BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

USING UNLABELED DATA FOR RETINAL SEGMENTATION

VYUŽITÍ NEOZNAČENAČENÝCH DAT PRO SEGMENTACI SÍTNICE

BACHELOR'S THESIS
BAKALÁŘSKÁ PRÁCE

AUTHOR
AUTOR PRÁCE

Andrii Shemshur

SUPERVISOR
VEDOUCÍ PRÁCE

Ing. Tomáš Vičar, Ph.D.

BRNO 2024

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

Bachelor's Thesis

Bachelor's study program **Biomedical Technology and Bioinformatics**

Department of Biomedical Engineering

Student: Andrii Shemshur ***ID:*** 240561

Year of study: 3 ***Academic year:*** 2023/24

TITLE OF THESIS:

Using unlabeled data for retinal segmentation

INSTRUCTION:

1) Through the literature, learn about methods for segmentation using convolutional neural networks and the possibilities of pre-training these networks on unlabeled data. 2) Create a simple network for segmenting blood vessels in retinal images. 3) Based on the literature search, select the most appropriate methods for retraining this network on unlabeled data and implement one of the methods. 4) Implement at least one other method. 5) Evaluate the methods and discuss the results.

RECOMMENDED LITERATURE:

[1] RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015. p. 234-241.

[2] RUBIN, Moran, Omer STEIN, Nir A. TURKO, et al. TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set. Medical Image Analysis [online]. 2019, 57, 176-185 [cit. 2023-09-04]. ISSN 13618415.

Date of project specification: 5.2.2024 ***Deadline for submission:*** 29.5.2024

Supervisor: Ing. Tomáš Vičar, Ph.D.

doc. Ing. Jana Kolářová, Ph.D.
Chair of study program board

ABSTRACT

This bachelor's thesis is concerned with the development and evaluation of advanced methods for medical image segmentation in the context of limited training data. The study examines supervised learning techniques employing Convolutional Neural Networks (CNNs), transfer learning with pre-trained models, and semi-supervised learning strategies.

A supervised convolutional neural network (CNN) model based on the U-Net architecture was employed as the baseline, achieving a Dice coefficient of 77.6% and an intersection over union (IoU) of 63.4%. The application of transfer learning using a ResNet34 encoder pre-trained on ImageNet led to a notable improvement in performance, with a Dice coefficient of 81.9%, an IoU of 69.3%, and an accuracy of 96.7%.

Furthermore, semi-supervised learning strategies, including pseudo-labeling and denoising pretraining, were employed to enhance the model's performance. The pseudo-labeling approach yielded a Dice coefficient of 81.7% and an IoU of 69.1%, thereby demonstrating the efficacy of leveraging unlabeled data. The denoising pretraining approach demonstrated robust performance, achieving a Dice coefficient of 80.3% and an IoU of 67.0%, even in the presence of noisy and unlabeled data.

These outcomes underscore the potential of transfer learning and semi-supervised methods to enhance segmentation accuracy in medical image analysis. They provide a robust foundation for future research in this field.

KEYWORDS

Medical Image Segmentation, Convolutional Neural Networks, Transfer Learning, Semi-supervised Learning, Pseudo-labeling, Denoising Pretraining, U-Net, ResNet34, Retinal Images

ABSTRAKT

Tato bakalářská práce se zabývá vývojem a hodnocením pokročilých metod pro segmentaci lékařských snímků v kontextu omezených trénovacích dat. Studie zkoumá techniky učení pod dohledem využívající konvoluční neuronové sítě (CNN), přenosové učení s předtrénovanými modely a strategie učení s částečným dohledem.

Jako základní model byl použit model konvoluční neuronové sítě (CNN) s dohledem založený na architektuře U-Net, který dosáhl koeficientu Dice 77,6% a průniku nad sjednocením (IoU) 63,4%. Použití přenosového učení pomocí kodéru ResNet34 předtrénovaného na síti ImageNet vedlo k výraznému zlepšení výkonu s koeficientem Dice 81,9%, IoU 69,3% a přesností 96,7%.

Kromě toho byly ke zvýšení výkonu modelu použity strategie učení s částečným dohledem, včetně pseudoznačení a předtrénování denoizace. Přístup pseudoznačení přinesl koeficient Dice 81,7% a IoU 69,1%, čímž prokázal účinnost využití neoznačených dat. Přístup před tréninkem denoizace prokázal robustní výkonnost a dosáhl koeficientu Dice 80,3% a IoU 67,0%, a to i v přítomnosti zašuměných a neoznačených dat.

Tyto výsledky podtrhují potenciál transferového učení a poloprovozních metod pro zvýšení přesnosti segmentace při analýze lékařských snímků. Poskytují solidní základ pro budoucí výzkum v této oblasti.

KLÍČOVÁ SLOVA

Segmentace lékařských snímků, konvoluční neuronové sítě, učení s přenosem, učení s částečným dohledem, pseudoznačení, předtrénování denoisingu, U-Net, ResNet34, snímky sítnice

ROZŠÍŘENÝ ABSTRAKT

Tato práce zkoumá potenciál konvolučních neuronových sítí (CNN) pro segmentaci lékařských snímků se zvláštním zaměřením na použití neoznačených dat. Přesná segmentace sítnicových cév má zásadní význam pro diagnostiku diabetické retinopatie (DR), která je hlavní příčinou slepoty. Nástup konvolučních neuronových sítí (CNN), zejména architektury U-Net, změnil tuto oblast a umožňuje přesnější a automatizovanější analýzu než tradiční metody strojového učení.

Úvodní fáze studie zahrnuje přehled vývoje CNN a jejich rostoucího významu v analýze lékařských obrazů. Jako základ slouží model U-Net, který je známý svou účinností a přesností při biomedicínské segmentaci. Architektura U-Net, zahrnující kontrakční a expanzní cesty, umožňuje extrakci a lokalizaci detailních rysů, což z ní činí účinný nástroj pro segmentaci obrazu sítnice.

Ústředním tématem této práce je zkoumání technik pro zvýšení výkonu CNN, zejména v kontextu omezených označených dat. To zahrnuje strategie učení s přenosem a částečně řízeného učení. Přenosové učení využívá kodér ResNet34 , který byl předem natrénován na datové sadě ImageNet, což výrazně zlepšuje segmentační metriky. Model dosáhl koeficientu Dice 81,9%, Intersection over Union (IoU) 69,3% a přesnosti 96,7%, což dokazuje výhody využití předtrénovaných modelů pro specializované lékařské úlohy.

Tato práce rozsáhle analyzuje strategie učení s polopřímým dohledem. Využití pseudoznačení, které zahrnuje předem vycvičený model generující štítky pro neoznačená data, přineslo koeficient Dice 81,7% a IoU 69,1%. Použití předtrénování denoizace, které zahrnovalo trénování modelu k odstranění šumu z obrázků před jemným doladěním pro segmentaci, přineslo koeficient Dice 80,3% a IoU 67,0%. Tyto metody ukazují potenciál využití neoznačených dat ke zvýšení robustnosti a přesnosti modelu.

Součástí práce je také komplexní analýza segmentace snímků sítnice z databáze High-Resolution Fundus Image Database (HRF), která obsahuje snímky s rozlišením 3504x2336 pixelů. Model U-Net prokázal schopnost rozlišovat mezi cévními a necévními oblastmi s průměrnou přesností 95,9%. Metriky přesnosti (precision) a odvolání (recall) byly 82,6%, resp. 90,8%, s koeficientem Dice 77% a IoU 62,7%. Tyto hodnoty potvrzují účinnost modelu při vytváření přesných segmentací, které se přesně shodují s manuálními anotacemi.

Integrace konvolučních neuronových sítí (CNN) do zobrazování sítnice představuje významný pokrok v lékařské diagnostice. Jejich schopnost zpracovávat složité obrazy a poskytovat přesné výsledky segmentace z nich činí účinný nástroj pro analýzu stavů, jako je diabetická retinopatie. Pokračující vývoj v této oblasti naznačuje, že konvoluční neuronové sítě (CNN) se mohou stát standardním nástrojem v lékařském zobrazování, který nabídne lepší diagnostické schopnosti a lepší

výsledky pro pacienty.

Závěrem tato práce poskytuje komplexní hodnocení metod založených na konvolučních neuronových sítích (CNN) pro segmentaci obrazu sítnice, přičemž se výrazně zaměřuje na využití neoznačených dat. Zejména strategie učení s přenosem a částečně řízeného učení prokázaly značný potenciál pro zvýšení segmentačního výkonu, což z nich činí cenné nástroje pro analýzu lékařských obrazů. Budoucí výzkum by se měl zaměřit na zdokonalení těchto metod, zkoumání hybridních modelů a optimalizaci parametrů učení s cílem dále zvýšit přesnost segmentace a použitelnost těchto metod v celé řadě úloh lékařského zobrazování.

SHEMSHUR, Andrii. *Using unlabeled data for retinal segmentation*. Bachelor's Thesis. Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Biomedical Engineering, 2024. Advised by prof. Ing. Tomáš Vičar, Ph.D.

Author's Declaration

Author:	Andrii Shemshur
Author's ID:	240561
Paper type:	Bachelor's Thesis
Academic year:	2023/24
Topic:	Using unlabeled data for retinal segmentation

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno ....

author's signature*

*The author signs only in the printed version.

ACKNOWLEDGEMENT

I would like to thank my thesis advisor, Ing. Tomáš Vičar, Ph.D. for his valuable comments and help with developing a theoretical model and experimental verification

Contents

List of Figures

List of Tables

Introduction

The evolution of medical image segmentation has been profoundly influenced by advancements in machine learning, mainly through the development of Convolutional Neural Networks (CNNs). A significant milestone in this field was the introduction of the U-Net architecture by Ronneberger, Fischer, and Brox[4], designed explicitly for biomedical image segmentation. This model, characterized by its unique architecture that includes a contracting path to capture context and a symmetric expanding approach for precise localization, has been widely used due to its efficiency in various medical imaging applications, including retinal image segmentation.

Diabetic Retinopathy (DR) is a primary application area for these advanced segmentation techniques. As a complication of diabetes, DR is a leading cause of vision impairment and blindness globally, particularly affecting the working-age population. Conventional methods for DR segmentation, primarily based on manual analysis, are not only time-consuming but also cause errors due to the complex nature of retinal images. The U-Net and similar CNN models offer an opportunity to automate and enhance the accuracy of this segmentation process, which is essential for early diagnosis and treatment.

However, the reliance on extensively labeled datasets poses a significant limitation in the broader application of these models. Creating such datasets is often resource-intensive and impractical, especially in medical fields where expert annotation is required. That's why this thesis explores the potential of using unlabeled data for retinal segmentation in DR. Inspired by studies like Rubin et al.'s work[5] on TOP-GAN for cancer cell classification using deep learning with minimal labeled data, one of these thesis goals is to analyze and expand methods for effectively training CNN models with limited or no labeled data.

The potential of using unlabeled data in deep learning could not only increase resource efficiency but also contribute to leveraging the vast amounts of unlabeled medical images that are currently underutilized. Studies have shown that models trained on large, diverse datasets, even if unlabeled, can achieve remarkable accuracy and generalization capabilities. This approach aligns with the current trend in deep learning, which emphasizes the importance of data quantity and diversity over extensive labeling.

