

Refining Class Confusion with Pixel Adjacency and Continual Learning in Medical Image Segmentation

Hyunji Lee
School of Computer Science
and Engineering,
Kyungpook National
University,
Daegu, South Korea
hyunji0483@knu.ac.kr

Pawat Puttimit
Department of Computer
Science,
Kasetsart University,
Bangkok, Thailand
pawat.put@ku.th

Wassaphas Thongsopa
Department of Computer
Science,
Kasetsart University Sriracha
Campus,
Chon Buri, Thailand
wassaphas.t@ku.th

Seongmin Roh
School of Computer Science
and Engineering,
Kyungpook National
University,
Daegu, South Korea
smroh@knu.ac.kr

Yu Rim Lee
Department of Internal
Medicine,
College of Medicine,
Kyungpook National
University
Daegu, South Korea
deblue00@naver.com

Soo Young Park
Department of Internal
Medicine,
College of Medicine,
Kyungpook National
University
Daegu, South Korea
psyong0419@gmail.com

Won Young Tak
Department of Internal
Medicine,
College of Medicine,
Kyungpook National
University
Daegu, South Korea
wytak@knu.ac.kr

Soon Ki Jung*
School of Computer Science
and Engineering,
Kyungpook National
University,
Daegu, South Korea
skjung@knu.ac.kr

Abstract—In medical image segmentation, challenges arise when attempting to differentiate between closely adjacent anatomical structures, where pixel adjacency frequently leads to boundary confusion. This paper presents a continual learning-based approach that addresses this issue by initially merging classes with high pixel adjacency and subsequently separating them to refine boundary distinctions. Experiments conducted across three diverse datasets—KNUH Body Composition, MM-WHS, and FLARE22—revealed improvements in segmentation performance when classes prone to boundary confusion were merged and then separated. Specifically, classes with distinctly separated Hounsfield Unit (HU) values showed performance improvements, as the model was able to effectively learn during the merging and separation process. Conversely, classes with similar HU and significant size differences experienced performance degradation, as the model tended to focus more on the larger organ during separation. Additionally, merging organs with high structural variability between slices posed challenges for boundary learning, which impacted segmentation accuracy after separation. These results emphasize the importance of considering both pixel adjacency and HU distinction, as well as structural consistency across slices, when designing class merging and separation strategies to optimize medical image segmentation.

Keywords—medical image segmentation, continual learning, pixel adjacency, class merging and separation

I. INTRODUCTION

Medical image segmentation provides critical information for physicians to make diagnoses and formulate treatment plans. Due to its direct contribution to medical judgment, medical image segmentation research focuses on achieving high accuracy segmentation performance, among other things. In particular, accurately distinguishing between neighboring anatomical structures is a major challenge. When these structures are spatially closely connected and the pixel adjacencies between classes are high, confusion between classes can easily occur, leading to poor performance.

For example, in the body composition segmentation [1], which consists of muscle (MUS), intermuscular fat (IMAT), subcutaneous fat (SAT), and visceral fat (VAT), as shown in Fig. 1, IMAT is mainly surrounded by MUS, with about 70% of IMAT pixels being adjacent to MUS. This high pixel

adjacency significantly increases the probability of IMAT being misclassified as MUS. This study introduces a method where classes with high pixel adjacency, which are often confused, are initially merged for training. After this, the Continual Segmentation [2] technique is applied to separate the previously merged classes. By first learning the shared boundaries between the merged classes, the model can then refine the separation of each class more precisely. This approach allows for more accurate segmentation compared to training all classes at once.

In this study, we applied the Continual Segmentation technique to separate the merged classes while maintaining performance on the other classes. In addition, we applied a state-of-the-art technique [3] that reduces the amount of computation while maintaining high performance by utilizing class-specific heads that independently handle the weights between classes and CLIP [11] that generates text embeddings for each class. Unlike many existing studies [4, 5], this approach does not rely on knowledge distillation of features, offering a more computationally efficient solution.

