

ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 [Online]

URL: <https://dergipark.org.tr/ij3dptdi>

3D MEDICAL IMAGE SEGMENTATION WITH DEEP LEARNING METHODS

Yazarlar (Authors): Sezin Barın , Uçman Ergün , Gür Emre Güraksın 




Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): Barın S., Ergün U., Güraksın G. E., “3D Medical Image Segmentation with Deep Learning Methods” *Int. J. of 3D Printing Tech. Dig. Ind.*, 9(1): 73-91, (2025).

DOI: 10.46519/ij3dptdi.1571288

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): <https://dergipark.org.tr/en/pub/ij3dptdi/archive>

3D MEDICAL IMAGE SEGMENTATION WITH DEEP LEARNING METHODS

Sezin Barın^a , Uçman Ergün^a , Gür Emre Güraksın^b 

^a Afyon Kocatepe University, Engineering Faculty, Biomedical Engineering Department, Afyonkarahisar/Turkey

^b Afyon Kocatepe University, Engineering Faculty, Computer Engineering Department, Afyonkarahisar/Turkey

* Corresponding Author: sbarin@aku.edu.tr

(Received: 21.10.24; Revised: 05.02.25; Accepted: 19.03.25)

ABSTRACT

With advancements in technology, three-dimensional (3D) medical imaging has become vital in modern medicine, contributing to more accurate diagnosis, treatment planning, and personalized medicine. However, segmenting abdominal organs remains a challenging task due to anatomical variations, limited labeled data, and image noise. This study investigates the impact of deep learning-based architectures and preprocessing techniques on 3D organ segmentation using the publicly available Multi-Atlas Labeling Beyond the Cranial Vault (BTCV) dataset. To achieve this, 3D U-Net, UNETR, and SwinUNETR models were employed, and the effects of various preprocessing techniques and loss functions, including Dice Loss, Focal Loss, and Cross-Entropy Loss, were systematically analyzed. The findings reveal that combining Dice Loss with Cross-Entropy Loss significantly enhances segmentation performance. Additionally, preprocessing techniques improved segmentation accuracy by 1.19%, further optimizing model performance. Among the evaluated models, 3D U-Net achieved the highest overall segmentation performance, with an average Dice score of 0.8397, outperforming SwinUNETR and UNETR. These findings underscore the importance of selecting appropriate preprocessing methods and loss functions in 3D medical image segmentation. The results contribute to more precise and efficient medical image analysis, with potential applications in clinical decision support systems. Future research should focus on optimizing hybrid architectures, integrating advanced augmentation strategies, and expanding evaluation across multiple datasets to improve the robustness and real-world applicability of automated segmentation methods.

Keywords: Deep Learning, Image Processing, 3D Image Segmentation, Medical Image Analysis, 3D U-Net, UNETR, SwinUNETR.

1. INTRODUCTION

Medical imaging systems and medical images have long been a fundamental part of medicine. Developments in medical imaging systems have enabled the development of new approaches in early diagnosis of diseases, treatment planning, and monitoring of the treatment process[1]. However, medical image analysis, interpretation, and reporting are usually time-consuming and require expertise. In particular, manual segmentation applications in medical images are labor-intensive, prone to inter-observer variability, and subject to inconsistencies[2]. For this reason, studies on automatic analysis of medical images have continued to be popular for years.

Advancements in artificial intelligence, particularly in deep learning, have led to significant progress in automatic image analysis[3]. Unlike traditional image processing and machine learning methods, deep learning techniques automatically extract distinctive features from images, making them highly effective in segmentation, classification, detection, and registration tasks[4]. Thanks to this advanced feature, deep learning models exhibit superior performance in challenging tasks such as segmentation, classification, detection, and registration of images. The success of deep learning in medical image analysis has led to an increasing number of studies applying these methods for automatic segmentation[3,5-6]. According to the

literature, segmentation studies have a small proportion compared to other tasks in the field of medical image analysis. Compared to tasks like classification and detection, medical image segmentation remains a less explored area due to challenges such as limited annotated datasets[7], high computational costs[6], the complexity of medical image features[3], evaluation difficulties[8], and the integration of automated methods into clinical workflows[9].

Segmentation is a process used to separate targeted anatomical structures (organs, lesions, etc.) from other structures in medical images[10]. This process is the most critical step in many clinical applications, such as disease diagnosis, treatment planning, surgical interventions, and monitoring of the disease process, providing quantitative and objective information[11]. Accurate segmentation facilitates better treatment decisions and enables more efficient monitoring of disease progression. While manual segmentation can be time-consuming and subjective, automatic segmentation methods reduce the workload and provide more consistent results[12]. Deep learning models, particularly Convolutional Neural Networks (CNNs) and their derivatives, have demonstrated superior performance over conventional segmentation approaches[13]. Accurate and reliable organ segmentation in medical images is important in clinical applications in disease diagnosis, treatment planning, treatment process monitoring, surgical planning, and navigation systems [14]. Despite their success, medical image segmentation models still face challenges, such as variations in organ shapes, poor image contrast, and noise-related artifacts. In organ segmentation, it is possible to analyze medical images in 2D and 3D. Literature studies show that 2D image analysis is more widely preferred than 3D. Possible reasons for this situation can be listed as follows [3,7,10,11,15];

- 2D images can be easily obtained, and therefore, data sets can be easily created,
- 2D image analysis has lower computational costs than 3D images,
- 3D images need more preprocessing,
- Ease of comparison due to testing old methods on 2D images,
- Ease of visualization and interpretation of 2D images

In addition, there are some advantages of segmenting medical images in 3D instead of 2D segmentation that provide more successful segmentation. For example[14,16,17];

- 3D segmentation methods provide more consistent and accurate segmentation by preserving volumetric information and spatial relationships between neighboring slices and using contextual information.
- Preserving contextual information between neighboring pixels increases segmentation performance in noisy or low-contrast regions .
- When 3D medical images are analyzed in 2D, errors caused by shifts between slices are not considered. More accurate results can be obtained by reducing projection effects with 3D segmentation methods. This is especially important in surgical planning or navigation systems .

Since 2D segmentation methods work independently for each slice, they need to be repeated on the 3D volume. This increases computational cost and is an inefficient process. 3D segmentation methods work directly on the 3D volume using efficient architectures such as 3D convolutional neural networks and significantly reduce computational time [18].

However, these advantages of 3D images bring some difficulties. Handling 3D data in parallel necessitates the use of high computation hardware. Moreover, 3D organ segmentation is challenging due to the large data size, variations in the shape and appearance of organs, and image noise and artifacts[10]. From another perspective, deep learning-based approaches have the potential to automatically learn meaningful features from 3D medical images and increase segmentation accuracy[9].

The primary purpose of this study is to perform high-performance 3D abdominal organ segmentation using deep learning methods. As mentioned, abdominal organ segmentation is an essential problem in medical image analysis. Still, it is a challenging task due to the complex anatomy of organs, image quality issues, and high variability between organs [14]. Existing 2D segmentation methods show limited

performance in 2D organ segmentation since they cannot fully capture 3D volumetric information. Therefore, 3D segmentation methods that take into account 3D contextual information and volumetric morphology of organs contribute to the development of existing automatic segmentation systems. In the scope of the study, the performances of 3D U-Net [16], UNETR [19], and SwinUNETR [20] models, which are deep learning-based 3D segmentation architectures that have been widely used in recent years, were comparatively evaluated. These models aim to perform organ segmentation by extracting meaningful volumetric features from 3D medical images using 3D convolutional neural networks (CNN) and transformer-based approaches. In addition to deep learning-based 3D segmentation architectures, the effects of morphological image preprocessing and different loss functions on segmentation performance were also investigated in the study. In addition, the impact of data augmentation techniques on 3D organ segmentation was also analyzed in the scope of the study. Data augmentation methods can increase the generalization ability of deep learning models, especially in cases with limited training examples. Therefore, the aim is to make the models robust against different variations by applying different data augmentation methods.

The rest of this paper is structured as follows: Section 2 presents the related work, summarizing previous research on 3D medical image segmentation and highlighting key advancements in deep learning-based approaches. Section 3 describes the materials and methods used in this study, detailing the dataset, preprocessing techniques, and model configurations. Section 4 presents the experimental results, comparing different segmentation approaches and analyzing their effectiveness. Finally, Section 5 concludes the paper by summarizing key findings and discussing potential future research directions.

2. RELATED WORKS

In recent years, deep learning-based approaches, especially convolutional neural networks (CNN) and transformer architectures, have shown impressive results in medical image segmentation[3, 5].

