Department of Information and Communication Technology

Machine Learning in Medicine

Retina Blood Vessel Segmentation

ID - Student Name:

22BI13472 - Nguyễn Bá Vinh

22BI13220 - Nguyễn Minh Khôi

22BI13227 - Trần Trung Kiên

22BI13462 - Chu Hoàng Việt

22BI13351 - Nguyễn Ngọc Nhi

Lecturer:

Prof. Tran Giang Son

Academic year: 2022 - 2025

Hanoi, March 2025

1 Introduction

Retinal blood vessel segmentation is an essential step in the diagnosis and treatment of eye diseases. Early detection of these conditions using automated segmentation methods can help doctors manage the diseases more effectively, improve patient care, and reduce the burden on healthcare systems.

In this report, we trained the R2U-Net model with images from the three different datasets and compared its performance with several other leading segmentation methods.

2 Dataset

This section presents the **DRIVE** (Digital Retinal Images for Vessel Extraction) dataset, a widely used benchmark for the development and evaluation of retinal vessel segmentation algorithms. It plays a crucial role in advancing research related to automated diagnosis of diabetic retinopathy and other retinal disorders.

The **DRIVE** dataset consists of 40 color fundus images, each with a resolution of 584 × 565 pixels. Every image includes a circular field of view (FOV) with a diameter of approximately 540 pixels, clearly defining the region of interest. The dataset is divided into two equal subsets: 20 images for training and 20 for testing.

Manual vessel annotations are provided by two human observers. For evaluation purposes, the segmentation by the first observer is used as the ground truth, while the second observer's annotation serves to measure inter-observer variability. All annotations were performed by trained ophthalmological professionals, ensuring high-quality reference masks.

The DRIVE dataset is specifically designed to support the development of automated vessel segmentation systems, and it remains one of the most frequently cited and utilized datasets in this field. It is publicly available and can be accessed at: https://drive.grand-challenge.org/.

3 Model

The following section describes the internal architecture of the R2-U-Net model, detailing how its layers, blocks, and connections are structured to enhance segmentation performance.

3.1 Model Overview

The Recurrent Residual U-Net (R2U-Net) is an enhanced version of the traditional U-Net architecture, specifically designed to improve feature extraction in medical image segmentation tasks. It is particularly well-suited for applications that demand precise boundary detection, such as retina blood vessel segmentation. R2U-Net introduces two key enhancements: recurrent connections and residual connections. The recurrent connections allow the model to iteratively refine spatial features over multiple time steps, effectively capturing more detailed contextual information. Meanwhile, the residual connections facilitate better gradient flow during training, enabling the construction of deeper networks without suffering from vanishing gradients. Together, these innovations lead to improved segmentation accuracy and more robust feature representation.

3.2 Model Architecture

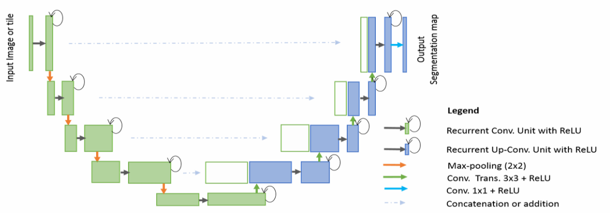

Figure 1: R2U-Net architecture

R2U-Net follows an encoder-decoder structure, extending the conventional U-Net by incorporating Recurrent Residual Convolutional Blocks (RRCNN) to enhance feature extraction and refinement. Residual blocks help in training deep networks by avoiding vanishing gradients, while recurrent connections are introduced in each convolutional block to allow iterative feature refinement.

The encoder path consists of repeated Recurrent Residual Convolutional Blocks (RRCNN), each followed by max-pooling layers. Each RRCNN block integrates Recurrent Convolutional Layers (RCL) and residual connections, progressively reducing spatial resolution while extracting high-level features for effective segmentation.

