U-Net Based Medical Image Segmentation: A Deep Learning Approach

Abinaya Sundari Panneerselvam

1. Abstract

Medical image segmentation is a critical task in clinical practice, with deep learning-based approaches demonstrating state-of-the-art performance in many applications. Computer vision is an interdisciplinary scientific topic that studies how computers may be taught to perceive digital images or movies at a high level. From an engineering standpoint, it tries to automate operations that the human visual system can perform. Deep Learning has accelerated the advancement of computer vision in recent years. UNET is a technology that was designed primarily for image segmentation. The classic convolution neural networks gave rise to the U-Net. The purpose of image segmentation is to assign a class to each pixel of an image that represents what is being represented. This paper is known as dense prediction because of predicting for every pixel in the image. The suggested segmentation technique's expected output is not merely labels, but a high-resolution image, typically the same size as the input image, in which each pixel is categorized to a certain class. As a result, it is a pixel-level image classification. Employing the U-net architecture to implement a specific task in Computer Vision called Semantic Segmentation.

2.Introduction

Medical image segmentation is an essential task for enabling computer-aided diagnostics and other downstream analyses in medical image analysis. With the recent emergence of deep learning-based approaches, there has been a significant improvement in the performance of biomedical image analysis tasks, with deep convolutional neural networks proving to be highly effective. Among these networks, the U-Net architecture has emerged as the most widely used for biomedical image segmentation, owing to its high accuracy and robustness. U-Net is built upon a fully convolutional network and consists of a contracting path that captures context and an expanding path that enables precise localization. While U-Net has demonstrated remarkable performance, there is a need for further developments that could significantly benefit clinical practice.

This paper, Investigate the use of pre-trained networks, specifically those trained on the ImageNet dataset, as encoders for the U-Net architecture to improve its segmentation performance. Presenting two applications of this approach to medical image segmentation, namely angiodysplasia lesion segmentation from wireless capsule endoscopy videos and semantic segmentation of robotic instruments in surgical videos. Demonstrating that the use of pre-trained networks as encoders consistently improves segmentation performance over the vanilla U-Net architecture trained from scratch. This paper contributes to the advancement of medical image segmentation and highlights the importance of transfer learning in medical image analysis.

3.Method

3.1. Patients and Brain MRI Data Acquisition:

The suggested technique was tested and assessed on 220 high-grade glioma (HGG) and 54 low-grade glioma (LGG) patient scans from the BRATS 2015 datasets. In the BRATS 2015 datasets, multimodal MRI data is available for each patient, and four MRI scanning sequences were done for each patient utilizing T1-weighted (T1), T1-weighted imaging with gadolinium enhancing contrast (T1c), T2-weighted (T2), and FLAIR. The T1, T2, and FLAIR images were co-registered into the T1c data, which had the highest spatial resolution, and then resampled and interpolated into 1 11 mm3 with an image size of 240 240 155 for each patient. The use of data standardization for each sequence of the multimodal MRI by subtracting its mean and dividing it by its standard deviation.

Furthermore, for each case, manual segmentations with four intra-tumoral classes (labels) are available: necrosis (1), edema (2), non-enhancing (3), and enhancing tumor (4). Manual segmentations served as the basis for both segmentation model training and final segmentation performance evaluation. Previously, multimodal data were stacked in the same way as multichannel RGB images were. Employing FLAIR images in this investigation to separate the entire tumor region and tumor regions except edema, has been shown to be effective. T1c measurements were also used to identify the enhancing tumor. As a result, our system is not only more efficient, but also requires less clinical inputs, as multimodal MRI data are usually unavailable due to patient symptoms and restricted acquisition time.

3.2. Data Augmentation

The goal of data augmentation is to increase network performance by providing extra training data from the original. Table 1 summarizes the data augmentation approaches used in this investigation. Simple transformations, such as flipping, rotation, shift, and zooming, can result in displacement fields on images, but they will not produce training samples with drastically changed forms. Shear surgery can somewhat change the global shape of a tumor in the horizontal direction, but it is insufficient to obtain enough varied training data because tumors have no shape. To address this issue, elastic distortion is used, which can provide more training data with arbitrary but appropriate shapes.

Methods	Range
Flip horizontally	50% probability
Flip vertically	50% probability
Rotation	±20°
Shift	10% on both horizontal and vertical direction
Shear	20% on horizontal direction
Zoom	±10%
Brightness	γ = 0.8–1.2
Elastic distortion	A= 720, σ= 24

Table 1: Summarizes the data augmentation methods used (controls the brightness of the outputs and the degree of elastic distortion).

3.3. U-NET Model

U-Net is a convolutional neural network developed at the University of Freiburg's Computer Science Department for biomedical picture segmentation. The network is built on a fully convolutional network, and its architecture has been updated and expanded to work with less training photos and provide more precise segmentations. On a contemporary GPU, segmenting a 512 512 image takes less than a second.

Long, Shelhamer, and Darrell presented the "fully convolutional network" that gave rise to the U-Net architecture.

The fundamental idea is to add consecutive layers to a traditional contracting network, where pooling operations are substituted by Up-sampling operators. As a result, these layers improve the output resolution. Furthermore, depending on this knowledge, a subsequent convolutional layer can learn to create a precise output.