U-Net[13] and V-Net[21] are considered the leading CNN architectures in medical image

segmentation. Çiçek et al. [16] introduced 3D U-Net, an extension of the U-Net model, designed for volumetric medical image segmentation. The study aimed to improve segmentation performance in 3D medical imaging, particularly in cases with sparse annotations where manual labeling is limited. The proposed model replaces 2D convolutions with 3D convolutional layers, enabling better feature extraction for volumetric data. The model was evaluated on electron microscopy (EM) data for neuron segmentation and magnetic resonance imaging (MRI) data for brain tumor segmentation. Experimental results demonstrated that 3D U-Net significantly outperformed traditional 2D approaches, particularly in segmenting small and complex structures. The model achieved a Jaccard score (IoU) of 0.853 in neuron segmentation and a Dice Similarity Coefficient (DSC) of 0.897 in brain tumor segmentation. These results highlight the effectiveness of 3D U-Net in handling volumetric medical images, even when trained on limited labeled data, making it a valuable tool for automated medical image analysis.

Milletari et al. [21] introduced V-Net, a fully convolutional neural network (FCN) designed for volumetric medical image segmentation, particularly prostate segmentation in MRI scans. The study aimed to overcome the limitations of 2D CNNs by utilizing 3D convolutional layers, allowing the model to learn spatial context across entire volumetric images. A key innovation of V-Net is the introduction of a Dice loss function, which is optimized directly during training. This loss function effectively addresses the class imbalance issue, which is common in medical image segmentation, by prioritizing foreground voxels without requiring manual weighting. The model was trained and evaluated on the PROMISE12[22] prostate MRI dataset, consisting of 50 training and 30 test volumes. The experimental results demonstrated that V-Net achieved a Dice similarity coefficient (DSC) of 0.869, outperforming standard CNN-based approaches. The study also highlighted that V-Net significantly reduced segmentation time, achieving inference in just 1 second per MRI volume, making it suitable for real-time clinical applications.

Recently, Transformer architectures have shown remarkable performance in medical

image segmentation. Hatamizadeh et al. [19] proposed UNETR, a transformer-based model for 3D medical image segmentation, addressing the limitations of CNNs in capturing long-range dependencies. Unlike traditional methods, UNETR employs a Vision Transformer (ViT) encoder, processing 3D volumes as sequential patches, which enhances global context understanding. The model connects the transformer encoder to a CNN-based decoder via skip connections for precise segmentation. Evaluated on BTCV and MSD datasets, UNETR achieved 0.856 Dice in BTCV multi-organ segmentation, 0.964 in spleen segmentation, and 0.711 in brain tumor segmentation, outperforming state-of-the-art CNN and hybrid models. These results highlight UNETR's superior performance in volumetric medical image segmentation, establishing it as a strong candidate for future transformer-based segmentation models.

Chen et al. [23] introduced TransUNet, a hybrid model combining CNNs and Transformers for medical image segmentation. The study aimed to overcome CNNs' limitations in capturing long-range dependencies while maintaining precise localization through U-Net-like skip connections. Evaluated on the Synapse multi-organ CT dataset, TransUNet achieved 77.48% Dice score, outperforming CNN-based and transformer-only methods. Similarly, on the ACDC cardiac segmentation dataset, it achieved 89.71% Dice score, surpassing competing models. These results highlight TransUNet's effectiveness in balancing global context understanding with fine-grained spatial details, making it a strong alternative to traditional FCN-based segmentation models.

Cao et al. [20] introduced SwinUNETR, a U-Net-like architecture incorporating Swin Transformer blocks to enhance medical image segmentation. The study aimed to overcome the locality limitations of CNNs while reducing the high computational cost of standard Transformers. Unlike conventional U-Net models, SwinUNETR leverages hierarchical shifted window attention to capture both local and global dependencies efficiently. The model was evaluated on Synapse multi-organ segmentation (CT) and ACDC cardiac segmentation (MRI) datasets. SwinUNETR achieved a Dice similarity coefficient (DSC) of 79.13% on Synapse and 90% on ACDC,

outperforming CNN-based U-Net variants and Transformer-based architectures. These results highlight SwinUNETR's effectiveness in balancing spatial precision and computational efficiency, making it a promising alternative for high-accuracy medical image segmentation.

Isensee et al. [18] introduced nnU-Net, a self-configuring deep learning framework for biomedical image segmentation, addressing the challenge of manually optimizing deep learning models for diverse datasets. Unlike conventional approaches, nnU-Net automatically adapts its preprocessing, network architecture, training strategies, and post-processing to any given segmentation task. It systematically categorizes parameters into fixed, rule-based, and empirical decisions, reducing the need for expert intervention. nnU-Net was extensively tested on 23 public datasets across 53 segmentation tasks, achieving state-of-the-art performance in most cases. Notably, it outperformed highly specialized models in numerous international biomedical segmentation challenges. This study demonstrates nnU-Net's effectiveness as an out-of-the-box solution, making high-quality segmentation accessible without requiring expert knowledge or extensive computational resources.

Recently, models such as U-Mamba [24] and SegMamba [25], called the Mamba family, have also attracted considerable attention in segmentation studies. Ma et al. [24] introduced U-Mamba, a hybrid CNN-State Space Model (SSM) architecture for biomedical image segmentation, aiming to enhance long-range dependency modeling while maintaining computational efficiency. Unlike CNNs, which struggle with global context, and Transformers, which are computationally expensive, U-Mamba integrates Mamba blocks (a variant of SSMs) to efficiently capture both local and global features. The model employs a self-configuring mechanism, similar to nnU-Net, allowing it to automatically adapt to various datasets without manual tuning. Evaluated on four diverse segmentation tasks—3D abdominal CT and MRI segmentation, endoscopy instrument segmentation, and microscopy cell segmentation—U-Mamba consistently outperformed state-of-the-art CNN-based (nnU-Net, SegResNet) and Transformer-based (UNETR, SwinUNETR)

models. Notably, in 3D abdominal CT segmentation, U-Mamba achieved a Dice score of 0.8683, surpassing nnU-Net (0.8615), and in 3D MRI segmentation, it achieved 0.8501, outperforming SwinUNETR and UNETR. These results demonstrate U-Mamba's ability to balance computational efficiency with high segmentation accuracy, positioning it as a promising alternative to existing deep learning architectures in biomedical imaging.

Xing et al. [25] introduced SegMamba, a novel 3D medical image segmentation model that integrates Mamba-based State Space Models (SSMs) to enhance long-range dependency modeling while maintaining computational efficiency. Unlike CNN-based methods, which struggle with global context, and Transformer-based models, which suffer from high computational costs, SegMamba employs Tri-orientated Spatial Mamba (ToM) blocks to effectively capture global information in volumetric medical images. The model also incorporates a Gated Spatial Convolution (GSC) module for improved spatial feature representation and a Feature-level Uncertainty Estimation (FUE) module to refine multi-scale feature integration. Evaluated on BraTS 2023 (brain tumor segmentation), AIIB 2023 (airway segmentation), and CRC-500 (colorectal cancer segmentation) datasets, SegMamba achieved Dice scores of 91.32% on BraTS 2023, 88.59% on AIIB 2023, and 48.02% on CRC-500, outperforming state-of-the-art CNN and Transformer models. These results demonstrate SegMamba's ability to efficiently model long-range dependencies while maintaining high segmentation accuracy, positioning it as a strong alternative to existing deep learning architectures in medical imaging.

Automatic preprocessing and data augmentation techniques also play an important role in 3D medical image segmentation. Zhao et al. [26] proposed a learning-based data augmentation method to address the challenge of one-shot medical image segmentation, where only a single labeled scan is available. Unlike traditional augmentation techniques that rely on random transformations, this method learns spatial and intensity transformations from unlabeled medical images and applies them to generate realistic synthetic training examples. The model captures anatomical and imaging variations by learning spatial deformation fields

and intensity mappings, enabling robust augmentation beyond simple rotations or flips. Evaluated on MRI brain segmentation, the proposed method significantly outperformed state-of-the-art one-shot segmentation approaches, including single-atlas segmentation and traditional augmentation-based supervised segmentation. The study demonstrated that using learned transformations improved Dice scores by up to 0.056, bringing performance closer to fully supervised models while requiring only minimal labeled data. These results highlight the potential of learning-based augmentation to enhance segmentation accuracy in low-data medical imaging scenarios.

Despite significant advancements, 3D medical image segmentation continues to face challenges, particularly in achieving high accuracy for small and complex anatomical structures. CNN-based models excel at capturing local spatial details but struggle with modeling long-range dependencies. On the other hand, Transformer-based architectures address this limitation effectively; however, they often demand substantial computational resources and large amounts of labeled data. Hybrid approaches, such as U-Mamba and SegMamba, propose alternative mechanisms to balance these limitations, yet their generalizability across diverse datasets remains uncertain.

