BIOMEDICAL IMAGE SEGMENTATION MODEL USING U-NET

1st Karthik R[1], 2nd Abhimanyu Jai Menon[2],3rd Hardik Arora[3]
[123] Computer Science and Engineering,Vellore Institute of Technology Chennai,India

Abstract – **Medical image segmentation is essential for accurate diagnosis and treatment planning. This study utilizes a UNet-based deep learning model to segment medical images, focusing on "benign" and "malignant" cases with annotated masks. By preprocessing data and optimizing the architecture, the model achieved strong performance with a mean IoU of 84 percent. Results demonstrate precise delineation of target regions, underscoring the model's potential in real-world medical applications. This work offers a foundation for future research in advanced segmentation tasks, including 3D and multimodal imaging.**

Keywords – **Medical Image Segmentation, UNet Architecture, Deep Learning, Tumor Segmentation, Convolutional Neural Networks (CNNs)**

I. INTRODUCTION

Medical image segmentation is a crucial functionality in today's health practice since it equips physicians with precise and automated procedures to identify and outline the existing structures in medical images that include tumors, organs, and other regions of interest. The accuracy in the segmentation done requires high precision in diagnosis, treatment planning, and studying the course of the disease, especially where entities like cancer, neurological disorders, and cardiovascular diseases are concerned. Nonetheless, manual segmentation is a time-consuming, subjective process and prone to human error, thus making an automated, accurate, and reliable segmentation system indispensable.

Recently, methods from the field of deep learning, such as CNNs, have proved promising in answering these questions. Among them, the UNet architecture is one of the most popular frameworks used for medical image segmentation, as it can learn hierarchical spatial features and, with its symmetric architecture, it enables effective downsampling while accurately upsampling image data.

It was originally applied in biomedical image segmentation but has been accordingly applied in many successful applications ranging from tumor segmentation in CT and MRI scans to retinal vessel segmentation in fundus images. UNet typically consists of an encoder that captures contextual information through its contracting path, followed by a symmetric decoder for the expanding path; hence, it is well suited for segmentation applications that require pixel accuracy. Notwithstanding the success of UNet, challenges remain in the context of medical image segmentation. An important challenge has to do with class imbalances, which means that the classes with healthy tissues or benign lesions may dominate the dataset, which consequently leads to biased predictions. More complex inherent properties of diversity in medical images formed by acquisition techniques variability, patient conditions variability, and image quality variability hinder generalization. Although various modifications of the UNet model including attention mechanisms and hybrid models have also been proposed, development to ensure strong performance across several datasets with high segmentation accuracy is an open field of research.

This paper finally presents the approach of an improved UNet-based medical image segmentation approach that could yield more accurate segmentations of benign and malignant tumors. We use the very well-preprocessed dataset of images, considering only such images for which the corresponding ground truth masks exist; therefore, these images have been ensured to guarantee that the data is reliable while training the model. Utilizing the benefits of the UNet architecture combined with optimized training strategies may enhance segmentation precision and generalization performance. The results of the created model are evaluated with both quantitative metrics: Dice Coefficient and Intersection over Union (IoU). The visuals imply promising results. Its contributions include the engineering of a robust deep learning model for the purpose of medical image segmentation, class imbalances mitigation, and insights into the model's performance in real-world medical contexts. Building on this work provides a basis upon which future gains in automated medical image analysis can be built, enabling potentially significant applications beyond clinical practice.

II. RELATED WORKS

Medical image segmentation is at the heart of today's medical imaging and healthcare. It allows for the precise delineation of anatomical structures, pathological regions, and functional areas, which are keys to various diagnostic and therapeutic applications, including disease detection, surgical planning, and treatment monitoring. Segmentation transforms raw medical data into clinically meaningful representations to help health-care professionals make informed decisions. Conventionally, the area of segmentation was built upon handcrafted features, techniques involving thresholding, and methods based on regions. Traditional methods suffered from some serious limitations: sensitivity to noise, inability to adapt easily to different datasets, and failure to account for complex variability in anatomy.

The emergence of deep learning has revolutionized the field, introducing a paradigm shift in how medical images are processed and analyzed. Convolutional neural networks (CNNs) have become central to this transformation, leveraging hierarchical feature extraction to capture intricate spatial and contextual details in images. CNN-based architectures, particularly U-Net and its derivatives, have demonstrated

This footnote will be used only by the Editor and Associate Editors. The edition in this area is not permitted to the authors. This footnote must not be removed while editing the manuscript.

exceptional performance in medical image segmentation, achieving state-of-the-art results across diverse imaging modalities.

