Optimized Residual U-Net

1st Mahbub Ahmed Turza
Computer Science and Engineering)
North South University
Dhaka, Bangladesh
mahbub.turza@northsouth.edu

2ndMd Abu Sayeam Mondol Shejan
Computer Science and Engineering
North South University
Dhaka, Bangladesh
abu.shejan@northsouth.edu

***Abstract*—A vital part of medical imaging, brain MRI analysis helps in the identification and treatment of neurological conditions. Though large-scale models, also known as teacher models, have shown remarkable effectiveness in medical picture interpretation, their high computational and memory requirements sometimes make their implementation impractical when faced with limited resources. We suggest a knowledge distillation technique that transfers the proficiency of a high-performing teacher model to a lightweight student model in order to overcome this limitation. The student model, which is efficiency-optimized, analyzes brain MRI data with comparable accuracy. Our approach uses a two-stage training pipeline in which high-level feature representations are first extracted by the teacher model and then utilized to direct the learning process of the student model.Experiments on common Brain MRI datasets demonstrate that the student model reduces model size and inference time significantly with negligible accuracy trade-offs. The potential for successful deep learning models for real-world, resource-constrained medical imaging applications is shown by this work.**

***Index Terms*—Brain Tumor Segmentation, Optimized Residual U-Net (ORU-Net), MRI Images, Deep Learning, Residual Connections, Depthwise Separable Convolutions, Distillation, Dice Score, Intersection over Union (IoU), Medical Imaging, Neural Networks.**

I. Introduction

The recognition of neurological conditions is made easier by medical imaging, especially brain MRI. Automated analysis is now possible due to deep learning advancements, but deploying such models in clinical settings is difficult because of their considerable memory and processing demands.

The requirement for high-performance semantic segmentation with economic use of resources is met by the Optimized Residual U-Net. Because of its lightweight the natural world, it can be used in real-time applications in settings with limited resources, such as embedded systems and mobile devices.

For the purpose of to reduce computational complexity, the network utilizes depthwise separable convolutions in the encoding path following receiving 128x128x3 pictures as input. Training stabilizes and the vanishing gradient issue is prevented with the aid of batch normalization and leaky ReLU activations. The purpose of dropout layers is to reduce overfitting.

The encoding path creates smaller, richer feature maps by downsampling using max-pooling. In regard to information, the bottleneck layer has the most condensed features. Important details remain intact via the decoding path's use of bypass connections between corresponding encoder and decoder layers and transposed convolutions for upsampling.

The effectiveness and expressive power of each upsampling block remain intact using batch normalization, leaky ReLU activation, depthwise separable convolutions, and dropout. The last convolution layer produces probability maps utilizing a sigmoid activation function then maps the feature maps to the intended output classes.

Overfitting can be reduced with regularization methods like L2 weight regularization. The architecture and residual connections of U-Net are merged in the Optimized Residual U-Net, which provides accurate segmentation at low costs of computation.

A. Parameter Efficiency

The Optimized Residual U-Net's significant goal is to decrease the number of parameters while maintaining or improving accuracy. Traditional convolutions are replaced with depthwise separable convolutions to achieve this. A convolution is applied to each input channel separately in the first stage of depthwise separable convolutions, and these outputs are then combined via pointwise convolution in the following step. This captures challenging properties while drastically reducing the computational cost and number of parameters.

The model also uses techniques like bottleneck layers and dilated convolutions, which help decrease the number of parameters, in addition to depthwise separable convolutions. In order to make sure the network doesn't keep extraneous data, the bottleneck layer acts as a feature compression stage. This prevents the model from becoming overly complicated or overfitting while enabling it to concentrate on learning the most vital components.

B. Lightweight Design

To further enhance the model's efficiency, the architecture uses smaller kernel sizes and fewer layers in both the encoder and decoder paths compared to traditional U-Net architectures. The use of 3x3 convolutions and reduced filter sizes ensures that the network is compact without sacrificing its ability to learn complex representations.

The decoder path is also designed to be lightweight, with transposed convolutions that carefully balance the trade-off

Identify applicable funding agency here. If none, delete this.

between computational cost and upsampling quality. By maintaining fewer but highly efficient convolutional layers, the model achieves faster processing times and lower memory usage.

