed the challenge of processing blurred and noisy data.
6. Evaluate performance on real and synthetic datasets to ensure high accuracy and reliability.
7. The key novelty of this work lies in integrating a pre-trained ResNet50 encoder into the Pix2Pix GAN architecture to enhance feature extraction for vein segmentation, combined with a tailored post-segmentation matching process using Akaze.

Image acquisition, pre-processing, segmentation, and matching are steps in the finger vein detection process. An image sensor and an infrared light source are the two components of the FV image acquisition equipment used in this process. The obtained FV images are pre-processed to facilitate feature extraction. Common image pre-processing techniques include data enrichment, noise reduction, and image contrast enhancement.

Finger vein recognition is a biometric identification method that uses near-infrared light to capture images of a person's unique vein patterns beneath their skin. This method offers high security and accuracy, as it is difficult for anyone to spoof vein patterns. By utilizing advanced image processing techniques, including feature extraction and pattern recognition, finger vein recognition systems can reliably identify individuals, making them a favorable choice for many applications, from access control and border security to financial transactions.

3.Methodology

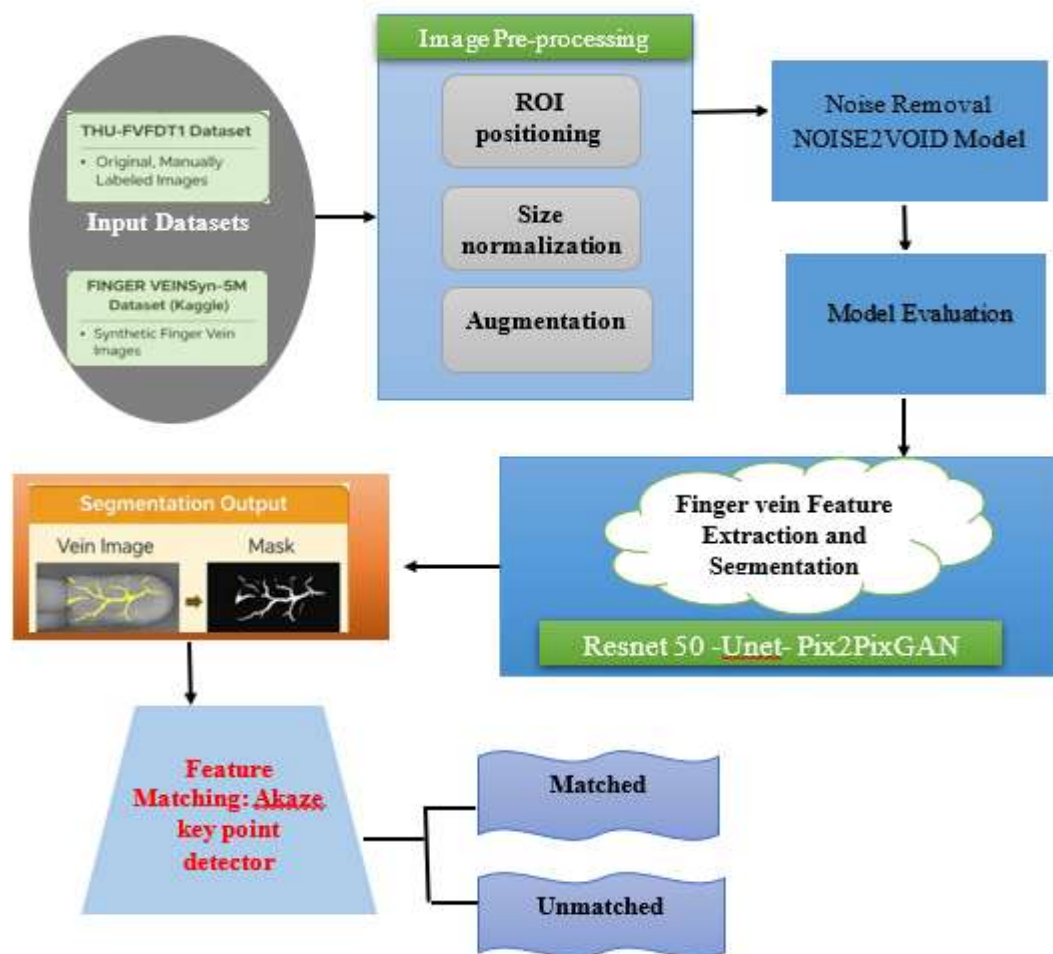


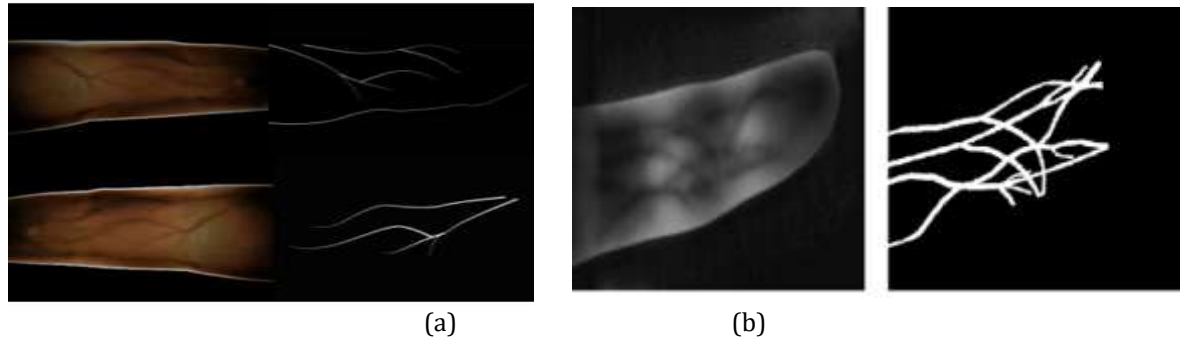
Fig.1.Overall Architecture

Figure 1 illustrates the overall framework of the proposed approach. Initially, finger-vein images are acquired from the dataset and subjected to pre-processing, which includes region of interest (ROI) extraction, size normalization, and data augmentation to ensure uniformity and enhance generalization. Subsequently, image denoising is performed using the Noise2Void model to suppress noise while preserving structural details. The denoised images are then provided as input to a Pix2Pix-based Generative Adversarial Network (GAN) incorporating a ResNet-50 encoder. This network performs feature extraction and segmentation simultaneously, generating both segmented finger-vein images and their corresponding mask representations. Finally, feature matching is carried out using the Akaze keypoint detection and description algorithm. A matching score is computed based on descriptor similarity, which serves as the decision metric for identification.

3.1Image Acquisition

The primary dataset used in our study is THU-FVFD1 [18], which includes raw images of FV and finger dorsal images. THU-FVFD1 was named before June 9, 2014. Most of the images were provided by students and staff volunteers from Tsinghua University in Shenzhen. The raw images have a resolution of 720x576

pixels. The ground-truth annotations for the dataset were generated manually to ensure accurate labeling. In addition to THU-FVFD1, a second dataset, FINGER VEINSyn-5M [19], was used to evaluate the robustness of the proposed framework. This dataset is publicly available on Kaggle and consists of a large collection of synthetically generated finger-vein images designed to simulate realistic vascular patterns.