In light of these challenges, this study aims to bridge existing gaps by developing a segmentation framework that balances local and global feature extraction while optimizing computational efficiency. Specifically, this research focuses on improving 3D abdominal organ segmentation performance by leveraging advanced deep learning models and tailored preprocessing techniques. Considering the inherent difficulties of medical image segmentation, such as anatomical variability, image noise, and limited training datasets, this study contributes to the field by systematically evaluating the effectiveness of different segmentation methodologies. By addressing key limitations in current approaches, this work provides valuable insights for optimizing deep learning architectures for medical imaging applications.

3. MATERIAL AND METHOD

This study comparatively evaluates the performances of three widely used deep learning architectures for 3D medical image segmentation: 3D U-Net [16], UNETR [19], and SwinUNETR [20]. The impact of image preprocessing, post-processing, and loss

functions on the performance of these models is systematically examined. The objective is to identify the optimal configuration that maximizes segmentation accuracy. The overall study workflow is illustrated in Figure 1.



Figure 1. Flowchart of the Study

3.1. Dataset Used

The study utilizes the BTCV (Beyond the Cranial Vault) dataset [27], which was specifically designed for abdominal organ segmentation. The dataset was created for the "Multi-Atlas Labeling Beyond the Cranial Vault" competition held at the MICCAI 2015 conference. The images consist of 30 contrast-enhanced abdominal and pelvic CT scans. Each scan consists of 85 to 198 slices. The slice thickness is 2.5 mm, and the slice interval is 2.5 mm. The image size is 512 x 512 pixels. All scans were taken in the portal venous contrast phase. HU (Hounsfield Unit) values are in the

range of [-1024, 3071]. It is in NIFTI (.nii.gz) file format. Each CT scan in the dataset contains reference images manually segmented by expert radiologists. There are 13 segmented organs: spleen, right kidney, left kidney, gallbladder, esophagus, liver, stomach, aorta, inferior vena cava, portal vein and splenic vein, pancreas, retinitis adrenal gland, left adrenal gland. In the study, 24 of these 30 CT images were used for training, while six were used for validation and testing. Figure 2 shows a sample image from the dataset and the corresponding labeled image.

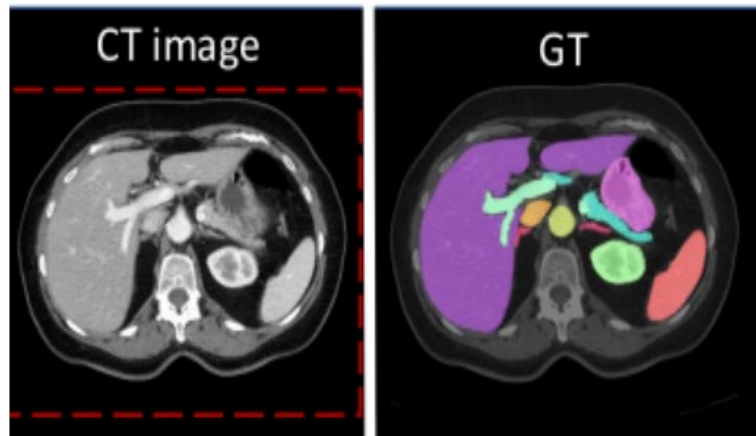


Figure 2. Dataset Sample Image and Ground Truth Image.

3.2. Applied Deep Learning Models

The study trained and compared three different deep-learning architectures with the dataset. First, the 3D U-Net architecture proposed by Çiçek et al. [16] was used. 3D U-Net was explicitly designed for volumetric medical image segmentation by extending the traditional U-Net architecture with three-dimensional convolutions. This architecture comprises an encoder with consecutive 3D convolutional layers, ReLU activations, and max pooling

operations, and a decoder with upsampling layers followed by 3D convolutions and ReLU activations. At each downsampling step, the number of feature channels is doubled. Skip connections from encoder to decoder enable low-level and high-level features to be effectively combined. This structure can produce detailed segmentation maps even in cases where there is limited training data. A standard 3D U-Net architecture is given in Figure 3. [16].

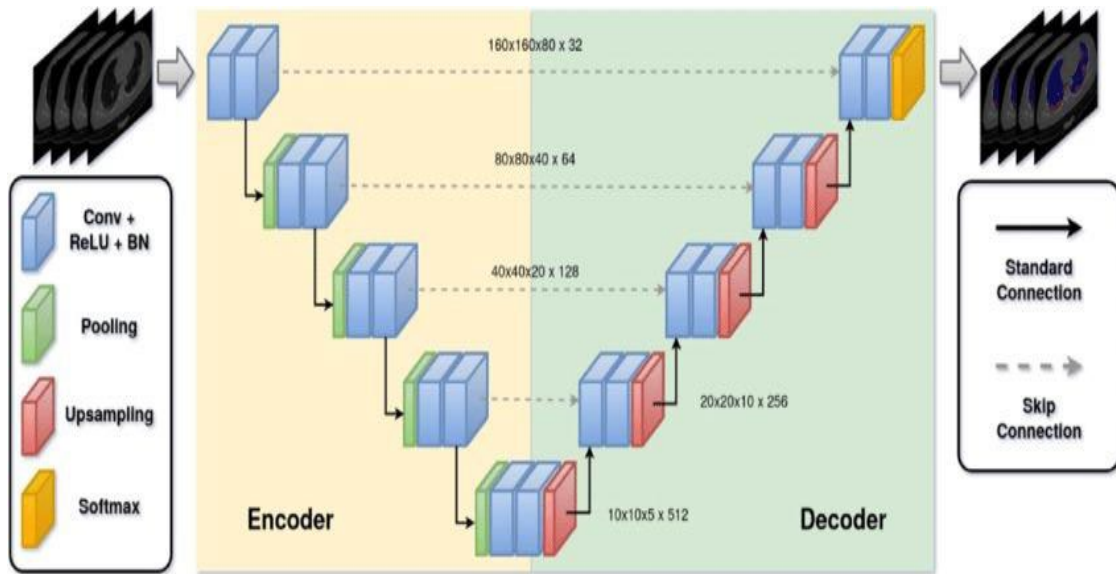


Figure 3. Standard 3D U-Net architecture

Secondly, the UNETR (UNet-Transformers) architecture developed by Hata-mizadeh et al. [19] is implemented. UNETR is a hybrid architecture that uses a transformer-based encoder and a CNN-based decoder. The encoder section first divides the 3D image into small volumetric patches with regular intervals and passes these patches through a linear projection layer. Then, position codings are added to these projections, and the resulting sequences are passed through a series of transformer blocks. Each transformer block has a multi-head self-attention mechanism and a feed-forward neural

network layer. This structure can effectively capture long-range spatial dependencies. The decoder section uses a CNN structure that combines features from different encoder layers and gradually amplifies them. This hybrid structure of UNETR combines the global context capturing ability of transformers with the local feature extraction power of CNNs, achieving successful results, especially for the segmentation of complex anatomical structures. Figure 4 shows a model showing the working flow of UNETR.

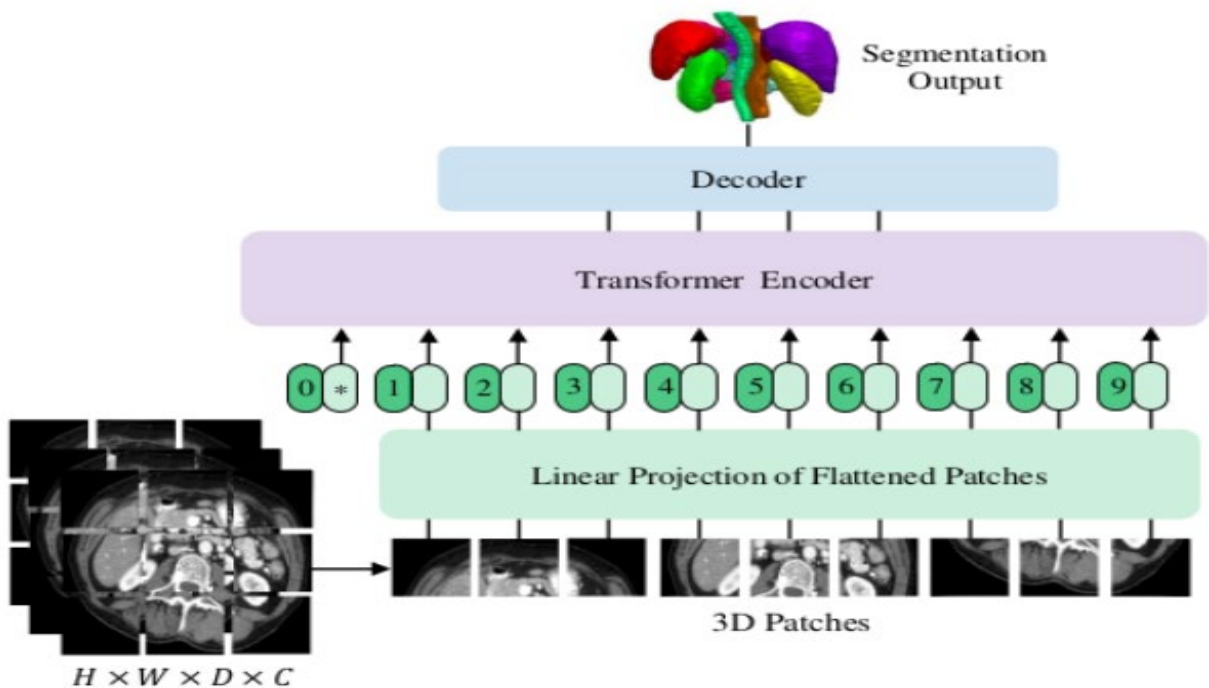


Figure 4. UNETR sample workflow [28]

