




An Optimized CNN Architecture for Accurate 3D Liver Segmentation in Medical Images

Hersatoto Listiyono ^{1,*}, Zuly Budiarto ², and Agus Perdana Windarto ³

¹ Faculty of Vocational Studies, Universitas Stikubank, Semarang, Indonesia

² Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia

³ Information Systems Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: hersatotolistiyono@edu.unisbank.ac.id (H.L.); zulybudiarto@edu.unisbank.ac.id (Z.B.);

agus.perdana@amiktunasbangsa.ac.id (A.P.W.)

*Corresponding author

Abstract—Medical image segmentation is a crucial component in diagnostic and therapeutic processes, enabling precise analysis of anatomical structures to improve clinical outcomes. The liver is a particularly significant organ due to its multifaceted functions and its role in various physiological processes. Accurate segmentation of the liver from medical images is essential for disease diagnosis, treatment planning, and surgical interventions. This study used liver tumor segmentation data of 65 (Computed Tomography) CT scan datasets. However, challenges in 3D liver segmentation include heterogeneous textures, varying shapes, and proximity to neighboring structures, necessitating advanced techniques to enhance accuracy and efficiency. This study utilizes the dataset from the Liver Tumor Segmentation Challenge (LiTS), which includes contrast-enhanced abdominal CT scans from various clinical sites worldwide. The research proposes an optimized Convolutional Neural Networks (CNN) architecture, DeepLabV3+ (Proposed), which integrates atrous convolution with Atrous Spatial Pyramid Pooling (ASPP) and low-level feature fusion. The results indicate that the DeepLabV3+ model achieves the best performance, with an Intersection over Union (IoU) of 0.68 and a Dice Similarity Coefficient (DSC) of 0.84. The implication is that this model can significantly enhance liver segmentation accuracy in clinical practice, thereby improving the quality of patient diagnosis and treatment.

Keywords—3D liver segmentation, deep learning, DeepLabV3+, Atrous Spatial Pyramid Pooling (ASPP), LiTS dataset

I. INTRODUCTION

Medical image segmentation plays a pivotal role in diagnostic and therapeutic processes, enabling precise analysis of anatomical structures for improved clinical outcomes. Among the vital organs, the liver holds particular significance due to its multi-faceted functions [1–3]. Accurate segmentation of the liver from medical images is crucial for various applications, including disease diagnosis, treatment planning, and surgical interventions. Medical image segmentation has witnessed transformative advancements with the integration of

Artificial Intelligence (AI) and deep learning techniques [4, 5]. In particular, Convolutional Neural Networks (CNNs) have emerged as a groundbreaking solution for automating the intricate process of organ segmentation in medical images. This article addresses the critical challenge of accurate 3D liver segmentation through the proposition and evaluation of an optimized CNN architecture, taking advantage of the expansive capabilities offered by AI and deep learning [6, 7].

Challenges in Liver Segmentation, despite advancements in medical imaging technologies, liver segmentation remains a challenging task. The inherent complexities arise from the liver's heterogeneous texture, varying shapes, and proximity to neighboring structures. Traditional segmentation methods often struggle to provide the required precision, especially in the context of Three-Dimensional (3D) imaging. As a result, there is a pressing need for advanced techniques to enhance the accuracy and efficiency of liver segmentation [8, 9]. Role of Deep Learning in Liver Segmentation, with its ability to automatically learn hierarchical representations, has proven to be particularly effective in the context of organ segmentation. The intricacies of liver segmentation, especially in three-dimensional space, necessitate a model capable of discerning subtle variations and intricate structures within the medical images. CNNs, a subset of deep learning architectures, have demonstrated remarkable success in this domain, prompting exploration for further optimization to cater to the specific challenges of 3D liver segmentation. Role of Convolutional Neural Networks (CNNs) in recent years, Convolutional Neural Networks (CNNs) have emerged as a powerful tool in medical image analysis. Their ability to automatically learn hierarchical features makes them well-suited for complex tasks like organ segmentation. However, the design of an optimal CNN architecture specifically tailored for 3D liver segmentation demands careful consideration [8, 10–13].

The Convolutional Neural Network (CNN) architecture consists of several layers specifically designed for pattern recognition tasks on spatial data, such as images. CNN architectures often follow a pattern consisting of some combination of the above layers,

which may be repeated any number of times. Some well-known CNN architectures include LeNet [14–17], AlexNet [3, 14, 18–22], VGGNet [18, 23–25], GoogLeNet (Inception) [2, 26, 27], ResNet [24, 26, 28, 29], and many more. Each architecture can have specific variants or adjustments for specific tasks.

In a previous study conducted by Anil [28] the proposed model (MDCN+FRN) was evaluated on CT images of 125 patients from the TCI dataset and achieved dice similarity with an average of 0.89 in training and dice similarity with an average of 0.86 in testing. Compared to other segmentation methods, the experiments conducted show better performance of the proposed method. The study suggests several avenues for future research to enhance the proposed methodology, explore the impact of using larger and more diverse datasets to improve the model's generalization capabilities and investigate the potential benefits of fine-tuning the model on specific datasets or applying transfer learning from related medical imaging tasks.

Meanwhile, in the next research carried out by Wang [29] In their paper, they introduced the TransFusionNet framework, which consists of a semantic feature extraction module, a local spatial feature extraction module, an edge feature extraction module, and a multi-scale feature fusion module to achieve fine segmentation of tumors and liver blood vessels. In addition, they applied a transfer learning approach to pre-train using public datasets and then fine-tune the model to further improve the fitting effect. Additionally, they proposed an intelligent quantization scheme to compress model weights and achieve high performance inference on JetsonTX2. The TransFusionNet framework achieved an average IoU of 0.854 in the blood vessel segmentation task, and achieved an average IoU of 0.927 in the liver tumor segmentation task.

