August 2022
Master's Degree Thesis

Parallel UNet Architecture with Late Fusion Residual Network for Segmentation of Brain Tissues

Graduate School of Chosun University

Department of Information and Communication Engineering

Rukesh Prajapati

Parallel UNet Architecture with Late Fusion Residual Network for Segmentation of Brain Tissues

August 26, 2022

Graduate School of Chosun University

Department of Information and Communication Engineering

Rukesh Prajapati

Parallel UNet Architecture with Late Fusion Residual Network for Segmentation of Brain Tissues

뇌 조직 세그멘테이션을 위한 Late Fusion Residual Network을 갖는 병렬 UNet 구조

Advisor: Prof. Young-Sik Kim

This thesis is submitted to Chosun University in partial fulfillment of the requirements for a Master's degree

April 2022

Graduate School of Chosun University

Department of Information and Communication Engineering

Rukesh Prajapati

This is to certify that the master's thesis of

Rukesh Prajapati

has been approved by examining committee for the thesis requirement for the master's degree in Engineering.

Committee Chairperson

Prof. Jae-Young Pyun _____

Committee Member

Prof. Young-Sik Kim _____

Committee Member

Prof. Goo-Rak Kwon _____

May 2022

Graduate School of Chosun University

Table of contents

List of figures

List of tables

초 록

뇌 조직 세그멘테이션을 위한 Late Fusion Residual Network을 갖는 병렬 UNet 구조

프라자파티 루케스
지도 교수: 김영식
정보통신공학과
조선대학교

신경퇴행성 만성질환, 알츠하이머 병 특정 뇌 질환에 대한 확실한 진단은 뇌의 해부학적 변화를 분석함으로써 가능하다. 뇌 해부학의 변화를 평가하기 위해 세분화의 정확도가 중요하다. 최근 연구에서 UNet 기반의 설계는 생물학적 영상 분할에서 딥러닝 아키텍처를 대체할 정도로 성능이 능가하는 것으로 나타난다. 그러나 의료영상의 해상도가 낮고 데이터가 부족하여 분할 정확도를 높이는 것이 어렵다. 본 논문에서는 잔류 네트워크를 사용하여 3개의 병렬 UNet을 결합하는 고유 아키텍처를 제안한다. 이 디자인은 세 가지 면에서 기본 접근 방식을 개선한다. 먼저 단일 이미지가 아닌 연속 3장의 이미지를 입력으로 활용한다. 결과적으로 제안된 모델은 3장의 이미지 사이에서 자유롭게 학습할 수 있다. 또한 3개의 고유한 UNet을 사용하여 개별적으로 압축 및 압축해제를 통해 모델의 이미지의 특성을 결합하는

것을 방지한다. 마지막으로 UNet의 출력은 잔류 네트워크 설계에 따라 스킵 연결에 의해 출력에 해당하는 이미지의 특성이 증대되는 방식으로 병합된다. 뇌 조직 분할을 위해 제안된 방법은 공개적으로 사용 가능한 데이터 셋에서 테스트 된다. 제안된 모델은 실험데이터에 따라 뇌척수액(CSF)의 경우 94.9%, 회백질(GM)의 경우 95.8%, 백질(WM)의 경우 95.93%의 분할 정확도를 달성했다. 이는 제안된 방법이 자동으로 안정적이며 정확한 방식으로 뇌 MRI 영상을 분할할 수 있음을 보여준다. 제안된 방법은 다양한 변형된 UNet뿐만 아니라 기존의 단일 UNet보다 성능이 높다.

키워드: 조직 세그멘테이션, 병렬 UNet, 잔류 네트워크, 뇌 MRI.

Abstract

Parallel UNet Architecture with Late Fusion Residual Network for segmentation of Brain Tissues

Rukesh Prajapati
Advisor: Prof. Young-Sik Kim
Dept. of Information and Communication Engineering
Chosun University

A reliable diagnosis of particular brain illnesses can be made by analyzing changes in brain anatomy. For assessing changes in brain anatomy, segmentation accuracy is critical. UNet-based designs have been shown to outperform alternative deep learning architectures in biological picture segmentation in recent research. However, due to the poor resolution of medical pictures and a lack of data, enhancing segmentation accuracy is difficult. In this paper, a unique architecture is offered that uses a residual network to join three parallel UNets. In three ways, this design improves on the basic approaches. To begin, three successive photos are utilized as input instead of a single image. As a result, the proposed model is free to learn from nearby photos as well. The photos are also compressed and decompressed separately using three distinct UNets, preventing the model from combining the images' characteristics. Finally, the outputs of the UNets are merged in such a way that the characteristics of the picture corresponding to the output are augmented by a skip connection, as per the residual network design. For brain tissue segmentation, the suggested approach was tested on publicly

available datasets. The proposed model achieves segmentation accuracies of 94.9% for cerebrospinal fluid (CSF), 95.8% for gray matter (GM), and 95.93% for white matter (WM), according to the experimental data. This demonstrates that the suggested method can automatically segment brain MRI images in a reliable and exact manner. The suggested design outperformed a single traditional UNet as well as various UNet variations.

Keywords: Tissue segmentation, parallel UNet, residual network, brain MRI.

1. Introduction

Medical image processing technology has advanced at a rapid pace, benefiting mankind and playing a vital role in clinical diagnostics. Medical imaging has advanced to the point that it can now be used to observe the human body in order to diagnose and monitor medical disorders [1-2]. Image information is provided by imaging modalities such as ultrasound (US), magnetic resonance imaging (MRI), and X-ray imaging, which the radiologist must study and assess extensively in a shorter amount of time [2]. MRI is commonly used for structural analysis because it produces pictures with high soft-tissue contrast and spatial resolution while posing minimal health risks. Different brain illnesses, such as epilepsy, schizophrenia, Alzheimer's disease, and other degenerative disorders, are diagnosed using quantitative brain MRI scans [3]. For diagnosis and postoperative analysis, MRI is also required to detect and pinpoint diseased tissues and healthy structures. As a result, segmenting aberrant tissues from medical pictures is critical for the research and treatment of many disorders. The observer can follow up on the person if brain problems are detected early. As a result, the primary goal is to develop better tools that aid in the interpretation of pictures.

Section 1.1 of this chapter gives a brief overview of MRI and how it is used to diagnose various brain illnesses. The suggested work's summary and motivation are presented in Section 1.2. The research aims and key contributions of the thesis are presented in Section 1.3. The thesis outline is explained in Section 1.4.

1.1. A Brief Review of MRI

MRI is a medical imaging technique that uses a magnetic field and computer-generated radio waves to obtain detailed pictures of organs and tissues [2]. The bulk of MRI machines employ large, tube-shaped magnets. While we are inside an MRI scanner, the magnetic field realigns the water molecules in our bodies for a brief duration. Similar to slices in a loaf of bread, these aligned atoms generate small signals that are utilized to form cross-sectional MRI images. In addition, the MRI scanner can provide three-dimensional pictures that may be seen from various angles. An MRI scan is a non-invasive way for a doctor to examine the organs, tissues, and skeletal system. It generates high-resolution pictures of the interior of the body to assist in the diagnosis of a variety of illnesses. The most frequent imaging test for the brain is MRI. The functional MRI of the brain (fMRI) is a one-of-a-kind MRI technique [3]. It creates pictures of blood flow to specific parts of the brain. It may be used to examine the brain's anatomy and determine which parts of the brain are in charge of critical functions. Functional MRI can also be used to diagnose damage after a head injury or disorders like Alzheimer's disease.

1.2. MRI for Brain Disorders Diagnosis

MRI of the brain and brain stem is a non-invasive, painless procedure that produces comprehensive pictures of the brain and brain stem. The value of MRI in brain diagnosis is widely understood, and it has been incorporated into a number of novel brain diagnostic criteria [4-5]. An MRI scan can reveal cysts, tumors, bleeding, edema, developmental and anatomical anomalies, infections, inflammatory illnesses, and blood vessel issues. High-resolution MRI may detect the presence and degree of brain atrophy, which can help with Alzheimer's disease diagnosis in vivo [6] and even indicate the presence of

neurofibrillary tangles (NFTs), which are considered a hallmark pathology of the illness [7]. While MRI-measured brain shrinkage is a reliable and sensitive predictor of neurodegeneration in general [8], it may also be utilized to distinguish between a variety of dementias with comparable atrophy patterns [9].

1.3. Overview and Motivation

Neurodegenerative diseases cause the brain and nerves to degrade over time. These illnesses have the ability to change people's personalities and lead them to become confused. They can also harm the brain's cells and neurons. Alzheimer's disease, for example, is a progressive brain illness that affects people as they age [10]. Over time, they can cause havoc with the memory system and brain functions. Other disorders, such as Tay-Sachs disease, are hereditary and appear early in life.

While certain therapies might temporarily relieve symptoms, none will stop or reverse the disease's progression. In today's lifestyle, prevention is the key to minimizing the prevalence of neurological disorders and, as a result, the number of global mortality. As a result, a strong predictor of clinical diagnosis is an early sign of a greater chance of neurodegenerative illness [10]. Furthermore, medical test findings, clinical history, and medical picture collection are all used to detect neurological brain illnesses. The capacity of an expert to understand a large amount of data in a short amount of time might make diagnosis more difficult in some cases [11]. Medical imaging technologies and computer software advancements are being used to help doctors swiftly diagnose and analyze disorders. These systems diagnose the illness exclusively on the basis of visual data [12]. One of the most difficult problems in the medical profession is developing software for radiological

image processing. Because accurate sickness detection is based on both picture capture and interpretation, modern diagnostic systems are built using cutting-edge computer and data processing technologies. Despite the availability of commercially accessible computer-based diagnostic equipment, completely automated techniques are still lacking in the literature [10-12]. They are difficult to use for diagnostic purposes as a result of this flaw. The loss of brain tissue is the first indicator of the development of a brain ailment.