The success of these models relies on foundational frameworks such as TensorFlow and Keras that provide powerful tools for deep learning system design, training, and deployment. This synthesis aims to integrate and expand on the state-of-the-art advancements in deep learning for medical image segmentation, placing emphasis on the innovative architectures, task-specific adaptations, and emerging techniques toward the ultimate goal of challenges and enhanced clinical outcomes.

A. Foundational Frameworks: Foundations of TensorFlow and Keras as Enablers

Recent years have experienced remarkable advancements in medical image segmentation, all thanks to the development of robust and scalable frameworks in machine learning. Some of the most influential tools are TensorFlow and Keras, as they have given the computational backbone for how sophisticated neural networks can be implemented and trained.[1]

B. TensorFlow: A Scalable Framework for Deep Learning

TensorFlow is a versatile open-source platform designed by Abadi et al. for large-scale machine learning and deep learning applications. The architecture of computational graphs supports developers in defining and optimizing complex neural networks with a great degree of flexibility. Such design supports parallel processing, thus making TensorFlow suitable for model training on large data sets spread across multiples CPUs or even GPUs. For medical image segmentation, the capability of handling high-dimensional and multi-modal data is very important using TensorFlow. The scalability of TensorFlow is highly beneficial to models such as U-Net, which often work with volumetric datasets like 3D MRI or CT scans. TensorFlow also includes tools for real-time visualization, debugging, and performance monitoring. For example, TensorBoard allows researchers to monitor training metrics, visualize computation graphs, and analyze model performance in terms of iteration. In medical imaging, model optimization usually involves setting the tradeoff between accuracy, computational efficiency, and interpretability, making this feature essential in practice. TensorFlow is adaptable even to edge or mobile platforms where point-of-care segmentation models can be deployed-a handheld ultrasound device or a portable diagnostic tool.[2]

C. Keras: Simplifying Model Development

While TensorFlow makes available a robust backend, Keras interfaces at a higher level of abstraction and simplifies designing and training deep learning models. Keras abstracts away low-level complexities so that researchers can prototype and iterate quickly. Modular in architecture, Keras enables easy integration of layers, loss functions, and optimizers. With such ease, it has become a preferred option for medical image segmentation. Such architectures tend to involve devising and fine-tuning encoder-decoder types like U-Net. Besides this, Keras is extensible- supports building custom layers and operations. This is a very important need in the field of medical imaging where domain-specific requirements often translate into specific tailored solutions. For instance, Keras provides the possibility of using more complex loss functions, such as Dice loss or Tversky index, specifically designed to face the problem of imbalanced datasets and small target structures in segmentation tasks.[4] TensorFlow and Keras together are the backbone of modern medical image segmentation research, providing tools essential to address some of the complex challenges and forward advance the field.

D. U-Net and Its Evolution: A Benchmark for Segmentation

With the introduction of U-Net by Ronneberger et al., medical image segmentation reached a turning point. Actually, U-Net was initially conceived for biomedical imaging. Using the encoder-decoder architecture associated with skip connections, this model captures both low-level spatial details and high-level contextual information, which makes it the gold standard for pixel-wise segmentation tasks. Applications stretch from organ segmentation up to tumor delineation.[5]

E. Standard U-Net: Design and Impact

U-Net is the encoder-decoder architecture, comprising a contracting path (encoder) and an expansive path (decoder). During the process, the encoder extracts features at different scales by reducing the spatial resolution, which gives insight into capturing high-level semantics. The decoder reconstructs the segmentation map by upsampling these features, while skip connections bridge corresponding layers of the encoder and decoder and retain information of spatial dimensions. This combination of local and global features enables high accuracy in the segmentation of small and complex structures.[6] The success of U-Net has led to many adaptations and improvements that specifically address particular challenges and even expand upon its applicability to various clinical domains.

F. Recurrent Residual U-Net: Sharpening Feature Representations

The recurrent residual U-Net designed by Alom et al. is an advancement over the basic architecture with recurring connections as well as residual learning. This contributes to a better capture of spatial dependencies and the sharpening of feature representations. Recurrent blocks enable the network to iterate over the input, capturing long-range dependencies so important for the segmentation of elongated or overlapping structures. Residual learning enables information flow between layers, alleviating the vanishing gradient and allowing deeper architectures.[7] This variant has proved particularly strong for applications involving accurate boundary definition, such as liver segmentation from CT scans or cardiac segmentation from MRI. The recurrent residual U-Net iteratively refines the feature maps to reach higher accuracy and robustness compared to standard U-Net.