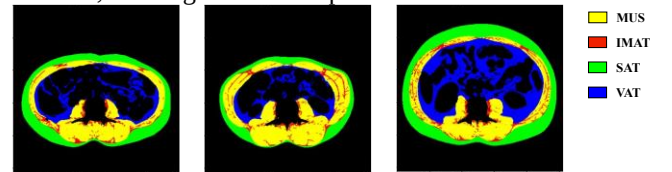


Fig. 1. KNUH Body Composition Dataset Ground Truth

This methodology is particularly useful for clearly distinguishing boundaries between adjacent anatomical structures when analyzing body components in CT images. This method, which improves performance with minimizing confusion between classes, was validated on in-house dataset, the KNUH Body Composition dataset provided by Kyungpook National University Hospital, as well as the publicly available abdominal CT multi-organ segmentation dataset Flare22 [6] and the heart segmentation dataset MM-WHS [7] of lung CT. The proposed method was compared to a traditional U-Net [8] that simultaneously trains all classes. Performance improvements were observed for many classes, particularly when the HU between merged classes were clearly distinct. However, mixed results were noted for classes with similar HU or high structural variability, with

some experiencing performance degradation. These findings suggest that while the proposed method effectively reduces boundary ambiguity and confusion in specific segmentation tasks, a more cautious approach is needed for classes with complex characteristics.

II. Related work

Medical image segmentation has been the subject of much recent research because it provides useful information for doctors to diagnose diseases and plan treatment. However, accurately segmenting complex anatomical structures is very challenging, especially distinguishing the boundaries between neighboring anatomical structures due to high pixel adjacency. To solve this problem, existing researchers have proposed various methods such as boundary-based segmentation [9].

Continual segmentation is an effective learning technique that can maintain performance when adding new classes to an already trained model, without requiring access to the data used in previous training. This is especially beneficial in medical contexts, where data access can be limited due to privacy concerns or other restrictions. Continual segmentation prevents catastrophic forgetting, ensuring that new information can be learned while preserving the performance of previously learned classes.

In the context of Continual Learning, feature-based knowledge distillation [4, 5] plays a critical role in retaining model performance when new data is introduced. This technique transfers knowledge from the feature vectors extracted by the current model's encoder to the feature vectors learned by the previous model, allowing the new data to be learned without degrading the performance of existing classes.

However, knowledge distillation is computationally expensive. To address this, alternative methods have been proposed, such as the Pseudo Label approach [10], which leverages predictions from the previous training phase as labels for the current training phase. Additionally, Contrastive Language-Image Pretraining (CLIP) [11] learns associations between text and image pairs using contrastive learning, generating image feature vectors that similar with specific text embeddings. Finally, class-specific heads are employed to minimize interference between classes by handling each class independently, thus preserving the performance of previously learned classes when new classes are introduced. This approach [3] significantly reduces computational costs—by approximately half—compared to traditional knowledge distillation, while still maintaining high performance.

III. METHOD

To mitigate the model confusion caused by high pixel adjacencies, we propose to merge frequently confused classes into a single class for initial training. Subsequently, the Continual Segmentation approach is employed to separate these merged classes. This methodology enables the model to first acquire a comprehensive understanding of the shared boundaries between the merged classes, which, during the separation phase, facilitates a more finely differentiation of the previously confused class boundaries by leveraging the learned boundary information and associated weights.

To compute the pixel adjacencies between classes, we first applied binary dilation to the ground truth of each class, expanding the pixels in the top, bottom, left, right, and diagonal directions by one pixel using a 3x3 kernel. Next, a logical AND operation was performed between the expanded region and the ground truth of the neighboring class to identify the overlapping pixels, which were then defined as the pixel adjacency between the respective classes. To avoid bias due to differences resulting from class size disparities, we normalized by dividing by the total number of pixels in each class.