This thesis expands the domain of medical image analysis, particularly retinal segmentation for DR. As a result, a novel, efficient approach to retinal segmentation that could significantly impact the early detection and management of DR was developed.

1 Literature Review

1.1 Theoretical Foundations of CNNs in Diabetic Retinopathy and Retinal Imaging

The evolution of Convolutional Neural Networks (CNNs) has had a profound impact on medical image analysis. Originating in the 1970s, CNNs experienced a resurgence with the rise of deep learning technologies. This resurgence was marked by their enhanced capability to process and interpret complexutilization, surpassing traditional machine learning methods. The significance of CNNs in medical imaging, particularly in the segmentation of retinal images, is huge. They offer an automated, efficient, and accurate approach to image analysis, which is crucial in diagnosing conditions like Diabetic Retinopathy (DR).[6]

DR, a direct consequence of vascular anomalies related to diabetes in the retina, is a leading cause of blindness. The segmentation of retinal blood vessels in fundus images is a critical step in diagnosing and treating DR. However, this process is challenging: variability in vessel size, low image contrast, and the presence of pathologies such as hemorrhages make segmentation a complex task. Traditional segmentation methods, including various machine learning and morphological approaches, have been used but with limitations in terms of accuracy and efficiency.[6]

Deep learning, particularly through CNNs, has revolutionized retinal vessel segmentation. These networks can autonomously learn from raw image data, making them well-suited for processing complex retinal images. Unlike traditional methods, which often involve manual feature coding, CNNs can automatically extract and process relevant features from retinal images. This capability significantly enhances the accuracy and efficiency of the segmentation process. Studies have shown that CNNs, when trained on large datasets, even outperforms existing algorithms in terms of classification accuracy and Area Under Curve (AUC) the Receiver Operating Characteristic (ROC).[7]

The advancements in CNN-based segmentation are evident in their superior performance on standard benchmarks such as the DRIVE, STARE, and CHASE databases.[6] These networks demonstrate not only high accuracy but also the ability to generalize across different datasets, a crucial factor in medical diagnostics. However, there are challenges, such as the risk of overfitting when training on limited data. Data augmentation techniques, including geometric transformations and contrast adjustments, are commonly used to minimize this risk. Another key point is the computational efficiency of CNNs. While the training process can be time-consuming, requiring substantial computational resources, the application of trained

models for new image segmentation is relatively fast, enhancing their practical use in a clinical area.[6]

The application of CNNs in retinal vessel segmentation has expanded the research and development area. The ability of these networks to detect fine structures like capillaries suggests their potential applicability in other areas of ophthalmological imaging. The increasing availability of medical imaging data combined with the necessity for accurate and efficient diagnostic tools highlights the importance of continued research and development in this area. The opportunities provided by CNNs in medical imaging, particularly with the ongoing advancements in deep learning, suggest a future where their full potential in enhancing diagnostic capabilities and patient care is fully realized.[6]

In conclusion, the integration of CNNs into the field of retinal imaging marks a significant advancement in medical diagnostics. Their ability to efficiently process complex images and provide accurate segmentation results makes them an outstanding technology in the analysis of conditions like DR. The ongoing developments in this field point to a future where CNNs could become a standard tool in medical imaging, offering enhanced diagnostic capabilities and improved patient outcomes.

1.2 Overview of Segmentation Techniques

1.2.1 Supervised Learning Techniques

In medical imaging, supervised learning is frequently employed for tasks that necessitate high accuracy and specificity, such as blood vessel or tumour segmentation. The models utilised, such as U-Net, possess an encoder-decoder architecture with skip connections, which enables efficient processing of medical images even when the training datasets are limited. These models are trained on annotated data, which ensures high accuracy in distinguishing between regions of interest. However, the primary limitation of this approach is its dependence on the quantity and quality of manual annotation, which can be resource-intensive. These methods are most suitable for image processing tasks where the accurate extraction of anatomical structures or pathological changes is required[6].

An illustrative case of supervised learning applied to retinal blood vessel segmentation is presented by Liskowski and Krawiec[**?**]. Their research utilized a deep neural network based on the U-Net architecture to segment blood vessels in retinal images from the DRIVE, STARE, and CHASE databases. The network was trained on a substantial number of examples (up to 400,000), which underwent preprocessing through global contrast normalization, zero-phase whitening, and were

augmented with geometric transformations and gamma corrections. The model notably surpassed previous algorithms in terms of the area under the ROC curve and classification accuracy. This case highlights the effectiveness of supervised learning techniques in achieving precise segmentation in medical imaging tasks.

- **Features:** Efficient training on small data.
- **Advantages:** High accuracy and specificity.
- **Disadvantages:** Dependent on high-quality annotation.
- **Applications:** Segmentation of complex structures, e.g. in oncology or cardiology.

1.2.2 Techniques Utilizing Unlabeled Data for Segmentation

Interactive Learning

The Interactive Learning approach combines machine learning with graph-based segmentation algorithms such as Graph Cuts. This allows the learning and segmentation process to be tailored in real time, taking into account the characteristics of a particular medical image. This interactivity is particularly valuable when dealing with abnormal images, where standard methods may produce erroneous results. However, the disadvantage is the need for active user participation, which can increase the time required to analyse the data[8].

An illustrative example of using interactive learning in medical image segmentation is presented by Wang et al.[8]. Their research introduced a deep learning-based interactive segmentation framework that integrates CNNs with a bounding box and scribble-based segmentation process. The suggested approach includes image-specific fine-tuning to adapt the CNN model to a particular test image. This fine-tuning can be performed either unsupervised or supervised, with additional user-provided scribbles. The framework was utilized for 2D segmentation of various organs from fetal MRI slices and 3D segmentation of brain tumors from different MR sequences. The findings demonstrated that the model greatly enhanced segmentation accuracy with fewer user interactions compared to conventional methods, establishing it as a robust and efficient solution for clinical applications.

- **Features:** Highly adaptable to image features.
- **Advantages:** Improved segmentation accuracy.
- **Disadvantages:** Requires user interaction.
- **Applications:** Use in complex clinical cases with non-standard pathology.

Weakly Supervised Learning

Weakly supervised learning methods reduce the reliance on fully annotated data by utilising incomplete or inaccurate labels. This is particularly pertinent in contexts where the cost or time required to fully annotate data is prohibitive. While such methods may be less accurate than fully supervised methods, they offer significant advantages in terms of reducing the time and cost of data preparation[9].

Multiple-Instance Learning (MIL) represents a type of weakly supervised learning. In one investigation, MIL was applied to the examination of retinal images for the detection of diabetic retinopathy[9]. MIL algorithms are capable of learning from images that are labeled solely with a diagnosis, eliminating the necessity for precise manual segmentations. This method identifies relevant patterns in the images automatically and utilizes them for overall classification, making it especially effective for large datasets where manual annotation is not feasible. The research showed that MIL not only streamlines the data preparation process but also enhances classification accuracy when compared to conventional single-instance learning techniques.

- **Features:** Use of incomplete labels.
- **Advantages:** Reduced annotation costs.
- **Disadvantages:** Potential reduction in accuracy.
- **Applications:** Processing large amounts of medical data, image preprocessing.

Transfer Learning

Transfer learning involves the application of models that have been pre-trained on large and diverse datasets to specific medical imaging tasks. The advantage of this approach is that it utilises generalised features, which can significantly enhance the accuracy of models on limited or specific medical data. However, additional task-specific model tuning may be required[10].

The application of transfer learning in medical image segmentation is exemplified by the study on retinal vessel segmentation using DeepLabv3+[10]. This study fine-tuned a pre-trained DeepLabv3+ model for the task of segmenting blood vessels in retinal images from the DRIVE dataset. The network was modified to accept single-channel images and perform two-class pixel-based classification (vessel and non-vessel). The segmented output images were then refined using a morphological closing operation. The results demonstrated that the proposed method achieved high accuracy, sensitivity, and specificity, outperforming many other methods in the field.

- **Features:** Use of pre-trained models.
- **Advantages:** Improved performance on specific tasks.

- **Disadvantages:** Need for model customisation.
- **Applications:** Improvement of existing models for specialised medical tasks.

Semi-supervised Learning

Semi-supervised learning is a distinctive approach that integrates the strengths of supervised and unsupervised learning. It utilises both annotated and unannotated data to train models, enabling the training of models on a greater volume of data. This is particularly advantageous in medical imaging, where annotation can be costly and time-consuming. Techniques such as the use of pseudo-labels can increase the number of training examples and improve the generalisability of the model without the need for additional annotation costs[11].

A semi-supervised framework based on U-Net was developed with the objective of reducing the workload associated with data annotation. The framework comprises three distinct stages:

1. Training the U-Net with enhanced ground truth labels.
2. Using the trained network to predict unlabeled data and taking the filtered prediction results as pseudo-labels.
3. Combining data amplification and dropout strategies to update the training set, iterating until the predetermined number of iterations is reached.

The framework demonstrated enhanced performance in comparison to fully supervised learning with an equivalent quantity of labelled data, thereby substantiating its efficacy in reducing the burden of data labelling while maintaining a high degree of segmentation accuracy[11].

- **Features:** Using a combination of annotated and unannotated data.
- **Advantages:** Expand the training dataset without additional cost.
- **Disadvantages:** There may be lower reliability of annotations due to the use of not fully validated data.
- **Applications:** Suitable for scenarios with limited annotation budget but large amount of raw data, e.g., pre-filtering large medical image sets.

Self-supervised Learning

Self-supervised learning is an approach that allows models to learn without standard annotations, using instead tasks that the model can perform on its own by creating its own labels. This can include tasks such as reconstruction or predicting the next state based on the current input. This method is particularly useful in medical imaging for pre-training models on large amounts of unlabelled data, allowing the extraction of rich data representations that can then be used for refinement with fewer annotated examples[12].

A self-supervised framework was developed using multimodal data to generate training labels automatically. In particular, vessel maps were generated from angiography images using straightforward image processing techniques. Subsequently, the generated vessel maps were employed to train a convolutional neural network (CNN) with the objective of segmenting vessels in retinal images. This method permitted the utilisation of a substantial quantity of unlabelled data, thereby markedly expanding the training dataset without the necessity for manual annotation. The results demonstrated that the self-supervised approach achieved competitive segmentation performance on public datasets such as DRIVE and STARE, thereby illustrating the potential of self-supervised learning to enhance segmentation accuracy in medical imaging[12].

- **Features:** Training without explicit annotations, using internally generated labels.
- **Advantages:** Reduced reliance on annotated data.
- **Disadvantages:** Requires careful selection of pre-tasks to ensure usefulness of extracted features.
- **Applications:** Pre-training models on large medical datasets before performing finer tuning on smaller annotated subsets.

Unsupervised Learning

In the field of medical imaging, unsupervised learning frequently employs techniques such as Generative Adversarial Networks (GANs) for the purpose of segmentation. These techniques are particularly advantageous in the identification of hitherto unknown patterns or anomalies in the data, without the necessity for labelled training data. GANs, in particular, are capable of generating high-resolution images and can be employed to enhance image quality and segmentation accuracy by learning the underlying distribution of the data.

