Image Segmentation Using Deep Learning

by

Nasrin Akbari

A thesis submitted in fulfillment of the requirements
for the degree of Master of Applied Science

in the Department of Electrical and Computer Engineering

University of Victoria

Image Segmentation Using Deep Learning

by

Nasrin Akbari

B.Sc., K. N. Toosi University of Technology, 2016
M.Sc., University of Tehran, 2018

Supervisory Committee

Dr. Amirali Baniasadi , Co-Supervisor
(Department of Electrical and Computer Engineering)

Dr. Ibrahim Numanagić, Co-Supervisor
(Department of Computer Science)

Abstract

The image segmentation task divides an image into regions of similar pixels based on brightness, color, and texture, in which every pixel in the image is assigned to a label. Segmentation is vital in numerous medical imaging applications, such as quantifying the size of tissues, the localization of diseases, treatment planning, and surgery guidance. This thesis focuses on two medical image segmentation tasks: retinal vessel segmentation in fundus images and brain segmentation in 3D MRI images. Finally, we introduce LEON, a lightweight neural network for edge detection.

The first part of this thesis proposes a lightweight neural network for retinal blood vessel segmentation. Our model achieves cutting-edge outcomes with fewer parameters. We obtained the most outstanding performance results on CHASEDB1 and DRIVE datasets with an F1 measure of 0.8351 and 0.8242, respectively. Our model has few parameters (0.34 million) compared to other networks such as ladder net with 1.5 million parameters and DCU-net with 1 million parameters.

The second part of this thesis investigates the association between whole and regional volumetric alterations with increasing age in a large group of healthy subjects (n=6739, age range: 30–80). We used a deep learning model for brain segmentation for volumetric analysis to extract quantified whole and regional brain volumes in 95 classes.

Segmentation methods are called edge or boundary-based methods based on finding abrupt changes and discontinuities in the intensity value. The third part of the thesis introduces a new Lightweight Edge Detection Network (LEON). The proposed approach is designed to integrate the advantages of the deformable unit and DepthWise Separable convolutions architecture to create a lightweight backbone employed for efficient feature extraction. Our experiments on BSDS500 and NYUDv2 show that LEON, while requiring only 500000 parameters, outperforms the current lightweight edge detectors without using pre-trained weights.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor, Prof. Amirali Baniasadi, for understanding and helping me throughout my research. This thesis would not have been possible without his guidance and support. I would also like to acknowledge Dr. Ibrahim Numanagić from the Department of Computer Science at the University of Victoria as the second reader of this thesis. I am gratefully indebted to him for providing facilities and his precious comments on the thesis.

I would like to dedicate this thesis to my wonderful parents for their encouragement, endless love and support.

Nasrin Akbari

Chapter 1

EFL-Net: AN Efficient Lightweight Neural Network Architecture For Retinal Vessel Segmentation

1.1 Introduction

The shape of retinal vessels reflects a patient's overall health and aids in diagnosing several conditions, including diabetes and hypertension. Patients' blindness can be avoided with timely diagnosis and treatment of specific disorders. Deep learning algorithms have recently achieved the best results compared to other techniques for retinal vessel segmentation. However, a major drawback with these methods is that these models require a large number of parameters and computations. In this paper, we proposed a lightweight neural network for retinal blood vessel segmentation named efficient and fast lightweight network (EFL-Net). EFL-Net introduces the ResNet branches shuffle block (RBS block), which has a high capacity to extract features from several granularities, and the Dilated Separable Down block (DSD block), which enhances the network's receptive field. Both suggested blocks are lightweight that can be plugged into the state-of-the-art backbone of CNN models. In addition, we adopt PixelShuffle in the decoder of our model as an upsampling layer, which has a greater capacity than deconvolution and interpolation approaches for learning features. We provide the findings from two datasets: Drive and CHASEDB1. Our model achieves cutting-edge outcomes with fewer parameters. We gained the most significant performance results on CHASEDB1 and DRIVE datasets with an F1 measure of 0.8351 and 0.8242, respectively. Compared to other networks, such as ladder net with 1.5 million parameters and DCU-Net with 1 M parameters, our model has fewer parameters (0.340 M).

1.1.1 Retina

The retina is the light-sensitive nerve tissue layer at the back of the eye that receives images and transmits them to the brain as electric signals via the optic nerve [1]. Retinal and optic nerve changes may indicate some diseases such as hypertension and glaucoma [2]. A part of Highly Active Antiretroviral Therapy (HAART) is Ritonavir. It is a potent inhibitor of HIV disease protease that causes retinal impairment in long-term use. Primary eye care (PEC) can use a funduscopy examination to give an early screening for drug-induced retinal toxicity. [3]. In the procedure of funduscopy examination, An ophthalmologist looks at structures of the retina, retinal blood vessels, and optic nerve head (disk) of the eye [4].

Hypertensive Retinopathy (HR) is another disease that can blur sight [5]. The aging process augments the oxidative load, which leads to higher levels of oxidative

stress which in the retina causes pathologies such as aging macular degeneration or neuropathic complications of diabetes in the eye [6]. Diabetic retinopathy (DR) is a condition that can occur in persons who have diabetes. It gradually damages the retina, the light-sensitive lining at the back of the eye. Diabetic retinopathy is a major sight-threatening consequence of diabetes. Diabetes impairs the body's capacity to utilize and retain glucose (glucose). Too much sugar in the blood causes the condition, damaging the body, including the eyes. Diabetes causes tiny blood vessels throughout the body, including the retina, to deteriorate over time. It occurs when these tiny blood vessels leak blood and other fluids. It causes the retinal tissue to swell and cloudy or blurred vision. As described, retinal vessel images acquired from the fundus can be used to detect severe diseases.

There are several ways to analyze retinal images and find diseases. One of them is retinal image segmentation, which is divided into manual and automatic segmentation. Manual segmentation needs expert technical staff and is time-consuming. Automated algorithms are valuable for the early detection and treatment of diseases related to the eyes due to benefits over manual segmentation, such as increased accuracy, reduced cost, and quicker speed.

One of the automatic models applied to retina images to achieve corresponding vessel segmentation is U-Net [7]. It is one of the successful medical and biomedical image segmentation methods based on deep neural networks. Many efforts based on it have been proposed to leverage the encoder-decoder paradigm for retinal vessel segmentation.

In human sight, discrimination between images containing blood vessels and their corresponding distorted background is a difficult task. Hence, detecting diseases is much more challenging. So developing practical algorithms to identify vessel images and surroundings would be useful. [8].

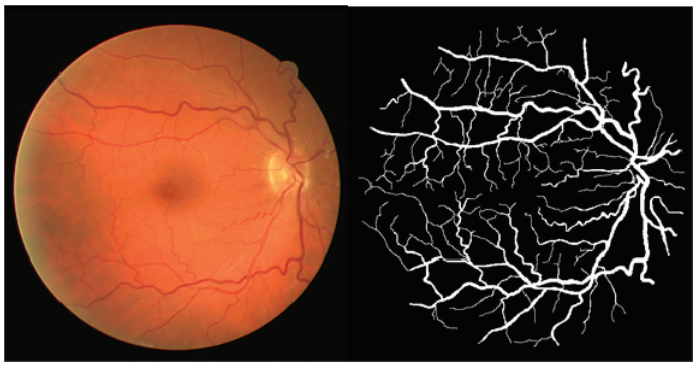

FIGURE 1.1: A retinal image from DRIVE dataset (left) and retinal vessel segmentation (right) [9]

1.1.2 Problems In Retinal Images

Images of the retina are taken with a unique camera tool known as a fundus camera, which is categorized as one of the non-invasive imaging modalities. There are many problems in retinal images obtained from the fundus. Many fundus images contain low contrast, noise, and other problems resulting from poor acquisition processes. However, the fundamental problem is the poor contrast from uneven lighting of the unevenly curved retina surface. Low contrast and noise issues can be resolved using image processing techniques. Moreover, these issues must be primarily fixed in the first place in order to detect and analyze retinal vascular changes that assist in disease diagnosis.

Researchers presented various algorithms to enhance retinal vessel detection, concentrating on the segmentation and extraction of specific data from connected vessel stems and terminals. Segmentation of the tiny vessels is difficult, and most approaches miss them throughout the segmentation process. Consequently, accurate retinal vascular segmentation is critical for identifying eye disorders.

1.1.3 Research Objectives

Deep neural networks (DNNs) approaches can automatically learn reliable, complex features from raw data without requiring feature engineering. These techniques have achieved enormous success in computer vision and played a significant role in medical health. A considerable number of research papers [7, 10, 11, 12] on retinal vessel segmentation have been published. Many of those researches suggest different deep neural network architectures for retinal vessel segmentation. Still, according to our observation, most of them are not optimal in architecture and number of parameters. We believe more research is needed to improve the number of parameters and accuracy.

The main objective of our research is to design a lightweight CNN architecture with fewer parameters to segment retinal vessels with the same or better segmentation results of cutting-edge networks.

1.2 Literature Review

1.2.1 Retinal Segmentation Importance

Ophthalmologists can learn more about the condition of patients' eyes by segmenting retinal blood vessels. An experienced ophthalmologist reviews the collected retinal images by manually isolating the vessels from their background. During manual segmentation, the observer will unavoidably make a significant intervention. [13, 14]. This approach takes much time, and its accuracy varies. As a result, early treatment choices are delayed [15].

Adopting an automated retinal image segmentation is a more effective substitute. Automatic segmentation of blood vessels relies on a good strategy for dealing with changes, and low contrast, noise, and central light reflex are required. Automatic retinal vessel segmentation and detection can be helpful in several diagnostic processes, including creating a DR grading system, evaluating early retinopathy, determining the foveal avascular zone, and identifying additional anomalies. Thus segmenting retinal blood vessels from images of the retinal fundus has become one of the fields of study that has gained more popularity recently.

1.2.2 Retinal Vessel Segmentation Methods: Image-Processing

Some essential image-processing techniques for retinal vessel segmentation are pattern recognition, multi-scale retinal, mathematical morphology, matched filtering, and model-based retinal vessel extraction techniques.

Pattern-recognition algorithms classify retinal vessel and non-vessel pixels. The method categorizes into supervised and unsupervised techniques. Manual labeling is required for supervised techniques to identify the vessel and non-vessel pixels. Unsupervised techniques do not require labels to find vessel pixels.

In the supervised approaches, Gaussian mixture models (GMM) [16, 17] and the KNN [17], SVM [18, 19] are typically utilized as classifiers.

A supervised technique based on extracting a feature vector for each pixel in the green channel of retinal images was proposed by Niemeijer [20]. Their research uses second-order derivatives at several different scale factors and the Gaussian matching filter for feature extraction. After that, the KNN method is applied as a classifier. It classifies pixels based on whether or not they include vessels. These methods are effective for high-quality images but not for noisy images.

When it comes to unsupervised approaches, the vessel pixels are categorized without the use of any human-applied labels. In [21], Salem et al. suggested the radius-based clustering algorithm as an unsupervised vascular segmentation. They map the distribution of image pixels using the distance-based principle in this method. Their algorithms' performance is validated on the STARE database [22], where they achieve an average sensitivity of 82.1% without reporting their accuracy. The key benefit of Salem's proposed method is that it is a semi-supervised approach that produces retinal segmentation without needing the training set. They employ a hybrid approach combining supervised and unsupervised techniques to produce the feature vector of each vessel pixel in the image. Then, they apply a modified KNN algorithm to cluster the pixels in the image according to the feature vector.

The morphological image-processing methods are a group of digital image-processing techniques based on mathematical morphology. The morphology method is based on specific operations, such as top-hat and bottom-hat, and employs structural elements. Zana and Klein proposed a strategy for the identification of retinal vessels [23]. It utilizes mathematical morphological curvature validation. The authors of the study characterized the vessels as having a brilliant pattern that was piecewise connected and had a locally linear pattern. They used the mathematical morphological and cross-curvature evaluation technique to obtain a binary vessel image. They use a four-step process. The first phase deals with noise reduction, while the second step deals with the linear pattern with the vessels' Gaussian profiles. Evaluation of the cross-curvature comes in as the third phase, and the final vessel images are obtained in the fourth stage, which is based on linear filtering. By minimizing false positive rates, the strategy enhanced performance. However, their technique misses small vessels like previous retinal vessel algorithms. Morphological-based methods suffer from common problems, such as lack of vessel enhancement, background noise, and lack of detection of tiny vessels.

Multiscale retinal vessel extraction methods use vessel width at different scales. The vessel's width narrows as it moves out from the optic disc and increases as it approaches inside. Multiscale approaches are proposed to better segment vessel images by representing vessels in their proper size. Nguyen et al. [24] created a retinal vessel segmentation technique employing a primary line detector. This strategy overcomes each line detector's weaknesses. The final segmented image combines line filter responses at different scales. This approach is tested on REVIEW, DRIVE, and STARE [22]. It is accurate even in low-contrast images. However, it does not detect tiny vessels.

Hou et al. [25] modified Nguyen et al. [24] 's multiscale approach and provided an improved one. The author presented a multi-directional morphological top-hat transform by rotating structuring elements to make the background consistent and enhance vessels. Afterward, he employed a multiscale line detector. Line detectors at different scales give diverse line responses. The improved multiscale line detector averaged all responses at different scales by modifying their weights to create the segmented image. The results of these experiments demonstrated that this strategy was capable of segmenting close vessels and also could deal with central reflex vessels. Their technique decreases false positives on pathology-containing images but

needs refinement for detecting small vessels.

Multiscale approaches have three issues. First, the segmented image is noisy. Second, their performance on pathology-containing images is poor, and third, the tiny vessel detection needs further attention. These approaches need improvement to eliminate noisy pixels, work on images with pathologies, and find tiny vessels.

The segmentation of retinal vessels using the model-based approach relies on mathematical modeling and image-processing methods to produce segmented vessel images. Vermeer et al. [26] introduced a model-based retinal segmentation to increase the procedure's effectiveness; the classification stage is added after the first phases, which are based on Laplace and thresholding techniques [27]. The STARE dataset [22] is used to validate the approach, which results in an accuracy of 92.8%. By adjusting the approach to account for specific blood vessel characteristics, it is possible to provide the right vessel direction for segmentation.

Karunanayak et al. [28] suggested a model-based retinal segmentation which consists of three steps. Adjusting point operators fix uneven lighting. Second, compute vessel edge features and enhance them with a Gabor filter. The last step uses region-growing with automatic seed point selection to segment retinal vessel images. On DRIVE, the method achieves 94% accuracy. This method works well because it solves two major retinal image problems. First, they overcame differences in retinal image illumination to observe the vascular network. Second, they enhanced retinal vessel visibility. Their method needs more validation on challenging images and observing tiny vessels.

Before introducing deep learning techniques, traditional approaches such as thresholds, clustering, and histogram-based algorithms have been suggested to address image segmentation issues. While neural networks have been around for years, deep learning is now a cutting-edge technique verified by the scientific community [29]. Deep learning refers explicitly to developing a neural network model to address imaging issues. The concept of having several layers of image/signal processing units is implied by the word "deep," which is comparable to the organization of neurons in the brain. Deep learning's primary benefit is automatically identifying a training dataset's best features. By doing this, the laborious process of manually choosing the characteristics of a particular job, such as classification or regression, is avoided.

1.2.3 Deep Learning Techniques in Retinal Fundus Images

One of the first scientists who used neural networks to identify retinal arteries was Zhang et al. [30]. They trained the network to locate retinal vessels using a self-organizing map (SOM) as a preprocessing strategy, and their unsupervised suggested method produced accurate segmentation of retinal vessels. They follow a three-step process. First, they took the green channel of the intensity of the RGB retinal image and used it to make a multidimensional feature vector. Then, to make the vessel an enhanced intensity feature, a morphological procedure is used. Second, a self-organizing map SOM, a kind of unsupervised neural network using the idea of pixel clustering, is employed as a classifier. In order to achieve the goal of final vessel segmentation, the Otsus threshold approach is used to categorize each neuron of the output layer of the SOM as a retinal vessel neuron or a retinal non-vessel neuron.

Tan et al. introduced a seven-layer CNN approach [31], recognizing not only retinal vessels but also the optic disc and fovea in the retinal image. The simultaneous detection of many features significantly impacts the precision of segmenting the retinal arteries. They applied the normalization approach before segmentation to

obtain consistent background illumination and contrast to eliminate noise and non-uniform backdrop from retinal images. Before being fed into the subsequent CNN, their contribution included the selection of efficient vessel pixels in fundus images of three color channels. Four neurons in the output layer stand for the backdrop, the optic disc, the fovea, and the blood vessels, respectively.

A framework for identifying retinal vessels and optic disc segmentation was introduced by Manini et al. [32]. They developed a CNN and called it deep retinal image understanding. Its CNN architecture is based on VGGNet, which is mostly employed for image classification. The FC layers at the network's edge are taken out using their technique. Four max-pooling layers are employed between the convolutional layers to divide the architecture into five phases. Feature maps are produced using convolutions with various filters of the same size between the pooling layers. The deeper network, the more information on the characteristics is provided. Because they noticed that the coarse feature maps at the final stage of CNN did not help with vessel detection, and thin vessels were not appropriately detected. Therefore, they downscaled the image layer by layer and achieved two feature maps for retinal vessels and the optic disc segmentation. Their technique produces an accurate final vessel image based on a probability map identifying pixels as vessels or discs. However, optical disc shadow images produced by their technology do not demonstrate improved outcomes.

In order to retinal vessel detection, Mahapatra et al. [33] designed a technique for retinal vessel recognition based on the local saliency map and CNN architecture. Saliency maps from unsupervised and supervised data from the trained CNN are combined to get the final output. In order to give both global and local image information, their approach is unusual in that it calculates the saliency values of each image pixel at various scales. The CNN can obtain extra information from such vessel pixel data. Compared to other approaches, the method's performance is improved by combining information. Their method's primary objective is to evaluate image quality for analyzing the retinal vessels. Due to their method's simple computation, it can swiftly assess image quality and advise the patient to seek treatment as soon as possible.

Several works have effectively used the U-Net [7] to segment blood vessels [34, 35]. It performs exceptionally well in a variety of biomedical segmentation applications. It requires a tiny number of annotated images and has a good training time. An expansive path (left side of the U-Net structure) and a contracting path make up the structure (right side of the U-Net structure). The contracting path adheres to a convolutional network's typical architecture. It entails applying two 3×3 convolutions (unpadded convolutions) repeatedly, each time being followed by a rectified linear unit (ReLU), a 2×2 max pooling operation, and a stride of 2 downsampling operation. They multiply the number of feature channels by two with each downsampling step. An upsampling of the feature map is followed by a 2×2 convolution ("up-convolution") that cuts the number of feature channels in half, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU, at each stage of the expansive path. Due to the loss of border pixels in each convolution, cropping is required. Each 64-component feature vector is mapped to the required number of classes at the top layer using a 1×1 convolution. The network includes 23 convolutional layers in total. It is crucial to choose the input tile size to apply all 2×2 max pooling operations to a layer with an even x and y-dimension. It will enable a smooth tiling of the output segmentation map. However, the results in vessel segmentation are unsatisfactory since no attempt was made to achieve the optimal trade-off between accuracy

and the number of network parameters. Deep networks' retinal vessel segmentation has improved more in the past five years (from 2018 to 2022) [36, 10, 11, 37, 12, 38, 39, 40]. Reza Azad et al. proposed BCDU-Net (Bi-directional ConvLstm U-Net with densely connected convolutions) in [36]. They demonstrated that the network could capture more discriminative information and produce more accurate segmentation results by integrating BConvLSTM in the skip connection and adding densely connected convolutional blocks. In addition, they found that employing batch normalization after the up-convolution layer allowed them to accelerate the network by around six times. Compared to state-of-the-art alternatives, the experimental findings on three publicly available benchmark datasets demonstrated considerable gains in semantic segmentation. Their BConvLSTM helps to non-linearly combine the feature maps acquired from the respective encoding path and the prior decoding up-convolutional layer in place of a simple concatenation in the skip connection of U-Net. In the last convolutional layer of the encoding pipeline, they employed densely coupled convolutions to improve feature propagation and promote feature reuse. Their suggested model is assessed on three datasets, segmentation of skin lesions, lung nodules, and retinal blood vessels.

Yuqian Zhou et al. are improving Retinal Vessel Segmentation Trained with Noisy labels [10] to increase the resilience of the model trained on noisy labels. They suggest the K-fold Study Group Learning (SGL) better deal with noisy labels in limited datasets, notably for the retinal vascular segmentation challenge. Their idea was inspired by the K-fold cross-validation scheme and knowledge distillation methodologies. With this technique, the entire training set was initially randomly and averaged into several subgroups. They train all of these subsets using pixel-wise binary cross-entropy loss, similar to the cross-validation scheme. They train a model from scratch by optimizing the ground truth vessel labels and the pseudo-derived label set after optimizing the model sets. They then infer the estimated segmentation label as the pseudo label. Since each model learned using a partial training set may be viewed as a "Study Group," they named their learning strategy Study Group Learning or SGL. By combining and integrating the knowledge from several study groups, the final model is developed through group debate. Theoretically, this plan might be seen as a regularization to prevent inaccurately and overfitting annotations. Their model performs well on noisy data.

