Critique of group 17 presentation - CV-Segmentation-MRI Tumor
Critiques by group 50.

Q: The project used two approaches to classify MRI images of Tumors. An autoencoder network was used as well as a convolution network. I didn't fully understand the workflow of the convolution network. There was some mention about feature extraction of the convolution network but this was unclear. Overall interesting subject and the expected amount of work was done.
A: Thank your suggestion. About the About the convolution network, you can find more information in our report in "VGG8-FCN16" and "FCN8-ResNet34". As for feature extraction, we first downsample datasets and then use data augmentation to increase the diversity of data, that is what we

Q: The methodologies could be compared and explained better. I didn't get a sense of which one performed better other than mentioning that one overfit. More analysis should be given to this topic. Also overfitting can be corrected, but I didn't see this presented. How could you fix that?
A: We have more comprehensive result and comparison in our report. As for the overfitting, our purpose is to highlight Unet, and one of the advantages of unet is to correct the problem of overfitting.

Q: The IOU metric should be explained better as you use it for both of your neural networks. I didn't get a clear sense of what your IOU graphs indicated. The IOU plot also has a floating point value which has more than 5 decimal places that should be formatted in exponential representation.
A: Thank you for your reminding, we will fix that and we provide more details in our report about IoU.

Q: Your performance looks good for segmenting the tumor area, but when does your neural network fail. Are there specific tumors that it is worse at predicting? Maybe provide a confusion matrix. Were there false positives or true negatives?
A: Sorry about the ignoring this part, and yes, there are false positives and true negatives in tumor detection. But the dataset is rare in such special condition.

Q: The code was well formatted and easy to read through and was presented well. The images of the MRI with and without masks were also a nice visualization of the problem at hand.
A: Thank you.

Critiques by Group 9
Q: Very good explanation of model structures.
Q: Clearly went into depth about the different types of network structures used and compared and contrasted the types generally
Q: Could spend more time and provide more info on the results of the networks' performances.
A: We provide more in the report/
● More importantly, while the models are explained very well and understood, why did some of them not work as well? Insight into the specifics of why certain models might've failed and
Q: why certain models performed better. Even if we can't understand the model, trying to understand why is good. For example, VGG-FCN8 has very poor validation loss. What could

this have been caused by? Is there indication of the model overfitting? How could that be mitigated? Etc. etc.
A: this is because of the limitation of the CNN model, we can easily find it will lose lots of information during the upsample procedure, while Unet can use a lot of skip connection to remain as much as the information of original input
Q: Also what is the advantage and usefulness of jump connections in networks?
A: Combining layers that have different precision helps retrieving fine-grained spatial information, as well as coarse contextual information. More can be found in report.
Q: It would be interesting to see more computer vision or other methods used for more data preprocessing. For example, extracting handcrafted features from the data and seeing if it improves performance or utilizing biological aspects of the tumors that are visible in the data to create new features.
A: Great idea.
Q: More time could be used to also explain a lot of design decisions. What are the reasons for downsampling the data? Why use dice loss and IOU metrics? Are these from prior works in the field or due to specific data related reasons?
A: downsampling eaquals feature extraction, dice loss and IoU are widely used in segmentation task, yes
Q: This might be a limitation of the data but it may also be interesting to see how the model performs on brains without tumors to possibly provide insight into the network's failures and shortcomings.

Critiques by Group 31
Q: From the result of FCN8 model, the models were still overfitted. Please try other methods solving overfitting?
A: The reason we were keeping the result of FCN8 is that we use Unet as our main model, overfitting is acceptable and can be compared to the Unit.

Q: Please explain why use VGG and Resnet for feature extraction?
A: We use VGG and resent for feature extraction because as deep CNNs, they outperforms baselines on many tasks and datasets, also they are the most used image-recognition architectures. We use Unet as our main model because

Q: Please explain why choose Unet model and why it is the main model?
A: As a competition winner model, UNet performed upsampling 4 times and used skip connection in the same stage instead of directly supervising and loss-reverting on high-level semantic features, which ensured the final result capture more low-level features, which also allows the fusion of features of different scales, so that multi-scale prediction and deep supervision can be performed. We think Unet is a great model for medical image segmentation task.