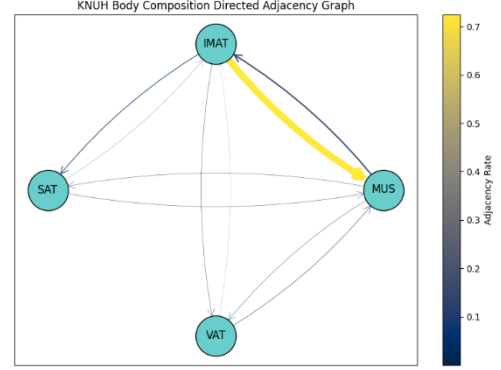


Fig. 2. Directional graph of pixel adjacency by class in the KNUH Body Composition dataset

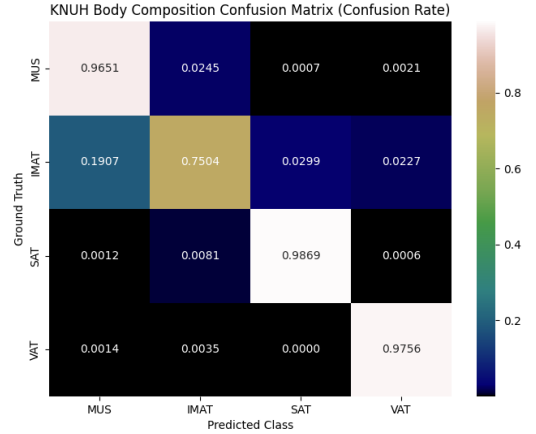


Fig. 3. Confusion matrix for U-Net models trained on the KNUH Body Composition dataset

The visualization of class adjacency calculated through this process is presented in Fig. 2. Higher adjacency values are depicted using lighter colors and thicker edges. As shown, approximately 70% of IMAT, a class with a relatively small area, is adjacent to MUS. This indicates that IMAT is more likely to be misclassified as MUS. Conversely, since MUS occupies a larger area, the proportion of its pixels that are adjacent to IMAT is small, resulting in only a modest impact on segmentation performance for MUS.

The confusion between MUS and IMAT was verified through a confusion matrix derived from the results of a U-Net model trained using the traditional method of learning all classes simultaneously. As depicted in Fig. 3, the confusion matrix for the KNUH Body Composition Dataset indicates the proportion of IMAT pixels that were incorrectly predicted as MUS, with the IMAT row and MUS column representing the number of pixels predicted as MUS when IMAT is the ground truth, divided by the total number of IMAT pixels.

The matrix reveals a high rate of misclassification of IMAT as MUS. Moreover, when separating merged classes, we have found empirically that separating low-confusion classes (e.g., MUS) into new classes performs better than separating high-confusion classes (e.g., IMAT). Therefore, we separate low-confusion classes into new classes.

To determine which classes to merge and separate, We utilized a confusion matrix to exclude classes that were highly connected in the class adjacency graph but rarely confused by the model due to their distinct shapes. Then, we selected classes with high pixel adjacencies and frequently misclassified by the model for merging.

To separate the merged classes, we use the latest continual segmentation technique [3] that utilizes pseudo-labels and integrates CLIPs with class-specific heads. This approach allows the model to learn the weights for each class independently, providing a more efficient alternative to commonly used knowledge distillation methods. It leverages pseudo-labels and lightweight class-specific heads from previous learning stages to flexibly accommodate newly introduced classes and enable independent predictions for both new and previously learned classes. This strategy effectively minimizes the impact of learning new classes on the performance of existing classes. However, in this study, we did not implement the pseudo-label strategy due to the accessibility of previous training data. Instead, we focused on merging and separating classes. In addition, we combine Contrastive Language-Image Pretraining (CLIP) embeddings with class-specific head operations to further convey the semantic information of each class.