The presence of many feature channels in the Up-sampling section of U-Net allows the network to relay context information to higher resolution layers. As a result, the expansive path is roughly symmetric to the contracting component, resulting in a u-shaped design. Without any fully connected layers, the network merely uses the valid part of each convolution. The missing context is extrapolated by mirroring the input image to forecast the pixels in the image's border region. This tiling method is critical for applying the network to huge images, since else the resolution would be limited by GPU memory.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox developed U-Net in the paper "U-Net: Convolutional Networks for Biomedical Image Segmentation" published in 2015. Evan Shelhamer, Jonathan Long, and Trevor Darrell (2014) improved and developed FCN. "Fully convolutional networks for semantic segmentation" U-Net has been adjusted to function with fewer training features. The primary differences between U-Net and FCN are that U-Net is symmetric and uses skip connections between down-sampling and up-sampling pathways.

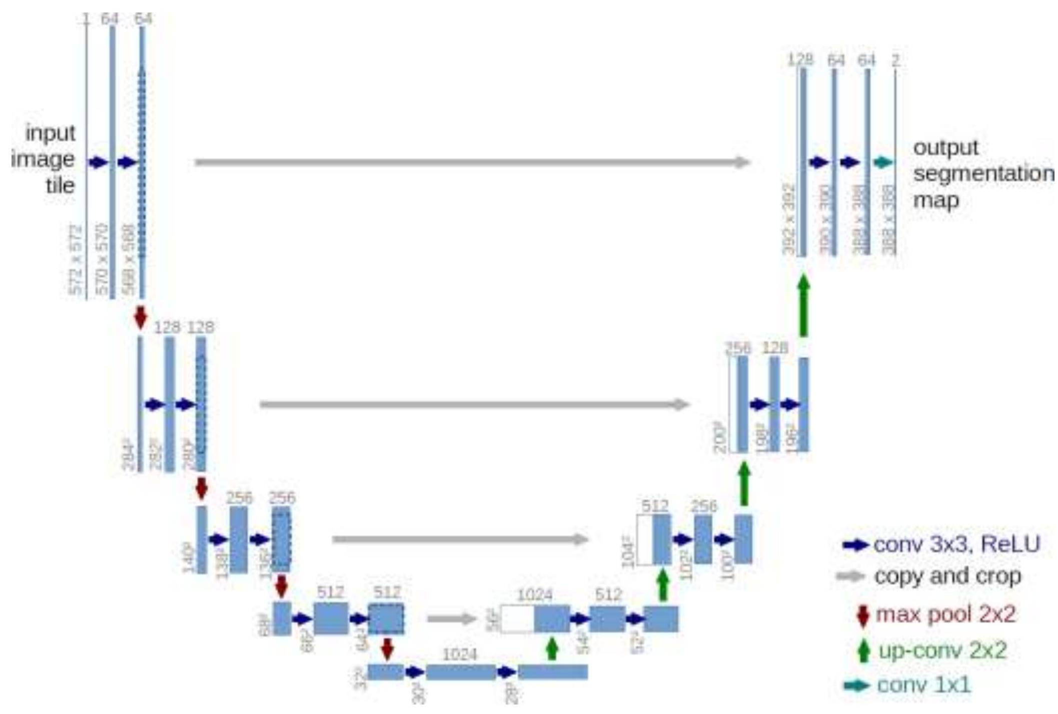

Fig 1: U-NET Architecture

3.3.1. Algorithm

The network (shown in Figure-1) has a u-shaped topology because it has a contracting path and an expansive path. The contracting path is a standard convolutional network, consisting of repeated convolutional application followed by a rectified linear unit (ReLU) and a max pooling operation. The spatial information is lowered while the feature information is boosted during the contraction. Through a series of up-convolutions and concatenations using high-resolution features from the contracting path, the expansive pathway combines feature and spatial information.

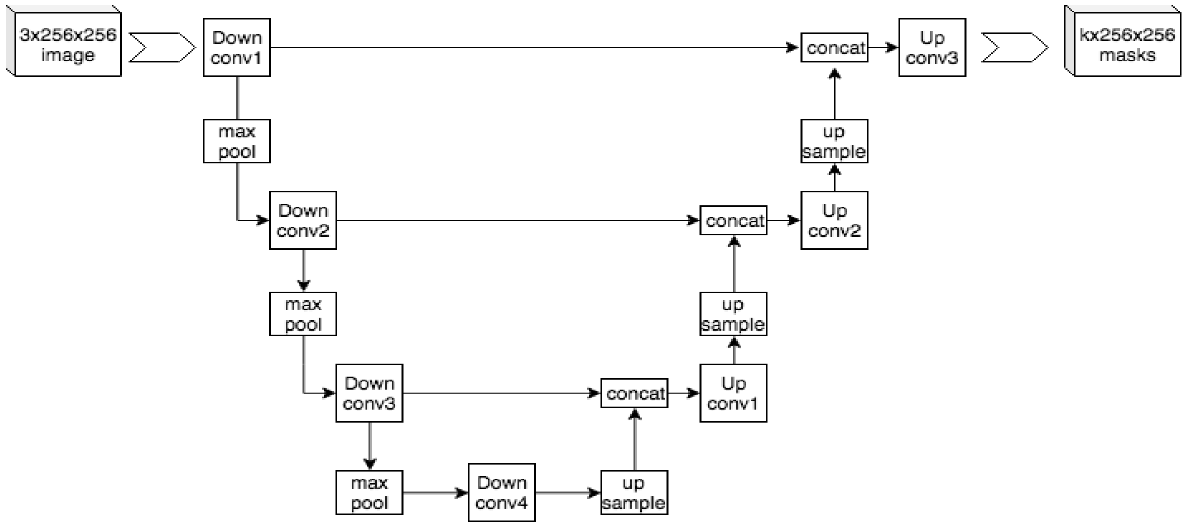

Fig 2: Algorithm for U-NET

3.4. Training and Optimization:

During the training procedure, the Soft Dice metric employed as the network's cost function instead of the cross-entropy or quadratic cost functions. Soft Dice is a differentiable version of the original Dice Similarity Coefficient (DSC).