In the decoder path, transposed convolutions (upsampling layers) restore spatial resolution, while RRCNN blocks refine extracted features. Skip connections from the encoder path help preserve fine-grained spatial details essential for accurate segmentation.

The key innovation in R2-UNet is the **Recurrent Residual Convolutional Block (RRCNN)**, which improves feature representation and refinement through iterative processing. The **Recurrent Convolutional Layer (RCL)** applies sequential convolutions over discrete time steps, allowing feature evolution over iterations. The output at time step t is computed as:

$$O_{ijk}^{l}(t) = (\mathbf{w}_k^f)^T * \mathbf{x}_l^{f(i,j)}(t) + (\mathbf{w}_k^r)^T * \mathbf{x}_l^{r(i,j)}(t-1) + b_k$$

Where $\mathbf{w}_k^f$ and $\mathbf{w}_k^r$ are weights for the current and recurrent inputs. The output is then processed through a ReLU activation function:

$$\mathcal{F}(\mathbf{x}_l, \mathbf{w}_l) = f(O_{ijk}^l(t)) = \max(0, O_{ijk}^l(t))$$

Residual connections further enhance feature reuse and stabilize training. The final feature representation is computed as:

$$x_{l+1} = x_l + \mathcal{F}(x_l, \mathbf{w}_l)$$

By combining recurrent processing with residual learning, RRCNN blocks ensure efficient feature extraction and refinement, making R2-UNet highly effective for medical image segmentation.

4 Data Flow

This section describes how data flows through the R2U-Net model, detailing the transformation of input images as they pass through different processing stages to produce a segmented output. The data flow occurs in three main phases: encoding, bottleneck, and decoding.

4.1 Input

The R2U-Net model takes as input high-resolution images of retinal blood vessels, which are typically captured using advanced medical imaging devices such as fundus cameras. These cameras provide a detailed view of the back of the eye, highlighting critical anatomical structures like the retina, optic disc, and most importantly, the intricate network of blood vessels. Segmenting these vessels is crucial in medical practice, as it aids in the early detection and monitoring of a wide range of ophthalmic and systemic conditions, including diabetic retinopathy, glaucoma, hypertension, and cardiovascular diseases.

For this study, we utilize the **DRIVE** (Digital Retinal Images for Vessel Extraction) dataset, a widely adopted benchmark in retinal vessel segmentation research. The dataset consists of 40 high-resolution color fundus images, which are cropped during preprocessing to a final resolution of 565 × 565 pixels. Each image contains a circular field of view (FOV) with a diameter of approximately 540 pixels, defining the region of interest for vessel segmentation.

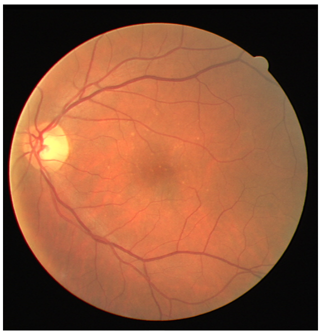

Figure 2: Retina Blood vessel

4.2 Encoding Phase

In the encoding phase, the input image is passed through a series of Recurrent Residual Convolutional Blocks (RRCNN). Each block consists of Recurrent Convolutional Layers (RCL) and residual connections, allowing the network to iteratively refine feature maps. This phase captures spatial dependencies and contextual information while progressively reducing spatial resolution using max-pooling layers. As the spatial dimensions decrease, the network focuses on extracting high-level semantic features, preparing the data for further processing in the next stage.

4.3 Bottleneck Phase

At the bottleneck phase, the model reaches its lowest spatial resolution, where features are highly abstract. An additional RRCNN block further processes these features, ensuring that both fine and coarse details are retained for segmentation. This phase acts as a bridge between feature extraction and reconstruction, compressing the most relevant information before it is passed to the decoder for upsampling.

4.4 Decoding phase

In the decoding phase, the compressed feature maps from the bottleneck are progressively upsampled using transposed convolutions, restoring spatial resolution. Each upsampling step is followed by an RRCNN block, which refines the extracted features. To recover the spatial details lost during encoding, skip connections from the encoder path are used, merging low-level spatial information with high-level semantic features. This fusion helps enhance segmentation accuracy, ensuring that fine structures in the image are well-preserved in the final output.