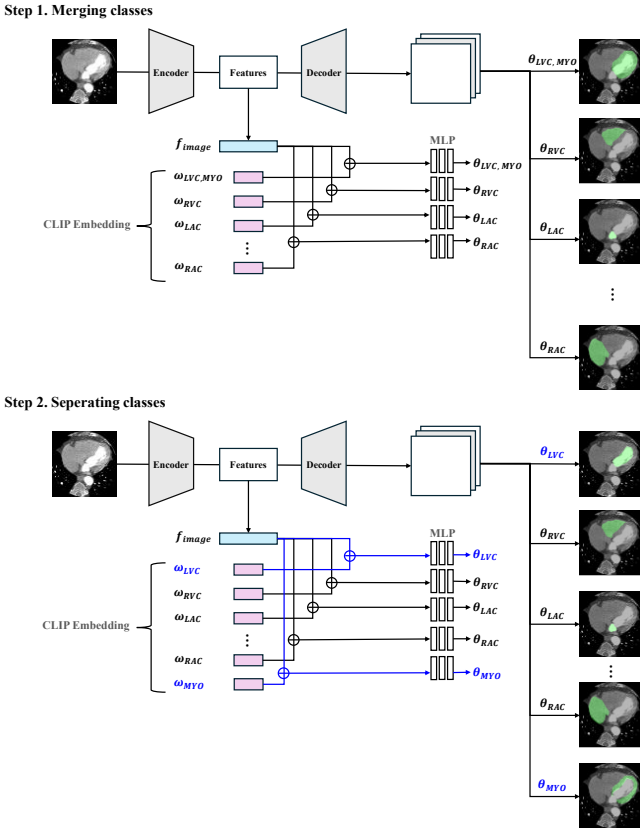


Fig. 4. Overview of Proposed Method

Fig. 4 illustrates the process where two classes with high confusion rates are merged to train the initial model, followed by the separation of the merged classes. During this process,

the weights from the merged classes are transferred to the newly introduced class-specific heads. This approach leverages Continual Learning, allowing for the preservation of the performance on previously well-learned classes while enabling finer boundary distinctions in the newly separated classes. The weights from the merged class-specific heads contribute to the learning of the boundaries in the new classes, thereby guiding the model to achieve more precise segmentation.

The input image is processed through an encoder, extracting features from multiple layers, which are then compressed using Global Average Pooling (GAP). Following this, text embeddings for each class are derived from descriptive sentences using CLIP, and these embeddings are combined with the compressed features from the final encoder output. The resulting combined vectors are passed through a multi-layer perceptron (MLP) to generate kernels for the class-specific heads. These kernels act as filters for the convolution operations, producing class-specific predictions. This method enables class-independent predictions, effectively minimizing the impact of newly introduced predictions on existing class predictions.

By combining Dice Loss and Binary Cross Entropy Loss for each class, we generate independent loss values per class. This approach reduces confusion between classes and prevents performance degradation in previously learned classes during Continuous Segmentation. The loss function is defined as follows:

$$\text{Dice}(P, T) = \frac{2 \cdot (P \cdot T) + \epsilon}{P + T + \epsilon}, \quad (1)$$

$$L_{\text{Dice}} = 1 - \text{Dice}(P, T), \quad (2)$$

$$L_{\text{MultiDice}}(P, T) = \frac{1}{B} \sum_{b=1}^B \left(\frac{1}{C} \sum_{c=1}^C (1 - L_{\text{Dice}}(P_{b,c}, T_{b,c})) \right), \quad (3)$$

$$L_{\text{BCE}}(P, T) = -(T \cdot \log(P) + (1 - T) \cdot \log(1 - P)), \quad (4)$$

$$L_{\text{MultiBCE}}(P, T) = \frac{1}{B} \sum_{b=1}^B \left(\frac{1}{C} \sum_{c=1}^C L_{\text{BCE}}(P_{b,c}, T_{b,c}) \right), \quad (5)$$

$$L_{\text{Total}}(P, T) = L_{\text{MultiDice}}(P, T) + L_{\text{MultiBCE}}(P, T), \quad (6)$$

where P represents the predicted value, T is the ground truth, and ϵ is a very small value to avoid the division by zero errors. Also, B denotes the batch size, and C represents the number of classes. Thus, $P_{b,k}$ and $T_{b,c}$ refer to the prediction and target for the c -th class in the b -th sample of the batch, respectively.

IV. EXPERIMENTS

A. Dataset

To validate the effectiveness of our proposed method, we utilized one in-house dataset and two publicly available datasets.

