Kidney Image Segmentation

Kai Li
Faculty of Applied Science & Engineering
University of Toronto
Toronto, Canada
kaii.li@mail.utoronto.ca

Hanson Liu
Faculty of Applied Science & Engineering
University of Toronto
Toronto, Canada
hansonzh.liu@mail.utoronto.ca

Benjamin Mah
Faculty of Applied Science & Engineering
University of Toronto
Toronto, Canada
benjamin.mah@mail.utoronto.ca

***Abstract*—In the domain of medical imaging, manual annotation of blood vessels in human kidney images is often found to be a labor-intensive and non-scalable task, highlighting the need for automated segmentation solutions. While U-Net architectures have been commonly employed for biomedical image segmentation tasks due to their efficacy in various contexts, their basic architecture faces limitations in handling the variability in kidney shapes and sizes, leading to suboptimal segmentation accuracy. To address these challenges, our strategy transitions from the initially proposed U-Net to a Feature Pyramid Network (FPN) architecture, enhanced with a SE-ResNeXt-50 (32 × 4d) backbone incorporating squeeze and excitation blocks. This shift leverages FPN's abilities in capturing multi-scale spatial information, a capability that is particularly important given our training on the 3D Hierarchical Phase-Contrast Tomography (HiP-CT) dataset [1]. The performance of this model is evaluated and assessed with the DICE loss and DICE score metrics, revealing that the refined FPN architecture improves the accuracy of kidney blood vessel segmentation. This progression highlights the effectiveness of integrating a pre-trained backbone, squeeze and excitation blocks, and DICE score, further proving the FPN architecture's utility in kidney image segmentation.**

I. Introduction

The development of a Common Coordinate Framework (CCF) has emerged as a significant advancement, particularly through the adoption of the Vasculature Common Coordinate Framework (VCCF). This approach uses the network of blood vessels as a navigational system to map the trillions of cells throughout the entire human body. The VCCF, especially with its application in localizing kidney cells, allows researchers and medical professionals to understand cellular functions and relationships on the premise that every cell is within a proximate distance to the vasculature [2]. This allows for a clear representation from the organ level down to individual cells.

However, the goal of accurately mapping the vasculature to contribute to the VCCF faces many challenges, one of them being the labor-intensive and slow task of manually annotating vascular structure, which often requires over six months for a single dataset completion [1]. This constraint significantly affects the pace of research and the application of machine learning models, which may struggle to generalize across the evolving imaging techniques like Hierarchical Phase-Contrast Tomography (HiP-CT). Furthermore, the reliance on expert annotators for this meticulous task is not only extremely costly, but also introduces variability, highlighting the need for an automated and accurate segmentation solution [3].

Acknowledging the challenges inherent in manual annotation for vasculature mapping, it becomes crucial to explore other solutions, such as the task of machine segmentation in the field of machine learning. Image segmentation is the task of dividing an image into sub-regions based on extracted features such as color or texture. This process has significantly enhanced precision medicine through the development of computer-aided diagnosis systems, which leverage a multitude of imaging techniques and modalities, including Magnetic Resonance Imaging (MRI) and the previously mentioned HiP-CT. An example of a machine learning model architecture that has been tailored for biomedical image segmentation is U-Net, which is a deep convolutional neural network that employs an encoding and decoding path for precise localization [4]. Similarly, the Feature Pyramid Network (FPN) builds upon the foundation laid by U-Net, particularly in addressing the challenges of handling the variability of medical imaging. The FPN architecture is able to integrate high-level semantic information with the spatial information from shallow layers. Further refining the capabilities of FPNs has demonstrated potential for improving segmentation accuracy, as shown by studies achieving up to a 99.1% accuracy in brain tumor segmentation [5].