Finally, the SwinUNETR architecture proposed by Cao et al.[20] is implemented. SwinUNETR is a model that uses Swin Transformer blocks in a U-Net-like architecture. The distinctive feature of the Swin Transformer is the shifted window approach. In this approach, self-attention calculations are first performed in small, non-overlapping windows, which are then shifted to establish connections between windows. This method effectively captures features at different scales while reducing computational complexity. The encoder section

of SwinUNETR consists of successive Swin Transformer blocks, and the number of channels increases while the feature resolution decreases in each block. The decoder section is a structure that gradually amplifies and combines the features coming from the encoder. This architecture is expected to perform highly, especially in abdominal organ segmentation, where extensive contextual information is important. Figure 5 shows the workflow of SwinUNETR.

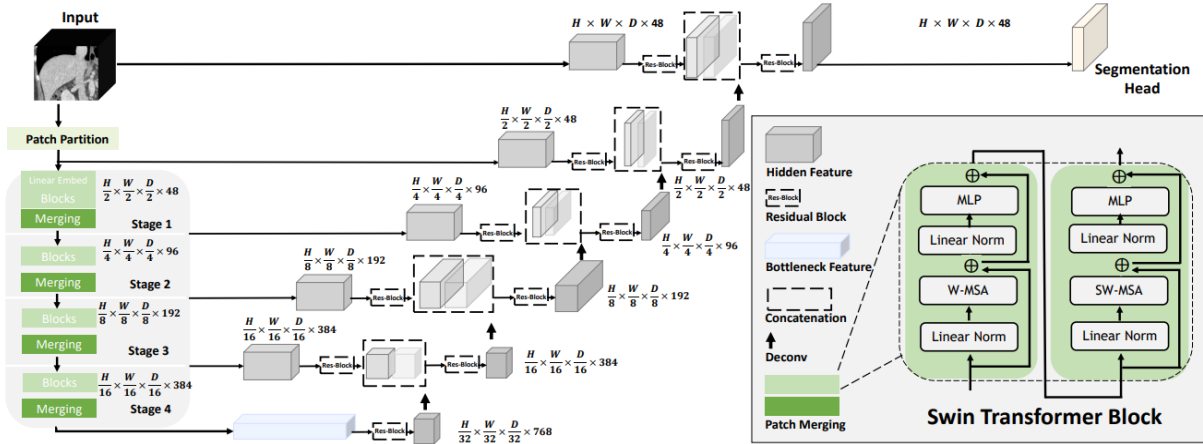


Figure 5. SwinUNETR sample study flow

3.3. Evaluated Loss Functions

The study further examined the effect of different loss functions on the model performance. The loss functions used include Dice Loss, Cross-Entropy Loss, Focal Loss, Dice+Cross-Entropy Loss, and Dice+Focal Loss.

Dice Loss was proposed by Milletari et al.[21]. Dice Loss is a loss function used primarily in medical image segmentation. It tries to maximize the overlap ratio between the estimated segmentation mask and the ground truth. Thus, it aims to increase the segmentation accuracy [29]. It is a popular choice, especially for imbalanced datasets, because it prevents the model from ignoring minority classes and can focus on the overlapping regions between the estimated and ground truth masks. The formula of Dice loss (Equation (1)) is;

$$L_{dice} = 1 - \frac{1}{c} \sum_{c=0}^{C-1} \frac{2 \sum_{n=1}^N t_n^c y_n^c}{\sum_{n=1}^N (t_n^c + y_n^c)} \quad (1)$$

Cross-entropy loss[30] is a common loss function that optimizes the correct classification

of each pixel. It measures the difference between two probability distributions for a random variable. It is used in segmentation tasks to measure how well the model's predictions match the target labels[29]. Cross-entropy loss is particularly effective in improving pixel-wise accuracy. The formula for Cross-Entropy loss (Equation (2)) is as follows;

$$L_{CE(y,t)} = - \sum_{n=1}^N \log(t_n \cdot y_n) \quad (2)$$

Focal loss, proposed by Lin et al. [31], is an improved version of the cross-entropy loss function that assigns different weights to easy and complex examples. Complex examples are those that are misclassified with a high probability, while easy examples are those that are correctly classified with a high probability. This helps balance the effect of easy and complex examples on the overall loss [29]. The formula for Focal loss (Equation (3)) is:

$$L_{focal(y,t,\gamma)} = - \sum_{n=1}^N (1 - t_n \cdot y_n)^\gamma \log(t_n \cdot y_n) \quad (3)$$

Here, γ is a non-adjustable positive hyperparameter. A flat cross-entropy loss is

obtained when γ is set to 0 for all samples. Focal loss is especially effective in cases where the imbalance between the background and the organ of interest is high. Dice + Focal loss, Focal+Cross-Entropy loss, and Dice + Cross-Entropy loss combinations were also used in the study. These hybrid loss functions aim to optimize both the overall shape similarity and pixel-based accuracy by combining the advantages of different loss functions. The formulas of these loss functions (Equations (4)-(5)) are as follows;

$$L_{DiceCE} = w_{ce} * L_{CE}(p, y) + w_{dice} * (1 - L_{Dice}(p, y)) \quad (4)$$

$$L_{DiceFocal} = L_{Focal}(p, y) + \lambda * (1 - L_{Dice}(p, y)) \quad (5)$$

The choice of loss functions directly impacts segmentation performance, particularly for imbalanced datasets. Dice Loss is preferred for its ability to mitigate class imbalance by maximizing region overlap, while Cross-Entropy Loss focuses on pixel-wise classification accuracy. Focal Loss enhances this by assigning higher weights to hard-to-classify samples, making it particularly effective for datasets where small anatomical structures are underrepresented. The study also evaluates hybrid loss functions (Dice+CE, Focal+CE, Dice+Focal) to balance spatial consistency and pixel-level accuracy

3.4. Applied Image Preprocessing Methods

The study investigated the effect of preprocessing applied to images in the dataset on the performance of trained models.

The intensity rescaling process linearly scales image intensities to a specific range. This process facilitates model training by normalizing the intensity differences between images from different scanners. Images obtained from a CT scanner have intensity values in Hounsfield Units (HU). Scale Intensity Ranged scales image intensity values to a specific range and thus reduces the intensity differences between different images.

The crop foreground process automatically determines the region containing the organ of interest by detecting non-zero voxels in the image and discards unnecessary background information. This increases computational

efficiency and allows the model to focus on the regions of interest.

The reorientation process is a preprocessing step used to align the orientation of the image and label data according to a standard coordinate system. Image orientations may differ in medical imaging due to different imaging devices or protocols. These differences can cause image processing errors and negatively impact the performance of deep learning models. This process standardizes the orientation of image and label data according to the "RAS" (Right-Anterior-Superior) coordinate system.

Spatial Resampling is a preprocessing step used to standardize an image's pixel dimensions (spacing) and label data to a certain value. In medical imaging, pixel dimensions may vary due to different imaging devices or protocols. This method reduces pixel size differences between different images by standardizing the spatial resolution of images.

Random Cropping Based on Positive and Negative Label Sampling is a data augmentation step that randomly cuts 3D patches from image and label data. The method helps alleviate the class imbalance problem by cutting equal numbers of patches from both positive (containing the target anatomical structure) and negative (not containing the target anatomical structure) regions. It was also used in the study to adjust the input dimensions.

Based on the conducted experiments, these methods have demonstrated successful outcomes and were deemed necessary at all stages of training, leading to their implementation throughout the entire process. The methods whose effects on performance are evaluated are as follows;

- Random Axis Flip is a data augmentation step that randomly flips images and labels data along specified axes. The primary purpose of this step is to expand the dataset and make the model more robust against rotational changes. Random 90-degree Rotation is a data augmentation step that randomly rotates the image and labels data by 90, 180, or 270 degrees.
- Random Intensity Shift is a data augmentation technique in which the

intensity of the image is randomly increased or decreased. It helps the model to be robust to different lighting conditions.