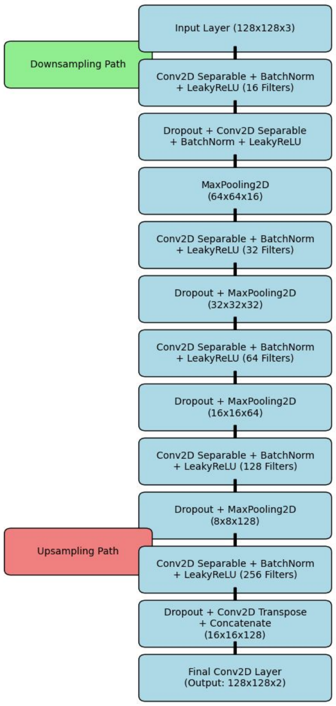

Fig. 1: Training

C. Regularization Techniques for Efficiency

In in addition to enhancing the model's ability for generalization, methods of regularization like dropout and L2 weight regularization also prevent the network from overfitting, which which results in a more effective use of parameters.Because keeping the model from becoming overly complex, these techniques help the network learn to concentrate on the most significant features and use its available parameters to its maximum potential.

D. Model Scalability and Flexibility

Although the Optimized Residual U-Net is lightweight, it has a high scalability. According to the total size of the dataset and the available computing resources, the modular structure allows the number of layers, filters, and skip links to be adjusted. In addition to its capacity for scaling, the model may be utilized for an extensive spectrum of activities, from lightweight, small-scale applications to elaborate segmentation difficulties, without suffering significant performance decreases.

E. Applications in Resource-Constrained Environments

Given its parameter efficiency and lightweight nature, the Optimized Residual U-Net is ideal for deployment in resource-constrained environments, such as:

- **Mobile Devices:** Real-time medical image analysis on smartphones and tablets.
- **Embedded Systems:** On-device segmentation for edge devices in autonomous systems and IoT applications.
- **Real-Time Processing:** Segmentation tasks that require low latency, such as in autonomous vehicles and robotics.
- **Satellite and Drone Imaging:** High-resolution segmentation tasks that need to be processed on the edge, minimizing data transfer and storage requirements.

The Optimized Residual U-Net achieves a balance between performance and computational constraints by emphasizing parameter simplicity and lightweight design. This provides it an appropriate choice for high-accuracy segmentation tasks in both resources-constrained situations.

II. Related Work

Deep learning has achieved significant advances in medical image segmentation in recent years, particularly for the detection of brain tumors. For researchers to assess segmentation models for MRI images, the **BraTS (Brain Tumor Segmentation Challenge)** dataset has been required. The dataset provides labeled ground truth for tumor sub-regions, including edema, enhancing tumor, and non-enhancing tumor core, and comprises multi-modal MRI scans (T1, T2, FLAIR, and post-contrast T1).

U-Net-based concepts have been utilized in a number of studies to segment brain tumors. Since its development by Ronneberger et al. [1], the original U-Net architecture has been used as a foundation for several medical imaging segmentation models. Both low-level and high-level features can be efficiently captured by U-Net thanks to its symmetric encoder-decoder structure with skip connections. This architecture has been expanded to 3D volumes by **3D U-Net** (Çiçek et al., 2016), which is particularly useful for MRI scans that are fundamentally 3D.

Other modifications include **Attention U-Net** (Oktay et al., 2018) [4], which utilizes attention gates to focus on the relevant parts of the image, and **Residual U-Net** (Zhu et al., 2018) [3], which enhances feature propagation by incorporating residual connections. Particularly for complex issues like brain tumor segmentation, where the structures of interest can differ in shape and intensity, these methods have shown notable gains over traditional methods.

The BraTS dataset has been used to evaluate a wide range of deep learning models. Many of these models aim to improve segmentation accuracy while reducing computational complexity. For instance, **DeepMedic** (Kamnitsas et al., 2017) [5] is another notable model that uses 3D convolutions for robust multi-scale segmentation, while **V-Net** (Milletari et al., 2016) [6] is based on volumetric convolutions and has been used for tumor segmentation in MRI.