(a) (b)
Fig. 2 Sample dataset Masks for

(a) Thu-Fvfdt1 Dataset And (b) Finger Veinsyn-5m Dataset

3.2 Noise Removal by Noise2Void Model

The Noise2Void model is used to remove noise from the image after the preprocessing stage. This model uses the U-Net as its backbone architecture and employs the Blind Spot mechanism during training. This deep learning model, called Noise2Void (N2V), was specifically designed to remove noise from images without requiring clean target data. By training on the actual faulty input, this model can effectively detect and remove noise. This process aims to improve image quality by eliminating unwanted, random variations in color and brightness that could obscure important details.

3.3 Image formulation

Let $y = (y_1, \dots, y_n)$ is considered as the noisy image and $s = (s_1, \dots, s_n)$ is the underlying clean signal. The denoised output is estimated using a parametric function

$\hat{s} = f(y; \theta)$ Assuming that the noise affecting each pixel is conditionally independent, the likelihood of the observed image can be expressed as:

$$p(y|s) = \prod_{i=1}^n p(y_i|s_i). \quad (1)$$

$p(y_i|s_i)$ will be referred to as the observation likelihood and termed as a random noise model. This noise model is typically computed using a simple stack of calibration images. It is unaffected by the contents of the images to be denoised. The creators of Noise2Void demonstrate that training remains feasible even in the absence of noisy training pairs. They utilize single images to generate input and target data during training. In a normal case, the central pixel is also included in the prediction, but in N2V, the central pixel is not included, so the network learns from data from the neighbouring pixel. As a result, pixel wise prediction of a trained N2V network can be viewed as

$$\hat{s}_i = \tilde{g}(\mathbf{x}_{RF}(i); \theta) \approx s_i, \quad (2)$$

The model is implemented using a U-Net-based encoder-decoder architecture with skip connections, employing 3×3 convolutions, ReLU activation, max-pooling for downsampling, and transposed convolutions for upsampling, followed by a 1×1 convolution for output reconstruction. During training, input images are divided into 64×64 patches, and random pixels are masked to enforce the blind-spot constraint. The model is optimized using the Adam optimizer with mean squared error (MSE) loss. Compared to supervised approaches such as Noise2Clean and Noise2Noise, Noise2Void eliminates the need for clean or paired noisy data, making it highly suitable for medical imaging applications. The denoised output improves vein visibility and enhances the performance of subsequent GAN-based segmentation and feature extraction stages.

Noise2Void [20] Denoising provides benefits over traditional DL methods, such as noise-to-low-noise and noise-to-noise. Each pixel in the same image is replaced by another pixel in Noise2Void models. N2V model training typically doesn't require many epochs. Training patches and validation patches for both datasets are shown in Figure 3

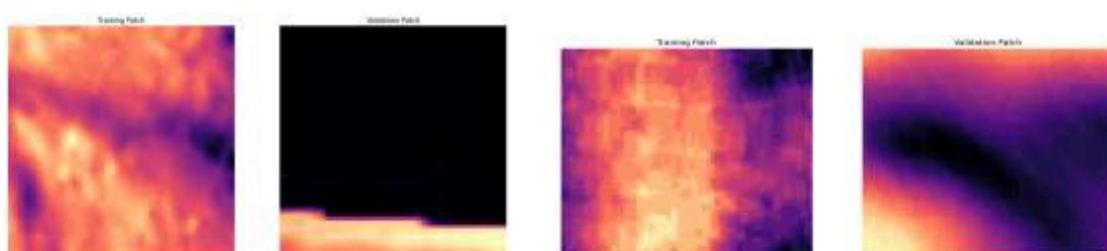


Fig.3 Training and validation patch for

(a)Thu-Fvfdt1 Dataset And (b)Finger Veinsyn-5m Dataset

3.4 Vein detection

In medical imaging, vein segmentation is a crucial task, but manual segmentation can be laborious and error-prone. Recent advances in deep learning have shown promising results in image segmentation challenges. This paper proposes a Resnet50 -Unet Pix2Pix GAN-based method for vein segmentation [21]. The basic Unet generator model in the Pix2Pix GAN has been replaced with a ResNet-50 encoder-generator model, and their performances are compared. The architecture of the Pix2Pix GAN includes a discriminator network that distinguishes generated images from real ones and a generator network that takes an input image and produces a matching output image (Figure 5). To enable accurate localisation using skip connections, the decoder component adheres to a U-Net-style architecture.

3.5 Pix2Pix GAN

Traditional Pix2Pix GANs use a standard U-Net architecture in the generator; this work introduces a modified Residual structure that includes a ResNet50 encoder in the Generator Model. This innovative architecture combines the strengths of both ResNet and U-Net, allowing the generator to produce more accurate and detailed images. The ResNet 50-based U-Net architecture maintains the core U-Net structure while integrating. ResNet blocks improve feature extraction capabilities. The vanishing gradient problem is reduced, and feature learning is improved by the residual functions that the ResNet blocks are designed to learn. Steps in Pix 2Pix GAN are represented in Figure 4. This combination allows the generator to:

- 1) ResNet blocks allow the generator to capture complex contextual relationships between image features.
- 2) The U-Net architecture guarantees that spatial information remains intact throughout the generator.
- 3) The combined architecture enables the generator to produce highly detailed and realistic images.

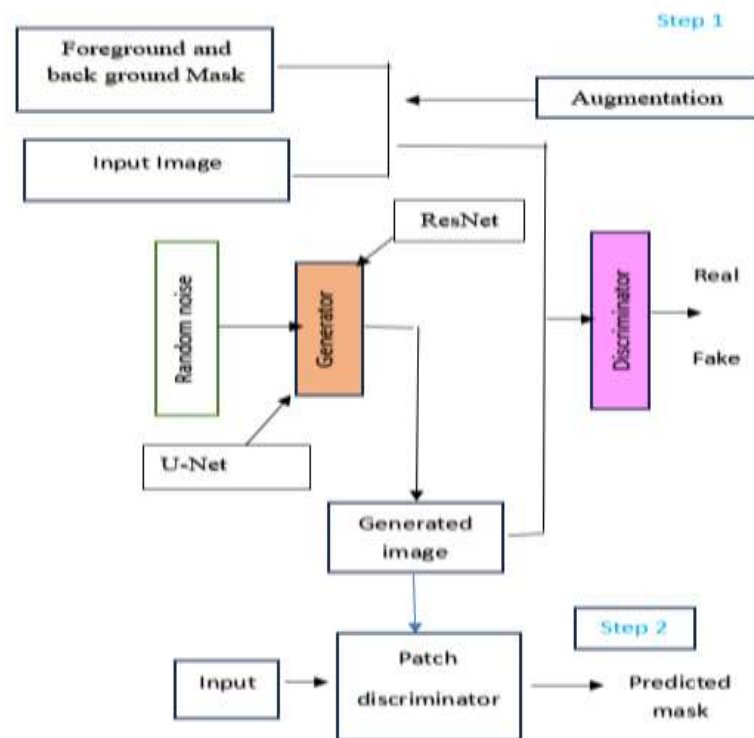


Fig 4. Steps in Pix2Pix GAN

3.6 Resnet 50 Encoder -Pix2Pix Model Structure

The encoder architecture consists of a U-Net architecture [22] with an image size of 512x512. The Unet consists of an encoder part and a Decoder part. Pix2Pix GAN, which utilizes a ResNet-50 Encoder architecture, is highly effective for image-to-image translation tasks. The proposed architecture combines a ResNet50 encoder [23], pretrained on ImageNet, with a U-Net style decoder for Segmentation. The ResNet 50 encoder extracts deep features from the input image, while the U-Net decoder reconstructs the output image using skip connections to retain spatial information. Convolution 1 reduces the resolution to 256x256, followed by residual blocks conv2_x, conv3_x, conv4_x, and conv5_x (16x16). ResNet50 employs stride-2 convolutions to decrease spatial dimensions. The decoder up-samples the resolution and features a basic U-Net decoder structure. At each stage, skip connections concatenate the up-sampled feature maps with the corresponding encoder features, enhancing spatial information preservation. The decoder then reconstructs the full-resolution output segmentation map (512x512). Convolutional blocks use 3x3

convolutions with ReLU activation, with upsampling performed via transposed convolutions or interpolation, followed by concatenation and additional convolutions.