The use of Generative Adversarial Networks (GANs) has been demonstrated to enhance the resolution of optical coherence tomography (OCT) images, thereby improving the accuracy of retinal layer segmentation. The GAN model generates high-resolution images from low-resolution inputs, which are then segmented using either traditional or deep learning-based segmentation methods. This approach not only improves the quality of the images but also significantly enhances the segmentation performance, providing clearer and more detailed structures for analysis.[13].

Another illustrative example is the utilisation of GANs for the classification of cancer cells free from stains. This technique involves training a GAN to generate high-fidelity synthetic images of cancer cells, which are then used to augment the training dataset for a classification model. The GAN-generated images facilitate the

model's learning of more effective features, even when there is a limited quantity of real, labelled data. This results in an improvement in the accuracy of cancer cell classification.[5].

- **Features:** Exploration of data without pre-labelling, using GANs for data enhancement and segmentation.
- **Advantages:** Ability to detect unknown or unexpected patterns, improve image resolution, and enhance segmentation accuracy.
- **Disadvantages:** Lack of control over the quality of the generated data, potential for misleading artifacts.
- **Applications:** Research tasks, pre-diagnosis, analysis of complex or poorly understood medical conditions, such as enhancing OCT images for better segmentation of retinal layers and generating synthetic training data for cancer cell classification.

2 Theoretical description of the algorithm components

2.1 CNN architecture

The field of biomedical research has witnessed a profound transformation in the past decade, largely due to the advent of deep learning, with convolutional neural networks (CNNs) playing a pivotal role. Among the plethora of CNN architectures, the U-Net model, designed with medical segmentation tasks in mind, has emerged as a particularly noteworthy example[14].

Main CNN architectures in medical segmentation

A plethora of CNN architectures have been employed in the field of medical segmentation, including specialized models such as SegNet and DeepLab. Each architecture has demonstrated the capacity to utilise distinctive image processing techniques. However, U-Net stands out for its ability to process high-resolution data with a relatively small amount of training data. This is achieved through a distinctive structural configuration that enables accurate localization and utilization of the context at all depth levels[14].

U-Net architecture

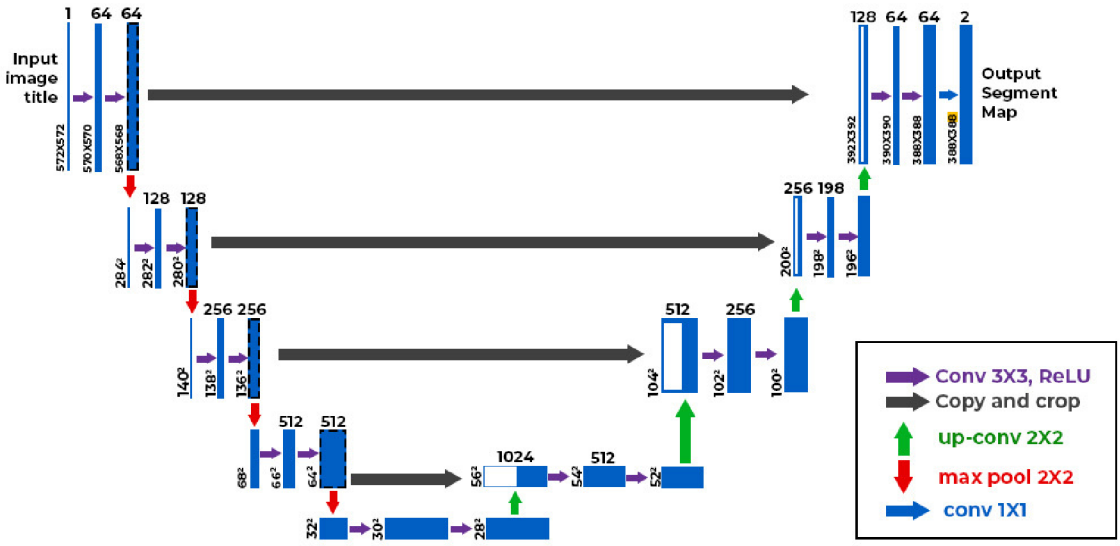


Fig. 2.1: U-Net Architecture. [1]

The U-Net convolutional neural network has been designed to handle small amounts of data and requires fewer training examples through the use of data augmentation. The U-Net architecture comprises two main parts: the compressive path (encoder) and the extensional path (decoder)[14]. As illustrated in Figure 2.1, the architecture effectively processes the input image through these paths to generate an accurate segmentation map.

The encoder is comprised of repeating modules, each of which includes two convolutions, an activation function (typically ReLU) and subsampling (max pooling). This allows the network to progressively reduce the spatial dimensions of the image and increase the depth of context.

Decoder: The U-Net expansion path is constituted by a sequence of layers that perform an upsampling operation on the previous feature maps. This involves the application of transpose convolution or up-convolution operations, which serve to recover the details and spatial dimensions of the image. A crucial element is the utilisation of skip connections, which transfer information from the compressive path directly to the corresponding layers of the expanding path. This ensures the preservation of context and the accuracy of localisation.

The final component of the architecture is a convolutional layer that classifies each pixel of the image, thereby generating a segmentation map.

Features:

- U-Net uses a skip-join mechanism to preserve the context of information at different depth levels, which is critical for accurate segmentation.
- Asymmetric structure with a deeper encoder and a shallower decoder, which optimizes both computational resources and segmentation quality.
- Efficient data utilisation due to augmentation, making U-Net particularly suitable for training on small datasets[14].

Advantages:

- Excellent generalisation ability, allowing the model to perform well even with limited data.
- High accuracy in segmentation tasks, thanks to accurate reconstruction of object locations.
- Flexibility in use, U-Net can be adapted for different medical segmentation tasks by changing only the input layers and loss functions.

Disadvantages:

- Difficulty in training at very large image sizes due to memory limitations.

- Need for careful tuning of the training process to achieve optimal results.
- Possibility of overtraining on overly detailed local features without proper regularisation or augmentation.

Applications:

- **Medical segmentation:** U-Net is widely used for segmentation tasks of various types of medical images, such as segmentation of tumours, blood vessels, and other tissues. It can also be used for other image-to-image applications, such as denoising and modality translation.
- **Different image modalities:** It effectively handles different types of medical images, including MRI, CT, and ultrasound images.

Conclusion

The U-Net model exhibits remarkable efficiency and versatility in medical segmentation tasks. It consistently achieves high levels of segmentation accuracy and quality under a range of conditions, making it a preferred choice for numerous clinical applications where the processing of medical images in an accurate and expedient manner is of paramount importance.

2.2 Data augmentation

Theoretical introduction to data augmentation

The augmentation of data is a crucial aspect of the training of convolutional neural networks, particularly in the context of medical imaging applications where data volumes may be limited. This process entails the generation of modified copies of the original images, thereby artificially expanding the training dataset. Techniques such as rotation, scaling, changing illumination, and the addition of noise facilitate the diversification of the data without the necessity of collecting additional images[15].

Types of data augmentation

The two main types of augmentation used in this study include pixel transformations and affine transformations:

- **Pixel Transformations :** These methods change the intensity and colours of pixels in an image. These methods include contrast correction, saturation, noise addition and blurring. These transformations help the model to better adapt to variations in the data that may occur due to differences in lighting or image quality[15].

- **Affine transformations:** Include rotating, reflecting, shifting, and rescaling images. These methods distort the image in a geometric sense, which allows the model to be trained to be invariant to the orientation and scale of objects in the images[15].

Importance of data augmentation

The application of data augmentation is of significant importance in improving the generalisability of the model and preventing overfitting, particularly when the available amount of data is limited. In the context of medical images, where each pixel may contain important diagnostic information, augmentation allows the model to better adapt to real-world clinical conditions. This is of particular importance in order to ensure diagnostic accuracy under varying imaging conditions and differences in patient characteristics[15].

Use of augmentation in U-Net

The U-Net architecture incorporates data augmentation techniques to enhance the accuracy of retinal vessel segmentation. These techniques, including contrast enhancement and noise addition, enable the model to more effectively recognise both thick and thin vessels, which is crucial for the accurate diagnosis of diseases such as diabetic retinopathy and glaucoma[15].

Empirical evidence demonstrates that data augmentation is an effective approach for significantly improving segmentation quality[15]. Studies have shown that the combined use of pixel and affinity augmentation methods can lead to notable improvements in segmentation performance. The combined use of pixel and affine transformations can achieve greater accuracy and model robustness, enabling the segmentation of medical images acquired under a wider range of conditions.

Key points of data augmentation in U-Net

Using U-Net for retinal vessel segmentation, data augmentation improves model training in several ways:

1. **Increased data diversity:** Augmentation allows the model to see images in different variations, which is crucial for training deep neural networks working with medical data.
2. **Improved generalisability:** The model becomes less sensitive to specific features of the training dataset, which reduces the likelihood of overtraining and improves its ability to deal with new, previously unseen images.

3. **Fine vessel segmentation:** Particularly important for conditions where retinal structure needs to be accurately determined. Data augmentation, especially methods that increase contrast and add noise, help in training the model to recognise thin and difficult to discern vessels

Conclusion

The application of data augmentation not only expands the size and diversity of the training set without the need for additional data collection costs, but also facilitates the training of more efficient and versatile models for medical segmentation. In the context of U-Net and retinal vascular segmentation, this is particularly valuable as it provides high diagnostic accuracy in a variety of clinical settings. The utilisation of combined augmentation methods, such as pixel and affine transformations, has been demonstrated to yield superior outcomes, enabling the model to adapt to the intricacies of medical images and enhancing its capacity to segment both thick and thin vessels. These attributes render data augmentation a pivotal tool in contemporary deep learning techniques, particularly in the domain of medical imaging, where the quality and accuracy of outcomes can be of paramount importance.

2.3 Loss Functions

Introduction to loss functions

The role of loss functions in the training of neural networks is of particular importance in the context of medical image segmentation tasks. The accuracy and ability of the model to discriminate between classes can have a significant impact on clinical outcomes. The measurement of the error between the predicted output and the actual labels allows the training process to be guided towards the minimisation of this error. The choice of an appropriate loss function is of critical importance in order to achieve high segmentation accuracy and specificity[16].

The main types of loss functions

In the context of medical image segmentation, the principal types of loss functions can be classified as follows:

1. **Distribution-based losses:** These functions, such as cross-entropy, measure the discrepancy between predicted probabilities and true labels based on information theory.

2. **Region-based losses:** Loss functions such as Dice loss evaluate the similarity or overlap between predicted and true segmented regions, which is particularly useful in medical segmentation.
3. **Boundary-based losses:** These functions, such as Euclidean or Hausdorff distance based functions, focus on minimising the distances between the boundaries of predicted and true labels.
4. **Compound losses:** Combinations of the previous types, such as combining Dice and cross-entropy, can account for several aspects of segmentation errors[16].

Importance of selecting a loss function

The selection of an appropriate loss function is of paramount importance in the context of managing problems such as unbalanced classes. In such cases, certain classes may be less represented in the data, which may result in an undertrained model. In order to address this issue, loss functions can be adapted or weighted to increase the importance of sparse classes during the training process. This approach helps to achieve a more balanced performance of the model in different classes[16].