In the task of kidney biomedical image segmentation, the process to accurately identify structures is complicated by the organ's complex vasculature and variability in its anatomical composition. These complexities pose significant challenges and may lead to low accuracy. Therefore, the application of a machine learning model, more specifically incorporating refined FPN architectures, emerges as a compelling strategy. This approach is especially effective in dealing with the intricate vascular system of the kidney, as opposed to traditional methods that may fail to capture important details in the images. By evaluating the performance of various machine learning models with the task of biomedical image segmentation through existing literature, this paper aims to develop a clear problem statement that addresses the current challenges of this task for kidney vasculature analysis. This approach paves the way for a comprehensive literature review and positions our research in the evolving landscape of medical imaging.

II. Literature Review

A. U-Net: Convolutional Networks for Biomedical Image Segmentation

The "U-Net: Convolutional Networks for Biomedical Image Segmentation" paper by O. Ronneberger, P. Fischer, and T. Brox presents a network that leverages data augmentation to effectively utilize limited annotated samples, enhancing performance in biomedical image segmentation where large datasets are scarce. The U-Net model is fast, processing 512x512 images in less than a second on modern GPUs, showcasing the model's practicality for real-time applications. Furthermore, the model's versatility is demonstrated across various biomedical segmentation tasks, highlighting the model's adaptability and potential in diverse medical fields [6].

B. MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation

The "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation" paper introduces MultiResUNet, enhancing U-Net by integrating MultiRes blocks with adaptively increasing filters, determined by the parameter W=×U, with =1.67. This architecture adjustment, along with the replacement of U-Net's connections with Res paths, uses Sigmoid and ReLU functions for activation and optimizes performance across different layers. The model ran tests on five medical imaging datasets: Endoscopy, Dermoscopy, Fluorescence microscopy, MRI, Electron microscopy, achieving improvements of 10.57%, 5.07%, 2.63%, 1.41%, 0.62%, respectively [7].

C. Loss odyssey in medical image segmentation

The paper "Loss odyssey in medical image segmentation" by J. Ma et al. highlights the critical role of loss functions in medical image segmentation, noting the absence of their systematic study. It examines 20 loss functions across four 3-dimensional segmentation tasks using data from six public datasets. The study found that compound loss functions like Hausdorff distance loss, boundary loss, focal loss, and DICE combined with TopK loss demonstrated robustness and high accuracy [8].

D. DRU-net: a novel U-net for biomedical image segmentation

The paper "Dru-Net: A novel U-Net for biomedical image segmentation" by X. Hu and H. Yang critiques U-Net's limitations with varying biomedical shapes and sizes due to its standard convolutions' inability to handle geometric transformations effectively. It presents DRU-net, featuring a deformable encoder for better geometric transformation learning and a reshaping upsampling convolution decoder for improved resolution and feature fusion. In addition, DRU-net employs focal loss to address class imbalance and achieves competitive accuracy on various biomedical image segmentation datasets with fewer parameters than U-Net [9].

E. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation

The paper "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation" by H. Huang et al. enhances UNet and UNet++ by introducing UNet 3+, which incorporates full-scale skip connections to merge low-level details with high-level semantics, crucial for accurately capturing the variable sizes of kidneys. Furthermore, UNet 3+ achieves computational efficiency through reduced network parameters. The model further employs a hybrid loss function, enhancing organ boundary delineation and minimizing over-segmentation in non-organ images, thus improving segmentation accuracy [10]

F. EG-TransUNet: a transformer-based U-Net with enhanced and guided models for biomedical image segmentation

The "EG-TransUNet: a transformer-based U-Net with enhanced and guided models for biomedical image segmentation" paper by S. Pan et al. explores the use of attention-based Transforms in both the encoder and decoder, leading to the development of EG-TransUNet. This upgraded UNet architecture integrates three Transformer-enhanced modules—progressive enhancement, channel spatial attention, and semantic guidance attention—to improve spatial detail and semantic accuracy in image segmentation. EG-TransUNet demonstrated superior generalization across datasets, achieving DICE scores of 93.44% and 95.26% on the Kvasir-SEG and CVC-ClinicDB datasets, respectively [11].