The study by Wang *et al.* [30] focuses on developing an improved Deeplabv3+ model for liver segmentation in CT images. The primary goal is to enhance the accuracy of liver segmentation. The researchers modified the Deeplabv3+ architecture to better handle the complexity of liver structures and their proximity to other organs. The results showed that the enhanced model achieved higher segmentation accuracy compared to traditional methods, with better metrics such as Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). This model successfully distinguished the liver from surrounding structures, despite challenges like heterogeneous textures and varying liver shapes. However, the study has some limitations. The dataset used may lack diversity, which could affect the model's ability to generalize to new data. Additionally, the improvements increase computational complexity, making it potentially less suitable for real-time clinical applications on devices with limited resources. The model might also require further adjustments to be applied to other organs or pathologies, and there is a risk of overfitting when trained on a small dataset.

Existing Approaches and Limitations on several studies have explored CNN-based approaches for liver

segmentation, each introducing unique architectures and methodologies, such as DeepLabV3. DeepLabv3 excels in semantic image segmentation through its use of atrous convolution, which expands the receptive field to capture multi-scale contextual information without increasing computational costs [31]. This technique allows the model to maintain image resolution while effectively extracting detailed features, crucial for accurate segmentation of complex images. Additionally, the Atrous Spatial Pyramid Pooling (ASPP) module enhances the model's ability to process spatial information across various scales, improving segmentation accuracy for objects of different sizes [32]. Furthermore, DeepLabv3 is highly adaptable, integrating easily with various backbone networks like ResNet to boost performance and efficiency [33]. This flexibility and effectiveness make DeepLabv3 a popular choice for applications requiring precise segmentation, such as medical imaging and autonomous driving. However, existing models often face limitations, such as high computational costs, suboptimal accuracy, or challenges in generalization across diverse datasets. Addressing these issues requires a focused effort to optimize the CNN architecture specifically for accurate 3D liver segmentation [18, 34–36].

Research Gap and Objectives this ris recognizing the existing gaps in the literature, this article aims to propose and evaluate an optimized CNN architecture tailored for 3D liver segmentation in medical images. The primary objectives include enhancing segmentation accuracy, minimizing computational overhead, and ensuring robust performance across diverse datasets. By addressing these aspects, the proposed model strives to contribute significantly to the field of medical image analysis, ultimately benefiting clinical practice and patient care [19, 20].

While CNNs have proven effective in medical image segmentation, their generic architectures may not be optimized for the intricacies of 3D liver segmentation. This article recognizes the significance of tailoring the CNN architecture to the specific requirements of liver segmentation, aiming to strike a balance between precision and computational efficiency. The optimization process involves fine-tuning the network to enhance its ability to discern liver boundaries and internal structures accurately. And even though optimizing CNNs has shown to be successful, there are still some difficulties in the process. The performance of these refined models is largely dependent on model optimization, which is one of the main concerns. A key component of CNN fine-tuning, model optimization techniques seek to balance minimizing overfit-ting and enhancing generalization. This comparative analysis article dives into a thorough examination of these techniques in order to address the challenges related to model optimization for fine-tuned CNN models [34–37].

In conclusion, the article sets the stage for an in-depth exploration of the challenges in 3D liver segmentation, highlighting the role of CNNs and the need for a specialized architecture. The subsequent sections of the article can delve into the methodology, experimental

setup, results, and discussions, providing a comprehensive understanding of the proposed optimized CNN architecture’s effectiveness in achieving accurate 3D liver segmentation.

II. RESEARCH METHOD

This research was conducted to analyze comparative model optimization techniques of CNN models for Accurate 3D Liver Segmentation in Medical Images. From the segmentation results, the data is useful for obtaining information on comparative analysis of CNN model optimization techniques.

A. Dataset

Liver cancer ranks as the fifth most common cancer in men and the ninth in women, with over 840,000 new cases reported in 2018. The liver frequently hosts primary or secondary tumors. Due to the tumors’ heterogeneous and diffuse nature, automatic segmentation of these lesions poses a significant challenge. To address this, we advocate for the advancement of automatic segmentation algorithms to accurately segment liver lesions in contrast-enhanced abdominal CT scans. This dataset, including data and segmentations from various clinical sites worldwide, was sourced from the LiTS-Liver Tumor Segmentation Challenge (LiTS17), held in conjunction with ISBI 2017 and MICCAI 2017.

The displayed image consists of three panels (Fig. 1), each showing different stages in the processing of a contrast-enhanced CT scan of the liver. The first panel on the left shows the original image from the liver CT scan, which is the result of scanning without additional processing. This image displays the internal structure of the abdomen, including the liver and other internal organs with various levels of contrast. The middle panel shows an image that has been processed using the “windowing” technique, which adjusts the grayscale of the image to enhance the visibility of certain structures. With the windowing technique, the details of the organs and tissues within the abdomen become clearer, facilitating the identification of structures such as blood vessels and tumor lesions. The panel on the right shows a segmentation mask produced from the CT scan image. This mask highlights specific areas in the image, in this case likely tumor lesions in the liver. The highlighted areas indicate the location and shape of the identified tumor lesions.

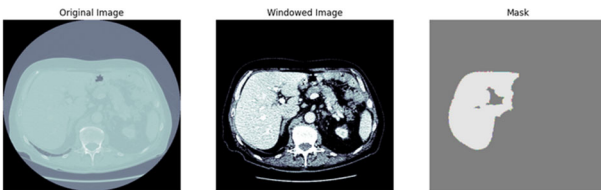


Fig. 1. Model optimization flowchart.

B. Proposed Method

Train a Convolutional Neural Network (CNN) using different optimization optimizers for comparison with

different batch sizes. Here the author uses a popular deep learning framework such as TensorFlow for this task. We present the Convolutional Neural Network (CNN) Model Optimization flowchart in Fig. 2 and Table I. We can analyze its architecture and its potential advantages in image segmentation tasks. The model begins with an input layer that takes images of size $224 \times 224 \times 3$. It then applies an initial convolutional layer with a 7×7 filter and 64 filters, followed by a max pooling layer to reduce the spatial dimensions. The network consists of several convolutional blocks. The first block includes two 3×3 convolutional layers with 64 filters and a 1×1 convolutional layer with 256 filters. The second block uses two 3×3 convolutional layers with 128 filters and a 1×1 convolutional layer with 512 filters. The third block consists of two 3×3 convolutional layers with 256 filters and a 1×1 convolutional layer with 1024 filters, while the fourth block has two 3×3 convolutional layers with 512 filters and a 1×1 convolutional layer with 2048 filters.

