Medical Image Segmentation with Deep Learning

Yatharth Patil, Manthena Harshith reddy, Madhira Jayendra Shankar, Jatothu Vamshi, Apurv Kushwaha,

***Abstract*—This project presents a novel approach to advancing the accuracy and efficiency of medical image segmentation. The conventional convolutional blocks are replaced with resilient residual blocks, inspired by ResNet architectures, to alleviate the vanishing gradient problem and facilitate the training of deeper networks. Additionally, an extra layer is strategically introduced at the bottleneck for long-range feature extraction, capturing contextual information over larger spatial regions. These architectural modifications empower the neural network to learn intricate representations, improving its ability to discern complex patterns in medical images. The proposed methodology is implemented and evaluated on the 'Anatomical Tracings of Lesions After Stroke' (ATLAS) dataset, showcasing superior segmentation performance compared to existing approaches. This project contributes to the evolving landscape of medical image analysis, providing a robust solution for accurate and context-aware segmentation.**

***Index Terms*—Medical Image Segmentation, Residual Blocks, Convolutional Neural Networks, Deep Learning, ResNet Architecture, Long-Range Feature Extraction, Image Analysis, Feature Learning**

I. Introduction

A. Background

Assessing the location and extent of lesions caused by chronic stroke is crucial for medical diagnosis, surgical planning, and prognosis. However, the intricate shapes and small sizes of these lesions in MRI images pose challenges for clinicians in visualizing their spread and severity. Manual lesion segmentation from MR images is time-consuming and subjective. Therefore, there is a pressing need for automatic methods in clinical practice to efficiently measure stroke lesions. Despite this demand, the task remains challenging due to variations in lesion shape, scale, size, and location, limiting the accuracy of automatic segmentation. Additionally, some lesions have fuzzy boundaries, making it difficult to distinguish between stroke and non-stroke regions. To address these challenges, advanced and precise image segmentation techniques are required for the successful implementation of a deep learning project focused on automating stroke lesion segmentation.

B. Challenges and Motivation

Medical image segmentation faces challenges rooted in the diversity of lesion characteristics, ranging from varied shapes to unclear boundaries, and the uneven distribution of data in training sets. The complexity of deep learning models, compounded by the scarcity of well-annotated datasets, adds to the intricacy of the task. Despite these challenges, the motivation to enhance medical image segmentation is driven by the profound impact it can have on clinical workflows and patient outcomes.

Improving segmentation directly addresses the need for more efficient and accurate diagnoses. Streamlining clinical workflows becomes possible, allowing healthcare professionals to swiftly analyze and interpret medical images, thereby expediting decision-making processes. The accurate segmentation of medical images isn't just about facilitating quicker diagnoses; it plays a pivotal role in optimizing surgical planning. Detailed insights into the extent and location of lesions enable more precise and tailored interventions, ultimately contributing to improved patient outcomes.

Moreover, accurate segmentation has prognostic value, aiding healthcare professionals in predicting disease progression and refining long-term treatment strategies. Beyond the immediate clinical benefits, advancements in medical image segmentation contribute to the broader landscape of technological innovation in healthcare. By pushing the boundaries of segmentation methodologies, the field is propelled towards the development of more sophisticated and effective medical imaging solutions. In essence, the challenges in medical image segmentation are met with a resounding motivation to revolutionize clinical practices and enhance the overall landscape of healthcare technology.

II. Related Works

Medical image segmentation has been a focal point in research, driven by the need for accurate and efficient analysis of complex anatomical structures and pathological regions in various imaging modalities. Over the years, researchers have explored diverse approaches to tackle the challenges inherent in this domain.

A. Prior-based

1) **Traditional Image Processing Techniques:** Early efforts involved handcrafted features and traditional image processing methods such as thresholding, region growing, and edge detection. While these approaches showed promise in simple cases, they struggled with the complexity of medical images, especially in the presence of noise, variability, and ambiguous boundaries.
2) **Graph-Cut and Level Set Methods:** Graph-cut algorithms and level set methods emerged as popular techniques for medical image segmentation. These methods utilized graph theory and mathematical models to capture image properties and evolve contours. While effective in certain scenarios, they faced limitations in handling variations in lesion characteristics and required extensive parameter tuning.
3) **Machine Learning Approaches:** With the rise of machine learning, researchers explored techniques like support vector machines (SVM) and random forests for

medical image segmentation. These methods demonstrated improved performance by learning discriminative features from the data. However, their success heavily relied on feature engineering and might struggle with capturing intricate patterns.

