

# Improving Cardiac Computed Tomography Scan Segmentation Using a U-Net Model with Continual Learning Techniques

Wanida Khamprapai<sup>1</sup>, Wassaphas Thongsopa<sup>1</sup>, Chayanon Deejaiwong<sup>1</sup>, Jirawan Charoensuk<sup>1</sup>,  
Seksan Mathulaprangsan<sup>2</sup>, Chalothon Chootong<sup>1,\*</sup>

<sup>1</sup>Department of Computer Science and Information, Faculty of Science at Sriracha, Kasetsart University, Chonburi, Thailand

<sup>2</sup>Department of Computer Engineering, Faculty of Engineering at Kamphaeng Sean, Kasetsart University,  
Kamphaeng Sean, Thailand

Received 17 September 2025; received in revised form 20 January 2026; accepted 28 January 2026

DOI: <https://doi.org/10.46604/peti.2026.15705>

## Abstract

Accurate segmentation of cardiac structures in computed tomography (CT) scans is challenging due to the proximity and similar intensity of adjacent organs. This study introduces an enhanced U-Net-based approach incorporating continual learning, class merging, and separation strategies to improve cardiac CT segmentation. Anatomically related structures are first merged and later separated through class-specific heads, reducing boundary misclassification. Furthermore, pixel adjacency is employed to improve the delineation of complex cardiac regions. The proposed method is evaluated on the MM-WHS 2017 dataset, focusing on seven components: left ventricular cavity (LVC), right ventricular cavity (RVC), left atrium cavity (LAC), right atrium cavity (RAC), myocardium (MYO), ascending aorta (AA), and pulmonary artery (PA). Experimental results show that the proposed model achieves a dice score coefficient (DSC) of 94.08% and an intersection over union (IoU) of 92.03%, outperforming baseline U-Net models. These findings demonstrate the effectiveness of structure-aware learning in advancing cardiac CT segmentation.

**Keywords:** medical image segmentation, CT scan segmentation, U-Net model, continual learning

## 1. Introduction

Medical image segmentation is an essential process in healthcare that can assist physicians in analyzing the structure of internal organs. Computed tomography (CT) provides high-resolution images of body structures that physicians usually use in diagnostic procedures. Many researchers study deep learning in medical image analysis, which enables more accurate segmentation and classification tasks [1-3]. However, CT image segmentation faces a significant challenge due to the similarity of organ structures. This similarity in density and texture between neighboring anatomical structures, particularly in cardiac imaging, where chambers and vessels are closely positioned, complicates accurate delineation using traditional intensity-based methods. Complex anatomical structures cause class confusion in medical image segmentation and represent a limitation of deep learning techniques. Boundary overlap is a common challenge in organ segmentation, as neighboring organs often exhibit unclear or fuzzy boundaries, leading to misclassification or segmentation inaccuracies.

In recent years, deep learning approaches have been applied to improve the segmentation performance of CT images. Among these techniques, U-Net and its variants learn hierarchical features that help distinguish subtle differences between similar-appearing structures. U-Net is a foundation of medical image segmentation; it was introduced by Ronneberger et al.

---

\* Corresponding author. E-mail address: [chootong.c@ku.th](mailto:chootong.c@ku.th)

[4]. It is designed with contracting and expansive paths for multi-channel feature maps. The encoder layers reduce the spatial resolution of the feature maps as the network depth increases in the contracting path. In contrast, the expansive path decodes the encoded data and locates the features while maintaining the spatial resolution of the input.

The decoder layers in the expansive path upsample the feature maps while also performing convolutional operations. The skip connections from the contracting path help preserve the spatial information lost in the contracting path, allowing the decoder layers to locate features more accurately. Zhou et al. [5] proposed a nested U-Net architecture for enhancing medical image segmentation, called UNet++. This model redesigns the skip connection scheme with nested dense skip pathways connecting the encoder and decoder at multiple resolutions. The semantic gap between feature maps in the original U-Net is reduced, which can improve boundary delineation between anatomically similar structures.

In this study, continual learning and pixel adjacency classes are applied to help U-Net and U-Net Partial models better differentiate boundaries between adjacent organs. Moreover, class merging and separation techniques are utilized to reduce errors caused by boundary overlap and improve accuracy in medical image segmentation. The Multi-Modality Whole Heart Segmentation 2017 (MM-WHS 2017) [6] dataset is used for model training and evaluation. This dataset consists of seven classes: left ventricular cavity (LVC), right ventricular cavity (RVC), left atrium cavity (LAC), right atrium cavity (RAC), myocardium (MYO), ascending aorta (AA), and pulmonary artery (PA). It is widely used for fully supervised learning and serves as a benchmark for cardiac image segmentation. Dice score coefficient (DSC) and intersection over union (IoU) are employed to report model accuracy.

## **2. Related Studies**

To provide a foundation for this study, it is essential to review previous research on the related subjects. This section highlights three primary areas: medical image segmentation, Contrastive Language–Image Pretraining (CLIP), and continual learning.

### *2.1. Medical image segmentation*

Medical image segmentation is the process of dividing a CT image into different regions corresponding to object boundaries. The challenges of CT image segmentation include complex anatomical structures, noise, artifacts, and image variations [7]. Deep learning methods, including U-Net, U-Net Partial, a fully convolutional network, and others, have been successfully applied to CT image segmentation. U-Net outperforms other deep learning models because its structure is specifically designed for segmentation with skip connections and achieves high performance even with limited data.