- Local Brightness Adjustment is a method that adjusts the brightness in some regions of an image. It is usually used to increase local details or to eliminate unbalanced lighting.
- Gaussian Noise Addition is a technique that adds random pixel changes to the image according to a Gaussian distribution. It increases the model's robustness to noisy data.
- Random Contrast Adjustment is a technique in which the difference between the bright and dark areas of the image is randomly adjusted. It increases the model's robustness to different lighting conditions.
- Gaussian Smoothing is a filter that blurs the image by averaging pixel values with their neighbors. It is used to reduce high-frequency noise.
- Gaussian Sharpening is a technique that highlights the edges and fine details in the image. It highlights essential features by increasing the pixel contrast at the edges.
- Random Cropping by Label Classes focuses on the areas of the image labeled with certain classes and performs random cropping in these parts. It has been tried as an alternative to the Random Cropping by Positive and Negative Labels method.

4. RESULTS AND DISCUSSION

The loss functions and preprocessing methods used in the study were tested on 3D U-Net due to the low computational cost, and the successful results obtained were applied to SwinUNETR and UNETR models.

First, the loss functions were evaluated. In determining the loss functions, eight pieces of $48 \times 48 \times 48$ images were created from each image for training the model. The Random Cropping Based on Positive and Negative Label Sampling algorithm was used for this. Table 1 shows the performance values obtained from 3D U-Net in different loss functions. When the table is examined, it is seen that the use of loss functions together has a positive effect on the model performance. When the performance values are examined, it is seen that the Cross-Entropy and Dice loss functions generally give the most successful results. The Cross-Entropy +

Dice(Dice+CE) loss function was used in the following steps of the study.

In the continuation of the study, the effect of input size on performance was investigated using the DiceCE loss function. For this purpose, the model was trained with different-sized image patches from each image, including eight images of $48 \times 48 \times 48$, eight images of $96 \times 96 \times 96$, and four images of $96 \times 96 \times 96$. The performance values obtained from the 3D-UNet model for different input sizes are presented in Table 2. When analyzing the obtained values, it is observed that the results are quite close to each other. However, overall, the model trained with four images of $96 \times 96 \times 96$ achieved the best performance. Experiments conducted with larger input sizes (e.g., $128 \times 128 \times 128$) resulted in excessive memory consumption, particularly with SwinUNETR. To address this issue, alternative input sizes were explored, and it was determined that $96 \times 96 \times 96$ provides the optimal balance between contextual information and computational efficiency. Similar strategies have been employed in previous studies. For instance, LoGoNet[32] adopts a patch-based segmentation approach, avoiding large input sizes to reduce computational overhead. These findings are consistent with the observations in this study, demonstrating that smaller patches improve memory efficiency while maintaining segmentation accuracy. Since achieving maximum performance is the priority in this study, all subsequent image preprocessing methods were applied using the most optimal input size of $96 \times 96 \times 96$.

In the continuation of the study, the effects of image preprocessing methods on model performance were examined. First, data augmentation methods were applied. For this, random and 90-degree rotation operations were applied to the images on the axes. As a result of the positive effect of this operation on performance, this operation was added to all subsequent preprocesses. The following operations differed in each training, and their effects on performance were examined. Table 3 shows the applied methods and their effects on segmentation studies. The numbers given to the applied preprocesses are as follows;

1. Data Augmentation(to increase training diversity and prevent overfitting)

- Random Flip (to introduce rotational invariance and improve robustness to anatomical variations)
 - 90-degree Rotation (to simulate different scanning orientations and improve generalization)
2. Local Brightness Adjustment (to compensate for scanner-specific intensity variations)
 3. Gaussian Noise Addition (to make the model more robust to imaging noise and artifacts)
 4. Random Contrast Adjustment (to enhance organ boundaries and improve segmentation clarity)
 5. Gaussian Smoothing (to reduce high-frequency noise and improve stability in predictions)
 6. Gaussian Sharpening (to enhance edge clarity and refine organ contours)
 7. Random Cropping by Label Classes (to focus learning on underrepresented structures and balance class distributions)

Table 3 confirms that data augmentation techniques significantly enhance segmentation accuracy by improving generalization. Additionally, Gaussian Sharpening and Random Contrast Adjustment contributed to performance gains by enhancing edge clarity and improving organ boundary delineation. However, Random Cropping by Label Classes produced inconsistent results, suggesting that targeted cropping does not always benefit segmentation performance. According to these findings, the most effective preprocessing methods were selected and applied to SwinUNETR and UNETR. The final preprocessing pipeline consisted of:

- Random Flip (to introduce rotational invariance)
- Random Intensity Shift (to improve robustness against intensity variations)
- Gaussian Sharpening (to enhance boundary clarity)

While preprocessing techniques generally improved segmentation performance, some models showed a slight decline in Dice scores of some classes after applying certain transformations. This effect was more prominent in Transformer-based architectures

such as SwinUNETR and UNETR, which are highly sensitive to intensity and contrast variations. Unlike CNN-based models, which primarily rely on local spatial details, Transformers incorporate long-range dependencies, making them more vulnerable to alterations in intensity distribution caused by preprocessing steps. Additionally, some augmentations, such as Gaussian Sharpening and Contrast Adjustment, may have unintentionally altered the natural organ boundaries, leading to minor segmentation inconsistencies in certain cases. These findings indicate that preprocessing strategies should be carefully tailored for different model architectures to ensure optimal performance.

The results indicate that 3D U-Net achieved the best trade-off between segmentation accuracy and computational efficiency. Although SwinUNETR produced comparable results to 3D U-Net, its training time was significantly longer (7298 sec vs. 3555 sec). Similarly, UNETR required 5564 sec for training but yielded slightly lower Dice scores. These findings suggest that CNN-based models like 3D U-Net are more computationally efficient than Transformer-based architectures (SwinUNETR and UNETR) for this segmentation task.

The comparative training times for each model were as follows:

- SwinUNETR: 7298.58 sec
- UNETR: 5564.40 sec
- 3D U-Net: 3555.69 sec

These results reinforce that while transformer-based models can capture long-range dependencies, they require significantly higher computational resources compared to CNN-based models.

Table 5 compares the segmentation performance of several methods, including 3D U-Net, Swin-UNETR, and UNETR tested in this study, against reference methods reported in the literature for 13 abdominal organs. The metrics are Dice scores, which indicate segmentation accuracy, and the last column presents the average performance across all organs.

The performance variability across organ types suggests that no single architecture universally outperforms others. Instead, combining specialized architectures, loss functions, and preprocessing techniques tailored to specific organs may yield optimal results. The findings reinforce the need to develop more robust methods for handling small organ segmentation, which is critical for applications requiring detailed anatomical analysis. This detailed comparison highlights the strengths and limitations of various deep learning models for 3D abdominal organ segmentation. The results show that CNN-based models such as 3D U-Net are computationally efficient and achieve competitive accuracy, while Transformer-based models such as UNETR are superior in capturing complex anatomical structures. However, the critical difference between the dice values of the UNETR model trained and tested on the same dataset by different researchers shows that the model has some problems with stabilization.

Also, the proposed 3D U-Net model demonstrates superior performance in pancreas segmentation, achieving a Dice score of 0.823, the highest reported value among all compared methods. This result highlights the effectiveness of the applied preprocessing techniques and model architecture in handling small and low-contrast structures. Given that pancreas segmentation remains a challenging task due to its anatomical variability and low contrast with surrounding tissues, this improvement is particularly significant.

In conclusion, the presented comparison table shows that all methods exhibit difficulties in segmentation of small organs and underlines the need for future research focusing on region-specific improvements and advanced magnification techniques.

Figure 6 presents sample segmentation results from the trained models. A detailed qualitative analysis was performed to evaluate segmentation accuracy, organ boundary preservation, and common misclassification patterns.

3D U-Net and SwinUNETR produced visually similar segmentation results, while UNETR showed slightly lower segmentation accuracy, particularly for small anatomical structures. All models generated false-positive organ segmentations in regions where no ground truth annotations exist. This issue was more pronounced in UNETR, suggesting that transformer-based architectures may introduce excessive spatial dependencies, leading to misclassifications in low-contrast areas.

3D U-Net exhibited the most stable segmentation boundaries across different organs, whereas SwinUNETR tended to capture more detailed structures at the cost of minor over-segmentation artifacts. Small organs such as the adrenal glands and gallbladder remained the most difficult to segment accurately across all models. The reduced contrast in these structures, along with their small size, likely contributed to the under-segmentation observed in multiple cases.

