VU-NET: AN EXPLAINABLE AI APPROACH FOR LIVER SEGMENTATION

Ramakrishnananda
Department of Computer Science and Engineering
Tezpur University - 784028
Tezpur, Assam, India
rknanda.0111@gmail.com

Pranjal Singh Katiyar
Department of Computer Science and Engineering
Tezpur University - 784028
Tezpur, Assam, India
pranjalsinghkatiyar@gmail.com

Rosy Sarmah
Department of Computer Science and Engineering
Tezpur University - 784028
Tezpur, Assam, India
rosy8@tezu.ernet.in

***Abstract*—Medical image segmentation, particularly in tumor identification, is a critical aspect of modern healthcare. In this study, we propose our own VU-Net model to address liver tumor segmentation, utilizing the LiTS17 dataset and GradCAM visualization to provide insights into the model's decision-making. Our contributions include a fine-tuned VU-Net model, extensive performance evaluation, and application of Grad-CAM for interpretability. These findings enhance understanding of deep learning models in medical imaging, promising advancements in tumor segmentation methodologies and contributing to improved patient care.**

***Index Terms*—Grad-CAM, U-Net, V-Net, Semantic Segmentation, Explainable AI**

I. Introduction

Deep learning [1], a powerful subset of machine learning, has emerged as a transformative force in the realm of artificial intelligence. At its core, deep learning revolves around the concept of neural networks, inspired by the intricate structure and functioning of the human brain. Unlike traditional machine learning algorithms, deep learning excels at handling vast amounts of unstructured data by automatically learning hierarchical representations through multiple layers of neural networks [2]. Images, with their rich and diverse content, encapsulate a wealth of information, from the subtle nuances of facial expressions to the intricate structures of celestial bodies. Harnessing the power of deep learning, researchers and practitioners have unlocked new avenues in image recognition, classification, and generation. Through convolutional neural networks (CNNs) [3] [4], deep learning algorithms can parse through massive image datasets with remarkable efficiency, extracting meaningful features and patterns with unparalleled accuracy. Within the realm of image analysis, segmentation stands as a pivotal task, aiming to partition images into semantically meaningful regions. Traditional segmentation methods often falter in complex scenarios, where intricate boundaries and diverse textures abound [6]. In this scenario, deep learning emerges as a beacon of hope, offering robust solutions for semantic and instance segmentation [7]. By leveraging architecture like U-Net, deep learning models excel in delineating objects and delineating boundaries with remarkable precision.

The application of segmentation techniques in medical imaging holds immense promise, revolutionizing diagnostics, treatment planning, and disease monitoring. From identifying tumours in MRI (Magnetic Resonance Imaging) scans to delineating organs in CT (Computed Tomography) images, from Ultrasounds to X-Rays, deep learning-based segmentation empowers healthcare professionals with invaluable insights, facilitating timely interventions and personalized treatments. By automating tedious tasks and augmenting human expertise, medical image segmentation not only enhances diagnostic accuracy but also expedites workflows, ultimately improving patient outcomes. Furthermore, the benefits of leveraging deep learning for medical image segmentation extend beyond mere efficiency gains. It enables the development of predictive models, paving the way for early disease detection and prognostic assessments. The U-Net and it variations are effective for small-scale medical image data but struggles with 2D segmentation's limited inter-slice information [8]. To address this, 3D images were utilized in the VNet network [9], employing 3D convolution operations for spatial information. While these models yield accurate results, challenges in data acquisition hinder dataset scalability. The goal is to estimate data sample distribution and generate new samples, necessitating further exploration to enhance U-Net's segmentation accuracy. On basis of this we propose our own architecture. The proposed architecture, VU-Net modifies the V-Net framework to accommodate 2D input images, enhancing its compatibility with our specific multi-class segmentation tasks while maintaining the core strengths of V-Net. Key to this adaptation is the final output layer, which incorporates three filters and employs a softmax activation function to accurately map input features to distinct classes. Drawing inspiration from seminal research, the architecture includes a connecting pathway designed to capture detailed contextual information, thus improving pattern recognition capabilities. Additionally, the decoder segment from the established U-Net architecture is integrated, enhancing the model's ability to reconstruct fine details from high-level features. This integration leverages U-Net's proven effectiveness in image segmentation, resulting in a robust and technically refined framework that excels in multi-class segmentation for image analysis.