Furthermore, the U-Net is able to handle noise and artifacts in CT images. U-Net consists of both a contracting path (Encoder) and an expansive path (Decoder) [4]. The contracting path captures important information from the image while reducing the spatial dimensions. The expansive path utilizes the patterns learned in the contracting path to generate a segmentation map with the same size as the original input. A key feature of U-Net is the use of skip connections to transfer spatial information and details from the contracting path directly to the expansive path [8]. These skip connections improve segmentation accuracy and help prevent the loss of important information.

U-Net Partial [9] is an interesting model for CT image segmentation. U-Net Partial is an improvement on U-Net, designed to reduce class overlap between adjacent objects. U-Net Partial employs class merging and separation strategies. The merging class groups adjacent objects into a single category so that the model can learn the characteristics of the object group and reduce confusion between items with similar features. Subsequently, the objects within each group are separated so that the model can clearly learn the characteristics of each object. The separating class assists the model in accurately identifying the boundaries of each object, especially those with unclear boundaries.

## 2.2. Contrastive Language–Image Pretraining (CLIP) in medical image explanation

CLIP is a technique for training AI models to understand the relationship between text descriptions and images by learning from pairs of images and their captions [10-11]. The model consists of separate image and text encoders that convert input data into vector representations. The model is then trained to recognize that vectors of related image–text pairs have high similarity, whereas vectors of unrelated images and texts have low similarity.

CLIP has been applied to many research areas. For example, CLIP was combined with principal component analysis (PCA) to enhance the efficiency of managing images from text descriptions [12]. The features of CLIP, VGG-19, and ClinicalBERT were integrated to automatically generate captions from surgical images [13]. CLIP was employed to connect travel images with travel-related emotions [14]. The findings of these studies indicate that CLIP is capable of efficiently managing the relationship between captions and images. In addition, CLIP can be implemented in combination with other models to improve performance.

## 2.3. Continual learning in image segmentation

Continual learning enables an AI system to continuously learn new data without forgetting what it previously knew. It allows the model to adapt to new data while avoiding catastrophic forgetting [15-16]. In sequential learning, the model acquires new data by overwriting or deleting previous knowledge. Continual learning aims to retain old knowledge while learning new information. For instance, the replay-based method involves storing and replaying previous training data together with new training data [17]. The regularization-based method adds a penalty to the loss function to prevent changes to significant parameters learned from previous tasks [18]. The parameter isolation-based method protects previously acquired knowledge by assigning different sets of parameters to each task [19].

## 3. Dataset Preparation

The MM-WHS 2017 dataset consists of 20 sets of 3D CT scans, which include LVC, RVC, LAC, RAC, MYO, AA, and PA in .nii.gz format. Fig. 1 shows an example CT scan from the MM-WHS 2017 dataset. The figure illustrates the complex anatomical structures of the heart, including chambers and vessels with similar intensity distributions and unclear boundaries. This example highlights the segmentation challenges caused by anatomical proximity and low contrast between adjacent cardiac structures.

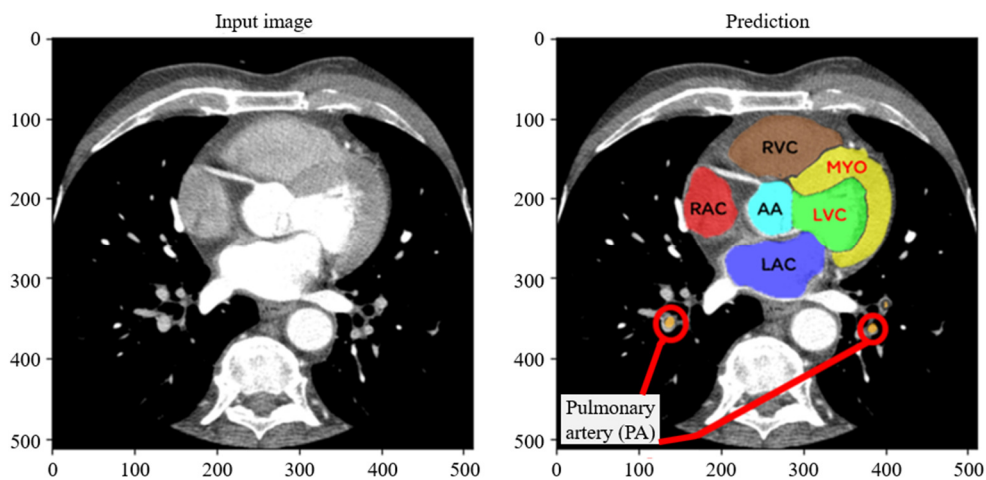


Fig. 1 Representative axial CT slice from the MM-WHS 2017 cardiac dataset

Fig. 2 demonstrates the dataset preprocessing process. Several preprocessing and augmentation techniques were applied. Contrast enhancement was performed using linear intensity scaling within the selected Hounsfield unit (HU), followed by normalization to the [0, 1] range. Opacity adjustment was applied by randomly scaling pixel intensities with a factor in the

range of 0.9–1.1 to simulate acquisition variability. Synthetic noise was introduced to enhance robustness, including Gaussian noise with zero mean and a standard deviation of 0.01–0.03 and salt-and-pepper noise with a probability of 0.01. To mitigate excessive noise, Gaussian filtering (kernel size 3×3,  $\sigma = 1.0$ ) and median filtering (kernel size 3×3) were applied.