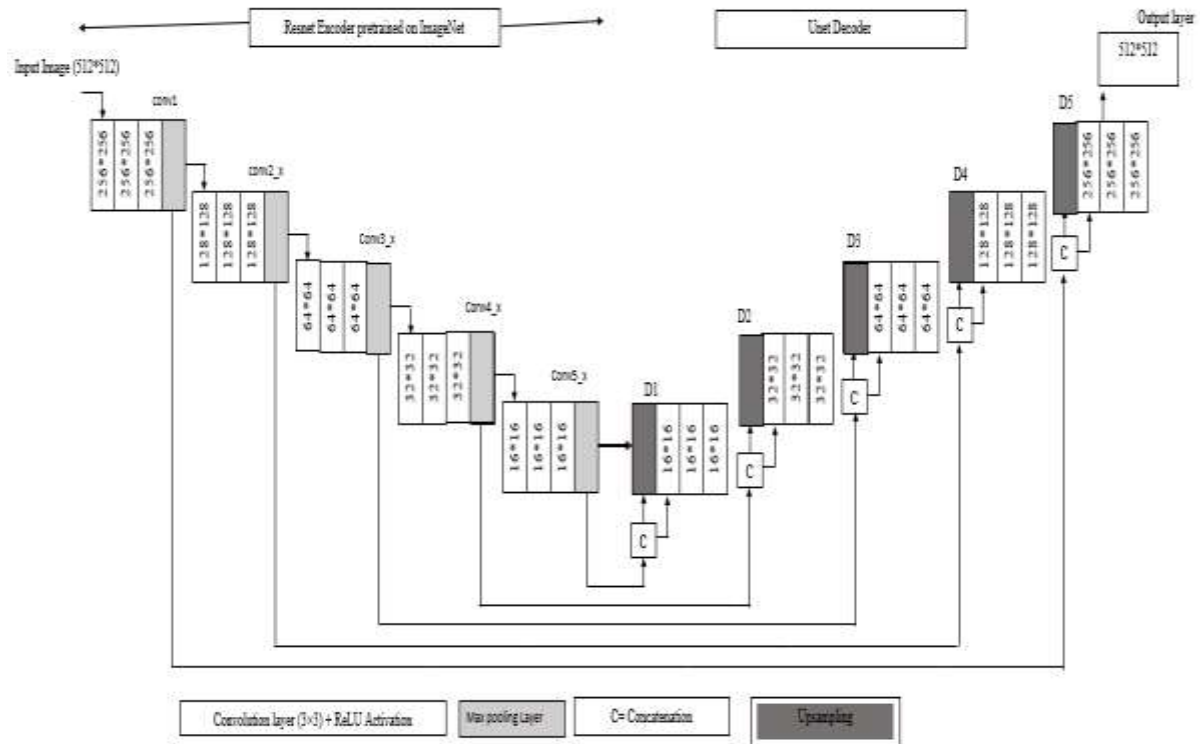


Fig.5. ResNet 50 Encoder Generator Model

During Training, the generator minimises both adversarial loss (to fool the discriminator) and reconstruction loss, ensuring that the generated output closely approximates the ground truth image. Simultaneously, the discriminator is trained to differentiate between real and generated images, thereby guiding the generator to produce more realistic outputs. Figure 5 represents the ResNet 50 Unet Pix2Pix GAN model employed in this study.

3.7 Discriminator

A Patch GAN architecture includes many transposed convolutional blocks used by the discriminator. This setup determines whether small NxN regions of the input image are real or fake. A probabilistic map showing the authenticity of each patch is generated by applying the Patch GAN discriminator [24,25] convolutionally across the entire input image. Figure 6 represents the architecture of the Patch GAN discriminator.

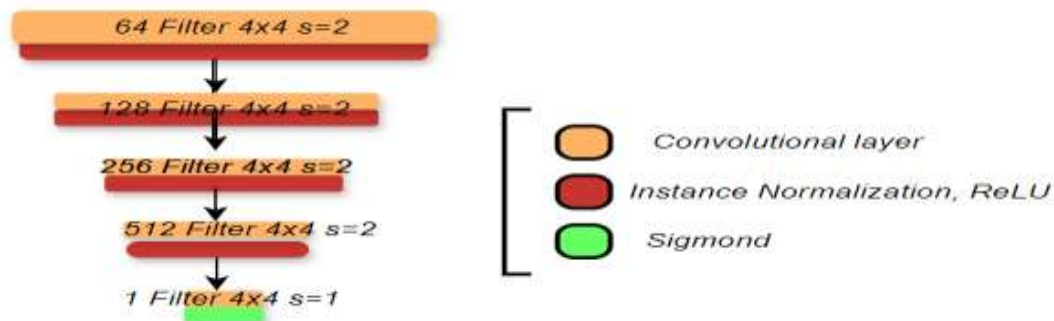


Fig.6. Patch GAN discriminator

D.Loss

The patch-wise predictions are averaged to determine the discriminator's final output, D1. GANs' loss can be represented as

$$L_{GAN}(G, D1) = E_{a,b}[\log D1(a, b)] + E_{x,z}[\log(1 - D1(a, G(a, z)))] \quad (3)$$

Here, G1 tries to minimize this loss against an adversarial D1 that attempts to maximize it. In the generator and D1(a,b) is the Discriminator probability that the image pair is real. We also contrast with an

unconditional variation where the discriminator does not attempt to assess the significance of conditioning the discriminator:

Equation 4, a is the input image, and b is the ground truth image. $G(a, z)$ is the generated image from the

$$L_{GAN}(G, D1) = E_y[\log D1(b) + E_{x,x}[\log(1 - D1(G(a, z)))] \quad (4)$$

$$L_{L1}(G) = E_{a,b,z}[\|y - g(a, z)\|_1] \quad (5).$$

Equation 5 represents the Reconstruction Loss, also known as L1 Loss. The discriminator's [26] function remains the same, but the generator's goal is to fool the discriminator while also producing outputs that are close to the ground truth in terms of L2 distance. Additionally, we consider using L1 distance instead, which typically results in less blurring than L2. Our final loss is

$$G^* = \arg \min_G \max_D L_{GAN}(G, D1) + (G) \quad (6)$$

In the absence of z , the network can still train a mapping from x to y , but it produces deterministic results and cannot match any distribution other than a delta function. Earlier GANs recognised this by including Gaussian noise z as generator input alongside x .