II. Background

Computer vision is a multidisciplinary field aimed at enabling computers to interpret visual data, akin to human visual perception. It harnesses artificial intelligence and machine learning to analyze images or videos, performing tasks such as object identification, face recognition, and anomaly detection. Medical Image Segmentation, a subset of computer vision, involves partitioning medical images into distinct segments for detailed analysis. Recent advances in deep learning, particularly convolutional neural networks, have revolutionized computer vision, simplifying tasks like medical image classification and segmentation, thus advancing medical applications. Convolutional Neural Networks (CNNs) [10] have revolutionized deep learning and computer vision, particularly in image processing tasks. They leverage convolutional layers, pooling layers, and fully connected layers to extract hierarchical features from input images, forming the backbone of various computer vision applications. Convolutional layers use filters to capture spatial hierarchies and patterns, performing element-wise multiplications and summations across the input image to generate feature maps. This convolution operation is fundamental in CNNs, extracting features and patterns crucial for image processing.

Medical image segmentation is vital for diagnostic and treatment planning, where accurate delineation of anatomical structures is essential. Convolutional Neural Networks (CNNs) have revolutionized this field by autonomously learning from raw pixel data, enhancing segmentation accuracy. They excel in discerning intricate patterns in medical images, enabling both semantic and instance-level segmentation tasks. CNN-based segmentation models have demonstrated exceptional performance in tumour detection [37] [11], organ segmentation [12], and lesion localization [13], improving diagnostic accuracy and treatment planning [14]. Their adaptability to diverse imaging modalities and segmentation challenges has propelled innovations in personalized medicine [15], ultimately benefiting patient care and clinical outcomes [16]. A more detailed study on deep learning in medical images may be found in the survey [17].

In our research, we selected the U-Net segmentation model [18] for its renowned accuracy and adaptability across diverse imaging datasets. Its distinctive encoder-decoder architecture effectively preserves spatial details, driving advancements in medical imaging and environmental analysis. Notably, U-Net's efficacy extends to brain tumor segmentation and cardiovascular image analysis [19]–[21], enhancing disease detection and patient care. While U-Net excels with small-scale medical image datasets, challenges arise in 2D segmentation tasks due to limited inter-slice information. Recent work [9] introduces a novel strategy for 3D image segmentation, addressing some limitations, yet scalability remains an issue. Further investigation is needed to enhance U-Net's segmentation accuracy. Our proposed model is based on VNet [26] which comprises an encoder with multiple stages maintaining the same resolution and a decoder for gradual decompression, producing an output image of the original size. Inheriting U-Net's jump connection, VNet mitigates information loss during feature extraction. Additionally, it adopts the short-circuit connection mechanism from ResNet [13], adding input and output at each stage to learn the residual function. The authors in [9] demonstrates an application of VNet in medical imaging, specifically in 3D image segmentation utilizing a volumetric, fully convolutional neural network. The model is trained end-to-end on MRI volumes of the prostate, enabling it to predict segmentation for the entire volume simultaneously.

From our limited survey, we can understand that the integration of Convolutional Neural Networks (CNNs) into segmentation models represents a paradigm shift in image analysis, offering unprecedented capabilities for precise and efficient segmentation tasks. The versatility and adaptability of CNN architectures, coupled with the sophistication of segmentation models like U-Net and VNet have propelled advancements in medical imaging. By harnessing the power of CNNs, researchers and practitioners can achieve superior segmentation accuracy, enabling deeper insights and more informed decision-making. As we continue to explore and innovate in this field, the synergy between segmentation models and CNNs promises to unlock new possibilities and drive transformative changes in image analysis and beyond. While various methods exist, trained models may still not consistently produce desired outputs for all input data. Most Deep learning based segmentation methods operate as black boxes, lacking explanations for their predictions. Explainable deep learning techniques aim to clarify model decision-making through visualization and interpretation. These methodologies are often denoted as interpretable deep learning or explainable artificial intelligence (XAI) [30]. In medical imaging, researchers are progressively adopting XAI techniques to elucidate algorithmic outcomes. An explanation is deemed effective if it provides understanding of the neural network's decision-making process or renders the decision comprehensible. A survey on XAI used in deep learning-based medical image analysis can be found in [31]. In the survey, XAI techniques used in medical image analysis have been categorized into three types: visual, textual, and example-based. Gradient-weighted class activation mapping (Grad-CAM), introduced by Selvaraju et al. [32], extends the concept of CAM. Unlike CAM, Grad-CAM can be applied to any CNN architecture to generate post hoc local explanations without the need for global average pooling. Additionally, the authors introduced guided Grad-CAM, which combines guided backpropagation with Grad-CAM through element-wise multiplication. Both Grad-CAM and guided Grad-CAM have found applications in medical image analysis. In [34], Ji utilized Grad-CAM to indicate the areas on histology lymph node sections where a classifier based its decision on metastatic tissue presence. Grad-CAM was also employed in [35] to identify small bowel enteropathies on histology. Gunashekar et al [35] employed an interpretable deep learning model to analyze a convolutional neural network's (CNN) predictions for segmenting prostate tumours. The CNN, based on a U-Net architecture trained on multi-parametric MRI data, automati-