The number of images in the dataset was increased to 200 3D CT scans. All images were resized to a consistent resolution before training. Next, organ segmentation and labeling were used to annotate LVC, RVC, LAC, RAC, MYO, AA, and PA. Each 3D CT volume was resized to a fixed dimension of (64, 256, 256), resulting in 64 axial slices per volume. To reduce computational cost, four representative 2D slices were selected from each resized 3D volume for training and evaluation. Consequently, the final dataset contained 800 2D slices. These slices were divided into a training set of 640 slices (80%) and a testing set of 160 slices (20%), which were saved in .npz format.

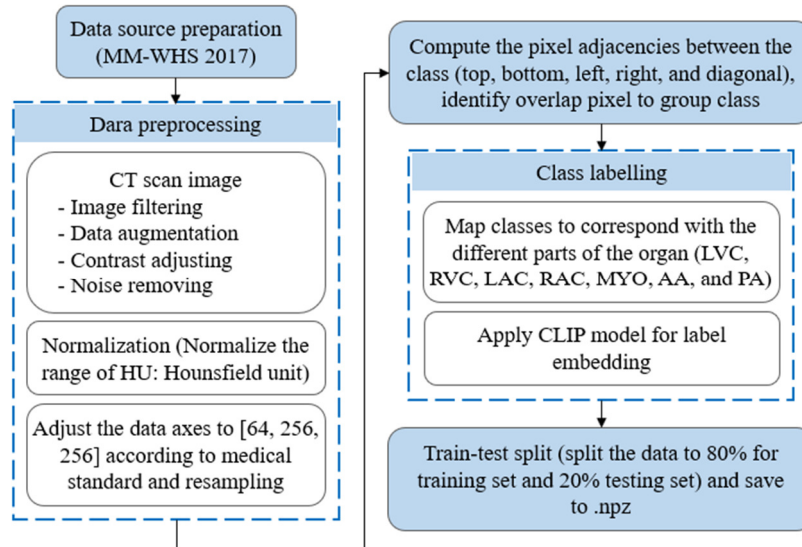


Fig. 2 Overview of the data preprocessing pipeline

A HU is a standardized quantitative scale used in CT imaging to describe tissue density. The Hounsfield scale is calibrated with water, air, and bone densities. This standardized scale allows radiologists and image processing algorithms to distinguish between tissue types based on density features. In medical image segmentation, especially for cardiac CT scans, distinguishing between neighboring structures with similar Hounsfield values is often challenging. This study uses a visualization range of  $-260$  to  $+340$  HU. Table 1 lists the HU values suitable for segmenting cardiac structures from CT scan images. Using HU values from  $-260$  to  $+340$  helps the model focus on tissues and organ structures while eliminating potential confusion from air ( $-1000$  HU) and bone ( $+1000$  HU), thereby improving segmentation accuracy.

Table 1 Hounsfield unit values for cardiac CT segmentation

Tissue	Common HU range	HU range used in this study
Air	$-1000$	Excluded
Fat	$-190$ to $-30$	Values from $-260$ and above
Soft tissue	$-29$ to $+150$	Full range used
Myocardium	$+40$ to $+60$	Full range used
Aorta	$+100$ to $+300$	Full range used
Bone	$+700$ to $+3000$	Excluded

## 4. Segmentation Method

In this section, the details of the segmentation method for cardiac CT images are presented. Section 4.1 explains the Class Merging and Separation strategy implemented within a modified U-Net framework to address boundary confusion. Section 4.2 outlines the specific model configuration, including the integration of CLIP-based embeddings. Finally, Section 4.3 describes the model training process, including the loss functions and optimization strategies.

#### 4.1. Class merging and separation

The anatomical structures in cardiac CT imaging, particularly the right atrium, right ventricle, and myocardium, are challenging to segment due to their similar spatial positioning and pixel-level characteristics. This anatomical proximity creates significant classification challenges that compromise model performance. In this study, class merging and separation methods are employed to enhance cardiac CT image segmentation. This technique systematically addresses boundary delineation issues inherent in cardiac structure segmentation.

Each cardiac structure is labeled as follows: Class 1 = LVC, Class 2 = RVC, Class 3 = LAC, Class 4 = RAC, Class 5 = MYO, Class 6 = AA, and Class 7 = PA. For the merging step, Class 1 (LVC) and Class 5 (MYO) are merged into Group 1, Class 5 (MYO) and Class 2 (RVC) are merged into Group 2, Class 7 (PA) and Class 2 (RVC) are merged into Group 3, and Class 4 (RAC) and Class 2 (RVC) are merged into Group 4. The merging structure is shown in Fig. 3. Class merging is the main idea behind grouping anatomically or spatially related organs—such as those located near each other or those with unclear boundaries—into the same category. The objective of this technique is to allow the model to learn shared features of organ groups more effectively and reduce confusion between organs with similar characteristics.

After the merging step, the continual learning technique was applied to help the model learn the unique characteristics of each organ without forgetting previously acquired information. In addition, the class-specific head technique enhances the clarity of learning for each organ by reducing interference between closely related classes and ensuring more specialized feature learning for each class. By separating the classes, the model can more accurately identify the boundaries of each organ, especially in cases where organs have complex boundaries or share similar structural and pixel-level characteristics.