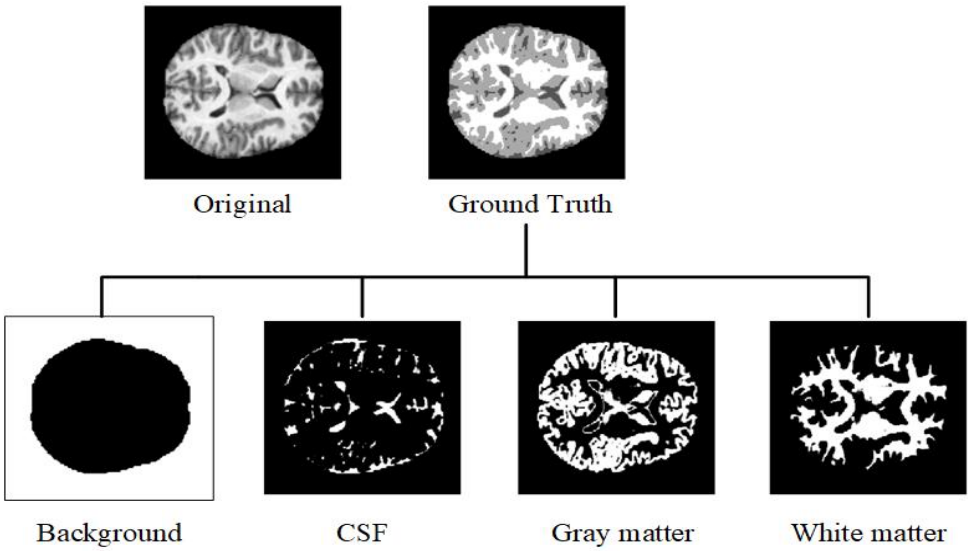


Figure 1-1. Binary map of the brain MRI ground truth image and its segmented structures.

For example, MRI is a non-invasive and accurate method of assessing brain tissue for the diagnosis of brain diseases. As a result, the expert's main goal is obtaining meaningful information on brain tissue segmentation. The usage of highly qualified specialists is required for this activity. The brain MRI ground truth picture and its segmented structures are shown in Figure 1-1. On the other hand, manual tracing of brain components including grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) yields conflicting findings. A

CNN-based automated segmentation of brain regions from brain MRI, such as WM, GM, and CSF, is proposed which assists in the diagnosis and successful treatment of various brain illnesses.

1.4. Objectives

The utilization of numerous frames to provide coherent results and parallel structure to depict extremely sophisticated interactions between surrounding slices motivates the suggested innovative architecture for brain MRI image segmentation. Instead of a conventional way of using single input for the model, the input consists of three 2D slices labeled "late," "middle," and "early," with the "middle" slice in the middle expected.

The following is a list of the proposed model's major contributions:

- To extract associated information from many slices, using several slices is suggested as input, including nearby slices.
- A unique parallel UNet that preserves each input slice's distinct spatial information.
- To boost performance, merging the filters from UNets at the end with a residual network.
- Experimentation around with scaling pictures generated using OASIS data. The suggested technique does not employ additional or extra procedure, or pre- or post-processing of the photographs, except from scaling the 2D images.
- Examination of the most up-to-date approaches, including traditional UNet and a three image input accepting modified UNet. The suggested strategy outperforms the other approaches.

1.5. Thesis Layout

There are five chapters in this thesis. Following the introduction in Chapter 1, Chapter 2 gives an overview of sematic segmentation algorithms for medical and natural photos. The suggested approach and pipeline are explained in Chapter 3, notably the parallel UNet architecture and fusion with residual network. The experimental results are compared to the results of state-of-the-art algorithms in Chapter 4. Finally, Chapter 5 wraps up with a brief overview of the study's findings.

2. Related Works

In image processing, semantic segmentation assigns a distinct class label to each pixel for localization [13]. MRI is the most widely used non-invasive technique for studying the brain in medical image processing, producing a contrast picture in the tissue for the characteristics of interest by repeating various excitations [37]. MRI can identify disorders that affect the brain, such as Alzheimer's disease (AD) and multiple sclerosis (MS) [38]. The term "tissue atrophy" is frequently used to describe the symptoms of Alzheimer's disease. It's critical to accurately identify and categorize the sick tissue and its surrounding healthy tissues when diagnosing disorders like Alzheimer's. For a more accurate diagnosis, a vast amount of data is necessary. Physicians, on the other hand, may find manually evaluating vast and complex MRI datasets and extracting critical information challenging. In addition, manual brain MRI image analysis is time-consuming and error-prone [39]. As a result, for precise and dependable findings, an automated segmentation approach must be devised. Large datasets have recently been utilized to evaluate computer aided MRI segmentation to aid clinicians in making qualitative diagnoses. The brain's structural alterations may also be assessed by MRI segmentation at different time periods.

2.1. Medical Image Segmentation

Conventional approaches for picture segmentation, such as clustering and thresholding, were utilized before convolutional neural networks (CNN) to locate object boundaries with low-level information [15]. For pixel-level localization of scene labels, a number of graphical models have been utilized [15]. When it comes to segmenting neighboring class labels, these solutions fall short. Graphical models, such as Conditional Random Forests (CRFs) [16],

are still employed as refinement layers to increase performance. Deep learning algorithms from the beginning fine-tuned fully linked categorization layers [17]. Overfitting and insufficient depth for developing abstract features were solved using a refinement method in these investigations [17, 18]. CNN has been widely employed in several segmentation tasks in recent studies [14]. The limits of classical pixel classification have been solved by CNN. The capacity of deep convolutional neural networks to autonomously learn features has been shown to improve performance [20]. Patches, sliding windows, and fully connected CRFs, among other things, were used in previous CNN techniques to picture segmentation. These methods contain redundant calculations and are unable to learn global characteristics [21]. In the absence of completely connected layers from CNN, a fully convolutional network (FCN) architecture for supervised pixel-wise prediction with marginal number of weights in convolution layers was significantly quicker, avoiding the constraints of prior techniques [19]. This design was significant in segmentation research since it allowed for the generation of segmentation maps for pictures of any resolution [15]. The combining of information from various filter layers is possible with FCN and the "Skip" design [19].

2.2. Parallel UNet Architecture

UNet uses the same architecture as FCN and SegNet [1], which is entirely made up of convolutional layers. UNet has a symmetric design and comprises of an encoder and a decoder [22]. The encoder reduces the spatial dimension by using pooling layers, whereas the decoder recovers it [23]. The skip connections allow information to be sent at the same level from the encoder to the feature map of the decoder. Several research have recently proposed several UNet versions to increase the performance of medical picture

segmentation [24]. Most of the research [25, 26, 27, 28] employed single UNet topologies with a variety of changes such batch normalization, data augmentation, and patch-wise segmentation. Few designs have been described in recent years that use more than one UNet. For X-ray picture recognition and localization, a two-parallel UNet technique was presented [29]. Multi-Inputs UNet (MI-UNet) is another variation that uses several inputs to extract parcellation information from brain MRI [30]. When many UNets are used, merging the output or layers within them becomes a difficult operation. One method involves feeding the output of one parallel UNet to the watershed algorithm as a seed for segmenting the output of another UNet [31]. To make use of multimodal data, inputs were contracted separately and then concatenated before being decoded into a single output [32]. TMD-UNet consists of three concurrent sub-UNet models with modified node topologies [33]. TMD-UNet, unlike the typical UNet architecture, uses all of the convolutional units' output features as input for the following nodes.

2.3. Late Fusion Approach

As the neural network's depth grows, the accuracy becomes saturated and eventually deteriorates. The residual network [35] proposed a paradigm for addressing the deterioration problem. In this strategy, the shortcut connections conduct identity mapping, and the outputs of these connections are added to the stacked layer's outputs. Identity shortcut connections can be readily created and taught end-to-end with backpropagation [35], with no additional parameters or computational complexity. ResUnet-a modified residual blocks to replace the UNet architecture's building blocks [36]. The change allowed for the semantic segmentation of high-resolution pictures to be labeled.

3. Proposed Method

A novel architecture is offered for the segmentation of brain MRI images, motivated by the use of multiple frames to produce coherent findings and multi-path parallel architecture to describe very complicated interactions between surrounding slices. Rather of using a single slice as input, three sequential 2D slices is employed labeled 'early, middle, and late,' with the central 'middle' slice predicted as shown in Figure 3-1(a) and -(c).