TUMOR DETECTION IN MR IMAGES

Xibo Zhang, Zihan Wang, and Chenhao Zhou

*University of California San Diego, La Jolla, CA 92093-0238

ABSTRACT

MRI (magnetic resonance images) has become a main detection tool in the modern medical system. However, It is difficult to robustly distinguish tumors from surrounding tissue. Thereby, we proposed a deep learning tool to help tumor detection in clinic practice. We tested three models including Unet Model, VGG-FCN8 and ResNet-FCN8 and conducted comparative experiments. The experimental results are based on IoU evaluation criteria, and it turns out that the validation of Unet has higher accuracy.

***Index Terms*—** Tumor MRI, Unet, FCN, VGG-FCN8, ResNet-FCN8, IoU evaluation criteria

1. INTRODUCTION

In 1971, Raymond Damadian proposed a tumor detection method by nuclear magnetic resonance [1]. MRI (magnetic resonance images) has become a main detection tool in the modern medical system. In 2018, there are more than 1.76 million new cases of cancer in the United States. The total number of cancers is expected to be 29.5 million by 2040. At the same time, MRI has also played a critical role in tumor detection. However, the development of medical devices has reached a bottleneck, and it is hard to break through the current tumor detection technology. In addition to improving medical staff's ability to analyze MRI, machine learning has become the best auxiliary method.

Unet was published in 2015[2] and belongs to a variant of FCN. The original intention of Unet was to solve the problem of biomedical images. Since the effect was really good, it was later widely used in various directions of semantic segmentation, such as satellite image segmentation. Industrial defect detection. Unet is based on the Encoder-Decoder structure and realizes feature fusion through stitching. The structure is simple and stable. In order to better study the performance of Unet, we used the classic FCN model and selected two network structures VGG and ResNet.

FCN(Fully Convolutional Networks) was first published in CVPR 2015 paper "Fully Convolutional Networks for Semantic Segmentation". This is the first work to train FCNs end-to-end for pixelwise prediction and from supervised pretraining[3]. The difference with the CNN network is that the CNN usually ends up being a fully connected layer But FCN replaces the fully connected layer with convolutional layers, so that the every network output is a heatmap rather than feature map. Each layer of the FCN is a convolution filter and the output image of the FCN is restored to the original size by Deconvolution. The purpose we construct fcn model for our dataset is for comparison between classic fcn model and our Unet model. We also completed the construction of FCN and selected some pictures for testing. The purpose we construct fcn model for our dataset is for comparison between classic fcn model and our Unet model. This comparison exists in many ways with Unet, such as cost of storage space, accuracy of results, and robustness of the model.

2. RELATED WORK

In recent years, there have been many articles about tumor detection. Most are focused on MRI technology and corresponding improvement programs. We have classified the articles into three categories: image processing methods, network model reconstruction, and non-MRI detection methods.

Articles based on image processing often focus on filtering the image. For examples, in "Automatic Analysis of Brain Tumor from Magnetic Resonance Images based on Geometric Median Shift", they propose an automated approach based on the geometric median shift algorithm over Riemannian manifolds, for the brain tumor detection and segmentation in magnetic resonance images[4]. In "Brain Tumor Localization and Segmentation Based on Pixel-Based Thresholding with Morphological Operation", they proposed a technique which is mainly based on the preprocessing step for de-noising input MRI, thresholding, and morphological operation and calculating performance parameters for validation[5]. Our project just use the mask of MRI to be trained, and we did not take this into consideration much. As for network construction, there are couples of papers are similar to ours. Like "UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation", this study classifies the ROI and non-ROI using fully convolutional network with new architecture, namely UNet-VGG16. Their model or architecture is a hybrid of U-Net and VGG16 with transfer Learning to simplify the U-Net architecture[6]. And this paper "A Customized VGG19 Network with Concatenation of Deep and Handcrafted Features for Brain Tumor Detection" confirms that

the VGG19 with SVM-RBF helped to attain better classification accuracy with Flair (>99%), T2 (>98%), T1C (>97%) and clinical images (>98%)[7]. There are also articles using new technologies to position the tumor, from biological perspective, like "Fibroblast Activation Protein (FAPI) PET for diagnostics and advanced target volume delineation in head and neck cancer". This analysis aims to introduce an approach of tumor detection and contouring for radiotherapy using visualization of cancer associated fibroblast: PET-CT with 68Ga-radiolabeled inhibitors of Fibroblast Activation Protein (FAPI)[8].

In general, the combination of machine learning and image processing methods is the topics of most articles in recent years, because it has unlimited possibilities and can bring unexpected results.