G. Evaluating Kidney Segmentation Using Different Attention U-Net Architectures

The paper "Evaluating Kidney Segmentation Using Different Attention U-Net Architectures" by V. Alevizos and M. Hon investigates automatic kidney detection using machine learning, testing various Attention U-Net architectures with backbones like VGG19, ResNet152V2, and EfficientNetB7 on a dataset of 200 kidney samples. Their analysis using Jaccard index/IoU and DICE coefficient metrics revealed that backbone-less models outperformed those with backbones, though complex data benefits from ImageNet architectures with non-frozen weights and Attention U-Net. The study emphasizes the importance of considering dataset complexity in model selection for segmentation [12].

H. MDA-Net: Multi-Dimensional Attention-Based Neural Network for 3D Image Segmentation

The paper "MDA-Net: Multi-Dimensional Attention-Based Neural Network for 3D Image Segmentation" by R. Gandhi and Y. Hong introduces MDA-Net, a U-Net based network for efficient 3D image segmentation in biomedical tasks, addressing the high memory and computational demands of traditional methods. MDA-Net employs slice-wise, spatial, and channel-wise attention mechanisms, and uses a squeeze and excitation approach to integrate information from adjacent slices, maintaining essential data while reducing complexity. Tested on the MICCAI iSeg and IBSR datasets for 3D brain

scans, MDA-Net showed superior performance, with potential applicability in other biomedical segmentation tasks like kidney imaging [13].

I. Mixed-Sized Biomedical Image Segmentation Based on U-Net Architectures

The paper "Mixed-Sized Biomedical Image Segmentation Based on U-Net Architectures" by P. Benedetti et al. explores the segmentation of variably sized organs, noting decreased accuracy with sparse binary masks in the 3D-IRCADb-01 dataset. The authors implemented an automated zooming technique to focus on specific organ areas, improving segmentation for organs with sparse masks. This method resulted in a 20% increase in segmentation accuracy as measured by the DICE coefficient [14].

J. CMM-Net: Contextual multi-scale multi-level network for efficient biomedical image segmentation

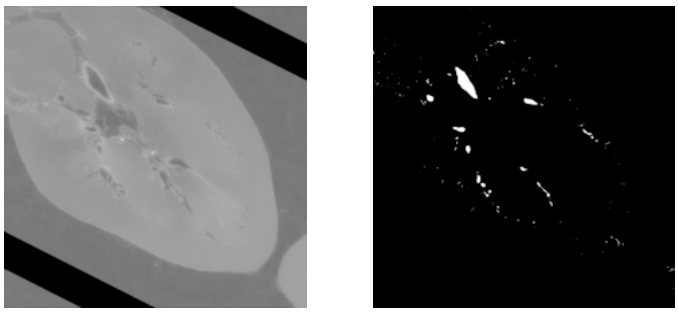
The "CMM-Net: Contextual multi-scale multi-level network for efficient biomedical image segmentation" paper by M. Al-masni and D.-H. Kim presents CMM-Net, an enhanced U-Net integrating global contextual features and a dilated convolution module for adapting to the complex anatomy of organs. The paper also introduces Inversion Recovery, an augmented testing method using logical operations to ensure comprehensive and accurate segmentation. CMM-Net's effectiveness is demonstrated on the ISIC 2017, DRIVE, and BraTS datasets, achieving DICE coefficients of 85.78%, 80.27%, and 88.96% respectively, indicating its superiority over traditional U-Net models [15].

III. Problem Statement

The objective of this paper is to develop and evaluate a refined Feature Pyramid Network (FPN) model that improves the segmentation of blood vessels in kidney images, crucial in automating and scaling the annotation processing in medical imaging. Despite the commonly implemented U-Net architecture in biomedical segmentation tasks, its limitations in handling the complex structures of the kidney vasculature contribute to the shift toward the FPN architecture. This paper will build upon the FPN architecture by integrating a backbone model that includes squeeze and excitation blocks. The evaluation will involve the 3D Hierarchical Phase-Contrast Tomography (HiP-CT) dataset [1], using both the DICE loss and DICE score metrics to assess the model's performance.