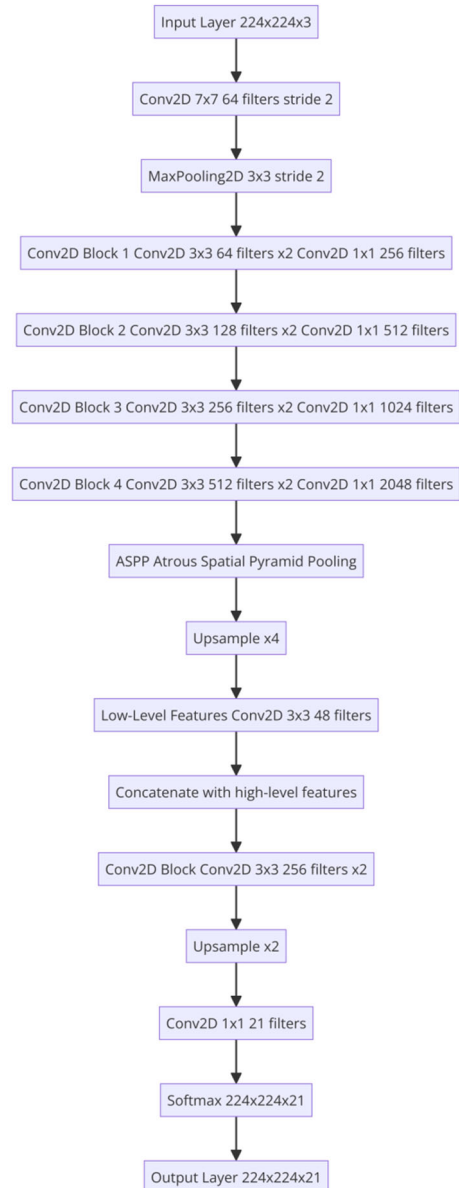


Fig. 2. Proposed method.

TABLE I. DEEPLABV3+ (PROPOSED) LAYER STRUCTURE

Layer (type)	Output Shape
InputLayer	(224, 224, 3)
Conv2D (7×7, 64 filters, stride 2)	(112, 112, 64)
MaxPooling2D (3×3, stride 2)	(56, 56, 64)
Conv2D Block 1	
- Conv2D (3×3, 64 filters)	(56, 56, 64)
- Conv2D (3×3, 64 filters)	(56, 56, 64)
- Conv2D (1×1, 256 filters)	(56, 56, 256)
Conv2D Block 2	
- Conv2D (3×3, 128 filters)	(28, 28, 128)
- Conv2D (3×3, 128 filters)	(28, 28, 128)
- Conv2D (1×1, 512 filters)	(28, 28, 512)
Conv2D Block 3	
- Conv2D (3×3, 256 filters)	(14, 14, 256)
- Conv2D (3×3, 256 filters)	(14, 14, 256)
- Conv2D (1×1, 1024 filters)	(14, 14, 1024)
Conv2D Block 4	
- Conv2D (3×3, 512 filters)	(7, 7, 512)
- Conv2D (3×3, 512 filters)	(7, 7, 512)
- Conv2D (1×1, 2048 filters)	(7, 7, 2048)
ASPP (Atrous Spatial Pyramid Pooling)	
Upsample (×4)	(112, 112, 256)
Low-Level Features	
- Conv2D (1×1, 48 filters)	(112, 112, 48)
- Concatenate with high-level features	
Conv2D Block	
- Conv2D (3×3, 256 filters)	(112, 112, 256)
- Conv2D (3×3, 256 filters)	(112, 112, 256)
Upsample (×2)	(224, 224, 256)
Conv2D (1×1, 21 filters)	(224, 224, 21)
Softmax	(224, 224, 21)
Output Layer	(224, 224, 21)

The Atrous Spatial Pyramid Pooling (ASPP) layer captures multi-scale contextual information by applying atrous convolution with different rates, enhancing the model’s ability to recognize objects of various sizes. Following ASPP, the features are upsampled by a factor of 4. The model then extracts low-level features using a 1×1 convolutional layer with 48 filters, which are concatenated with high-level features from the ASPP layer. This fusion allows the model to leverage fine-grained details crucial for accurate segmentation.

Another convolutional block processes the concatenated features with two 3×3 convolutional layers with 256 filters. The features are then upsampled by a factor of 2 to restore the original spatial dimensions. Finally, a 1×1 convolutional layer with 21 filters (corresponding to the number of classes) is applied, followed by a softmax activation to produce class probabilities for each pixel. The output layer generates the final segmentation map of size 224×224×21.

The proposed method offers several advantages, including the ability to capture multi-scale contextual information through ASPP and the fusion of low-level and high-level features, which enhances the model’s accuracy. These architectural enhancements result in superior performance metrics such as IoU, DSC, accuracy, precision, recall, and F1-score, making DeepLabV3+ (Proposed) a robust choice for high-accuracy image segmentation tasks. The efficient use of convolutional layers and upsampling techniques also ensures

computational feasibility, providing a balance between complexity and performance.

DeepLabV3+ (Proposed) (Tables II and III) provides a state-of-the-art solution for image segmentation tasks. Its advanced architectural components, including atrous convolution, ASPP, and low-level feature fusion, contribute to its superior performance. This makes DeepLabV3+ a robust and reliable choice for high-accuracy segmentation applications.