3.8

3.9 The proposed AKAZE algorithm:

Extraction had a major effect on the final accuracy and effectiveness of the FV matching procedure. Most FV identification systems use the SURF technique to improve FV validation, and lines and regional borders in images suggest similar space areas. The Accelerated-KAZE (AKAZE) extraction procedure with function-matching scores and effective repeatability was thus utilised in our work. The KAZE's accelerated version (AKAZE), aids in reducing the amount of time required for feature extraction and the identification process. It is also good for comparing images taken before and after an activity in a single block since it shows less variation in matching scores across incidence angles. The Hessian matrix algorithm was utilized to extract the AKAZE [27] feature detection L_{xy}^i , L_{yy}^i and L_{xx}^i are represented in Eq. (7).

$$L_{Hessian}^i = \sigma_{i,norm}^2 (L_{xx}^i L_{yy}^i - L_{xy}^i L_{xy}^i) \quad (7)$$

here, $\sigma_{i,norm}^2$ is the normalized scale factor defined as $\sigma_{i,norm}^2 = \sigma_i / 2^{0i}$ of the function. Here L_{xy}^i , L_{yy}^i and L_{xx}^i denote the second Gaussian derivatives in the horizontal, vertical and cross directions, respectively. When the Hessian response at the current scale exceeds the predefined threshold, $T_A - KAZE$, the algorithm identifies extrema by comparing the candidate point with its neighbors within a $3 \times 3 \times 3$ neighborhood across scales, using the Hessian determinant criterion.

A search using σ_i sample steps was carried out within a radius of $6\sigma_i$ to centralize the KAZE[28] features. The differential values of all neighboring points along the circle around the center were used to compute Gaussian-weighted positions in the first phase. These values represent the sector-wise pixel response of the image. The sector region with the highest value was provided as the main orientation for the function's aspect points after the entire circle had been completed. Furthermore, it detects a large number of inlier features in the function and generates a non-linear scale-space instead of using Gaussian blurring

The first step is to use the Hessian matrix to identify the FV image's important points $H(x; \sigma)$ is presented in Eq. (8).

$$H(x; \sigma) = \begin{pmatrix} L_{xx}(x; \sigma) & L_{xy}(x; \sigma) \\ L_{xy}(x; \sigma) & L_{yy}(x; \sigma) \end{pmatrix} \quad (8)$$

$$L_{xx}(x; \sigma) = 1(x) \times \frac{\partial^2}{\partial x^2} g(\sigma) \quad (9)$$

$$L_{xy}(x; \sigma) = 1(x) \times \frac{\partial^2}{\partial x \partial y} g(\sigma) \quad (10)$$

$$L_{yy}(x; \sigma) = 1(x) \times \frac{\partial^2}{\partial y^2} g(\sigma) \quad (11)$$

Here $L_{xx}(x; \sigma)$ is the Gaussian convolution's second derivative, $g(\sigma)$, at scale σ . "i", the medium $H(x; \sigma)$ is distinct at $H(x)$ and scale σ is indicated in Eqs. (9-11)). The box filter is used to simulate the Gaussian second-order derivative in order to reduce computation time. The factors such as $D_{xx}(x; \sigma)$, $D_{xy}(x; \sigma)$, and $D_{yy}(x; \sigma)$ are used to reduce processing time and speed up convolutional computations. Eq. (9) provides the estimated value of the Hessian's determinant.

$$\det(H_{approx}) = D_{xx} D_{yy} - (0.9 D_{xy})^2 \quad (12)$$

To determine the orientation of each recognized feature, the first method is to convolve the feature pixel with its surrounding area using the straight and perpendicular Haar wavelet strainers. Step two of the system involves creating a keypoint descriptor

The efficiency of AKAZE as an image-to-detection method is attributed to the fact that its matching score fluctuates less across all incidence angles. AKAZE finds a much higher number of inlier characteristics than other methods, and the fraction of inliers corresponding to aspects increases as the number of related

aspects increases[29]. It produced excellent results for cumulative point errors, extracted gesture errors, and the function's translational and rotational components in every interframe.

4.Results And Discussion

The proposed model was implemented on a system with an AMD Ryzen 5 processor, NVIDIA GTX-series GPU, 16 GB RAM, and 1 TB storage, running on Windows 11. The model was developed using Python. In Noise to Void training, each training patch in the dataset used to train the model is of size (64, 64).

The training and validation patches of Noise to Void are shown in Figure 3. Each training patch creates eight blind spots during training. The training runs for one hundred epochs. Based on the training data, the model's weights are updated after each epoch. Mean Squared Error is used as the loss function, and two metrics are monitored: mean absolute error and MSE. Validation losses are also low, indicating that the model converged successfully. The PSNR, SSIM, MSE, and RMSE values for 10 images in N2V training are listed in Table 1.

Table 1 Mse, Psnr, Ssim, And Rmse Values For Ten Images

Image	PSNR	SSIM	MSE	RMSE
1.	31.68	0.9930	0.44	0.66
2.	32.19	0.9954	0.39	0.62
3.	31.98	0.9936	0.41	0.64
4.	32.86	0.9916	0.33	0.57
5.	31.56	0.9924	0.45	0.67
6.	32.84	0.9955	0.33	0.57
7.	32.24	0.9941	0.38	0.61
8.	33.22	0.9950	0.30	0.54
9.	32.09	0.9944	0.40	0.63
10.	32.01	0.9939	0.40	0.63

4.1Evaluation Metrics

A finger vein recognition system's performance metrics are shown in Table 2. The following measures are used to assess the FVR system's performance:

Precision is the proportion of correctly identified finger veins (True positives, TP) to the total of both false positives (FP) and TP. Recall is the ratio of missed finger veins (false negatives) to the total of TP. The F1-score provides a balanced metric by calculating the harmonic mean of precision and recall.

We used the Intersection over Union (IoU) to assess our model's performance. [30] These metrics were calculated by comparing the FV segmentation prediction (F) with the ground truth in the THU-FVFD1 dataset and the Kaggle dataset. A higher IoU indicates a greater overlap between the predicted and reference FV regions, which signifies better segmentation accuracy. Accuracy is defined as the proportion of correctly classified pixels (both object and background) out of all pixels in the image.