3. DATASET AND FEATURES

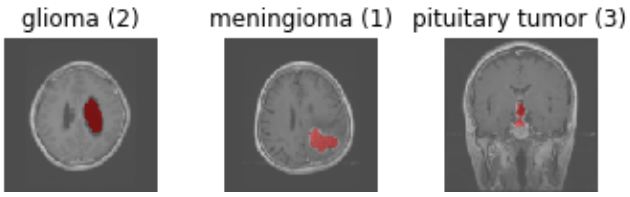

Fig. 1. Different types of brain tumor

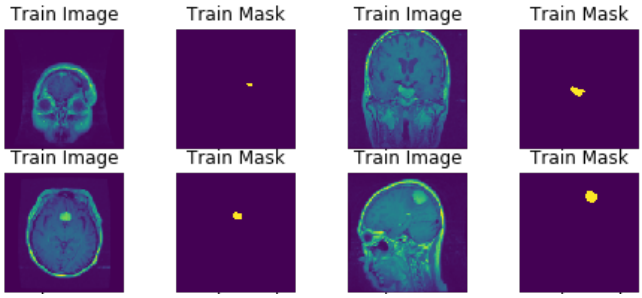

Fig. 2. Brain tumor MR Images and their corresponding fully annotated tumor segmentation masks. Tumors are marked yellow in the train masks.

In this project, we used a tumor MR images dataset, in which it contains 3064 MRI images and each images comes with a corresponding fully annotated tumor segmentation map(see Fig. 2). There are 3 types of brain tumor in the dataset. Fig. 1(1) shows meningioma, a tumor that forms on membranes that cover the brain and spinal cord just inside the skull. Fig. 1(2) shows glioma, a type of tumor that starts in the glial cells of the brain or the spine. Fig. 1(3) shows pituitart, a small gland found inside the skull just below the brain and above the nasal passages.

The images and masks are 512 * 512 pixels originally. After down sampling, the image size used for training is 128*128. The train set and test set are initially splited as 2451 and 613. We used data augmentation in pre-process to use the dataset more efficiently. Data augmentation is a technique that can significantly increase the diversity of data. In our case, in order to have better interpretation of possible shift and rotation invariance in MR images, we randomly displace vectors in the images on a coarse 3 *3 grid to simulate smooth deformations. The displacements obey Gaussian distribution. After augmentation, There are 4902 samples in train set, while the size of test set remains the same.

4. METHODS

4.1. UNet

4.1.1. Network Architecture

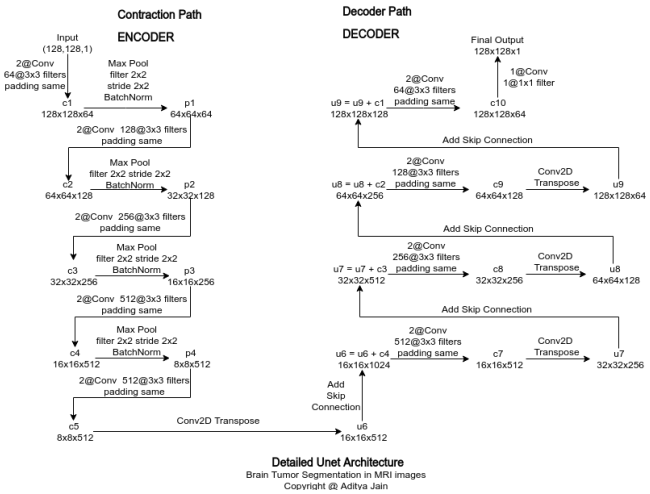

Fig. 3. U-net architecture (examples of 128*128 pixels images).

The network architecture is illustrated in Fig. 3[9]. An encoder (for downsampling) - decoder (for upsampling) structure and a jump connection are highlights of this model[9].

The model consists of a 23-layer convolutional network. In the encoder part, There is a repeated application of two convolutions, each of which are followed by a ReLU layer and a max-pooling operation. One combination as mentioned above is a downsampling step, and can double the number of feature channels. Then in the decoder part, the resolution is increased sequentially through the upsampling operation, which consists of two 3*3 convolutions followed by a ReLU. At the final layer a 1x1 convolution is used to output desired number of classes.

In UNet, the downsampling operations are done 4 times. Symmetrically, its decoder also upsamples 4 times accord-

ingly, restoring the advanced semantic feature map obtained by the encoder to the resolution of the original picture. The network also connects the upsampling result with the output of submodule in the encoder if they have the same resolution, and then use it as the input of the next submodule in the decoder.