Impact of loss functions on model performance

The utilisation of weighted loss functions has been demonstrated to markedly enhance the efficacy of segmentation in scenarios characterised by a high degree of class imbalance. In particular, loss functions that have been adapted to incorporate distance or predictive probability have exhibited superior performance in comparison to traditional loss functions, such as cross-entropy and Dice, in contexts where the accurate segmentation of sparse classes is a necessity[16].

Specific Loss Functions Used

In the practical part of this Bachelor's thesis, the following specific loss functions will be utilized: Dice Loss, Focal Loss, BCEWithLogits Loss, and MSE Loss. These functions are selected based on their suitability for various aspects of medical image segmentation and enhancement tasks.

Dice Loss: The Dice Loss function is particularly effective for medical image segmentation tasks, especially when the objective is to maximize the overlap between the predicted and actual segmented regions. Its effectiveness is highlighted by its ability to handle imbalanced class distributions commonly found in medical datasets.

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} g_i p_i}{\sum_{i=1}^{N} g_i + \sum_{i=1}^{N} p_i} \tag{2.1}$$

where g_i is the binary ground truth value (0 or 1) for pixel i, p_i is the predicted probability for pixel i, and N is the total number of pixels.

Focal Loss: This loss function is beneficial in addressing class imbalance by focusing on hard-to-classify examples, thus improving the model's performance on minority classes. It was chosen for its top ranking in binary-class segmentation tasks as highlighted in the referenced study[16].

$$L_{Focal} = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{2.2}$$

where α_t is a weighting factor for class t, p_t is the predicted probability for the class t, and γ is a focusing parameter that adjusts the rate at which easy examples are down-weighted.

BCEWithLogitsLoss: This function is preferred for scenarios involving pseudo-labeling, as it combines a sigmoid layer and the binary cross-entropy loss in a single class. It effectively bridges the gap between probability outputs and ground truth labels, facilitating better model performance in semi-supervised learning contexts.

$$L_{BCEWithLogits} = -\frac{1}{N} \sum_{i=1}^{N} [g_i \log(\sigma(p_i)) + (1 - g_i) \log(1 - \sigma(p_i))] \tag{2.3}$$

where $\sigma(p_i)$ is the sigmoid of the predicted probability for pixel i, g_i is the binary ground truth value for pixel i, and N is the total number of pixels.

MSELoss: This loss function is employed during the denoising pre-training phase, ensuring a cleaner dataset for subsequent training stages. It calculates the mean squared error between predicted and actual values, thus optimizing the model for noise reduction.

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^{N} (p_i - g_i)^2 \tag{2.4}$$

where p_i is the predicted value for pixel i, g_i is the ground truth value for pixel i, and N is the total number of pixels.

Conclusion

The selection of the loss function and its weighting strategy should be based on the specific task requirements and data characteristics. Experimental results demonstrate that the integration of adapted loss functions can significantly enhance segmentation quality, resulting in higher accuracy and reliability in medical applications. Therefore, the design and selection of suitable loss functions represents a pivotal task in machine learning, particularly in the context of medical imaging[16].

2.4 Performance measures

The evaluation of a model's performance in the context of medical image segmentation is based on the use of several key metrics. The following section provides a detailed description of each of these metrics.

Accuracy

The degree of accuracy is determined by the proportion of correct predictions and is calculated using the following formula:

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i), \tag{2.5}$$

where:

- y — true labels,
- $\hat{y}$ — predicted labels,
- n — number of samples in the data,
- $1(\hat{y}_i = y_i)$ — indicator function that is 1 when the predicted label $\hat{y}_i$ is equal to the true label y_i, and 0 otherwise.

Alternatively, accuracy can also be expressed in terms of true positives, true negatives, false positives, and false negatives using the following formula:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2.6}$$

where:

- TP — True Positives (correctly predicted positive cases),
- TN — True Negatives (correctly predicted negative cases),
- FP — False Positives (incorrectly predicted positive cases),
- FN — False Negatives (incorrectly predicted negative cases).

The metric of accuracy is of paramount importance in evaluating the overall performance of a model in medical image segmentation. This metric indicates the

proportion of correctly classified instances out of the total instances. High accuracy is essential for ensuring reliable predictions, which is particularly important in medical applications where incorrect predictions can lead to misdiagnosis or inappropriate treatment decisions, potentially having serious consequences for patient health.

Precision and Recall

Precision and recall are defined as follows:

$$\text{precision} = \frac{TP}{TP + FP}, \quad \text{recall} = \frac{TP}{TP + FN},$$

where:

- TP — True Positives (correctly predicted positive cases),
- FP — False Positives (incorrectly predicted positive cases),
- FN — False Negatives (incorrectly predicted negative cases).

It is evident that the metrics of precision and recall are crucial for the process of medical diagnosis. Precision is defined as the proportion of selected objects that are, in fact, relevant. Recall, in turn, indicates the proportion of actual relevant objects that were selected by the model.

Jaccard Index or IoU

The Jaccard index is a measure of the degree of overlap between the predicted and actual segmentation areas.

$$\text{J}(y, \hat{y}) = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|},$$

where:

- $|y \cap \hat{y}|$ — the number of elements in the intersection of the ground truth and predicted sets,
- $|y \cup \hat{y}|$ — the number of elements in the union of the ground truth and predicted sets.

This metric is of particular importance in segmentation tasks, as it directly reflects the quality of matching of segmented regions, whereas the absolute number of correctly classified background pixels is not considered. Consequently, this metric is more relevant than simple precision.

F1 score or Sørensen–Dice coefficient

F1-score represents the harmonic mean of precision and recall:

$$F_1 = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}},$$

where:

- precision — the ratio of the number of true positives to the sum of true positives and false positives,
- recall — the ratio of the number of true positives to the sum of true positives and false negatives.

The F1-score can also be expressed in terms of true positives (TP), false positives (FP), and false negatives (FN) as follows:

$$F_1 = 2\frac{TP}{2TP + FP + FN}.$$

Another equivalent form, known as the Sørensen–Dice coefficient, using set operations is:

$$F_1(y, \hat{y}) = \frac{2|y \cap \hat{y}|}{|y| + |\hat{y}|},$$

where:

- $|y|$ — the number of elements in the ground truth set,
- $|\hat{y}|$ — the number of elements in the predicted set,
- $|y \cap \hat{y}|$ — the number of elements in the intersection of the ground truth and predicted sets.

In the context of medical segmentation, this metric is of paramount importance as it considers both accuracy and completeness. The metric strikes a balance between detecting all relevant cases (high completeness) and minimising false positives (high accuracy). This is of particular importance in the context of medical applications, where the omission of pathology can have serious consequences for patient health.

3 Practical part

3.1 Supervised CNN Implementation

Practical Application of CNN Architecture The research model is based on the U-Net framework and includes standard components to ensure high efficiency and precision in segmentation tasks. The model architecture was adapted to manage the high-resolution and complex nature of fundus images effectively.

Implementation of U-Net Architecture: The U-Net model employed in this study incorporates standard components designed to enhance the extraction and segmentation of features.

- **Double Convolution Block (DoubleConv):** Each DoubleConv block consists of two consecutive convolution operations followed by batch normalization and ReLU activation. This configuration enhances feature extraction by intensifying the convolutional processing, which allows the model to capture more intricate details within the image data, crucial for accurate medical diagnostics.
- **Downscaling Block (Down):** This block integrates max pooling with a DoubleConv block to reduce the spatial dimensions of the input images while simultaneously deepening the feature maps. The reduction in dimensionality is essential for abstracting complex features at lower resolutions, facilitating a more comprehensive analysis of potential pathologies.
- **Upscaling Block (Up):** For spatial expansion, the model employs either bilinear upscaling or transposed convolution, coupled with a DoubleConv block. This arrangement is crucial for reconstructing higher resolution feature maps from the condensed feature representations, ensuring precise localization and delineation of segmentation targets.
- **Output Convolution Block (OutConv):** Features a single convolution layer with a kernel size of 1, designed to transform the deep feature representations into the final segmentation output, mapping directly to the required number of output classes. This block plays a critical role in delivering clear, actionable segmentation results.
- **U-Net Core (UNet):** The core of the model comprises an encoder (downscaling path) and a decoder (upscaling path), interconnected via a bottleneck. This setup utilizes the Down, Up, and OutConv blocks in a symmetric layout, ensuring that comprehensive feature capture and efficient segmentation are achieved across various image conditions. The model supports adjustable

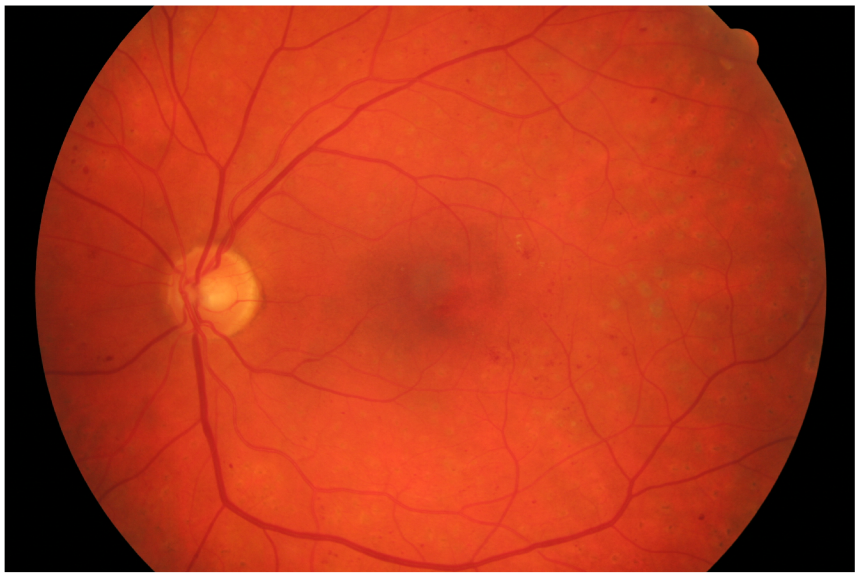
input channels and output classes, making it versatile across different segmentation challenges.

These components and configurations align with the standard U-Net architecture, ensuring robustness and accuracy in segmenting high-resolution fundus images.