Authors in [11] suggest using a context encoder network to retain spatial information and collect additional high-level information for 2D medical image segmentation. A feature encoder module, a context extractor, and a feature decoder module are the three core parts of CENet. The fixed feature extractor is a pre-trained ResNet block. A recently designed dense atrous convolution (DAC) block and residual multi-kernel pooling (RMP) block combine to make the context extractor module. They used the planned CE-Net to segment various 2D medical images.

Lei Mou et al. created the CS2-Net, a novel curvilinear structure segmentation network that works with 2D images and 3D volumes in [37]. Using a self-attention method on high-level characteristics in the channel and spatial dimensions, their CS2-Net enhances the inter-class discriminating and intra-class aggregation characteristics. The encoder module, the channel and spatial attention module (CSAM), and the decoder module make up this network. The characteristics of the input data are extracted using the encoder module. In order to create channel-spatial attention-aware expressive characteristics, these features are put into two parallel attention blocks, the channel attention block (CAB) and a spatial attention block (SAB). The SAB weighted features in all spatial locations are used to selectively aggregate the

features in each spatial location, allowing the model to capture the features' long-range dependencies and ensure that comparable characteristics are associated with one another regardless of distance. The CAB, meanwhile, ensures that the entire space is utilized to represent, normalize, and hence increase the contrast of the features in various channels, enabling the model to be built with enhanced discriminating characteristics. The segmentation outcome is produced after the decoder module has been used to recover curvilinear features. They offer a feature decoder module that recovers the dimensions of the high-level semantic features layer by layer instead of immediately up-sampling the CSAM features to the original image dimensions. Each of the four blocks in the encoder and decoder modules uses a residual network (ResNet) as its backbone, with a max-pooling layer placed on top to broaden the receptive field and improve the extraction of global features. A skip link is inserted between each layer of the encoder and decoder, similar to the U-shaped network, to combine the features at various levels and compensate for the information loss brought on by the max-pooling processes. The final segmentation map is created by applying a 1×1 kernel ($1\times1\times1$ kernel in the 3D phase) convolutional layer and a sigmoid layer to the encoder output at the conclusion of the CS2-Net.

In [12], the Authors introduced a modified U-Net model that can locate hidden vessel features from the segmented vessel image itself instead of the raw input image. IterNet is made up of several mini-UNet iterations, each of which can be four depths deeper than a typical UNet. In order to make training easier, IterNet additionally uses weight-sharing and skip-connection capabilities. As a result, even with such a big architecture, IterNet can still learn from just 10 to 20 annotated images without any pre-training or prior knowledge. They add an iteration of mini-UNets after UNet and utilize the result of UNet as the input of the next mini-UNets to correct mistakes highly prevalent in existing vessel segmentation models, such as inconsistent vessels and missing pixels, etc. They successfully provide IterNet the capacity to identify potential defections in the intermediate outcomes and fix them rationally by incorporating weight sharing in mini-UNets and skip-connections.

LadderNet is suggested in [38], which may be thought of as a chain of several U-Nets. LadderNet features skip connections between every pair of adjacent decoder and decoder branches in each level, in addition to having several pairs of encoder-decoder branches as opposed to only one pair in a U-Net. Additionally, they employed modified residual blocks, which combine two convolutional layers into a single block and share weights. A LadderNet may be considered an ensemble of Fully Convolutional Networks and provides extra channels for information flow due to skipping connections and residual blocks (FCN). The shared weights within each residual block help to reduce the number of parameters, while the equivalence to an ensemble of FCNs increases segmentation accuracy.

Md Zahangir Alom et al. present a recurrent U-Net design in [39] that maintains the original U-compactness. The initial convolutional unit in this configuration comprises eight feature channels. The number of channels doubles after each pooling layer in the encoder, as per the original U-Net technique. Unlike bilinear interpolation, the decoder uses transposed convolutions to boost the model's representational strength. Its core is the segmentation network's several encoding and decoding layers, which are covered by the recurrent units.

Xin Yang et al. [8] present a feature extraction module based on deformable convolution and residual structure, which may significantly enhance the modeling capability for the vessel. Two U-Net make up the DCU-overall Net's framework. Additionally, they employ a residual channel attention module (RCAM) to connect

U-Net models to increase the efficiency of information flow between them. This technique improves the network's capacity to express features and makes model improvement easier. In the last five years, the segmentation of retinal vessels using deep networks has improved substantially. In [40], In order to perform adaptive feature refinement, SA-UNet incorporates a spatial attention module that infers the attention map along the spatial dimension and multiplies the attention map by the input feature map. In order to avoid overfitting, the suggested network substitutes structured dropout convolutional blocks for the original U-Net convolutional blocks.

Currently, there is a trend in retinal segmentation toward creating a more lightweight network while preserving segmentation accuracy. We show in Figure 1.2 the F1 measure and complexity of many well-known DNN-based techniques [36, 10, 11, 37, 12, 38, 39, 40, 8]. By looking at the number of parameters in their suggested networks, it is clear that additional study and development are required to construct a network with fewer parameters with the same or higher accuracy.

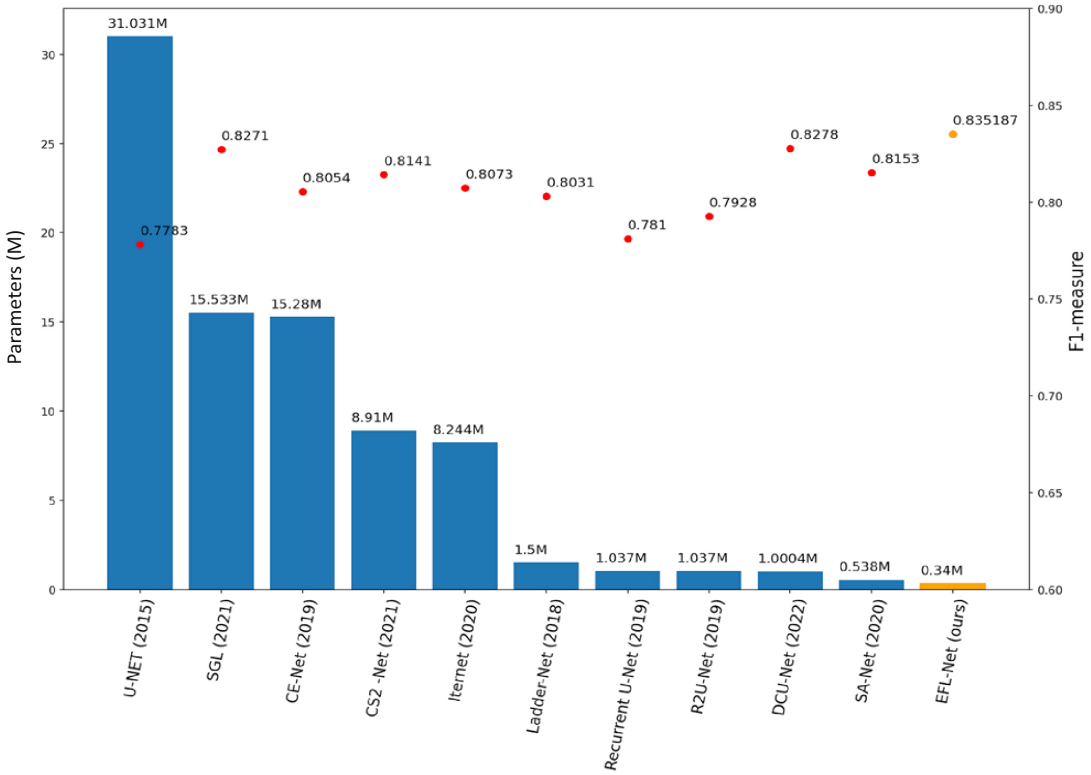

FIGURE 1.2: Accuracy and number of parameters of several retinal vessel segmentation papers in five past years (CHASEDB1 dataset [41])

1.3 Methodology

The U-Net is a convolutional neural network that was designed to perform image segmentation in the field of biomedicine [7]. U-net is an enhancement on the previously developed FCN — "Fully convolutional networks for semantic segmentation" [42]. Because it performed so well and could use tiny training datasets, U-Net became the most reliable architecture for the semantic segmentation of biological images.

U-Net architecture comprises four encoder blocks in the left part, called the contracting path, and four decoder blocks in the right part called an expansive path. The encoder captures features from the input image and decreases its resolution via pooling layers. And the decoder part reconstructs the image and restores object details via skip connections between the encoder and decoder layers as shown in Figure 1.3.

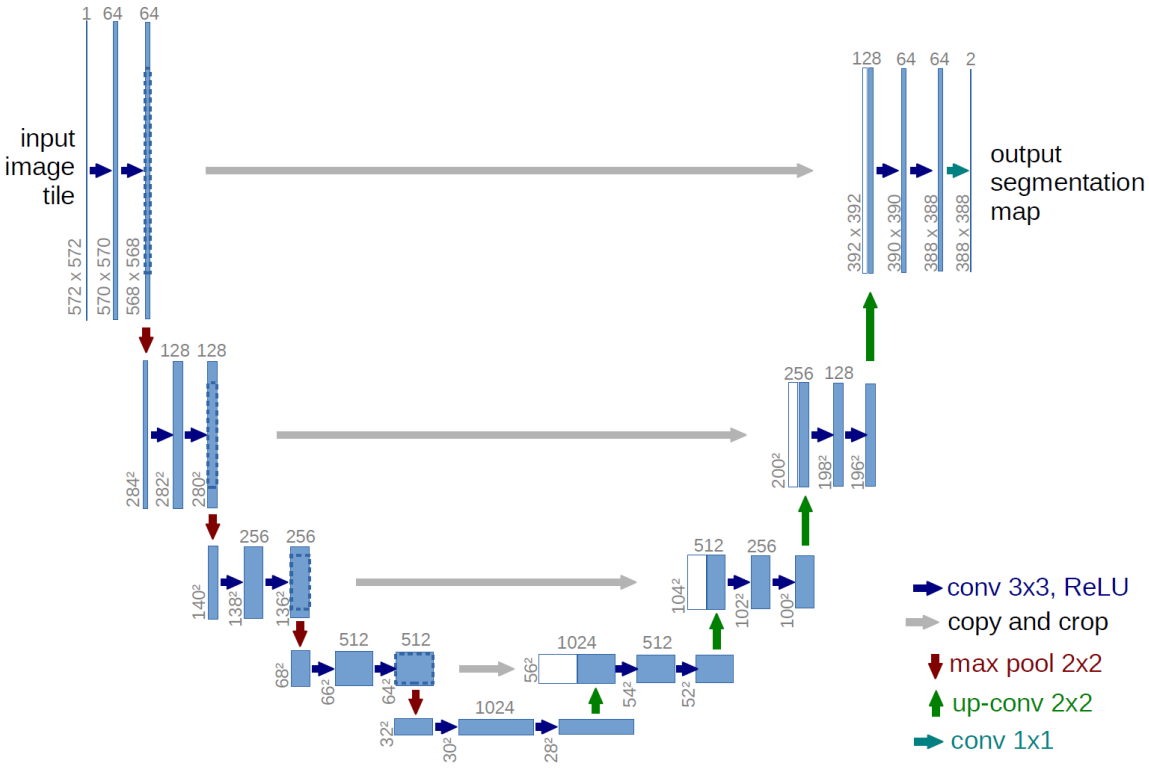

FIGURE 1.3: U-Net architecture [7].

Despite its achievements, this model has several drawbacks, such as a large number of parameters (31.031 Million) and poor performance on the task of retinal vessel segmentation (F1 score of 0.7783 On CHASEDB1 dataset). Consequently, we tried to improve this architecture for retinal vessels by searching and investigating other papers and their blocks to address the mentioned problems. Finally, we introduce our solution, focusing mainly on developing two blocks for enhanced feature extraction: the Resnet Branches Shuffle Block (RBSB) and the Dilated Separable Down Block (DSDB). Besides, we use efficient layers such as pixel shuffle of deconvolution and interpolation techniques in the decoder path of U-Net [43].

1.3.1 Our Proposed Architecture

Inspired by the U-Net [7], ShuffleNetV2 [44], we propose our model for retinal vessel segmentation, which is named Efficient and Fast Lightweight Neural Network (EFL-Net). We aim to establish a lightweight and accurate deep learning model to segment the retinal blood vessels. To increase the receptive field of Unet, we add the resnet branches shuffle block (RBS block) and the dilated separable down block (DSD block) to our architecture. We detail the proposed blocks in the next sections. A diagram of our architecture is presented in Figure. 1.4. There is a Dropout Block layer after each RBS block, and within this block, there is a batch normalization layer

after each convolution layer. The encoder path of U-Net has four stages, and each stage is made up of EFL-Net, RBS, and DSD blocks for extracting features and down-sampling. We use the PixelShuffle layer [43] instead of the deconvolution layer in the decoder path of U-Net for upsampling. Furthermore, rather than concatenating the encoder and decoder features, we add them together to reduce computation.

In the subsequent section, we will discuss the idea in greater detail and break it down into the core concepts that form the structure of our architecture.

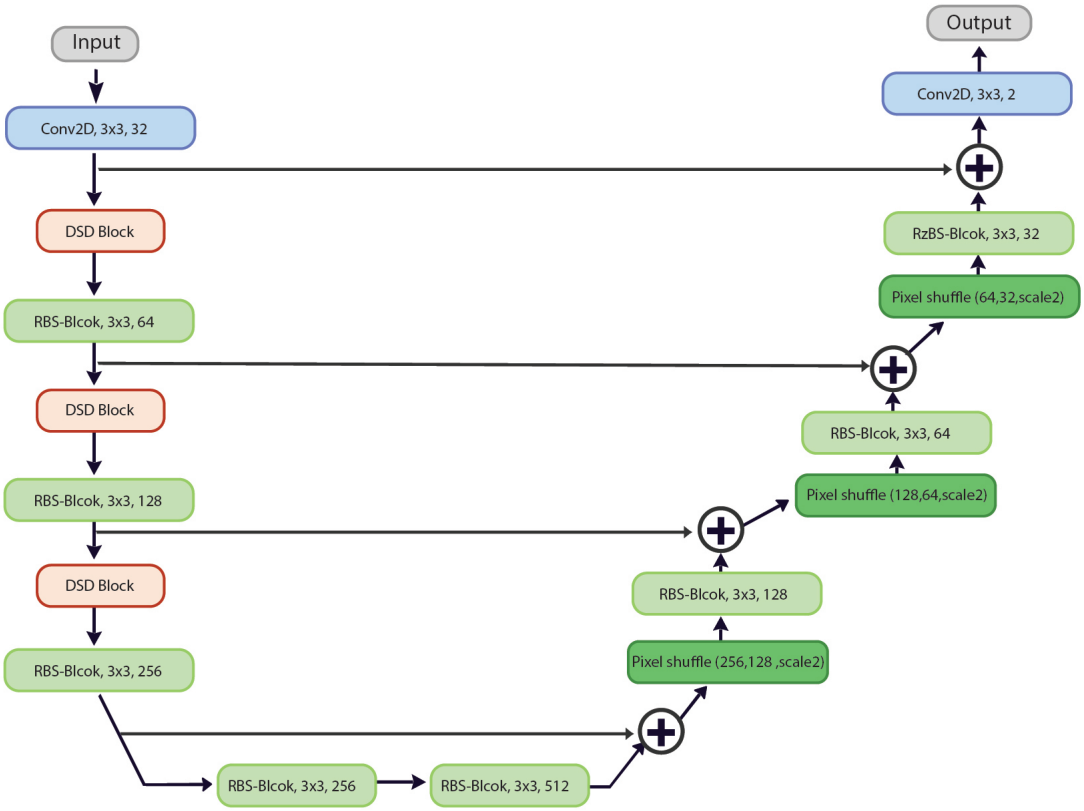

FIGURE 1.4: EFL-Net Architecture

1.3.2 Resnet Branches Shuffle Block

In order to improve the feature extraction,Inspired by Res2Net [45], We made several changes to the shuffleNetv2 basic unit [44] which is shown in Figure 1.6. Compared to the original basic unit of ShuffleNetV2, we changed the number of split channels in the input feature maps in the RBS block (Figure 1.5). In this block, the input with n channels is divided during the forward pass into n/4 groups with the same number of channels each, which is indicated by x_i, where $i \in \{1,2,3,4\}$. The spatial size of each feature subset x_i is the same. Each x_i has a matching original bottleneck unit of the ShuffleNetV2 basic unit, indicated by $C_i()$, with the exception of x_1. The result of $C_i()$ is used to represent y_i. The output of $C_{i-1}()$ is added to the feature subset i, and then it is passed into $C_i()$. The output result of $C_i()$ has a bigger receptive field than x_i. Additionally, the channels $x_i(i > 1)$aggregates to feature subset x_1 to reuse the features. The remainder of the block is identical to the ShuffleNetV2 convolution block. We plugged an RBS block into our network as a lightweight feature extractor in our architecture's encoder and decoder path. The RBS block is more effective than

the original shuffleNetv2 basic unit due to its superior multi-scale feature extraction capabilities and fewer parameters. To compare the strength of the RBS block and the ShuffleNetV2 block, we did some ablation experiments and provided the results in the experimental result.

According to Figure 1.5, our designed architecture has several concepts and blocks, which we will explain in detail in the following sections.

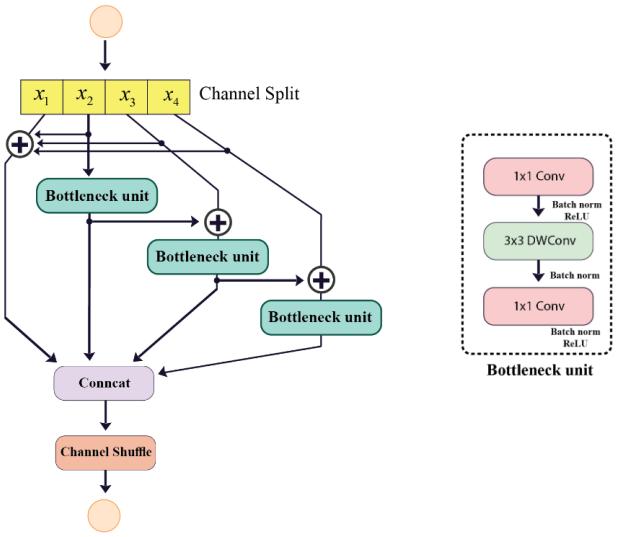

FIGURE 1.5: a) ShuffleNetV2 basic unit [44], b) Resnet Branches Shuffle Block (RBS block), DWConv stands for depth-wise convolution

1.3.3 ShuffleNetV2 Basic Unit

To achieve the goal of increasing the number of channels for feature extraction without increasing the total number of calculations, the ShuffleNetV2 [44] basic unit makes use of both pointwise [46] and group convolutions [46]. Also, the authors of ShuffleNetV2 introduce a simple operator known as channel split. At the beginning of the shuffeNet block, the channel split is used to separate the input feature map into two branches. The branch on the left remains as the identity for feature reuse. The other branch consists of three convolutions, two pointwise convolutions, and a group-wise convolution with the same number of input and output channels. After convolutions, the two branches are concatenated together. Therefore, the number of channels remains the same as input feature maps. Next, we use the same "channel shuffle" method to enable information communication between two branches [47], which will be explained more in the next section.

In order to keep things simple and describe our RBS block, we call the three convolutions (two pointwise, one depthwise) "bottleneck unit" (The dashed rectangle in Figure 1.6). We will discuss the depthwise and pointwise convolutions in the next section.

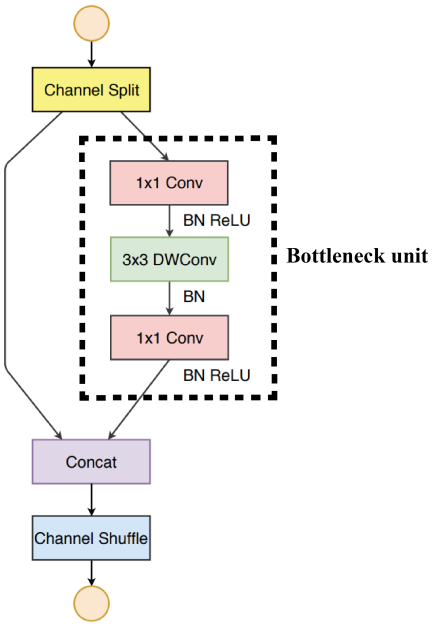

FIGURE 1.6: SuffleNetV2 basic block

1.3.4 Depthwise Separable Convolution

Depthwise separable convolution is an increasingly well-known technique that has been implemented in neural network design [46]. In deep learning frameworks like TensorFlow and Keras, a depthwise separable convolution is referred to as a "separable convolution." The depth-wise separable convolution is a modification of the standard convolution that enables it to perform the same operations with fewer parameters. A depthwise separable convolution separates the regular convolution process into 2 parts: a depthwise convolution and a pointwise convolution (Figure 1.7).