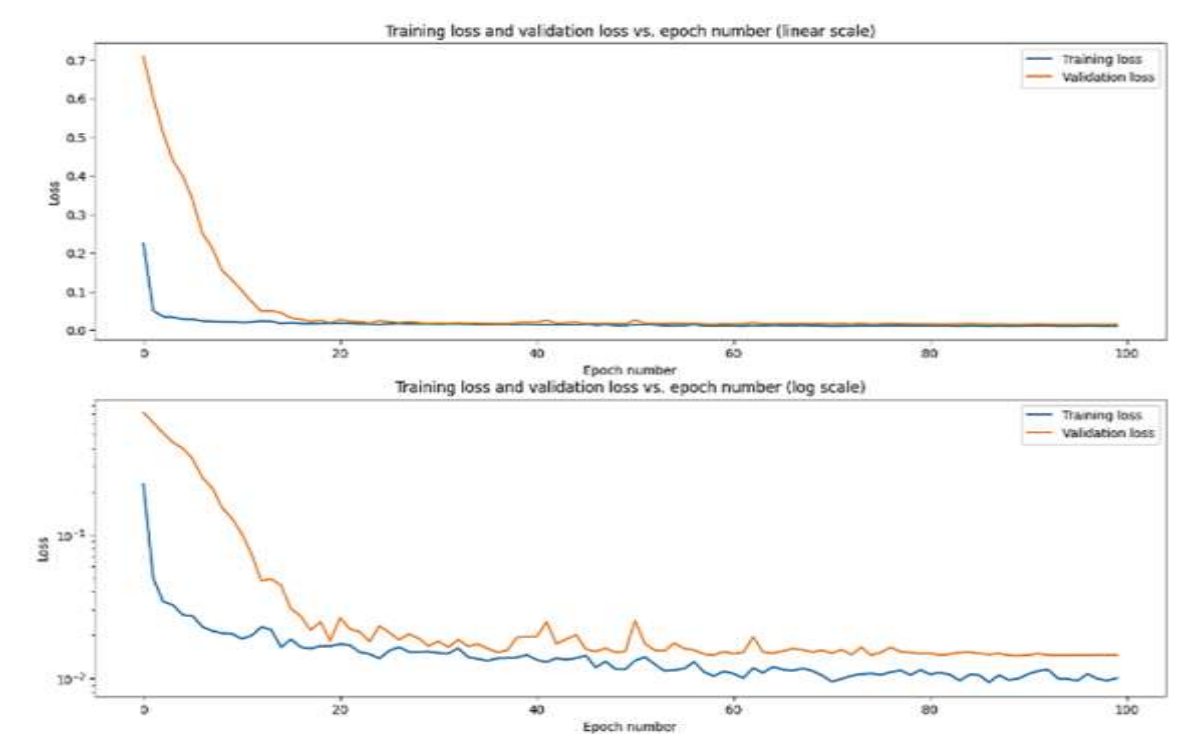


Fig. 8. Graph represents Training Loss and Validation Loss

Figure 8 shows the Loss on Training and Validation. The gap between the two losses suggests some overfitting, especially if validation loss starts to rise or stays higher while training loss keeps decreasing. Both losses show a downward trend on the graph. However, the fluctuations in validation loss point to a need for potential improvements to boost the model's stability and its ability to generalize.Noise2Void improves image quality by denoising while maintaining structural details. N2V removes random noise without requiring clean ground truth images for training. The output image is grayscale because N2V typically works on intensity patterns rather than preserving color. Figure 9presents the N2V input and output images for the same input.

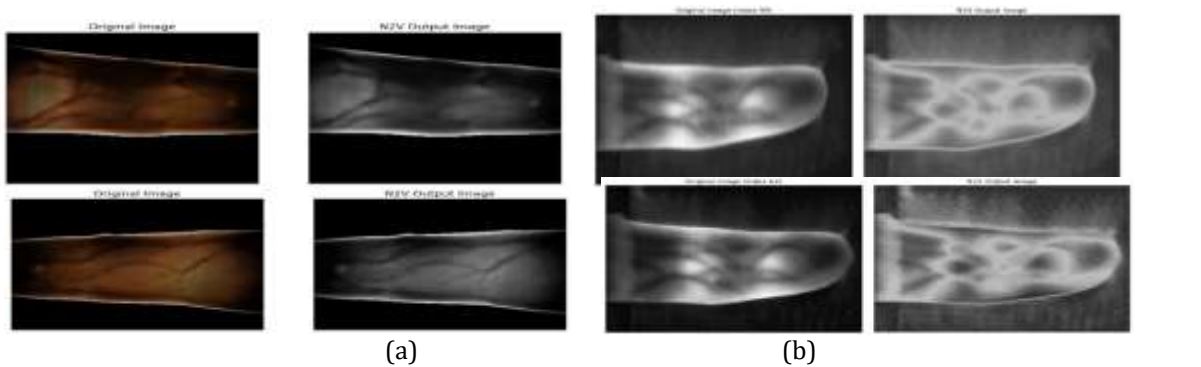


Fig. 9. Noise2Void image output for the input image(a)THU-FVFD1 Dataset and (b)FINGER VEINSyn-5M Dataset

Dataset	Original Image	Ground Truth Mask	Predicted Mask
THU-FVFD1			
	(a)	(b)	(c)

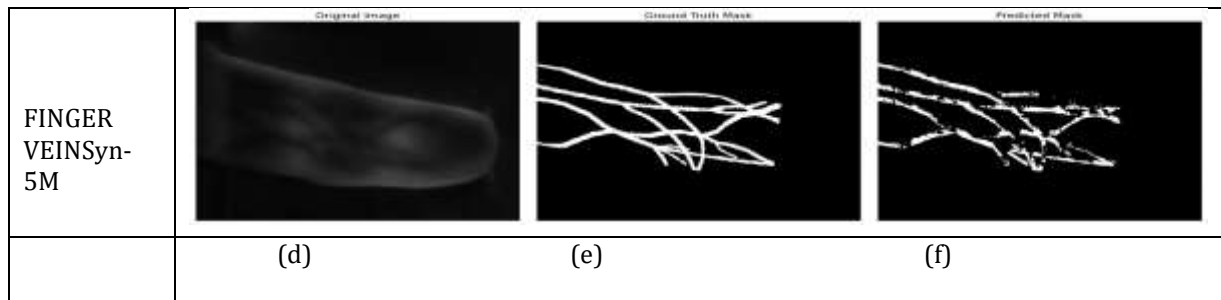


Fig. 10 Results of vein detection for THU-FVFD1 Dataset and FINGER VEINSyn-5M Dataset (a) Raw Image (b) Ground Truth Mask (c) Segmented output Resnet 50 Unet Pix2Pix GAN (d) Raw Image (e) Ground Truth Mask (f) Segmented output Resnet 50 Unet Pix2Pix GAN

The results presented in Fig. 10 demonstrate that the proposed model is capable of accurately extracting the primary vein structures across both the THU-FVFD1 and Syn-5M datasets. The predicted masks exhibit strong structural alignment with the ground truth annotations, particularly for major vein branches, indicating that the model effectively captures dominant spatial patterns. On the THU-FVFD1 dataset, the segmentation output preserves most of the main vascular pathways with minimal background leakage. However, minor discontinuities are observed in thinner vein regions, suggesting that the model may have limited sensitivity to low-contrast or extremely fine structures. For the Syn-5M dataset, the model demonstrates good generalization capability. The predicted masks closely follow the synthetic ground truth structure, confirming robustness across data domains. Nevertheless, slight over-segmentation and fragmented thin branches are noticeable, indicating occasional false positives and reduced continuity in small-scale features.