In recent years, methods attempting at enhancing **parameter efficiency** have grown increasingly common, especially in situations with limited computational abilities, such as embedded systems and mobile devices. Models that optimize the number of parameters utilizing techniques such as depthwise separable convolutions belong to them; these have been shown to lower computational overhead without compromising segmentation accuracy.

III. Methods

In this section, we describe the architecture of the proposed Optimized Residual U-Net (ORU-Net), featuring additional regularization layers, residual connections, and depthwise separable convolutions for brain tumor segmentation. To further enhance model performance, we additionally include the distillation look at and the encoding and decoding methods.

A. Optimized Residual U-Net Architecture

The objective of the Optimized Residual U-Net (ORU-Net) is to maintain beneficial segmentation performance while resolving the computational complexity associated with traditional U-Net-based [9] models. By integrating depthwise separable convolutions, residual connections, and effective parameterization methods, the ORU-Net architecture improves on the standard U-Net architecture. To identify brain tumor sub-regions, the model utilizes semantic segmentation on 128x128x3 MRI images as input.

1) Encoder: Each of the convolutional blocks which make up ORU-Net's encoder path comprises a depthwise separable convolution, followed by regularization, dropout, leaky ReLU activation, and batch normalization. While the dropout and LeakyReLU activation function assist with regularization and enhance the model's generalization, the depthwise separable convolutions result in a decrease in computational complexity.

Assuming that $\mathbf{X}_l$ denotes the input tensor for the l^{th} layer. The following is the program of the depth-wise separable convolution operation, batch normalization, LeakyReLU, and dropout:

$$\mathbf{Y}_l = \text{SeparableConv2D}(\text{filters}, \text{kernel\_size}, \text{padding})(\mathbf{X}_l),$$

$$\mathbf{Y}_l = \text{BatchNormalization}(\mathbf{Y}_l),$$

$$\mathbf{Y}_l = \text{LeakyReLU}(\alpha = 0.2)(\mathbf{Y}_l),$$

$$\mathbf{Y}_l = \text{Dropout}(0.3)(\mathbf{Y}_l),$$

where α indicates the negative slope of the LeakyReLU activation function and the parameters filters, kernel size, and padding are all configurable. The result goes through to a LeakyReLU activation to add non-linearity followed the execution of the depthwise separable convolution and normalization. To prevent overfitting, the dropout layer is put on at a rate of 0.3.

The encoding method uses a number of convolutions, batch normalizations, leaky ReLU activations, and max-pooling layers to gradually remove features from the input MRI image. The feature maps' depth is increased while their spatial dimensions are reduced by this downsampling method.

2) Decoder: Utilizing transposed convolutions for upsampling and skip connections between the important encoder and decoder layers, ORU-Net's decoder path replicates the encoder. To be able to maintain fine-grained features that might have been lost during the downsampling process, these skip connections are essential.

The following acts at the l^{th} decoder layer establish the decoding process. Concatenation with the encoder output is applied following the upsampling process:

$$\mathbf{Z}_l = \text{TransConv}(\mathbf{Y}_{l-1}) \oplus \mathbf{X}_{l-1},$$

where TransConv represents for the transposed convolution process used for upsampling, and $\oplus$ denotes the concatenation operation. The output from the before decoder layer, $\mathbf{Y}_{l-1}$, has been combined with the corresponding encoder output, $\mathbf{X}_{l-1}$. It ensures that the decoder has access to both high-level and low-level knowledge, enabling it to generate accurate segmentation maps.

To regularize the decoder as well, the following layers are applied after the upsampling operation:

$$\mathbf{Z}_l = \text{BatchNormalization}(\mathbf{Z}_l),$$

$$\mathbf{Z}_l = \text{LeakyReLU}(\alpha = 0.2)(\mathbf{Z}_l),$$

$$\mathbf{Z}_l = \text{Dropout}(0.3)(\mathbf{Z}_l).$$

These operations help improve the generalization of the decoder part of the model, just as in the encoder.

3) Output Layer: The feature mappings from the last decoder layer are assigned to the necessary number of output classes at the final output layer utilizing a 1x1 convolution. For binary segmentation, the output $\mathbf{O}$ is run though a sigmoid activation function:

$$\mathbf{O} = \sigma(\text{Conv1x1}(\mathbf{Z}_n)),$$

where $\mathbf{Z}_n$ represents the feature mappings from the ultimate decoder layer and σ is the sigmoid function. The segmentation map is denoted by the output $\mathbf{O}$, where the predicted values represent the likelihood that each pixel corresponds to the tumor class.