IV. Description of Dataset

This paper uses a dataset consisting of high-resolution 3D images of kidneys alongside 3D segmentation masks of their vasculature, derived from the Hierarchical Phase-Contrast Tomography (HiP-CT) imaging conducted by the European Synchrotron Radiation Facility (ESRF) and provided by the Human Organ Altas (HOA) [1]. The dataset contains images of kidneys from a wide demographic, accounting for discrepancies between age, sex, and BMI, allowing for a comprehensive perspective on human kidney anatomy. HiP-CT is able to capture high-resolution 3D data, ranging from $1.4\,\mu\text{m}$ to $50\,\mu\text{m}$. Each subset within the dataset is composed of TIFF scans representing 2D slices of 3D volumes, arranged to facilitate vertical stacking along the z-axis. Accompanying these images are blood vessel segmentation masks in TIFF format, as shown in Fig. 1. Out of the 2,780 images, 2,280 were allocated for training and 500 for validation, resulting in a split of approximately 82% for training and 18% for validation.



(a) 2D Kidney Image (b) 2D Kidney Image Mask

Fig. 1. Sample kidney image and mask

V. Methodology and Model

A. Data Pre-Processing

As with most machine learning tasks, particularly within the domain of image segmentation, data preprocessing is a necessary step to ensure effective learning in models. Initially, images are transformed into grayscale format using OpenCV, aiming to concentrate on intensity values while disregarding color information. Subsequently, these images are resized to dimensions of 1024 × 1024 pixels. This resizing is achieved by involving equal padding from all sides for images that fall short of the specified size. Following this, images are transformed from NumPy arrays to PyTorch tensors. These tensors undergo min-max normalization, a process in which pixel values are scaled to a normalized range of [0, 1] based on the tensor's minimum and maximum values. The normalized tensor is then rescaled to the [0, 255] range and transformed into an 8-bit unsigned integer format.

In parallel, the label or mask preprocessing involves a binarization step, where all non-zero pixel values are set to 255. This effectively segments the mask into background pixels (0) and object pixels (255), the latter representing blood vessels in the context of this paper.

The final step of preprocessing involves the application of random transformations to the training images. These transformations, which are controlled by the hyperparameter p_{augm}, dictate the likelihood of applying a given transformation. The assortment of transformations, including rotations, scaling, cropping, adjustments in contrast, Gaussian blur, motion blur, and grid distortion, is designed to improve the robustness of the model by mimicking the variability encountered in real-world scenarios. In this particular paper, the hyperparameter was set to a value of 0.1, or 10%. This preprocessing not only augments the data but it contributes to the model's

ability to generalize from the training data to unseen images, improving its applicability in practical settings. By applying augmentations, the number of images was increased from 2,780 to 4,574.

B. Model Architecture

The selection of the Feature Pyramid Network (FPN) as the baseline model for this study is due to its prevalent application in target detection tasks. This choice is justified by the model's demonstrated effectiveness in biomedical image segmentation tasks, where an improved FPN model achieved a 99.1% accuracy rate and a DICE score of 92% in the segmentation of brain tumors, as evidenced in [5]. The basic FPN architecture is built around three key components: encoding and decoding pathways, connected by horizontal connections. This configuration, similar to the U-Net model, reduces the spatial dimensions of the input while extracting high-level features during the encoding stage. The culmination of this path implements a 1 × 1 convolution layer as well as two 3 × 3 convolution layers to lay the groundwork for the segmentation feature map. During the decoding stage, the image is upsampled using nearest neighbor interpolation and is subsequently added element-wise with the feature maps from the encoding phase from the horizontal connections. This stage is further refined with the addition of two 3 × 3 convolution layers. All feature maps are then concatenated to create a 512-channel output. This output is processed through a 512 3 × 3 convolution filter, incorporating batch normalization and ReLU activation, finishing with a 1 × 1 convolution to derive the final feature map, as shown in Fig. 2 [5].