G. ConvLSTM U-Net: Temporal Information Integration

Sometimes medical imaging is dynamic or sequential in nature, such as video-based endoscopy or time-series MRI. To treat these cases, Azad et al. formulated the ConvLSTM U-Net, which integrates convolutional long short-term memory units into the structure of a U-Net. These units allow the model to track temporal dependencies through a form of memory that can keep track of its previous inputs, thereby increasing the consistency of segmentations between frames.[8] ConvLSTM U-Net model has demonstrated excellent performance in applications like organ motion tracking, in which temporal coherence is highly important. For instance, for cardiac imaging, it can accurately segment heart chambers at various time points, further allowing functional analysis and disease diagnosis.[10]

H. Recalibration-Based Frequency U-Net: Towards Enhanced Texture Analysis

Azad et al. developed further extension of U-Net with a frequency domain analysis, leading to DFR-U-Net. This architecture recalibrates the frequency components of input images by relieving challenges that emerge due to texture and contrast variations. It balances low- and high-frequency information to better obtain finer segmentation results in tasks like skin lesion segmentation and retinal vessel detection.

I. Attention-Enhanced Variants: Focusing on Salient Features

Among the many tools developed for ameliorating model performance on complex segmentation tasks, attention mechanisms have emerged as particularly potent. Attention mechanisms allow a network to focus on relevant regions and suppress irrelevant features, and attention-enhanced U-Net variants obtain better accuracy and robustness compared with other baseline approaches. This is particularly helpful in those cases where target structures are small or poorly defined, such as tumour segmentation from PET scans or lesion detection from dermoscopic images.[15]

J. Applications of Deep Learning to Clinical Practice

Deep learning has helped make tremendous progress in the automatic segmentation of medical images and has found various applications across different clinical areas. These applications range from polyp detection, skin lesion analysis, all the way to nucleus segmentation.

K. Polyp Detection in Colonoscopy

Polyp segmentation in colorectal cancer screening is crucially essential as early detection drastically improves survival. Challenges were addressed by using FCNs and statistical appearance models that utilize spatial hierarchies and texture analysis to [17]identify polyps with various shapes and sizes. Such methods proved to be very effective in automating polyp detection, decreasing reliance on manual review, and improving diagnostic efficiency.

L. Skin Lesion Analysis

An accurate segmentation of skin lesions is important for melanoma and other dermatological conditions. Deep learning models that include full-resolution convolutional networks (FRCNs), as well as architectures enhanced by attention, have shown excellent performance in dermoscopic images segmentation. These models preserve spatial details and focus on boundary accuracy, allowing for early and accurate detection of malignant lesions.

M. Microscopy: Nucleus Segmentation

Nucleus segmentation is an elementary task in biological research, giving insights into cellular structures and functions. CNN-based architectures have so far been successfully applied to the task. Some of the factors that the network must contend with include noise, variability in imaging conditions, and differences in resolution.

N. Advances in Multi-Scale and Contextual Learning

The most impactful direction in deep learning-based segmentation is developing methods that capture multi-scale features and contextual information. All these innovations help in dealing with problems like overlapping structures, variable sizes of objects, and ambiguous boundaries, mainly in medical images.[20]

III. METHODOLOGY

A. Dataset Selection and Preprocessing

The BUSI dataset was selected for its rich content, including ultrasound images and segmentation masks for three classes: normal, benign, and malignant.

- Preprocessing: Images were resized to (128×128x128) x (128x128x128) pixels for consistency.
- Grayscale conversion and normalization (pixel range [0, 1]) were applied for faster convergence.
- Binary masks were created for the benign and malignant classes, excluding the normal class due to the absence of segmentation masks.

B. Data Visualization and Validation

- Sample images, masks, and overlays were visualized using Matplotlib to validate preprocessing and class distribution.
- Data imbalances were addressed by excluding irrelevant classes.

C. Data Splitting

- The dataset was divided into training (90 percent) and testing (10 percent) sets to ensure robust model training and validation. Stratified sampling maintained class balance.