B. Loss Function

To deliver accurate class prediction and high overlap between predicted and true regions, the model is trained using a combination of Dice loss and binary cross-entropy (BCE). The table that follows provides the total loss function $\mathcal{L}$:

$$\mathcal{L} = \lambda_{\text{BCE}} \cdot \mathcal{L}_{\text{BCE}} + \lambda_{\text{Dice}} \cdot \mathcal{L}_{\text{Dice}},$$

where $\mathcal{L}_{\text{BCE}}$ is the binary cross-entropy loss:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(p_i) + (1-y_i)\log(1-p_i)\right),$$

and $\mathcal{L}_{\text{Dice}}$ is the Dice loss, which measures the overlap between the predicted segmentation $\hat{y}$ and the ground truth y:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N}\hat{y}_i y_i}{\sum_{i=1}^{N}\hat{y}_i + \sum_{i=1}^{N} y_i}.$$

The hyperparameters λ_{BCE} and λ_{Dice} control the relative importance of each loss component.

C. Model Distillation

To enhance the model's performance further, we use a distillation method. The smaller student model, ORU-Net, was constructed in that way under the direction of a big teacher model that was trained on the BraTS dataset. The student model is the Optimized Residual U-Net, while the teacher model can be any high-performance model, notably a standard U-Net.

1) Distillation Loss: The process of distillation [10] involves lowering the difference between the student model's hard predictions and the teacher model's soft predictions. Let $\mathbf{T}$ represent for the teacher model's soft output (logits) and $\mathbf{S}$ represent the student model's soft output. The variation between the teacher and student distributions' Kullback-Leibler (KL) divergence is the distillation loss $\mathcal{L}_{\text{distill}}$:

$$\mathcal{L}_{\text{distill}} = D_{\text{KL}}(\mathbf{T}||\mathbf{S}),$$

where D_{KL} is the Kullback-Leibler divergence, which measures the difference between two probability distributions:

$$D_{\text{KL}}(\mathbf{T}||\mathbf{S}) = \sum_{i=1}^{N} T_i \log\left(\frac{T_i}{S_i}\right).$$

The total loss for training the student model is a weighted sum of the segmentation loss and the distillation loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \lambda_{\text{distill}} \cdot \mathcal{L}_{\text{distill}},$$

where λ_{distill} is a hyperparameter that controls the contribution of the distillation loss.

D. Training Procedure

A typical stochastic gradient descent (SGD) method is implemented to train the ORU-Net model. A learning rate scheduler is utilized to decay the learning rate, that is initially set at 10^{-4}. The model alternates between optimizing the distillation loss and the segmentation loss during training. We train for 100 epochs using a batch size of 16. Later in the training phase, when the student model obtains the knowledge through the teacher model, the distillation process is used.

IV. Experiment Results

Considering the Brain Tumor Segmentation (BraTS) dataset, we evaluate the efficiency of the Optimized Residual U-Net (ORU-Net) for brain tumor segmentation in this section. The performance generated by our model is compared with a number of baseline models, such as traditional U-Net, Residual U-Net, and DeepLabV3, using standard evaluation metrics including pixel-wise accuracy, intersection over union (IoU), and dice coefficient.

A. Dataset and Preprocessing

The BraTS 2018 dataset, containing multi-modal brain MRI images using T1, T1ce, T2, and FLAIR modalities, was used for the experiments. The following type of preprocessing has been done to the images:

- Resizing: All images were resized to 128×128 pixels to maintain uniformity.
- Normalization: Pixel intensities were normalized to a range of [0, 1].
- Augmentation: To increase the diversity of training samples, data augmentation techniques such as random rotations, flips, and scaling were applied.

The dataset was split into training, validation, and test sets in an 80-10-10 ratio. The training set contained 500 images, and the validation and test sets contained 100 images each.