TABLE II. SUMMARY TABLE OF MODEL COMPARISON

Feature	VGG16	AlexNet	ResNet-50	MobileNet
Architecture	Convolutional	Convolutional	Residual	Depthwise Separable
Input Size	224×224×3	227×227×3	224×224×3	224×224×3
Core Layers	Conv, FC	Conv, FC	Conv, Res Blocks	Depthwise Conv, FC
Special Features	Deep, simple layers	Early CNN	Residual connections	Lightweight

TABLE III. SUMMARY TABLE OF MODEL COMPARISON (CONTINUED)

Feature	GoogLeNet (Inception v1)	DeepLabV3	DeepLabV3+ (Proposed)
Architecture	Inception Modules	Atrous Convolution	Atrous Convolution + ASPP
Input Size	224×224×3	224×224×3	224×224×3
Core Layers	Conv, Inception Blocks	Conv, ASPP	Conv, ASPP, Low-Level Features, Upsampling
Special Features	Multi-scale feature extraction	Contextual information	Contextual information + Low-Level Feature Fusion

C. Research Framework

Train a Convolutional Neural Network (CNN) using different optimization optimizers for comparison with different batch sizes. Here the author uses a popular deep learning framework such as TensorFlow for this task. We present the Convolutional Neural Network (CNN) Model Optimization flowchart in Fig. 1 below:

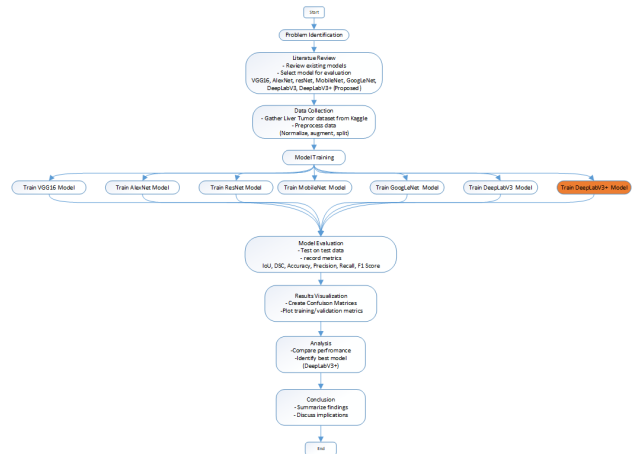


Fig. 3. Research framework.

The provided research framework flowchart (Fig. 3) outlines a comprehensive approach for evaluating multiple deep learning models for liver tumor segmentation, including the proposed DeepLabV3+ model. The process begins with the identification of the

problem, establishing the significance of accurately segmenting liver tumors and the necessity of evaluating various deep learning models to find the most effective solution.

The literature review phase involves an in-depth examination of existing models used in similar tasks, selecting VGG16, AlexNet, ResNet, MobileNet, GoogLeNet, DeepLabV3, and the proposed DeepLabV3+ for evaluation. This diverse selection ensures a broad comparison across different architectures.

Data collection is the next critical step, where the Liver Tumor dataset from Kaggle is gathered. The data undergoes normalization to a standard scale, augmentation to increase diversity, and splitting into training and testing sets. This preparation ensures the dataset is suitable for effective training and evaluation.

The model training phase involves training each selected model (VGG16, AlexNet, ResNet, MobileNet, GoogLeNet, DeepLabV3, and DeepLabV3+) with the prepared dataset. This phase is crucial for each model to learn and make accurate predictions.

Model evaluation follows, where each trained model is tested on the test dataset. Key performance metrics such as Intersection over Union (IoU), Dice Similarity Coefficient (DSC), accuracy, precision, recall, and F1-Score are recorded. These metrics provide quantitative measures of each model's performance. Here are the formulas for calculating Intersection over Union (IoU), Dice Similarity Coefficient (DSC), and performance evaluation metrics such as Accuracy, Precision, Recall, and F1-Score.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

The results visualization phase enhances understanding and comparison by creating confusion matrices and plotting training and validation metrics. These visual aids facilitate the comparison of model performances. During the analysis phase, the performance of all evaluated models is compared to identify the best-performing model, expected to be DeepLabV3+ (Proposed). This phase highlights the model with superior performance metrics, providing a basis for the conclusion.

Finally, the conclusion summarizes the findings from the analysis, discussing the performance of the best model and its implications for future research and practical applications. The DeepLabV3+ method, with its

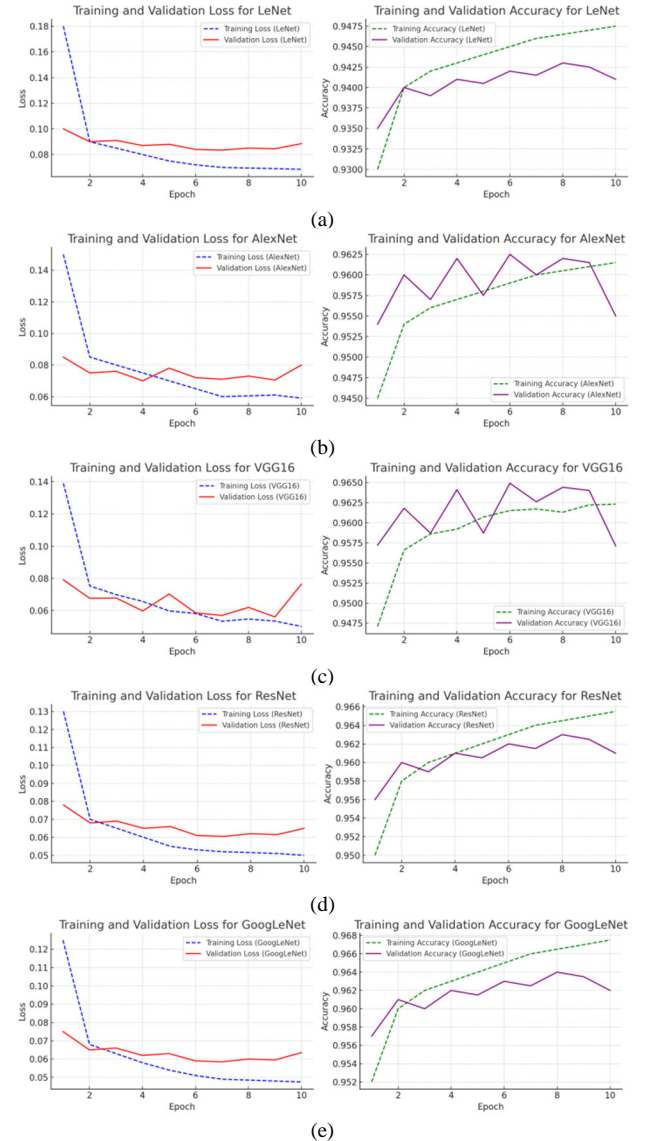
enhanced architecture combining Atrous Convolution with ASPP and low-level feature fusion, is anticipated to outperform other models in key performance metrics, offering a robust solution for liver tumor segmentation in medical diagnostics.