D. Model Architecture

1) U-Net Design::

- Encoder: Feature extraction via convolutional layers followed by max-pooling.
- Decoder: Upsampling combined with skip connections to restore spatial details.
- Output: A single-channel mask generated by a (1×11)x(11×1) convolution layer with a sigmoid activation function.

E. Model Compilation

The model was compiled with the following settings:

- Loss Function: Binary Crossentropy for binary segmentation.
- Optimizer: Adam for adaptive learning.

- Metrics: Accuracy and Dice Coefficient to evaluate segmentation performance.

F. Training and Evaluation

- The model was trained for 100 epochs using early stopping to prevent overfitting.
- Predictions were validated using ground truth masks, with performance assessed through IoU, Precision, Recall, and F1-Score.

IV. DATASETS

The dataset applied for the biomedical segmentation of images includes ultrasounds differentiated into three classes: benign, malignant, and normal. It is well-oriented towards breast cancer detection. It aims to differentiate between benign and malignant tumors and to identify normal tissues that are free from abnormalities.

- Normal Class: The class consists of images with no existing tumors or abnormalities. These images are very important to help identify the healthy tissues from possible problematic areas in medical imaging. More importantly, the normal class does not have its corresponding segmentation masks that in return makes it inappropriate for direct segmentation tasks; however, is appropriate for classification purposes.

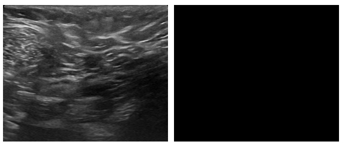


Fig. 1. Image and its mask for normal class

- Benign Class: Images under this class contain tumors that are non-cancerous. Benign lesions are less threatening as compared to carcinomas but still need to be identified correctly for proper medical diagnosis and treatment. The dataset contains masks for the segmentation of benign tumors, which highlight regions of interest in the images. These masks help the models learn how to delineate a benign growth correctly from the surrounding tissues.

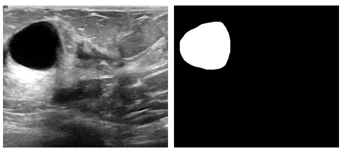


Fig. 2. Image and its mask for benign class

- Malignant Class: Malignant images represent cancerous tumors, which are critical for early detection and intervention. As with the benign class, segmentation masks are provided, marking the malignant areas. These masks help the model to learn the boundaries of cancerous growths, allowing it to distinguish between malignant and non-malignant regions in unseen images. Malignant tumors are a priority for detection in clinical practice, as early detection significantly increases the chances of successful treatment.

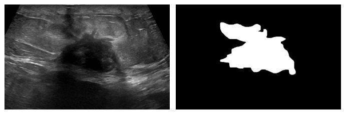


Fig. 3. Image and its mask for Malignant class

A. Data Characteristics

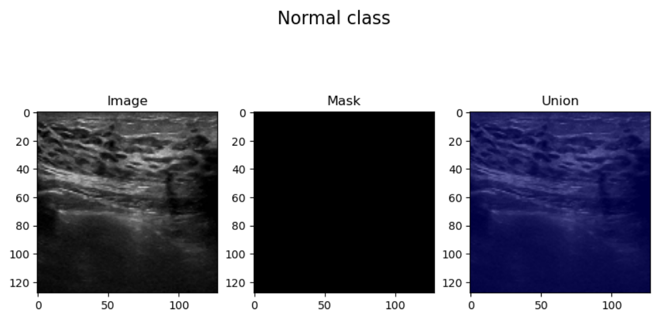
The dataset consists mainly of grayscale ultrasound images, with average resolution for visual inspection and training deep learning models. Images are annotated with masks at pixel-level accuracy indicating areas of abnormality, such as benign and malignant tumors. The respective pair comprising an image and the pixel mask is part of the ground truth from the segmentation process.