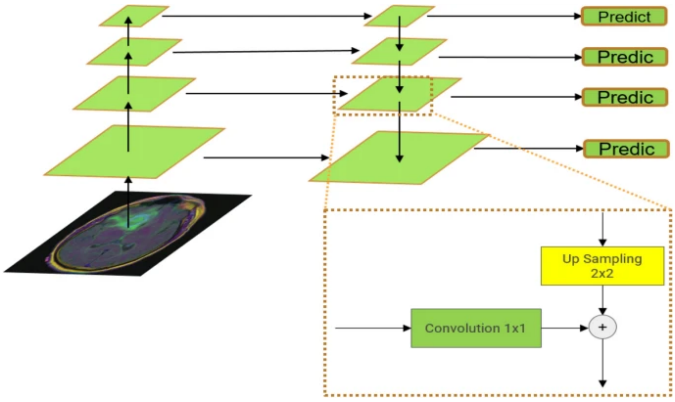


Fig. 2. Basic FPN Architecture

Building upon this foundation, this paper introduces a refined version of the FPN, incorporating the SE-ResNeXt-50 (32 × 4d) architecture as a backbone model. The SE-ResNeXt-50 (32 × 4d) incorporates squeeze-and-excitation (SE) blocks and grouped convolutions, allowing for the model to focus on informative features while suppressing the non-informative features and inversion incorporate—a technique that modifies the contrast in the kidney images, allowing our model to segment the blood vessels with greater ease. The CustomModel class defined in our codebase leverages the SE-ResNeXt-50 (32 × 4d) model as the encoder backbone, with the decoder specified by the PyTorch segmentation models (SMP) library. The architecture of the SE-ResNeXt-50 (32 × 4d) is fairly complex, as it is composed of 20 residual building blocks split into separate groups dependent on the output size [16]. The model initiates its feature extraction process with a layer of 64 convolution filters of size 7 × 7 with a stride of 2 to capture the broad features, followed by a 3 × 3 max pooling layer, also with a stride of 2, to reduce spatial dimensions. The first group of three residual building blocks each begin with a layer of 128 1 × 1 convolutions, followed by a layer of 128 3 × 3 convolutions, grouped into 32 channels (cardinality), enabling the network to process distinct features in separate channel groups. The block concludes with another layer of 256 1 × 1 convolutions, increasing the number of channels for the subsequent squeeze-and-excitation (SE) block. This SE block transforms the 56 × 56 feature maps through a fully-connected layer, denoted by fc, that has an input size of 16 and an output size of 256, fine-tuning the feature channels and allowing the model to focus on informative features. The structure of these blocks, as well as the remaining 17 blocks and the final output layer are detailed in Table I, adapted from [16].

TABLE I
Architecture of SE-ResNeXt-50 (32 × 4d)

Output Size	Residual Block Specifications
112 × 112	conv, 7 × 7, 64, stride 2
56 × 56	max pool, 3 × 3, stride 2
	[conv, 1 × 1, 128 conv, 3 × 3, 128 C = 32 conv, 1 × 1, 256 fc, [16, 256]] × 3
28 × 28	[conv, 1 × 1, 256 conv, 3 × 3, 256 C = 32 conv, 1 × 1, 512 fc, [32, 512]] × 4
14 × 14	[conv, 1 × 1, 512 conv, 3 × 3, 512 C = 32 conv, 1 × 1, 1024 fc, [64, 1024]] × 6
7 × 7	[conv, 1 × 1, 1024 conv, 3 × 3, 1024 C = 32 conv, 1 × 1, 2048 fc, [128, 2048]] × 3
1 × 1	global average pool, 1000-d fc, softmax

To enhance the model's capability further, it is initialized with pretrained weights sourced from the ImageNet database. This choice allows for the application of transfer learning techniques, which significantly accelerates the model's training process by leveraging the knowledge it has already acquired from a diverse collection of images. By doing so, the model is building upon a base of prelearned visual features while being

trained on the HiP-CT dataset.

C. Loss Function

To evaluate and train the model, we adopted the DICE loss and DICE score coefficient metrics due to their widespread acceptance and effectiveness in segmentation tasks [8].