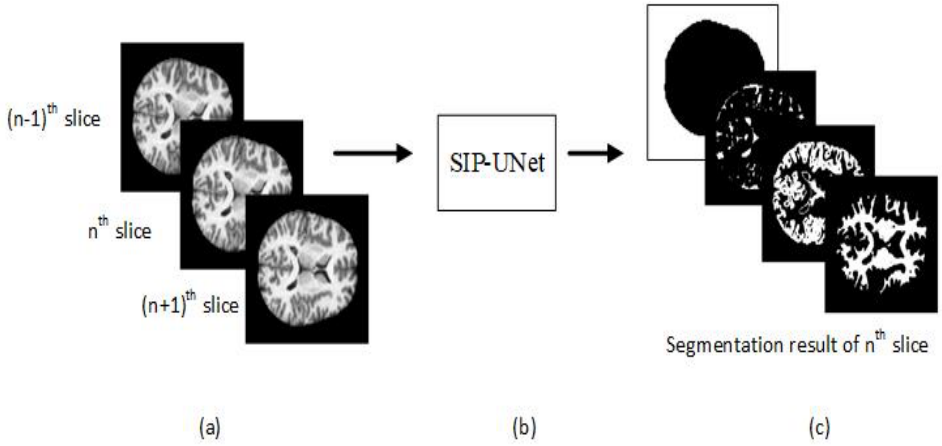


Figure 3-1. Input for the proposed model (a) 3 sequential input where $(n+1)^{th}$, n^{th}, and $(n-1)^{th}$ slice are later, central, and early images/slices respectively, (b) the proposed novel model, and (c) segmentation result generated for the n^{th} slice.

The spatial information is believed to be connected with the 'middle' slice is made up of the nearby slices 'early' and 'late.' The three slices are run through three distinct UNets before being fused to forecast the 'middle' slice. The model's multi-paths fusion is driven by *Nie et al*., who discovered that late fusion improves performance [42]. A unique way is presented for fusing parallel UNets with the help of a late merging residual network. To the stacked

outputs residual network, UNet outputs are added for the 'middle' slice. To improve performance, the model may learn from nearby layers as well as reinforce and maintain the attributes of the center slice.

3.1. Proposed Architecture

The proposed serial-input parallel UNet (SIP-UNet) differs from traditional UNet in that it may use and extract characteristics from surrounding slices or pictures. In a normal UNet, there is just one related picture that is utilized as input. However, in SIP-UNet, the segmentation result of the core slice is also obtained using two nearby slices ('early' and 'later'). MRI images are sliced for input data having three consecutive slices for each individual and then fed into SIP-UNet together. The difference between the input for a regular UNet and the proposed SIP-UNet is seen in Figure 3-2(a) and -(b).

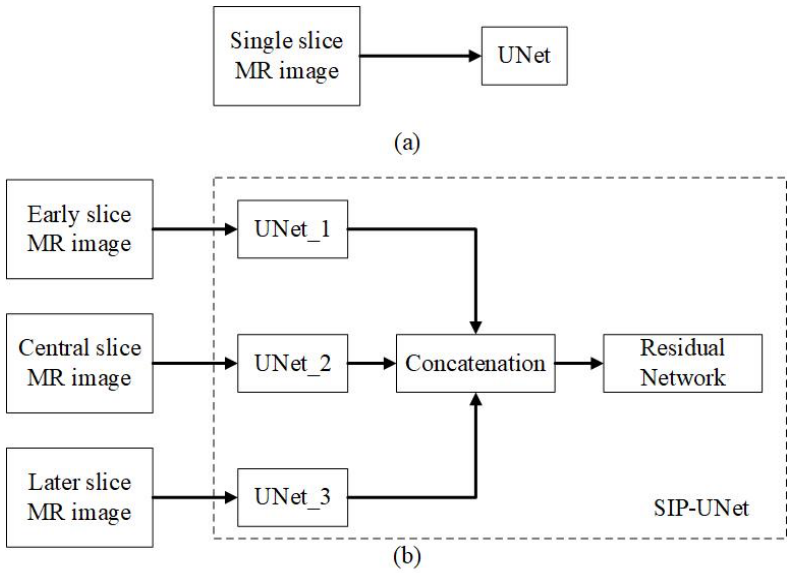


Figure 3-2. Illustration of difference between: (a) conventional UNet and (b) proposed model

A traditional UNet and the UNet utilized in the SIP-UNet are shown in Figure 3-3(a) and –(b). The proposed SIP-UNet is made up of two primary components: a parallel UNet and a residual network for late fusion. For different views (axial, sagittal, and coronal) of brain MRI, the model is trained and tested separately.

3.2. Parallel UNets

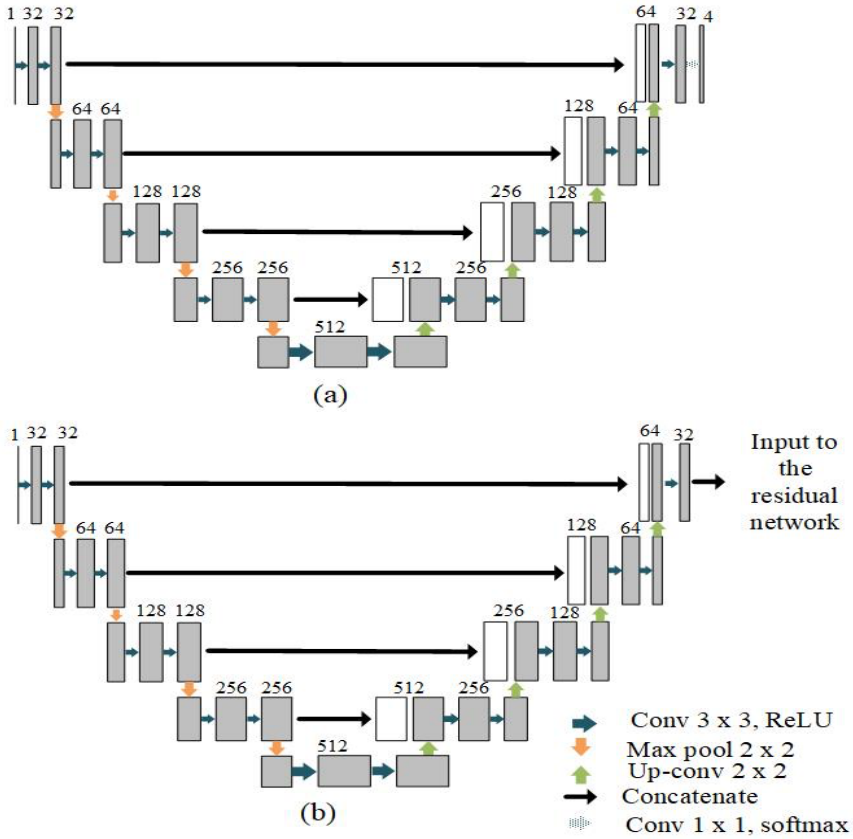


Figure 3-3. Architecture of: (a) a conventional segmentation UNet and (b) one of the UNets utilized to encode and decode single input in SIP-UNet.

The parallel UNet is made up of three typical UNets that are all the same. Each of the slices from the input data is passed to a separate UNet. Figure 3-3(a) and -(b) depicts the architecture of each unique UNet that constructs the parallel structure in Figure 3-2(b). An encoder path and a decoder path make up the UNet architecture. A completely convolutional network design is used in both the encoder and decoder. A 3 × 3 convolution is done twice in the decoder route, followed by a 2 × 2 max pooling operation, which doubles the number of feature channels at each down-sampling step. Alternatively, the decoder route includes a 2 × 2 up-convolution, which reduces the number of feature channels by half, followed by concatenation with the equivalent encoder path feature map and two 3 × 3 convolutions. At the end of the process, a feature map with 32 layers is created. In both the encoder and decoder parts, each convolutional operation is followed by a ReLU [31] activation. Table 1 shows the structural features of a typical UNet. The last convolution block, which is followed by the Softmax function, is omitted from the parallel UNet's construction block. Later, a suggested residual network is employed to merge the feature maps from the several UNet pathways.

3.3. Proposed Fusion Using Residual Network

Using a residual network, a new way for combining the features of parallel UNet designs is presented. To begin, each UNet's 32 feature maps are concatenated as illustrated in Figure 3-4. Two three-by-three convolutions are conducted on the concatenated feature maps, as illustrated in Figure 3-4. The feature maps of the core slice are then updated with the output of the stacked layers. The skip connection is only utilized for the core slice's feature maps. Skipping the link for the 'central' slice feature maps is believed to maintain

and reinforce the information while preventing the model from learning superfluous features from nearby slices.

Formally, the feature maps from 'early', 'central', and 'later' slices are denoted as x_e, x_c, and x_l respectively, and concatenated layers as x', and the concatenated layers are allowed to fit another mapping of $\mathcal{F}(x')$. The underlying mapping $\mathcal{H}(x')$ is defined as:

$$\mathcal{H}(x') = \mathcal{F}(x') + x_c \tag{1}$$

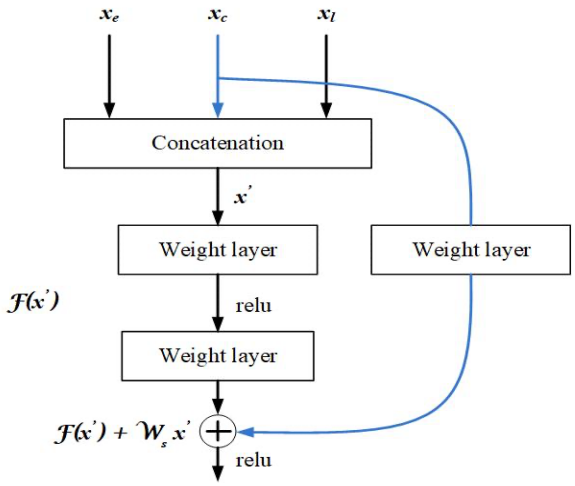


Figure 3-4. The building block of late fusion residual learning for merging the UNets. x_e represents early slice processing UNet's final features, x_l represents later slice processing UNet's final features, and x_c represents central slice processing UNet's final features, which is also passed using the skip connection in fusion model.