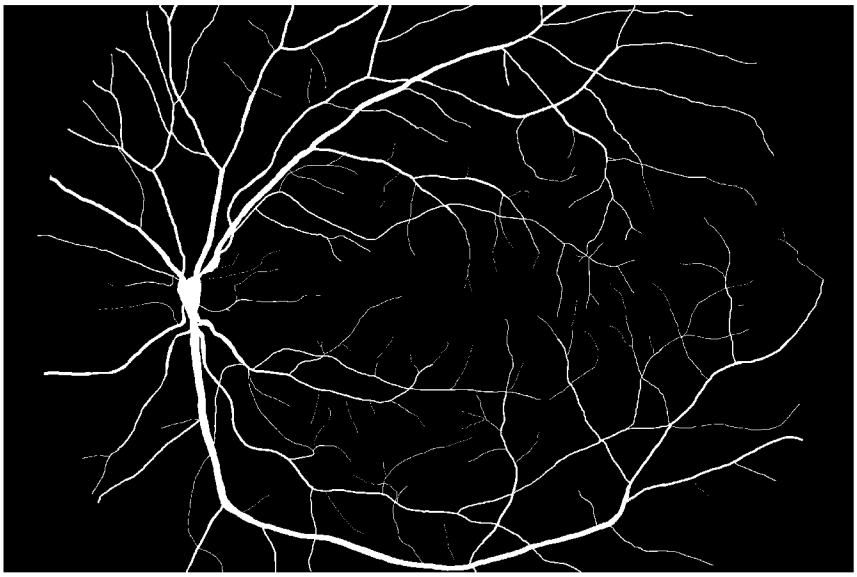
Materials

Dataset Description: The research employs the High-Resolution Fundus (HRF) Image Database[17] for the purpose of retinal segmentation. This database is selected specifically for its high-resolution images, measuring 3504x2336 pixels, which align with clinical standards. The dataset comprises 45 images in total, categorized into three groups: 15 images of healthy eyes, 15 images of eyes affected by diabetic retinopathy (DR), and 15 images of eyes with glaucoma. This varied collection supports a thorough comparative analysis and a robust assessment of the segmentation techniques used, highlighting the efficacy of each method across different pathological conditions. Figure 3.1 illustrates examples of the images along with their corresponding binary gold standards.

Significance of the Dataset: The HRF Image Database plays a crucial role in improving the precision of identifying key retinal structures, including the optic disc, macula, and blood vessels. It has been used in the past to evaluate other segmentation techniques, demonstrating its dependability and significance in medical image analysis. The high-resolution images facilitate advanced automated methods for extracting both vascular and non-vascular tissues, which are essential for training deep learning models. Additionally, the dataset offers a robust objective foundation for comparing automated segmentation outcomes with the benchmarks established by ophthalmology specialists.



(a)



(b)

Fig. 3.1: Examples of (a) images and (b) binary gold standards from the used dataset

Implementation Details of Data Augmentation

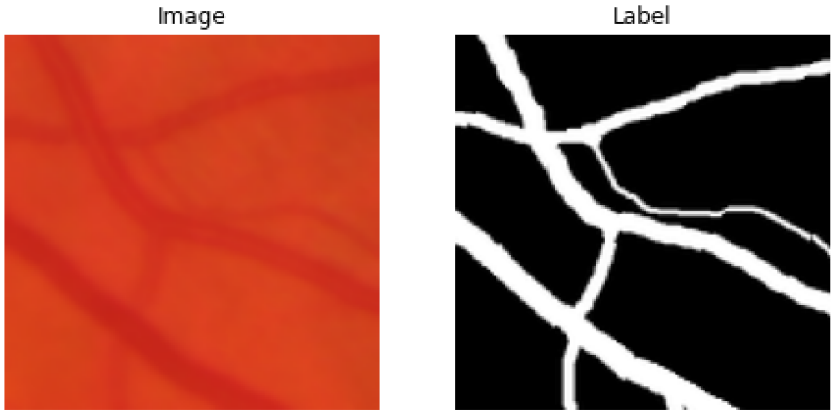
Data augmentation plays a pivotal role in enhancing the robustness and performance of the segmentation model, particularly given the complex nature of fundus images. The augmentation process is designed to introduce a realistic variability in the images, simulating conditions that the model will encounter in real-world medical settings.

Preprocessing and Augmentation Techniques: Initially, all images in the dataset undergo preprocessing using the *AspectRatioPreservingResize* class. This class adjusts each image to a target dimension of 1024 pixels while preserving the original aspect ratio. The resizing employs the high-quality LANCZOS algorithm[18], ensuring that clarity and detail are maintained, which is crucial for subsequent segmentation tasks.

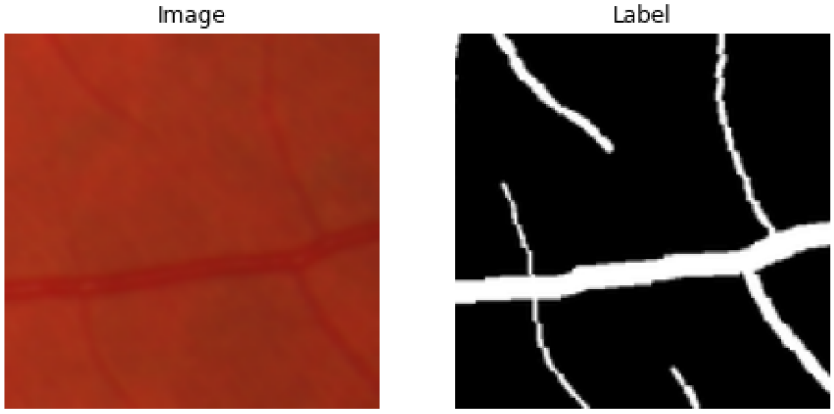
Following resizing, the *HRFDataset* class implements a robust augmentation strategy:

- **Random Cropping:** The dataset is expanded by extracting multiple 128x128 pixel segments from each image. This random cropping not only increases the number of training samples by a factor of ten but also ensures that the model encounters a wide array of image segments, enhancing its ability to generalize across different retinal conditions.
- **Geometric Transformations:** Images are subjected to random horizontal and vertical flips with a 20% probability. Additional affine transformations include rotations ($\pm$10 degrees), translations (up to $\pm$5% of the image size), scaling (between 90% and 110%), and shearing ($\pm$5 degrees). These transformations are critical for training the model to be invariant to the orientation and scale of anatomical features in the images.
- **Photometric Adjustments:** Adjustments to brightness and contrast are made, with levels varying between 90% and 110%. A Gaussian blur is applied using a 3x3 kernel with a 30% probability. These adjustments mimic variations in image quality due to different clinical imaging conditions.

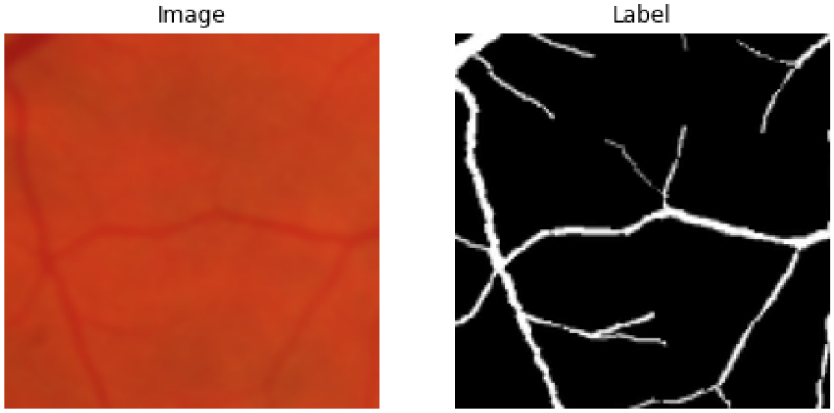
Impact of Augmentation on Model Training: Each augmentation technique is applied with specific probabilities, ensuring a balanced modification of the training dataset. This controlled variability prevents model overfitting and enhances its adaptability to diverse imaging scenarios encountered in clinical practice.



(a)



(b)



(c)

Fig. 3.2: Examples of images and binary gold standards cutouts with augmentation.

The results of these steps are illustrated in Figure 3.2, showing examples of images and binary gold standards cutouts with augmentation. This visualization helps in understanding the diversity and quality of the training samples created

through these augmentation techniques.

Practical Training and Validation

Training Setup and Parameters: The segmentation model leverages the ADAM optimizer[19], known for its efficacy in deep learning tasks. Training commenced with a learning rate of 0.01, adjusted over time through a decay schedule to enhance learning efficiency. Specifically, we used the following settings:

- **Optimizer:** ADAM with an initial learning rate of 0.01.
- **Learning rate scheduler:** MultiStepLR[20] with milestones at 20, 40, 60, and 80 epochs, and a decay factor (gamma) of 0.1.
- **Loss function:** Dice Loss 2.1 (binary mode) from the pytorch-toolbelt library[21].
- **Batch size:** 16.
- **Number of epochs:** 100.

The U-Net architecture was trained using these settings, balancing computational resources and effective learning progression.

Dataset Utilization and Augmentation: As described in Section 3.1, the High-Resolution Fundus (HRF) Image Database was systematically partitioned for training, validation, and testing. 70% of the images were used for training to expose the model to a broad spectrum of scenarios, enhancing its robust segmentation capability. The rest of the dataset was split evenly between validation and testing to thoroughly assess the model's performance.

Data augmentation, as detailed in Section 3.1, played a critical role in model training

Monitoring and Adjustments: Throughout the training process, which spanned 100 epochs, performance metrics were closely monitored. Adjustments were made responsively to optimize the training parameters, ensuring that the model's performance remained robust across all phases. Checkpoints were strategically used to save the best-performing models based on validation loss. The progression of training and validation loss over 100 epochs using the Dice Loss (2.1) function is illustrated in Figure 3.3.

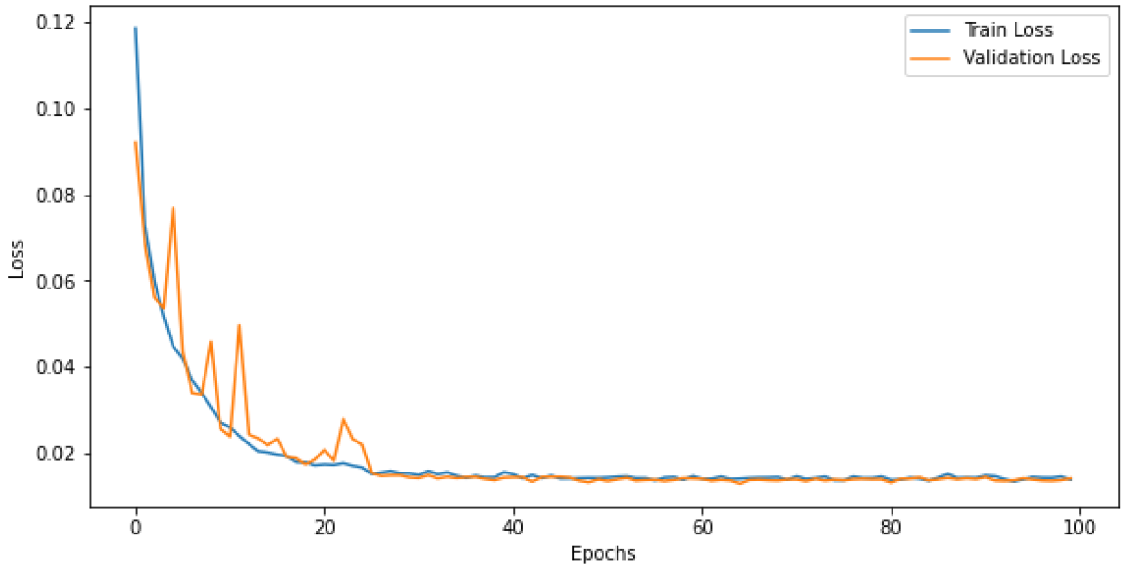


Fig. 3.3: Training and Validation Loss Progression of 100 epochs using the Dice Loss function.