4.1.2. Loss Function

The loss function we used in Unet is cross entropy loss function, which is computed as:

$$E = \sum_{x \in \Omega} w(x) log(p_{l(x)}(x))$$

where $p_{l(x)}$ a pixel-wise soft-max, and is computed as

$$pk(x) = exp(a_k(x))/(\sum_{k_t=1}^{K} exp(a_{k^0}(x)))$$

where $a_k(x)$ denotes the activation in feature channel k at the pixel position, and $w(x)$ is the weight map.

4.2. FCN

4.2.1. FCN Network Architecture

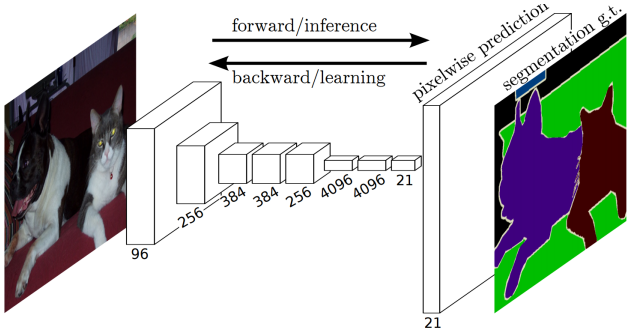

Fig. 4. FCN architecture

Fully Convolutional Networks (FCNs) owe their name to their architecture, (the network architecture is illustrated in Fig. 4.[10]) which is built only from locally connected layers, such as convolution, pooling and upsampling. Note that no dense layer is used in this kind of architecture. This reduces the number of parameters and computation time. Also, the network can work regardless of the original image size, without requiring any fixed number of units at any stage, givent that all connections are local. To obtain a segmentation map (output), segmentation networks usually have 2 parts :

Downsampling path : capture semantic/contextual information

Upsampling path : recover spatial information

The downsampling path is used to extract and interpret the context (what), while the upsampling path is used to enable precise localization (where). Furthermore, to fully recover the fine-grained spatial information lost in the pooling or downsampling layers, we often use skip connections.

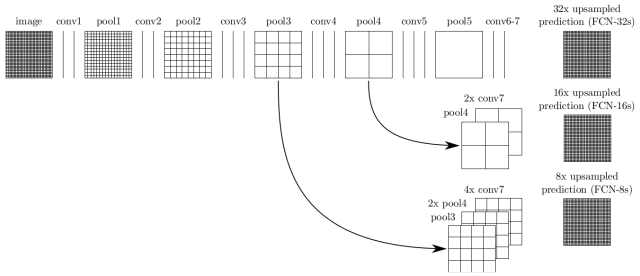

Fig. 5. variants of the FCN architecture

A skip connection is a connection that bypasses at least one layer. Merging features from various resolution levels helps combining context information with spatial information. So we can use different skip connection layers and strides for the last convolution, yielding different segmentation results(see Fig. 5[10]). Since combining layers that have different precision helps retrieving fine-grained spatial information, as well as coarse contextual information. Generally, FCN8 produces more precise segmentation maps. Therefore in the following experiments, we use FCN8 as our basic structure.

4.2.2. FCN8-VGG16 Network Architecture

The quality of the features we get from the network will significantly influence the prediction result. So here we want to use two different networks to extract features and see if they will influence the segmentation result.

Firstly we will use the VGG net, it is widely used in image classification. As shown in Fig. 6[11] VGG Net is a plain and straight forward CNN architecture. The idea of VGG architectures is to stack the convolutional layers with increasing filter sizes. use multiple 3x3 kernels. In the experiment we use VGG16.Add as discussed previously, we use FCN8 as main structure.

4.2.3. FCN8-ResNet34 Network Architecture

We also use ResNet to get the feature map. he basic idea of ResNet is to use the residual learning framework(in Fig. 7[12]) to avoid gradient vanishing in the deep network. These residual networks are easier to optimize, and can gain accuracy from considerably increased depth.