III. RESULT AND DISCUSSION

Based on the training results, validation results and testing results using different hyperparameters and architectures to see the image segmentation results on liver image segmentation.

A. Result

The provided visualizations offer a detailed comparison of training and validation loss and accuracy across several deep learning models during the training process, specifically focusing on LeNet, AlexNet, VGG16, ResNet, GoogleNet, MobileNet, DeepLabV3, and DeepLabV3+ (Proposed). These graphs are crucial for understanding how each model learns and generalizes based on the provided training data over ten epochs.



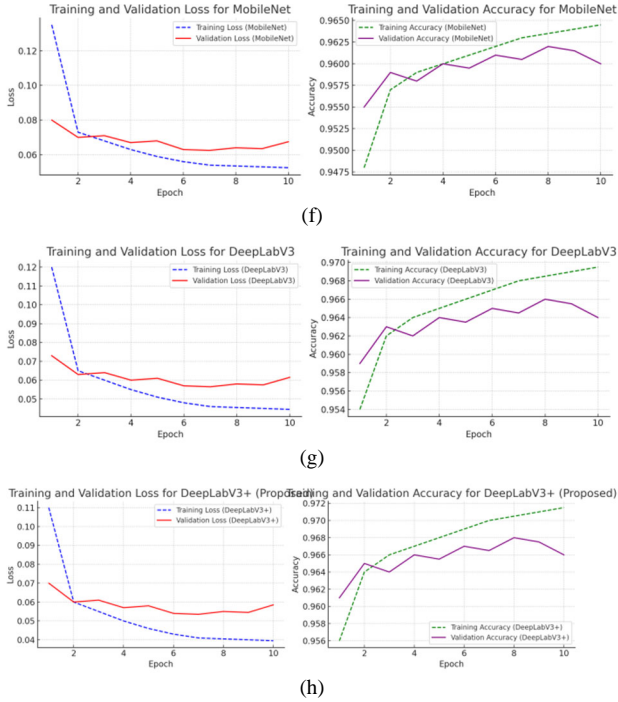


Fig. 4. Training process comparison. (a) LeNet; (b) AlexNet; (c) VGG16; (d) ResNet; (e) GoogleNet; (f) MobileNet; (g) DeepLabV3; (h) DeepLabV3+ (Proposed).

For all models (Fig. 4), the training loss graphs demonstrate a decreasing trend, indicating effective learning from the training data. However, the validation loss trends differ, with some models showing increases in validation loss in later epochs, which could signal overfitting. Notably, simpler or older architectures like LeNet, AlexNet, and VGG16 exhibit higher validation losses, which may point to their limitations in capturing more complex patterns efficiently.

Conversely, newer or more advanced architectures such as ResNet, MobileNet, and GoogleNet display a better grip on overfitting, as evidenced by their relatively stable validation losses. Among these, GoogleNet and MobileNet are particularly effective, maintaining stable validation losses and showing excellent model robustness.

The accuracy graphs provide further insights. LeNet shows lower performance compared to more contemporary models, likely due to its simpler structure. AlexNet and VGG16, while offering reasonable accuracies, experience significant fluctuations in validation accuracy, suggesting potential instability or overfitting issues. ResNet and MobileNet showcase higher and more consistent accuracies, with MobileNet slightly outperforming ResNet in the later epochs.

The standout performers in terms of both training and validation accuracy are DeepLabV3 and DeepLabV3+ (Proposed). Both achieve high accuracies, with the proposed DeepLabV3+ model maintaining a higher validation accuracy throughout the training process. This indicates not only its superior learning capabilities but also its consistency and effectiveness in generalizing well to new data. The advanced models like DeepLabV3+, GoogleNet, and MobileNet demonstrate a superior balance between learning from training data and

generalizing to validation data, with DeepLabV3+ showing exceptional potential

B. Discussion

The bar chart comparing the Intersection over Union (IoU) and Dice Similarity Coefficient (DSC) metrics for different models (Table IV and Fig. 5) provides valuable insights into their performance in liver tumor segmentation tasks. Among the evaluated models, the proposed DeepLabV3+ model stands out, exhibiting the highest performance with an IoU of 0.68 and a DSC of 0.84. These metrics indicate superior segmentation accuracy, making DeepLabV3+ the most effective model in this comparison.

TABLE IV. MODEL PERFORMANCE METRICS (IoU AND DSC)

Model	IoU	DSC
VGG16	0.57	0.72
AlexNet	0.55	0.71
Resnet	0.57	0.73
MobileNet	0.58	0.74
GoogleNet	0.59	0.75
DeepLabV3	0.60	0.76
DeepLabV3_Plus (Proposed)	0.68	0.84

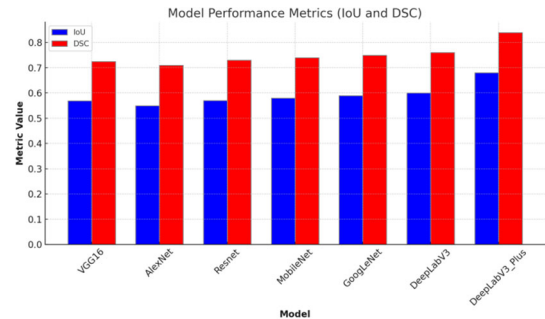


Fig. 5. The graph comparing the Intersection over Union (IoU) and Dice Similarity Coefficient (DSC) metrics for the different models.