cally segments the prostate gland and tumours. To interpret the CNN's segmentation, the authors generated heatmaps using Grad-CAM. The work in [36] utilized Grad-CAM to highlight the areas on brain MRI images that influenced the classifier's decision regarding tumour presence. In medical image segmentation, Grad-CAM serves to elucidate which regions of an input medical image contribute significantly to the segmentation outcome. By analyzing gradients flowing into the final convolutional layers of the segmentation network, Grad-CAM generates heatmaps pinpointing areas the network focuses on for segmentation decisions. For instance, in tumour segmentation from MRI or CT scans, Grad-CAM highlights regions indicative of tumour presence, aiding clinicians in understanding the model's segmentation decisions.

A. Motivation and Contributions

Liver tumour segmentation plays a pivotal role in early detection, diagnosis, and treatment planning in clinical settings. Despite advancements in medical imaging, accurate segmentation remains challenging due to tumour variability and complex anatomical structures. Traditional methods often struggle with accuracy in such scenarios. This underscores the need for advanced AI approaches. However, the lack of transparency in AI models impedes their clinical acceptance. Explainable AI techniques have emerged to enhance the transparency, interpretability, and explainability of opaque AI methods. A segmentation method that is explainable will likely be more trusted by experts. VU-Net, an explainable AI model adapted from V-Net architecture for 2D liver tumour segmentation, addresses this gap. By providing insights into model decisions, it enhances trust among healthcare professionals. Comparative analysis with U-Net allows for a comprehensive evaluation, highlighting VU-Net's advantages in accuracy, efficiency, and interpretability. Furthermore, exploring interpretability through techniques like Grad-CAM ensures reliability validation and integration into clinical workflows, fostering advancements in medical image analysis and healthcare AI. The location of the tumours can be better visualized from CT images using Gradient-weighted Class Activation Mappings (GradCAMs) based visualization tool.

The contributions of our work can be summarized as follows:

1) A Fine-tuned light weight VU-Net model having state-of-the-art performance.
2) Exhaustive performance comparison on a single dataset for segmentation of cancerous and healthy liver.
3) Multilabel segmentation of liver and tumour.
4) Application of Grad-CAM to VU-Net and U-Net models for enhanced interpretability in liver tumour segmentation.
5) Exploration of Grad-CAM's utility in providing insights into model decision-making, advancing the field of interpretability in medical image analysis.
6) Expansion of the research frontier by proposing a novel model architecture and methodology along with explainable AI in medical imaging, paving the way for future innovations in the field.

III. Methods

In this section, we introduce VU-Net, a new architecture designed for liver tumour segmentation, and compare its performance with U-Net. We also employ Grad-CAM for visual insights into model decisions. We detail VU-Net's architecture, training procedures, and comparative evaluation setup. Grad-CAM enhances interpretability by visualizing key regions influencing segmentation decisions. Our approach aims to advance liver tumour segmentation while ensuring transparency and reliability in AI-driven medical image analysis.

A. U-Net

The U-Net architecture, introduced in [18], is a convolutional neural network (CNN) optimized for semantic segmentation tasks, particularly in biomedical image analysis. It features a contracting encoder path with convolutional blocks—each containing two 3x3 convolutions followed by batch normalization and ReLU activation—alternating with max-pooling layers for spatial reduction. This is followed by a bottleneck that captures the densest input representation through similar convolutional arrangements. The expansive decoder path then upsamples these feature maps using transposed convolutions, also supplemented by batch normalization and ReLU activation. Skip connections from the encoder are concatenated with the decoder to retain spatial information lost during downsampling. The architecture ends with a 1x1 convolutional layer that applies sigmoid or softmax activation for binary or multiclass segmentation, respectively. Designed for images of arbitrary size, U-Net processes 128x128 images in this instance to output a segmentation map classifying each pixel according to the task. U-Net is widely regarded as the most popular image segmentation architecture due to its adaptability, modular design, and proven effectiveness across various medical imaging modalities. The authors in [22] provide a thorough review of medical image segmentation technologies, particularly U-Net variants, discussing their architecture, innovations, and efficiency. They categorize and evaluate relevant methodologies and introduce commonly used loss functions, evaluation metrics, and modules for medical imaging segmentation, offering valuable references for future research. Several improved U-Net architectures have recently been proposed for medical imaging segmentation [23], [24], [25].