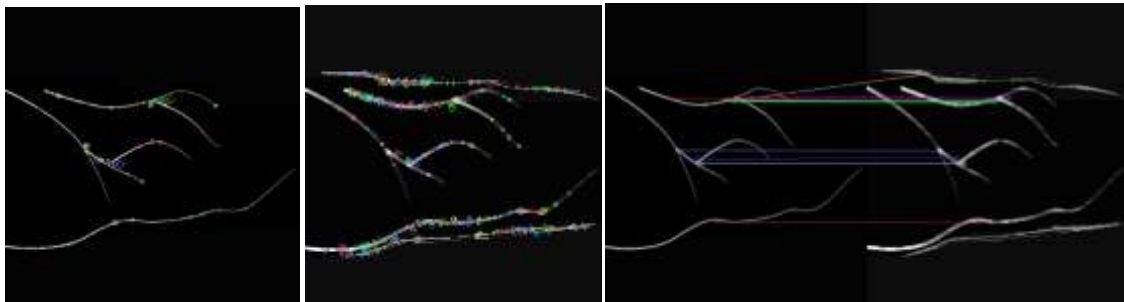


Figure 5: (a) Original image mask (b) Predicted image mask (c) AKAZE Feature matching

Figure 5 displays the feature matching outcomes of the AKAZE algorithm, which compares the predicted segmented vein masks and corresponding ground truth mask to determine how structurally similar the extracted patterns are. The multicoloured lines indicate the matched keypoints in the two masks. The branching points and curvature changes of the vein are represented by these keypoints. The successful detection and matching of these characteristic structures indicate a high degree of structural similarity between the predicted mask and the ground-truth mask, demonstrating the effectiveness of AKAZE for feature-based matching.

Table 2: Performance Evaluation Of Images in Thu-Fvfdt1 Dataset

Images	Precision	Recall	F1-score	IOU (Intersection over Union)
1	0.995	0.989	0.9917	0.9841
2	0.990	0.986	0.9868	0.9743
3	0.985	0.988	0.9860	0.9733
4	0.982	0.983	0.9820	0.9666
5	0.975	0.977	0.9752	0.9531

6	0.9746	0.963	0.9686	0.9400
Average	0.9836	0.9810	0.9817	0.9652

The proposed ResNet50-Pix2Pix model was trained using the Adam optimiser with a batch size of 1, momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and a fixed learning rate of 0.0002, aligning with the original Pix2Pix framework and promoting stable adversarial learning. Training occurred over 50 epochs, with the learning rate linearly decaying after the first 100 iterations to encourage convergence. The generator loss combined adversarial loss and an L1 reconstruction loss, weighted by $\lambda = 100$. The discriminator was optimized with binary cross-entropy loss within the 70×70 PatchGAN framework. The ResNet-50 encoder enhances feature representation through residual learning and deeper convolutional layers, resulting in better vein boundary discrimination against noise. The ResNet-50 U-Net model predicts masks that preserve vein continuity, reduce false positives, and more closely match the ground truth mask. This shows that adding a ResNet-50 encoder into Pix2Pix yields more accurate segmentation than the standard U-Net backbone. The finger vein recognition system exhibits excellent performance, with precision, recall, and F1-score indicating strong recognition capabilities. Although Image 6 displayed slightly lower recall (0.96) and F1-score (0.96), it still maintained high precision and specificity, highlighting the system's overall outstanding performance in finger vein recognition across all images. Table 3 presents each model's accuracy as a percentage.

Table 3: Performance Evaluation Of Images in Finger Veinsyn-5m Dataset

Images	Precision	Recall	F1-score	IOU (Intersection over Union)
1	0.971	0.922	0.946	0.962
2	0.962	0.914	0.937	0.953
3	0.968	0.919	0.943	0.959
4	0.960	0.912	0.935	0.951
5	0.969	0.921	0.944	0.960
6	0.963	0.916	0.939	0.955
Average	0.9655	0.9173	0.9407	0.9567

The confusion matrix in Fig. 11(a) assesses the segmentation performance of a Resnet 50-Pix2Pix UNet model using the THU-FVFD1 Dataset. More pixels are correctly identified in the foreground, indicating better segmentation performance. The model achieves a very high number of true negatives (257,449), indicating excellent background classification. However, the number of true positives (386) is relatively low compared to false negatives (1,311), suggesting that a significant portion of some actual vein pixels are misclassified as background. Figure 11(b) shows a confusion matrix for binary segmentation using a FINGER VEINSyn-5M Dataset. The model performs well in background detection as well as for the foreground, but it needs improvements in foreground detection, particularly in reducing false positives to enhance accuracy [32]. Figure 13 represents the loss graph for GAN models at first 100 Iterations. When calculating the LOSS, the residual GAN also performs well for first Dataset.

The GAN is trained with generator and discriminator loss values over 50 epochs. The training loss for the GAN is shown in Figures 14 and 15. The generator's loss fluctuates significantly, suggesting difficulties in consistent improvement, possibly due to strong discriminators or training imbalance. Discriminator losses are low and stable, indicating they effectively distinguish between fake and real data.

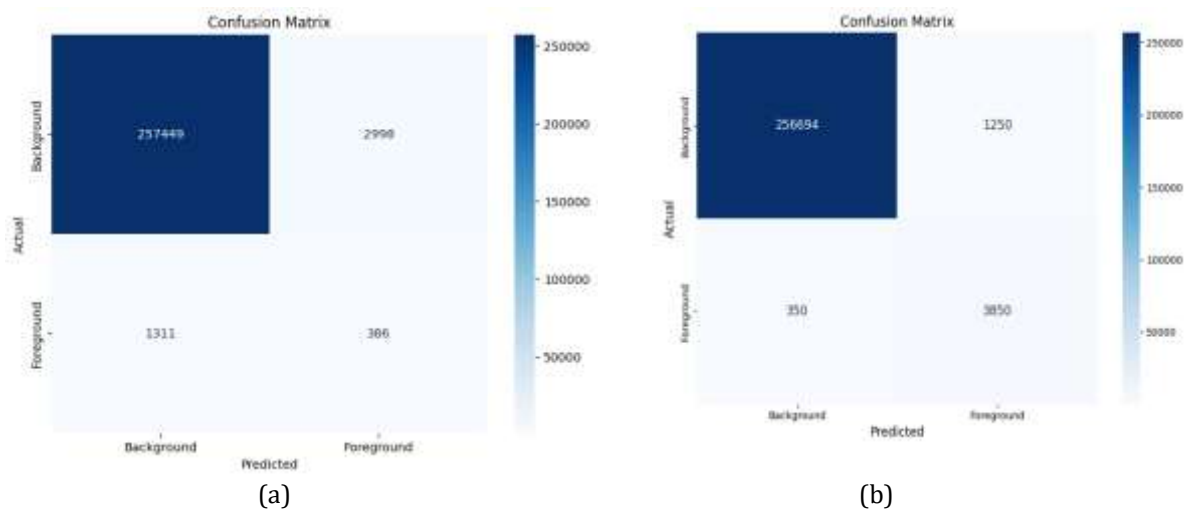


Fig.11 Confusion matrix for Pix2Pix Resnet50 Encoder Model for (a)THU-FVFD1 Dataset and (b)FINGER VEINSyn-5M Dataset