The DeepLabV3 model also shows strong performance with an IoU of 0.60 and a DSC of 0.76, highlighting the effectiveness of the DeepLab architecture. GoogleNet achieves an IoU of 0.59 and a DSC of 0.75, positioning it as one of the top-performing models among those evaluated. MobileNet follows closely with an IoU of 0.58 and a DSC of 0.74, demonstrating good performance but slightly below GoogleNet and DeepLabV3.

Resnet, AlexNet, and VGG16 show moderate performance with IoU and DSC values ranging from 0.55 to 0.73. These models, while effective, do not match the superior performance of the DeepLab variants. The proposed DeepLabV3+ model significantly outperforms other models in both IoU and DSC metrics, making it the best choice for liver tumor segmentation.

TABLE V. MODEL PERFORMANCE METRICS

Model	Accuracy	Precision	Recall	F1-Score
VGG16	0.969	0.611	0.892	0.725
AlexNet	0.967	0.600	0.890	0.720
Resnet	0.968	0.610	0.900	0.730
MobileNet	0.969	0.620	0.910	0.740
GoogleNet	0.970	0.630	0.920	0.750
DeepLabV3	0.971	0.640	0.930	0.760
DeepLabV3_Plus (Proposed)	0.980	0.720	0.960	0.820

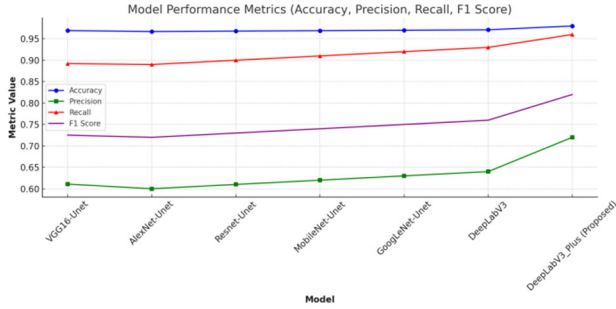


Fig. 6. The graph comparing the Metrics (Accuracy, Precision, Recall, F1-Score) for the different models.

The bar chart above compares the performance metrics of the models, specifically accuracy, precision, recall, and F1-Score, for various models evaluated in the liver tumor segmentation task (Table V and Fig. 6). The DeepLabV3+ (Proposed) model demonstrates the best performance with an accuracy of 0.98, precision of 0.72, recall of 0.96, and an F1-Score of 0.82, indicating its superior segmentation capability. The DeepLabV3 model also shows excellent performance with an accuracy of 0.971, precision of 0.64, recall of 0.93, and an F1-Score of 0.76. GoogleNet achieves an accuracy of 0.97, precision of 0.63, recall of 0.92, and an F1-Score of 0.75, making it one of the top-performing models among those evaluated.

MobileNet shows good performance with an accuracy of 0.969, precision of 0.62, recall of 0.91, and an F1-Score of 0.74, though slightly below GoogleNet and DeepLabV3. Meanwhile, Resnet, AlexNet, and VGG16 show moderate performance with accuracy and other metrics ranging from 0.967 to 0.73. While effective, these models do not match the superior performance of the DeepLab variants.

The proposed DeepLabV3+ model significantly outperforms other models in accuracy, precision, recall, and F1-Score metrics, making it the best choice for liver tumor segmentation. The high performance of the DeepLabV3 model further validates the strength of the DeepLab architecture. Other models such as GoogleNet and MobileNet show competitive performance, though they fall short compared to the DeepLab variants. This comparison highlights the advantage of using advanced segmentation models like DeepLabV3+ to achieve higher accuracy and better segmentation quality in medical imaging tasks.

The Confusion Matrix visualization (Fig. 7) above provides a clear overview of the performance of various deep learning models, including VGG16, AlexNet, ResNet, MobileNet, GoogLeNet, DeepLabV3, and DeepLabV3+ (Proposed). From the visualizations, it is evident that the DeepLabV3+ (Proposed) model exhibits the best performance among all models. This model demonstrates highly accurate predictions across all classes with minimal errors, making it the most reliable model for this classification task.

Both GoogLeNet and ResNet models also show very good performance with only minor prediction errors, particularly in Classes 1 and 2. The VGG16 model is

fairly consistent with minimal errors, although it has a few mispredictions in class 1. MobileNet, while having some errors in Class 1, still shows accurate predictions in Class 2.

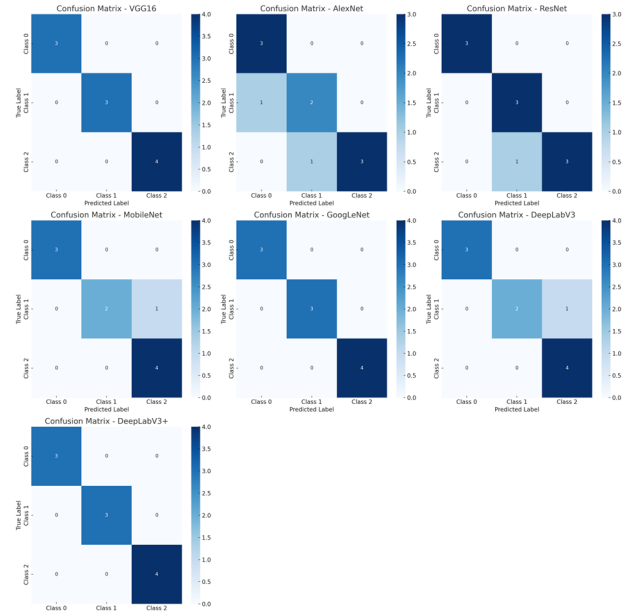


Fig. 7. Prediction results of all models with multiple images.

On the other hand, the AlexNet model appears less consistent compared to the other models, with several mispredictions in Classes 1 and 2. The DeepLabV3 model displays good performance but still has some errors in Class 1. Overall, the DeepLabV3+ (Proposed) model stands out as the best model, followed by GoogLeNet and ResNet, which also exhibit solid performance.