B. VU-Net

The VU-Net architectural framework builds upon the robust foundation of the V-Net [26], albeit with tailored adjustments to seamlessly accommodate 2D input images, a departure from the native 3D input characteristic of V-Net. This modification ensures compatibility with the specific requirements of our segmentation task while retaining the inherent strengths of the V-Net architecture.In particular, the final layer of our architecture, serving as the pivotal output layer, is meticulously crafted

to wield three filters. This strategic choice aligns with the exigencies of our multi-class segmentation objective. To facilitate the intricate mapping of input features to distinct classes, the softmax activation function is judiciously employed in this concluding layer. The architectural prowess is further enriched through an intelligently designed connecting pathway, drawing inspiration from an influential and meticulously referenced research paper. This pathway plays a pivotal role in capturing intricate contextual information across the input data, thereby enhancing the model's ability to discern subtle patterns and features. Moreover, the decoder segment of our architecture is seamlessly integrated from the well-established U-Net architecture. This deliberate adoption contributes significantly to the holistic structure of our model, leveraging the U-Net's proven efficacy in image segmentation tasks. The decoder's role in progressively reconstructing intricate details from high-level features obtained in earlier stages complements the overall segmentation process. In essence, the amalgamation of V-Net's foundational strength, a tailored adjustment for 2D inputs, a sophisticated connecting pathway, and the incorporation of U-Net's decoder segment results in a comprehensive and technically sound architecture. The careful selection of design elements ensures that our model is well-equipped to handle the intricacies of multi-class segmentation tasks, making it a robust and versatile tool for image analysis. Below is the Overview of the architecture.

Encoding Pathway: The network comprises five convolution blocks, each structured with 2 convolution kernels, normalization, and a ReLU activation function, all using 32 in sequence. Convolution kernels have 16, 32, 64, 128, 256 filters with a stride of 1 to extract distinctive features.

Bottleneck: The V-Net's bottleneck distills essential features from the encoding pathway for precise segmentation. Comprising densely connected convolutional layers with optional dilation, it captures intricate patterns and mitigates the vanishing gradient issue. Positioned between encoding and decoding, it refines high-level representations.The architecture incorporates a dual-layer convolutional block, complemented by Batch Normalization and Rectified Linear Unit (ReLU) activation functions, enhancing feature extraction and promoting non-linearity within the model.

Decoding Pathway: The decoder path involves upsampling feature maps. Each block consists of two 3x3 convolutions, batch normalization which is done during concatenation, and ReLU activation. Upsampling is done through transposed convolutions.

Skip Connections: Connecting pathways include ReLU activation, pooling module with a 5x5 filter (16 to 256 filters), ReLU activation, and another convolution layer. Outputs are combined for the first and second modules, producing the final output.

Output Layer/Final Layer: The decoding pathway ends with a convolutional layer for mapping deep features to the desired classes (as in segmentation masks) and a softmax function for multiclass semantic segmentation.

Input/Output The architecture is designed to process medical images of various sizes, with a focus on images measuring 128x128 pixels. It accommodates varying channel counts based on specific application needs. The output from the system is a segmented image where each pixel is accurately classified into distinct categories, reflecting the different classes present in the dataset. This functionality allows for precise segmentation crucial for detailed medical image analysis.

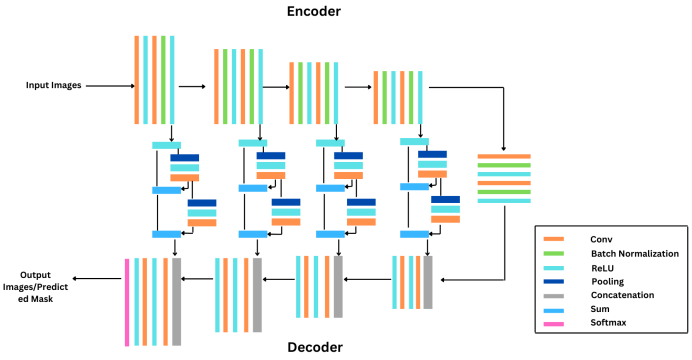

Fig. 1. VU-Net Architecture

IV. Results

The study and analysis were conducted on the dataset which will be mentioned below, utilizing various both segmentation models. The experiment followed these steps:

- Thorough examination and pre-processing of each selected dataset to align with model-specific input requirements and optimize training efficiency.
- Application and training of different Segmentation models with the implementation of early stopping to prevent overfitting.
- Calculation of various performance metrics on both training and testing data.
- Visualization of mask images generated for specific testing dataset images.
- Generation of informative graphs depicting performance metrics during training epochs determined by the early stopping procedure.
- Generation of heatmaps for each class using Grad-CAM
- Supemimposing the heatmaps on the original input image.