4.5 Output

The output of the R2U-Net model is a clean, segmented version of the original retinal image, where the blood vessels are clearly highlighted and separated from the background. This result comes in the form of a binary mask — essentially, each pixel is labeled to show whether it belongs to a blood vessel or not. Think of it as the model sketching out the entire vessel network, making it much easier to see and analyze. This automated segmentation saves a lot of time and effort compared to doing it manually and provides consistent, accurate results that are extremely helpful for medical diagnosis and research. It allows clinicians to quickly focus on the areas of interest and spot any abnormalities in the retinal vasculature with much greater ease.

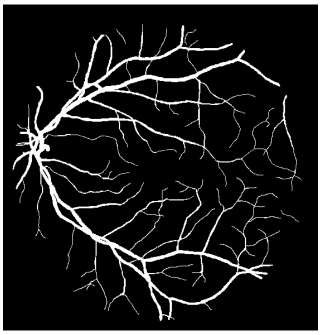

Figure 3: Segmentation of Retina Blood Vessel

5 Workflow

This section outlines the full implementation workflow of the R2U-Net model for retinal vessel segmentation, from preprocessing to model evaluation and visualization.

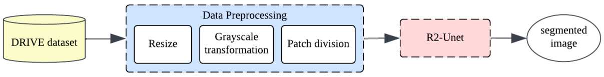

Figure 4: Work Flow

5.1 Preprocessing

The preprocessing stage involves three main steps to prepare the retinal images for model input. First, each image is centrally cropped from its original resolution of 584 × 565 pixels to 565 × 565 pixels, ensuring uniform input dimensions while preserving the important retinal regions. Second, the color fundus images are converted into grayscale, simplifying the data and reducing computational complexity without losing critical structural information. Lastly, the images are divided into smaller patches to facilitate localized learning and improve training efficiency, particularly when handling high-resolution data.

5.2 Dataset Splitting

The processed dataset is split into training and validation sets with a ratio of 90% for training and 10% for validation. This split allows the model to learn from the majority of the data while preserving a portion for evaluating generalization performance during training.

5.3 Model Initialization and Training Setup

The R2U-Net model is initialized with a deep encoder-decoder structure consisting of three **encoder** blocks, a single **bottleneck** block, and three **decoder** blocks. This architecture allows the model to extract hierarchical features while preserving spatial details through skip connections. The Binary Cross-Entropy (BCE) loss function is used as the training criterion, which is suitable for binary segmentation tasks. The Adam optimizer is employed to update the model's parameters efficiently.

The training process is configured to run for a maximum of 10 epochs. Early stopping is applied with a patience value of 5; if the validation loss does not improve for five consecutive epochs, training is terminated to prevent overfitting. During each epoch, key evaluation metrics—including accuracy, sensitivity (recall), and Dice coefficient—are computed and printed to monitor the model's performance.

5.4 Testing and Visualization

After training, the model is tested on a sample image from the test set to verify its functionality and ensure it produces meaningful segmentation results. This step serves as a preliminary check rather than a full evaluation of the model's performance. The segmented vessel mask is then visualized alongside the original input image, allowing for a qualitative inspection of the model's output.