Table 1. The Effect of Loss Functions on Model Performance (a:IoU, b: Recall, c:Precision, d: Dice)

Loss Functions	Parameters	Background	Spleen	Right Kidney	Left Kidney	Gallbladder	Esophagus	Liver	Stomach	Aorta	Inferior Vena Cava	Portal Vein and Splenic Vein	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Average
Cross Entropy Loss	d	0.9964	0.9018	0.918	0.9029	0.5597	0.6582	0.9588	0.8267	0.8779	0.7961	0.6769	0.7789	0.4731	0.5093	0.7739
	c	0.995	0.8771	0.9625	0.9222	0.7424	0.8081	0.9685	0.9157	0.9107	0.8827	0.7693	0.8277	0.7472	0.6825	0.858
	b	0.9978	0.9345	0.8785	0.8873	0.4677	0.5981	0.9495	0.7653	0.8501	0.7284	0.6133	0.7387	0.3736	0.4247	0.7291
	a	0.9928	0.8244	0.849	0.8259	0.358	0.497	0.9209	0.7099	0.7827	0.6625	0.5145	0.6382	0.3151	0.348	0.6599
Focal Loss	d	0.9963	0.9225	0.9272	0.9276	0.5991	0.5814	0.9467	0.8304	0.863	0.7898	0.6187	0.6935	0.4422	0.4208	0.7542
	c	0.9957	0.9041	0.9572	0.9581	0.595	0.8531	0.9281	0.8726	0.9307	0.8371	0.833	0.9215	0.702	0.7663	0.861
	b	0.9969	0.9455	0.8993	0.8992	0.5402	0.4602	0.9666	0.8009	0.8101	0.7565	0.5076	0.5596	0.3665	0.3117	0.7015
	a	0.9926	0.8579	0.8643	0.8653	0.4001	0.4161	0.8991	0.7165	0.7599	0.6545	0.4548	0.5339	0.2984	0.2718	0.6418
Dice Loss	d	0.9964	0.937	0.927	0.9183	0.5378	0.6603	0.9566	0.8234	0.8686	0.7959	0.6041	0.7453	0.4716	0.505	0.7677
	c	0.9952	0.9287	0.942	0.959	0.5723	0.7832	0.955	0.9099	0.9359	0.8609	0.8075	0.7945	0.6954	0.5807	0.8371
	b	0.9977	0.9476	0.9135	0.8812	0.5245	0.6085	0.9584	0.7689	0.8138	0.7457	0.4896	0.7079	0.3897	0.4838	0.7308
	a	0.9929	0.8824	0.8642	0.8492	0.3295	0.4945	0.9169	0.7086	0.7683	0.6618	0.4366	0.5965	0.3142	0.3439	0.6542
Focal+Cross Entropy Loss	d	0.9929	0.6477	0.7494	0.7673	0.4939	0.5735	0.8859	0.781	0.7926	0.5868	0.3156	0.6624	0.5145	0.4933	0.6612
	c	0.9901	0.5431	0.9908	0.9875	0.4435	0.7809	0.9678	0.8732	0.9389	0.9131	0.8389	0.8124	0.6252	0.5405	0.8033
	b	0.9957	0.8583	0.6095	0.6324	0.5022	0.4843	0.8253	0.7208	0.6983	0.4431	0.2066	0.5932	0.4763	0.4898	0.6097
	a	0.9859	0.4951	0.6058	0.6265	0.3067	0.4107	0.8026	0.646	0.6599	0.4225	0.1977	0.4991	0.3503	0.3395	0.5249
Dice + Cross Entropy Loss	d	0.997	0.9386	0.9357	0.9331	0.6518	0.6768	0.9617	0.8592	0.8855	0.8042	0.6342	0.7931	0.6406	0.5347	0.8033
	c	0.9965	0.9481	0.9386	0.939	0.5291	0.8206	0.9608	0.8687	0.9345	0.8958	0.8079	0.8302	0.657	0.7319	0.847
	b	0.9975	0.9307	0.9333	0.9273	0.7343	0.6042	0.9628	0.8536	0.8462	0.7336	0.5352	0.7643	0.6352	0.4348	0.7781
	a	0.994	0.8851	0.8793	0.875	0.4265	0.5133	0.9263	0.759	0.7956	0.6737	0.4721	0.658	0.4727	0.3714	0.693
Dice +Focal Loss	d	0.9966	0.903	0.8746	0.8921	0.5182	0.6599	0.9539	0.8022	0.8973	0.8488	0.7221	0.7824	0.6642	0.6143	0.795
	c	0.9951	0.9068	0.9379	0.8561	0.7717	0.8963	0.9709	0.9112	0.933	0.8682	0.7468	0.817	0.7331	0.7099	0.861
	b	0.9981	0.9037	0.833	0.9401	0.4236	0.5275	0.9382	0.7494	0.8661	0.834	0.7032	0.7568	0.6169	0.5493	0.76
	a	0.9933	0.8256	0.7886	0.8098	0.3389	0.4978	0.9122	0.6919	0.8149	0.7381	0.5662	0.6443	0.4979	0.4552	0.6839

Table 2. The Effect of Input Size on Model Performance (a:IoU, b: Recall, c:Precision, d: Dice)

Input Size	Parameter	Background	Spleen	Right Kidney	Left Kidney	Gallbladder	Esophagus	Liver	Stomach	Aorta	Inferior Vena Cava	Portal Vein and Splenic Vein	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Average
48x48x48x8	d	0.997	0.9386	0.9357	0.9331	0.6518	0.6767	0.9617	0.8592	0.8855	0.8042	0.6342	0.7931	0.6406	0.5347	0.8033
	c	0.9965	0.9481	0.9386	0.939	0.5291	0.8205	0.9608	0.8687	0.9345	0.8958	0.8079	0.8302	0.657	0.7319	0.847
	b	0.9975	0.9307	0.9333	0.9273	0.7344	0.6042	0.9628	0.8536	0.8462	0.7336	0.5352	0.7643	0.6352	0.4348	0.7781
	a	0.994	0.8851	0.8793	0.875	0.4266	0.5133	0.9263	0.759	0.7956	0.6737	0.4721	0.658	0.4727	0.3714	0.693
96x96x96x8	d	0.9964	0.9195	0.9301	0.929	0.6866	0.672	0.9586	0.8003	0.8864	0.8113	0.6901	0.7644	0.5824	0.5664	0.7995
	c	0.9951	0.9212	0.9402	0.9305	0.5609	0.8133	0.9691	0.8883	0.9148	0.8337	0.7266	0.825	0.7683	0.6444	0.838
	b	0.9977	0.9221	0.9212	0.9279	0.7131	0.5797	0.9486	0.7604	0.8616	0.7999	0.6649	0.7204	0.4731	0.5277	0.7727
	a	0.9928	0.8527	0.8695	0.8677	0.4635	0.5069	0.9206	0.6828	0.7967	0.6844	0.5276	0.6204	0.4125	0.4028	0.6858
96x96x96x4	d	0.9968	0.9483	0.9389	0.9367	0.7448	0.7035	0.9648	0.8189	0.9048	0.8406	0.7201	0.7963	0.664	0.6392	0.8298
	c	0.9956	0.9635	0.9337	0.9329	0.7352	0.8765	0.9662	0.9297	0.9189	0.8733	0.771	0.8561	0.7304	0.7755	0.8756
	b	0.9981	0.934	0.9452	0.9408	0.6695	0.5969	0.9636	0.7588	0.8927	0.8149	0.6795	0.7512	0.6115	0.5517	0.7934
	a	0.9936	0.9018	0.8851	0.8813	0.5078	0.5449	0.9321	0.711	0.8264	0.726	0.5645	0.6632	0.4976	0.4749	0.7222

Table 3. The Effect of Image Preprocessing on Model Performance

Preprocess	Background	Spleen	Right Kidney	Left Kidney	Gallbladder	Esophagus	Liver	Stomach	Aorta	Inferior Vena Cava	Portal Vein and Splenic Vein	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Average
Unprocessed	0.9968	0.9483	0.9389	0.9367	0.7448	0.7035	0.9648	0.8189	0.9048	0.8406	0.7201	0.7963	0.664	0.6392	0.8298
1	0.9969	0.9538	0.9355	0.937	0.7731	0.7528	0.964	0.8423	0.9036	0.8584	0.719	0.7769	0.6674	0.6304	0.8365
1_7	0.9966	0.954	0.9328	0.936	0.7478	0.7539	0.9613	0.8193	0.8945	0.8153	0.6988	0.7443	0.6495	0.6124	0.8226
1_3	0.9967	0.9312	0.9222	0.9289	0.7228	0.7435	0.957	0.8298	0.9008	0.8397	0.6678	0.7709	0.6773	0.6741	0.8259
1_2	0.9967	0.947	0.9361	0.9347	0.7118	0.74	0.9602	0.8138	0.8876	0.8529	0.7183	0.7928	0.6655	0.6155	0.8266
1_4	0.9969	0.9521	0.9384	0.9349	0.7402	0.7436	0.9622	0.8372	0.8904	0.8542	0.7119	0.8056	0.6491	0.6276	0.8317
1_5	0.9966	0.903	0.8746	0.8921	0.5182	0.6599	0.9539	0.8022	0.8973	0.8488	0.7221	0.7824	0.6642	0.6143	0.795
1_4_5	0.9967	0.9492	0.9343	0.9374	0.6315	0.7318	0.9581	0.8024	0.9006	0.8434	0.7254	0.816	0.6451	0.5896	0.8187
1_6	0.9968	0.9483	0.9389	0.9367	0.7448	0.7035	0.9648	0.8189	0.9048	0.8406	0.7201	0.7963	0.664	0.6392	0.8298
1_4_6	0.9969	0.9538	0.9355	0.937	0.7731	0.7528	0.964	0.8423	0.9036	0.8584	0.719	0.7769	0.6674	0.6304	0.8365