Evaluation of Model Performance

The model's performance was rigorously evaluated upon completion of the 100th epoch. Metrics such as Dice coefficient, Intersection over Union (IoU), Accuracy, Recall, and Precision were computed to assess the quality of the segmentation results. The following table summarizes the obtained metrics:

Table 3.1: Performance metrics at the 100th epoch

Metric	**Dice**	**IoU**	**Accuracy**	**Recall**	**Precision**
Value	0.776	0.634	0.960	0.912	0.832

Additionally, a comparison was made between the results of models trained using Dice Loss (2.1) and Focal Loss (2.2) over 50 epochs. The results are summarized in the following table:

Table 3.2: Performance metrics using Dice Loss and Focal Loss at the 50th epoch

Loss Function	**Dice**	**IoU**	**Accuracy**	**Recall**	**Precision**
Dice Loss	0.738	0.585	0.953	0.910	0.828
Focal Loss	0.707	0.547	0.952	0.909	0.827

These metrics indicate a robust segmentation capability of the model, which demonstrates particularly high accuracy and recall. The Dice coefficient and IoU

values provide evidence of the model's efficiency in the overlap between the predicted and actual segmentation, essential for reliable medical image analysis.

Discussion: The high Dice coefficient and IoU values are particularly noteworthy as they directly reflect the model's ability to produce segmentations that closely match the ground truth. A Dice coefficient of 0.776 and an IoU of 0.634 (as shown in Table 3.1) indicate a high level of overlap between the predicted and actual segmented areas, which is crucial for accurate medical diagnoses and treatment planning.

Comparing the results using Dice Loss and Focal Loss, as shown in Table 3.2, we observe that the Dice Loss outperforms the Focal Loss across most metrics. Specifically, the Dice coefficient and IoU are higher when using Dice Loss (0.738 and 0.585, respectively) compared to Focal Loss (0.707 and 0.547, respectively).

In terms of Accuracy, Recall, and Precision, Dice Loss also shows slightly better performance, with Accuracy at 0.953, Recall at 0.910, and Precision at 0.828, compared to Focal Loss which achieved Accuracy at 0.952, Recall at 0.909, and Precision at 0.827. Although the differences are marginal, they further support the effectiveness of Dice Loss in providing a more reliable segmentation.

The higher Dice and IoU values obtained with Dice Loss underscore its superiority in ensuring that the predicted segmentation areas have a significant overlap with the actual areas. This is essential in medical applications, where precise segmentation of anatomical structures can significantly impact clinical decisions and patient outcomes. The results demonstrate that using Dice Loss leads to more accurate and reliable segmentation performance, making it a preferred choice for medical image analysis.

3.2 Transfer Learning Enhancement

Practical Application of Transfer Learning The segmentation model leverages transfer learning to enhance its performance by using a pre-trained encoder. Specifically, the ResNet34[22] architecture, pre-trained on the ImageNet dataset[23], was used to improve feature extraction capabilities, which is crucial for effective segmentation of medical images. Examples of images from the ImageNet dataset are shown in Figure 3.4.

Fig. 3.4: Examples of images in ImageNet dataset[2]

Modifications to the U-Net Architecture: The U-Net model was modified by replacing its encoder with the ResNet34 pretrained on the ImageNet dataset. This modification enhances the model's ability to extract features from high-resolution fundus images, improving the accuracy and efficiency of the segmentation tasks.

Training Setup and Parameters: The training setup and parameters were the same as described in Section 3.1.

Dataset Utilization and Augmentation: The same dataset utilization and augmentation techniques described in Section 3.1 were applied.

Monitoring and Adjustments: The monitoring and adjustment techniques were also consistent with those described in Section 3.1.

Comparison of Different Encoders An evaluation of multiple encoders was performed to identify the best architecture for segmentation tasks. The Table 3.3 presents the performance metrics for various encoder models. The metrics for each encoder were derived using Focal Loss, highlighting the advantage of Dice Loss when compared with the results in Table 3.4.

Table 3.3: Comparison of Different Encoder Models

Encoder	Dice	IoU	Accuracy	Recall	Precision
ResNet34	0.796	0.661	0.963	0.908	0.824
ResNet50	0.782	0.642	0.958	0.906	0.822
ResNet152	0.782	0.643	0.958	0.902	0.814

Evaluation of Model Performance The performance of the model was thoroughly assessed with the metrics outlined in Section 3.1. The table below presents the gathered metrics:

Table 3.4: Performance metrics at the 100th epoch with ResNet34 encoder

Metric	**Dice**	**IoU**	**Accuracy**	**Recall**	**Precision**
Value	0.819	0.693	0.967	0.921	0.849

Discussion: The results demonstrate that the pre-trained model on ImageNet with the ResNet34 encoder provides significant improvements over the previous setup described in Section 3.1. With a Dice coefficient of 0.819 and an IoU of 0.693, the ResNet34 encoder outperformed the custom U-Net model. Additionally, the model achieved a higher accuracy of 0.967 and a recall of 0.921, indicating enhanced capability in identifying relevant features within the images. The precision of 0.849 also suggests a robust reduction in false positives compared to the previous approach (see Table 3.4).

A comparison of different encoders, as shown in Table 3.3, further highlights the effectiveness of the ResNet34 encoder. It achieved higher Dice and IoU values compared to ResNet50 and ResNet152, demonstrating its superior performance in terms of segmentation accuracy and overlap with ground truth areas.

Relative to the supervised CNN approach outlined in Section 3.1, employing the pre-trained model on ImageNet with the ResNet34 encoder through transfer learning demonstrates significant improvements in all performance metrics. This underscores the utility of transfer learning in utilizing pretrained models to boost outcomes, particularly in medical image segmentation where there is a scarcity of labeled data.

When comparing the results obtained using Dice Loss and Focal Loss, as discussed earlier, the Dice coefficient and IoU are consistently higher with Dice Loss. This further emphasizes the advantage of using Dice Loss for medical image segmentation tasks, as it optimizes for overlap-based metrics which are crucial for accurate segmentation.

The findings emphasize the importance of adopting transfer learning to enhance accuracy, recall, and the quality of segmentation overall. The higher Dice and IoU values obtained with the pre-trained model on ImageNet with the ResNet34 encoder indicate that it is particularly effective in ensuring that the predicted segmentation areas have a significant overlap with the actual areas. This is essential in medical applications, where precise segmentation of anatomical structures can significantly impact clinical decisions and patient outcomes. Overall, the results demonstrate

that using the pre-trained model on ImageNet with the ResNet34 encoder leads to more accurate and reliable segmentation performance, making it a preferred choice for medical image analysis.

3.3 Semi-supervised Learning Strategy Based On Pseudo-labels

Practical Application of Semi-supervised Learning The model employs semi-supervised learning to enhance its performance by integrating both labeled and unlabeled data. This approach entails training the model on labeled data, generating pseudo-labels for unlabeled data, and then retraining the model using both labeled and pseudo-labeled data (as illustrated in Figure 3.5).

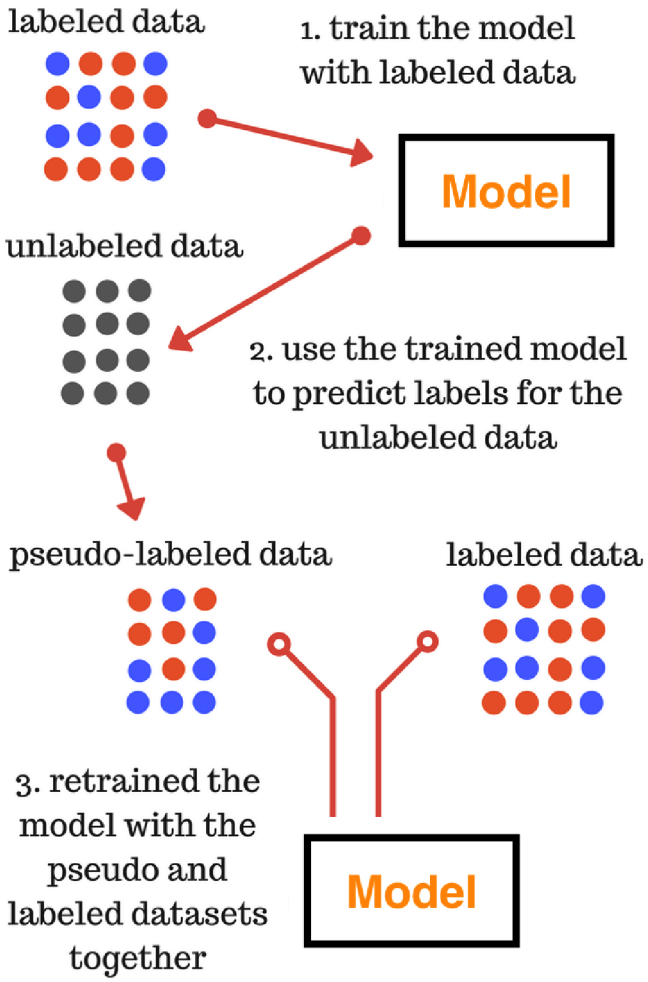


Fig. 3.5: Illustration of the semi-supervised learning process based on pseudo-labels[3]

Dataset for Pseudo-labeling: The Retinal Fundus Multi-Disease Image Dataset (RFMiD)[24] was chosen for the experimental application of unlabeled data due to its extensive collection of retinal images (example illustrated in Figure 3.6). The dataset includes 3,200 high-resolution images captured across diverse patient groups using various fundus cameras. It offers a broad spectrum of clinical scenarios, providing a solid basis for comparative analysis of various retinal diseases .

Fig. 3.6: Example of image from the RFMiD dataset

Pseudo-label Generation: To generate pseudo-labels, a pre-trained U-Net model from Section 3.1 was used. The following steps were taken:

1. **Patch Extraction:** Unlabeled retinal images were segmented into 128-pixel patches.
2. **Model Prediction:** Each patch was passed through the pre-trained model to predict the label.
3. **Reconstruction:** The predicted labels for each patch were reassembled to form pseudo-labeled images that match the original dimensions of the retinal images.

The generated pseudo-labels were saved for further processing (as shown in Figure 3.7 and 3.8).

Filtering Pseudo-labels: Since some pseudo-labeled images were suboptimal (see Figure 3.8), a filtering mechanism was implemented to ensure quality. The pseudo-labels were filtered based on the proportion of white pixels indicating confidence in the prediction. This process involved:

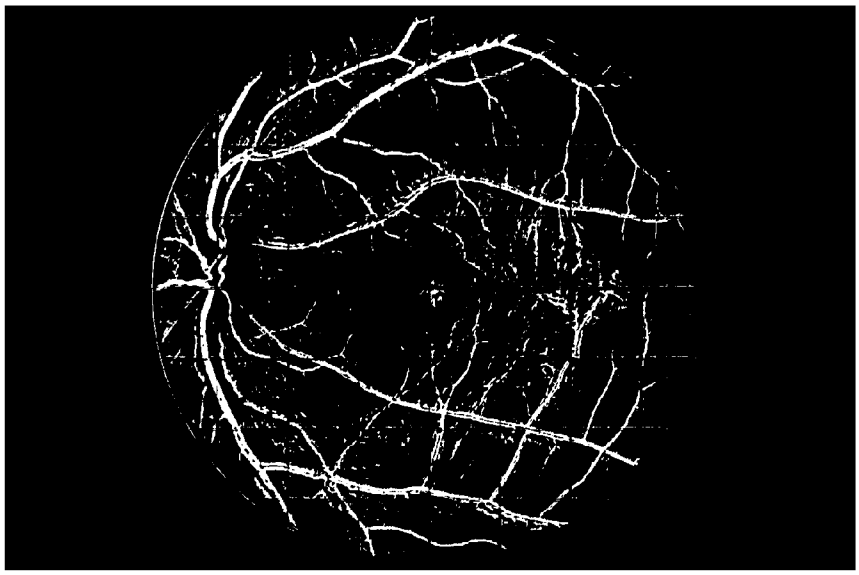


Fig. 3.7: Example of pseudo-label from filtered Pseudo-Labeled Dataset.