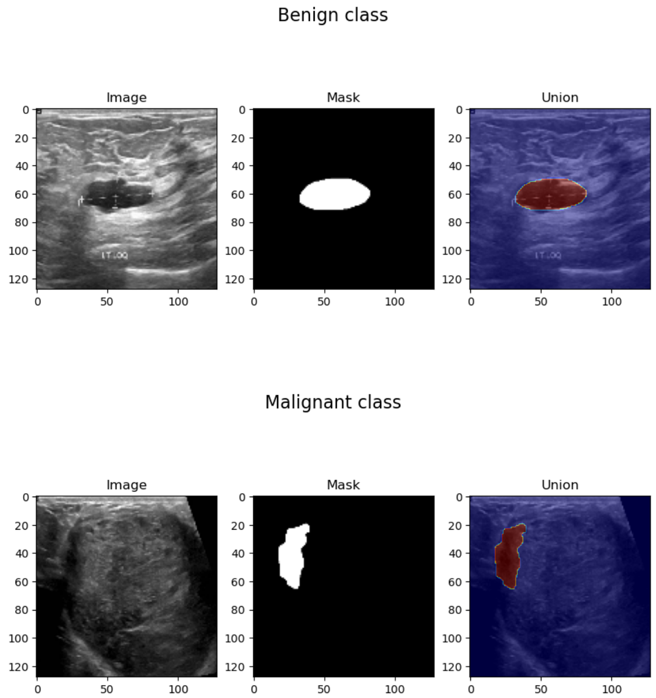
The masks themselves are binary images, where pixel values can be either 0 or 1, indicating the background and tumor areas, respectively. The importance of these binary masks is reflected during training of the model since these masks help the model learn the segmentation boundaries between the tumor and healthy tissue. Once applied to new, unseen data, the model then predicts these binary masks.

B. Preprocessing Steps

The preprocessing of the dataset often involves resizing images to a standard size, such as 128x128 or 256x256 pixels. It ensures consistency in model input. Images can also go through normalization by scaling their pixel values within the range [0, 1], which further helps models converge faster and generalize better. Data augmentation techniques like rotation, flipping, and zooming can be applied to increase artificially the diversity of the dataset, though care is taken not to distort the clinical features of the tumors.

The dataset is divided into training and testing subsets to ensure that the model is evaluated on data it has not encountered during training, thus assessing generalization capabilities. In general, the majority of the data is used for training (e.g., 80-90 percent), the remaining part is left for validation and testing for estimating performance.





C. Challenges and Importance

One challenge in this dataset is the inherent class imbalance of benign, malignant, and normal classes. The class with all the masks is least useful for the task of segmenting everything, and by implication, there are potentially less malignant samples than benign samples. Data balancing should be appropriately carried out, preferably through stratified sampling or weighting in the loss function, to balance the class imbalance to ensure that the model learns to spot malignant tumors highly accurately under such a skewed distribution.

This dataset is the first of its kind in the medical image setting to use such a technique to aid early detection of breast cancer by developing and refining algorithms for segmentation. Distinguishing and classifying between benign and malignant lesions by doing an accurate segmentation makes a very significant step toward automated diagnostic tools that can assist radiologists to make quicker and more reliable decisions.

V. RESULTS

In Fig. 4, the results obtained by the proposed segmentation model are presented. The given model's performance toward categorizing biomedical ultrasound images into three primary classes namely benign, malignant, and normal regions is evaluated. This figure shows examples of input ultrasound images, their corresponding ground truth masks, as well as the segmentation outputs that the model predicts.