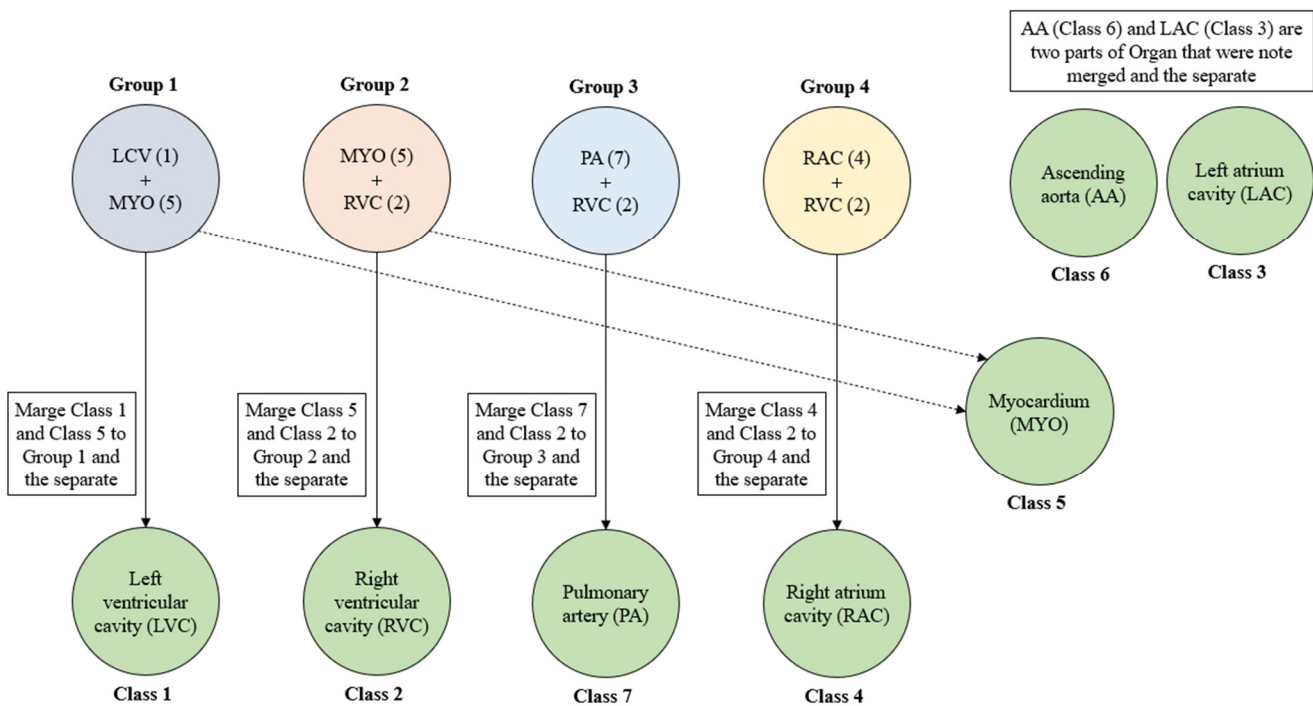


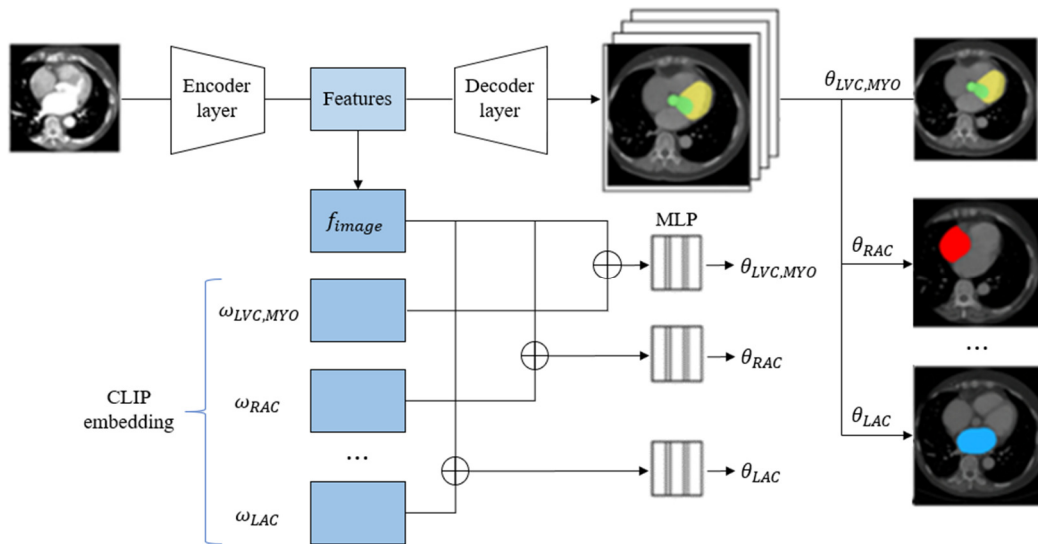
Fig. 3 Flowchart of class merging and separation