DICE loss is a metric to quantify the disparity between the predicted segmentation and true segmentation, guiding the model toward higher accuracy by penalizing discrepancies. The formula for DICE loss is given by:

$$\text{DICE loss} = 1 - \frac{2 \cdot \sum_i (y_{\text{true},i} \cdot y_{\text{pred},i})}{\sum_i y_{\text{true},i} + \sum_i y_{\text{pred},i}} \tag{1}$$

where y_{true} and y_{pred} represent the true and predicted segmentation maps, respectively.

DICE score, on the other hand, directly evaluates the overlap between the model's predictions and true segmentation map, serving as an intuitive measure of segmentation accuracy. The DICE score is calculated as:

$$\text{DICE score} = \frac{2 \cdot \sum_i (y_{\text{true},i} \cdot y_{\text{pred},i})}{\sum_i y_{\text{true},i} + \sum_i y_{\text{pred},i}} \tag{2}$$

These metrics range from 0 to 1, with a DICE score of 0 indicating no overlap, to 1, denoting a perfect agreement between the predicted and actual segmentation. Our methodology, making use of DICE loss for model optimization and DICE score for performance evaluation, forms a robust framework to accurately train and validate the model.

VI. Results

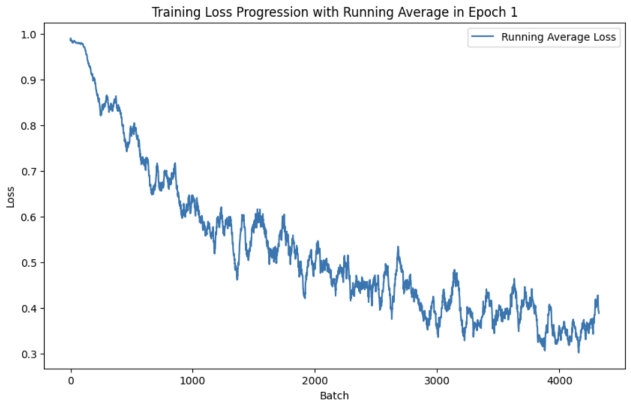


Fig. 3. Running Training Loss in Epoch 1

The training accuracy of the model converges after only 2 epochs using the hyperparamters listed in table II, after which there is negligible decrease in the DICE loss. As demonstrated in Fig. 3 and 4, the loss decreases steadily in the first epoch and begins to stagnate in the second epoch. The training was stopped early—after the second epoch—as the model began to exhibit overfitting behaviour. The final validation score achieved was 0.7490. A sample inference was performed on a kidney slice from the validation set, demonstrating the model prediction was indeed highly consistent with the ground truth.

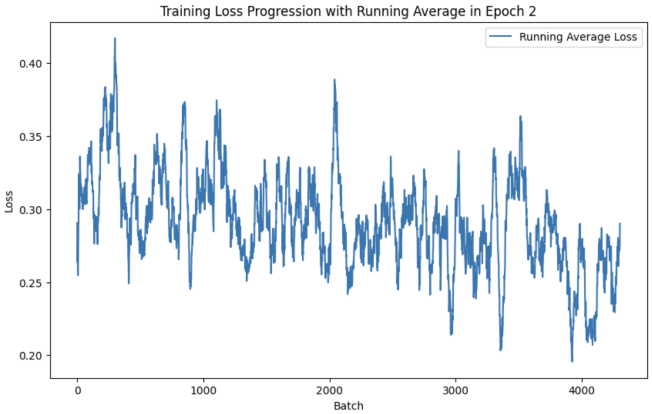


Fig. 4. Running Training Loss in Epoch 2

Fig. 5. Model Inference Against Ground Truth Mask

As shown in Fig. 5, although the model is able to pinpoint the locations of blood vessels with high accuracy, it sometimes struggles to identify the full region of the vessel, resulting in partial segmentations. This could result from a range of factors, including poor normalization techniques—causing certain areas of the image to be obscured—or the choice of the DICE loss function. It may be worthwhile to examine other loss functions used in medical segmentation literature including BCE, IoU, and focal loss, or any linear combination of these loss functions and potentially find a more holistic approach for measuring the loss for this task.