It performs better than vgg in most of the recognition tasks. So we want to know if better features can help us get better segmentation result.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig. 6. variants of the VGG architecture

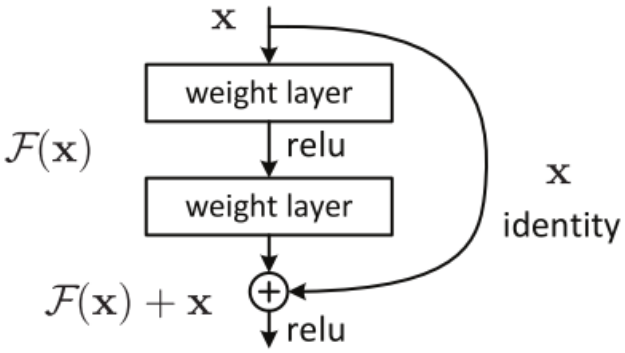

Fig. 7. residual framework

5. EXPERIMENTS/RESULTS/DISCUSSION

5.1. Evaluation Criteria

To evaluate the performance of the networks, the trained networks ran a forward pass on the test data and the IoU was computed. The Intersection-Over-Union (IoU), also known as the Jaccard Index, is one of the most commonly used metrics in semantic segmentation. the IoU is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth, as shown in Fig. 8[13]. This metric ranges from 0–1 with 0 signifying no overlap (garbage) and 1 signifying perfectly overlapping segmentation (fat dub).

In this project, we set a threshold to classify every pixel to the specific label(0 or 1 in our dataset). And we will plot the relation between the threshold and the IoU evaluation to find the best threshold value.

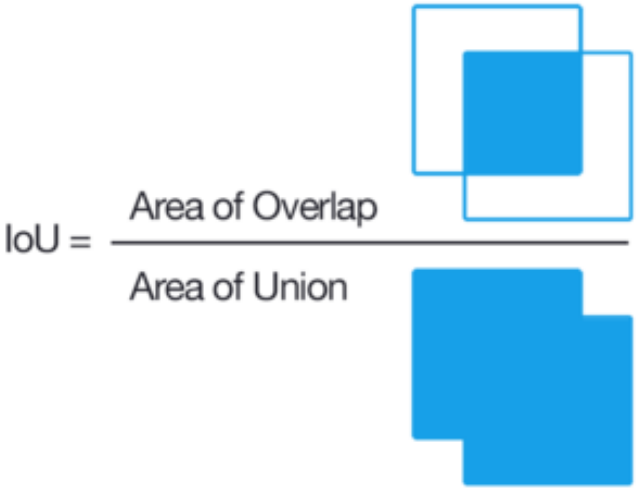

Fig. 8. IoU evaluation

5.2. Experiments and results

In the experiment, we construct three different models and split our dataset to training set and validation set to compare them. We adopted a early stopping strategy with validation loss as monitor to prevent over-fitting. Through the training procedure, we save the best model.

We plot the loss and the IoU for all three models to check if the model converges. The loss curve and the IoU curve of three experiments are shown in the Fig. 9, Fig. 11, Fig. 10. We can see from the figures that the loss of Unet is obviously lower than the other 2 models especially for the validation set.

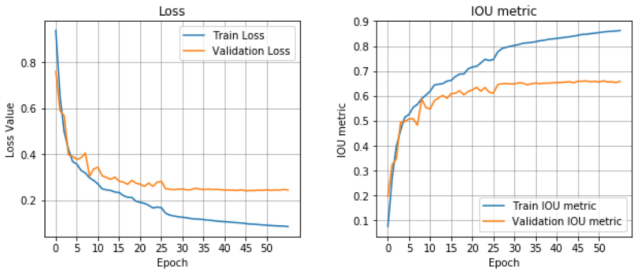

Fig. 9. Loss and IoU curve of Unet

In the next step, we applied the trained models to testing data and plot the relationship between IoU and threshold to get the best prediction result for all there models. The results for the three models are shown in Table 1. WHen we look at the validation set, the Unet achieves a IOU of 0.661, which is

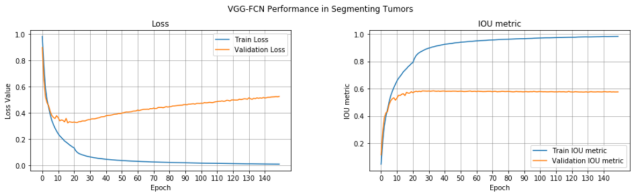

Fig. 10. Loss and IoU curve of FCN8-VGG16

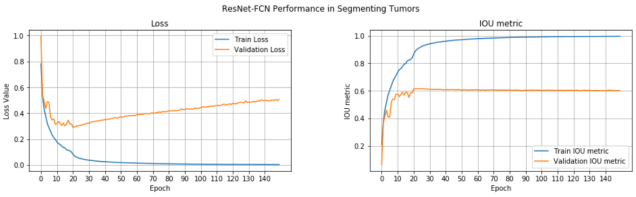

Fig. 11. Loss and IoU curve of FCN8-ResNet34

significantly better than FCN8-VGG16 at 0.579 and FCN8-ResNet34 at 0.617.