- *KNUH Body Composition*

We employed the in-house KNUH Body Composition dataset, provided by Kyungpook National University Hospital. This dataset contains 1,004 axial slices of abdominal CT scans, with a resolution of 512x512, labeled with body composition classes (MUS, IMAT, SAT, VAT) at the L3 vertebral level for 1,004 patients.

- *MM-WHS*

The MM-WHS dataset [7] consists of CT and MRI scans for major heart organ segmentation. For this study, we used CT scan data with a 512x512 resolution from 20 patients, including fine-grained labels for the left ventricle blood cavity (LVC), right ventricle blood cavity (RVC), left atrium blood cavity (LAC), right atrium blood cavity (RAC), myocardium of the left ventricle (MYO), and ascending aorta (AA).

- **FLARE22**

We also utilized the FLARE22 dataset [6], which provides abdominal CT scans from 50 patients, collected from multiple institutions. Each scan has a resolution of 512x512 and includes labels for 13 major abdominal organs, including the liver, kidneys, spleen, pancreas, aorta, inferior vena cava, adrenal glands, gallbladder, esophagus, stomach, and duodenum.

B. Selecting the classes to merge

For the KNUH Body composition dataset, as indicated by the pixel adjacency graph in Fig. 2 and the confusion matrix in Fig. 3, IMAT and MUS are highly adjacent, and the model frequently confuses IMAT with MUS. Consequently, we merged IMAT and MUS, and then separated MUS.

For the MM-WHS dataset, based on the adjacency and confusion matrices between classes shown in Fig. 4 and Fig. 5, we conducted experiments on four cases:

- **LVC-MYO**: LVC and MYO are adjacent, with LVC often misclassified as MYO. Thus, LVC and MYO were merged, and MYO was subsequently separated.
- **RVC-MYO**: RVC and MYO are adjacent, showing similar confusion rates, with MYO more frequently misclassified. Therefore, RVC and MYO were merged, followed by RVC separation.
- **RVC-PA**: RVC and PA are adjacent, with PA often confused as RVC. Thus, RVC and PA were merged, and RVC was then separated.
- **RVC-RAC**: RVC and RAC are adjacent, and RAC is frequently misclassified as RVC. Despite lower adjacency, their similar size and frequent confusion led to merging, followed by RVC separation.

For the FLARE22 dataset, based on the adjacency and confusion matrices between classes shown in Fig. 6 and Fig. 7, we conducted experiments on four cases:

- **Liver-IVC**: Liver and IVC are adjacent, with IVC frequently misclassified as Liver. Therefore, they were merged, followed by separating Liver.
- **Liver-Gallbladder**: Liver and Gallbladder are adjacent, with Gallbladder often confused with Liver. Thus, they were merged, and Liver was separated.
- **Pancreas-Duodenum**: Pancreas and Duodenum are adjacent, with Pancreas often confused as Duodenum. They were merged, followed by Duodenum separation.
- **Esophagus-Stomach**: Esophagus and Stomach are adjacent, and Esophagus is often confused with Stomach. They were merged, and Stomach was separated.