1. **White Pixel Count:** Counting the number of white pixels in each pseudo-label.
2. **Quality Threshold:** Only pseudo-labels with a white pixel proportion within a computed optimal range were retained.

As a result, a refined dataset of 803 high-quality pseudo-labeled images was obtained (as shown in Figure 3.7).

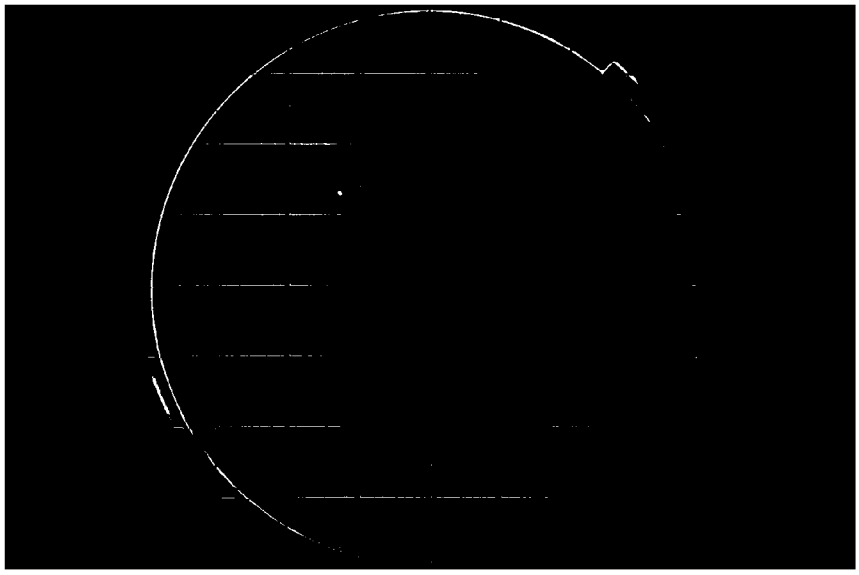


Fig. 3.8: Example of unsuccessful pseudo-label.

Retrain Stage: The retraining stage utilized a combined dataset of 310 real-labeled image crops and 803 pseudo-labeled image crops, totaling 1113. The same augmentation strategies used during the initial model training on real labels were applied to both datasets, ensuring consistency in the model's learning experience. Validation data consisted of 80 random crops from the real-labeled dataset, maintaining the integrity of the evaluation process. While most parameters were kept consistent with those used during the initial training phase, the loss function was changed to BCEWithLogitsLoss (2.3) to better handle the pseudo-labeled data. This approach to retraining, which utilizes a mixed dataset, is semi-supervised and designed to maximize the model's predictive performance by leveraging both verified and inferred label data.

Training Setup and Parameters: For the semi-supervised training with pseudo-labels, the model was trained using the following setup:

- **Optimizer:** ADAM with an initial learning rate of 0.001.
- **Learning rate scheduler:** MultiStepLR with milestones at 20, 40, and 60 epochs, and a decay factor (gamma) of 0.1.
- **Loss function:** BCEWithLogitsLoss (2.3).
- **Batch size:** 16.
- **Number of epochs:** 50.

Comparison of Ratios Between Real and Pseudo-labeled Crops: To further analyze the effect of different ratios of real to pseudo-labeled crops, additional retraining experiments were conducted for 20 epochs with varying ratios. The Table 3.5 summarizes the performance metrics for different ratios.

Table 3.5: Performance metrics for different ratios of real to pseudo-labeled crops

Ratio	Dice	IoU	Accuracy	Recall	Precision
0.5	0.817	0.691	0.959	0.902	0.814
0.6	0.789	0.652	0.963	0.912	0.831
0.75	0.780	0.639	0.964	0.922	0.850

Table 3.6: Performance metrics before and after retraining

Metric	Dice	IoU	Accuracy	Recall	Precision
Before Retraining	0.776	0.634	0.960	0.912	0.832
After Retraining	0.817	0.691	0.959	0.902	0.814

Results: The performance of the model before and after retraining on the combined dataset showed remarkable improvements. Initial training on a purely labeled dataset achieved an accuracy of 96%, a Dice coefficient of 77.6%, and an Intersection over Union (IoU) of 63.40% (see Table 3.6). After incorporating the combined dataset, the model showed improved performance metrics: accuracy slightly decreased to 95.90%, the Dice coefficient increased to 81.70%, and the IoU improved to 69.10%. Precision slightly decreased from 83.17% to 81.40%, while recall slightly decreased from 91.20% to 90.20%.

Discussion: The results indicate that the semi-supervised learning strategy using pseudo-labels has proven to be an effective method for improving the performance of the segmentation model. By leveraging unlabeled data, the model was able to better generalize and achieve higher accuracy and segmentation quality. This approach can be further refined by improving the quality of pseudo-labels and exploring different architectures and training parameters.

Comparing these results to the Transfer Learning Enhancement discussed in Section 3.2, the semi-supervised approach also shows competitive performance. The Transfer Learning model with ResNet34 and ImageNet achieved a Dice coefficient of 81.9% and an IoU of 69.3% (see Table 3.4), which are slightly higher than those obtained with semi-supervised learning. However, the semi-supervised approach still demonstrates substantial improvements over the purely supervised model, making it a valuable method in scenarios where additional unlabeled data is available.

Retraining experiments showed that different ratios of real to pseudo-labeled crops affected the model's performance. As demonstrated in Table 3.5, a higher ratio of real to pseudo-labeled data generally led to better performance. Specifically, the 0.5 ratio provided the best balance, resulting in a Dice coefficient of 81.7% and an IoU of 69.1%. This suggests that a balanced dataset with a sufficient number of high-quality pseudo-labels is crucial for optimal results. Further refinement and experimentation with different ratios and quality thresholds can help in achieving even better performance.

Overall, the findings emphasize the importance of adopting semi-supervised learning to enhance accuracy, recall, and the quality of segmentation, particularly in medical image analysis where labeled data is often scarce. The comparison with both the supervised CNN approach and the Transfer Learning Enhancement highlights the potential of semi-supervised methods to bridge the gap between limited labeled data and high-performance segmentation models.

3.4 Semi-supervised Denoising Pretraining

Practical Application of Semi-supervised Denoising Pretraining The model training process included a pre-training stage using denoising tasks to improve the extraction of features before the final segmentation task. This approach leverages a semi-supervised learning strategy to enhance the model's performance in handling noisy and unlabeled data, which is common in medical imaging.

Dataset for Denoising: The same dataset described in Section 3.3 was used. Synthetic noise was added to these images to simulate various real-world imaging conditions. The noise types included Gaussian, salt-and-pepper, and speckle noise. This preparation helps the model learn to identify and mitigate noise during the pre-training phase.

Implementation Details of Pretraining: The pretraining was implemented using a custom `DenoisingDataset` class, which generates noisy images and their corresponding clean targets. The noise types used were:

- **Gaussian Noise:** Additive Gaussian noise with a standard deviation scaled to the image intensity range.
- **Salt-and-Pepper Noise:** Randomly set pixels to 0 (salt) or 1 (pepper) with a probability proportional to the noise level.
- **Speckle Noise:** Multiplicative noise where random values are added to pixel intensities.

These noisy images were used to train the model to predict the clean image from the noisy input, thereby improving its robustness to noise.

Training Setup and Parameters: For the denoising pretraining, the model was trained using the following setup:

- **Optimizer:** ADAM with an initial learning rate of 0.01.
- **Learning rate scheduler:** MultiStepLR with milestones at 20, 40, 60, and 80 epochs, and a decay factor (gamma) of 0.1.
- **Loss function:** Mean Squared Error (MSE) Loss (2.4) for the pretraining phase.
- **Batch size:** 16.
- **Number of epochs:** 100.

The model architecture used in this phase was based on the U-Net framework, modified to handle the denoising task.

Dataset Utilization and Augmentation: The dataset was divided into two distinct sets: a training set and a validation set. The ratio of the two sets was 80/20. No additional data augmentation techniques were employed during the denoising pretraining phase.

Monitoring and Adjustments: The performance of the model during the pretraining phase was monitored using loss values in both training and validation sets. Checkpoints were saved based on the best validation loss to ensure optimal weights for the subsequent fine-tuning phase.

Fine-tuning for Segmentation

After the denoising pretraining, the model was fine-tuned for the segmentation task using the same U-Net architecture.

Training Setup and Parameters: For the fine-tuning phase, the following parameters were used:

- **Optimizer:** ADAM with an initial learning rate of 0.01.
- **Learning rate scheduler:** MultiStepLR with milestones at 20, 40, 60, and 80 epochs, and a decay factor (gamma) of 0.1.
- **Loss function:** Dice Loss (binary mode) from the `pytorch-toolbelt` library.
- **Batch size:** 16.
- **Number of epochs:** 100.

Dataset Utilization and Augmentation: The same dataset utilization and augmentation techniques described in Section 3.1 were applied during the fine-tuning phase.

Monitoring and Adjustments: The monitoring and adjustment techniques were also consistent with those described in Section 3.1.

Evaluation of Model Performance

The fine-tuned model was evaluated on the test set after 100 epochs. The following table summarizes the metrics obtained:

Table 3.7: Performance metrics at the 100th epoch after fine-tuning

Metric	Dice	IoU	Accuracy	Recall	Precision
Value	0.803	0.670	0.961	0.907	0.823

Discussion: Comparing results to the supervised CNN approach outlined in Section 3.1, the denoising pre-training approach shows comparable, if not slightly improved, performance. The supervised CNN approach achieved a Dice coefficient of 77.6% and an IoU of 63.4%(as shown in Table 3.1), whereas the denoising pretraining approach achieved a Dice coefficient of 80.3% and an IoU of 67.0%(as shown in Table 3.7). This demonstrates the effectiveness of incorporating semi-supervised denoising pretraining to enhance the model's robustness and segmentation performance.

The high accuracy of 96.1% and recall of 90.7% further highlight the model's ability to accurately identify relevant features within the images. While the precision value of 82.3% suggests that there is still room for improvement in reducing false positives, the overall metrics indicate a well-balanced performance.

Overall, the semi-supervised denoising pretraining strategy has proven beneficial in enhancing the model's robustness and performance in medical image segmentation tasks, making it a valuable approach in scenarios where labeled data is scarce and the quality of segmentation is crucial for clinical decision-making. The comparison with the supervised CNN approach underscores the potential of this strategy to achieve high-performance segmentation results.