B. Deep Learning-based

The advent of deep learning, particularly convolutional neural networks (CNNs), marked a significant breakthrough in medical image segmentation. Models like U-Net, FCN, and DeepLab demonstrated remarkable capabilities in automatically learning hierarchical features, enabling them to handle the complexity and variability in medical images. Transfer learning, where pre-trained models are fine-tuned on medical data, further boosted performance, especially in scenarios with limited annotated datasets.

DUNet: The study introduces ”D-UNet,” a novel architecture for chronic stroke lesion segmentation in medical imaging. D-UNet combines the benefits of 2D and 3D convolutions, outperforming 2D networks and requiring less computation than 3D counterparts. To address data imbalance, an innovative loss function, Enhanced Mixing Loss (EML), is proposed. Evaluations on the ATLAS dataset show D-UNet’s superior performance in terms of Dice Similarity Coefficient (DSC) and precision compared to three top-tier methods. The results highlight D-UNet’s potential for effective chronic stroke lesion segmentation in clinical applications.

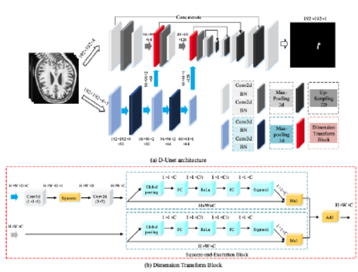

Fig. 1. D-Unet Architecture

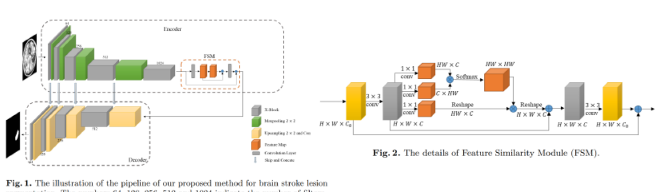

Fig. 2. X-Net Architecture

*1) **X-Net**:* **XNet:** X-Net i a neural network architecture designed to address challenges in brain stroke lesion segmentation. X-Net utilizes depthwise separable convolution and features a Feature Similarity Module (FSM) for capturing long-range dependencies. This design results in a reduced network size and enhanced efficacy in extracting dense contextual information, leading to improved segmentation. The evaluation on the ATLAS dataset demonstrates superior performance compared to six advanced approaches. The authors have provided their code and models for further use and development, emphasizing X-Net’s potential for advancing brain stroke lesion segmentation in medical imaging applications.

C. Our Contributions

We are looking to revamp our image segmentation approach with a couple of key enhancements. First off, we’re considering a **shift from the standard convolution block to the more robust residual block**. Inspired by architectures like ResNet, the inclusion of residual blocks introduces skip connections that help alleviate the vanishing gradient problem. This alteration has the potential to empower our network to learn more intricate features, ultimately improving its capability to discern complex patterns in medical images.

Additionally, we’re proposing the **integration of an extra layer at the bottleneck of the network**. This additional layer serves a specific purpose: long-range feature extraction. By strategically placing this layer, we aim to capture contextual information and dependencies over larger spatial regions. In the realm of medical image segmentation, where structures or abnormalities often span considerable distances within an image, this modification becomes particularly crucial. The goal here is to equip our network with a broader contextual understanding, enabling more accurate segmentation by considering relationships between pixels that are farther apart.

In essence, our proposed improvements involve embracing the power of residual blocks for enriched feature learning and introducing an extra layer at the bottleneck to harness long-range dependencies. Through these adjustments, we’re aiming to tackle the intricacies of medical image segmentation, allowing our network to learn more nuanced representations and contextual information for a potentially heightened segmentation performance.

III. Proposed Work

A. Network Architecture

Our Model is based on Unet++ architecture. There are 5 Blocks, with 2 Residual blocks in each. As seen in the Image, we have added a Ken Block, which is an inception block with a filter size of 1,3,5, and 7 with a skip connection.

B. Cost Functions

We used Adam as the optimizer and Sparse categorical-cross-entropy as the loss function.