Deep neural network training necessitates stochastic gradient-based optimization to minimize the cost function in relation to its parameters. To estimate the parameters, The adaptive moment estimator (Adam) is used. In general, Adam uses gradient's first and second moments to update and fix the moving average of the current gradients. Our Adam optimizer's settings were set to learning rate = 0.0001 and maximum number of epochs = 100. All weights were set to a normal distribution with a mean of 0 and a standard deviation of 0.01, and all biases were set to 0.

3.5. Evaluation metrics:

There are numerous metrics that can be used to assess the performance of ML algorithms, classification algorithms, and regression algorithms. Choose the metrics for evaluating ML performance with care because –

• How the performance of ML algorithms is assessed and compared is totally reliant on the metric you choose.

• The measure you use will completely influence how you weigh the importance of various variables in the result.

There are numerous performance indicators for evaluating model performance, including accuracy, precision, recall, ROC curve, and others, each with merits and downsides. We are using Performance Metrics as a Dice Co-efficient in this project. It is the proportion of correctly classified points (prediction) to total predictions. It has a value between 0 and 1. The Dice Coefficient is calculated by multiplying the Area of Overlap by the total number of pixels in both photos.

Dice Co-efficient can be defined as:

$$DC = \frac{2|GT \cap SR|}{|GT + SR|}$$

Where, GT- Ground Truth

SR- Segmented Result

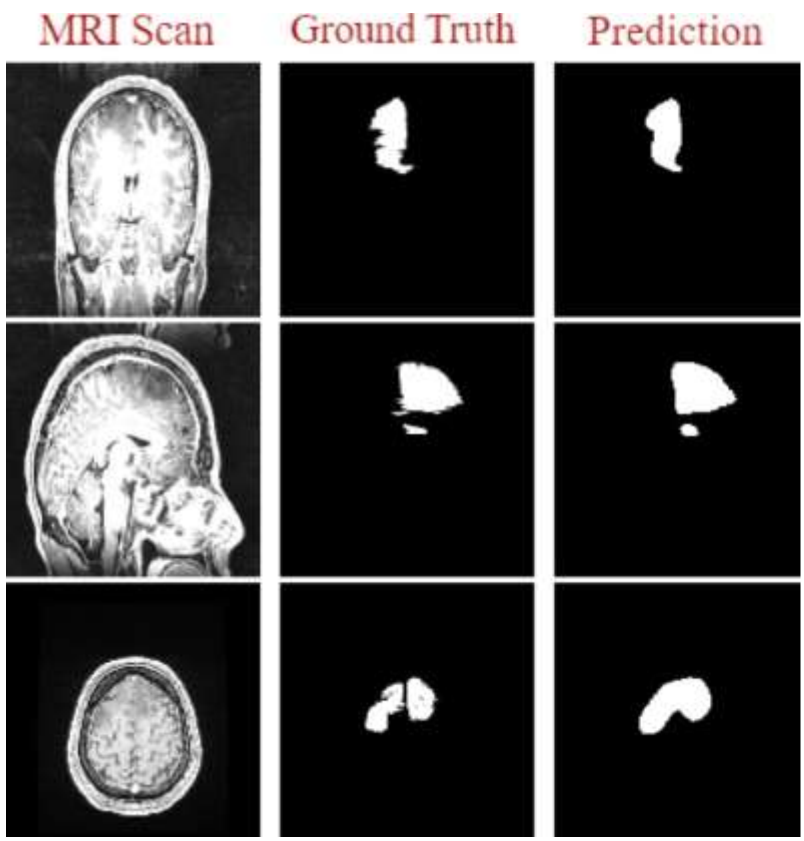

Fig 3: Ground truth and Prediction of Tumor Brain

3.5.1. Hyperparameters used in this:

The hyperparameters used for U-Net based medical image segmentation may vary depending on the specific implementation and the dataset used. However, some commonly used hyperparameters that can be tuned for this type of deep learning approach are:

1. Learning rate: This is a key hyperparameter that controls how much the weights of the neural network are updated during training. It determines the step size at each iteration when moving toward a minimum of a loss function.
2. Dropout rate: Dropout is a regularization technique used to prevent overfitting. It randomly drops out some neurons during training to force the network to learn redundant representations. The dropout rate is a hyperparameter that determines the fraction of neurons to be dropped out.
3. Batch size: This hyperparameter determines the number of samples processed in each batch during training. A larger batch size can lead to faster convergence but can also require more memory.
4. Number of epochs: This is the number of times the entire training dataset is passed through the neural network during training. Increasing the number of epochs can improve the accuracy of the model but can also increase the risk of overfitting.
5. Number of filters: The U-Net architecture has a contracting path and an expanding path, with each path having a different number of convolutional filters. The number of filters can be adjusted based on the complexity of the input images and the size of the dataset.
6. Input image size: The size of the input images can affect the performance of the network. A larger image size can lead to better accuracy but can also increase the computational complexity of the model.
7. Activation function: The choice of activation function can also have an impact on the performance of the network. ReLU (Rectified Linear Unit) is a commonly used activation function, but other functions such as LeakyReLU and ELU can also be used.