#### 4.2. Model configuration

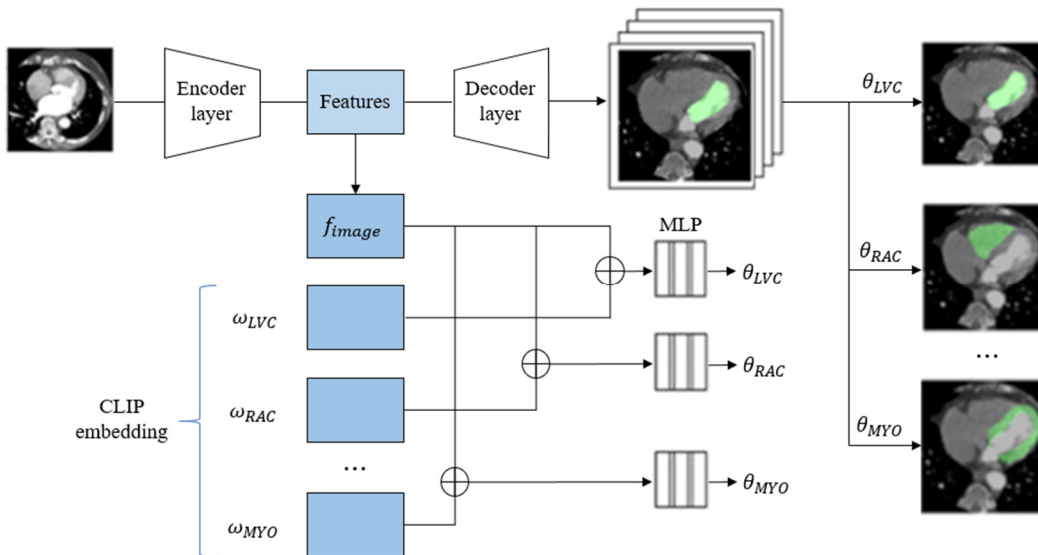
The U-Net model is widely used for biomedical image segmentation. Its main components are an encoder that extracts useful features from the input image and a decoder that is responsible for up-sampling intermediate features and producing the final output. The encoder and decoder parts are symmetrical and connected by paths. The encoder consists of repeated convolutional layers and max pooling to extract intermediate features. These extracted features are then up-sampled by the corresponding decoder, where saved copies of the encoder's features are concatenated onto the decoder's features via

connecting paths. The final layer produces the output, which is a segmentation mask. In this study, U-Net Partial is used, which applies class merging and separation with continual learning on the original U-Net to enhance the performance of organ segmentation. The segmentation model has two networks for merging and separating.

Fig. 4 illustrates the example process where LVC and MYO are merged since they have a high confusion rate. The weights from the merged classes are transferred to the newly introduced class-specific heads. In this process, the continual learning technique is applied to previously well-learned classes while enabling finer boundary distinctions in the newly separated classes. The weights from the merged class-specific heads layer contribute to learning the boundaries in the new classes, thereby guiding the model to achieve more precise segmentation.



(a) Merging network structure



(b) Separating network structure

Fig. 4 U-Net Partial with merging and separating network

In this study, a continual learning technique is applied so that the model is updated across multiple training stages rather than being trained from scratch. The learning process follows a predefined sequence: the model is first trained using merged class labels to learn shared representations and coarse anatomical boundaries and is subsequently refined to separate the merged classes for finer boundary delineation. There are two main stages:

- (1) Class merging, where anatomically adjacent structures with similar intensity profiles are grouped into a single label to facilitate the learning of shared global morphology and reduce inter-class confusion;

- (2) Class separation, where the model is progressively refined to segment individual anatomical structures using class-specific segmentation heads.

The primary objective of this sequential learning strategy is to mitigate boundary confusion caused by high pixel adjacency between anatomically neighboring structures. Unlike classical continual learning frameworks, which explicitly evaluate catastrophic forgetting across independent tasks, the proposed approach focuses on progressive boundary refinement while preserving previously learned segmentation behavior through stage-wise training and parameter isolation.

Each CT scan image is processed through an encoder. The image features ( $\omega$ ) are extracted from multiple layers and then compressed using global average pooling. CLIP is used to obtain label embeddings for each class. For each image's label, a fixed textual prompt corresponding to CLIP is used to generate label embeddings for each class. For each anatomical structure, a fixed textual prompt corresponding to the organ name is defined: 'LVC', 'RVC', 'LAC', 'RAC', 'MYO', 'AA', and 'PA'. For merged classes, the prompts are defined as 'LVC and MYO', 'MYO and RVC', 'PA and RVC', and 'RAC and RVC', respectively. These prompts are encoded using a pre-trained CLIP text encoder to obtain class-level semantic embeddings. These embeddings are generated once offline and then combined with compressed features from the final encoder output.

The feature vectors are combined via a multi-layer perceptron (MLP) to generate kernels for the class-specific heads ( $\theta$ ), which will be used to produce class-specific predictions. Dice loss and binary cross-entropy (BCE) loss are combined to generate independent loss values per class, which can reduce confusion between classes and prevent performance degradation in previously learned classes during continuous segmentation.

#### 4.3. Model training

To segment cardiac structures from CT scan images, the U-Net and U-Net Partial models are utilized. During the model training process, the model's performance at different learning rates, 0.0001 (1e-4) and 0.00001 (1e-5), is compared using the Adam and AdamW optimizers. Dice loss and BCE loss are used as loss functions to enhance model performance. Their formulations are presented in the following equations.

$$Dice\ loss = 1 - \frac{2 \times TP}{2 \times TP + FP + FN} \quad (1)$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively.

$$BCE\ loss = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \quad (2)$$

BCE loss is suitable for segmentation problems where the probability of each pixel needs to be considered. It helps reduce model errors in areas where organs overlap or in cases where pixel boundaries of organs are not clearly defined. Here,  $y$  denotes the actual pixel value, and  $\hat{y}$  denotes the predicted value from the model.

## 5. Experimentation Result

This study evaluates the results of models that employ differing learning rates and optimizers in terms of DSC and IoU. The experimental results for the merged group classes are shown in Table 2. The AdamW optimizer with a learning rate of 1e-4 achieved significantly higher DSC and IoU than the others in Groups 1, 2, and 3. However, in Group 4, Adam with a learning rate of 1e-4 outperformed AdamW with the same learning rate in both DSC and IoU. These results indicate that initially merging classes and then separating them improves segmentation accuracy. The experiment demonstrates that DSC and IoU are higher for both Adam and AdamW when the learning rate is 1e-4 compared to 1e-5 across all groups. Considering these metrics, Group 3, which merged PA and RVC, performed the best overall due to their anatomical proximity and frequent

misclassification in baseline models using the merge-and-separate class strategy, achieving the highest average DSC of 87.85%. The model can effectively learn adjacent organ boundaries and reduce inter-class confusion when the AdamW optimizer with a learning rate of  $1e-4$  is employed.