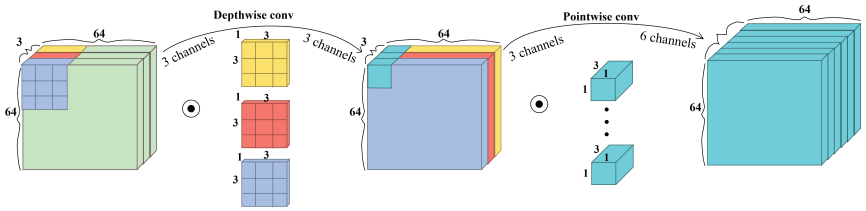

FIGURE 1.7: Depthwise separable convolution [48].

Depthwise Convolution

It is also known as "group-wise convolution." When performing a regular 2D convolution with multiple input channels, each filter has the same depth as the input (Figure 1.9), but depthwise convolutions ensure that each channel remains distinct from the others. In other words, in depthwise convolution, each filter has a depth of 1. For example, according to Figure 1.8, the "red" filter interprets the "redness" of each pixel in the input channel, and the "blue" filter interprets the "blueness" of each pixel in the blue channel, and the "yellow" filter interprets the "yellowness" of each pixel in the yellow channel. The steps involved in this technique are as follows:

after the input feature maps of three dimensions have been partitioned into distinct channels, the next step is to convolve the input of each channel with a 2D filter. Finally, the output of each channel is stacked together to obtain the output of the entire three-dimensional feature map.

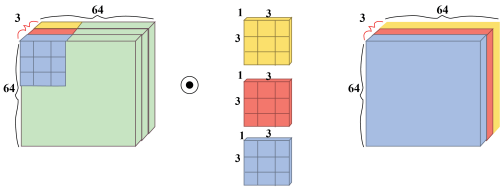

FIGURE 1.8: Depth-wise convolution (DWConv).

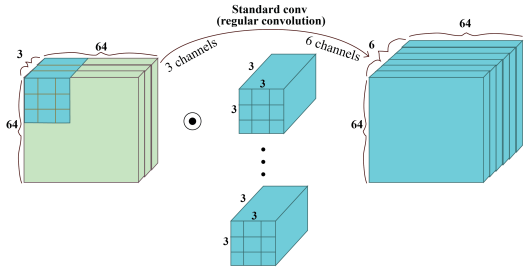

FIGURE 1.9: Regular convolution.

Pointwise Convolution

It is a convolution that iteratively processes each pixel with a 1x1 kernel. This kernel depth equals the number of channels in the input image. Together with depthwise convolutions, it yields an effective type of convolution called depthwise separable convolutions. (An example of using pointwise convolution is shown in Figure 1.10)

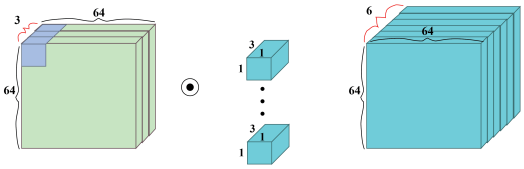

FIGURE 1.10: Point-wise convolution (1×1 Conv)

Depthwise separable convolution prevents overfitting since it requires fewer parameters than "regular" convolutional layers. Moreover, it is faster and more cost-effective in computation due to fewer parameters.

1.3.5 Channel Shuffle

It is an operation to transmit data between various channels to increase accuracy [47]. As we mentioned before, we divide the input channels into four groups, and we do depthwise separable convolution independently on each chunk of channels except the left one. Therefore, we will obtain several features that are independent of each other. Therefore, we will lose some information because our channels cannot communicate with each other, yet part of our input contains data crucial to another

group's features. One promising solution is to mix channels from different groups (Figure 1.17). Channel shuffle strengthens the representations of features as information flows freely between channel groups.

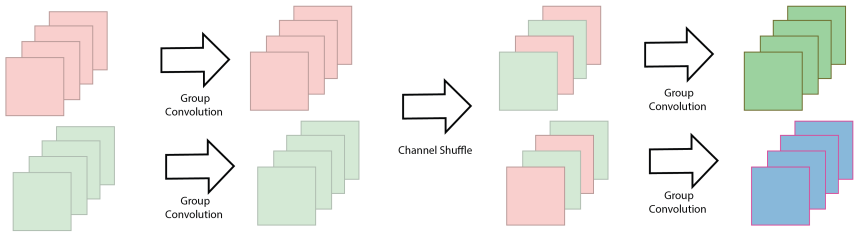

FIGURE 1.11: Channel shuffle operation on two subgroups of feature maps.

1.3.6 Dilated Separable Down Block (DSD)

Before we describe our DSD block, we need to understand the concept of the receptive field (RF) [49]. At each layer, the receptive field is defined as the size of a region in the input that contributes to generating a certain feature in the output. When we want to predict the boundaries of an object (i.e., an organ like the heart, a tumor, a vessel), we must provide the model access to all the relevant parts of the input image. In a CNN model, each neuron controls a specific part of the data. This lets the neurons learn patterns like lines, edges, and tiny details. During the convolution process, each neuron is exposed to different parts of input data, which fills a segmented area called the local receptive field. In Figure 1.12, three layers and corresponding receptive fields are shown.

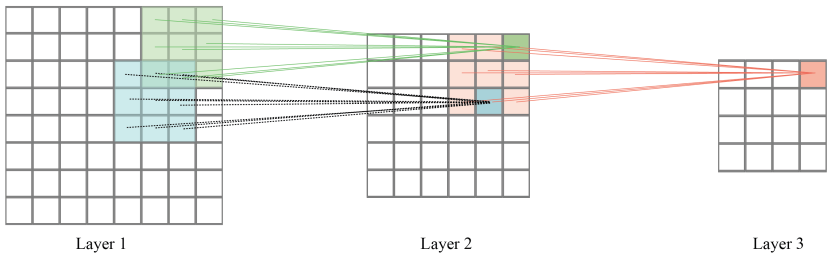

FIGURE 1.12: Receptive field.

The benefit of using convolutional layers with different sizes and strides is that it can learn specific features that might not be considered while using pooling layers with a fixed operation. We increase the model expressiveness by using convolution instead of a pooling layer. On the other hand, compared to regular convolutional operations with larger filters, dilated (atrous) convolution allows us to effectively enlarge the field of view of filters without increasing the number of parameters, so the computation cost remains untouched. Based on mentioned points, to increase the receptive field, we introduce the Dilated Separable Down block (DSD block) in which the input feature maps pass through two branches; each one has consecutive 3×3 group convolutional layers with a stride size of 2 and a pointwise convolutional layer. We apply different dilation rates to different groups to extract multi-scale features. As shown in Figure 1.13, we apply the size one dilation rate to the left branch and the size two dilation rate to the right, and then we concatenate the results

of the branches. In this way, we improve the ability of our network architecture to represent features.

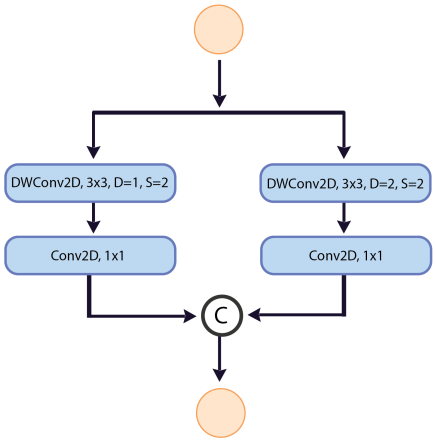

FIGURE 1.13: Dilated Separable Down Block

1.3.7 PixelShuffle

We use PixelShuffle [43] instead of deconvolution or interpolation methods for upsampling the features in the decoder path of U-Net. It improves our network's ability to segment retinal vessels and maintain as much information as possible. Upsampled feature maps are created by rearranging the pixels from the channels to the spatial dimension. Here, we upsample by a factor of r to transform a $H \times W \times Cr^2$ tensor into a $rH \times rW \times C$ tensor. In our design, we set $r = 2$ to upsample tensors by 2. An example of pixel shuffle is shown in Figure 1.14.

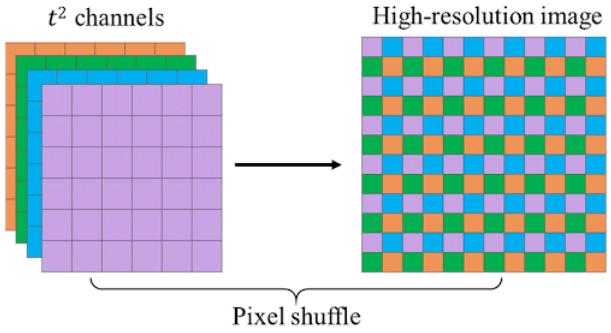

FIGURE 1.14: Pixel Shuffle layer [50]

1.4 Experimental Environment and Result

1.4.1 DATASETS and Data Preprocessing

DRIVE [41], and CHASEDB1 [41] are two retinal segmentation datasets that are available publicly, so we have tested our model on them. The DRIVE dataset has 40 2D RGB images, each with a resolution of 565 x 584 pixels, and the training and test sets each have 20 images.

The CHASEDB1 dataset contains 28 images, with one taken from 14 children's left and right eyes. Every image has a resolution of 999 x 960 pixels. The CHASEDB1 dataset has 20 training and eight testing images.

Two experts have carefully labeled the images to provide the ground truth. Most studies use the data from the first labeler as standard ground truth to train and test models. To increase the size of the dataset, we used data augmentation. Additionally, we used a mask or field of view (FOV) to ensure that extracted patches from the input image have vessels so that we do not have to do extra work on patches from the background. The DRIVE dataset already contains a binary mask. However, CHASEDB1 does not. Therefore, we created a mask for CHASEDB1 manually.

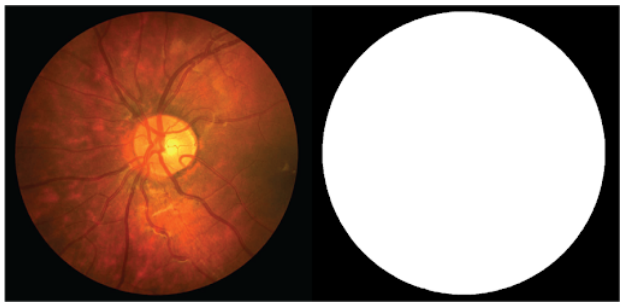

FIGURE 1.15: An image of a retinal vessel and its mask form DRIVE database [9].

It is crucial to pre-process the data before training the model, as fundus images exhibit noise and uneven lighting. So, initially, we picked the green channel of the RGB image since blood vessels show up better in green than other channels. In the next stage, data is normalized to have a mean of zero and a standard deviation of one for every pixel x in the image.

$$X_{norm} = \frac{X - \mu}{\delta}$$

Where μ and δ represent the mean value and standard variance of an image, then we scale the data value to [0,255].

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where X_{MIN} and X_{Max} are the minimum and maximum pixel values in the image X. Next, we use contrast limited adaptive histogram equalization (CLAHE) [51] and gamma adjustment to improve the contrast between the image foreground and background. Last, the image's intensity values are mapped to a range of 0 to 1.

We randomly pick 1250 64 x 64 patches from each fundus image during training. This makes a training dataset with 25000 DRIVE patches and 25000 CHASEDB1 patches. For testing, we take 16 sample patches from each test fundus image. Figure 1.16 is an illustration of input patches and corresponding label patches.

CLAHE

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a sort of Adaptive histogram equalization (AHE) that ensures that the contrast between different blocks of an image is the same and is not accentuated in any region of the image significantly [51]. The adaptive histogram partitions an image into different blocks.

OpenCV's default is eight by 8-pixel tiles. Thus it breaks this image into 8×8 tiles and then applies histogram equalization within each block or tile. In the end, the surrounding tiles are assembled using bilinear interpolation to eliminate the artificial boundaries.

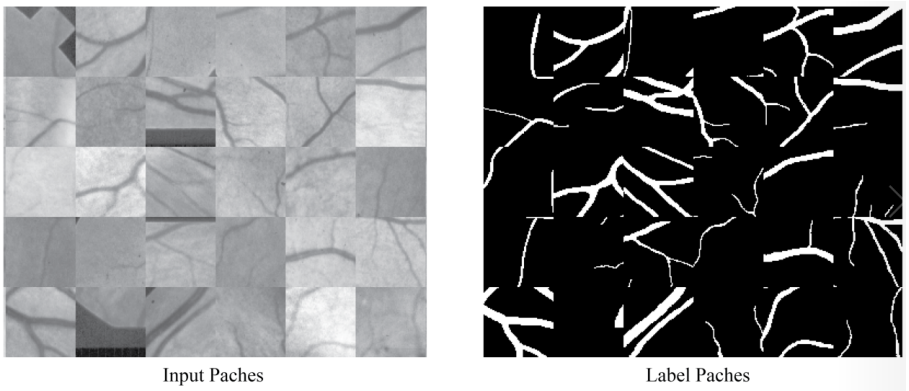

FIGURE 1.16: Examples of retinal vessel image patches (DRIVE database [9]).

1.4.2 Evaluation Approaches

By comparing the segmentation results to the ground truth (GT), it is possible to derive four scenarios. True positive (TP) measures the number of pixels that accurately classify the blood vessel into positive categories. False Positive (FP) denotes that the background was incorrectly assessed as having vessels. False Negative (FN) indicates the number of pixels that incorrectly segment the regions that have vessels into negative categories. In contrast, True Negative (TN) denotes the number of pixels correctly segmenting the background into negative categories. In addition to the four quantitative indicators already covered, the following six performance criteria can be used to evaluate the efficacy of a segmentation model: Sensitivity (SE), Specificity (SP), Accuracy (ACC), Precision (Pr), and F-Measure (F1). Their definitions are as follows:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The Specificity (SP) measures the ratio of correctly segmented background pixels to actual background pixels. In contrast, Sensitivity (SE) measures the ratio of

correctly segmented blood vessel pixels to actual blood vessel pixels. The Accuracy (Acc) value shows what percentage of the total image pixels were correctly segmented. Precision measures the quality of the model's positive prediction; however, recall measures the quality of negative predictions. The better architecture is trained on the given data, the greater the precision value is. F1 is the weighted harmonic mean of Precision and Recall.

1.4.3 Loss Function

We use a focal loss to address the class imbalance between foreground and background pixels.[52], which was suggested by Lin et al. (2017). Focal loss is defined formally as follows:

$$FL(\mathrm{p}_t) = -\alpha_t * (1 - \mathrm{p}_t)^{\gamma} * log(\mathrm{p}_t)\ where \begin{cases} p & if y = 1 \\ 1 - p & else \end{cases} \tag{1.1}$$

Where p indicates the predicted probability of the network output, the focusing parameter that can be adjusted is denoted by γ . In the focal loss, samples that are easy to classify contribute less to the loss values, while samples that are difficult to classify contribute more. This causes the model to focus more on difficult-to-classify samples. Hence it enhances the precision of challenging samples.

1.4.4 Experimental Environment And Parameters

We trained the network for 200 epochs from scratch and initialized the weights with random values. We use a batch size of 256. We selected the Adam optimizer with default parameters, and the initial learning rate is set to 0.001. We update the learning rate at each epoch by cosine function attenuation strategy. Instead of the most common loss function (cross-entropy loss) for the image segmentation task, we selected focal loss. It is used to address the issue of class imbalance problem. Since, in the segmentation of the retinal vessels, the number of pixels in the background of the images (negative class) is more than the vessel class (positive class). Focal loss helps in obtaining as much efficiency as possible. We use a server with Linux operating system,2.30 GHz processor, 128 GB RAM, NVIDIA TESLA P100 GPU, and Pytorch 1.7.0 framework for the implementation.

1.4.5 Experimental Result

We evaluate EFL-Net on the DRIVE and CHASEDB1 datasets. In order to evaluate the effectiveness of the model on test images, we divided the test images into blocks according to the sliding window with a size 96×96 and with the sliding step of 16. Then the predicted sub-blocks are combined to create the final image. Tables 1.1 and 1.2 show the performance comparison results of the proposed architecture with existing methods on the DRIVE and CHASEDB1 datasets using evaluation metrics of Se, Sp, Acc, AUC, and F1. It is clear from the results that our model performs best on CHASEDB1 and produces a Satisfactory result on DRIVE compared to other works. It achieves the highest F1 of 0.8351 and the high specificity of 0.9860 on the CHASEDB1 dataset, which indicates that the performance of our model is better than previous works on retinal vessel segmentation. By comparing the EFL-Net results with other networks, it is evident that our network has obtained results equivalent to the best networks with a fewer number of parameters in all evaluation criteria. Our lightweight network achieves these results with just 0.34 million

parameters compared to other networks such as ladder net with 1.5 million parameters and DCU-Net with one million parameters. The above results clearly show that the proposed model is an effective method for retinal vessel segmentation.

TABLE 1.1: Performance comparison between our EFL-Net and some state-of-the-art models on DRIVE

Methods	Year	F1	SE	SP	Acc	AUC	Parameters (M)
U-NET [7]	2015	0.7783	0.8288	0.9701	0.9578	0.9772	31.031
SGL [10]	2021	0.8271	0.869	0.9843	0.9771	0.992	15.533
CE-Net [11]	2019	0.8054	0.8093	0.9797	0.9641	0.9834	15.28
CS2-Net [37]	2021	0.8141	0.8329	0.9784	0.9651	0.9851	8.91
Iternet [12]	2020	0.8073	0.797	0.9823	0.9655	0.9851	8.244
Ladder-Net [38]	2018	0.8031	0.7978	0.9818	0.9656	0.9839	1.5
Recurrent U-Net [39]	2019	0.781	0.7459	0.9836	0.9622	0.798	1.037
R2U-Net [39]	2019	0.7928	0.7756	0.982	0.9634	0.9815	1.037
DCU-Net [8]	2022	**0.8278**	0.8075	0.9841	0.9664	0.9872	1.0004
SA-Net [53]	2020	0.8153	0.8573	0.9835	0.9755	0.9905	0.538707
EFL-Net (ours)	2022	**0.8242**	0.7957	**0.9802**	0.9567	0.9803	0.340

TABLE 1.2: Performance comparison between our EFL-Net and some state-of-the-art models on CHASEDB1

Methods	Year	F1	SE	SP	Acc	AUC	Parameters (M)
U-NET [7]	2015	0.8174	0.7537	0.982	0.9531	0.9755	31.031
BCDU-Net (d=3) [36]	2019	0.8224	0.8007	0.9786	0.956	0.9789	20.659
SGL [10]	2021	0.8316	0.838	0.9834	0.9705	0.9886	15.533
CE-Net [11]	2019	0.8243	0.8276	0.9735	0.9545	0.9794	15.28
CS2-Net [37]	2021	0.8228	0.8154	0.9757	0.9553	0.9784	8.91
Iternet [12]	2020	0.8205	0.7735	0.9838	0.9573	0.9816	8.244
Ladder-Net [38]	2018	0.8202	0.7856	0.981	0.9561	0.9793	1.5
Recurrent U-Net [39]	2019	0.8155	0.7751	0.9816	0.9556	0.9782	1.037
R2U-Net [39]	2019	0.8171	0.7792	0.9813	0.9556	0.9784	1.037
DCU-Net [8]	2022	0.8272	0.8115	0.978	0.9568	0.981	1.0004
SA-Net [53]	2020	0.8263	0.8212	0.984	0.9698	0.9864	0.538707
EFL-Net (ours)	2022	**0.8351**	0.7977	0.9860	0.9651	**0.9868**	0.340

We perform ablation experiments on the CHASEDB1 dataset. In the first step, we examined the relative performance of networks with and without DSD blocks (Max-Pooling used instead) in the encoder. The outcomes of the experiments are presented in rows 1 and 2 of Table 1.3. By contrasting the F-scores of the two scenarios, it is clear that the DSD block is superior to max-pooling. Improved F1 performance can be achieved with DSD blocks. Also, we switched RBS Block with the shuffleNetV2 basic unit to see how well the RBS block components worked in EFL-net compared to the ShuffleV2 Basic unit. By comparing the performance of EFL-Net (ShuffleNetV2 basic unit + DSD block) and EFL-Net (RBS block + DSD block), we can see that the RBS block can give noticeable improvements, and the model f1 value goes up by about 1.47 %.

As a result, the model performance is significantly better with the RBS and the DSD blocks, proving their effectiveness for the retinal vessel segmentation tasks.

TABLE 1.3: Ablation experiment on RBS and DSD blocks. The EFL-Net is trained and evaluated on CHASEDB1.