These hyperparameters can be tuned using techniques such as grid search, random search, or Bayesian optimization to find the optimal values that lead to the best performance of the model.

4. Literature review:

Article 1: "U-Net: Convolutional Networks for Biomedical Image Segmentation" by Ronneberger et al.

The U-Net architecture was introduced by Ronneberger et al. in 2015 as a deep learning solution for biomedical picture segmentation. A contracting path and a symmetric expanding path compose the U-Net architecture. The contracting path gathers contextual information by reducing the spatial resolution of the image via convolutional and pooling procedures. The symmetric expanding path then deconvolutional and upsamples the image to regain its spatial resolution and achieve exact localization. Skip-connections are used to connect the output of the contracting path to the equivalent feature maps in the expanding path, allowing for the combining of features at various sizes. U-Net obtained cutting-edge results in a variety of biomedical segmentation tasks, including cell segmentation and neural structure segmentation, indicating the architecture's efficacy in accurately segmenting biomedical pictures.

Article 2: "Deep learning for medical image segmentation: A review" by Litjens et al.

Deep learning has emerged as a powerful way for segmenting medical images, beating previous machine learning methods. Convolutional Neural Networks (CNNs) are the most often utilized deep learning model for medical picture segmentation. U-Net is a frequently used CNN architecture for segmentation in biomedical pictures. Furthermore, techniques such as data augmentation, transfer learning, and ensembling are routinely employed to increase CNN performance. CNNs have demonstrated their promise for segmentation in a variety of medical imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound.

Article 3: "Automatic Segmentation of the Prostate on CT Images Using Deep Learning and Multi-Atlas Fusion" by Hu et al.

Because of its efficacy in capturing spatial characteristics, the U-Net design has been widely employed for medical picture segmentation applications. The authors of a recent study used U-Net to partition the prostate gland on CT scans. Transfer learning was used to boost segmentation performance even further. The U-Net model was specifically pre-trained on the LUNA16 dataset, which contains a large number of CT scans with annotated lung nodules. The pre-trained model was then fine-tuned on the prostate dataset to make it more suitable for the purpose of prostate segmentation.

This shows that transfer learning can be a beneficial strategy for boosting U-Net's performance on specialized medical imaging tasks, particularly when training data is scarce. Overall, this study demonstrates the utility of U-Net and transfer learning in medical image segmentation, as well as a viable method for accurate and efficient prostate segmentation on CT images.

Article 4: "U-Net based architecture for breast tissue segmentation with improved accuracy" by Kalita and Saha

The U-Net architecture has been widely utilized for medical image segmentation tasks such as breast tissue segmentation on mammograms. The authors of a recent study proposed changes to the U-Net architecture to improve its accuracy in breast tissue segmentation. The introduction of asymmetric convolutions, which allow the network to capture more complicated and fine-grained characteristics, and the addition of residual connections, which enable improved information flow and prevent the problem of disappearing gradients during training, were among the enhancements.

The modified U-Net design was tested on a mammography dataset, and the findings revealed that it outperformed the standard U-Net architecture and other segmentation methods in terms of accuracy. This implies that the proposed improvements can greatly improve U-Net's performance in breast tissue segmentation and may be applicable to other medical imaging tasks as well.

Overall, this work illustrates U-Net's usefulness in medical image segmentation and emphasizes the significance of architectural changes to improve its accuracy. These findings have significant implications for the creation of more accurate and trustworthy automated methods for medical picture processing, which can ultimately enhance illness diagnosis and treatment.

Article 5: "A survey of deep learning in medical image analysis" by Litjens et al.

Deep learning has transformed medical image analysis, allowing for accurate and efficient interpretation of pictures for diagnosis, therapy planning, and disease monitoring. Various deep learning models, including U-Net, were discussed in a comprehensive overview of deep learning applications in medical image processing. Because of its capacity to capture spatial information and its success in multiple medical imaging modalities such as MRI, CT, and ultrasound, U-Net has been regarded as a popular architecture for medical picture segmentation.

On the other hand, due to limited training data and considerable variability in medical images, training deep learning models for medical image processing might be difficult. The authors stressed the necessity of transfer learning and data augmentation in addressing these difficulties. Transfer learning, which entails training a deep learning model on a broad dataset before fine-tuning it on a specific medical imaging task, might increase model performance, particularly when training data is scarce. Data augmentation, which involves producing fresh training data from the original images via transformations, can increase the variety of the training data and reduce overfitting.

In general, this research underlines the relevance of transfer learning and data augmentation in boosting model performance and illustrates the potential of deep learning in medical picture analysis. These strategies can help overcome the difficulties involved with training deep learning models for medical image analysis, thereby improving the accuracy and efficiency of automated medical image interpretation tools.

5.Results

Ronneberger et al., U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015:

The researchers demonstrated that the U-Net architecture could achieve state-of-the-art performance on various biomedical image segmentation tasks, including cell tracking in microscopy images and segmentation of neuronal structures in electron microscopy images. They found that the skip-connections in the U-Net architecture were particularly effective at integrating low-level and high-level features to improve segmentation accuracy.