Table 2 DSC and IoU values of merged class

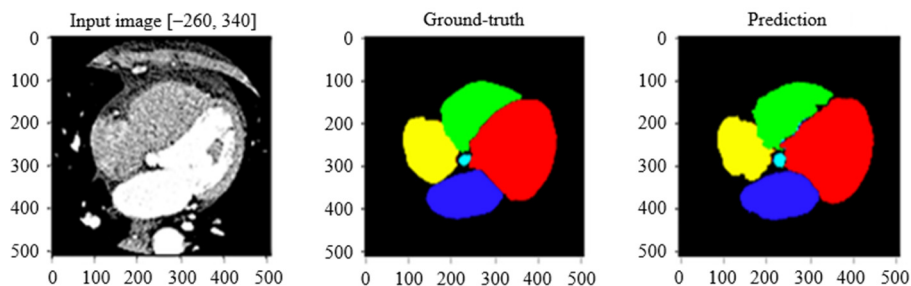
Group class	1e-5 / Adam		1e-4 / Adam		1e-5 / AdamW		1e-4 / AdamW	
	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU
LVC + MYO (Group1)	83.23	80.57	83.56	80.73	82.17	79.42	<b>83.76</b>	<b>81.46</b>
MYO + RVC (Group2)	81.34	76.12	82.52	77.69	83.05	78.04	<b>84.91</b>	<b>80.35</b>
PA + RVC (Group3)	83.51	75.94	87.54	80.90	85.44	78.18	<b>87.85</b>	<b>81.05</b>
RAC + RVC (Group4)	75.54	70.45	<b>81.23</b>	<b>76.01</b>	77.97	72.81	79.74	74.96

Table 3 presents the experimental results for separate classes, comparing the performance of the model in each class using DSC and IoU. In addition, the results obtained by varying the optimizers (Adam, AdamW) and learning rates ( $1e-4$ ,  $1e-5$ ) were compared with the baseline (U-Net). The AdamW optimizer with a learning rate of  $1e-4$  performed best across multiple classes. For instance, in the AA class, AdamW with a learning rate of  $1e-4$  achieved a DSC of 94.75 and an IoU of 92.73, which are the highest. This indicates that AdamW, with a learning rate of  $1e-4$ , prevents overfitting and enhances performance. DSC and IoU were significantly improved in several classes when the learning rate increased from  $1e-5$  to  $1e-4$ . The model performs poorly for organs with complex structures, such as PA and LAC.

Table 3 DSC and IoU for separated classes

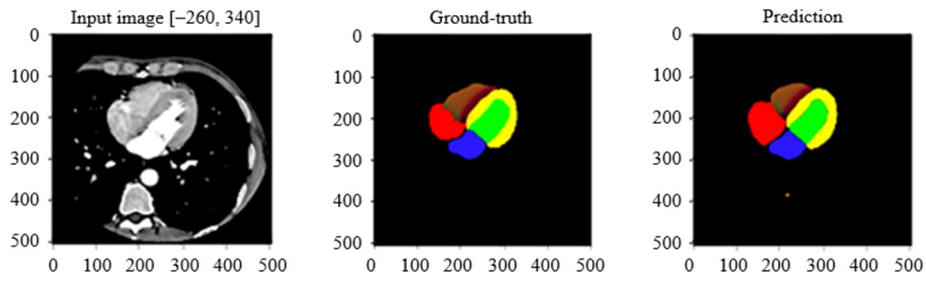
Separate class	Baseline (U-Net)		1e-5 / Adam		1e-4 / Adam		1e-5 / AdamW		1e-4 / AdamW	
	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU	DSC	IoU
RAC (4)	81.63	76.77	69.19	64.20	<b>83.43</b>	<b>78.54</b>	76.28	71.24	81.41	76.57
LVC (1)	89.33	86.93	81.69	79.25	<b>93.04</b>	<b>90.68</b>	82.09	77.66	88.98	86.69
LAC (3)	90.02	87.40	82.66	79.58	<b>90.03</b>	87.21	78.94	76.10	82.73	80.09
MYO (5)	80.19	86.18	80.58	76.25	<b>89.60</b>	85.26	82.09	77.66	80.72	76.68
AA (7)	91.64	89.91	93.93	91.87	91.81	89.94	88.35	86.18	94.75	92.73
PA (6)	<b>76.79</b>	<b>73.76</b>	63.81	60.82	74.72	71.89	66.29	63.05	75.39	72.52
RVC (2)	84.10	78.78	82.73	77.07	84.82	79.43	75.74	69.84	<b>84.96</b>	<b>79.77</b>

In particular, PA, when utilizing the AdamW optimizer with a learning rate of  $1e-5$ , achieved a DSC of only 66.29%, below the baseline of 76.79%. PA is anatomically smaller and exhibits higher inter-subject shape variability compared to other cardiac structures. These characteristics result in fewer representative pixels during training, making its segmentation more sensitive to minor feature distortions introduced during class merging and separation. In addition, boundary contrast for LAC and PA is often weak in CT scans, as their HU distributions overlap with surrounding tissues and blood-filled regions. While the proposed method effectively reduces confusion between large adjacent organs, it may inadvertently suppress subtle boundary cues that are critical for accurately segmenting small or thin structures.