Method	F1	SE	SP	Acc	AUC	Parameters (M)
EFL-Net (Standard convolution + maxpooling2D)	0.747399	0.6839	0.9815	0.9489	0.82892	1.936738
EFL-Net (Standard convolution + DSD block)	0.770642	0.7031103	0.986	0.957	0.979272	1.788498
EFL-Net (ShuffleNetV2 basic unit + DSD block)	0.815429	0.739147	0.990829	0.962996	0.985666	0.410922
EFL-Net (RBS block + DSD block + cross entropy loss)	**0.801086**	**0.950963**	0.947362	**0.9648**	**0.988168**	0.340950
EFL-Net (RBS block + DSD block+ Focal loss)	**0.8351**	**0.7693**	0.9891	**0.9648**	**0.9871**	0.340950
EFL-Net (2 layer architecture)	**0.825407**	**0.782398**	0.985897	**0.963387**	**0.985175**	0.120804

The outcomes of our model segmentation on the DRIVE and CHASEDB1 datasets are displayed in Figure 1.17. From left to right, the original images of retinal vessels are depicted in the first column. Our network outputs for vessels are shown in the second column, and the third column is the binary result which uses a threshold to clean the network output. The ground truth of each input image is shown in the last column.

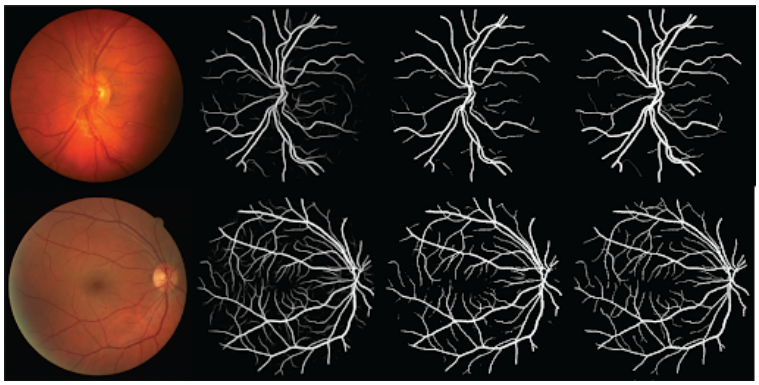

FIGURE 1.17: Retinal Vessel images, first row, an image of CHASEDB1 dataset. Second row, an image of DRIVE

1.5 Conclusion

In order to improve the accuracy and speed of blood vessel segmentation, we propose a lightweight network which is called EFL-Net. We chose modules with a few parameters while enhancing the network receptive field. Thus after extensive investigation, we proposed the ResNet Branches Shuffle Block (RBS block) and the Dilated Separable Down block (DSD block), both of which have a high capacity to extract features from several levels. RBS block is created by making changes to the convolution block of the shuffle Net block. In addition, rather than using 2D max pooling, which loses much information, we developed the DSD block, which can significantly expand the network's reception field while reducing feature size without sacrificing essential features. Besides, In an upsampling path, we use PixelShuffle, which is better than deconvolution and interpolation at learning better features. Two datasets are used to assess the suggested model. The results of our network show that our results are comparable to state-of-the-art results and have a lightweight structure with 0.340 million parameters that keep computational costs low. On the DRIVE datasets, the results were satisfactory (F1 is 0.8242), and the best performance was seen on the CHASEDB1 dataset (F1 is 0.835187).

Chapter 2

A Cross-Sectional Study of Regional Brain Volume Analysis on Large Healthy Dataset

2.1 Introduction

The structure of the brain is constantly changing throughout its lifetime. The volumetric analysis of structural brain high-resolution magnetic resonance images (MRI) may provide quantitative evidence of neurological diseases and help radiologists to identify patients at risk for rapid volume loss in the earliest stages. Several studies have investigated the effect of age and gender on brain structures. However, most of these studies' findings are based on small datasets, and many of them have not normalized the brain volume by head size in their analysis, which can lead to misleading results. In this study, to clarify the effect of age and gender on regional and global brain volumes, we used high-resolution 3D T1-weighted MRI data from a large cross-sectional population (n = 6739, 25–75 years). We segment all the subcortical structures of the brain with the help of the FastSurfer network into 95 regions and measure the volume of the critical brain structures—hippocampus, ventricles, and other brain structures. Doctors and researchers require access to age/gender-related normative percentiles of brain volume to reliably distinguish brain disorders such as Alzheimer's, multiple sclerosis, epilepsy, and traumatic brain injuries from the consequences of normal aging.

As a person ages, changes occur in all body parts, including the brain. Brain aging is inevitable to some extent, but it is not uniform; it affects everyone, or every brain, differently. In less than two decades, it is estimated that the number of people over the age of 65 years is projected to outnumber those under 18 years old for the first time in U.S. history [54]. Unfortunately, common neurological disorders affecting older adults, such as Alzheimer's, strokes, neuropathy, cerebrovascular disorders, and Parkinson's disease, are expected to increase commensurately [55, 56]. Consequently, the demand for precise biomarkers to assist in distinguishing neurodegenerative diseases from the consequences of normal aging is growing.

In the past few years, high-resolution magnetic resonance imaging (MRI) has provided the opportunity to measure brain alterations associated with aging. It has been demonstrated that aging is associated with a decrease in whole-brain volume [57] and essential to gray matter (GM) volume in the temporal, frontal, and parietal regions of the brain. Moreover, aging affects GM deficits in the insula, cerebellum, basal ganglia, and thalamus. Although there is growing evidence showing an extensive decrease in several WM structures of the brain with aging [58, 59, 60], there are some reports proving that aging leads to an increase in WM of the brain [61, 62].

Ge et al. [63] discuss possible standard patterns of global GM and WM volume across adult life in healthy individuals from 20–86 years. To adjust for variation in the head size, it uses normalized volumes as a percentage of intracranial volume (%GM and %WM). The results of this study indicate that GM volume loss seems to be a steady, linear function of age throughout adulthood. However, WM volume loss appears to be delayed until middle adulthood, and Both seem independent of gender. Ziegler et al. [64] shows similar findings. Salat et al.[65] demonstrate that prefrontal white matter volume reduction in old age is much more significant than grey matter volume decline, especially in the ninth decade, which shows that white matter is more susceptible to normal aging.

The results of [66] indicate that the frontal lobe suffers more significant volumetric reductions with aging than the temporal lobe. Age-related declines in both regions were larger in males than in women, indicating that sexual dimorphisms in human neuroanatomy are not fixed and continue to alter throughout maturity. In contrast to [67, 68] reporting static white matter volume after adolescence, the finding in [69] suggests that white matter volumes in the frontal and temporal lobes grow until the fifth decade. After that, they start to decline.

Several studies [70, 71, 72] have associated Alzheimer's disease (AD) and ageing. Experiments in [73] show that hippocampal volume in clinical AD patients was 15–40% smaller than in healthy controls (HC). Moreover, [74] reports annual hippocampal atrophy rates 4.66% for AD patients, and 1.41% for healthy control group. However, hippocampal atrophy is known to lack specificity for AD, as it can occur in dementias other than AD, such as vascular dementia, semantic dementia, Parkinson's dementia, and frontotemporal lobar degeneration [75]. In addition, [76, 77] show significant volume shrinkage of the thalamus in mild cognitive impairment (MCI) and in AD patients.

One of the best cases of age-related disease is Parkinson's disease (PD) [78]. Authors in [79] observe a considerable grey matter atrophy bilaterally in the occipital lobe of Parkinson's disease patients with dementia (PDD) compared with PD patients. In addition, considerable temporal lobe atrophy was found in Alzheimer's disease compared to PDD, including the hippocampus and parahippocampal gyrus. Moreover, impairments in the basal ganglia have been associated with PD [80].

Azevedo et al. [81] identifies top-ranked brain regions affected by Multiple Sclerosis (MS)-specific atrophy, which are all grey matter structures. The results show that in thalamic and whole-brain, the rate of MS-specific atrophy may decrease, and the contribution of normal aging may increase with age, whereas age does not affect the putamen and caudate atrophy associated with normal aging or MS.

The author in [82] shows that women are less vulnerable to age-related changes in brain size. However, other reports present the opposite results [83, 84]. Additionally, a few studies have examined the impact of gender differences on the aging-related shrinkage of brain subregions in a small dataset. Therefore, the results are still controversial, and more specific and comprehensive studies are needed to clarify these changes better.

A recently published paper in Nature [85] reports that the results of most published brain-wide association studies are unreliable because they involved too few participants. Also, one of the tough challenges for all researchers in this domain is addressing inter-individual variation in brain morphology and head size which complicates volumetric brain imaging studies, and the decision as to whether to use raw or intracranial volume (ICV)-adjusted volumetric data will impact the understanding of any statistical analysis carried out in conjunction with a morphometric study [86]. Hence, we use a large healthy dataset for the analyses to address these

limitations and clarify the effect of age and gender on regional and global brain volumes.

In the present MRI study, to clarify the effect of gender on age-related changes in brain sub-regions by MR imaging, we investigated the association between whole and regional volumetric alterations with increasing age in a large group of healthy subjects (n=6,739, age range: 30–80). For volumetric analysis, we used the FastSurferCNN [87] network to extract quantified whole and regional brain volumes in a total of 95 classes: frontal lobe, temporal lobe, putamen, precentral, insula, cuneus, hippocampus, amygdala, etc.

To model the influence of gender and age on brain volumes (i.e., whole and regional brain volumes), we adjusted volumetric measures by dividing the volumes of all the individual brains by the intracranial volume (ICV) of the respective subjects and performed a regression analysis.

Many age-related neural disorders, such as Alzheimer's disease [88] and Parkinson's disease [89], may benefit from studies of volumetric changes in large populations, which may help in understanding the mechanisms underlying brain atrophy and monitoring the clinical treatment effects of these diseases.

2.2 Background

2.2.1 What is MRI?

A magnetic resonance imaging (MRI) scan is a non-invasive procedure that creates high-quality images of the organs and structures found inside the body. In order to get such precise images, magnetic resonance imaging (MRI) uses a powerful magnet, radio waves, and a computer. It does not make use of X-rays in its procedures [90], because it does not use X-rays or any other form of radiation, magnetic resonance imaging (MRI) is the imaging test of choice when people will need regular imaging for the diagnosis or treatment monitoring of their brain, in particular [90].

2.2.2 How Does MRI Work?

MRIs use protons found in high concentrations within the human body. Every proton has a spin, resulting in a slight magnetic charge. When a significant magnetic field is present in an MRI machine, the protons will align themselves with the magnetic field. After that, the MRI operator will inject a radiofrequency pulse that will force the protons into realigning themselves at either 90 or 180 degrees with respect to the static magnetic field. When the radio frequency pulse that pushed the proton against its natural state is switched off, the protons realign themselves with the magnetic field as they were intended to be in the first place, and they release electromagnetic energy along the way. The MRI may detect this energy, and different types of tissues can be distinguished from one another based on the rate at which they release energy once the pulse is switched off [90]

2.2.3 Some Basic Terminology

Anatomical Coordinate System

The anatomical coordinate system is made up of three planes that describe a human's standard anatomical position:

- Sagittal: the sagittal plane is a longitudinal plane that divides the body into two parts, right and left.
- Coronal: The coronal plane is a longitudinal plane that divides the body into anterior (front) and posterior (back) sections.
- Axial: The axial (or transverse plane) is a horizontal plane that divides the body into superior (upper) and inferior (lower) sections.

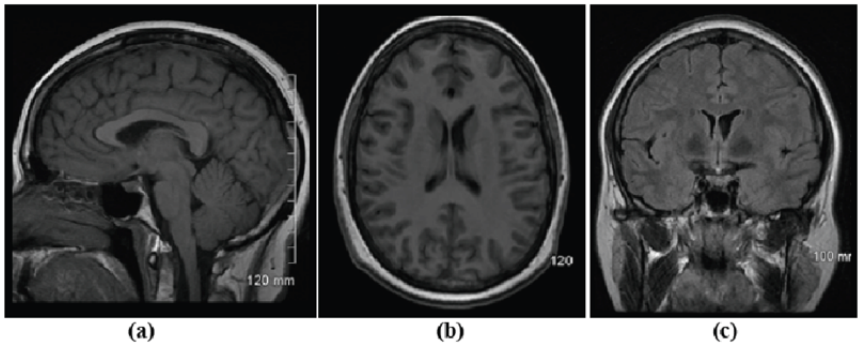

FIGURE 2.1: MRI planes for MRI head scan (a) sagittal, (b) axial and (c) coronal [91]

Voxel-Based Morphometry (VBM)

Each MRI scan provides details about a physical volume in three dimensions. The spatial resolution of an MRI image is determined by the size of its voxels. In computer graphics, a voxel is the smallest data unit that may be used to represent a point in three-dimensional space. Like a pixel in a 2D digital image, a 3D voxel only has one value. The spatial resolution of the data can be inferred from the voxel size, with a higher resolution corresponding to smaller voxels. Structural MRI typically uses voxels 1 mm 1 mm x 1 mm in size.

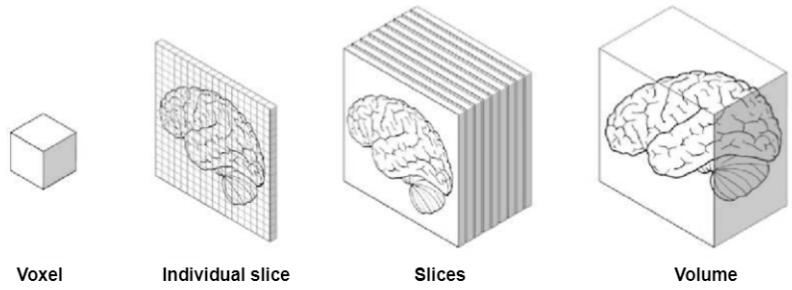

FIGURE 2.2: The basic concept of MRI image acquisition, Voxels are gathered into slices of varying thickness, which are subsequently gathered into a volume that contains the complete sample[92].

Comparing the amount of tissue in a specific anatomical region is a typical method for studying individual differences in brain structure. Structural MRI imaging can be done objectively using a technique called voxel-based morphometry (VBM).

Total Intracranial Volume (ICV)

Intracranial volume is the space occupied by the brain, meninges, and cerebrospinal fluid (CSF) [93].

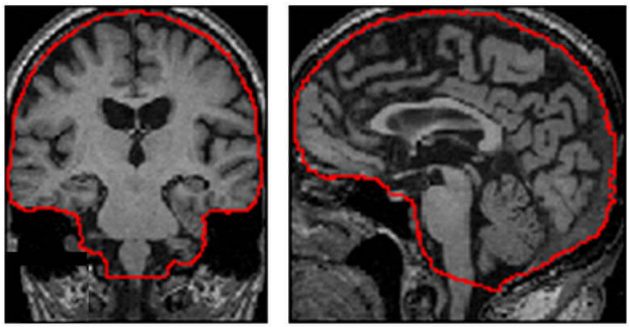

FIGURE 2.3: A segmentation example using the intracranial mask on the coronal and sagittal planes

The ability to compare the volumes of patients' brain structures to healthy controls' brain structures is a powerful tool for helping physicians assess the possibility of neurodegeneration or the progression of a neurodegenerative disorder.

For volumetric analysis, to account for variations in brain size, sex, and head size across a population, the volumes of brain structures are normalized (or scaled) to ICV because ICV is a very stable measurement across the lifetime of a patient.

2.2.4 FastSurferCNN Network

FastSurfer is a time-efficient alternative to traditional deep-learning-based methods for analyzing structural MRI automatically. FastSurferCNN can divide an entire brain MRI image into 95 classes (not counting the background) in less than 1 minute when using a GPU (or about 14 minutes when using a CPU). FastSurferCNN is made up of a series of F-CNNs that operate on coronal, axial, and sagittal 2D slices and a final view aggregation step. All three F-CNNs share the same fundamental structure, depicted in Figure 2.4, which consists of a series of 4 dense encoder and decoder blocks separated by a bottleneck layer. To aid in data recovery and expand network connectivity, they incorporate two enhancements into FastSurferCNN: competitive dense blocks and spatial view aggregation. These components will be broken down and elaborated upon in the next section [87]

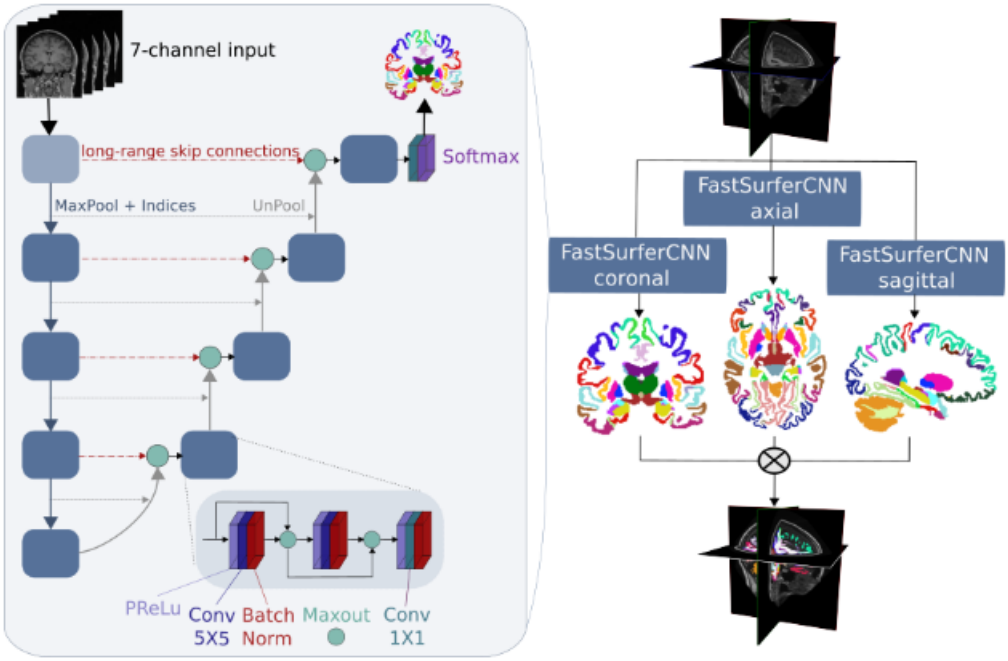

FIGURE 2.4: FastSurfer Network Architecture [87].

Competitive Dense Block

At each stage, before transferring the inputs to the side output layers (from left to right), we do a Maxout operation instead of the standard concatenation block. Maxout activation can significantly reduce the number of parameters compared to the classical dense blocks. Instead of stacking the output of previous layers at each stage on top of each other, we only keep the maximum value at each position by inducing competition between feature maps and accelerating network convergence. (Figure 3.6).

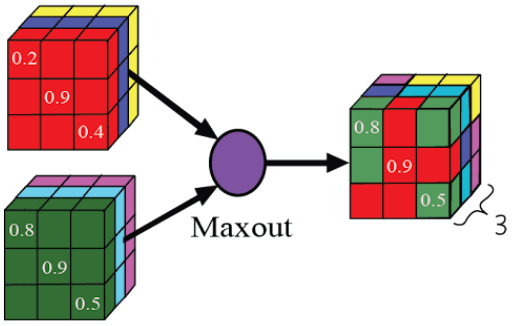

FIGURE 2.5: Visual Representation of Maxout Layer [87]

2.2.5 View Aggregation

Since the brain has a complex 3D structure, it is necessary to train one F-CNN per anatomical plane and aggregate the views into a single output. Each network, thus, learns the anatomical representation of the brain structures in the coronal, axial, or sagittal view, depending on the orientation of the 2D slices. Each model's probability

maps are aggregated using a weighted average to get the final segmentation. In addition, view aggregation acts as a regularizer to reduce erroneous predictions [87].

2.3 Methods

2.3.1 MRI Acquisition and Participants

We used an in-house dataset containing 6739 MRI scans and related information for healthy individuals. The MRI scans were acquired from different scanners. Figure 2.6 and 2.7 show the age distribution of the samples used in this study (n=6739).The voxel size is 1mm x 1mm x 1mm. Demographic details of the cohort are presented in Table 2.1. Further, we compare two independent male and female populations to ensure the difference in the means is insignificant.