Given the abundance of metrics and data, only the most relevant and informative ones are presented. We selected key hyperparameters: a 0.001 learning rate, Adam optimizer, softmax for multiclass segmentation, ReLU for intermediate layers, and a batch size of 16. Experiments were conducted on Google Colaboratory (Colab), utilizing its free GPU resources to accelerate training and testing, enabling efficient handling of large datasets and rapid research prototyping.

A. Dataset Preprocessing

The LiTS17 dataset [27], sourced from the 2017 Liver Tumour Segmentation Challenge, comprises 130 CT scan images formatted in NII standard. The dataset's 512 x 512 pixel

resolution images are categorized into Background, Liver, and Tumour, totaling 49.9 GB in size. The initial dataset of 130 NII format images was converted to PNG format using the nilabel Python library, resulting in 58,638 individual PNG images. Among these, 39,475 lacked liver representation, while 19,163 featured liver and/or liver tumor. From the latter subset, 1,916 test images were randomly selected, while the remaining 17,247 images constituted the training set, with 10% allocated for validation. Each 512x512 PNG image was downscaled to 128x128 for computational efficiency without compromising image quality.

B. Performance Evaluation

Performance metrics in deep learning and machine learning measure model effectiveness, guiding optimization for robust algorithms. They offer insights into a model's behavior, aiding in assessing strengths and weaknesses for a given task. For Image segmentation the performance metrices that we will be using following metrices:

Confusion Matrix [28] is crucial for evaluating model performance by comparing predictions with actual outcomes in both binary and multiclass scenarios. It breaks down predictions into True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), representing correct and incorrect predictions. In image segmentation, these elements correspond to accurately identified targets, erroneously highlighted non-targets, correctly identified backgrounds, and missed targets, respectively.

Accuracy [28] (calculated as given in Table I) assesses the overall correctness of model predictions in machine learning and segmentation tasks and offers a comprehensive measure of segmentation quality. In image segmentation, it gauges the model's ability to classify pixels accurately, providing essential insights into its effectiveness.

Precision [28] assesses the accuracy of positive predictions by dividing true positives by the sum of true positives and false positives (Table I. High precision indicates reliable positive predictions, reducing false alarms and improving the model's effectiveness in identifying positive cases.

Recall [28], or sensitivity, gauges a model's ability to detect all relevant class instances (Table I). It's vital where missing positives carry significant costs, ensuring comprehensive coverage and accuracy in classification.

IoU (Intersection over Union) [29] assesses segmentation accuracy by comparing predicted and ground truth regions' overlap. A high IoU indicates precise region delineation, crucial for tasks requiring accurate spatial alignment.

Dice Coefficient [13] evaluates segmentation accuracy by comparing predicted and ground truth regions, balancing precision and recall. A high score indicates precise segmentation while minimizing errors.

Categorical Cross Entropy [1] Categorical cross-entropy is vital for training neural networks in multi-class classification, optimizing model parameters to accurately estimate probabilities and handle imbalanced class distributions.

TABLE I
PERFORMANCE METRICS

Metric	Formulae
Accuracy	$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP + TN + FP + FN}}$
Recall	$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}}$
Precision	$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}}$
IoU	$\text{IoU} = \frac{\text{TP}}{\text{TP + FP + FN}}$
Dice Coefficient	$Dice = \frac{2\times TP}{2\times TP+FP+FN}$
Categorical Cross Entr.	$H(y,p) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{i,j}\log(p_{i,j})$

TABLE II
TEST RESULTS FOR LIVER DATASET USING U-NET AND VU-NET ARCHITECTURES

Metric	**U-Net**	**VU-Net**
Accuracy	0.9974	0.9934
Recall	0.9960	0.9895
Precision	0.9962	0.9908
IoU	0.9062	0.7964
Dice Coefficient	0.9961	0.9901
Categorical Cross Entr.	0.0096	0.0428

In the evaluation of the liver dataset, the VU-Net architecture exhibits commendable performance compared U-Net. A specific benchmark is drawn against the U-Net model, the VU-Net architecture demonstrates competitive results with only slight deviations in key performance metrics. Notably, the differences in accuracy (acc), recall, precision, and the dice coefficient are marginal, with difference values of 0.004, 0.0065, 0.0054, and 0.006, respectively. However, the intersection over union (IoU) exhibits a more pronounced difference, registering at 0.10 compared to U-Net.