6 Result

This section presents the results of the trained R2U-Net model on the validation set, highlighting its segmentation performance through key metrics such as validation loss, Dice score, accuracy, and sensitivity. Below is the table and the line chart which show the summarized results:

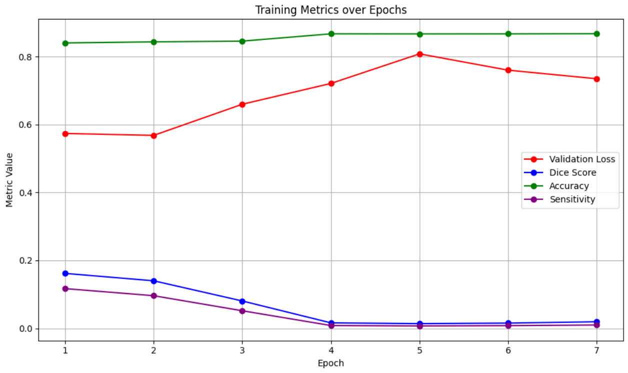

Figure 5: Training Metric over Epochs

Epoch	Val Loss	Dice Score	Accuracy	Sensitivity
1	0.5737	0.1617	84.03%	11.69%
2	0.5681	0.1399	84.34%	9.60%
3	0.6597	0.0805	84.56%	5.17%
4	0.7210	0.0161	86.71%	0.82%
5	0.8081	0.0138	86.69%	0.70%
6	0.7601	0.0154	86.70%	0.79%
7	0.7348	0.0193	86.74%	0.98%

Table 1: Validation Metrics Across Training Epochs

Here is an example of segmented result from the input image:

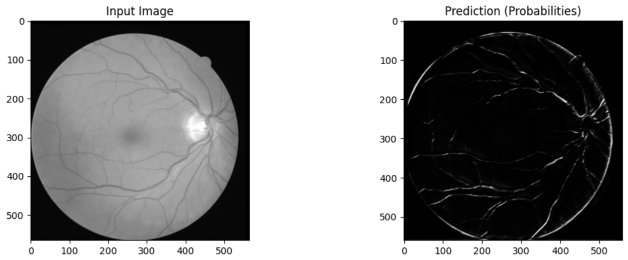

Figure 6: Result

From the results presented in Table 1, the model achieved its best validation performance at **epoch 2**, with the lowest validation loss of **0.5681** and a Dice score of **0.1399**. This suggests that the model effectively learned meaningful features at this stage. However, after epoch 2, the validation loss began to increase steadily, while the Dice score and sensitivity declined significantly.

By epoch 7, the validation loss had risen to **0.7348**, while the Dice score dropped to **0.0193**, indicating that the model was struggling to maintain its segmentation capability. Sensitivity also declined sharply, suggesting a reduced ability to correctly identify vessel structures. Despite this, accuracy showed a slight improvement, reaching **86.74%**, likely due to the model correctly classifying more background pixels rather than improving segmentation quality.

Given the lack of improvement in validation loss for **five consecutive epochs**, early stopping was triggered at **epoch 7** to prevent further overfitting. This suggests that while the model initially captured useful features, it failed to generalize well in later epochs, reinforcing the need for potential refinements in training strategies.

7 Future Work

Although the R2U-Net model performed reasonably well, there's still room for improvement. One key challenge was detecting small and thin blood vessels, as shown by the lower Dice and sensitivity scores. In the future, using attention mechanisms or multi-scale feature extraction could help the model focus better on these details.

It would also be beneficial to apply more diverse data augmentation techniques, such as elastic distortions or brightness changes, to make the model more adaptable. Training with larger datasets like STARE or CHASE_DB1, or using transfer learning, could also lead to better generalization.

Finally, adding post-processing steps—like morphological operations—to clean up the segmentation output, and exploring ways to integrate the model into real-world medical tools, could make it more useful in practical settings.

References

[1] Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A., & van Ginneken, B. (2004). *IEEE Transactions on Medical Imaging*, *23*(4), 501–509. https://doi.org/10.1109/tmi.2004.825627

[2] Grand Challenge. (n.d.). *DRIVE - Digital Retinal Images for Vessel Extraction*. Retrieved from https://drive.grand-challenge.org/

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint* https://arxiv.org/pdf/1505.04597

[4] Micromachines. (2021). *Retina U-Net: A Novel CNN for Biomedical Image Segmentation*. Retrieved from https://www.mdpi.com/2072-666X/12/12/1478

[5] Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv preprint* https://arxiv.org/pdf/1802.06955v5