TABLE 2.1: Demographic Information

Variable	N	Mean	SD	Difference(Female - Male)	$P-value$
Female	3359	53.900268	11.688655	0.3103	0.2808
Male	3380	53.589941	11.929353		

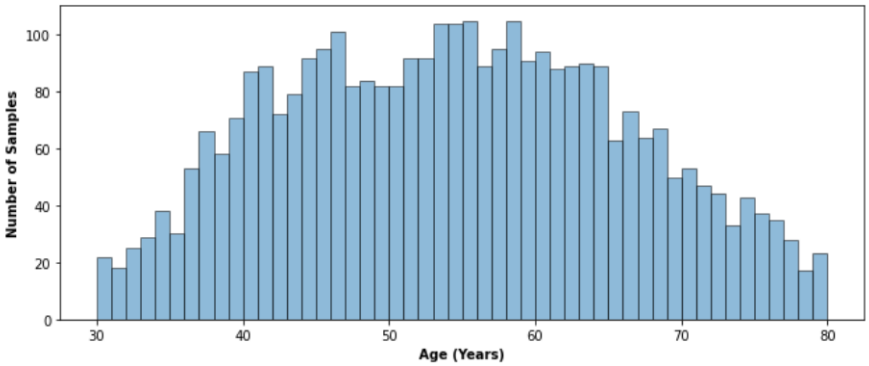

FIGURE 2.6: Histogram showing the age distribution for the female samples used in this study (n=3359)

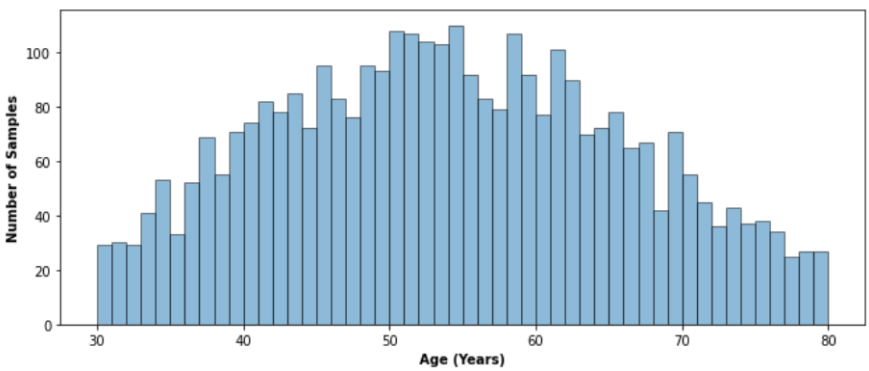

FIGURE 2.7: Histogram showing the age distribution for the male samples used in this study (n=3380)

2.3.2 MRI Pre-Processing

Figure 2.8 shows the overall pipeline of our methodology. For the analyses, we used the FastSurfer segmentation network [87] to generate the respective regional volumetric variables from 3D T1-weighted MRI scans of all of the subjects. We select a total of 95 regions of the brain for the segmentation. FastSurfer [87] is a fast and substantially verified deep learning pipeline that can analyze structural MRIs of the human brain in a completely automated fashion. As a result, it produces segmented output compatible with FreeSurfer and enables scalable big-data analysis as well as time-sensitive clinical applications such as the localization of structures during the process of image acquisition or the extraction of quantitative measures. FastSurfer-CNN is composed of three fully convolutional neural networks that operate on coronal, axial, and sagittal 2D slice stacks and a final view aggregation that combines the advantages of 3D patches and 2D slices. FastSurfer trained over 134 participants, age range 27-66. An example of the output of the FastSurfer model is shown in Figure 2.9

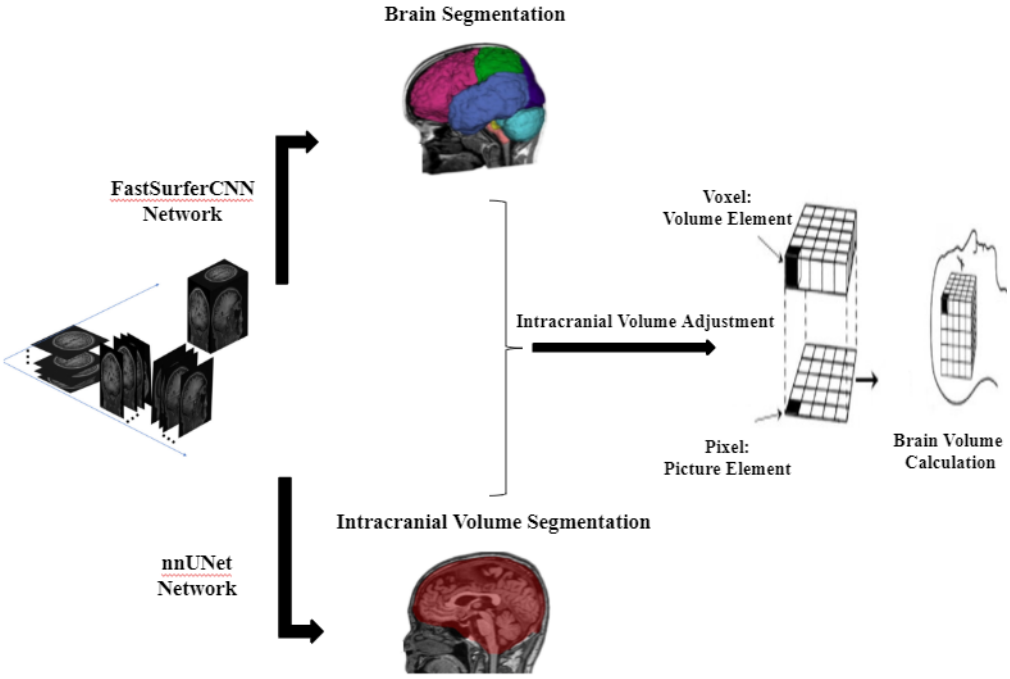

FIGURE 2.8: This diagram shows the process flow of our methodology. First we use FastSurferCNN and nnUNet networks to generate a segmentation mask and ICV, respectively. Then, the normalized brain volume is computed by dividing the segmentation mask by ICV. Finally, we count the number of voxels in each brain region to get the regional brain volume. [94]

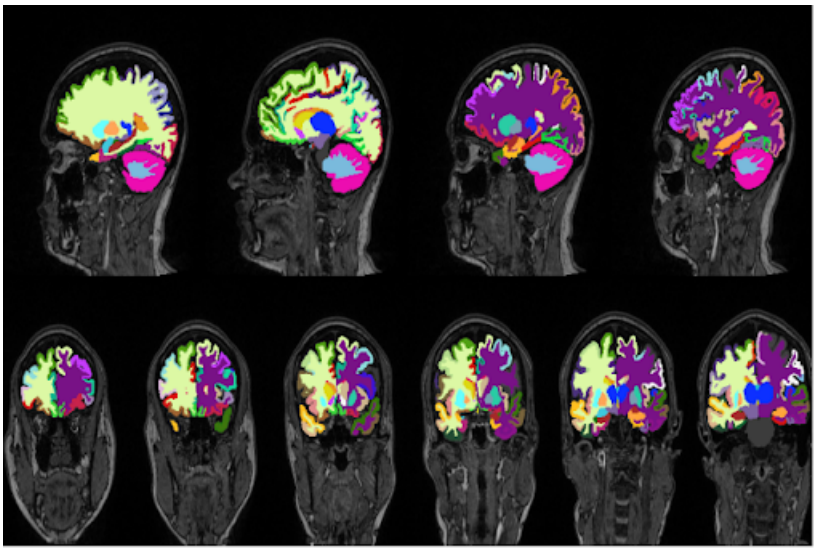

FIGURE 2.9: An example of the FastSurfer segmentation model.

As subjects have different head sizes, intracranial volume (ICV) is employed as a control variable in the partial correlation between age and various ROIs to reduce interindividual variation between patients. To estimate the total intracranial volume, we used 60 subjects and annotated the intracranial part of them according to Figure 2.10. We used these labeled data to train the nnUNet [95] for intracranial mask generation. nnUNet is a self-configuring method for deep learning-based medical image segmentation. It includes preprocessing, network architecture, training, and post-processing for any new task.

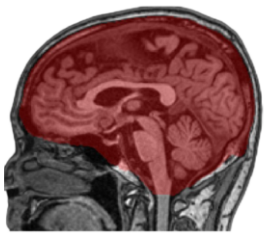

FIGURE 2.10: An example of intracranial volume (ICV) mask segmented by nnUNet model. Visualization overlay on raw MRI image was done in MITK tool [96].

We investigate various regression models (i.e., linear, quadratic, and cubic) to identify the best-fitting model to evaluate the relationships between the global and regional changes in brain volume with aging. We select the best regression model with the high R^2 and have equally distributed residuals around zero to display the data. Also, for regression analysis, we use normalized GM, WM, etc., values by dividing the individual brains' volumes by the ICV of the respective subjects. We also used ICV as a covariate to find partial correlations between the proportional compartmental volumes and age and sex.

2.4 Results

This section summarises the effect of age and gender on global and regional brain volumes. This can help distinguish brain disorders such as Alzheimer's, multiple

sclerosis, epilepsy, and traumatic brain injuries from the consequences of normal aging. Because of commercial issues, we can not include results related to the percentiles and reduction rate per region per age.

Comparison of Brain Volumes Between Sexes

According to Table 2.2, males have considerably bigger ICV, WM volume, and GM volume values than females. However, after adjusting for ICV, there is statistically significant evidence that women had more GM volume. Also, when the data were normalized using ICV, the differential in WM between males and females disappeared.

TABLE 2.2: ICV, WM, and GM volumes before and after the use of ICV adjustment method.

Variable	Female Mean ±Std	Male Mean±Std	Difference (Male - Female)	P-value
ICV	1773.310015±110.409040	2029.264389 ±130.111684	255.9544	<0.00001
GM	514.166756 ±38.733265	570.451695 ±45.719739	56.2849	<0.00001
WM	466.949521 ±45.777167	528.887977 ±53.072465	61.9385	<0.00001
nGM	0.290138 ±0.016110	0.281057 ±0.017693	-0.0091	<0.00001
nWM	0.263212 ±0.018371	0.260283 ±0.019357	-0.0029	<0.00001

Regression Analysis

As tables 2.3 and 2.4 indicate, there is a linear negative association between normalized and raw volume of cerebral, cerebellum GM/WM and age. In addition, our data reveal that in the regions listed in the tables, the volume of males is decreasing more quickly than the females. WM and GM shrink faster than the cerebellum regions.

Brainstem shows a similar pattern of results in [97], which shows that both males and females have stable brainstem volumes throughout adulthood.

TABLE 2.3: The regression results between age and listed of raw volumes

Variable	Gender	R^2 Linear	R^2 Quadratic	R^2 Cubic
Cerebral_GM	Female	0.23424	0.23487	0.23577
	Male	0.25145	0.25333	0.25333
Cerebellum_GM	Female	0.07044	0.07227	0.07285
	Male	0.11822	0.11931	0.12079
GM	Female	0.23654	0.23757	0.23859
	Male	0.27316	0.27479	0.2748
Cerebral_WM	Female	0.06871	0.08575	0.0858
	Male	y 0.07415	0.09793	0.09859
Cerebellum_WM	Female	0.08836	0.10668	0.10683
	Male	0.10958	0.12985	0.12991
WM	Female	0.07358	0.09158	0.09162
	Male	0.08347	0.10494	0.10606
Brainstem	Female	0.01923	0.02901	0.02901
	Male	0.0236	0.03526	0.03544

TABLE 2.4: The regression result between age and listed of normalized volumes by ICV

Variable	Gender	R^2 Linear	R^2 Quadratic	R^2 Cubic
Normalized_Cerebral_GM	Female	0.33785	0.33827	0.34143
	Male	0.34891	0.34903 0.34903	0.34926
Normalized_Cerebellum_GM	Female	0.05201	0.05204	0.05331
	Male	0.11275	0.11276	0.11595
Normalized_GM	Female	0.33981	0.34008	0.34374
	Male	0.3717	0.37178	0.3723
Normalized_Cerebral_WM	Female	0.09525	0.11363	0.11419
	Male	0.10986	0.13961	0.13963
Normalized_Cerebellum_WM	Female	0.08629	0.10127	0.10131
	Male	0.11337	0.1298	0.13002
Normalized_WM	Female	0.10295	0.12265	0.12314
	Male	0.12224	0.15062	0.15098
Normalized_Brainstem	Female	0.01125	0.01672	0.01682
	Male	0.02062	0.02857	0.02994

Tables 2.5 and 2.6 show the results of lobes analysis for raw and normalised volume values, respectively. Our data suggests that frontal lobe shrinks faster than other lobes, and the occipital and parietal lobes shows a non-significant age-related volume change.

TABLE 2.5: The regression result between age and listed of lobes

Variable	Gender	R^2 Linear	R^2 Quadratic	R^2 Cubic
Frontal_Lobe	Female	0.26244	0.26251	0.26304
	Male	0.26412	0.26437	0.26471
Temporal_Lobe	Female	0.1948	0.19684	0.19796
	Male	0.21756	0.22095	0.22121
Parietal_Lobe	Female	0.18398	0.18422	0.18472
	Male	0.18623	0.18847	0.18878
Occipital_Lobe	Female	0.0606	0.06895	0.07032
	Male	0.07949	0.08503	0.08504

TABLE 2.6: The regression result between age and listed of normalized lobes by ICV

Variable	Gender	R^2 Linear	R^2 Quadratic	R^2 Cubic
Normalized_Frontal_Lobe	Female	0.33421	0.33726	0.33896
	Male	0.34055	0.34126	0.3413
Normalized_Temporal_Lobe	Female	0.25545	0.25557	0.25886
	Male	0.28956	0.29031	0.29094
Normalized_Parietal_Lobe	Female	0.23321	0.2338	0.2355
	Male	0.24275	0.24298	0.24373
Normalized_Occipital_Lobe	Female	0.05173	0.05586	0.05861
	Male	0.07943	0.08142	0.08151

Correlation Analysis

Table 2.7 and 2.8 show the Pearson correlation results between age and global regions and lobes, respectively. Except for the occipital lobe, the results reveal a considerably high correlation between age, GM, and lobes. Furthermore, the findings show that the correlations of GM and WM volumes with age are not different between males and females.Table 3.3 in the Appendix provides the correlation analysis for 95 sub-regions.

TABLE 2.7: The Pearson partial correlations result between age and listed of global regions controlling for the effect of ICV

Variable	Gender	r	CI95%	P-value
GM	Female	-0.605	[-0.63, -0.58]	<0.00001
	Male	-0.62163	[-0.64, -0.6]	<0.00001
WM	Female	-0.313	[-0.34, -0.28]	<0.00001
	Male	-0.34638	[-0.38, -0.32]	<0.00001

TABLE 2.8: The Pearson partial correlations result between age and listed of lobes controlling for the effect of ICV

Lobe	Gender	r	CI95%	P-value
Frontal_Lobe	Female	-0.58578	[-0.61, -0.56]	<0.00001
	Male	-0.59021	[-0.61, -0.57]	<0.00001
Temporal_Lobe	Female	-0.51347	[-0.54, -0.49]	<0.00001
	Male	-0.54532	[-0.57, -0.52]	<0.00001
Parietal_Lobe	Female	-0.48767	[-0.51, -0.46]	<0.00001
	Male	-0.49554	[-0.52, -0.47]	<0.00001
Occipital_Lobe	Female	-0.24304	[-0.27, -0.21]	<0.00001
	Male	-0.28872	[-0.32, -0.26]	<0.00001

2.5 Conclusion

We used magnetic resonance imaging (MRI) scans and other information from a group of 6739 healthy people between the ages of 30 and 80 to figure out how age and gender affect changes in whole and regional brain volumes. The results show that a linear line is the most suitable model for describing the patterns of global and regional brain changes that occur with advancing age. Also, there are significant differences in the volume and shrinkage rate of brain structures between men and women.

. We provide normative values for 95 regions of the brain and total grey/white matter volumes as a function of age for use in research and clinical studies. These norm values, in conjunction with statistical analysis such as percentiles, can provide a quick and objective assessment of the patient's brain volume status.

Currently, few large-sample studies have evaluated brain volume variations between sexes during the aging process and have specifically investigated participants over 50. The unadjusted brain volumes of GM and WM were found to be more prominent in men in most of these studies [98, 99, 86]. Our finding in table Table 2.2 demonstrates the same result; However, after normalization by head size (ICV), the evidence indicates that women have considerably more GM volume than men. This needs to be studied more in terms of hormones and molecular genetics to figure out why and how this happens[98].

Our present findings contribute to understanding how structures in the human brain change over the adult years and will help address the pathological age-related neural changes in age-related neural disorders such as Parkinson disease and Alzheimer disease. While the current findings are based on cross-sectional data, a longitudinal study will provide valuable insight into the role of brain atrophy in aging.

Chapter 3

LEON: LightWeight Edge Detection Network

3.1 Introduction

Edge detection is the process of finding meaningful transitions in an image. This is done by detecting discontinuities in texture, color, brightness, etc. Edges provide boundaries between different regions in the image. Detecting these boundaries is the first step in many computer vision tasks such as edge-based face recognition, edge-based target recognition, scene understanding, image segmentation, object proposal, and object detection [100].

Over the past few years, deep learning techniques have greatly promoted edge detection research. Traditional approaches to the BSDS500 dataset often achieve a 0.59 ODS F-measure. On the other hand, deep methods can achieve 0.828 ODS [100]. Although recently proposed architectures achieve high accuracy, they are computationally inefficient. This makes developing lightweight networks that reduce the number of parameters while maintaining the detection accuracy critical. Figure 3.1 shows several well-known deep learning-based methods' detection accuracy and complexity (model size). As shown in Figure 3.1, the orange dot indicates how well our model matches human perception in terms of accuracy with a few parameters.

Many deep learning-based edge detectors use Visual Geometry Group (VGGNet) [101] as their feature-based extractor because of its excellent performance. However, VGGNet has a pretty extensive backbone and employs many parameters, making it appropriate to fit more complex tasks such as image segmentation and object recognition. This work is motivated by the fact that edge detection is a low-level image processing task and does not need complex networks for feature extraction.

To decrease the number of parameters and floating point operations (FLOPs), we take advantage of depthwise separable convolutions [102] which disentangles the spatial and channel interaction that is mixed in a regular convolutions operation. However, it reduces the performance in comparison to conventional convolution. To compensate for the reduced performance, we increase the receptive field by carefully choosing proper lightweight components for edge detection. We explain the details in section 3.

The rest of this paper is organized as follows. Section 2 reviews related works. The proposed network architecture is described in section 3. Section 4 presents experimental results and compares them to the state-of-the-art edge detector networks using (Berkeley Segmentation Dataset 500) BSDS500 [103] and NYUDv2 [104] datasets. In section 5, we offer concluding remarks and discuss future research directions.

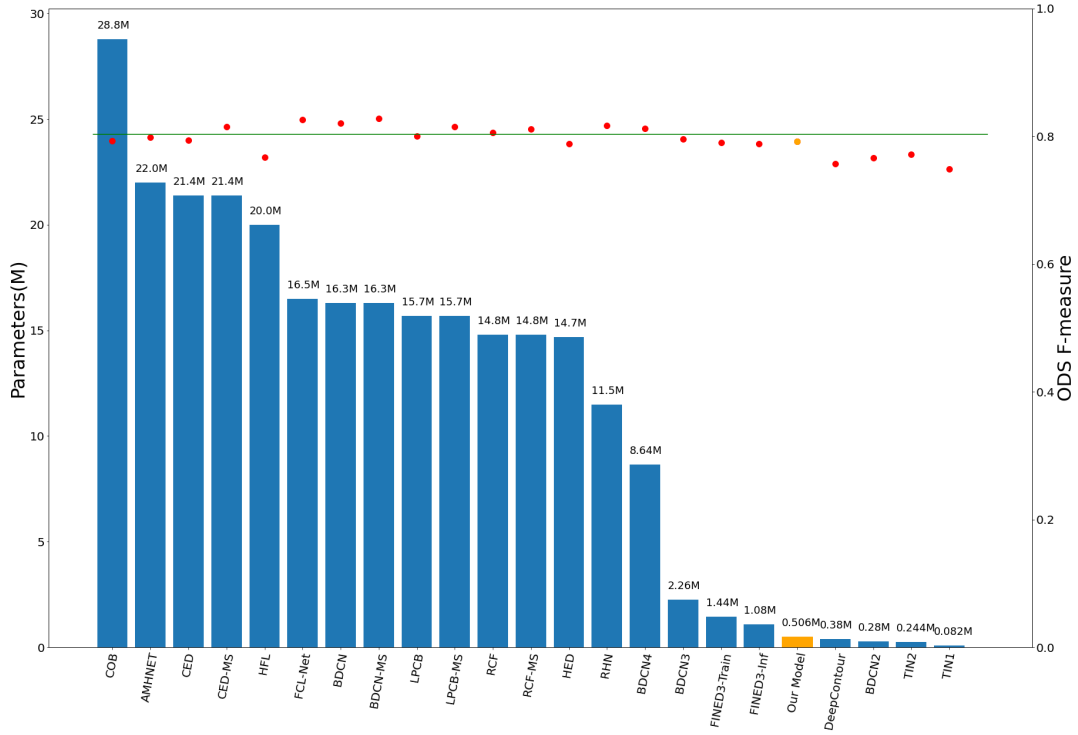

FIGURE 3.1: Comparison of complexity and accuracy performance among various edge detection schemes. Our proposed methods (orange).