C. Grad-CAM based Visualization

Class Activation Mapping (CAM) [33] enhances model interpretability, localizes objects, and identifies discriminative features. GradCAM generates heatmaps, aiding in diagnosis by highlighting regions of interest. These visualizations, including intermediate activation maps, mitigate the opacity of ML and CNNs, providing valuable insights. GradCAM clarifies model distinctions with its effective differentiation of outcomes.

Based on the GradCAM visualizations of both models from Fig. 2 and Fig 3, noteworthy distinctions emerge in their approach to identifying liver and tumor regions. VU-Net form Fig. 3 demonstrates a refined focus on the liver area, effectively capturing its boundaries while occasionally considering features beyond it. Conversely, U-Net from Fig. 2 exhibits a more constrained focus solely on the liver, omitting extraneous features. Similarly, in tumor identification, VU-Net displays commendable precision in targeting tumor regions akin to U-Net. These observations underscore the nuanced differences in segmentation strategies between the two models,

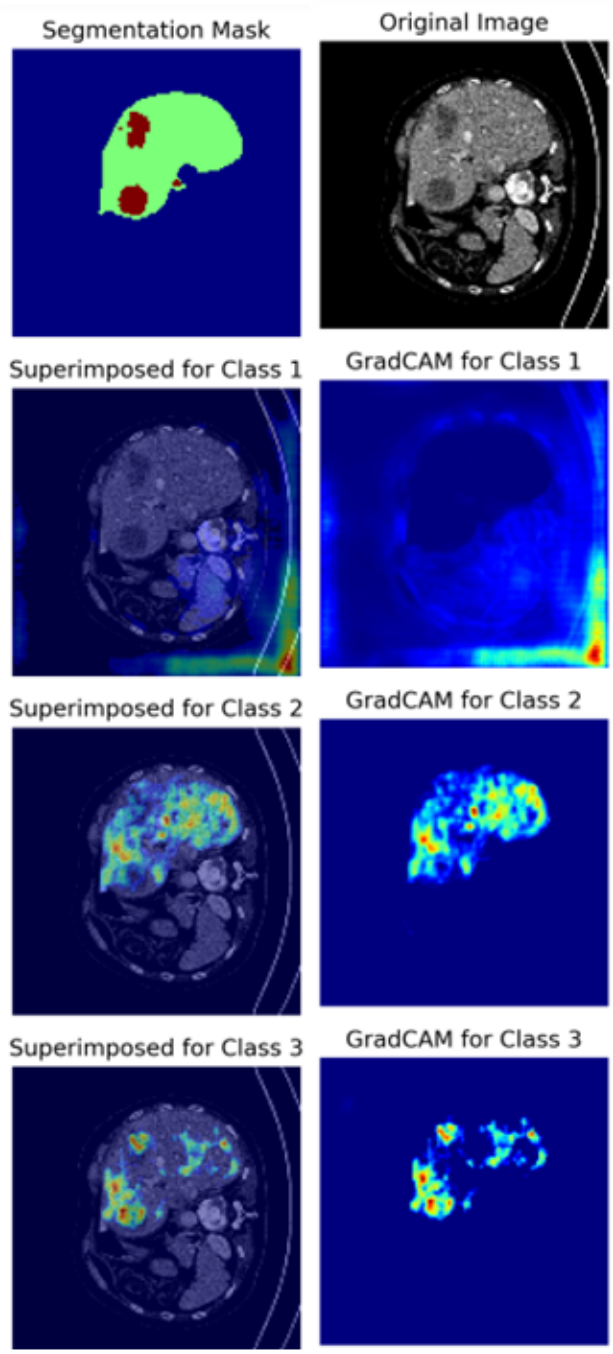

Fig. 2. U-Net Segmentation Results

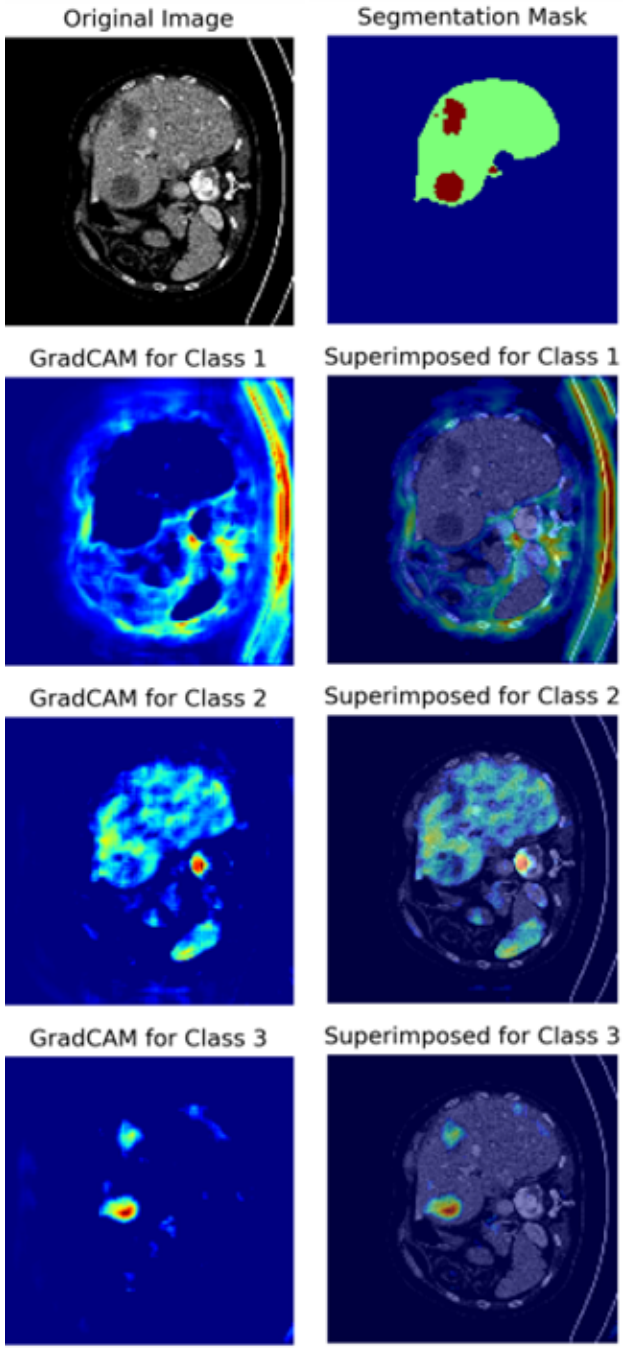

Fig. 3. VU-Net Segmentation Results

shedding light on their respective strengths and areas for refinement.

Conclusion

This paper presents a comprehensive evaluation of liver tumor segmentation utilizing the LiTs17 dataset. Leveraging the novel VU-Net architecture alongside U-Net, we observed commendable performance metrics, with VU-Net demonstrating results, particularly evident in the Accuracy, Precision and Recall which were at par with UNet. GradCAM