(a) Heart CT segmentation - Case 1: merging

Fig. 5 Segmentation result from the proposed model for class merging and class separation



(b) Heart CT segmentation - Case 2: separation

Fig. 5 Segmentation result from the proposed model for class merging and class separation (continued)

Fig. 5 shows an example of a segmentation result for class merging and class separation. The experimental results demonstrate that the merge-and-separate class approach can significantly improve model performance in segmenting organs from medical images. This approach helps the model learn organ structures more effectively and reduces confusion between adjacent organs. Moreover, to evaluate the segmentation results, a system has been developed that enables users to upload files in .nii.gz or .npz formats via a web application. Users can view the segmentation results for U-Net Partial across various slices, allowing for easy comparison. Fig. 6 illustrates an example of a prediction result.

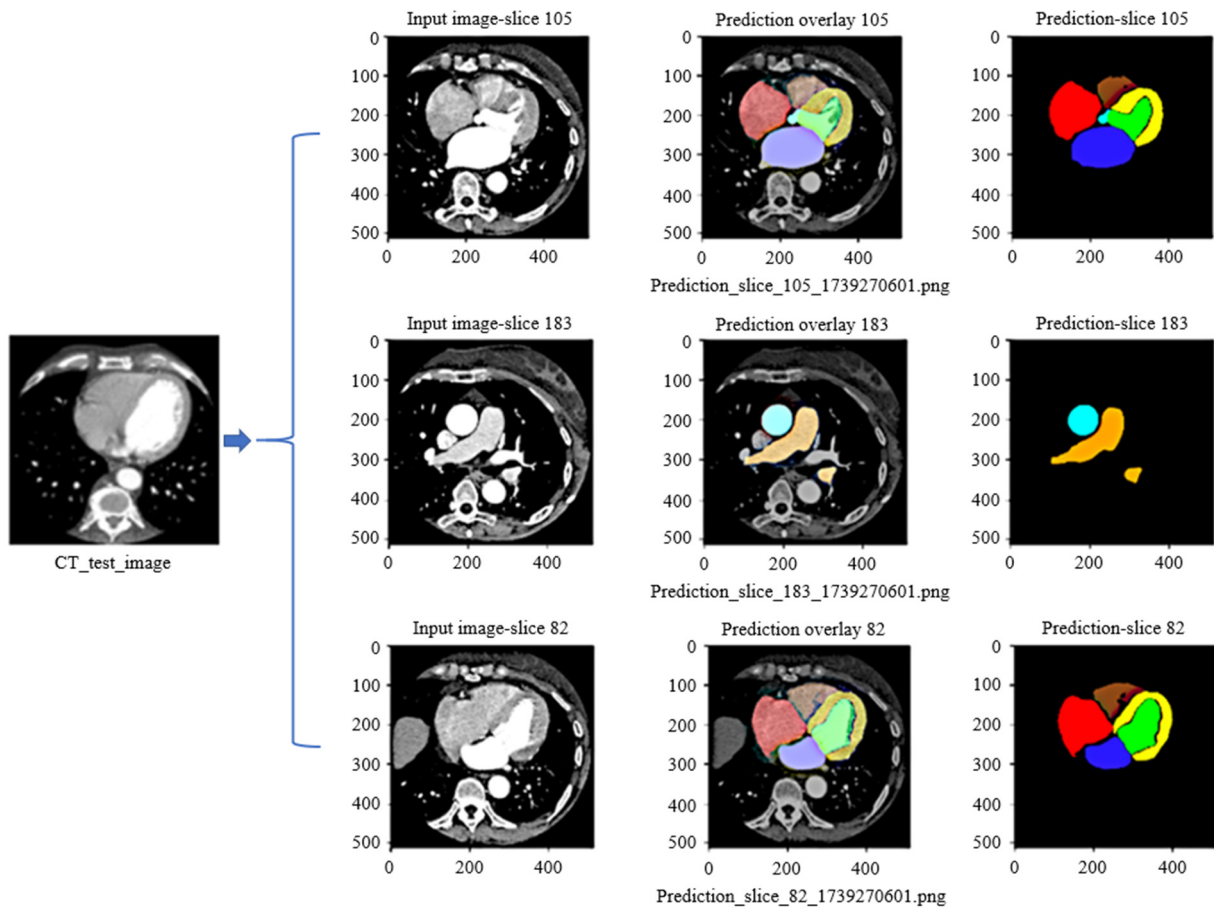


Fig. 6 Examples of prediction results

## 6. Limitations

Several limitations of the present study should be acknowledged, including the following:

- (1) The MM-WHS 2017 dataset contains a limited number of CT volumes. This limitation may affect the generalizability of the proposed method and its robustness across diverse anatomical variations.
- (2) The experiments rely on a single public dataset. Future work will include validation on additional multi-center datasets to better assess the generalization capability of the proposed approach.

- (3) Although the proposed method was compared with a baseline U-Net, a comprehensive comparison with a wider range of state-of-the-art segmentation methods was not performed. This is partly due to differences in experimental settings and label definitions across existing studies. Incorporating more extensive comparisons with recent state-of-the-art approaches will be an important direction for future research.