IV. Experimental Details

A. Datasets

For the Dataset, we are using ATLAS (Anatomical Tracings of Lesions After Stroke) Release 2.0, an open-source dataset consisting of 304 T1-weighted MRIs. Initially, we preprocessed the Atlas dataset, chose different brain images, and then created a cropped image that contained the lesion. As each of the cropped images had a different size, we pulled them to the (124,124) size, as it was seen that almost for every image, all the lesions were occupied in the frame of the

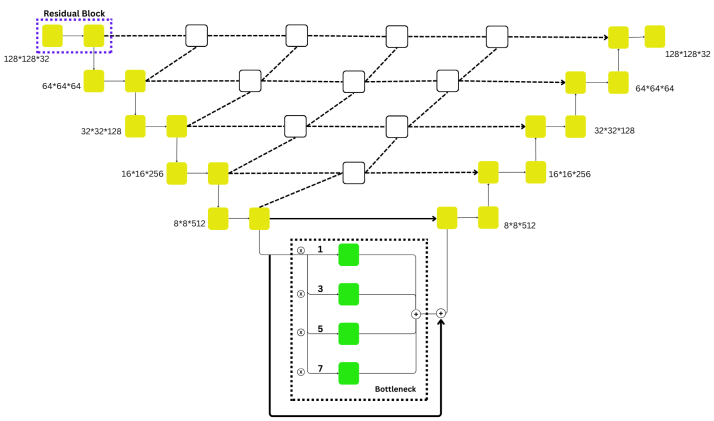

Fig. 3. Network Architecture

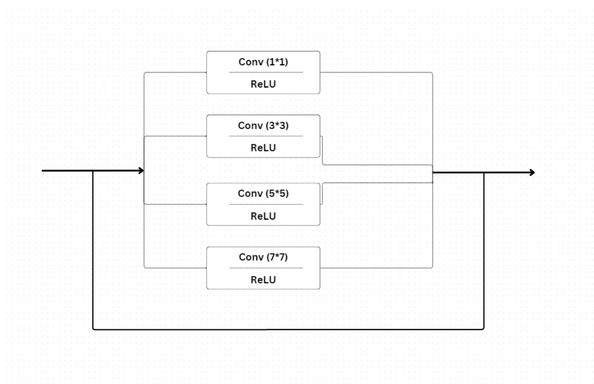

Fig. 4. Ken Block.

size (124,124). As (124,124) is not a power of 2, while going down in the Unet++, we would only be able to get a depth of 4. Therefore, we also padded a layer of 4 pixels, on the horizontal and vertical sides of the image, such that we get, a (128,128) size image.

B. Training Details

Initially, we ran the test on our Unet++ with residual blocks, hoping for a result; we tried with different layer or width combinations, like [8,16,32,64,128,256] and [2,4,8,16,32,64], to test whether having a deep network will be a good choice or not. But initially, for the case of [8,16,32,64,128,256], we could not converge it. We tried with RMSprop and Adam Optimizer. The loss function used was Sparse Categorical-Crossentropy. We also tried different combinations of learning rate and weight decay(in the case of the Adam optimizer).

In the case of a Very deep network, even if it was converging, results were not good. For the case of [2,4,8,16,32,64], it was not expected to get results, and we did not get it. But every time the test ran, a pattern was noticed that, even if the network is not deep, if the **last layer filter size is high** , we were getting good results. Hence, again, trials with layers of the **pattern [*,*,*,*,512]** were tried, and good results were achieved.

In the next phase of the trial, It was also noticed that with **further processing on the bottleneck**, we were able to get very good results. Hence, we added an inception layer, with a skip connection to the bottleneck. Filters of size 1,3,5,7 were chosen. Our Final model has the architecture of Unet++, with widths as [32,64,128,256,512], with the ken block at the bottleneck. **We used Adam as the optimizer and Sparse categorical-cross-entropy as the loss function**. Again, with hit and trial, we were able to converge the model for the **learning rate of 1e-6 and weight decay of 1e-4.**

C. Baseline Methods

As the baseline model, we are using Unet. It is widely used for image segmentation; It has an **Encoder** arm and **Decoder** arm. Going from dimension reduction to dimension increment, the network learns about the structure of the image, and thus, it helps in the image segmentation task.

D. Evaluation Metrics

As an evaluation metric, we are using the Dice coefficient, as it is one of the widely accepted metrics for checking the accuracy between the original mask and the generated mask. The dice coefficient is calculated as,

$$(A \cap B)/n(A) * n(B) \tag{1}$$

Here, the intersection term denotes, a similar pixel between the generated and the original mask, while the n(A) and n(B), denote the number of pixels in both the images.

V. Results

For training the model, batch size was kept at 1(It is the requirement of the model), and the model was trained for 40 epochs. For testing, we took 10-15 images and tried to generate a segmentation mask for it. An average dice coefficient of 0.7 was achieved. As the testing dataset, is smaller, we might be getting biased results.