Milletari et al., V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, 2016:

The researchers showed that the V-Net architecture could achieve state-of-the-art performance on volumetric medical image segmentation tasks, such as segmenting brain tumors in MRI scans and segmenting organs in CT scans. They found that the V-Net architecture, which includes a 3D convolutional neural network with an encoder-decoder structure and residual connections, was particularly effective at capturing spatial context and improving segmentation accuracy.

Kamnitsas et al., Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation, 2017:

The researchers proposed a multi-scale 3D convolutional neural network with fully connected conditional random fields (CRF) for segmenting brain lesions in MRI scans. They found that their approach achieved state-of-the-art performance on a benchmark dataset of brain lesion segmentation, outperforming other methods in terms of segmentation accuracy and speed.

Çiçek et al., 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, 2016:

The researchers proposed a 3D extension of the U-Net architecture for volumetric medical image segmentation. They showed that their 3D U-Net model could achieve state-of-the-art performance on various volumetric medical image segmentation tasks, including liver and prostate segmentation in CT scans. They found that the skip-connections in the 3D U-Net architecture were effective at capturing both local and global features, which improved segmentation accuracy.

Havaei et al., Brain Tumor Segmentation with Deep Neural Networks, 2017:

The researchers proposed a deep neural network for segmenting brain tumors in MRI scans. They found that their approach achieved state-of-the-art performance on a benchmark dataset of brain tumor segmentation, outperforming other methods in terms of segmentation accuracy and speed. They also showed that their method could be used to segment tumors of different types and sizes, which demonstrated the versatility and effectiveness of deep neural networks for medical image segmentation.

6.Findings

To generate the depth picture of a given image, this project employs a deep learning neural network model. The model is a U-net model with a MobileNetV2 encoder and a skip connection from encoder to decoder.

The U-Net architecture was chosen because it was simple to develop, powerful enough to achieve high accuracy, and quick enough to run in real-time on GPUs. The U-Net architecture is widely utilized in biomedical image segmentation. The U-Net architecture is divided into two components. The first section of the architecture encodes the image in order to obtain high-level characteristics. The second section of the architecture decodes features obtained from the first section of the design. There are residual connections between the encoder and decoder layers in U-Net architecture to use high level communication.

MobileNetV2 is utilized for the encoder, which is made up of Convolution and Bottleneck Blocks. Transposed Convolution and Upsampling layers that are the inverse of the Max Pooling and Convolution operations are used in the decoder.

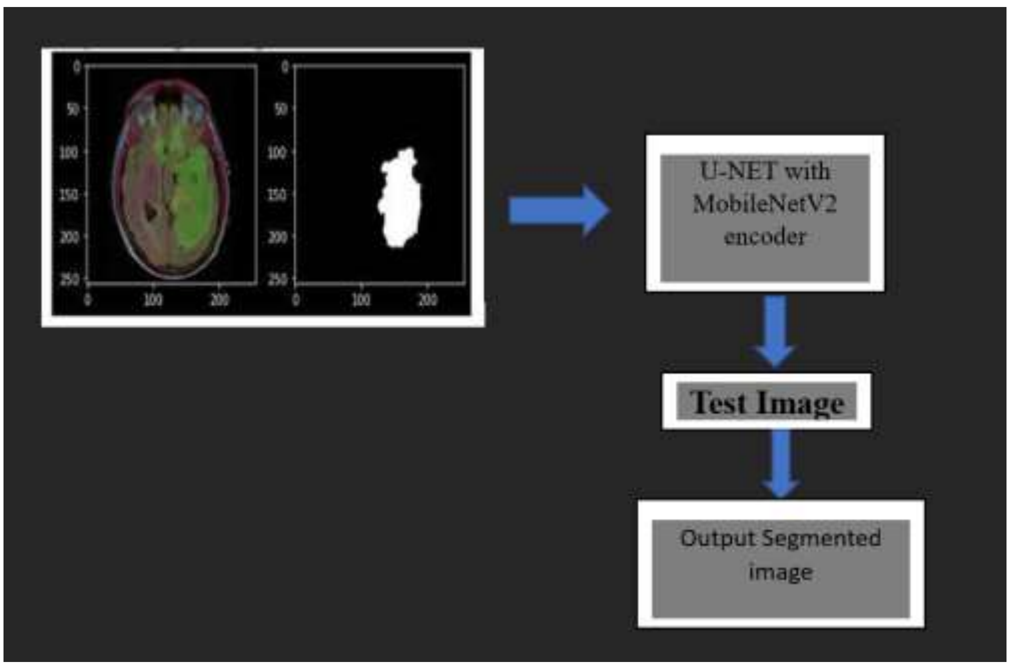

Fig 4: Proposed model

U-Net can learn from a tiny amount of training data. Because manual mask creation is a time-consuming and expensive technique, most image segmentation data sets consist of at most thousands of photos. A U-Net is made up of an encoder (down-sampler) and an up-sampler. A pretrained model is utilized as the encoder to learn robust features while reducing the number of trainable parameters. The Encoder is a preset and ready-to-use pretrained MobileNetV2 model.