Table 4. Comparison of Model Dice Performance

Model	Preprocess	Background	Spleen	Right Kidney	Left Kidney	Gallbladder	Esophagus	Liver	Stomach	Aorta	Inferior Vena Cava	Portal Vein and Splenic Vein	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Average
3D U-Net	Unprocessed	0.9968	0.9483	0.9389	0.9367	0.7448	0.7035	0.9648	0.8189	0.9048	0.8406	0.7201	0.7963	0.664	0.6392	0.8298
	Pre-Processed	0.997	0.9595	0.9434	0.9422	0.7225	0.7419	0.9664	0.7942	0.894	0.8456	0.7604	0.8232	0.7126	0.6522	0.8397
SwinUNETR	Unprocessed	0.9968	0.9447	0.9396	0.9373	0.6359	0.7043	0.9632	0.8009	0.8919	0.8462	0.7325	0.7958	0.6401	0.591	0.8157
	Pre-Processed	0.9969	0.9508	0.9379	0.9356	0.6812	0.7154	0.9657	0.8247	0.8872	0.8493	0.7306	0.8182	0.6891	0.6305	0.8295
UNETR	Unprocessed	0.9966	0.903	0.8746	0.8921	0.5182	0.6599	0.9539	0.8022	0.8973	0.8488	0.7221	0.7824	0.6642	0.6143	0.791
	Pre-Processed	0.9961	0.8326	0.9279	0.9161	0.6653	0.7332	0.9487	0.7942	0.8655	0.8147	0.6932	0.7489	0.6668	0.6057	0.8006

Table 5. Quantitative comparisons of the performance of segmentation studies with basic models on the BTCV dataset in the literature

Methods	Referans	Spleen	Right Kidney	Left Kidney	Gallbladder	Esophagus	Liver	Stomach	Aorta	Inferior Vena Cava	Portal Vein and Splenic Vein	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Average
SETR+PUP [33]		0,929	0,893	0,892	0,649	0,764	0,954	0,822	0,869	0,742	0,715	0,714	0,618		0,797
nnUNet [18]		0,942	0,894	0,910	0,704	0,723	0,948	0,824	0,877	0,782	0,720	0,680	0,616		0,802
ASPP [34]		0,935	0,892	0,914	0,689	0,760	0,953	0,812	0,918	0,807	0,695	0,720	0,629		0,811
TransUNet [23]		0,952	0,927	0,929	0,662	0,757	0,969	0,889	0,920	0,833	0,791	0,775	0,637		0,838
UNETR [19]		0,968	0,924	0,941	0,750	0,766	0,971	0,913	0,890	0,847	0,788	0,767	0,741		0,856
UNETR [32]		0,912	0,940	0,938	0,693	0,690	0,954	0,754	0,891	0,830	0,703	0,734	0,660	0,577	0,790
SwinUNETR [32]		0,952	0,947	0,945	0,790	0,770	0,963	0,755	0,901	0,850	0,771	0,760	0,702	0,659	0,828
nnUNet [32]		0,859	0,944	0,924	0,796	0,755	0,960	0,781	0,894	0,849	0,756	0,776	0,675	0,663	0,818
3D-Unet Ours		0,960	0,943	0,942	0,723	0,742	0,966	0,794	0,894	0,846	0,760	0,823	0,713	0,652	0,840
SwinUNETR Ours		0,951	0,938	0,936	0,681	0,715	0,966	0,825	0,887	0,849	0,731	0,818	0,689	0,631	0,830
UNETR Ours		0,833	0,928	0,916	0,665	0,733	0,949	0,794	0,866	0,815	0,693	0,749	0,667	0,606	0,801

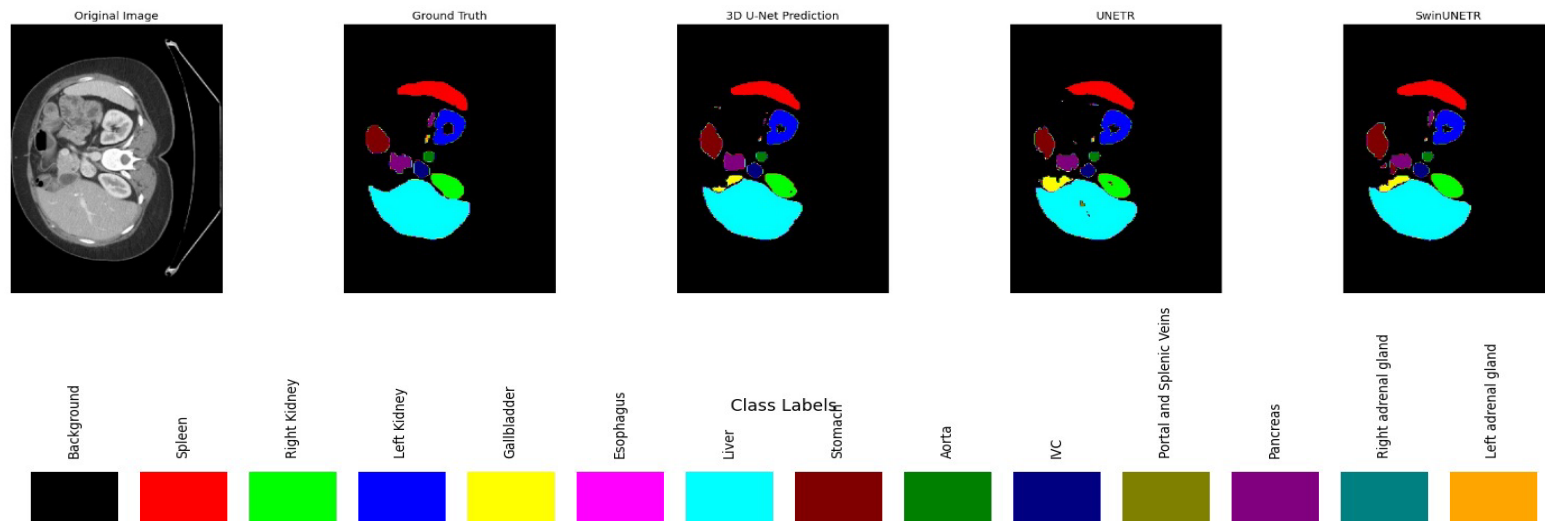


Figure 6. Segmentation results of the model

5. CONCLUSION AND FUTURE WORKS

Advancements in 3D medical imaging technology have significantly improved diagnosis, treatment planning, and patient monitoring. However, these advancements also introduce challenges in processing and analyzing large-scale medical data. Artificial intelligence, particularly deep learning-based approaches, has emerged as a transformative tool in medical imaging, demonstrating the ability to automate and enhance segmentation tasks, thereby reducing manual effort and improving diagnostic precision.

This study evaluated three deep learning architectures—CNN-based 3D U-Net, hybrid SwinUNETR, and Transformer-based UNETR—using the BTCV dataset for the segmentation of 13 different abdominal organs. Additionally, the impact of various preprocessing techniques and loss functions on model performance was analyzed. The findings highlight the importance of model architecture selection, preprocessing strategies, and loss function optimization in achieving high segmentation accuracy.

Among the models tested, 3D U-Net outperformed SwinUNETR and UNETR in both segmentation accuracy and efficiency. The highest Dice score was achieved with the $96 \times 96 \times 96$ input size, balancing spatial context and memory constraints effectively. This result aligns with previous literature, reinforcing that moderate patch sizes maintain both computational feasibility and sufficient anatomical context. Preprocessing techniques such as Random Flip, Intensity Shift, and Gaussian Sharpening improved segmentation performance, enhancing robustness to variations in image acquisition. Despite these optimizations, the segmentation of small, mobile organs such as the adrenal glands and gallbladder remained a challenge, primarily due to their low contrast, anatomical variability, and limited training samples.