To maintain equal dimension of x_c and $\mathcal{F}$, a linear projection W_s within a skip connections is performed. The building block of the residual network considered in this paper is defined as:

$$y = \mathcal{F}(x', \{W_i\}) + W_s x_c \quad (2)$$

Here, x' and y are the considered building block's input and output vectors respectively. $\mathcal{F}(x', \{W_i\})$ function represents a residual mapping to be learned. In Figure 3-4, to omit biases for simplification of notations, we get $\mathcal{F} = W_2 r(W_1 x')$, where r denotes the ReLU [43]. In proposed model, the flexible residual function $\mathcal{F}$ has two layers as shown in Figure 3-5. Even though more layers are possible, it will represent a linear layer while using a single layer. For the linear layer: $y = W_1 x' + W_s x_c$ there are no observed advantages in the residual network [37]. The function $\mathcal{F}(x', \{W_i\})$ in Equation 2 represents convolutional layers. The structure detail of the proposed residual block is shown in Table 2.

3.4. Loss Function

The goal of this research is to categorize brain MRI pictures down to the pixel level. The model is trained to predict which of the four groups each pixel belongs to. After the final convolution layer, the softmax activation function is applied for the multi-class prediction model. In the ground truth, the truth labels are integer encoded. Background is 0, CSF is 1, GM is 2, and WM is 3. The sparse categorical cross-entropy loss function is the most widely utilized loss function for this type of multi-class segmentation problem. The following is a definition of cross-entropy:

$$\mathcal{L} = -\frac{1}{|P|} \sum_{p \in P} log\ (\sigma_i) \quad (3)$$

Table 1. Architecture of a single UNet

Layer Name	Output Shape	Connected To
Input_1	256 × 256 × 1	
Conv2d	256 × 256 × 32	Input_1
Conv2d_1	256 × 256x 32	Conv2d
Max_pooling2d	128 × 128 × 32	Conv2d_1
Conv2d_2	128 × 128 × 64	Max_pooling2d
Conv2d_3	128 × 128 × 64	Conv2d_2
Max_pooling2d_1	64 × 64 × 64	Conv2d_3
Conv2d_4	64 × 64 × 128	Max_pooling2d_1
Conv2d_5	64 × 64 × 128	Conv2d_4
Max_pooling2d_2	32 × 32 × 128	Conv2d_5
Conv2d_6	32 × 32 × 256	Max_pooling2d_2
Conv2d_7	32 × 32 × 256	Conv2d_6
Max_pooling2d_3	16 × 16 × 256	Conv2d_7
Conv2d_8	16 × 16 × 512	Max_pooling2d_3
Conv2d_9	16 × 16 × 512	Conv2d_8
Conv2d_transpose	32 × 32 × 256	Conv2d_9
Concatenate	32 × 32 × 512	Conv2d_transpose Conv2d_7
Conv2d_10	32 × 32 × 256	Concatenate
Conv2d_11	32 × 32 × 256	Conv2d_10
Conv2d_transpose_1	64 × 64 × 128	Conv2d_11
Concatenate_1	64 × 64 × 256	Conv2d_transpose_1 Conv2d_5

Layer Name	Output Shape	Connected To
Conv2d_12	64 × 64 × 128	Concatenate_1
Conv2d_13	64 × 64 × 128	Conv2d_12
Conv2d_transpose_2	128 × 128 × 64	Conv2d_13
Concatenate_2	128 × 128 × 128	Conv2d_transpose_2 Conv2d_3
Conv2d_14	128 × 128 × 64	Concatenate_2
Conv2d_15	128 × 128 × 64	Conv2d_14
Conv2d_transpose_3	256 × 256 × 32	Conv2d_15
Concatenate_3	256 × 256 × 64	Conv2d_transpose_3 Conv2d_1
Conv2d_16	256 × 256 × 32	Concatenate_3
Conv2d_17	256 × 256 × 32	Conv2d_16
Conv2d_18	256 × 256 × 4	Conv2d_17

All “conv2d” corresponds to a 3x3 convolution block followed by ReLU activation function except for the last convolution block which is followed by the Softmax Function.

Table 2. Architecture of the proposed residual network for fusion of the parallel UNets

Layer name	Output shape	Connected to
Concatenate_12	256 × 256 × 96	Conv2d_16 Conv2d_33 Conv2d_50
Conv2d_51	256 × 256 × 64	Concatenate_12
Conv2d_52	256 × 256 × 64	Conv2d_51
Conv2d_53	256 × 256 × 64	Conv2d_33
Add	256 × 256 × 64	Conv2d_52 Conv2d_53
Conv2d_54	256 × 256 × 32	Add
Conv2d_55	256 × 256 × 32	Conv2d_54
Conv2d_56	256 × 256 × 32	Add
Add_1	256 × 256 × 32	Conv2d_55 Conv2d_56
Conv2d_57	256 × 256 × 4	Add_1

Two convolution blocks: conv2d_51 and conv2d_54 are followed by ReLU activation function. Similarly, two addition blocks: add and add_1 are also followed by ReLU activation function. The final convolution block "conv2d_57" is followed by Softmax function and generates segmented output.

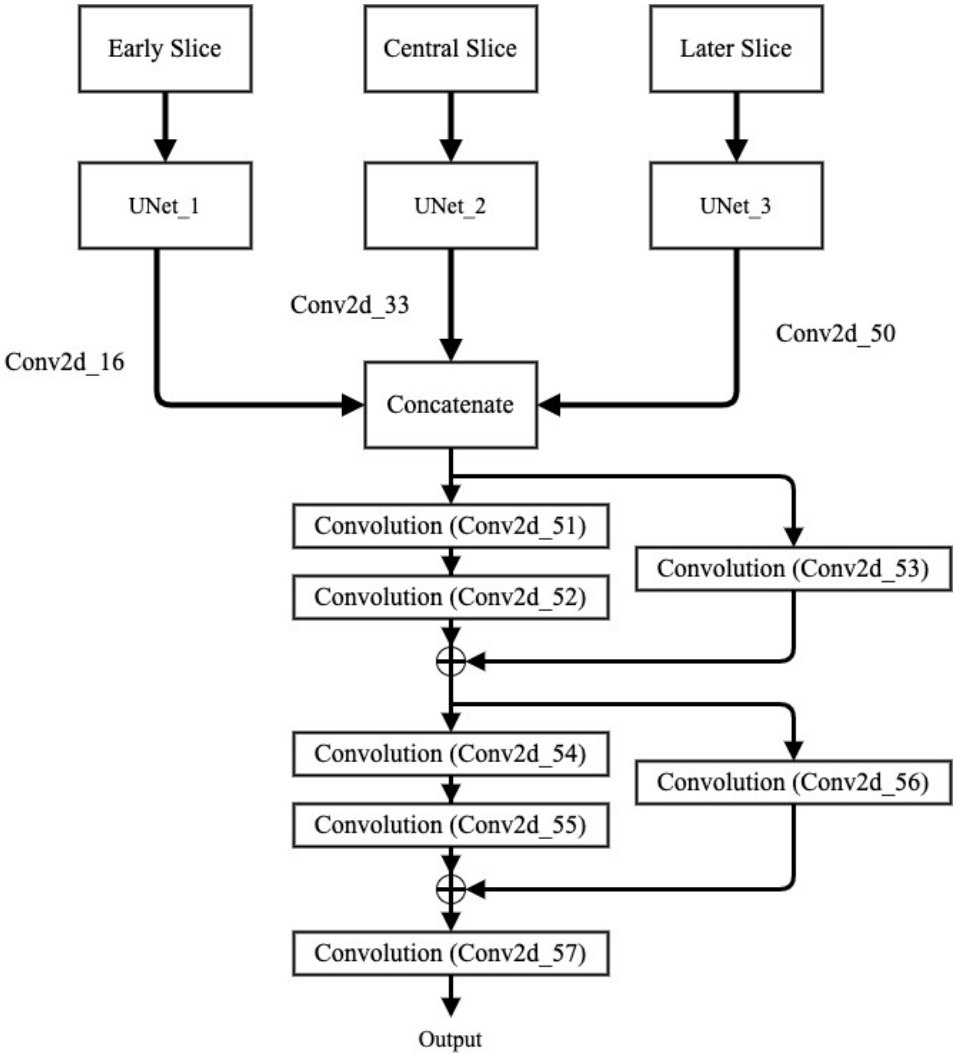


Figure 3-5. The proposed residual network for the fusion of the filters from parallel UNets. The corresponding layer names are from the Table 2. Showing the connections of the convolutional layers.

4. Experimental Analysis

4.1. Datasets

The proposed method was trained and evaluated using OASIS dataset. A detailed description of the dataset is discussed as follow.