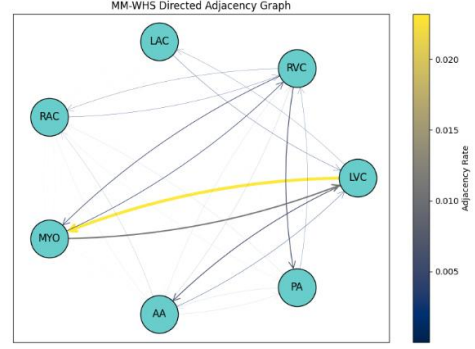


Fig. 4. Directional graph of pixel adjacency by class in the MM-WHS

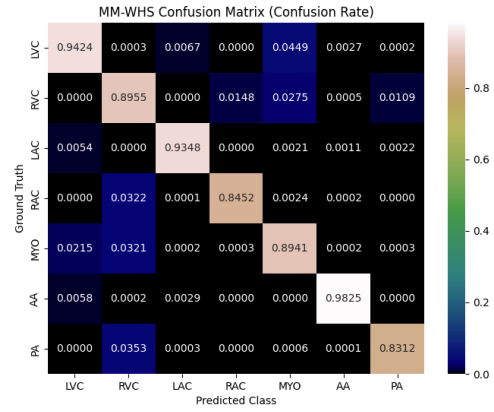


Fig. 5. Confusion matrix for U-Net models trained on the MM-WHS

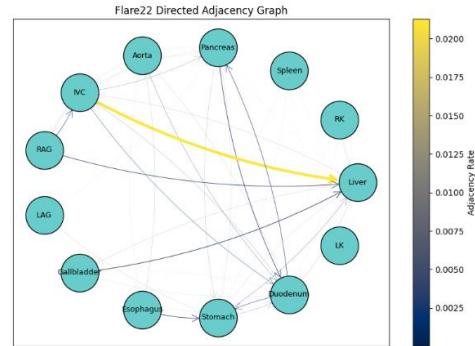


Fig. 6. Directional graph of pixel adjacency by class in the Flare22

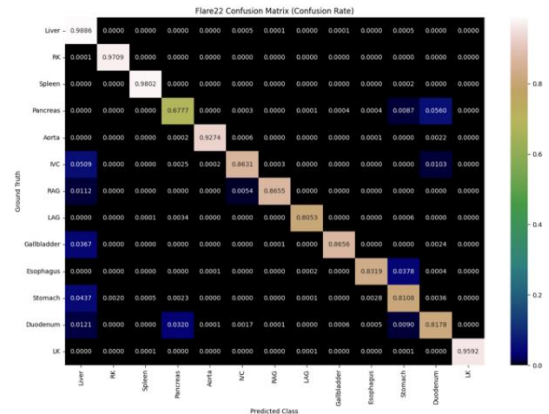


Fig. 7. Confusion matrix for U-Net models trained on the Flare22

C. Evaluation Metrics

To evaluate the performance of the model, we used the Dice Score Coefficient (DSC) and Intersection over Union (IoU), which are frequently used in medical image segmentation, It is defined as follows:

$$DSC = \frac{2TP}{2TP+FP+FN}, \quad (7)$$

$$IoU = \frac{TP}{TP+FP+FN}, \quad (8)$$

where TP, FP, and FN are the number of true positives, false positives, and false negatives, respectively. These measures result in an output between 0 and 1, with closer to 1 indicating better performance.

V. RESULT

Each experiment is labeled by a merged class, and the classes written below are the class that will be split into new classes.

TABLE I. PERFORMANCE (DSC) COMPARISON OF THE PROPOSED METHODS FOR KNUH BODY COMPOSITION

	DSC(%)				
	MUS	IMAT	SAT	VAT	AVG
Baseline	95.83	77.55	98.86	96.70	92.23
IMAT MUS	95.80	77.85	98.95	96.77	92.34

TABLE II. PERFORMANCE (IoU) COMPARISON OF THE PROPOSED METHODS FOR KNUH BODY COMPOSITION

	IoU(%)				
	MUS	IMAT	SAT	VAT	AVG
Baseline	92.02	63.68	97.75	93.74	86.80
MUS IMAT	91.97	64.04	97.93	93.88	86.95

The Experimental results on the KNUH body composition dataset show that the proposed method improves the segmentation performance of confusable classes that frequently overlap with MUS in terms of pixel adjacencies, especially IMAT. Although the overall improvement in the average performance metrics (DSC and IoU) is marginal, the improvement in the segmentation performance of confusable classes such as IMAT over other classes demonstrates the effectiveness of the proposed class merging and separation strategy.