This study makes a significant contribution to the field of 3D medical image segmentation by systematically evaluating different architectures and preprocessing techniques, offering insights into their comparative advantages. This research provides a thorough evaluation of multiple architectures, preprocessing techniques, and input size optimization,

offering a detailed analysis of their combined effects on segmentation performance. By systematically assessing these factors, this study highlights key elements that contribute to improved segmentation accuracy and computational efficiency. Additionally, it bridges the gap between CNN-based and Transformer-based models, highlighting their respective strengths and limitations.

Although this study achieved promising results, several areas require further investigation. Small organ segmentation continues to be difficult due to factors such as low contrast, shape variability, and data imbalance. Future research should explore specialized refinement techniques tailored to small organs, such as region-aware loss functions or attention mechanisms, to improve segmentation accuracy. Additionally, incorporating multi-scale feature extraction approaches could provide finer detail representation, enhancing performance across different organ sizes.

Further research could explore several enhancements to mitigate false positives and improve the segmentation of small organs. Developing specialized segmentation refinements for challenging organs, such as the adrenal glands and gallbladder, may enhance model precision in these regions. Implementing anatomical and structural constraints in the segmentation process could improve the delineation of organ boundaries and reduce false positives. Additionally, region-aware data augmentation techniques could enhance model robustness for small anatomical structures by simulating realistic variations in medical imaging. Exploring more effective training paradigms, such as curriculum learning or self-supervised learning, could improve segmentation performance, particularly in underrepresented organ classes.

Expanding the dataset with more diverse imaging modalities, including MRI and PET scans, could improve model generalizability across different clinical applications. Integrating domain adaptation techniques or contrastive learning methods may further improve segmentation performance in cross-domain applications. The development of self-supervised learning frameworks could reduce reliance on extensive annotated datasets while maintaining model robustness.

By addressing these challenges, future advancements in deep learning-based 3D segmentation will help bridge the gap between automated medical image analysis and real-world clinical implementation. This study contributes to the growing body of literature by demonstrating the effectiveness of deep learning models in multi-organ segmentation while outlining key areas for further development, ultimately supporting improved patient outcomes and assisting medical professionals in their decision-making processes.

ACKNOWLEDGES

This research was supported by the TÜBİTAK 2211-A and Afyon Kocatepe University BAP-24.FEN.BİL.02. Additionally, this article is derived from the PhD thesis titled 'Segmentation of Three-Dimensional Abdominal CT Images with Deep Learning Methods'.

REFERENCES

1. Rueckert D. and Schnabel J. A., "Model-Based and Data-Driven Strategies in Medical Image Computing," *Proceedings of the IEEE*, Vol. 108, Issue 1, Pages 110–124, 2020.
2. Wachinger C., Reuter M. and Klein T., "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, Vol. 170, Pages 434–445, 2018.
3. Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., van der Laak J. A. W. M., van Ginneken B. and Sánchez C. I., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, Vol. 42, Pages 60–88, 2017.
4. Lecun Y., Bengio Y. and Hinton G., "Deep learning," *Nature*, Vol. 521, Issue 7553, Pages 436–444, 2015.
5. Shen D., Wu G. and Suk H. Il, "Deep Learning in Medical Image Analysis," *Annual Review of Biomedical Engineering*, Vol. 19, Issue Volume 19, 2017, Pages 221–248, 2017.
6. Greenspan H., Van Ginneken B. and Summers R. M., "Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique," *IEEE Transactions on Medical Imaging*, Vol. 35, Issue 5, Pages 1153–1159, 2016.
7. Tajbakhsh N., Jeyaseelan L., Li Q., Chiang J. N., Wu Z. and Ding X., "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Medical Image Analysis*, Vol. 63, Page 101693, 2020.
8. Taha A. A. and Hanbury A., "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Medical Imaging*, Vol. 15, Issue 1, Pages 1–28, 2015.
9. Asgari Taghanaki S., Abhishek K., Cohen J. P., Cohen-Adad J. and Hamarneh G., "Deep semantic segmentation of natural and medical images: a review," *Artificial Intelligence Review*, Vol. 54, Issue 1, Pages 137–178, 2021.
10. Pham D. L., Xu C. and Prince J. L., "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, Vol. 2, Issue 2000, Pages 315–337, 2000.
11. Sharma N., Ray A. K., Shukla K. K., Sharma S., Pradhan S., Srivastva A. and Aggarwal L., "Automated medical image segmentation techniques," *Journal of Medical Physics*, Vol. 35, Issue 1, Pages 3–14, 2010.
12. Norouzi A., Rahim M. S. M., Altameem A., Saba T., Rad A. E., Rehman A. and Uddin M., "Medical Image Segmentation Methods, Algorithms, and Applications," *IETE Technical Review*, Vol. 31, Issue 3, Pages 199–213, 2014.
13. Ronneberger O., Fischer P. and Brox T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9351, Pages 234–241, 2015.
14. Roth H. R., Shen C., Oda H., Sugino T., Oda M., Hayashi Y., Misawa K. and Mori K., "A Multi-scale Pyramid of 3D Fully Convolutional Networks for Abdominal Multi-organ Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 11073 LNCS, Pages 417–425, 2018.
15. Hesamian M. H., Jia W., He X. and Kennedy P., "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *Journal of Digital Imaging*, Vol. 32, Issue 4, Pages 582–596, 2019.

16. Çiçek Ö., Abdulkadir A., Lienkamp S. S., Brox T. and Ronneberger O., “3D U-net: Learning dense volumetric segmentation from sparse annotation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9901 LNCS, Pages 424–432, 2016.
17. Chen S., Roth H., Dorn S., May M., Cavallaro A., Lell M. M., Kachelrieß M., Oda H., Mori K. and Maier A., “Towards Automatic Abdominal Multi-Organ Segmentation in Dual Energy CT using Cascaded 3D Fully Convolutional Network,” 2017.
18. Isensee F., Jaeger P. F., Kohl S. A. A., Petersen J. and Maier-Hein K. H., “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, Vol. 18, Issue 2, Pages 203–211, 2021.
19. Hatamizadeh A., Tang Y., Nath V., Yang D., Myronenko A., Landman B., Roth H. R. and Xu D., “UNETR: Transformers for 3D Medical Image Segmentation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, Pages 574–584.
20. Cao H., Wang Y., Chen J., Jiang D., Zhang X., Tian Q. and Wang M., “Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation,” in *European Conference on Computer Vision*, 2023, Vol. 13803 LNCS, Pages 205–218.
21. Milletari F., Navab N. and Ahmadi S. A., “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, Pages 565–571, 2016.
22. Litjens G., Toth R., van de Ven W., Hoeks C., Kerkstra S., van Ginneken B., Vincent G., Guillard G., Birbeck N., Zhang J., Strand R., Malmberg F., Ou Y., Davatzikos C., Kirschner M., Jung F., Yuan J., Qiu W., Gao Q. et al., “Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge,” *Medical Image Analysis*, Vol. 18, Issue 2, Pages 359–373, 2014.
23. Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y., Lu L., Yuille A. L. and Zhou Y., “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,” 2021.
24. Ma J., Li F. and Wang B., “U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation,” 2024.
25. Xing Z., Ye T., Yang Y., Liu G. and Zhu L., “SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation,” 2024.
26. Zhao MIT A., Balakrishnan MIT G., Durand MIT F., Guttag MIT J. V and Dalca MIT A. V, “Data Augmentation Using Learned Transformations for One-Shot Medical Image Segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, Pages 8543–8553.
27. Landman B., Xu Z., Igelsias J., Styner M., Langerak T. and Klein A., “MICCAI multi-atlas labeling beyond the cranial vault-workshop and challenge,” 2015.
28. “MONAI - Home,” Available: <https://monai.io/index.html> [Accessed: 20 September 2024]
29. Azad R., Heidary M., Yilmaz K., Hüttemann M., Karimijafarbigloo S., Wu Y., Schmeink A. and Merhof D., “Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook,” 2023.
30. Kline D. M. and Berardi V. L., “Revisiting squared-error and cross-entropy functions for training neural network classifiers,” *Neural Computing and Applications*, Vol. 14, Issue 4, Pages 310–318, 2005.
31. Lin T.-Y., Goyal P., Girshick R., He K. and Dollar P., “Focal Loss for Dense Object Detection,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, Pages 2980–2988.
32. Karimi Monsefi A., Karisani P., Zhou M., Choi S., Doble N., Ji H., Parthasarathy S. and Ramnath R., “Masked LoGoNet: Fast and Accurate 3D Image Analysis for Medical Domain,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Pages 1348–1359, 2024.
33. Zheng S., Lu J., Zhao H., Zhu X., Luo Z., Wang Y., Fu Y., Feng J., Xiang T., Torr P. H. S. and Zhang L., “Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, Pages 6881–6890.
34. Chen L.-C., Zhu Y., Papandreou G., Schroff F. and Adam H., “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,” *arxiv*, 2018.