TABLE II
Final Hyperparameters for Model Training

Hyperparameter	Value
Image Input Size	1024
Learning Rate	1e-4
Batch Size	1
Number of Epochs	2
Optimizer	AdamW
Learning Rate Scheduler	OneCycleLR
Percentage Start (for OneCycleLR)	0.1
Data Augmentation	Random Scale and Rotate, Random Crop Random Gamma, Random Contrast Grid Distortion, Gaussian Blur

The training DICE Loss, validation DICE loss, and validation score, which is calculated using the DICE coefficient are shown for each training epoch. The performance of the model suggests that the SE-ResNeXt-50 is a powerful encoder for segmentation applications and furthermore demonstrates the

TABLE III
Model Performance Across Epochs

Epoch Number	Training DICE Loss	Validation DICE Loss	Validation Score
Epoch 1	0.5165	0.5576	0.4847
Epoch 2	0.2862	0.2749	0.7490

efficacy of transfer learning via pretrained ImageNet weights. This model is intended to serve as a foundational baseline; with sufficient hyperparameter tuning this model should be able to exploit a greater number of training epochs and ultimately improve its prediction performance substantially. However, due to limitations with time and computational resources, extensive exploration of optimization strategies was curtailed. In the future, it may be worthwhile to tune the learning rate scheduler to ensure a more gradual but steadier convergence. Given access to more powerful compute resources, it is also recommended to explore the possibility of using larger batch sizes to further stabilize training. Furthermore, employing more data augmentation techniques such as random noise could reduce overfitting and improve generalization.

VII. Conclusion

In conclusion, the utilization of the Feature Pyramid Network (FPN) architecture in the task of kidney blood vessel segmentation has demonstrated its efficacy and underscores the robust performance of the state of the art SE-ResNeXt50 encoder. The encoder, pretrained on ImageNet weights, facilitates effective transfer learning, thereby leveraging the knowledge gained from a diverse and comprehensive dataset from one domain to enhance feature extraction abilities specific to another domain—medical imaging. The inherent architectural advantages of FPN over UNet, namely the ability to use multi-scale spatial information efficiently, make it particularly suitable for this segmentation of complex blood vessel structures in kidney slices. The performance is further bolstered by the SE-ResNeXt50 encoder, which combines the strengths of Squeeze-and-Excitation networks and ResNeXt architectures, providing greater representational capacity and increasing the attention to relevant features.

The model's current performance serves as a solid baseline. Its accuracy and efficiency can certainly be bolstered with further fine-tuning. Sufficient hyperparameter tuning, expansion of the dataset, and novel approaches to preprocessing and augmentation could significantly enhancing the model's ability to generalize.

References

[1] "Sennet + Hoa - hacking the human vasculature in 3D," Kaggle, https://www.kaggle.com/competitions/blood-vessel-segmentation/overview.

[2] G. M. Weber, Y. Ju, and K. Börner, "Considerations for using the vasculature as a coordinate system to map all the cells in the human body," Frontiers in Cardiovascular Medicine, vol. 7, Mar. 2020. doi:10.3389/fcvm.2020.00029

[3] M. J. Willemink et al., "Preparing medical imaging data for machine learning," Radiology, vol. 295, no. 1, pp. 4–15, Apr. 2020. doi:10.1148/radiol.2020192224

[4] H. Seo et al., "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications," Medical Physics, vol. 47, no. 5, May 2020. doi:10.1002/mp.13649

[5] H. Sun et al., "Brain tumor image segmentation based on improved FPN," BMC Medical Imaging, vol. 23, no. 1, Oct. 2023. doi:10.1186/s12880-023-01131-1

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Lecture Notes in Computer Science, pp. 234–241, 2015. doi:10.1007/978-3-319-24574-4_28

[7] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-net architecture for multimodal biomedical image segmentation," Neural Networks, vol. 121, pp. 74–87, Jan. 2020. doi:10.1016/j.neunet.2019.08.025