TABLE III. PERFORMANCE (DSC) COMPARISON OF THE PROPOSED METHODS FOR MM-WHS

	DSC(%)							
	LVC	RVC	LAC	RAC	MYO	AA	PA	AVG
Baseline	89.33	84.10	90.02	81.63	80.19	91.64	76.79	84.81
LVC MYO	93.12	85.44	84.03	82.10	90.46	93.84	70.76	85.68
MYO RVC	91.08	83.37	78.63	80.39	67.65	90.32	70.96	80.34
PA RVC	91.08	81.51	83.79	76.77	83.43	93.66	65.06	82.19
RAC RVC	92.41	84.73	74.87	83.09	62.79	94.08	64.50	79.50

TABLE IV. PERFORMANCE (IoU) COMPARISON OF THE PROPOSED METHODS FOR MM-WHS

	IoU(%)							
	LVC	RVC	LAC	RAC	MYO	AA	PA	AVG
Baseline	86.93	78.78	87.40	76.77	76.07	89.81	73.76	81.36
LVC MYO	90.95	80.08	81.20	77.40	86.18	92.03	68.01	82.26
MYO RVC	88.60	77.54	75.79	75.79	63.17	88.39	68.10	76.78
PA RVC	89.73	76.27	81.14	72.08	79.22	91.59	62.02	78.87
RAC RVC	90.25	79.10	72.10	78.43	58.44	92.00	60.91	75.89

The results of experiment on the MM-WHS dataset demonstrate that classes with sharply differentiated HU (e.g., LVC and MYO) can achieve significant performance improvements through effective boundary learning after merging, which also leads to slight performance gains in neighboring classes. Additionally, we found that classes with similar HU but well-defined boundaries (e.g., RVC and RAC) can benefit from improved performance after merging. Conversely, for classes with similar HU and significant structural variation across slices (e.g., PA and LAC), performance may degrade due to the model's focus on the merged class or blurred boundaries with the merged class. These findings suggest that an effective merging and separation strategy should account not only for pixel adjacency between classes, but also for the distinctiveness of HU and the structural consistency between slices.

TABLE V. PERFORMANCE (DSC) COMPARISON OF THE PROPOSED METHODS FOR FLARE22

	DSC(%)						
	Liver	RK	Spleen	Pancreas	Aorta	IVC	RAG
Baseline	91.65	94.25	95.79	78.79	94.11	88.16	92.10
IVC Liver	91.10	93.75	95.16	74.22	92.85	88.41	90.32
Gallbladder Liver	93.60	93.25	95.43	82.90	92.52	88.35	86.44
Pancreas Duodenum	92.15	93.25	96.16	75.62	91.51	88.22	88.28
Esophagus Stomach	92.92	93.73	96.54	82.21	93.74	88.70	86.75
	LAG	Gallbladder	Esophagus	Stomach	Duodenum	LK	AVG
Baseline	91.24	95.21	92.62	89.53	81.48	90.13	90.39
IVC Liver	86.82	88.69	90.19	83.50	79.74	88.70	87.96
Gallbladder Liver	88.02	91.19	89.10	90.48	80.02	91.65	89.46
Pancreas Duodenum	92.30	93.50	91.86	86.79	80.65	91.19	89.34
Esophagus Stomach	77.98	93.52	88.76	89.70	75.40	91.65	88.58

TABLE VI. PERFORMANCE (IoU) COMPARISON OF THE PROPOSED METHODS FOR FLARE22

	IoU(%)						
	Liver	RK	Spleen	Pancreas	Aorta	IVC	RAG
Baseline	90.13	93.23	95.10	75.08	91.44	84.51	90.37
IVC Liver	89.52	92.64	94.44	70.30	90.19	84.69	88.72
Gallbladder Liver	91.98	92.23	94.74	79.04	89.85	84.77	84.84
Pancreas Duodenum	90.51	92.21	95.54	71.56	88.96	84.62	86.68
Esophagus Stomach	91.29	92.63	95.76	78.38	91.06	84.99	85.06
	LAG	Gallbladder	Esophagus	Stomach	Duodenum	LK	AVG
Baseline	89.67	94.10	90.61	87.53	77.54	88.93	88.33
IVC Liver	85.25	87.42	88.25	81.17	75.54	87.43	85.81
Gallbladder Liver	86.36	90.21	87.26	88.36	75.78	90.52	87.38
Pancreas Duodenum	90.72	92.48	89.97	84.70	76.79	89.96	87.28
Esophagus Stomach	76.47	92.52	86.55	87.80	71.20	90.39	86.47

The experimental results on the Flare22 dataset indicate

that for the Liver and IVC, despite the substantial size difference between the two organs, the distinct difference in HU enables clear boundary learning for the IVC even after merging, resulting in improved performance for the smaller organ, IVC. In contrast, for the Liver and Gallbladder, as well as the Esophagus and Stomach, although the size difference is significant, the similarity in HU causes the model to prioritize the larger organ (e.g., Liver, Stomach), leading to a performance decline for the smaller organ (e.g., Gallbladder, Esophagus) while improving the performance of the larger organ.