A. Comparison with State-of-the-art Methods

For the accuracy measurement of our model, we decided to use the Dice coefficient. It measures the similarity between the original mask and the generated mask. It used the formula as (intersection between the two images) / (product of the number of the mask pixels in both images). It is used as it is widely accepted and is accurate.

Model	Dice-Coefficient
ResUNet	0.4702
DeepLab v3+	0.4609
2D Dense-UNet	0.4741
PSPNet	0.3571
SegNet	0.2767
U-net	0.4606
Proposed net	0.7

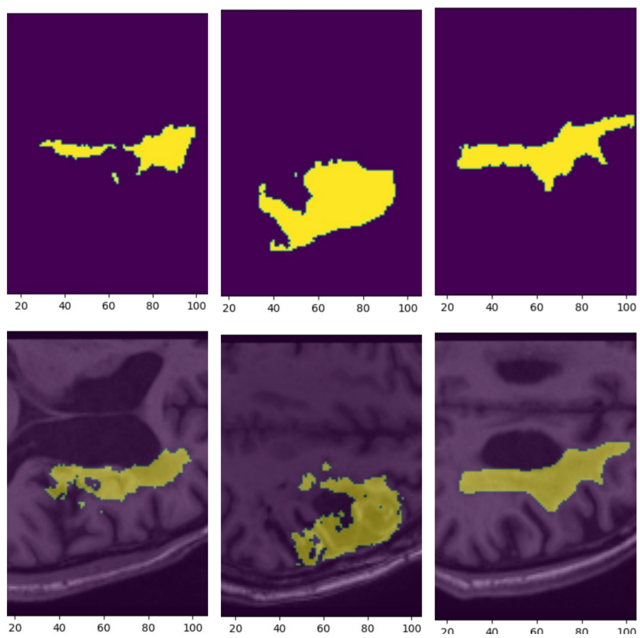

Fig. 5. Results

1) Qualitative Analysis: Using deep networks is not a good choice in this case, as they are very tricky to train as they do not converge easily. Hence, We kept the depth of the network as 5, with filters as [32,64,128,256,512]. Many different layer combinations were tried, but it was found that having a large filter size at the bottleneck helped improve the results, and another reason for going from 32 filters to 512 is that having a moderately deep model helps the model to learn about the noise, if any present in the image. And at the end of the network, we added a block with four, layers of filter size, 1,3,5, and 7, with a skip connection. Adding this block further increased the accuracy of the model.

VI. Ablation Study

We employ the Unet++, with dilated convolution, to capture long-range dependencies and obtain dense contextual information. To verify the effectiveness of this module, we tried it on many samples. An average of 0.7 dice coefficient was achieved. Using Unet++ helps to capture most of the structural dependencies. While computation at the bottleneck, allows us to get a feature map, that learns about lesions. Also, If trained for a higher number of epoch, much better results can be expected.

A. Effect of Various Network Modules

In our, module, we added skip dense skip connections, between the encoder block and the decoder block. This type of network is called Unet++. By doing this, many more spatial parts can be learned. Another thing that we changed is that we are using a dilated convolution block instead of a classical convolutional block. This helps us learn non-localized features. Another thing we added is a computational block at the bottleneck of the network. This block list is for extracting more complicated image features for the task.

B. Effect of Cost Functions

We are using Sparse-categorical Crossentropy as the loss function. This is one of the widely used loss functions for image segmentation. The main reason for choosing this loss function is that, in this case of segmentation, we are classifying each pixel, whether it is a part of the mask or not. Another benefit of using sparse-categorical cross-entropy is that there is no need to convert the image or mask to a one-hot encoded vector.

VII. Justification and Discussion

The model is trained n a robust dataset, with MRi images of different patients. To ensure that model is able to learn all the different distributions.

A. Failure Cases

The model is trained to work on a robust dataset. For that, we added noise to our original images so that, the model can be trained, to understand the distribution of the noise. But, for some, reason, in the case of noise in the input image, we are not able to get a good output mask. Also, the model is not trained on images with variable intensity. Hence, if the intensity of the input image is changed, we might get different results. It works nicely for any other type of input image.

VIII. Conclusion

We present an end-to-end model named Ken-model for brain stroke lesion segmentation. It can effectively extract informative features with fewer trainable parameters through the replacement of the traditional convolution with Dilated convolution. Furthermore, it can probe dense contextual information. The proposed method gracefully addresses the problems of the existing approaches, and the inefficiency in context capturing of long-range dependencies. Experiments on the ATLAS dataset demonstrate that our proposed Ken model could achieve better performance than existing models.