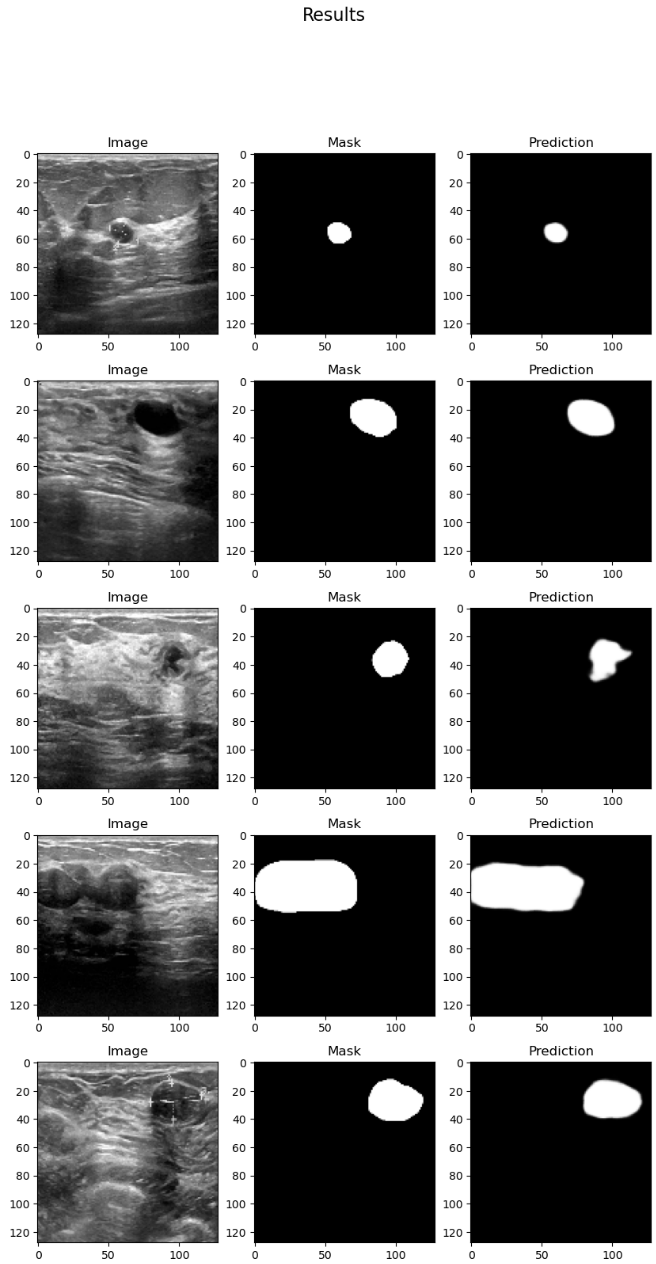


Fig. 4. Images, their masks and predictions

The proposed model for biomedical image segmentation is based on a U-Net architecture. This architecture is among the widely adopted deep learning models for medical imaging tasks due to its capability to perform accurate localization and segmentation. In Fig. 5 the architecture of the model is shown.