4.2. OASIS Dataset

The Open Access Series of Imaging Studios (OASIS) dataset [45] is used to test the suggested approach. In OASIS, there are 413 demented individuals and 20 non-demented participants. 50 participants were chosen at random for training and the remaining 386 subjects for testing the model out of the 436 total. The MRI scan and segmented 3D picture of each participant have dimensions of 176 × 208 × 176. Two-dimensional pictures are created in three planes: axial, coronal, and sagittal. There are certain empty photos that do not have information about the brain in all planes when converting each image to 2D data. To eliminate empty 2D pictures, only slices from 15 to 145 in the axial plane, 30 to 180 slices in the coronal plane, and 25 to 145 slices in the sagittal plane are used. All photos are scaled to 256 × 256 dimensions since these 2D slices vary in size across different dimensions and planes. Three successive photographs are concatenated after resizing them to produce an input for the proposed model. The size after concatenation is 256 × 256 × 3. Because colorless photos are utilized, the third dimension represents various slices rather than channels. The 'early' slice is represented by the first slice, the 'center' slice by the second slice, and the 'late' slice by the third slice.

For training, 50 patients are employed from the OASIS dataset. The models for various planes are trained using the same subjects. In the axial plane, 130 2D slices (ranging from 15 to 145) are collected from each individual. In the

coronal and sagittal planes, 150 2D slices (from 30 to 180) and 120 2D slices (from 25 to 145) are retrieved from each unique patient. The axial, coronal, and sagittal planes yield 6500, 7500, and 6000 pictures, respectively. For each plane, the models are trained and evaluated individually. The remaining 386 participants are utilized for testing and extract the photos in the same way as for training. In total 50180, 57900, and 46320 pictures are obtained as test data from the axial, coronal, and sagittal planes, respectively.

4.3. Evaluation Matrics

The Dice Similarity Coefficient (DSC) and the Jaccard Index are used to assess performance (JI). JI is the ratio of the expected and ground-truth pictures' overlapping area to the union area between them. Another statistic is the DSC, which is the ratio of the total number of pixels to two times the overlapping area of the ground truth and forecasted pictures.

For the ground truth segmentation map I and the predicted segmentation map I', JI and DSC are defined in Equation 4 and Equation 5 respectively,

$$\mathrm{JI} = \frac{|I \cap I'|}{|I \cup I'|} \tag{4}$$

$$\mathrm{DSC} = \frac{2|I \cap I'|}{|I| + |I'|} \tag{5}$$

Because the segmentation has four classes, JI and DSC are computed independently for each class. Performance on background segmentation is not tested among the four classes. Simpler models can also be used for background segmentation. For the remaining three classes, CSF, GM, and WM, the performance of the models are compared. The evaluation metrics for each class are determined individually in multi-class segmentation. If we wish to compute JI for CSF, for example, pixels connected to CSF are given a value

of 1, while the rest of the pixels are given a value of 0. For GM and WM, the same approach is used.

4.4. Experimental Setup

The Keras platform will be used to implement the proposed network. This experiment was conducted on a workstation with an Intel® Core™ i9-10900K CPU, 64 GB RAM, and an NVIDIA RTX 3090 GPU. The proposed model is designed for the problem of 2D segmentation. There are several views/planes in the 3D medical picture data: (i) axial, (ii) sagittal, and (iii) coronal. The model is trained for each plane separately and use it to generate predictions and test it. Initially 2D slices from 3D data are extracted before concatenating three successive slices. The 'center' slice is anticipated from these three slices. As a result, the 'central' slice's ground truth is taken as the training output. 255 × 255 × 3 is the input dimension for the proposed model. The 2D segmentation problem is what the proposed model is for. There are three types of views/planes in the 3D medical picture data: axial, sagittal, and coronal. For each plane, the model is trained separately and use it to generate predictions and test it. Two-dimensional slices are extracted from three-dimensional data before concatenating three slices together. The 'center' slice is calculated from these three segments. As a result, the 'central' slice's ground truth is used as the training output. The proposed model requires a 255 × 255 × 3 input dimension.

The plane that corresponds to the training plane is extracted and predicted during testing. When a model is trained using an axial plane, for example, the data connected to the axial plane is tested. The test input, like the training data, consists of three successive 2D slices that are delivered to the model. The 'central' slice's ground truth is also saved, and this is eventually used to evaluate the model by comparing it to the result.

4.5. Results and Discussions

In section 4.5.1, a research utilizing the input of a single slice and many slices is run in a basic UNet model and illustrate how SIP-UNet improves segmentation. In section 4.5.2, the segmentation performance is analyzed and compared with several existing models.

4.5.1. Analysis and Comparison with Single and Multiple slice UNet

When it comes to segmenting biological pictures, UNet outperforms the competition. For segmentation, first the UNet is put to the test using a single slice input. The goal of testing with a single slice input is to examine if the results are better with or without the surrounding slice's features as compared to testing with a multi-slices input. Later, the input layer of the same UNet is adjusted to make it appropriate for processing inputs with nearby slices. This UNet's input will consist of three 2D planes. The planes are made up of slices that are next to each other.

The majority of prior research [20, 23, 46] only used a limited number of slices. In [23, 46], the slices were alternated or only one slice was chosen from a group of many slices. Because the 2D slices are similar, the goal of predicting and training just a limited number of slices is to avoid duplicating information from surrounding slices. These approaches, however, are insufficient for quantifying brain alterations since they do not take into account the complete brain. All of the slices that make up portions of the brain are included in the investigation. This aids in measuring changes in each layer and, in turn, leads to matching 3D quantification findings.

Table 3 shows the results of the DSC and JI scores for UNet with single and multiple slices using the planned SIP-UNet. The results are the averages of the

test photos for the relevant plane. The DSC score in the SIP-UNet is marginally improved for the single slice and multi-slice UNets, as seen in the table. For WM, GM, and CSF, the DSC values obtained using the standard UNet are 0.953, 0.955, and 0.941, respectively, whereas the DSC scores obtained using the SIP-UNet are 0.954, 0.959, and 0.949, respectively. The DSC score of the UNet with single input and the UNet with multi-slices input is nearly comparable. The coronal and sagittal planes show a small improvement in DSC values, similar to the axial plane.

The enhanced JI scores in Table 3 for the identical testing pictures demonstrate the usefulness of the SIP-UNet. The JI of the GM and CSF in the axial plane is much superior than the other conventional UNets. The JI score improves by 1% for both tissue segmentations in this case. In the same way, WM and GM improve by the same proportion in the sagittal plane, while WM improves by the same amount in the coronal plane. In the JI of the GM and CSF in the coronal plane and CSF in the sagittal plane, the suggested SIP-UNet scores 2% higher. On the same data, the suggested SIP-UNet model has a higher JI, hence the results are better than the standard UNet model.

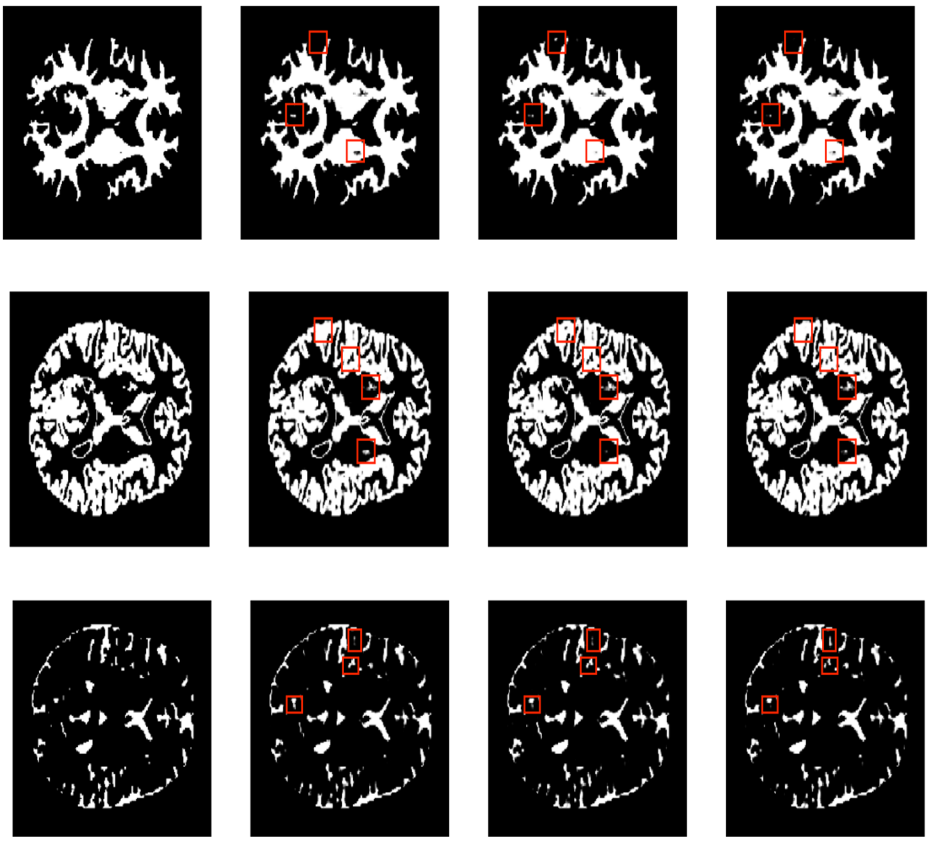
The outcomes are visually examined of specific slices of a random subject to assess the improvement in the SIP-UNet result. The binary mapping of the related tissue makes comparing particular tissue segmentation easy. A binary map of WM, GM, and CSF are built individually, then compare it to the ground truth and the outcomes of the various models. The result for the axial plane is shown in Figure 4-1. The ground truth and outcomes of the one slice input UNet, the multi-slice input UNet, and the suggested SIP-UNet are represented in the column. From top to bottom, the row shows a binary map of the various tissues: WM, GM, and CSF. The binary map in Figure 4-1 is a 70^{th} axial plane

slice from a randomly selected individual. A red box in the figures highlights the difference in the pattern of the binary map of the two models. The binary map of the multi-slice UNet output in the first row of WM contains a false prediction in a location marked by the red box at the top of the picture. In the other two models, however, there is no such incorrect prediction. The box in the center of the same row depicts the incorrect prediction in all three model outputs. In this scenario, there is no such tissue in the area covered by the middle red box in the ground truth, yet the incorrect prediction can be seen in all of the outputs. However, in close examination, it can be observed that the size of the incorrect projected tissue in the middlebox in the SIP-UNet result is far smaller. As a consequence, even in the incorrect prediction zone, the suggested SIP-UNet model's output is closer to the ground truth than the other two models. The location where the single slice UNet failed to predict the existence of tissue in that region, but the other two models were successful, is shown in the last red box from the top in the first row. The suggested SIP-UNet produces a result that is closer to the ground truth than the other two highlighted locations in the first row.