## 7. Conclusions

This study successfully developed and evaluated the U-Net Partial architecture, which integrates continual learning techniques and class merging and separation methods to improve cardiac CT scan segmentation. By addressing the inherent challenges of anatomical proximity and similar pixel-level characteristics between cardiac structures, the proposed method effectively reduces boundary confusion. The integration of CLIP text embeddings further improved the model's ability to distinguish between complex organs where traditional intensity-based methods often fail. This sequential learning approach provides a robust framework for medical image analysis. The experimental results on the MM-WHS 2017 dataset demonstrate that this methodology offers a significant advancement in automated cardiac diagnostics. The main conclusions of this study are as follows:

- (1) The class merging and separation technique improved DSC and IoU for each group, especially Group 3 (PA + RVC), which achieved optimal results, reaching 87.85%. Individual class performance demonstrated high accuracy, with LVC achieving 93.12% DSC and 90.95% IoU, while AA reached 94.08% DSC and 92.03% IoU.
- (2) The continual learning method reduces forgetting and clarifies distinctions between anatomically similar structures. Integrating CLIP text embeddings with compressed image features through an MLP enhances the model's ability to distinguish between different cardiac structures, particularly in cases where traditional intensity-based methods struggle due to similar HU values.
- (3) The efficiency of two optimizers, Adam and AdamW, was compared under different learning rates of  $1e-4$  and  $1e-5$ . The experimental results show that the AdamW optimizer with a learning rate of  $1e-4$  consistently outperforms other configurations across most cardiac structures.
- (4) Dice loss and BCE were used as loss functions to handle pixel-level classification in regions with unclear boundaries.

However, complex structures in CT scan images, such as the LAC, which have unclear boundaries, or small organs such as the PA, still pose challenges for accurate segmentation. In future work, segmentation performance may be further improved by incorporating dynamic class merging strategies and attention mechanisms designed to highlight salient anatomical structures and boundary details.

## Acknowledgments

This research was financially supported by the Faculty of Science at Sriracha and the Kasetsart University Research and Development Institute (KURDI), Kasetsart University.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- [1] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, et al., "Deep Learning in Medical Imaging and Radiation Therapy," *Medical Physics*, vol. 46, no. 1, pp. e1-e36, 2019.

- [2] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 2018.
- [3] X. Chen, J. K. Udupa, U. Bagci, Y. Zhuge, and J. Yao, "Medical Image Segmentation by Combining Graph Cuts and Oriented Active Appearance Models," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2035-2046, 2012.
- [4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention*, pp. 234-241, 2015.
- [5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3-11, 2018.
- [6] "MM-WHS: Multi-Modality Whole Heart Segmentation," <https://zmiclab.github.io/zxh/0/mmwhs/>, 2025
- [7] M. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, "Deep Learning for Medical Image Segmentation: State-of-the-Art Advancements and Challenges," *Informatics in Medicine Unlocked*, vol. 47, article no. 101504, 2024
- [8] B. Woo and M. Lee, "Comparison of Tissue Segmentation Performance between 2D U-Net and 3D U-Net on Brain MR Images," *International Conference on Electronics, Information, and Communication*, pp. 1-4, 2021.
- [9] H. Lee, P. Puttimit, W. Thongsopa, S. Roh, Y. R. Lee, S. Y. Park, et al., "Refining Class Confusion with Pixel Adjacency and Continual Learning in Medical Image Segmentation," *Proceedings of the 7th International Conference on Culture Technology*, pp. 86-92, 2024.
- [10] G. Arya, M. K. Hasan, A. Bagwari, N. Safie, S. Islam, F. R. A. Ahmed, et al., "Multimodal Hate Speech Detection in Memes Using Contrastive Language-Image Pre-Training," *IEEE Access*, vol. 12, pp. 22359-22375, 2024.
- [11] D. Chen, Z. Wu, F. Liu, Z. Yang, S. Zheng, Y. Tan, et al., "ProtoCLIP: Prototypical Contrastive Language Image Pretraining," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 1, pp. 610-624, 2025.
- [12] B. Liu, "Research on Image Classification and Retrieval Based on Contrastive Language-Image Pre-Training," *4th Asia-Pacific Conference on Communications Technology and Computer Science*, pp. 418-423, 2024.
- [13] S. Kütük, T. Çağlıkantar, and D. Sarıkaya, "Generating Automatic Surgical Captions Using a Contrastive Language-Image Pre-Training Model for Nephrectomy Surgery Images," *32nd Signal Processing and Communications Applications Conference*, pp. 1-4, 2024.
- [14] H. Q. Vu, B. Song, G. Li, and R. Law, "Exploring Emotional Aspects of Travel Concepts via Travel Photos Based on Contrastive Language-Image Pretraining," *Tourism Management*, vol. 108, article no. 105117, 2025.
- [15] J. Xu, J. Ma, X. Gao, and Z. Zhu, "Adaptive Progressive Continual Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6715-6728, 2022.
- [16] R. Gao and W. Liu, "Red Alarm: Controllable Backdoor Attack in Continual Learning," *Neural Networks*, vol. 188, article no. 107479, 2025.
- [17] Neeraj and P. Nandal, "Continual Learning Techniques to Reduce Forgetting: A Comparative Study," *2nd International Conference on Computational Intelligence, Communication Technology and Networking*, pp. 210-215, 2025.
- [18] B. Kann, S. Castellanos-Paez, and P. Lalanda, "Evaluation of Regularization-based Continual Learning Approaches: Application to HAR," *IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 460-465, 2023.
- [19] K. Hong, H. Jin, S. Suh, and E. Kim, "Exploration and Exploitation in Continual Learning," *Neural Networks*, vol. 188, article no. 107444, 2025.



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).