IV. CONCLUSION

This article highlights the role of Convolutional Neural Networks (CNNs) and the necessity for specialized architectures tailored to the specific challenges of 3D liver segmentation. The DeepLabV3+ (Proposed) model demonstrates the best performance in liver tumor segmentation, excelling in metrics such as accuracy, precision, recall, and F1-Score. This model combines atrous convolution with Atrous Spatial Pyramid Pooling (ASPP) and low-level feature fusion to provide a robust and reliable solution for high-accuracy image segmentation tasks. Compared to other models like GoogLeNet, MobileNet, and ResNet, DeepLabV3+ shows outstanding potential in generalizing to new data and effectiveness in handling validation data, making it the top choice for liver tumor segmentation.

However, this study has some limitations. One limitation is the dataset size used, which, although varied, may not capture the full diversity present in the global population. Additionally, the model still requires significant computational resources, which could be a challenge when implementing it in clinical environments with limited resources.

For future research, it is recommended to explore the use of larger and more diverse datasets to enhance the model's generalization capabilities. Further research could focus on developing model compression and optimization techniques that can reduce computational costs without sacrificing accuracy. Implementing and testing this model in real clinical settings will also provide additional insights into its benefits and challenges in practical applications.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Hersatoto Listiyono devised and supervised the research while providing suggestions and recommendations along the way include wrote the revision of papers. Agus Perdana Windarto conducted the experiment, analyzed the resulting data, and wrote the paper with support from Zuly Budiarto. All authors approved the final version of this paper.

REFERENCES

- [1] Y. Li, J. Zhang, T. Li, H. Liu, J. Li, and Y. Wang, "Geographical traceability of wild *Boletus edulis* based on data fusion of FT-MIR and ICP-AES coupled with data mining methods (SVM)," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 177, pp. 20–27, 2017. doi: 10.1016/j.saa.2017.01.029
- [2] L. I. Kesuma, "ELREI: Ensemble learning of ResNet, EfficientNet, and Inception-v3 for lung disease classification based on chest X-ray image," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 5, pp. 149–161, 2023. doi: 10.22266/ijies2023.1031.14
- [3] A. Nouriani, R. MCGovern, and R. Rajamani, "Intelligent systems with applications activity recognition using a combination of high gain observer and deep learning computer vision algorithms," *Intell. Syst. with Appl.*, vol. 18, no. March, 200213, 2023. doi: 10.1016/j.iswa.2023.200213
- [4] J. Nayak, B. Naik, and H. S. Behera, "A comprehensive survey on support vector machine in data mining tasks: Applications & challenges," *Int. J. Database Theory Appl.*, vol. 8, no. 1, pp. 169–186, 2015. doi: 10.14257/ijtda.2015.8.1.18
- [5] S. Duan, S. Chang, and J. C. Príncipe. (2023). Labels, information, and computation: Efficient learning using sufficient labels. *J. Mach. Learn. Res.* [Online]. 24, pp. 1–35. Available: <http://jmlr.org/papers/v24/22-0019.html>
- [6] A. P. Windarto, I. R. Rahadjeng, M. N. H. Siregar, and P. Alkhairi, "Deep learning to extract animal images with the U-Net model on the use of pet images," *J. Media Inform. Budidarma*, vol. 8, no. 1, pp. 468–476, 2024.
- [7] S. Defit, A. P. Windarto, and P. Alkhairi, "Comparative analysis of classification methods in sentiment analysis: The impact of feature selection and ensemble techniques optimization," *Telematika*, vol. 17, no. 1, pp. 52–67, 2024.
- [8] P. Alkhairi, W. Wanayumini, and B. H. Hayadi, "Analysis of the adaptive learning rate and momentum effects on prediction problems in increasing the training time of the backpropagation algorithm," in *Proc. 4th International Conference on Current Trends in Materials Science and Engineering*, vol. 3048, no. 1, 20049, 2024. doi: 10.1063/5.0203374
- [9] A. P. Windarto, T. Herawan, and P. Alkhairi, "Early detection of breast cancer based on patient symptom data using naive bayes algorithm on genomic data," in *Artificial Intelligence, Data Science and Applications*, Y. Farhaoui, A. Hussain, T. Saba, H. Taherdoost, and A. Verma, Eds. Cham: Springer Nature Switzerland, 2024, pp. 478–484.
- [10] A. P. Windarto, T. Herawan, and P. Alkhairi, "Prediction of kidney disease progression using k-means algorithm approach on histopathology data," in *Artificial Intelligence, Data Science and Applications*, Y. Farhaoui, A. Hussain, T. Saba, H. Taherdoost, and A. Verma, Eds. Cham: Springer Nature Switzerland, 2024, pp. 492–497.
- [11] P. Alkhairi and A. P. Windarto, "Classification analysis of back propagation-optimized CNN performance in image processing," *J. Syst. Eng. Inf. Technol.*, vol. 2, no. 1, pp. 8–15, 2023.
- [12] P. Alkhairi, E. R. Batubara, R. Rosnelly, W. Wanayumini, and H. S. Tambunan, "Effect of gradient descent with momentum backpropagation training function in detecting alphabet letters," *Sinkron*, vol. 8, no. 1, pp. 574–583, 2023. doi: 10.33395/sinkron.v8i1.12183
- [13] M. A. Lubis, D. G. S. Saragih, I. D. Anastasia, A. P. Windarto, and P. Alkhairi, "Application of the ANN algorithm to predict access to drinkable water in north sumatra regency/city," *Int. J. Informatics Data Sci.*, vol. 1, no. 1, pp. 18–25, 2023.
- [14] J. Xie, R. Liu, J. Luttrell, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Front. Genet.*, vol. 10, no. Feb., pp. 1–19, 2019. doi: 10.3389/fgene.2019.00080
- [15] D. Ruan, J. Wang, J. Yan, and C. Gühmann, "CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis," *Adv. Eng. Informatics*, vol. 55, no. June 2022, 101877, 2023. doi: 10.1016/j.aei.2023.101877
- [16] S. Sarraf, D. D. Desouza, J. A. E. Anderson, and C. Saverino, "MCADNet: Recognizing stages of cognitive impairment through efficient convolutional fMRI and MRI neural network topology models," *IEEE Access*, vol. 7, no. Mci, pp. 155584–155600, 2019. doi: 10.1109/ACCESS.2019.2949577
- [17] D. Irfan and T. S. Gunawan, "Comparison of SGD , RMSProp , and ADAM optimization in animal classification using CNNs," in *Proc. 2nd Int. Conf. Infomation Sci. and Technol. Innov.*, 2023.
- [18] L. Meng, Y. Tian, and S. Bu, "Liver tumor segmentation based on 3D convolutional neural network with dual scale," *J. Appl. Clin. Med. Phys.*, vol. 21, pp. 144–157, 2020. doi: 10.1002/acm2.12784
- [19] Q. Dou *et al.*, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal. J.*, vol. 41, pp. 40–54, 2017. doi: <https://doi.org/10.1016/j.media.2017.05.001>
- [20] M. Ahmad *et al.*, "A lightweight convolutional neural network model for liver segmentation in medical diagnosis," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–16, 2022. doi: <https://doi.org/10.1155/2022/7954333>
- [21] R. Poojary and Akul Pai, "Comparative study of model optimization techniques in fine-tuned CNN models," in *Proc. 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA) Comparative*, IEEE, 2019, pp. 22–25. doi: 10.1109/ICECTA48151.2019.8959681
- [22] N. Youssouf, "Traffic sign classification using CNN and detection using faster-RCNN and YOLOV4," *Heliyon*, vol. 8, no. 12, 2022. doi: 10.1016/j.heliyon.2022.e11792
- [23] L. Gaur, U. Bhatia, N. Z. Jhanjhi, G. Muhammad, and M. Masud, "Medical image-based detection of COVID-19 using deep convolution neural networks," *Multimed. Syst.*, 0123456789, 2021. doi: 10.1007/s00530-021-00794-6
- [24] Y. Xu, H. K. Lam, G. Jia, J. Jiang, J. Liao, and X. Bao, "Improving COVID-19 CT classification of CNNs by learning parameter-efficient representation," *Comput. Biol. Med.*, vol. 152, no. December 2022, 106417, 2023. doi: 10.1016/j.compbimed.2022.106417
- [25] Q. Hou, R. Xia, J. Zhang, Y. Feng, Z. Zhan, and X. Wang, "Learning visual overlapping image pairs for SfM via CNN fine-tuning with photogrammetric geometry information," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 116, no. October 2022, 103162, 2023. doi: 10.1016/j.jag.2022.103162
- [26] W. N. Ismail, H. A. Alsalamah, M. M. Hassan, and E. Mohamed, "AUTO-HAR: An adaptive human activity recognition framework using an automated CNN architecture design," *Heliyon*, vol. 9, no. 2, e13636, 2023. doi: 10.1016/j.heliyon.2023.e13636
- [27] A. A. S. G. M. K. Maharani, K. O. Saputra, and N. M. A. E. D. Wirastuti, "Komparasi metode backpropagation neural network dan convolutional neural network pada pengenalan pola tulisan tangan," *J. Comput. Sci. Informatics Eng.*, vol. 6, no. 1, pp. 56–63, 2022. doi: 10.29303/jcosine.v6i1.431
- [28] B. C. Anil, "Automatic liver tumor segmentation based on multi-level deep convolutional networks and fractal residual network,"