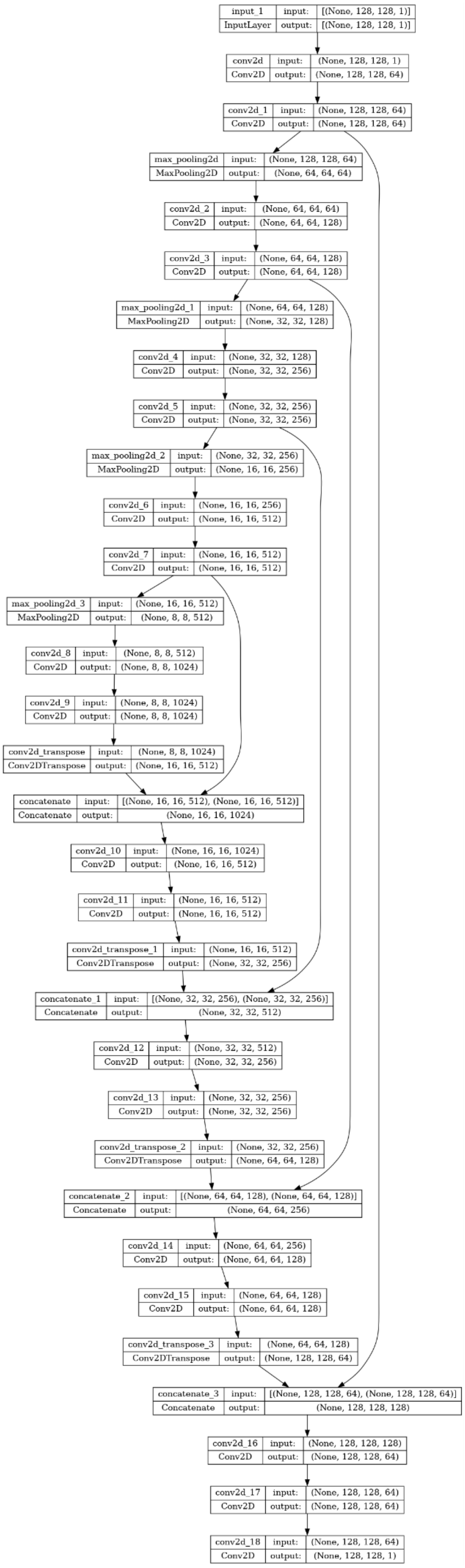


Fig. 5. U-Net Model

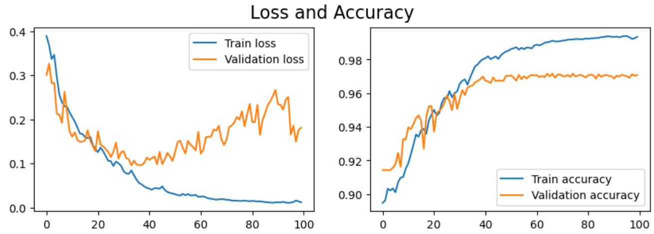


Fig. 6. Loss and Accuracy

VI. CONCLUSIONS

Developments of CNN-based architectures, including U-Net and its relatives, have dramatically progressed medical image segmentation to the point where diagnostic and clinical workflows are in transformation. The nature of these models, specifically at multi-scale analyses, contextual learning, and task-specific adaptations, greatly enhances the accuracy, efficiency, and the automation possibilities of segmentation tasks. Grounding here is laid by U-Net's encoder-decoder structure using skip connections, but variants in the form of recurrent, residual, and attention-enhanced U-Net further refine the performance for diverse clinical applications ranging from organ delineation to tumor detection.

Developments like the use of atrous convolutions for multi-scale feature extraction and gating mechanisms for contextual learning have improved segmentation model effectiveness in understanding complex anatomical structures along with ambiguous boundaries. Additional accuracy is achieved through attention mechanisms and frequency domain analysis, which concentrate on interest regions and textural differences as slight as that can be.

As these technologies mature, integration into clinical practice has the potential to revolutionize diagnostics by automating image analysis to reduce human error and increase diagnostic consistency, improving patient outcomes in general. However, this transformation calls for continued collaboration across academia, industry, and healthcare to make these models robust, interpretable, and effective for use in real-world scenarios. The medical image segmentation technology holds bright promise and much potential for transforming healthcare delivery and will continue to do so.

REFERENCES

[1] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M (2016) TensorFlow: a system for Large-Scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16), pp 265–283

[2] Akbari M, Mohrekesh M, Nasr-Esfahani E, Soroushmehr SR, Karimi N, Samavi S, Najarian K (2018, July) Polyp segmentation in colonoscopy images using fully convolutional network. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp 69–72

[3] Al-Masni MA, Al-Antari MA, Choi MT, Han SM, Kim TS (2018) Skin lesion segmentation

in dermoscopy images via deep full resolution convolutional networks. Comput Methods Prog Biomed 162:221–231

[4] Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK (2019) Recurrent residual U-Net for medical image segmentation. J Med Imaging 6(1):014006

[5] Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK (2018) Medical image analysis using convolutional neural networks: a review. J Med Syst 42(11):1–13

[6] Asadi-Aghbolaghi M, Azad R, Fathy M, Escalera S (2020) Multi-level context gating of embedded collective knowledge for medical image segmentation. arXiv preprint arXiv:2003.05056

[7] Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S (2019) Bi-directional ConvLSTM U-Net with densley connected convolutions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp 0–0

[8] Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S (2020, August) Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation. In: European Conference on Computer Vision. Springer, Cham, pp 251–266

[9] Azad R, Bozorgpour A, Asadi-Aghbolaghi M, Merhof D, Escalera S (2021) Deep frequency re-calibration U-Net for medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 3274–3283

[10] Bernal J, Sánchez J, Vilarino F (2012) Towards automatic polyp detection with a polyp appearance model. Pattern Recogn 45(9):3166–3182

[11] Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F (2015) WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Comput Med Imaging Graph 43:99–111. https://doi.org/10.1016/j.compmedimag.2015.02.007

[12] Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, Heng C, Becker T, Doan M, McQuin C, Rohban M (2019) Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nat Methods 16(12):1247–1253 Available: https://www.kaggle.com/c/data-scien ce-bowl-2018

[13] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848

[14] Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D (2018) DRINet for medical image segmentation. IEEE Trans Med Imaging 37(11):2453–2462

[15] Chen LC, Zhu Y, Papandreou G, Schrof F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp 801–818

[16] Chen LC, Papandreou G, Schrof F, Adam H (2019) Rethinking atrous convolution for semantic image segmentation. arXiv 2017. arXiv preprint arXiv:1706.05587

[17] Chollet F (2018) Keras: The python deep learning library. Astrophysics source code library ascl-1806

[18] Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, Kalloo A, Liopyris K, Mishra N, Kittler H, Halpern A (2018, April) Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th international symposium on biomedical imaging

[19] Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, Helba B, Kalloo A, Liopyris K, Marchetti M, Kittler H (2019) Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368

[20] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2020) Imagenet: A large-scale hierarchical image database, 2009. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 248–255