3.2 Related Work

Over the past few years, a number of edge detection solutions have been developed. Almost all edge detection approaches can be generally categorized into three groups, traditional edge detection, learning-based ones using handcrafted features, and deep learning networks. In the following paragraphs, we review some techniques developed in recent years.

Intensity and color gradients were the main focus of early pioneer edge detection methods. The Sobel [105] operator measures the 2-D spatial gradient of an image, emphasizing regions of high spatial frequency that correspond to edges. The Canny algorithm [106] is a multi-stage edge detector. In this algorithm, the intensity of the gradients is computed by employing a filter based on the derivative of a Gaussian. The Gaussian filter reduces the impact of image noise. Subsequently, removing non-maximum pixels of the gradient magnitude decreases possible edges to 1-pixel curves. Finally, edge pixels are kept or eliminated by applying the hysteresis threshold on the gradient magnitude. Zero-crossing theory based algorithms proposed by [107, 108]. Traditional approaches suffer from some limitations, including merely focusing on local intensity changes while failing to recognize and remove the non-edges texture.

The introduction of learning-based edge detectors made it possible to partially overcome challenges such as texture detection problems in traditional approaches.

In this group of detectors, handcrafted features are initially extracted. Later, classifiers trained using these features are applied to identify edges. Konishi et al. [109] proposed the first data-driven approach, which used images to learn the probability distributions of responses corresponding to the two sets of edge filters. In another work [110], random decision forests were applied to show the structure presented in local image patches. The structured forest uses color and gradient features for high-quality output edges.

The techniques mentioned above are developed according to handcrafted features, which mostly fail to provide high-level information for semantically meaningful edge detection and have a limited capability of capturing edges at different scales. Several convolutional neural network (CNN) algorithms with strong learning capabilities have been proposed to address these issues in recent years. One of the most influential DNN-based edge detection is HED[111]. This study uses fully convolutional neural networks and deeply-supervised nets to find the edge probability for every pixel. HED uses VGGNet [101] for the feature extraction and fuses all side outputs of VGGNet features to minimize the weighted cross-entropy loss function. Since then, various extensions based on HED and VGGNet have been developed, including CED [112],AMH-Net [113], RCF [114], LPCB [115], and BDCN [116].

While CNN is a very successful model, it often requires high computational power and resources. Hence, the current trend is to design efficient CNN structures that overcome such issues. Fined [117], dense extreme inception network [118], TIN[100] have proposed a lightweight architecture for edge detection. Although these networks are light and fast, they have a low detection accuracy. To achieve a better trade-off between accuracy and efficiency for edge detection, we need to optimize the architecture and initial parameters of deep learning models to consume fewer resources while maintaining accuracy. In this paper, we build our model by simplifying the backbone for feature extraction and carefully choosing the proper components. Therefore, we achieve good edge quality with a much simpler model compared to other studies.

3.3 Lightweight Edge Detection Network

In this section, we Introduce LEON. LEON is a light neural network with high running efficiency. It can address the inefficiency of the models outlined in the previous section at once. In Figure 3.2 we present LEON's architecture. We train the network from scratch. Below we review the components used by LEON.

3.3.1 Efficient Backbone

Most deep learning-based edge detectors [112, 113, 114, 115, 116] employ VGGNet as their feature extraction backbone. However, we believe that edge detection is a simple task and does not need to have an extensive backbone. We reduce the backbone's complexity while maintaining efficiency by using lightweight components. To resemble the pyramid structure, we stack up three stages and use a max-pooling operation for downsampling the features between them. As the stage number increases, the dimension of output feature maps decreases, and the feature channel number (the number of filters) increases. Stages 1, 2, and 3 have channel numbers 16, 64, and 256, respectively. The backbone is mainly composed of deformable and customized depthwise separable convolutions. To create the fused output, we use

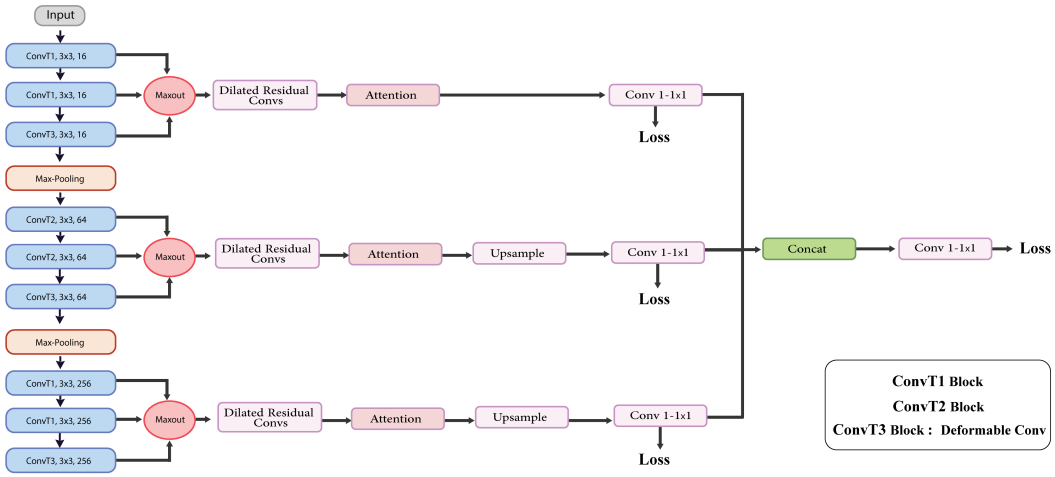

FIGURE 3.2: LEON architecture

FIGURE 3.3: Convt1 block (left) & Convt2 block (Right)

the standard bilinear interpolation to upsample the low-resolution features. Then, we combine all the stage outputs to form the fused output. We next elaborate on the layers and components used by LEON in detail.

Deformable Convolution

Geometric transformation and variations widely existing in natural images make feature extraction tasks challenging. Standard convolution kernels have a fixed structure and have limitations in capturing geometric transformations. Deformable convolutions can address this issue efficiently. This convolution type can change its kernel shape and parameters to adapt to the image content. This adds 2D offset kernels to the regular sampling location in the standard convolution, which enables the network to have different receptive fields according to the scale of the objects. These 2D offset kernels are learnable from the preceding feature maps using additional convolutional layers and can be trained end-to-end using normal back-propagation functions. We add this module at the end of each stage to keep our network light regarding parameters and computation. We can strengthen our features before transferring them to the next stage [119].

Depthwise Separable Convolution

Conventional convolution performs the channels and spatial-wise computation in one step. In contrast, Depthwise Separable Convolution reduces the number of parameters by splitting the computation into two steps: 1) depthwise convolution, which applies a single convolutional filter per input channel, and 2) pointwise convolution, which creates a linear combination of the output of the depthwise convolution [102]. This approach, however, degrades accuracy. To address this problem, we

strengthen the features by using additional side blocks while keeping the number of parameters as low as possible.

We use RELU activation after each pointwise convolution to add non-linearity to the model for making complex decisions (Figure 3.3 - Convt1).To increase the model's accuracy while keeping the number of parameters low, we modified Convt1 to Convt2 by adding pointwise convolution, which uses only 1×1 kernel to iterate through every single point between two RELU activations. In addition, to overcome the overfitting problem, after each ReLU activation, we employ a batch normalization technique as a regulariser.

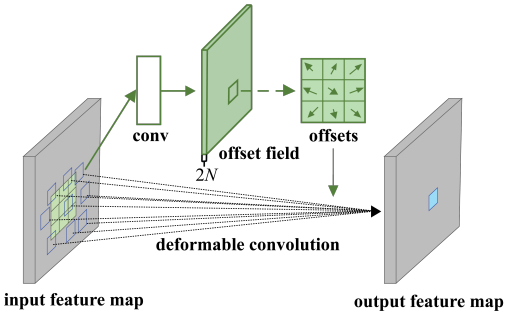

FIGURE 3.4: Visual Representation of deformable Convolution Operation [119]

3.3.2 Efficient Side Structure

MAXOUT Layer

At each stage, before transferring the inputs to the side output layers (from left to right), we do a Maxout operation instead of the standard concatenation block. Maxout activation can significantly reduce the number of parameters compared to the classical dense blocks. Instead of stacking the output of previous layers at each stage on top of each other, we only keep the maximum value at each position by inducing competition between feature maps and accelerating network convergence. (Figure 3.6)

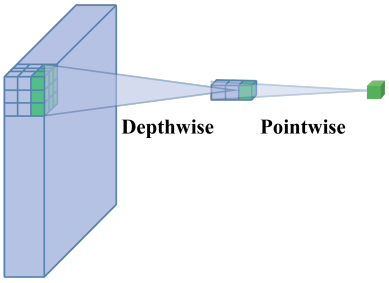

FIGURE 3.5: Visual Representation of depthwise separable convolution [102]

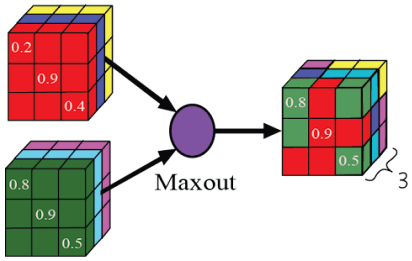

FIGURE 3.6: Visual Representation of Maxout Layer [87]

Dilated Residual Convolution Module

To enhance the extracted features by depth-wise separable convolution in the backbone, we connect every feature extraction layer to the dilated convolution module adopted in [100]. We use different dilation sizes to capture different levels of receptive fields in the image. The first dilation is 4, followed by 8, 12, and 16, and all the layers have 32 filters. After pixel-wise aggregation, we use hierarchical residual-like connections to improve the multi-scale representation ability at a more granular level. This block can be plugged into the state-of-the-art backbone without no effort. Figure 3.7 shows the design of DDR module.

Convolutional Block Attention Module (CBAM)

We use a lightweight spatial and channel attention module after the dilated residual convolution block to focus on the relevant features while diminishing the other parts [120].

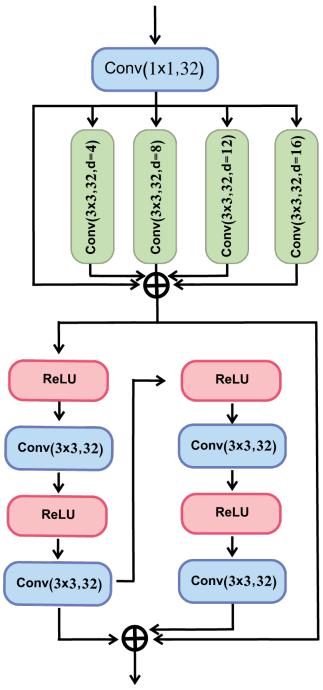

FIGURE 3.7: Visual Representation of dilated residual convolution module.

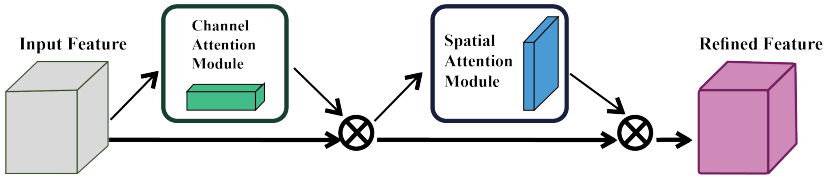

FIGURE 3.8: Visual Representation of attention module [120].

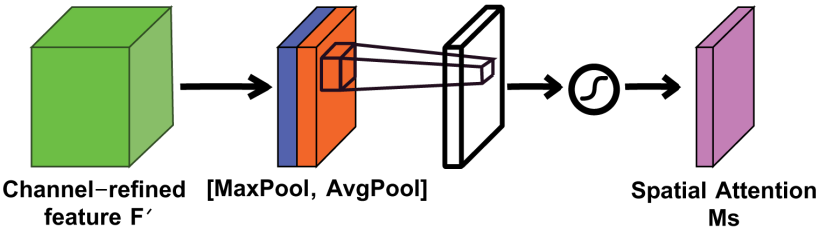

FIGURE 3.9: Visual Representation of spatial attention module [120].

The spatial attention extracts the inter-spatial relationship of features to find 'where' is an informative part of the image. To calculate this; we first apply average pooling and max pooling that summarizes the average presence of features and the most activated presence of a feature, respectively. Then, We use a convolution layer in addition to the concatenated feature descriptor to create a spatial attention map that specifies where to highlight or suppress features [120].

The channel attention block redistributes the channel feature responses to give higher importance to specific channels over others. In order to compute the channel attention, we squeeze the spatial dimension of the input feature map. [120].

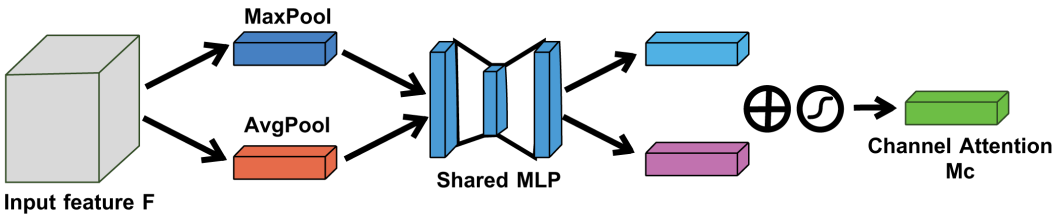

FIGURE 3.10: Visual Representation of channel attention module [120].

3.3.3 Loss Function

The edge and non-edge pixel data are not equally distributed in an image. CNN models can achieve pretty high accuracy by predicting the majority class, but they fail to capture the minority class. Unfortunately, this accuracy is misleading. We adopt the weighted Cross-Entropy loss function proposed in [114] to address this problem.

To train the network, we match all the stages and fused outputs to the ground truth. The following equation compares each pixel of each image to its label as [114].

$$L(x_i;W) = \begin{cases} \alpha.log(1-P(x_i;W)) & if y_i = 0 \\ 0 & if \leq y_i \leq \eta \\ \beta.logP(x_i;W) & otherwise \end{cases} \tag{3.1}$$

in which

$$\alpha = \lambda.\frac{|Y^+|}{|Y^+|+|Y^-|}$$
$$\beta = \frac{|Y^-|}{|Y^+|+|Y^-|} \tag{3.2}$$

$X, P(X), Y, W$, and , respectively, denote features extracted from the CNN network, the output of the standard sigmoid function, the ground truth edge probability, and all the parameters that will be learned in the CNN network, and the percentage of non-edge and edge pixels. The hyper-parameter is used to balance the number of positive and negative samples. Because multiple annotators label each image, and humans vary in cognition, the predefined threshold is used to distinguish between edge and non-edge pixels in the edge probability map. If a pixel is marked by fewer than of the annotators, it is considered a non-edge pixel. To generalize the loss function to all the pixels inside the image (I), at each stage (k) and fuse layer, we use the following equation [114]:

$$L(W) = \sum_{i=1}^{|I|}(\sum_{k=1}^{|K|} L(x_i^k;W) + L(x_i^{fuse};W)) \tag{3.3}$$

3.4 Experiment And Discussion

3.4.1 Implementation Details

We use PyTorch for implementation and initialize the stages of our backbone networks with Gaussian distribution with zero-mean and a standard deviation of 0.01. The learning rate starts from 0.01 and then is updated using a linear scaling factor, multiplying 0.1 for every two epochs. The optimizer is stochastic gradient descent, and the training process terminates at eight epochs. We conduct all the experiments on a single GPU, NVIDIA GeForce 2080Ti, with 11G memory.

3.4.2 Dataset

In order to have a fair comparison to other published works in table 3.1 and table 3.2, we evaluate our proposed network on the same Berkeley Segmentation (BSDS500) [103] and NYUDv2 [104] Dataset . BSDS500 consists of 200 training, 100 validation, and 200 test images. We combine the 200 training images and 100 validation images to create a training set. We adopt the data augmentation technique, same in RCF [114]. In addition, similar to RCF, we also added PASCAL VOC [121] dataset and its flipped images into our training set.

FIGURE 3.11: Example from BSDS500 dataset. From left to right: origin image and ground truth

The NYUD dataset comprises 1449 densely labeled pairs of aligned RGB and depth images (HHA). This dataset consists of video sequences from various indoor scenes captured by Microsoft Kinect's RGB and Depth cameras. It is divided into 381 training, 414 validation, and 654 testing images. Like RCF [114], we rotate the images and corresponding annotations to 4 different angles (0, 90, 180, and 270 degrees) and flip them at each angle.

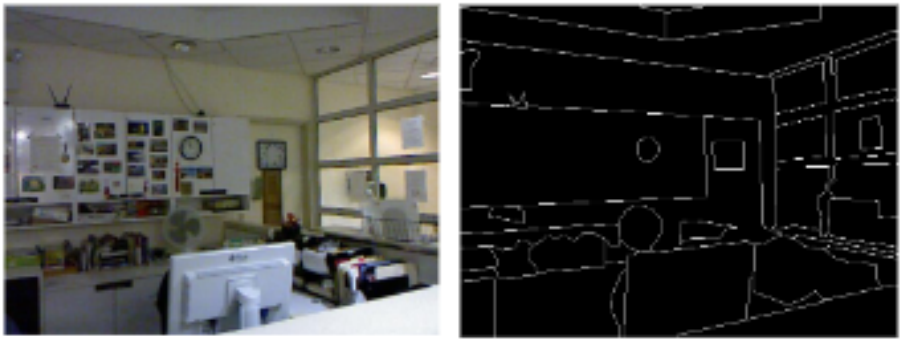

FIGURE 3.12: Example from NYUD dataset. From left to right: origin image and ground truth

3.4.3 Performance Metrics

Note that the share of edge pixels in each image is around 10%, whereas the share of non-edge pixels is 90%. Therefore, even when a model fails to predict edges, its accuracy is still 90%. As such, accuracy is a poor measure for evaluating imbalanced problems such as edge detection. Therefore, we use F-Score for the evaluation of our model. F-score combines the precision and recall of the model, where it reaches its best value at one and worst score at 0.

- Recall = TruePositives / (TruePositives + FalseNegatives)
- Precision= TruePositives / (TruePositives + FalsePositives)
- F-Measure = (2 * Precision * Recall) / (Precision + Recall)

We need a threshold to binarize the output of the CNN network to make it comparable to the ground truth (pixel-to-pixel), which is also binarized. There are two ways to compute the optimal threshold and corresponding F-score.

- Optimal Dataset Scale: Iterates over all possible thresholds and set one threshold for the entire dataset. The threshold that gives the best F-score for the dataset is used to calculate ODS score.
- Optimal Image Scale: Finds the best threshold and corresponding F-score for each image. The OIS F-score is calculated by averaging all of the F-scores for all images.

3.4.4 Comparison With State-Of-The-Arts

On BSDS500 dataset - We compare our methods in terms of F-score and number of parameters to prior edge detection approaches, including both traditional ones and recently proposed CNN-based models. According to Table 3.1 and Figure 3.13, we notice that our baseline model while using a significantly lower number of parameters, can even achieve outstanding results (ODS of 0.792 and OIS of 0.805) which are equal or better than most recent lightweight CNN models such as TIN1,TIN2 [100], BDCN2 [116], FINED3-Inf and FINED3-Train [117].

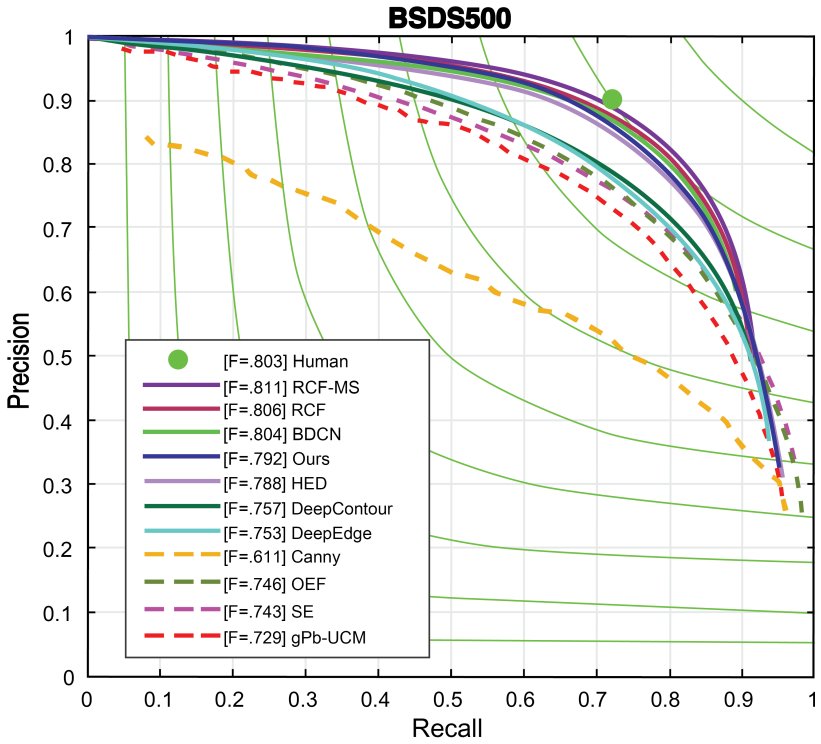

FIGURE 3.13: Precision-Recall curves of our models and some competitors on BSDS500 dataset.