visualization further enhanced interpretability, elucidating nuanced differences in segmentation strategies between the two models. Notably, VU-Net exhibited a refined focus on liver and tumor regions, shedding light on areas for refinement. Moving forward, there is ample scope for future improvements in refining the proposed architecture focusing on increasing the IoU score and leveraging Grad-CAM to advance the field of medical image analysis and explainable AI.

REFERENCES

[1] I. Goodfellow, Y. Bengio, & A. Courville (2016). Deep Learning. MIT Press

[2] W. McCulloch, & W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5 , 115–133.

[3] P. Mahapattanakul (2019). From human vision to computer vision - convolutional neural network (part 3/4).

[4] K. Fukushima (2007). Neocognitron. Scholarpedia, 2 (1), 1717.

[5] R. Nicole (in press). Title of paper with only first word capitalized. J. Name Stand. Abbrev.

[6] A. Rezaei, & F. Asadi (2024). Systematic review of image segmentation using complex networks.

[7] J. Long, E. Shelhamer, & T. Darrell (2015). Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[8] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, & O. Ronneberger (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

[9] F. Milletari, N. Navab, & S. -A. Ahmadi (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, pp. 565-571, doi: 10.1109/3DV.2016.79.

[10] A. Krizhevsky, I. Sutskever, & G. Hinton (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (NIPS).

[11] K. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, & H. Larochelle (2017). Deep Learning for Brain tumour Segmentation: A Comprehensive Review. Medical Image Analysis, 36, 231-248.