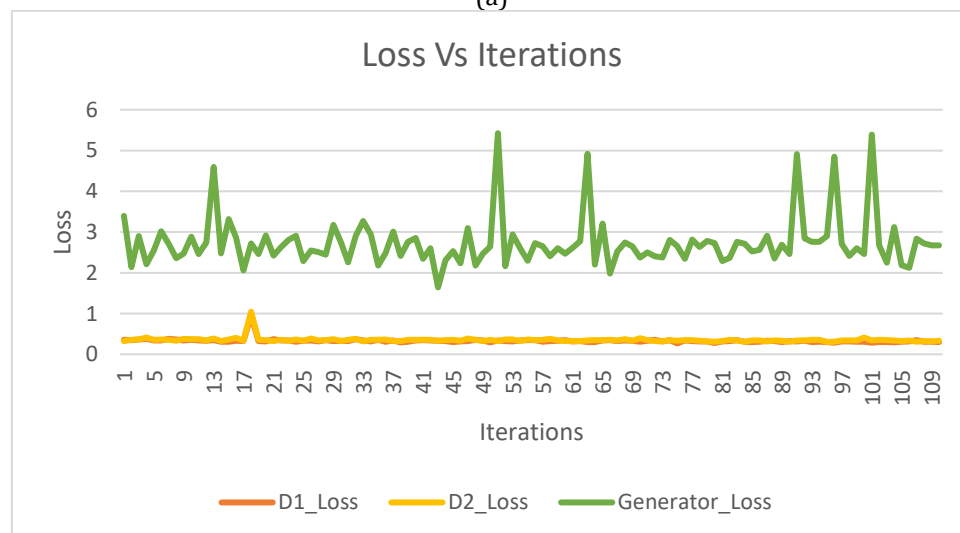
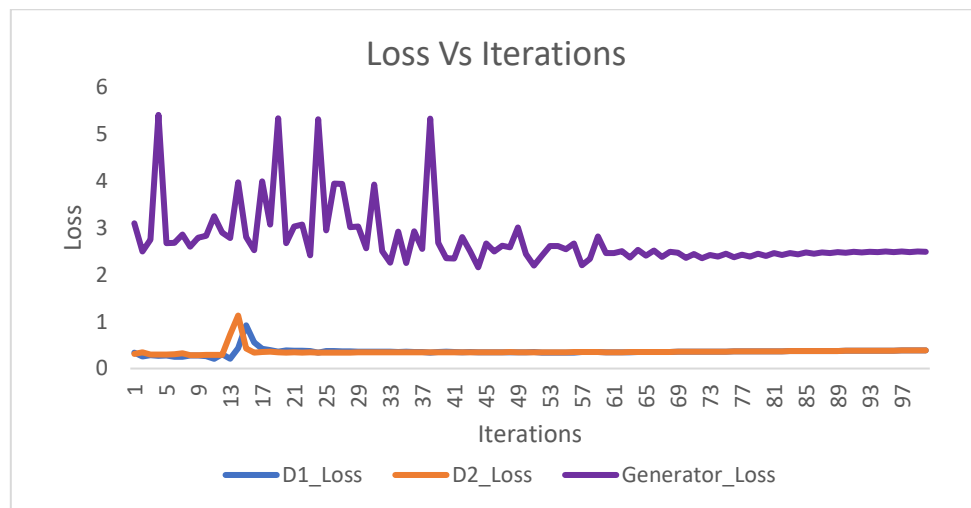


Fig. 12 Loss graph of ResNet Pix2Pix GAN for 110 Iterations for (a)THU-FVFD1 Dataset and (b)FINGER VEINSyn-5M Dataset

Table 3 displays various studies conducted by different authors on image segmentation using multiple models. Kansal et al. evaluated the U-Net v2 model on three datasets: ISIC, Kvasir-SEG, and CVC-Clinic DB,

achieving Dice scores of 92.48, 91.10, and 90.70, respectively, along with IoU values of 83.6, 85.9, 83.1, and 82.4. The enhanced skip connections in U-Net v2 facilitate better feature fusion between the encoder and decoder, resulting in improved segmentation boundaries across multiple medical image datasets. Multi-ResUNet, tested on two datasets, attained a Dice score of 0.956 and an IoU of 91.62 by employing multi-resolution feature aggregation, which captures structures at various scales and boosts segmentation accuracy. DiffVein applies two datasets related to finger veins (USM, THU-MVFDV3V) and achieves 92.22% accuracy with a Dice score of 86.47. On the MIMIC-CXR dataset, the SA-U-Net model (Zhong et al., 2023) attained 96.98% accuracy, a Dice score of 96.8, and an IoU of 91.9. Its spatial attention mechanism emphasizes key regions, enhancing segmentation accuracy and reducing false positives. The model reached 96.4% accuracy, with an IoU of 95.1%, Precision, Recall, and F1 Score of 97.9%, 97.4%, and 97.6%, respectively, when a Pix2Pix GAN with a U-Net encoder. Finally, replacing the encoder with ResNet50 in the Pix2Pix GAN improved performance on the same dataset, achieving 98.73% accuracy and a 96.7 IoU, illustrating that deeper feature extraction enhances segmentation of complex vein patterns. The hybrid ResNet50-U-Net-Pix2Pix GAN architecture achieves superior finger vein image segmentation performance on the THU-FVFD1 dataset when evaluated across all performance metrics.

Table 4: Performance Metrics With Other Models

Model	Year	Dataset	Accuracy	Dice / IoU	Other Metrics	Comments
U-Net v2[33]	2023	ISIC, Kvasir, CVC	-	ISIC Dice 92.48, Kvasir Dice 91.10 CVC - 90.70,	IoU=83.6,85.90, 83.10,82.40	Better skip connections
Multi-ResUnet [34]	2022	EchoNet Dynamics and Synthetic Dataset	-	Dice =0.956	Iou=91.62	Multi-resolution feature aggregation enhances segmentation accuracy
DiffVein [35]	2024	Finger vein (USM, THU-MVFDV3V)	92.22%	86.47, cIDice =0.93.16	AUC =96.1%	Diffusion+GAN hybrid
SA-Unet [36]	2023	MIMIC-CXR	96.98%	96.8	IoU=91.9 Sensitivity=82.12 Specificity=98.40	U-Net with spatial attention (SA-Unet)
Unet-Pix2Pix GAN	ours	Finger vein (THU-FVFD1)	96.4%	IoU=95.1	Precision =97.9, Recall =97.4, F1 Score=97.6	Uses Unet Encoder Generator Model
Hybrid-Resnet50-Unet-Pix2Pix GAN		Finger vein (THU-FVFD1)	98.73%	IoU =96.52	Precision =98.36, Recall =98.10 F1 score=98.17	ResNet50 encoder improves baseline Pix2Pix
Hybrid Resnet50-UnetPix2Pix GAN		Finger vein (FINGER VEINSyn-5M)	95.17%	IoU=95.67	Precision =96.56, Recall =91.8 F1 score=94.07	ResNet50 encoder improves baseline Pix2Pix

5.Conclusion

The study introduces an innovative Pix2Pix vein recognition method that aims to improve FV image segmentation for biometric applications. Initially, ROI positioning, size normalization, and augmentation enhance data consistency and robustness. The incorporation of the Noise2Void model effectively reduces noise without requiring clean target images, improving the quality of vein structures before segmentation. The hybrid ResNet50–U-Net–Pix2PixGAN architecture enables accurate feature extraction and segmentation, preserving both global contextual information and fine vascular details. This method produced good results for finger vein segmentation. Finally, an Akaze keypoint detector was proposed to perform feature matching by identifying descriptors in two images and calculating the similarity percentage to determine the segmentation performance. The use of the GAN model also facilitates data augmentation. The efficiency of the model is evaluated using several metrics in two datasets, with the ResNet-50 encoder combined with Pix2Pix GAN achieving the highest accuracy for the THU-FVFD1 Dataset. It ensures improved vein visibility, precise segmentation, and robust feature matching, making it suitable for real-world biometric authentication applications.

In the future, integrating multimodal biometrics could enhance accuracy and security in FV recognition by combining vein patterns with other biometric traits. The system's performance could be assessed and fine-tuned using various datasets to ensure it functions effectively in real-world applications.