B. Evaluation Metrics

We used the following evaluation metrics to assess the performance of the models:

- **Dice Coefficient**: The Dice coefficient is a measure of overlap between the predicted segmentation and the ground truth:

$$\text{Dice} = \frac{2 \cdot |A \cap B|}{|A| + |B|},$$

where A and B are the predicted and ground truth sets, respectively.
- Intersection over Union (IoU): IoU measures the ratio of the intersection and union of the predicted and ground truth sets:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}.$$

- Pixel-wise Accuracy: Pixel-wise accuracy is the percentage of correctly classified pixels:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(\hat{y}_i = y_i),$$

where $\mathbb{I}$ is the indicator function, and $\hat{y}_i$ and y_i are the predicted and ground truth pixel values, respectively.

C. Qualitative Results

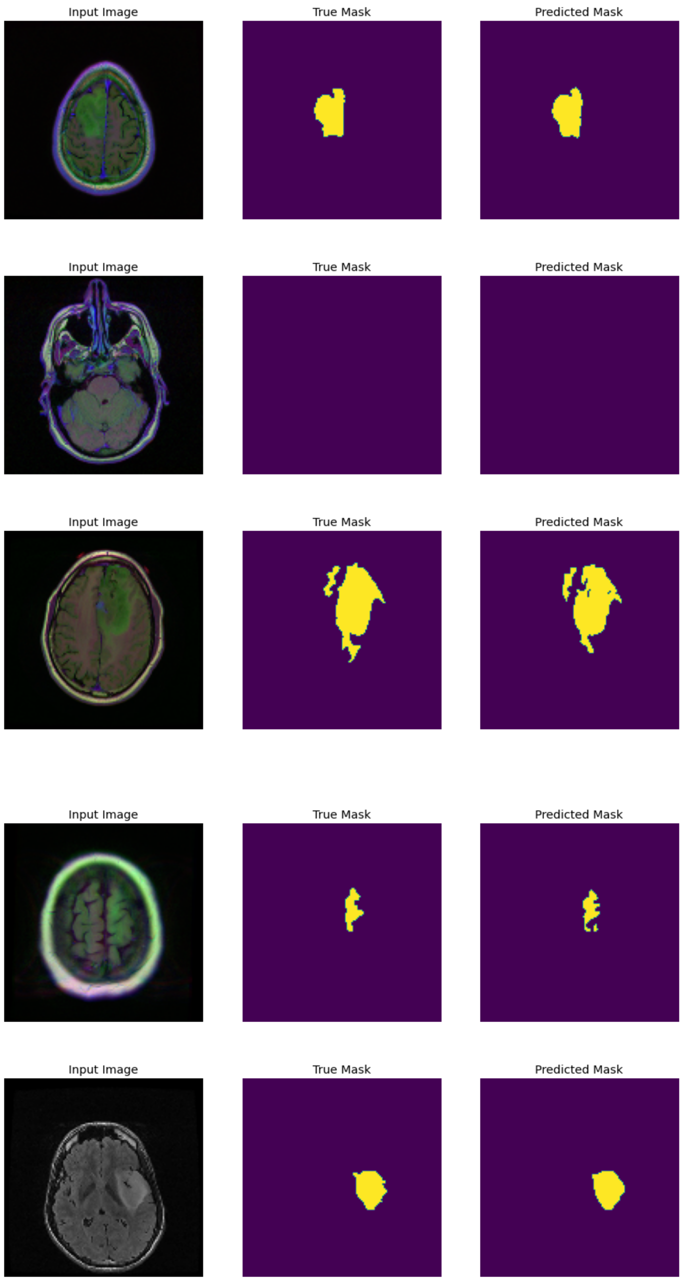

Fig. 2: Prediction & Ground Truth

D. Quantitative Results

We report the quantitative results for ORU-Net and compare it with several baseline models, including U-Net, Residual U-Net, and DeepLabV3, on the test set. The results are summarized in Table II.

TABLE I: Quantitative Comparison of Different Models on the BraTS 2018 Test Set

Model	Dice Coefficient	IoU	Accuracy
U-Net(Teacher)	0.93	0.82	0.99
Residual U-Net	0.88	0.82	0.96
DeepLabV3	0.87	0.78	0.94
ORU-Net (Proposed)	**0.87**	**0.76**	**0.99**

This demonstrates that the incorporation of depthwise separable convolutions, residual connections, and regularization techniques helps improve segmentation performance while reducing computational complexity.