Furthermore, in the case of the Pancreas and Duodenum, both organs exhibited significant inter-slice variability and complex structures, making boundary learning challenging after merging, which resulted in reduced performance for both organs. This demonstrates that merging organs with substantial inter-slice structural variability, in addition to HU similarities, can hinder effective boundary learning, leading to poor performance outcomes.

Therefore, when HU differences are distinct, clear boundary learning for smaller organs is possible, potentially enhancing their performance. However, when HU are similar and the size difference is large, the model tends to focus on the larger organ, which can result in performance degradation for the smaller organ. Additionally, given that boundary learning difficulties may arise when merging organs with significant inter-slice variability, it is suggested that inter-slice variability, along with HU differentiation, should be carefully considered in merging and separation strategies.

VI. CONCLUSION AND FUTURE WORK

In this study, we proposed a continual learning-based method to improve medical image segmentation by merging and separating classes with high pixel adjacency. Experiments conducted on the KNUH Body Composition, MM-WHS, and FLARE22 datasets demonstrated that merging and separating classes with high pixel adjacency reduced confusion and, in some cases, enhanced segmentation accuracy. Specifically, the merging and separating of classes with distinct HU led to better boundary learning and improved segmentation performance. However, for classes with similar HU and high structural variability, the results were inconsistent, and some classes experienced performance degradation.

These findings underline the need for a more carefully tailored merging and separation strategy, taking into account not only pixel adjacency but also HU differentiation and structural stability. Future work will focus on refining these strategies by using adaptive criteria for merging that rely on HU overlap, organ size, and effective feature extraction methods. Such an approach could help mitigate the issues faced by organs with similar HU or complex boundaries.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00360226) and also supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center)support program(IITP-2024-RS-2024-00437718) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)

REFERENCES

- [1] Ali, Shahzad, et al. "Abdominal CT Segmentation for Body Composition Assessment Using Network Consistency Learning." 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2023.
- [2] Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." *IEEE transactions on pattern analysis and machine intelligence* 40.12 (2017): 2935-2947.
- [3] Zhang, Yixiao, et al. "Continual learning for abdominal multi-organ and tumor segmentation." *International conference on medical image computing and computer-assisted intervention*. Cham: Springer Nature Switzerland, 2023.
- [4] Michieli, Umberto, and Pietro Zanuttigh. "Incremental learning techniques for semantic segmentation." *Proceedings of the IEEE/CVF international conference on computer vision workshops*. 2019.
- [5] Douillard, Arthur, et al. "Plop: Learning without forgetting for continual semantic segmentation." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.
- [6] FLARE Challenge. "FLARE22 Dataset." 2022. Grand Challenge, <https://flare22.grand-challenge.org/>.
- [7] Zhuang, Xiahai. "Multi-Modality Whole Heart Segmentation (MMWHS) Challenge." 2016. Zmic Lab, <https://zmiclalab.github.io/zxh/0/mmwhs/>.
- [8] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer International Publishing, 2015.
- [9] Hu, Shishuai, Jianpeng Zhang, and Yong Xia. "Boundary-aware network for kidney tumor segmentation." *Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 11*. Springer International Publishing, 2020.
- [10] Lee, Dong-Hyun. "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks." *Workshop on challenges in representation learning, ICML*. Vol. 3. No. 2. 2013.
- [11] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.
- [12] Loshchilov, Ilya, and Frank Hutter. "Sgdr: Stochastic gradient descent with warm restarts." *arXiv preprint arXiv:1608.03983* (2016).