[12] G. Wang, W. Li, S. Ourselin, M. J. Cardoso, & T. Vercauteren (2017). DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

[13] K. He, X. Zhang, S. Ren, & J. Sun (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[14] M. Vallières, C. R. Freeman, S. R. Skamene, & I. El Naqa (2017). Automated segmentation of head and neck organs at risk in CT images using convolutional neural networks. Radiotherapy and Oncology, 123(3), 498-503.

[15] N. D. Tran, S. Huang, Q. V. H. Nguyen, H. L. Vu, L. N. T. Nguyen, & H. T. Pham (2019). Deep learning for healthcare applications based on physiological signals: A review. Journal of Physics: Conference Series, 1343(1), 012103.

[16] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, et al. (2017). Deep learning for healthcare: Review, opportunities and challenges. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops

[17] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, et al. A survey on deep learning in medical image analysis. Med. Image Anal. 2017

[18] O. Ronneberger, P. Fischer, & T. Brox (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

[19] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, & H. Larochelle (2017). Automatic brain tumour detection and segmentation using U-Net based fully convolutional networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

[20] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, & K. H. Maier-Hein (2019). Fully automatic segmentation of cardiac images by deep learning. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

[21] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, et al. (2018). Automated segmentation of left ventricle in cardiac MRI using deep learning. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

[22] X. X. Yin, L. Sun, Y. Fu, R. Lu, & Y. Zhang (2022). U-Net-Based Medical Image Segmentation. J Healthc Eng. 2022 Apr 15;2022:4189781. doi: 10.1155/2022/4189781.

[23] A. AL Qurri, & M. Almekkawy (2023). Improved U-Net with Attention for Medical Image Segmentation. Sensors, 23(20), 8589. https://doi.org/10.3390/s23208589.

[24] Qing Xu, Z. Ma, N. He, W. Duan, DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation, Computers in Biology and Medicine, Volume 154, March 2023, 106626

[25] S. Jin, S. Yu, J. Peng et al., A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning, Sci Rep 13, 6762 (2023). https://doi.org/10.1038/s41598-023-33357-y

[26] J. Ma, Y. Deng, Z. Ma, K. Mao, Y. Chen, A liver segmentation method based on the fusion of VNet and WGAN, Computational and Mathematical Methods in Medicine, 2021, Article ID 5536903.

[27] LiTS – Liver Tumour Segmentation Challenge (LiTS17) organised in conjunction with ISBI 2017 and MICCAI. (2017).

[28] J. Han, M. Kamber, P. Pei, Data Mining Concepts and Techniques, Third Edition, Morgan Kaufmann Publishers, 2012. ISBN 978-0-12-381479-1.

[29] R. Shanmugamani, Deep Learning for Computer Vision, Packt Publishing, 2018.

[30] A. Adadi, M. Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), IEEE Access, 6 (2018), pp. 52138-52160, 10.1109/ACCESS.2018.2870052.

[31] B.H.M. Van der Velden, H.J. Kuijf, K.G.A. Gilhuijs, M.A. Viergever, Explainable artificial intelligence (XAI) in deep learning-based medical image analysis, Medical Image Analysis, Volume 79, 2022, 102470, ISSN 1361-8415, https://doi.org/10.1016/j.media.2022.102470.

[32] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, Proceedings of the IEEE International Conference on Computer Vision (2017), pp. 618-626.

[33] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016), pp. 2921-2929.

[34] J. Ji. Gradient-based Interpretation on Convolutional Neural Network for Classification of Pathological Images Proceeding of the International Conference on Information Technology and Computer Application, ITCA (2019), pp. 83-86, 10.1109/ITCA49981.2019.00026

[35] K. Kowsari, R. Sali, L. Ehsan, W. Adorno, A. Ali, S. Moore, B. Amadi, P. Kelly, S. Syed, & D. Brown. HMIC: hierarchical medical image classification, a deep learning approach Information, 11 (2020), 10.3390/INFO11060318

[36] P. Windisch, P. Weber, C. Fürweger, F. Ehret, M. Kufeld, D. Zwahlen, & A. Muacevic. Implementation of model explainability for a basic brain tumour detection using convolutional neural networks on MRI slices Neuroradiology (2020), 10.1007/s00234-020-02465-1

[37] N. Jitani, A. Basumatary, R. Sarmah, (2023). Deep Learning-Based Tumor Segmentation from CT Images. In: Borah, S., Gandhi, T.K., Piuri, V. (eds) Advanced Computational and Communication Paradigms . ICACCP 2023. Lecture Notes in Networks and Systems, vol 535. Springer, Singapore. https://doi.org/10.1007/978-981-99-4284-8_27