[8] J. Ma et al., "Loss odyssey in medical image segmentation," Medical Image Analysis, vol. 71, p. 102035, Jul. 2021. doi:10.1016/j.media.2021.102035

[9] X. Hu and H. Yang, "Dru-Net: A novel u-net for biomedical image segmentation," IET Image Processing, vol. 14, no. 1, pp. 192–200, Jan. 2020. doi:10.1049/iet-ipr.2019.0025

[10] H. Huang et al., "UNet 3+: A full-scale connected unet for medical image segmentation," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020. doi:10.1109/icassp40776.2020.9053405

[11] S. Pan, X. Liu, N. Xie, and Y. Chong, EG-transunet: Enhanced and guided U-net with transformer for biomedical image segmentation, Jul. 2022. doi:10.21203/rs.3.rs-1847306/v1

[12] V. Alevizos and M. Hon, Comparison of kidney segmentation under attention U-Net Architectures, Nov. 2021. doi:10.36227/techrxiv.16546281.v2

[13] R. Gandhi and Y. Hong, "MDA-net: Multi-dimensional attention-based neural network for 3D image segmentation," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Apr. 2021. doi:10.1109/isbi48211.2021.9434168

[14] P. Benedetti, M. Femminella, and G. Reali, "Mixed-sized biomedical image segmentation based on U-Net Architectures," Applied Sciences, vol. 13, no. 1, p. 329, Dec. 2022. doi:10.3390/app13010329

[15] M. Al-masni and D.-H. Kim, CMM-Net: Contextual Multi-Scale Multi-Level Network for efficient biomedical image segmentation, Jan. 2021. doi:10.21203/rs.3.rs-135359/v1

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, May 2018. doi:10.1109/cvpr.2018.00745

[17] H. Song, Y. Wang, S. Zeng, X. Guo, and Z. Li, "OAU-net: Outlined attention U-net for Biomedical Image segmentation," Biomedical Signal Processing and Control, vol. 79, p. 104038, Jan. 2023. doi:10.1016/j.bspc.2022.104038

[18] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for Medical Image Segmentation: A review of theory and applications," IEEE Access, vol. 9, pp. 82031–82057, 2021. doi:10.1109/access.2021.3086020

[19] P. C. Sau, M. Gupta, and A. Bansal, "Blood vessel segmentation using multitask U-Net," 2022 IEEE 6th Conference on Information and Communication Technology (CICT), Nov. 2022. doi:10.1109/cict56698.2022.9998004

[20] J. Guo, W. Zeng, S. Yu, and J. Xiao, "Rau-net: U-net model based on residual and attention for kidney and kidney tumor segmentation," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Jan. 2021. doi:10.1109/iccece51280.2021.9342530

[21] X. HE, P. Shun Leung, and L. Wang, "Progressively training an enhanced U-net model for segmentation of kidney tumors," Submissions to the 2019 Kidney Tumor Segmentation Challenge: KiTS19, 2019. doi:10.24926/548719.056

[22] J. Chen and B. Jin, "A 2D U-net for automated kidney and renal tumor segmentation," Submissions to the 2019 Kidney Tumor Segmentation Challenge: KiTS19, 2019. doi:10.24926/548719.089

[23] X. Chen and C. Xu, "Multi-encoder U-Net for automatic kidney tumor segmentation," Submissions to the 2019 Kidney Tumor Segmentation Challenge: KiTS19, 2019. doi:10.24926/548719.024

[24] B. Chen, "Segmentation of CT kidney and kidney tumor by MDD-Net," Submissions to the 2019 Kidney Tumor Segmentation Challenge: KiTS19, 2019. doi:10.24926/548719.010

[25] B. Bracke and K. Brinker, "U-Net based semantic segmentation of kidney and kidney tumours of CT images," Proceedings of the 15th

International Joint Conference on Biomedical Engineering Systems and Technologies, 2022. doi:10.5220/0010770900003123

[26] U. Tatli and C. Budak, “Biomedical image segmentation with modified U-Net,” Traitement du Signal, vol. 40, no. 2, pp. 523–531, Apr. 2023. doi:10.18280/ts.400211