The binary map of GM is compared in the second row of Figure 4-1. The first highlighted area depicts the area where UNet's multi-slices input predicted false. All three models were successful in the remaining three highlighted locations, however the shape and edge of the projected outcome in these regions varies in the single-slice and multi-slice input UNet. With the suggested SIP-UNet model, the proper edge and area of the tissues in these locations are right. The CSF findings are found on the third row. The single slice and multi-slice UNet models, like GM in the second row, predicted the tissues in the highlighted locations but were unable to provide a result with the

same area and edge as the ground truth. The SIP-UNet, on the other hand, predicts tissues in this region with a comparable edge and shape to the ground truth.

The segmentation results for the coronal plane are shown in Figure 4-2. The subject is chosen at random, and the 30th slice in the coronal plane of this subject is segmented using several models and shown in columns. The WM binary map is shown in the first row, and it can be seen that the SIP-UNet result (last column) can forecast a very tiny region containing WM, however the other two UNets with a single slice and multi-slice are unable to do so. A area in the upper left corner that was inaccurately predicted is highlighted in the lower two rows of GM and CSF. As illustrated in the last row, GM's miscalculation resulted in no CSF in that area. This misclassification, however, occurs in all three models. Despite the fact that the proposed model isn't ideal, it is a noticeable improvement in the highlighted areas. The single slice and multi-slice UNets are unable to predict the presence of smaller tissues in the remaining highlighted regions. The two highlighted locations to the right of the GM projected results (second row) demonstrate that the conventional UNet is incapable of detecting finer features. The SIP-UNet, on the other hand, worked admirably in these tiny areas, keeping the tissue margins near to the ground reality.



(a) (b) (c) (d)

Figure 4-1. Axial plane segmentation result (binary map) for WM, GM, and CSF (top to bottom): (a) ground truth, segmentation obtained by (b) UNet with single input, (c) multiple input UNet, and (d) suggested model (SIP-UNet).

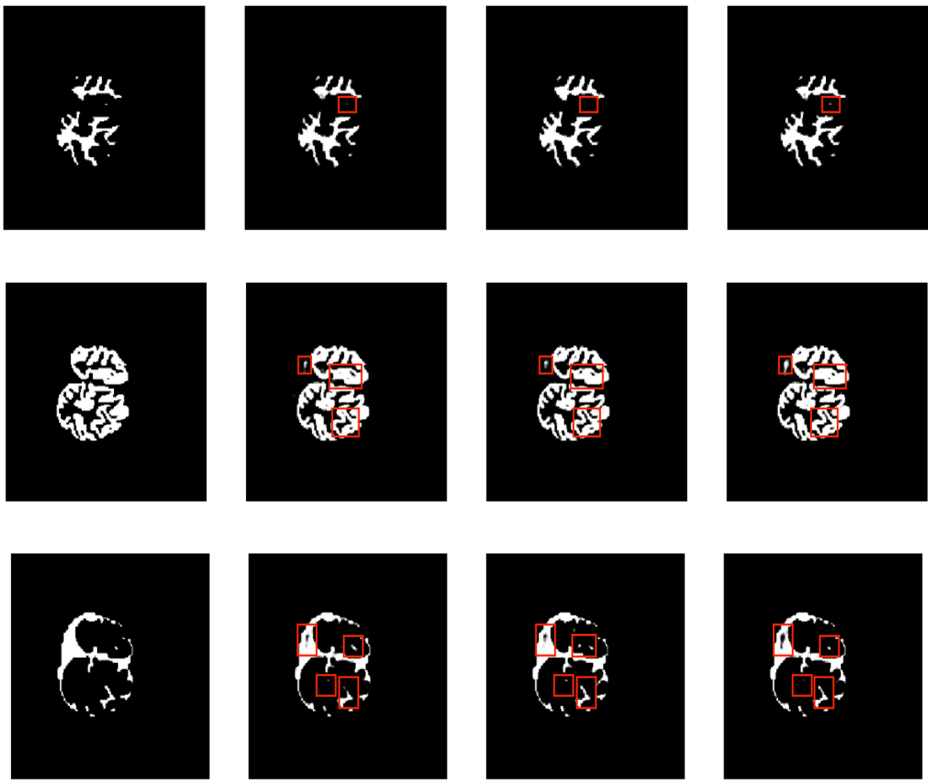
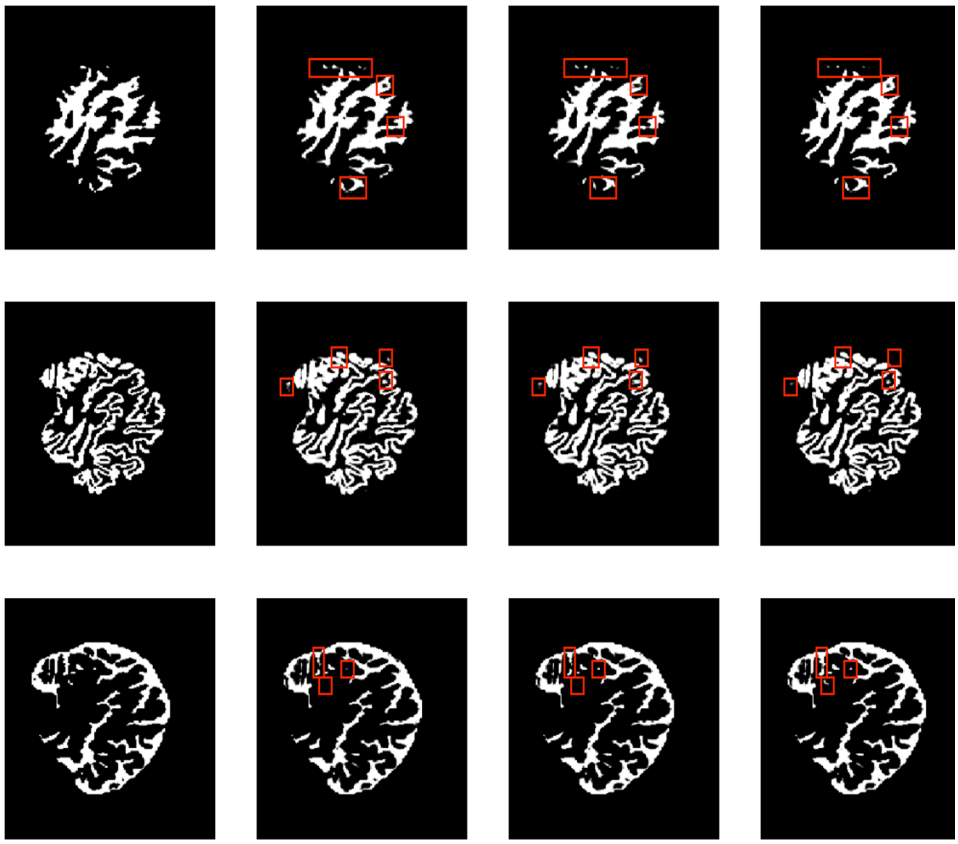


Figure 4-2. Coronal plane segmentation result (binary map) for WM, GM, and CSF (top to bottom): (a) ground truth, segmentation obtained by (b) UNet with single input, (c) multiple input UNet, and (d) suggested model (SIP-UNet).



(a) (b) (c) (d)

Figure 4-3. Sagittal plane segmentation result (binary map) for WM, GM, and CSF (top to bottom): (a) ground truth, segmentation obtained by (b) UNet with single input, (c) multiple input UNet, and (d) proposed model (SIP-UNet).