Fig. 12. one of the input images of Unet and its segmentation result with manual ground truth

Fig. 13. one of the input images of FCN8-VGG16 and its segmentation result with manual ground truth

Finally, we plot the segmentation result of our models in the original image to compare it with the ground truth. Some examples are shown in the following figures(Fig. 12, Fig. 13, Fig. 14). We can see that all 3 models can fulfill the segmentation tasks pretty good.

As we can see from the table and figures. Unet out performs the other two models significantly. Compared with FCN, etc., UNet performed upsampling 4 times and used skip connection in the same stage instead of directly supervising and loss-reverting on high-level semantic features, which ensured the final result capture more low-level features, which also allows the fusion of features of different scales, so that multi-scale prediction and deep supervision can be performed. The upsampling also makes the segmented image more precise considering edge restoration, etc.

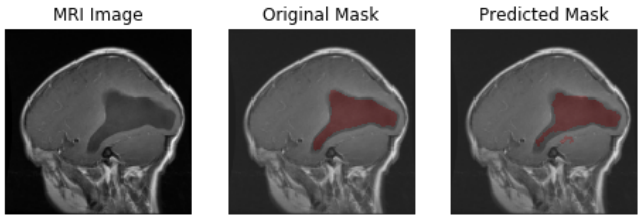

Fig. 14. one of the input images of FCN8-ResNet34 and its segmentation result with manual ground truth

6. CONCLUSION AND FUTURE WORK

As shown in results, Unet has better performance under the same dataset. In future work, we try to use a pyramid pooling module to aggregate different-region-based context, and that is Pyramid Scene Parsing Network (PSPNet)[14]. It has two major parts, a CNN to get the feature map of the given input image and a pyramid parsing module to harvest different sub-region representations.

Table 1. Performance of Unet, FCN8-VGG16, FCN8-ResNet34

Model Name	Train Threshold	Train IOU	Validation Threshold	Validation IOU
UNet	0.495	0.861	0.242	0.661
FCN8-VGG16	0.707	0.770	0.273	0.579
FCN8-ResNet34	0.525	0.892	0.172	0.617

A. CONTRIBUTION

1. *Xibo Zhang*: Find the brain tumor dataset, construct FCN8-VGG Model and FCN8-ResNet Model, train these two models and do the test.

2. *Chenhao Zhou*: Complete the code of data pre-processing and measure function, train the Unet Model and do the test.

3. *Zihan Wang*: Help with the report.

B. REFERENCES

[1] Raymond Damadian. Tumor detection by nuclear magnetic resonance. *Science*, 171(3976):1151–1153, 1971.

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

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

[4] M Gouskir, MA Zyad, and M Boutalline. Automatic analysis of brain tumor from magnetic resonance images based on geometric median shift. In *2020 IEEE 6th International Conference on Optimization and Applications (ICOA)*, pages 1–7. IEEE, 2020.

[5] Muhammad Yousuf, Khan Bahadar Khan, Muhammad Adeel Azam, and Muhammad Aqeel. Brain tumor localization and segmentation based on pixel-based thresholding with morphological operation. In *International Conference on Intelligent Technologies and Applications*, pages 562–572. Springer, 2019.

[6] Anindya Apriliyanti Pravitasari, Nur Iriawan, Mawanda Almuhayar, Taufik Azmi, Kartika Fithriasari, Santi Wulan Purnami, Widiana Ferriastuti, et al. Unet-vgg16 with transfer learning for mri-based brain tumor segmentation. *Telkomnika*, 18(3), 2020.

[7] Venkatesan Rajinikanth, Alex Noel Joseph Raj, Krishnan Palani Thanaraj, and Ganesh R Naik. A customized vgg19 network with concatenation of deep and handcrafted features for brain tumor detection. *Applied Sciences*, 10(10):3429, 2020.

[8] M Syed, P Flechsig, J Liermann, P Windisch, F Staudinger, S Akbaba, S Körber, C Freudlsperger, P Plinkert, J Debus, et al. Fibroblast activation protein (fapi) pet for diagnostics and advanced target volume delineation in head and neck cancer. *Nuklearmedizin*, 59(02):V50, 2020.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. pages 234–241, 2015.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2014.

[11] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[13] Adrian Rosebrock. Intersection over union (iou) for object detection. *Pyimageresearch*, 3(7), 2016.

[14] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.