4 Evaluation and Results

In this chapter, the results of various practical approaches for medical image segmentation, as implemented and tested in the previous chapters, are compared and analyzed. The performance metrics of each method are discussed in detail to highlight their strengths and weaknesses, providing a comprehensive evaluation of their effectiveness in the context of retinal image segmentation.

4.1 Comparison of Approaches

To facilitate a clear comparison, the performance metrics of all approaches are summarized in Table 4.1.

Table 4.1: Comparison of Performance Metrics for Different Approaches

Approach	**Dice**	**IoU**	**Accuracy**	**Recall**	**Precision**
Supervised CNN	77.6%	63.4%	96.0%	91.2%	83.2%
Transfer Learning (ImageNet + ResNet34)	**81.9%**	**69.3%**	**96.7%**	**92.1%**	**84.9%**
Semi-supervised Learning (Pseudo-labels)	81.7%	69.1%	95.9%	90.2%	81.4%
Denoising Pretraining	80.3%	67.0%	96.1%	90.7%	82.3%

4.2 Discussion

Supervised CNN vs. Transfer Learning Enhancement: The transfer learning approach using the ResNet34 encoder pre-trained on ImageNet showed significant improvements over the supervised CNN approach. The **Dice coefficient** and **IoU** values were higher, indicating better overlap and segmentation quality. The higher **accuracy** and **recall** further support the superior performance of the transfer learning approach in identifying relevant features within the images.

Supervised CNN vs. Semi-supervised Learning Based on Pseudo-labels: The semi-supervised learning approach using pseudo-labels demonstrated notable improvements in **Dice coefficient** and **IoU** after retraining. While the accuracy slightly decreased, the overall segmentation quality improved, highlighting the effectiveness of leveraging unlabeled data to enhance model performance. The semi-supervised approach also showed competitive performance compared to the transfer

learning enhancement, making it a valuable method in scenarios with additional unlabeled data.

Supervised CNN vs. Semi-supervised Denoising Pretraining: The semi-supervised denoising pretraining approach achieved comparable performance to the supervised CNN approach, with slightly improved **Dice coefficient** and **IoU** values. This strategy proved beneficial in enhancing the model's robustness and segmentation performance, particularly in handling noisy and unlabeled data.

Overall Comparison: Among the approaches tested, the transfer learning enhancement with the ResNet34 encoder pre-trained on ImageNet achieved the best overall performance. The semi-supervised learning strategies, both based on pseudo-labels and denoising pretraining, also demonstrated substantial improvements over the purely supervised CNN approach. These findings underscore the potential of transfer learning and semi-supervised methods to enhance the **accuracy**, **recall**, and quality of segmentation in medical image analysis.

Conclusion

The comprehensive evaluation and comparison of various approaches for medical image segmentation in this thesis have revealed significant advancements and practical implementations in the field of retinal image analysis. The study investigated the efficacy of supervised convolutional neural networks (CNNs), transfer learning, and semi-supervised learning strategies, which collectively enhance the accuracy and efficiency of medical image segmentation.

The supervised CNN approach, constructed upon the robust U-Net architecture, served as a baseline, exhibiting substantial segmentation capabilities with high accuracy and recall. However, the necessity for extensive labeled datasets places limitations on the scalability of the approach in environments with scarce annotated data.

Transfer learning, which employs a pre-trained ResNet34 encoder trained on ImageNet, emerged as the most effective method. This approach demonstrated a notable enhancement in performance metrics across all evaluated categories, thereby substantiating the efficacy of leveraging pre-trained models to enhance feature extraction and segmentation accuracy. The efficacy of this approach underscores the significance of employing generalized features derived from expansive, heterogeneous datasets to address specific medical imaging tasks.

The implementation of semi-supervised learning, through the use of pseudo-labeling and denoising pretraining, also demonstrated promising results. These strategies effectively incorporated unlabeled data, thereby expanding the training dataset and improving model generalization. The pseudo-labeling method demonstrated particularly notable improvements in the Dice coefficient and IoU, indicating enhanced segmentation quality even in the context of a reduction in the dependence on labeled data. The denoising pretraining strategy additionally enhanced the model's robustness, particularly in the context of handling noisy and variable imaging conditions that are commonly encountered in clinical practice.

In conclusion, this thesis provides a comprehensive analysis of different segmentation techniques, demonstrating the potential of advanced methodologies to overcome the limitations of traditional supervised learning. The transfer learning approach, which has demonstrated superior performance, is particularly recommended for medical image segmentation tasks. Furthermore, semi-supervised learning strategies, which are capable of leveraging unlabeled data, also present valuable alternatives, especially in data-scarce environments.

Future research should concentrate on the refinement of these advanced techniques, the exploration of hybrid models that combine the strengths of multiple approaches, and the optimization of training parameters with the objective of further

enhancing segmentation performance. Furthermore, extending the scope to encompass other medical imaging modalities and conditions can facilitate the generation of more comprehensive insights and the development of more versatile applications. Ultimately, this will contribute to the creation of more effective diagnostic tools and to enhanced patient outcomes.

The findings of this thesis demonstrate the necessity of adopting innovative and advanced learning strategies in medical image segmentation, thereby paving the way for future developments and improvements in clinical diagnostics.

Bibliography

[1] U-net architecture. https://media.geeksforgeeks.org/wp-content/uploads/20220614121231/Group14.jpg, 2022. Available at: https://www.geeksforgeeks.org/u-net-architecture-explained/.

[2] Bhagyasri Aggarwal, Arjun Acharya, and Aarav Laghari. A survey on image datasets for computer vision and deep learning convolutional neural network tasks, 10 2020.

[3] Gregor Gunčar, Matjaž Kukar, Tim Smole, Sašo Moškon, Tomaž Vovko, Simon Podnar, Peter Černelč, Miran Brvar, Mateja Notar, Manca Köster, Marjeta Jelenc, and Marko Notar. Differentiating viral and bacterial infections: A machine learning model based on routine blood test values, 05 2023.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[5] Moran Rubin, Omer Stein, Nir A. Turko, Yoav Nygate, Darina Roitshtain, Lidor Karako, Itay Barnea, Raja Giryes, and Natan T. Shaked. Top-gan: Stain-free cancer cell classification using deep learning with a small training set. *Medical Image Analysis*, 57:176–185, 2019.

[6] Paweł Liskowski and Krzysztof Krawiec. Segmenting retinal blood vessels with deep neural networks. *IEEE Transactions on Medical Imaging*, 35(11):2369–2380, 2016.

[7] Martina Melinscak, Pavle Prentasic, and Sven Loncaric. Retinal vessel segmentation using deep neural networks. In *VISAPP (1)*, pages 577–582, 2015.

[8] Guotai Wang, Wenqi Li, Maria A. Zuluaga, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Tom Doel, Anna L. David, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Transactions on Medical Imaging*, 37(7):1562–1573, 2018.

[9] Gwenolé Quellec, Guy Cazuguel, Béatrice Cochener, and Mathieu Lamard. Multiple-instance learning for medical image and video analysis. *IEEE Reviews in Biomedical Engineering*, 10:213–234, 2017.

[10] Michael Chi Seng Tang, Soo Siang Teoh, and Haidi Ibrahim. Retinal vessel segmentation from fundus images using deeplabv3+. In *2022 IEEE 18th International Colloquium on Signal Processing Applications (CSPA)*, pages 377–381, 2022.

[11] Yaning Li, Zijun Pei, Jiaguang Li, and Dali Chen. Semi-supervised learning framework in segmentation of retinal blood vessel based on u-net. In *2021 33rd Chinese Control and Decision Conference (CCDC)*, pages 5972–5978, 2021.

[12] Álvaro S. Hervella, José Rouco, Jorge Novo, and Marcos Ortega. Self-supervised deep learning for retinal vessel segmentation using automatically generated labels from multimodal data. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2019.

[13] Paria Jeihouni, Omid Dehzangi, Annahita Amireskandari, Ali Rezai, and Nasser M. Nasrabadi. Gan-based super-resolution and segmentation of retinal layers in optical coherence tomography scans. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 46–50, 2021.

[14] Jasjit S. Suri, Mrinalini Bhagawati, Sushant Agarwal, Sudip Paul, Amit Pandey, Suneet K. Gupta, Luca Saba, Kosmas I. Paraskevas, Narendra N. Khanna, John R. Laird, Amer M. Johri, Manudeep K. Kalra, Mostafa M. Fouda, Mostafa Fatemi, and Subbaram Naidu. Unet deep learning architecture for segmentation of vascular and non-vascular images: A microscopic look at unet components buffered with pruning, explainable artificial intelligence, and bias. *IEEE Access*, 11:595–645, 2023.

[15] Sabri Deari, İlkay Öksüz, and Sezer Ulukaya. Importance of data augmentation and transfer learning on retinal vessel segmentation. In *2021 29th Telecommunications Forum (TELFOR)*, pages 1–4, 2021.

[16] Takaaki Sugino, Toshihiro Kawase, Shinya Onogi, Taichi Kin, Nobuhito Saito, and Yoshikazu Nakajima. Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks. *Healthcare (Basel)*, 9(8), July 2021.

[17] Attila Budai and Jan Odstrcilik. High resolution fundus (hrf) image database, 2013.

[18] Cornelius Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Natl. Bur. Stand. B*, 45:255–282, 1950.

[19] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

[20] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[21] Eugene Khvedchenya. Pytorch toolbelt. `https://github.com/BloodAxe/pytorch-toolbelt`, 2019.

[22] Pavel Iakubovskii. Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch`, 2019.

[23] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[24] Samiksha Pachade, Prasanna Porwal, Dhanshree Thulkar, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, Luca Giancardo, Gwenolé Quellec, and Fabrice Mériaudeau. Retinal fundus multi-disease image dataset (rfmid): A dataset for multi-disease detection research. *Data*, 6(2), 2021.

[25] Dong-Hyun Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*, 07 2013.

[26] Meng Cai, Wei-Qiang Zhang, and Jia Liu. Improving deep neural network acoustic models using unlabeled data. In *2013 IEEE China Summit and International Conference on Signal and Information Processing*, pages 137–141, 2013.

[27] Chenguang Zhang, Tianjiao Yang, and Yan Zhang. Learning medical image segmentation via u-net based positive and unlabeled learning method. In *2021 IEEE 7th International Conference on Big Data Intelligence and Computing (DataCom)*, pages 1–6, 2021.

[28] V. Sujatha, Bommasamudram Sri Anitha, Gullipalli Tulasi Rama, Nelluri Niharika, and Anchula Sahithi. Convolutional neural network (cnn) based blood

vessel segmentation from ocular images. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 518–523, 2023.

[29] T. M. Sheeba, S. Albert Antony Raj, and M. Anand. Analysis of various image segmentation techniques on retinal oct images. In *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pages 716–721, 2023.

[30] Bambang Krismono Triwijoyo, Ahmat Adil, Dian Syafitri Chani Saputri, and Anthony Anggrawan. Morphology approach for segmentation of blood vessels in retinal images. In *2023 6th International Conference of Computer and Informatics Engineering (IC2IE)*, pages 346–350, 2023.