Table 3. Segmentation result comparison between the Multiresnet, SegNet, single slice input UNet, multi-slice UNet, and the proposed SIP-UNet

Methods	No. of input slices	WM		GM		CSF	
		DSC	JI	DSC	JI	DSC	JI
Axial plane							
Multiresnet [14]	1	0.692±0.174	0.551±0.168	0.753±0.068	0.608±0.079	0.728±0.081	0.579±0.091
SegNet [13]	1	0.843±0.045	0.731±0.065	0.871±0.050	0.775±0.076	0.843±0.045	0.731±0.065
Unet	1	0.953±0.067	0.915±0.084	0.955±0.027	0.915±0.042	0.941±0.033	0.890±0.053
Unet (modified)	3	0.951±0.069	0.912±0.087	0.955±0.028	0.915±0.043	0.944±0.032	0.895±0.052
Proposed method	3	0.954±0.066	0.917±0.083	0.959±0.027	0.922±0.042	0.949±0.031	0.904±0.050
Coronal plane							
Multiresnet [14]	1	0.732±0.079	0.583±0.090	0.758±0.048	0.613±0.060	0.740±0.060	0.591±0.073
SegNet [13]	1	0.899±0.038	0.818±0.0595	0.882±0.031	0.790±0.048	0.858±0.041	0.753±0.061
Unet	1	0.963±0.023	0.929±0.039	0.952±0.022	0.909±0.035	0.935±0.034	0.880±0.054
Unet (modified)	3	0.964±0.024	0.932±0.0395	0.956±0.022	0.917±0.035	0.945±0.032	0.897±0.051
Proposed method	3	0.965±0.023	0.933±0.038	0.958±0.022	0.921±0.035	0.949±0.031	0.904±0.049
Sagittal plane							
Multiresnet [14]	1	0.732±0.103	0.587±0.112	0.766±0.037	0.622±0.046	0.733±0.053	0.581±0.065
SegNet [13]	1	0.866±0.063	0768±0.092	0.862±0.034	0.759±0.051	0.834±0.041	0.716±0.058
Unet	1	0.954±0.031	0.914±0.050	0.949±0.022	0.904±0.036	0.937±0.030	0.882±0.046
Unet (modified)	3	0.955±0.032	0.916±0.051	0.952±0.022	0.909±0.035	0.940±0.0297	0.888±0.46
Proposed method	3	0.959±0.031	0.923±0.049	0.957±0.021	0.918±0.034	0.949±0.028	0.903±0.043

In the sagittal plane, the SIP-UNet performed better than the axial and coronal planes. The segmented results of the various models are placed in separate columns in Figure 4-3, with the rows representing the various tissues as previously. In the top highlighted region of the first row, the tissue segmentation miss-prediction (WM binary map) can be observed. In this location, all models have the incorrect segmented output, but if checked closely, the segmented WM in this region is comparably lower in the SIP-UNet result, indicating that it is close to the ground truth. In the first row, the bottom highlighted region demonstrates how effectively the edge is anticipated in the SIP-UNet. The multi-slice input UNet has unconnected tissue in the same region, but the SIP-UNet has a region that is near to the ground truth. Similarly, both the single-slice and multi-slice input UNets have unconnected tissue in the top left highlighted region of the last row (CSF binary map) of Figure 4-3, but the SIP-UNet output has linked tissue that fits the ground truth. The rest of the highlighted locations in Figure 4-3 second and third rows show how effectively the SIP-UNet operates in areas where the standard UNet fails. In summary, the suggested architecture is capable of extracting smaller tissue details and edges than previous implemented models and traditional UNets. Despite the fact that the DSC scores of standard UNets are almost comparable to those of the suggested approaches, the proposed method's JI score is higher, suggesting greater performance.

4.5.2. Comparison with Other Methods

On the same dataset, Table 3 compares the performance of Multiresnet, SegNet, the traditional UNet, and the proposed UNet. Same data is used to train and test all of the models indicated in Table 3. The code from GitHub is

Table 4. Comparison of different approaches for brain structure segmentation

Authors	Methods	DSC score		
		WM	GM	CSF
Zhang *et al.* [48]	CNN	86.4%	85.2%	83.5%
Nie *et al.* [49]	FCN	88.7%	87.3%	85.5%
Khagi *et al.* [20]	SegNet	81.9%	74.6%	72.2%
Lee *et al.* [47]	Patch-wise UNet	94.33%	93.33%	92.67%
Yamanakkanavar *et al.* [50]	Patch-wise Mnet	95.17%	94.32%	93.60%
Proposed method	SIP-UNet	95.93%	95.8%	94.9%

used to implement Multiresnet and SegNet in this table. The multi-slice input UNet is a variant of the single-slice input UNet in which the model's input layer is switched from one 2D input at a time to three 2D inputs at the same time.

When compared to Multiresnet and SegNet, the suggested model has the greatest mean DSC score. In all planes, the suggested method's DSC score is between 94% and 96% for all three classes. The objective of the proposed method is to extract information from surrounding slices without using any data augmentations or further pre- or post-processing beyond shrinking the picture to fit in the model. Multiresnet only scored between 69% and 76% DSC without any extra processing, while SegNet scored between 83% and 89% DSC. Table 3 shows that all of the models have a lower DSC value than the recommended technique. In addition, among the implemented Multiresnet and SegNet, the suggested technique has the greatest JI score. The suggested method's average JI score is in the range of 90% to 94%, whereas Multiresnet's average JI score is in the range of 55% to 62% and SegNet's average JI score is in the range of 71% to 81%. The output for all slices of the brain and all

three planes in the developed models has worse performance for both evaluation matrices. This suggests that the proposed model is far superior.

The results in Table 4 for comparison are taken straight from the published studies. Four of the methods in this table have yet to be implemented. Three of the methods (CNN, FCN, and SegNet) in Table 4 are taken from Khagi et al. [20]. Lee et al. [47] provide the outcome of patch-wise UNet. The average score of the three separate planes for each tissue is used to determine the DSC score for each tissue in Lee et al. [47] and the suggested technique. The patch-wise Mnet scores are obtained straight from Yamanakkanavar et al. [50].

Table 4 shows that deep learning architectures based on UNet outperform alternative segmentation models. The proposed SIP-UNet has the highest mean DSC score, with 95.93%, 95.8%, and 94.9% for the WM, GM, and CSF, respectively. Patch-wise UNet outperforms other approaches and comes close to the suggested method in terms of performance. However, according to Lee et al. [47], only three-slice intervals are employed, resulting in 48 slices per individual. The patch-wise approach's outcome for all of the slices is unknown, and despite the fact that just 48 slices per subject each subject were employed in that study, the DSC score is lower than the suggested method. Another technique employing patch-wise Mnet, similar to the patch-wise UNet method, employs 48 slices per subject. The DSC scores of all deep learning-based approaches are lower than those of UNet-based deep learning systems. In conclusion, regardless of the number of slices or the plane chosen, the suggested technique may greatly improve segmentation performance in all three planes.

5. Conclusion

The ability to measure changes in the brain relies heavily on brain tissue segmentation. In this paper, a completely automated brain tissue segmentation algorithm is offered that extracts associated information from surrounding slices. Unlike traditional segmenation models, which use single input, the proposed SIP-UNet takes use of surrounding slices by extraction of additional data from those slices. For the axial, coronal, and sagittal 2D planes, SIP-UNet can get a superior segmentation result. The proposed segmentations are both qualitatively and quantitatively trustworthy, according to both qualitative and quantitative examination. The proposed technique outperforms existing deep segmentation models in all three planes, with average DSC and JI scores that are much higher. For the three classes/tissues GM, WM, and CSF, the average DSC values in OASIS dataset are 95.80%, 95.93%, and 94.9%, respectively. In the testing OASIS dataset, the average JI scores for WM, GM, and CSF are 92.43%, 92.03%, and 90.37%, respectively. The suggested technique provides a JI score that is comparable to that of traditional UNets. However, the suggested technique is equivalent to others in terms of DSC score, and there is yet opportunity for improvement. The objective is to improve by extracting information from nearby slices. The proposed model is made up of many UNets that are then combined with a residual network. This method’s goal is to extract characteristics from individual slices, without suppressing the middle/current slice's features. Achievement in extracting features from surrounding slices, led to an improvemed segmentation result, based on the evaluation matrix scores and visual comparison. The suggested technique performed well in terms of extracting fine details about edges and recognizing tiny tissue sections.

The suggested model takes processing time and more memory in comparison with a single UNet, which might be a major restriction of this research. However, the proposed technique has demonstrated good segmentation results. In the future, the computational efficiency can be increased of the proposed SIP-UNet architecture by modifying the UNet utilized in parallel. In addition, segmentation efficiency can be boosted even further by combining features from parallel UNets using a different technique. One possible approach could be increasing the skip connections. Also, establishing a deeper residual network could be a viable solution.

References

[1] M. Thompson and L. G. Apostolova, "Computational anatomical methods as applied to aging and dementia," *Br J Radiol*, vol. 80 Spec No 2, pp. S78-91, Dec. 2007.

[2] Whitwell JL, Przybelski SA, Weigand SD, Knopman DS, Boeve BF, Petersen RC, and Jack CR "3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease," *Brain*, vol. 130, no. Pt 7, pp. 1777–1786, Jul. 2007.

[3] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Mild Cognitive Impairment: Clinical Characterization and Outcome," *Arch Neurol*, vol. 56, no. 3, pp. 303–308, Mar. 1999.

[4] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, and Phelps CH, "Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers Dement*, vol. 7, no. 3, pp. 280–292, May 2011.

[5] Tábuas-Pereira, M., Baldeiras, I., Duro, D., Santiago, B., Ribeiro, M. H., Leitão, M. J., Oliveira, C., and Santana, I "Prognosis of Early-Onset vs. Late-Onset Mild Cognitive Impairment: Comparison of Conversion Rates and Its Predictors," *Geriatrics (Basel)*, vol. 1, no. 2, Apr. 2016.

[6] Mosconi L, Mistur R, Switalski R, Tsui WH, Glodzik L, Li Y, Pirraglia E, De Santi S, Reisberg B, Wisniewski T, and de Leon MJ., "FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 36, no. 5, pp. 811–822, May 2009.