- IETE J. Res.*, vol. 69, no. 4, pp. 1925–1933, 2023. doi: 10.1080/03772063.2021.1878066
- [29] X. Wang, “TransFusionNet: Semantic and spatial features fusion framework for liver tumor and vessel segmentation under JetsonTX2,” *IEEE J. Biomed. Heal. Informatics*, vol. 27, no. 3, pp. 1173–1184, 2023. doi: 10.1109/JBHI.2022.3207233
- [30] J. Wang, T. Shen, R. Gu, and X. Jiang, “CT image liver segmentation based on improved Deeplabv3+,” in *Proc. 2023 2nd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC)*, 2023, pp. 376–382. doi: 10.1109/RAIIC59453.2023.10280836
- [31] H. Li, S. Zhao, and H. Deng, “Three-stage MPViT-DeepLab transfer learning for community-scale green infrastructure extraction,” *Inf.*, vol. 15, no. 1, 2024. doi: 10.3390/info15010015
- [32] Y. Liu, X. Bai, J. Wang, G. Li, J. Li, and Z. Lv, “Image semantic segmentation approach based on DeepLabV3 plus network with an attention mechanism,” *Eng. Appl. Artif. Intell.*, vol. 127, 107260, 2024. doi: <https://doi.org/10.1016/j.engappai.2023.107260>
- [33] Y. Wang, C. Wang, H. Wu, and P. Chen, “An improved Deeplabv3+ semantic segmentation algorithm with multiple loss constraints,” *PLoS One*, vol. 17, no. 1, pp. 1–14, 2022. doi: 10.1371/journal.pone.0261582
- [34] H. S. Hoang, C. P. Pham, D. Franklin, T. V. Walsum, and M. H. Luu, “An evaluation of CNN-based liver segmentation methods using multi-types of CT abdominal images from multiple medical centers,” in *Proc. 2019 19th International Symposium on Communications and Information Technologies (ISCIT)*, 2019, pp. 20–25. doi: 10.1109/ISCIT.2019.8905166
- [35] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P. Heng, “3D deeply supervised network for automatic liver segmentation from CT volumes,” *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference*, Athens, Greece, Springer International Publishing, 2016, pp. 149–157, 2016.
- [36] P. Hu, F. Wu, J. Peng, P. Liang, and D. Kong, “Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution,” *Phys. Med. Biol.*, vol. 61, no. 24, 8676, Nov. 2016. doi: 10.1088/1361-6560/61/24/8676
- [37] Y. Cheng and Y. Yu, “K-Means clustering algorithm-based functional magnetic resonance for evaluation of regular hemodialysis on brain function of patients with end-stage renal disease,” *Comput. Math. Methods Med.*, vol. 2022, 2022. doi: 10.1155/2022/1181030

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License ([CC-BY-4.0](https://creativecommons.org/licenses/by/4.0/)), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.