Acknowledgements

We acknowledge the creators of the THU-FVFD1 Tsinghua database and FINGER VEIN Syn-5M Datasets. Data Availability: The dataset used is included in the article.

Authors Contributions

Anju Vincent: Writing – review & editing, Methodology, Analysis.

Dr A Anitha: Supervision, Correction, and Editing

References

1. G. Nayar and T. Thomas, Partial palm vein-based biometric authentication, *J. Inf. Secur. Appl.* 72 (2023) 103390.
2. R. Nathan and S. Sakthivel, Finger vein extraction and authentication for security purpose based on CNN, in *Proc. Int. Conf. Adv. Data Eng. Intell. Comput. Syst. (ADICS)* (IEEE, 2024), pp. 1–6.
3. C. Y. Liu, S. J. Ruan, Y. R. Lai, and C. Y. Yao, Finger-vein as a biometric-based authentication, *IEEE Consum. Electron. Mag.* 8(6) (2019) 29–34.
4. A. Krishnan, T. Thomas, and D. Mishra, Finger vein pulsation-based biometric recognition, *IEEE Trans. Inf. Forensics Secur.* 16 (2021) 5034–5044.
5. L. Yang, X. Liu, G. Yang, J. Wang, and Y. Yin, Small-area finger vein recognition, *IEEE Trans. Inf. Forensics Secur.* 18 (2023) 1914–1925.
6. Z. C. Wan, L. Chen, T. Wang, and G. C. Wan, An optimization algorithm to improve the accuracy of finger vein recognition, *IEEE Access* 10 (2022) 127440–127449.
7. Y. Lu, S. Wu, Z. Fang, N. Xiong, S. Yoon, and D. S. Park, Exploring finger vein-based personal authentication for secure IoT, *Future Gener. Comput. Syst.* 77 (2017) 149–160.
8. B. Chawla, S. Tyagi, R. Jain, A. Talegaonkar, and S. Srivastava, Finger vein recognition using deep learning, in *Proc. Int. Conf. Artif. Intell. Appl.* (2021), pp. 69–78.
9. A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, Pixel recurrent neural networks, in *Proc. Int. Conf. Mach. Learn. (ICML)* (2016), pp. 1747–1756.
10. W. Yang, C. Hui, Z. Chen, J. H. Xue, and Q. Liao, Fv-GAN: Finger vein representation using generative adversarial networks, *IEEE Trans. Inf. Forensics Secur.* 14(9) (2019) 2512–2524.
11. A. Krull, T. O. Buchholz, T. Vičar, M. Prakash, M. Lalit, and F. Jug, Probabilistic Noise2Void: Unsupervised content-aware denoising, *Front. Comput. Sci.* 2 (2020) 5.
12. J. Cai et al., CT image enhancement using stacked generative adversarial networks and transfer learning for lesion segmentation improvement, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)* (2018), pp. 1–9.
13. M. Wang, H. Li, and F. Li, Generative adversarial network based on ResNet for conditional image restoration, *arXiv preprint arXiv:1707.04881* (2017).
14. H. J. Seong, J. Choi, S. G. Kim, M. Owais, and K. R. Park, Inf-GAN: GAN for illumination normalization of finger-vein images, *Mathematics* 9(20) (2021) 2613.
15. J. Choi, K. J. Noh, S. W. Cho, S. H. Nam, M. Owais, and K. R. Park, Modified conditional GAN-based optical blur restoration for finger-vein recognition, *IEEE Access* 8 (2020) 16281–16301.
16. Y. Li, H. Lu, Y. Wang, R. Gao, and C. Zhao, ViT-Cap: A novel Vision Transformer-based capsule network model for finger vein recognition, *Appl. Sci.* 12(20) (2022) 10364.

17. H. F. AL-Shahad, R. Yaakob, N. M. Sharef, H. Hamdan, and H. A. Hassan, An improved Pix2Pix generative adversarial network model to enhance thyroid nodule segmentation, *J. Adv. Inf. Technol.* 16(1) (2025) 37–48.
18. Tsinghua University, THU-FVFDt Dataset, (2025). Available: <https://www.sigs.tsinghua.edu.cn/labs/vipl/thu-fvfdt.html>.
19. Y. Wang et al., FingerVeinSyn-5M: A Million-Scale Dataset and Benchmark for Finger Vein Recognition, (2025). Available: <https://www.kaggle.com/datasets/evanwang98/fingerveinsyn-5m>.
20. A. Krull et al., Noise2Void: Learning denoising from single noisy images, *IEEE Trans. Neural Netw. Learn. Syst.* 32(9) (2019) 3186–3196.
21. D. Luo, J. Huang, W. Yang, M. S. Shakeel, and W. Kang, RSNet: Region-Specific Network for contactless palm vein authentication, *IEEE Trans. Inf. Forensics Secur.* 20 (2025) 2734–2747.
22. O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)* (2015), pp. 234–241.
23. K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2016), pp. 770–778.
24. J. Choi et al., Restoration of motion blurred image by modified DeblurGAN for finger-vein recognition, *Sensors* 21(14) (2021) 4635.
25. B. Hou and R. Yan, Triplet-classifier GAN for finger-vein verification, *IEEE Trans. Instrum. Meas.* 71 (2022) 1–12.
26. J. Zhang, Z. Lu, M. Li, and H. Wu, GAN-based image augmentation for finger-vein biometric recognition, *IEEE Access* 7 (2019) 183118–183132.
27. P. F. Alcantarilla, J. Nuevo, and A. Bartoli, Fast explicit diffusion for accelerated features in nonlinear scale spaces, *British Machine Vision Conference (BMVC)* (2013) 13.1–13.11.
28. P. F. Alcantarilla, A. Bartoli, and A. J. Davison, KAZE features, *Lecture Notes in Computer Science (ECCV)* 7577 (2012) 214–227.
29. E. Karami, S. Prasad, and M. Shehata, Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images, *arXiv preprint arXiv:1710.02726* (2017).
30. [H. Qin, R. Hu, M. A. El-Yacoubi, Y. Li, and X. Gao, Local attention transformer-based full-view finger-vein identification, *IEEE Trans. Circuits Syst. Video Technol.* 33(6) (2022) 2767–2782.
31. S. Lee, J. Lim, M. Choi, et al., Improved finger vein recognition using generative adversarial networks and transfer learning, *J. Neural Eng.* 19(10) (2025) 101101.
32. W. Nimalsiri et al., CXLSeg dataset: Chest X-ray with lung segmentation, in *Proc. Int. Conf. Cyber Manag. Eng. (CyMaEn)* (2023), pp. 327–331.
33. Y. Peng, M. Sonka, and D. Z. Chen, U-Net v2: Rethinking the skip connections of U-Net for medical image segmentation, *arXiv preprint arXiv:2311.17791* (2023).
34. N. Ibtehaz and M. S. Rahman, MultiResUNet: Rethinking the U-Net architecture for multiModel biomedical image segmentation, *Neural Netw.* 144 (2022) 128–143.
35. H. Li, Z. Zhang, X. Wang, and J. Li, DiffVein: Diffusion model and adversarial learning for robust finger vein segmentation, *IEEE Trans. Biometrics Behav. Identity Sci.*, early access (2024).