[7] N. Srivastava and R. Salakhutdinov, "Multimodal Learning with Deep Boltzmann Machines," *J. Mach. Learn. Res*., vol. 15, no. 1, pp. 2949–2980, Jan. 2014.

[8] J. B. Colby, J. D. Rudie, J. A. Brown, P. K. Douglas, M. S. Cohen, and Z. Shehzad, "Insights into multimodal imaging classification of ADHD," *Front Syst Neurosci*, vol. 6, p. 59, 2012.

[9] Charles Decarli and Steven T. DeKosky and Mony J. de Leon and Norman L. Foster and Nick and Fox and Richard Frank and Richard S. and Thies and W. Michael and Weiner and Zaven S. Khachaturian "The Use of MRI and PET for Clinical Diagnosis of Dementia and Investigation of Cognitive Impairment : *A Consensus Report*," 2004.

[10] Liu, C.-Y. Wee, H. Chen, and D. Shen, "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification," *Neuroimage*, vol. 84, pp. 466–475, Jan. 2014.

[11] P. Vemuri and C. R. Jack, "Role of structural MRI in Alzheimer's disease," *Alzheimers Res Ther*, vol. 2, no. 4, p. 23, Aug. 2010.

[12] Bernard, Olivier and Lalande, Alain and Zotti, Clement and Cervenansky, Frederick and Yang, Xin and Heng, Pheng-Ann and Cetin, Yoonmi and Patravali, Jay and Jain, Shubham and Humbert, Olivier and Jodoin, Pierre-Marc, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[13] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, Dec. 2017.

[14] N. Ibtehaz and M. Sohel Rahman, "MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation", arXiv preprint arXiv:1902.04049, 2019.

[15] I. Ulku, and E. Akagunduz, "A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images," arXiv preprint arXiv:1912.10230, 2019.

[16] J. Lafferty, A. McCalium, and F. CN Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," in *Proc. of the Eighteenth International Conference on Machine Learning,* 2001, pp. 282-289.

[17] Y. Ganin, and V. Lempitsky, "N4-Fields: Neural Network Nearest Neighbor Fields for Image Transforms", arXiv preprint arXiv:1406.6558, 2014.

[18] Feng Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou and P. E. Barbano, "Toward automatic phenotyping of developing embryos from

videos," in *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1360-1371, Sept. 2005.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation", arXiv preprint arXiv:1411.4038, 2014.

[20] B. Khagi, and, G. R. Kwon, "Pixel-Label-Based Segmentation of Cross-Sectional Brain MRI Using Simplified SegNet Architecture-Based CNN," in *Journal of Healthcare Engineering*, pp. 1-8. 10, Oct. 2018.

[21] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," in *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281-2292, Oct. 2019, doi: 10.1109/TMI.2019.2903562.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv preprint arXiv:1505.04597, 2015.

[23] N. Yamanakkanavar, J. Choi and B. Lee, "MRI Segmentation and Classification of Human Brain Using Deep Learning for Diagnosis of Alzheimer's Disease: A Survey," *Sensors*, vol. 20, no. 11, pp. 3243, 2020.

[24] N. Singh Punn, and S. Agarwal, "Modality specific U-Net variants for biomedical image segmentation: A survey," arXiv preprint arXiv:2107.04537, 2021.

[25] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, "nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation," arXiv preprint arXiv:1809.10486, 2018.

[26] P. Blanc-Durand, Axel Van Der Gucht, Niklaus Schaefer, Emmanuel Itti, and John O. Prior, "Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study," in *PLoS ONE,* vol. 13, no. 4, Apr. 2018, doi:10.1371/journal.pone.0195798.

[27] G. Tong, Y. Li, H. Chen, Q. Zhang, and H. Jiang, "Improved U-NET network for pulmonary nodules segmentation," *Optik*, vol. 174, pp. 460-469, 2018, Available: 10.1016/j.ijleo.2018.08.086.

[28] H. Dong, G. Yang, F. Liu, Y. Mo and Y. Guo, "Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional

Networks," *Communications in Computer and Information Science*, pp. 506-517, 2017.

[29] Q. Que, Z. Tang, R. Wang, Z. Zeng, J. Wang, M. Chua, T. Sin Gee, X. Yang, and B. Veeravalli, "CardioXNet: Automated Detection for Cardiomegaly Based on Deep Learning," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 612-615, doi: 10.1109/EMBC.2018.8512374.

[30] Y. Zhang, J. Wu, Y. Liu, Y. Chen, E. X. Wu and X. Tang, "MI-UNet: Multi-Inputs UNet Incorporating Brain Parcellation for Stroke Lesion Segmentation From T1-Weighted Magnetic Resonance Images," *in IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 526-535, Feb. 2021, doi: 10.1109/JBHI.2020.2996783.

[31] Y. Kong, H. Li, Y. Ren, G. Z. Genchev, X. Wang, H. Zhao, Z. Xie, and H. Lu, "Automated yeast cells segmentation and counting using a parallel U-Net based two-stage framework," *OSA Continuum*, vol. 3, no. 4, p. 982-992, 2020. Available: 10.1364/osac.388082 [Accessed 9 September 2021].

[32] J. Dolz, I. Ben Ayed, and C. Desrosiers, "Dense Multi-path U-Net for Ischemic Stroke Lesion Segmentation in Multiple Image Modalities," arXiv preprint arXiv:1810.07003, 2018.

[33] S.-T. Tran, C.-H. Cheng, T.-T. Nguyen, M.-H. Le, and D.-G. Liu, "TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation," *Healthcare*, vol. 9, no. 1, p. 54, Jan. 2021.

[34] M. Vu, G. Grimbergen, T. Nyholm and T. Löfstedt, "Evaluation of multislice inputs to convolutional neural networks for medical image segmentation", *Medical Physics*, vol. 47, no. 12, pp. 6216-6231, 2020.

[35] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[36] F.I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020. doi:10.1016/j.isprsjprs.2020.01.013.

[37] S. Bauer, R. Wiest, L. Nolte and M. Reyes, “A survey of MRI-based medical image analysis for brain tumor studies,” *Physics in Medicine and Biology*, vol. 58, no. 13, pp. R97-R129, 2013.

[38] C.-J. Hsiao, E. Hing, and J. Ashman, “Trends in Electronic Health Record System Use Among Office-based Physicians: United States, 2007-2012,” *National health statistics reports*, vol. 1, pp. 1-18, 2014.

[39] I. Despotović, B. Goossens, and W. Philips, W, “MRI segmentation of the human brain: challenges, methods, and applications,” *Computational and mathematical methods in medicine*, vol. 2015, 2015, doi:10.1155/2015/450341.

[40] H. Liu, and J. Jiang, “U-Net Based Multi-instance Video Object Segmentation,” arXiv preprint arXiv:1905.07826, 2019.

[41] F. Perazzi, A. Khoreva, R. Benenson, B. Schiele and A. Sorkine-Hornung, “Learning Video Object Segmentation from Static Images,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3491-3500, doi: 10.1109/CVPR.2017.372.

[42] D. Nie, L. Wang, Y. Gao and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 1342-1345, doi: 10.1109/ISBI.2016.7493515.

[43] V. Nair, and G. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” Proceedings of *ICML*, vol. 27, pp. 807-814, 2010.

[44] T. Rohlfing, "Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable," in *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 153-163, Feb. 2012, doi: 10.1109/TMI.2011.2163944.

[45] D.S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, “Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults,” *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498-1507, 2007.

[46] N. Yamanakkanavar and B. Lee, "A novel M-SegNet with global attention CNN architecture for automatic segmentation of brain MRI", *Computers in Biology and Medicine*, vol. 136, p. 104761, 2021. Available: 10.1016/j.compbiomed.2021.104761.

[47] B. Lee, N. Yamanakkanavar and J. Choi, "Automatic segmentation of brain MRI using a novel patch-wise U-net deep architecture", *PLOS ONE*, vol. 15, no. 8, p. e0236493, 2020.

[48] W. Zhang et al., "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation", *NeuroImage*, vol. 108, pp. 214-224, 2015.

[49] D. Nie, L. Wang, Y. Gao, & D. Shen, “Fully convolutional networks for multi-modality isointense infant brain image segmentation”, *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1342-1345, 2016.

[50] N. Yamanakkanavar and B. Lee, "Brain Tissue Segmentation using Patch-wise M-net Convolutional Neural Network," *2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, 2020, pp. 1-4, doi: 10.1109/ICCE-Asia49877.2020.9277239.

Acknowledgment

This thesis presents the conclusion of my master's degree at the Faculty of Information and communication, Chosun University. It has been an amazing couple of years, and I have a chance to learn much during this challenging period, yet jubilant time. My task has been highly exciting to deal with, however the work itself was sometimes hard and challenging, the numerous individuals who helped during this period are much helpful and supportive.

First, I would like to express my sincere gratitude to my advisor, for his encouragement, guidance and valuable knowledge shared throughout the development of this thesis. The exposure I have gained during my research would be a valuable treasure of my life.

I would also like to thank and appreciate all my professors, seniors, lab members, friends, and family, who have endured my partial absence, yet have been there for me.

Finally, I would like to thank OASIS and its collaborators for their great efforts, large amounts of work and willingness to share their data, without which this thesis and the original work described herein would not be possible.