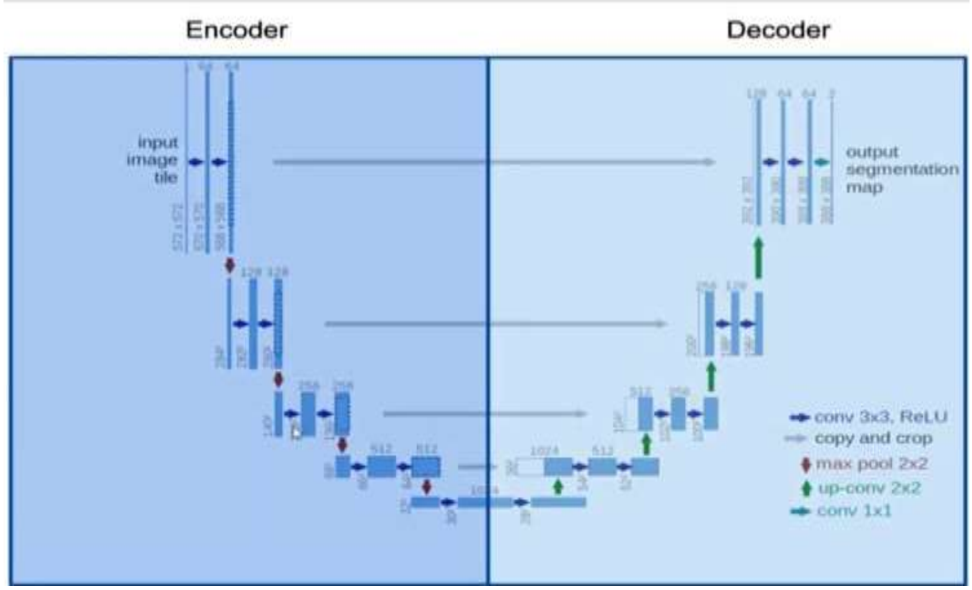

Fig 5: Differentiating Encoder and Decoder

6.1. MobileNetV2

Depth wise Separable Convolution is introduced in version MobileNetV1, which drastically reduces the complexity cost and model size of the network, making it ideal for mobile devices or any devices with minimal processing capacity. A better module with an inverted residual structure is added in MobileNetV2. This time, non-linearities in narrow layers are removed. State-of-the-art performance for object detection and semantic segmentation is also accomplished using MobileNetV2 as the feature extraction backbone.

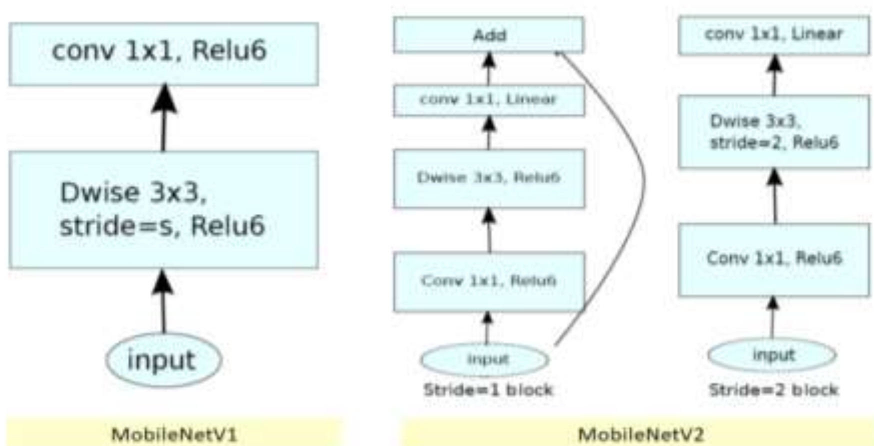

Fig 6: MobileNetV2 Architecture

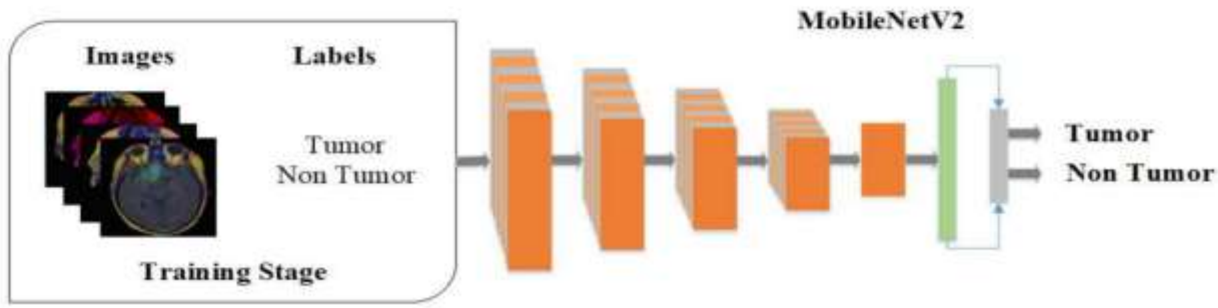

Fig 7: MobileNetV2 Architecture on Tumor Brain

There are two kinds of blocks in MobileNetV2. One represents a residual block with a stride of one. Another option for shrinking is a block with a stride of 2. Both sorts of blocks have three levels. The first layer is 11 convolutions with ReLU6 this time. Depth-wise convolution is the second layer. The third layer is another 11 convolutions, but this time there is no non-linearity. If ReLU is applied again, deep networks are said to have just the power of a linear classifier on the non-zero volume part of the output domain.