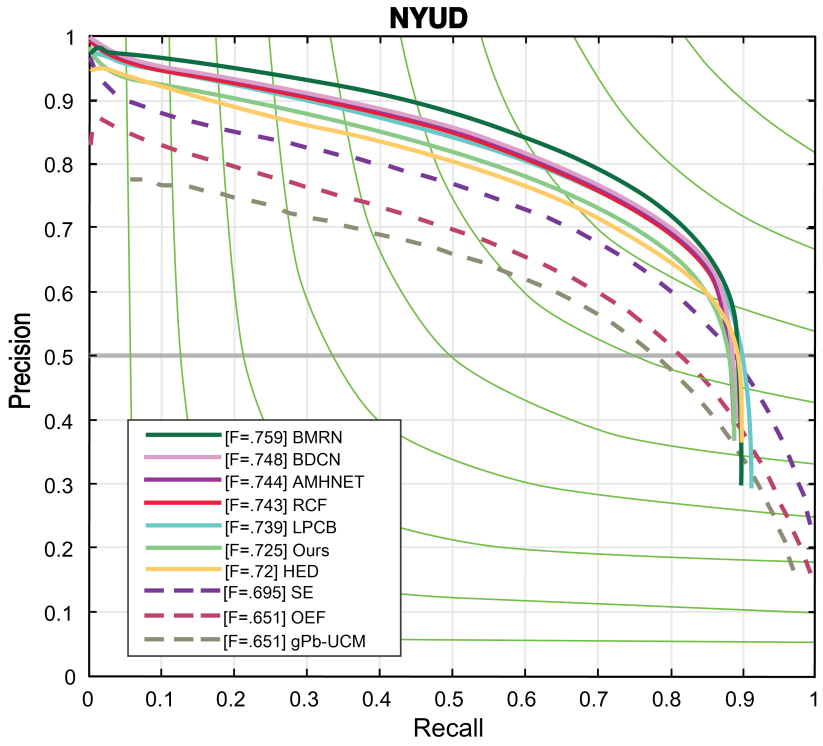

FIGURE 3.14: Precision-Recall curves of our models and some competitors on NYUD dataset.

On NYUD dataset - The comparison results on the NYUD dataset are illustrated in Table 3.2, and the precision-recall curves are depicted in Figure 3.14. For testing the model on NYUD, we use network settings similar to that used for BSDS500. Some studies use two separate models to train RGB images and HHA feature images of NYUD and report the evaluation metrics on the average for the outputs of the models. To evaluate results fairly, we contrasted our model's output with those of models only tested on RGB.

TABLE 3.1: Comparison with other methods on BSDS500 dataset.

Method	ODS	OIS	P(Million)
Canny [122]	0.611	0.676	-
OEF [123]	0.746	0.77	-
gPb-UCM [103]	0.72	0.755	-
SE [110]	0.743	0.763	-
AMHNET [113]	0.798	0.829	22
BDP-Net [124]	0.808	0.828	18.7
FCL-Nt [125]	0.826	0.845	16.5
LPCB [115]	0.815	0.834	15.7
BMRN [126]	0.828	0.81	+14.8
RCF [114]	0.806	0.823	14.8
HED [111]	0.788	0.808	14.7
COB [127]	0.793	0.82	28.8
RHN [128]	0.817	0.833	11.5
CED [112]	0.815	0.834	21.4
DeepEdge [129]	0.753	0.772	-
DeepContour [130]	0.757	0.776	0.38
BDCN [116]	0.82	0.838	16.3
BDCN2 [116]	0.766	0.787	0.48
BDCN3 [116]	0.796	0.817	2.26
BDCN4 [116]	0.812	0.83	8.69
TIN1 [100]	0.749	0.772	0.08
TIN2 [100]	0.772	0.792	0.24
FINED3-Inf [117]	0.788	0.804	1.08
FINED3-Train [117]	0.79	0.808	1.43
Our Model	0.792	0.805	0.506

TABLE 3.2: Comparison with other methods on NYUD dataset.

Method	ODS	OIS	P(Million)
OEF [123]	0.651	0.667	-
gpb-UCM [103]	0.632	0.661	-
SE [110]	0.695	0.708	-
SE+NG+ [131]	0.706	0.734	-
AMHNET [113]	0.744	0.758	22
BDCN [116]	0.748	0.763	16.3
LPCB [115]	0.739	0.754	15.7
RCF [114]	0.743	0.757	14.8
BMRN [126]	0.759	0.776	+14.8
HED [111]	0.72	0.734	14.7
Our Model	0.725	0.738	0.5

3.5 Conclusion

Most existing deep neural networks for edge detection use transfer learning from pre-trained models such as VGG16, which have many parameters and are trained for high-level tasks. Edge detection, however, has a simple set of features and does not require a large number of convolutional layers for feature extraction. Therefore, in this research, we introduced a new architecture that is lightweight and has state-of-the-art performance. Our network uses customized depth-wise separable and deformable convolutions to carry out edge detection. Besides, we use lightweight components to increase the receptive field of our model to produce high-quality edges. Our network architecture is extendable and can potentially be employed in other vision tasks such as s salient object detection and semantic segmentation.

Bibliography

[1] Helga Kolb. "Simple anatomy of the retina". In: (2012).

[2] John F Salmon. *Kanski's Clinical Ophthalmology E-Book: A Systematic Approach*. Elsevier Health Sciences, 2019.

[3] Ivana Beatrice Alberta, Salsha Alyfa Rahmani, et al. "RETINAL IMPAIRMENT ASSOCIATED WITH LONG-TERM USE OF RITONAVIR AMONG HIV PATIENTS: A SYSTEMATIC REVIEW FOR PRIMARY EYE CARE PRACTICE". In: *International Journal of Retina* 5.1 (2022), pp. 48–48.

[4] H Kenneth Walker, W Dallas Hall, and J Willis Hurst. "Clinical methods: the history, physical, and laboratory examinations". In: (1990).

[5] Samra Irshad and M Usman Akram. "Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy". In: *2014 Cairo International Biomedical Engineering Conference (CIBEC)*. IEEE. 2014, pp. 133–136.

[6] Andrew J Payne et al. "Antioxidant drug therapy approaches for neuroprotection in chronic diseases of the retina". In: *International journal of molecular sciences* 15.2 (2014), pp. 1865–1886.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[8] Xin Yang et al. "DCU-net: a deformable convolutional neural network based on cascade U-net for retinal vessel segmentation". In: *Multimedia Tools and Applications* 81.11 (2022), pp. 15593–15607.

[9] Joes Staal et al. "Ridge-based vessel segmentation in color images of the retina". In: *IEEE transactions on medical imaging* 23.4 (2004), pp. 501–509.

[10] Yuqian Zhou, Hanchao Yu, and Humphrey Shi. "Study group learning: Improving retinal vessel segmentation trained with noisy labels". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2021, pp. 57–67.

[11] Zaiwang Gu et al. "Ce-net: Context encoder network for 2d medical image segmentation". In: *IEEE transactions on medical imaging* 38.10 (2019), pp. 2281–2292.

[12] Liangzhi Li et al. "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks". In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2020, pp. 3656–3665.

[13] Helena M Pakter et al. "Measuring arteriolar-to-venous ratio in retinal photography of patients with hypertension: development and application of a new semi-automated method". In: *American journal of hypertension* 18.3 (2005), pp. 417–421.

[14] Tien Yin Wong et al. "Computer-assisted measurement of retinal vessel diameters in the Beaver Dam Eye Study: methodology, correlation between eyes, and effect of refractive errors". In: *Ophthalmology* 111.6 (2004), pp. 1183–1190.

[15] Cemil Kirbas and Francis Quek. "A review of vessel extraction techniques and algorithms". In: *ACM Computing Surveys (CSUR)* 36.2 (2004), pp. 81–121.

[16] João VB Soares et al. "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification". In: *IEEE Transactions on medical Imaging* 25.9 (2006), pp. 1214–1222.

[17] Diego Marín et al. "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features". In: *IEEE Transactions on medical imaging* 30.1 (2010), pp. 146–158.

[18] Elisa Ricci and Renzo Perfetti. "Retinal blood vessel segmentation using line operators and support vector classification". In: *IEEE transactions on medical imaging* 26.10 (2007), pp. 1357–1365.

[19] Xinge You et al. "Segmentation of retinal blood vessels using the radial projection and semi-supervised approach". In: *Pattern recognition* 44.10-11 (2011), pp. 2314–2324.

[20] Meindert Niemeijer et al. "Comparative study of retinal vessel segmentation methods on a new publicly available database". In: *Medical imaging 2004: image processing*. Vol. 5370. SPIE. 2004, pp. 648–656.

[21] Sameh A Salem, Nancy M Salem, and Asoke K Nandi. "Segmentation of retinal blood vessels using a novel clustering algorithm (RACAL) with a partial supervision strategy". In: *Medical & biological engineering & computing* 45.3 (2007), pp. 261–273.

[22] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response". In: *IEEE Transactions on Medical imaging* 19.3 (2000), pp. 203–210.

[23] Frederic Zana and J-C Klein. "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation". In: *IEEE transactions on image processing* 10.7 (2001), pp. 1010–1019.

[24] Uyen TV Nguyen et al. "An effective retinal blood vessel segmentation method using multi-scale line detection". In: *Pattern recognition* 46.3 (2013), pp. 703–715.

[25] Yanli Hou. "Automatic segmentation of retinal blood vessels based on improved multiscale line detection". In: *Journal of Computing Science and Engineering* 8.2 (2014), pp. 119–128.

[26] Koen A Vermeer et al. "A model based method for retinal blood vessel detection". In: *Computers in biology and medicine* 34.3 (2004), pp. 209–219.

[27] Rubén Posada-Gómez et al. "Digital image processing using LabVIEW". In: *Practical Applications and Solutions Using LabVIEW Software, InTech* (2011), pp. 297–316.

[28] Nalan Karunanayake, Nihal D Kodikara, et al. "An improved method for automatic retinal blood vessel vascular segmentation using gabor filter". In: *Open Journal of Medical Imaging* 5.04 (2015), p. 204.

[29] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[30] Jingdan Zhang et al. "Blood vessel segmentation of retinal images based on neural network". In: *International Conference on Image and Graphics*. Springer. 2015, pp. 11–17.

[31] Jen Hong Tan et al. "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network". In: *Journal of Computational Science* 20 (2017), pp. 70–79.

[32] Kevis-Kokitsi Maninis et al. "Deep retinal image understanding". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2016, pp. 140–148.

[33] Dwarikanath Mahapatra et al. "Retinal image quality classification using saliency maps and CNNs". In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2016, pp. 172–179.

[34] Yishuo Zhang and Albert Chung. "Deep supervision with additional labels for retinal vessel segmentation task". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2018, pp. 83–91.

[35] Qiangguo Jin et al. "DUNet: A deformable network for retinal vessel segmentation". In: *Knowledge-Based Systems* 178 (2019), pp. 149–162.

[36] Reza Azad et al. "Bi-directional ConvLSTM U-Net with densley connected convolutions". In: *Proceedings of the IEEE/CVF international conference on computer vision workshops*. 2019, pp. 0–0.

[37] Lei Mou et al. "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging". In: *Medical image analysis* 67 (2021), p. 101874.

[38] Juntang Zhuang. "LadderNet: Multi-path networks based on U-Net for medical image segmentation". In: *arXiv preprint arXiv:1810.07810* (2018).

[39] Md Zahangir Alom et al. "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation". In: *arXiv e-prints* (2018), arXiv–1802.

[40] Changlu Guo et al. "Sa-unet: Spatial attention u-net for retinal vessel segmentation". In: *2020 25th international conference on pattern recognition (ICPR)*. IEEE. 2021, pp. 1236–1242.

[41] Christopher G Owen et al. "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program". In: *Investigative ophthalmology & visual science* 50.5 (2009), pp. 2004–2010.

[42] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[43] Wenzhe Shi and Jose Caballero. "Ferenc Husz á r, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network". In: *Conf. on computer vision and pattern recognition (CVPR)*. Vol. 1883. 1874.

[44] Ningning Ma et al. "Shufflenet v2: Practical guidelines for efficient cnn architecture design". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 116–131.

[45] Shang-Hua Gao et al. "Res2net: A new multi-scale backbone architecture". In: *IEEE transactions on pattern analysis and machine intelligence* 43.2 (2019), pp. 652–662.

[46] François Chollet. "Xception: Deep learning with depthwise separable convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1251–1258.

[47] Xiangyu Zhang et al. "Shufflenet: An extremely efficient convolutional neural network for mobile devices". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 6848–6856.

[48] Atul Pandey. *Depth-wise convolution and depth-wise separable convolution*. 2018. URL: https://medium.com/@zurister/depth-wise-convolution-and-depth-wise-separable-convolution-37346565d4ec.

[49] Wenjie Luo et al. "Understanding the effective receptive field in deep convolutional neural networks". In: *Advances in neural information processing systems* 29 (2016).

[50] Mengjiao Qin et al. "Remote Sensing Single-Image Resolution Improvement Using A Deep Gradient-Aware Network with Image-Specific Enhancement". In: *Remote Sensing* 12 (Feb. 2020), p. 758. DOI: 10.3390/rs12050758.

[51] Karel Zuiderveld. "Contrast limited adaptive histogram equalization". In: *Graphics gems* (1994), pp. 474–485.

[52] Tsung-Yi Lin et al. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988.

[53] Jingfei Hu et al. "SA-Net: A scale-attention network for medical image segmentation". In: *PloS one* 16.4 (2021), e0247388.

[54] Nancy Krieger. "The US Census and the people's health: public health engagement from enslavement and "Indians not taxed" to census tracts and health equity (1790–2018)". In: *American Journal of Public Health* 109.8 (2019), pp. 1092–1100.

[55] Kevin A Matthews et al. "Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged 65 years". In: *Alzheimer's & Dementia* 15.1 (2019), pp. 17–24.

[56] Ivan I Kirov et al. "Global brain volume and N-acetyl-aspartate decline over seven decades of normal aging". In: *Neurobiology of aging* 98 (2021), pp. 42–51.

[57] Josephine Barnes et al. "Head size, age and gender adjustment in MRI studies: a necessary nuisance?" In: *Neuroimage* 53.4 (2010), pp. 1244–1255.

[58] Antonio Giorgio et al. "Age-related changes in grey and white matter structure throughout adulthood". In: *Neuroimage* 51.3 (2010), pp. 943–951.

[59] Catriona D Good et al. "A voxel-based morphometric study of ageing in 465 normal adult human brains". In: *Neuroimage* 14.1 (2001), pp. 21–36.

[60] Charles D Smith et al. "Age and gender effects on human brain anatomy: a voxel-based morphometric study in healthy elderly". In: *Neurobiology of aging* 28.7 (2007), pp. 1075–1087.

[61] Antonio Giorgio et al. "Longitudinal changes in grey and white matter during adolescence". In: *Neuroimage* 49.1 (2010), pp. 94–103.

[62] Hiroshi Matsuda. "Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease". In: *Aging and disease* 4.1 (2013), p. 29.

[63] Yulin Ge et al. "Age-related total gray matter and white matter changes in normal adult brain. Part I: volumetric MR imaging analysis". In: *American journal of neuroradiology* 23.8 (2002), pp. 1327–1333.

[64] David A Ziegler et al. "Cognition in healthy aging is related to regional white matter integrity, but not cortical thickness". In: *Neurobiology of aging* 31.11 (2010), pp. 1912–1926.

[65] David H Salat, Jeffrey A Kaye, and Jeri S Janowsky. "Prefrontal gray and white matter volumes in healthy aging and Alzheimer disease". In: *Archives of neurology* 56.3 (1999), pp. 338–344.

[66] Patricia E Cowell et al. "Sex differences in aging of the human frontal and temporal lobes". In: *Journal of Neuroscience* 14.8 (1994), pp. 4748–4755.

[67] Adolf Pfefferbaum et al. "A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood". In: *Archives of neurology* 51.9 (1994), pp. 874–887.

[68] Ruben C Gur et al. "Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance". In: *Journal of Neuroscience* 19.10 (1999), pp. 4065–4072.

[69] George Bartzokis et al. "Age-related changes in frontal and temporal lobe volumes in men: a magnetic resonance imaging study". In: *Archives of general psychiatry* 58.5 (2001), pp. 461–465.

[70] Mark J West et al. "Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease". In: *The Lancet* 344.8925 (1994), pp. 769–772.

[71] Stefano Tarantini et al. "Impaired neurovascular coupling in aging and Alzheimer's disease: contribution of astrocyte dysfunction and endothelial impairment to cognitive decline". In: *Experimental gerontology* 94 (2017), pp. 52–58.

[72] Anders M Fjell et al. "What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus". In: *Progress in neurobiology* 117 (2014), pp. 20–40.

[73] Lisette Bosscher and Philip Scheltens. "MRI of the medial temporal lobe for the diagnosis of Alzheimer's disease". In: *Evidence-based dementia practice. Osney Mead, Oxford* (2002), pp. 154–62.

[74] Josephine Barnes et al. "A meta-analysis of hippocampal atrophy rates in Alzheimer's disease". In: *Neurobiology of aging* 30.11 (2009), pp. 1711–1723.

[75] Lorenzo Pini et al. "Brain atrophy in Alzheimer's disease and aging". In: *Ageing research reviews* 30 (2016), pp. 25–48.

[76] DJA Callen et al. "Beyond the hippocampus: MRI volumetry confirms widespread limbic atrophy in AD". In: *Neurology* 57.9 (2001), pp. 1669–1674.

[77] GB Karas et al. "Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease". In: *Neuroimage* 23.2 (2004), pp. 708–716.

[78] John V Hindle. "Ageing, neurodegeneration and Parkinson's disease". In: *Age and ageing* 39.2 (2010), pp. 156–161.

[79] Emma J Burton et al. "Cerebral atrophy in Parkinson's disease with and without dementia: a comparison with Alzheimer's disease, dementia with Lewy bodies and controls". In: *Brain* 127.4 (2004), pp. 791–800.

[80] Fabio Blandini et al. "Functional changes of the basal ganglia circuitry in Parkinson's disease". In: *Progress in neurobiology* 62.1 (2000), pp. 63–88.

[81] Christina J Azevedo et al. "Contribution of normal aging to brain atrophy in MS". In: *Neurology-Neuroimmunology Neuroinflammation* 6.6 (2019).

[82] Jiang Xu et al. "Gender effects on age-related changes in brain structure". In: *American journal of neuroradiology* 21.1 (2000), pp. 112–118.

[83] Ioanis A Parashos, William E Wilkinson, and C Edward Coffey. "Magnetic resonance imaging of the corpus callosum: predictors of size in normal adults." In: *The Journal of neuropsychiatry and clinical neurosciences* 7.1 (1995), pp. 35–41.

[84] Edith V Sullivan et al. "Greater abnormalities of brain cerebrospinal fluid volumes in younger than in older patients with Alzheimer's disease". In: *Archives of neurology* 50.4 (1993), pp. 359–373.

[85] Debbie Overman. "For Accuracy, Brain Studies of Complex Behavior Require Thousands of People". In: *AXIS Imaging News* (2022).

[86] Olga Voevodskaya et al. "The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease". In: *Frontiers in aging neuroscience* 6 (2014), p. 264.

[87] Leonie Henschel et al. "Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline". In: *NeuroImage* 219 (2020), p. 117012.

[88] Lauren N Koenig et al. "Regional age-related atrophy after screening for preclinical Alzheimer disease". In: *Neurobiology of aging* 109 (2022), pp. 43–51.

[89] Eva Schaeffer et al. "Effects of exergaming on hippocampal volume and brain-derived neurotrophic factor levels in Parkinson's disease". In: *European journal of neurology* 29.2 (2022), pp. 441–449.

[90] Marinus T Vlaardingerbroek and Jacques A Boer. *Magnetic resonance imaging: theory and practice*. Springer Science & Business Media, 2013.

[91] PS Shijin Kumar and VS Dharun. "A study of MRI segmentation methods in automatic brain tumor detection". In: *Int. J. Eng. Technol* 8.2 (2016), pp. 609–614.

[92] Natasha Morales Drissi. *Brain Networks and Dynamics in Narcolepsy*. Vol. 1651. Linköping University Electronic Press, 2019.

[93] Spiros Sgouros et al. "Intracranial volume change in childhood". In: *Journal of neurosurgery* 91.4 (1999), pp. 610–616.

[94] https://slideplayer.com/slide/4262652/.