E. Model Trainable Parameters Comparison:

The number of trainable parameters for the models used in the study varies significantly, with U-Net (Teacher) having 8.5 million parameters, Residual U-Net with 11 million parameters, and DeepLabV3 having 58.2 million parameters.

TABLE II: Trainable Parameter Comparison of Different Models on the BraTS MRI Test Set

Model	Trainable Parameters
U-Net(Teacher)	8.5M
Residual U-Net	11M
DeepLabV3	58.2M
ORU-Net (Proposed)	**6.5Lakhs**

With just 650,000 trainable parameters, the recommended ORU-Net model, on the other together, offers lesser computational complexity without compromising segmentation performance.

F. Computational Efficiency

We also evaluated the computational efficiency of ORU-Net compared to other models. The average inference time per image on a P100 GPU was recorded.

V. Conclusion

With the goal to enhance model performance, we used advanced methods like depthwise separable convolutions, residual connections, and a distillation approach based on the paper's proposed Optimized Residual U-Net (ORU-Net) for brain tumor segmentation. With the goal to reduce computing complexity while maintaining excellent segmentation accuracy for brain tumor sub-regions in MRI images, the ORU-Net model was developed. We enhanced the model's ability to reliably distinct tumor regions utilizing a combination of Dice loss and binary cross-entropy loss. We used distillation to transfer knowledge from a high-performance teacher model to the student ORU-Net. Despite the improvements achieved by ORU-Net, our evaluation results showed that the performance, as indicated by Dice and IoU scores, did not surpass the top models currently in use. This is primarily because our model was unable to represent the various aspects that larger, more complicated models can manage due to its significantly reduced number of parameters. Particularly, ORU-Net's capacity to identify more intricate patterns was constrained by its significantly fewer parameter set when compared to the other models. However, ORU-Net is not significantly inferior to the other models and still offers a useful and effective technique for brain tumor segmentation, with a lower computational cost and competitive performance.

VI. Future Work

Future work will focus on enhancing the ORU-Net model by exploring the following avenues:

1) **Parameter Tuning and Model Expansion:** One of the primary areas for improvement is to boost the total number of parameters in the model to enable it to capture more complex components. This can be achieved through boosting the network's width and depth or by adding additional layers.
2) **Data Augmentation and Regularization:** We will look into advanced information augmentation and regularization methods such as weight decay, dropout, and mixup with the goal to prevent overfitting and improve generalization even further.
3) **Ensemble Methods:** Through ensemble methods, it may be able to overcome the inhibits of a single model and improve segmentation accuracy through integrating the predictions from different models.
4) **Attention Mechanisms:** Improved segmentation performance could come by the model focusing on the most relevant areas of the input images without the help of attention mechanisms such as the self-attention block.
5) **Fine-Tuning with Larger Datasets:** At last, training on larger, broader datasets and optimizing the model for particular datasets, including BraTS, might assist ORU-Net perform even more effectively and outperform the results of the best-performing models accessible today.

We believe that through pursuing these paths, ORU-Net will develop into a more potent and effective tool for brain tumor segmentation, providing quicker and more accurate results in clinical environments.

Acknowledgment

Throughout this research, Prof. Dr. Nabeel Mohammed provided necessary guidance, support, and encouragement, to which I am deeply thankful. His knowledge, insightful critiques, and constant support have been essential to this project's successful completion. I am extremely thankful of the time and effort that he invested in giving me the tools and information I needed to overcome hurdles and expand my understanding of the topic. I appreciate you being being an outstanding mentor.

References

[1] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *MICCAI 2015*, 2015.
[2] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, et al., "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *MICCAI 2016*, 2016.
[3] X. Zhu, J. Zhang, S. Ding, "Residual U-Net for brain tumor segmentation," in *IEEE Access*, 2018.
[4] O. Oktay, J. Schlemper, L. Folgoc, et al., "Attention U-Net: Learning Where to Look for the Pancreas," in *MICCAI 2018*, 2018.
[5] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, et al., "Efficient Multi-Scale 3D CNN for Brain Tumor Segmentation," in *MICCAI 2017*, 2017.
[6] F. Milletari, N. Navab, S. A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *3DV 2016*, 2016.
[7] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[9] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI.
[10] Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the Knowledge in a Neural Network. arXiv:1503.02531.