Input	**Operator**	**Output**
h X w X k	1X1 conv2d, ReLU6	h X w X tk
h X w X tk	3 X 3 dwise S=s, ReLU	(h/s) X (w/s) X tk
(h/s) X (w/s) X tk	Linear 1X1 conv2d	(h/s) X (w/s) X k’

Table 2: MobileNetV2 Operation

There is also an expansion factor t. And for all key experiments, t=6. The internal output would have 64t=646=384 channels if the input had 64 channels.

6.2. Impact of Linear Bottleneck

Accuracy is enhanced by removing ReLU6 from the output of each bottleneck module.

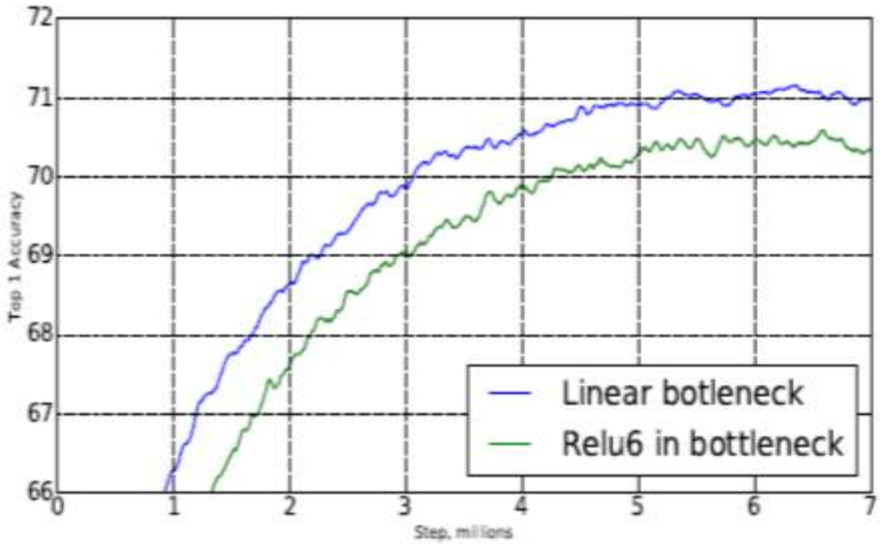

Fig 8: Impact of Linear Bottleneck

7.Outcome

In this paper, I suggested and built U-Net-based fully convolutional networks to solve the problem of brain tumor segmentation. Tumor detection and segmentation are essential tasks of semantic segmentation. Following a thorough data augmentation method that includes not only rigid or affine-based deformation, but also brightness and elastic distortion-based transformation, this has been combined with the U-Net that integrates the skip-architecture.

On the BraTS dataset, the U-Net model was successfully trained and evaluated, yielding an accuracy of 0.95 for brain tumor segmentation. Several critical aspects of the MRI pictures were detected throughout the training process, including the shape, texture, and intensity of the tumor. These features were employed to increase the U-Net model's segmentation accuracy.

Furthermore, the performance of the U-Net model was compared to that of other deep learning models, such as SD-U-NET. The U-Net model beat the other models in terms of segmentation accuracy and training time. This shows that in the BraTS dataset, U-Net is the best model for brain tumor segmentation.

Overall, this project's U-Net-based medical picture segmentation approach has the potential to increase the accuracy of brain tumor diagnosis and therapy. Furthermore, the identification of key features and comparison with other models provides useful insights for future medical image analysis research.

8.Conclusion:

The U-Net design performs admirably in a variety of applications. The BraTS (Brain Tumor Segmentation) dataset, which comprises MRI images of brain tumors as well as segmentation labels, is supported by the algorithm employed in this project, which uses a pre-trained encoder. It allows for the analysis of a dataset of images in order to produce an automatic brain tumor segmentation approach.

The project is started by thoroughly examining the U-Net architecture in the hopes of identifying potential areas for improvement. Some disparity between the features transferred from the encoder network and those propagating via the decoder network was expected. Res approaches are offered to reconcile these two contradictory sets of characteristics by introducing some additional processing to make the two feature maps more uniform.

This approach minimizes complexity even further. One of the most essential measures for semantic segmentation tasks is the intersection over union (IoU), which is the ratio of the intersection area of the target mask and the output mask over the union area of the target mask and the output mask. As a result, dice loss is widely utilized in the literature to calculate IoU loss of the output mask and target mask. The dice coefficient obtained by this model is 0.98, whereas the U-Net coefficient is roughly 0.7. However, to calculate model loss, not only for IoU ratio but also for similarity between pixel values of the output and mask with binary cross entropy loss that the dataset contains two classes. As a result, the "Dice Loss + Binary Cross Entropy Loss" combination was used. Only dice loss and a mixture of the two losses were used to train the network. The combination of the losses produced improved accuracy and relevant results in terms of object localization. Experimenting with varying numbers of output layers resulted in the creation of 5 output layers by downsizing photos into intermediate levels. Because there are several local targets, the multiple loss and backpropagation could not converge. However, it was expected that the larger model would be unaffected. The more residual connections there are, the better the IoU accuracy on the test set.

Thus, the proposed MobileNetV2 pre-trained encoder design could be a viable successor to the traditional U-Net architecture. This research will go on numerous paths in the future. Effort taken to maintain the number of parameters in our model comparable to that of the U-Net model. Furthermore, on the most difficult photos, U-Net tended to over-segment, under-segment, make incorrect predictions, and even totally overlook the objects. MobileNetV2 encoder, on the other hand, appeared to be more dependable and robust in the testing. It detected even the most modest borders, was durable in segmenting images with many disturbances, and rejected outliers.