[95] Fabian Isensee et al. "nnu-net: Breaking the spell on successful medical image segmentation". In: *arXiv preprint arXiv:1904.08128* 1.1-8 (2019), p. 2.

[96] Charlotte Debus et al. "MITK-ModelFit: A generic open-source framework for model fits and their exploration in medical imaging–design, implementation and application on the example of DCE-MRI". In: *BMC bioinformatics* 20.1 (2019), pp. 1–18.

[97] Xiaojing Long et al. "Healthy aging: an automatic analysis of global and regional morphological alterations of human brain". In: *Academic radiology* 19.7 (2012), pp. 785–793.

[98] Marina Buchpiguel et al. "Differences in total brain volume between sexes in a cognitively unimpaired elderly population". In: *Clinics* 75 (2020).

[99] Daniel L Greenberg et al. "Aging, gender, and the elderly adult brain: an examination of analytical strategies". In: *Neurobiology of aging* 29.2 (2008), pp. 290–302.

[100] Jan Kristanto Wibisono and Hsueh-Ming Hang. "Traditional method inspired deep neural network for edge detection". In: *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2020, pp. 678–682.

[101] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[102] Yunhui Guo et al. "Depthwise convolution is all you need for learning multiple visual domains". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 01. 2019, pp. 8368–8375.

[103] Pablo Arbelaez et al. "Contour detection and hierarchical image segmentation". In: *IEEE transactions on pattern analysis and machine intelligence* 33.5 (2010), pp. 898–916.

[104] Nathan Silberman et al. "Indoor segmentation and support inference from rgbd images". In: *European conference on computer vision*. Springer. 2012, pp. 746–760.

[105] O Rebecca Vincent, Olusegun Folorunso, et al. "A descriptive algorithm for sobel image edge detection". In: *Proceedings of informing science & IT education conference (InSITE)*. Vol. 40. 2009, pp. 97–107.

[106] Renjie Song, Ziqi Zhang, and Haiyang Liu. "Edge connection based Canny edge detection algorithm". In: *Pattern Recognition and Image Analysis* 27.4 (2017), pp. 740–747.

[107] Rajiv Mehrotra and Shiming Zhan. "A computational approach to zero-crossing-based two-dimensional edge detection". In: *Graphical Models and Image Processing* 58.1 (1996), pp. 1–17.

[108] James J. Clark. "Authenticating edges produced by zero-crossing algorithms". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.1 (1989), pp. 43–57.

[109] Scott Konishi et al. "Statistical edge detection: Learning and evaluating edge cues". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.1 (2003), pp. 57–74.

[110] Piotr Dollár and C Lawrence Zitnick. "Fast edge detection using structured forests". In: *IEEE transactions on pattern analysis and machine intelligence* 37.8 (2014), pp. 1558–1570.

[111] Saining Xie and Zhuowen Tu. "Holistically-nested edge detection". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1395–1403.

[112] Yupei Wang, Xin Zhao, and Kaiqi Huang. "Deep crisp boundaries". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3892–3900.

[113] Dan Xu et al. "Learning deep structured multi-scale features using attention-gated crfs for contour prediction". In: *Advances in neural information processing systems* 30 (2017).

[114] Yun Liu et al. "Richer convolutional features for edge detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3000–3009.

[115] Ruoxi Deng et al. "Learning to predict crisp boundaries". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 562–578.

[116] Jianzhong He et al. "Bi-directional cascade network for perceptual edge detection". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 3828–3837.

[117] Jan Kristanto Wibisono and Hsueh-Ming Hang. "Fined: Fast inference network for edge detection". In: *arXiv preprint arXiv:2012.08392* (2020).

[118] Xavier Soria Poma, Edgar Riba, and Angel Sappa. "Dense extreme inception network: Towards a robust cnn model for edge detection". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020, pp. 1923–1932.

[119] Jifeng Dai et al. "Deformable convolutional networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 764–773.

[120] Sanghyun Woo et al. "Cbam: Convolutional block attention module". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.

[121] Roozbeh Mottaghi et al. "The role of context for object detection and semantic segmentation in the wild". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 891–898.

[122] John Canny. "A computational approach to edge detection". In: *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), pp. 679–698.

[123] Sam Hallman and Charless C Fowlkes. "Oriented edge forests for boundary detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1732–1740.

[124] Kai Li et al. "Bi-directional pyramid network for edge detection". In: *Electronics* 10.3 (2021), p. 329.

[125] Wenjie Xuan et al. "FCL-Net: Towards accurate edge detection via Fine-scale Corrective Learning". In: *Neural Networks* 145 (2022), pp. 248–259.

[126] Shi-Shui Bao, You-Rui Huang, and Guang-Yu Xu. "Bidirectional Multiscale Refinement Network for Crisp Edge Detection". In: *IEEE Access* 10 (2022), pp. 26282–26293.

[127] Kevis-Kokitsi Maninis et al. "Convolutional oriented boundaries: From image segmentation to high-level tasks". In: *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017), pp. 819–833.

[128] Abdullah Al-Amaren, M Omair Ahmad, and MNS Swamy. "RHN: A residual holistic neural network for edge detection". In: *IEEE Access* 9 (2021), pp. 74646–74658.

[129] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. "Deepedge: A multi-scale bifurcated deep network for top-down contour detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 4380–4389.

[130] Wei Shen et al. "Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3982–3991.

[131] Saurabh Gupta et al. "Learning rich features from RGB-D images for object detection and segmentation". In: *European conference on computer vision*. Springer. 2014, pp. 345–360.

3.6 Appendix

TABLE 3.3: The Pearson partial correlations results between age and listed of regions controlling for the effect of ICV

Region	Gender	r	CI95%	P-value
Left-Cerebral-White-Matter	Female	-0.301	[-0.33 -0.27]	<0.00001
	Male	-0.338	[-0.37 -0.31]	<0.00001
Left-Lateral-Ventricle	Female	0.468	[0.44 0.49]	<0.00001
	Male	0.506	[0.48 0.53]	<0.00001
Left-Inf-Lat-Vent	Female	0.354	[0.32 0.38]	<0.00001
	Male	0.462	[0.43 0.49]	<0.00001
Left-Cerebellum-White-Matter	Female	-0.297	[-0.33 -0.27]	<0.00001
	Male	-0.358	[-0.39 -0.33]	<0.00001
Left-Cerebellum-Cortex	Female	-0.263	[-0.29 -0.23]	<0.00001
	Male	-0.378	[-0.41 -0.35]	<0.00001
Left-Thalamus-Proper*	Female	-0.565	[-0.59 -0.54]	<0.00001
	Male	-0.593	[-0.61 -0.57]	<0.00001
Left-Caudate	Female	-0.141	[-0.17 -0.11]	<0.00001
	Male	-0.222	[-0.25 -0.19]	<0.00001
Left-Putamen	Female	-0.4	[-0.43 -0.37]	<0.00001
	Male	-0.469	[-0.5 -0.44]	<0.00001
Left-Pallidum	Female	-0.36	[-0.39 -0.33]	<0.00001
	Male	-0.439	[-0.47 -0.41]	<0.00001
3rd-Ventricle	Female	0.54	[0.52 0.56]	<0.00001
	Male	0.572	[0.55 0.59]	<0.00001
4th-Ventricle	Female	0.129	[0.1 0.16]	<0.00001
	Male	0.133	[0.1 0.17]	<0.00001
Brain-Stem	Female	-0.121	[-0.15 -0.09]	<0.00001
	Male	-0.163	[-0.2 -0.13]	<0.00001
Left-Hippocampus	Female	-0.318	[-0.35 -0.29]	<0.00001
	Male	-0.376	[-0.4 -0.35]	<0.00001
Left-Amygdala	Female	-0.33	[-0.36 -0.3]	<0.00001
	Male	-0.356	[-0.39 -0.33]	<0.00001
CSF	Female	0.353	[0.32 0.38]	<0.00001
	Male	0.363	[0.33 0.39]	<0.00001
Left-Accumbens-area	Female	-0.386	[-0.41 -0.36]	<0.00001
	Male	-0.426	[-0.45 -0.4]	<0.00001
Left-VentralDC	Female	-0.438	[-0.47 -0.41]	<0.00001
	Male	-0.474	[-0.5 -0.45]	<0.00001
Left-choroid-plexus	Female	0.474	[0.45 0.5]	<0.00001
	Male	0.486	[0.46 0.51]	<0.00001
Right-Cerebral-White-Matter	Female	-0.298	[-0.33 -0.27]	<0.00001
	Male	-0.337	[-0.37 -0.31]	<0.00001
Right-Lateral-Ventricle	Female	0.483	[0.46 0.51]	<0.00001
	Male	0.514	[0.49 0.54]	<0.00001
Right-Inf-Lat-Vent	Female	0.277	[0.25 0.31]	<0.00001
	Male	0.38	[0.35 0.41]	<0.00001

Region	Gender	r	CI95%	P-value
Right-Cerebellum-White-Matter	Female	-0.293	[-0.32 -0.26]	<0.00001
	Male	-0.349	[-0.38 -0.32]	<0.00001
Right-Cerebellum-Cortex	Female	-0.254	[-0.29 -0.22]	<0.00001
	Male	-0.36	[-0.39 -0.33]	<0.00001
Right-Thalamus-Proper*	Female	-0.527	[-0.55 -0.5]	<0.00001
	Male	-0.559	[-0.58 -0.54]	<0.00001
Right-Caudate	Female	-0.113	[-0.15 -0.08]	<0.00001
	Male	-0.186	[-0.22 -0.15]	<0.00001
Right-Putamen	Female	-0.406	[-0.43 -0.38]	<0.00001
	Male	-0.47	[-0.5 -0.44]	<0.00001
Right-Pallidum	Female	-0.374	[-0.4 -0.34]	<0.00001
	Male	-0.429	[-0.46 -0.4]	<0.00001
Right-Hippocampus	Female	-0.304	[-0.33 -0.27]	<0.00001
	Male	-0.35	[-0.38 -0.32]	<0.00001
Right-Amygdala	Female	-0.265	[-0.3 -0.23]	<0.00001
	Male	-0.263	[-0.29 -0.23]	<0.00001
Right-Accumbens-area	Female	-0.368	[-0.4 -0.34]	<0.00001
	Male	-0.436	[-0.46 -0.41]	<0.00001
Right-VentralDC	Female	-0.447	[-0.47 -0.42]	<0.00001
	Male	-0.455	[-0.48 -0.43]	<0.00001
Right-choroid-plexus	Female	0.449	[0.42 0.48]	<0.00001
	Male	0.46	[0.43 0.49]	<0.00001
WM-hypointensities	Female	0.556	[0.53 0.58]	<0.00001
	Male	0.562	[0.54 0.58]	<0.00001
ctx-lh-caudalanteriorcingulate	Female	-0.263	[-0.29 -0.23]	<0.00001
	Male	-0.323	[-0.35 -0.29]	<0.00001
ctx-lh-caudalmiddlefrontal	Female	-0.322	[-0.35 -0.29]	<0.00001
	Male	-0.337	[-0.37 -0.31]	<0.00001
ctx-lh-cuneus	Female	-0.134	[-0.17 -0.1]	<0.00001
	Male	-0.211	[-0.24 -0.18]	<0.00001
ctx-lh-entorhinal	Female	-0.045	[-0.08 -0.01]	0.009
	Male	-0.129	[-0.16 -0.1]	<0.00001
ctx-lh-fusiform	Female	-0.253	[-0.28 -0.22]	<0.00001
	Male	-0.331	[-0.36 -0.3]	<0.00001
ctx-lh-inferiorparietal	Female	-0.306	[-0.34 -0.28]	<0.00001
	Male	-0.309	[-0.34 -0.28]	<0.00001
ctx-lh-inferiortemporal	Female	-0.3	[-0.33 -0.27]	<0.00001
	Male	-0.373	[-0.4 -0.34]	<0.00001
ctx-lh-isthmuscingulate	Female	-0.152	[-0.19 -0.12]	<0.00001
	Male	-0.2	[-0.23 -0.17]	<0.00001
ctx-lh-lateraloccipital	Female	-0.247	[-0.28 -0.21]	<0.00001
	Male	-0.258	[-0.29 -0.23]	<0.00001
ctx-lh-lateralorbitofrontal	Female	-0.394	[-0.42 -0.36]	<0.00001
	Male	-0.49	[-0.52 -0.46]	<0.00001
ctx-lh-lingual	Female	-0.159	[-0.19 -0.13]	<0.00001
	Male	-0.256	[-0.29 -0.22]	<0.00001
ctx-lh-medialorbitofrontal	Female	-0.337	[-0.37 -0.31]	<0.00001
	Male	-0.42	[-0.45 -0.39]	<0.00001
ctx-lh-middletemporal	Female	-0.437	[-0.46 -0.41]	<0.00001
	Male	-0.474	[-0.5 -0.45]	<0.00001

Region	Gender	r	CI95%	P-value
ctx-lh-parahippocampal	Female	-0.218	[-0.25 -0.19]	<0.00001
	Male	-0.276	[-0.31 -0.24]	<0.00001
ctx-lh-paracentral	Female	-0.261	[-0.29 -0.23]	<0.00001
	Male	-0.293	[-0.32 -0.26]	<0.00001
ctx-lh-parsopercularis	Female	-0.347	[-0.38 -0.32]	<0.00001
	Male	-0.389	[-0.42 -0.36]	<0.00001
ctx-lh-parsorbitalis	Female	-0.247	[-0.28 -0.21]	<0.00001
	Male	-0.331	[-0.36 -0.3]	<0.00001
ctx-lh-parstriangularis	Female	-0.297	[-0.33 -0.27]	<0.00001
	Male	-0.358	[-0.39 -0.33]	<0.00001
ctx-lh-pericalcarine	Female	-0.106	[-0.14 -0.07]	<0.00001
	Male	-0.154	[-0.19 -0.12]	<0.00001
ctx-lh-postcentral	Female	-0.354	[-0.38 -0.32]	<0.00001
	Male	-0.328	[-0.36 -0.3]	<0.00001
ctx-lh-posteriorcingulate	Female	-0.334	[-0.36 -0.3]	<0.00001
	Male	-0.375	[-0.4 -0.35]	<0.00001
ctx-lh-precentral	Female	-0.388	[-0.42 -0.36]	<0.00001
	Male	-0.407	[-0.44 -0.38]	<0.00001
ctx-lh-precuneus	Female	-0.238	[-0.27 -0.21]	<0.00001
	Male	-0.322	[-0.35 -0.29]	<0.00001
ctx-lh-rostralanteriorcingulate	Female	-0.213	[-0.24 -0.18]	<0.00001
	Male	-0.297	[-0.33 -0.27]	<0.00001
ctx-lh-rostralmiddlefrontal	Female	-0.419	[-0.45 -0.39]	<0.00001
	Male	-0.43	[-0.46 -0.4]	<0.00001
ctx-lh-superiorfrontal	Female	-0.544	[-0.57 -0.52]	<0.00001
	Male	-0.535	[-0.56 -0.51]	<0.00001
ctx-lh-superiorparietal	Female	-0.311	[-0.34 -0.28]	<0.00001
	Male	-0.31	[-0.34 -0.28]	<0.00001
ctx-lh-superiortemporal	Female	-0.415	[-0.44 -0.39]	<0.00001
	Male	-0.415	[-0.44 -0.39]	<0.00001
ctx-lh-supramarginal	Female	-0.37	[-0.4 -0.34]	<0.00001
	Male	-0.419	[-0.45 -0.39]	<0.00001
ctx-lh-transversetemporal	Female	-0.223	[-0.25 -0.19]	<0.00001
	Male	-0.212	[-0.24 -0.18]	<0.00001
ctx-lh-insula	Female	-0.211	[-0.24 -0.18]	<0.00001
	Male	-0.256	[-0.29 -0.22]	<0.00001
ctx-rh-caudalanteriorcingulate	Female	-0.146	[-0.18 -0.11]	<0.00001
	Male	-0.193	[-0.22 -0.16]	<0.00001
ctx-rh-caudalmiddlefrontal	Female	-0.298	[-0.33 -0.27]	<0.00001
	Male	-0.356	[-0.38 -0.33]	<0.00001
ctx-rh-cuneus	Female	-0.149	[-0.18 -0.12]	<0.00001
	Male	-0.251	[-0.28 -0.22]	<0.00001
ctx-rh-entorhinal	Female	0.065	[0.03 0.1]	<0.00001
	Male	-0.031	[-0.06 0.]	0.076
ctx-rh-fusiform	Female	-0.188	[-0.22 -0.15]	<0.00001
	Male	-0.277	[-0.31 -0.25]	<0.00001
ctx-rh-inferiorparietal	Female	-0.409	[-0.44 -0.38]	<0.00001
	Male	-0.434	[-0.46 -0.41]	<0.00001
ctx-rh-inferiortemporal	Female	-0.329	[-0.36 -0.3]	<0.00001
	Male	-0.385	[-0.41 -0.36]	<0.00001

Region	Gender	r	CI95%	P-value
ctx-rh-isthmuscingulate	Female	-0.153	[-0.19 -0.12]	<0.00001
	Male	-0.19	[-0.22 -0.16]	<0.00001
ctx-rh-lateraloccipital	Female	-0.218	[-0.25 -0.19]	<0.00001
	Male	-0.25	[-0.28 -0.22]	<0.00001
ctx-rh-lateralorbitofrontal	Female	-0.378	[-0.41 -0.35]	<0.00001
	Male	-0.467	[-0.49 -0.44]	<0.00001
ctx-rh-lingual	Female	-0.153	[-0.19 -0.12]	<0.00001
	Male	-0.226	[-0.26 -0.19]	<0.00001
ctx-rh-medialorbitofrontal	Female	-0.317	[-0.35 -0.29]	<0.00001
	Male	-0.393	[-0.42 -0.36]	<0.00001
ctx-rh-middletemporal	Female	-0.466	[-0.49 -0.44]	<0.00001
	Male	-0.481	[-0.51 -0.45]	<0.00001
ctx-rh-parahippocampal	Female	-0.246	[-0.28 -0.21]	<0.00001
	Male	-0.338	[-0.37 -0.31]	<0.00001
ctx-rh-paracentral	Female	-0.284	[-0.31 -0.25]	<0.00001
	Male	-0.273	[-0.3 -0.24]	<0.00001
ctx-rh-parsopercularis	Female	-0.317	[-0.35 -0.29]	<0.00001
	Male	-0.359	[-0.39 -0.33]	<0.00001
ctx-rh-parsorbitalis	Female	-0.289	[-0.32 -0.26]	<0.00001
	Male	-0.383	[-0.41 -0.35]	<0.00001
ctx-rh-parstriangularis	Female	-0.277	[-0.31 -0.25]	<0.00001
	Male	-0.338	[-0.37 -0.31]	<0.00001
ctx-rh-pericalcarine	Female	-0.065	[-0.1 -0.03]	<0.00001
	Male	-0.11	[-0.14 -0.08]	<0.00001
ctx-rh-postcentral	Female	-0.334	[-0.36 -0.3]	<0.00001
	Male	-0.309	[-0.34 -0.28]	<0.00001
ctx-rh-posteriorcingulate	Female	-0.256	[-0.29 -0.22]	<0.00001
	Male	-0.33	[-0.36 -0.3]	<0.00001
ctx-rh-precentral	Female	-0.35	[-0.38 -0.32]	<0.00001
	Male	-0.376	[-0.4 -0.35]	<0.00001
ctx-rh-precuneus	Female	-0.295	[-0.33 -0.26]	<0.00001
	Male	-0.359	[-0.39 -0.33]	<0.00001
ctx-rh-rostralanteriorcingulate	Female	-0.154	[-0.19 -0.12]	<0.00001
	Male	-0.238	[-0.27 -0.21]	<0.00001
ctx-rh-rostralmiddlefrontal	Female	-0.415	[-0.44 -0.39]	<0.00001
	Male	-0.428	[-0.45 -0.4]	<0.00001
ctx-rh-superiorfrontal	Female	-0.545	[-0.57 -0.52]	<0.00001
	Male	-0.529	[-0.55 -0.5]	<0.00001
ctx-rh-superiorparietal	Female	-0.31	[-0.34 -0.28]	<0.00001
	Male	-0.342	[-0.37 -0.31]	<0.00001
ctx-rh-superiortemporal	Female	-0.433	[-0.46 -0.4]	<0.00001
	Male	-0.435	[-0.46 -0.41]	<0.00001
ctx-rh-supramarginal	Female	-0.364	[-0.39 -0.33]	<0.00001
	Male	-0.378	[-0.41 -0.35]	<0.00001
ctx-rh-transversetemporal	Female	-0.293	[-0.32 -0.26]	<0.00001
	Male	-0.298	[-0.33 -0.27]	<0.00001
ctx-rh-insula	Female	-0.256	[-0.29 -0.22]	<0.00001
	Male	-0.307	[-0.34 -0.28]	<0.00001