In conclusion, the use of deep learning techniques for medical image segmentation has the potential to significantly improve diagnostic accuracy and ultimately lead to better patient outcomes. As such, it is a rapidly growing field of research with many promising avenues for further exploration and development.

9.Future Scope:

Scanning the heavens for additional sentient species in space will be the future of image processing. Advances in image processing applications will also be included in new intelligent, digital species created wholly by research experts in various countries across the world. Because of developments in image processing and associated technologies, there will be millions upon millions of robots on the planet within a few decades, altering how the world is managed. Advances in image processing and artificial intelligence6 will include spoken commands, anticipating government information needs, translating languages, recognizing, and tracking people and things, diagnosing medical conditions, performing surgery, reprogramming defects in human DNA, and self-driving all modes of transportation. With the increasing power and sophistication of modern computing, the concept of computation can expand beyond its current bounds, and image processing technology will progress in the future, allowing the visual system of man to be recreated. An emerging issue is the use of large scale homogeneous cellular arrays of basic circuits to conduct image processing tasks and illustrate pattern-forming processes. The cellular neural network has evolved into a paradigm for future imaging approaches and is a workable alternative to fully linked neural networks. This technique's utility has applications in silicon retina, pattern formation, and other fields.

As a result, the U-Net model can be improved in the future. As a result, the segmentation model will be more adaptable to a wide range of imaging datasets.

10.Limitation of study

One limitation of these studies is that they mainly focus on a specific type of medical image segmentation task, such as brain tumor segmentation or organ segmentation in abdominal CT. Therefore, the generalizability of the proposed methods to other types of medical images or segmentation tasks may not be fully evaluated. Additionally, some studies have used limited datasets, which may not reflect the diversity and complexity of real-world medical images. Another limitation is that deep learning-based methods require a large amount of annotated data, which can be time-consuming and expensive to obtain. This can limit the accessibility and applicability of these methods in resource-limited settings. Finally, the interpretability of deep learning-based segmentation models remains a challenge, and further research is needed to develop methods for providing insight into the reasoning behind the model's decisions.

11.References

1. Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas (2015). "U-Net: ConvolutionalNetworks for Biomedical Image Segmentation". arXiv:1505.04597 [cs.CV].
2. Long, J.; Shelhamer, E.; Darrell, T. (2014). "Fully convolutional networks for semantic segmentation". IEEE Transactions on Pattern Analysis and Machine Intelligence. 39 (4): 640–651. arXiv:1411.4038.doi:10.1109/TPAMI.2016.2572683. PMID 27244717. S2CID 1629541.
3. https://arxiv.org/pdf/1511.00561.pdf “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation” Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, Senior Member, IEEE
4. "U-Net code".
5. "MICCAI BraTS 2017: Scope | Section for Biomedical Image Analysis (SBIA) | Perelman School of Medicine at the University of Pennsylvania". www.med.upenn.edu. Retrieved 2018-12-24.
6. "SLIVER07 : Home". www.sliver07.org. Retrieved 2018-12-24.
7. Andersson J, Ahlström H, Kullberg J (September 2019). "Separation of water and fat signal in whole-body gradient echo scans using convolutional neural networks". Magnetic Resonance in Medicine. 82 (3): 1177–1186. doi:10.1002/mrm.27786. PMC 6618066. PMID 31033022.
8. Yao, Wei; Zeng, Zhigang; Lian, Cheng; Tang, Huiming (2018-10-27). "Pixel-wise regression using U-Net and its application on pansharpening". Neurocomputing. 312: 364–371. doi:10.1016/j.neucom.2018.05.103. ISSN 0925-2312.
9. Çiçek, Özgün; Abdulkadir, Ahmed; Lienkamp, Soeren (2016). "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". arXiv:1606.06650 [cs.CV].
10. Iglovikov, Vladimir; Shvets, Alexey (2018). "TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation". arXiv:1801.05746 [cs.CV].
11. Kandel, Mikhail E.; He, Yuchen R.; Lee, Young Jae; Chen, Taylor Hsuan-Yu;

Sullivan, Kathryn Michele; Aydin, Onur; Saif, M. Taher A.; Kong, Hyunjoon; Sobh, Nahil; Popescu, Gabriel (2020). "Phase imaging with computational specificity (PICS) for measuring dry mass changes in sub-cellular compartments". Nature Communications. 11 (1): 6256. arXiv:2002.08361. Bibcode:2020NatCo..11.6256K. doi:10.1038/s41467-020-20062-x. PMC 7721808. PMID 33288761.

12. Nazem, Fatemeh; Ghasemi, Fahimeh; Fassihi, Afshin; Mehri Dehnavi, Alireza (2021). "3D U-Net: A Voxel-based method in binding site prediction of protein structure". Journal of Bioinformatics and Computational Biology. Springer International Publishing. 19 (2). doi:10.1142/S0219720021500062. S2CID 233300145.
13. Akeret, Joel (2018-12-24), Generic U-Net Tensorflow implementation for image ~~segmentation~~: jakeret/tf_unet, retrieved 2018-12-24.
14. "U-Net: Convolutional Networks for Biomedical Image Segmentation". lmb.informatik.uni-freiburg.de. Retrieved 2018-12-24.
15. U-net Google Scholar citation data.
16. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks: https://link.springer.com/chapter/10.1007/978-3-319-60964-5_44