Computers in Biology and Medicine 168 (2024) 107695

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Adaptive t-vMF dice loss: An effective expansion of dice loss for medical image segmentation

Sota Kato [a,*], Kazuhiro Hotta [b]

[a] Department of Electrical, Information, Materials and Materials Engineering, Meijo University, Tempaku-ku, Nagoya, 468-8502, Aichi, Japan
[b] Department of Electrical and Electronic Engineering, Meijo University, Nagoya, Japan

ARTICLE INFO

Keywords:
Computer vision
Pattern recognition
Deep learning
Medical image
Polyp segmentation
Multi-organ segmentation
Dice loss
T-vMF similarity

ABSTRACT

Dice loss is widely used for medical image segmentation, and many improved loss functions have been proposed. However, further Dice loss improvements are still possible. In this study, we reconsidered the use of Dice loss and discovered that Dice loss can be rewritten in the loss function using the cosine similarity through a simple equation transformation. Using this knowledge, we present a novel t-vMF Dice loss based on the t-vMF similarity instead of the cosine similarity. Based on the t-vMF similarity, our proposed Dice loss is formulated in a more compact similarity loss function than the original Dice loss. Furthermore, we present an effective algorithm that automatically determines the parameter κ for the t-vMF similarity using a validation accuracy, called Adaptive t-vMF Dice loss. Using this algorithm, it is possible to apply more compact similarities for easy classes and wider similarities for difficult classes, and we are able to achieve adaptive training based on the accuracy of each class. We evaluated binary segmentation datasets of CVC-ClinicDB and Kvasir-SEG, and multi-class segmentation datasets of Automated Cardiac Diagnosis Challenge and Synapse multi-organ segmentation. Through experiments conducted on four datasets using a five-fold cross-validation, we confirmed that the Dice score coefficient (DSC) was further improved in comparison with the original Dice loss and other loss functions.

1. Introduction

In recent years, segmentation tasks that assign the class label to each pixel in an image have become important in the field of medicine [1–7] and cell biology [8–17]. Although various network architectures have been proposed to improve the segmentation accuracy [18–27], Dice loss [28] has been used in most cases. Dice loss brings regions consisting of all positive examples predicted by a model closer to regions consisting of the ground truth. The Dice Score Coefficient (DSC) is a measure of overlap that is widely used to assess the segmentation performance when a ground truth is available, and the Dice loss can be optimized the DSC directly. Then it is a powerful way to achieve semantic segmentation. It is also effective for datasets with many negative samples in an image because it calculates the loss as a percentage of the region that can be detected. In recent studies, various loss functions based on Dice loss have been proposed to further improve the DSC [29–34].

However, in the case of a multi-class segmentation task, the areas of predicted regions are often different for each class and tend to have an imbalanced segmentation. Then, the classes with larger regions converge in training loss faster than the classes with smaller regions, and the classes with large regions are too trained. As a result, the DSC of classes with small regions tends to be lower than that of classes with large regions as shown in Table 3. Dice loss is not a perfect solution because the accuracies of classes of small areas are not sufficient compared to other classes, and conventional loss functions based on Dice loss have also been unable to solve this problem yet. The main goal of our study is to improve DSC for imbalanced semantic segmentation by the improvement of Dice loss.

Therefore, we reconsider a Dice loss equation and confirm which components are important for the Dice loss by intentionally omitting each component. As a result, we discovered that the cosine similarity is an important component for the original Dice loss, and it is able to rewrite in the loss function using only the cosine similarity based on the dot product of the features by normalizing both the regions consisting of all positive examples and ground truth. In addition, in recent years, the t-vMF similarity [35], which is an extended method for cosine similarity, has also been proposed. In this study, we present t-vMF Dice loss based on the t-vMF similarity. The t-vMF Dice loss is formulated in a more compact similarity loss function than the original Dice loss. Furthermore, we present an algorithm that automatically determines

* Corresponding author.
E-mail addresses: 150442030@ccalumni.meijo-u.ac.jp (S. Kato), kazuhotta@meijo-u.ac.jp (K. Hotta).
URL: https://usagisukisuki.github.io/ (S. Kato).

https://doi.org/10.1016/j.compbiomed.2023.107695
Received 25 November 2022; Received in revised form 30 October 2023; Accepted 6 November 2023
Available online 27 November 2023

the parameter κ for the t-vMF similarity using the Dice coefficient for the validation set, which we named Adaptive t-vMF Dice loss. Using Adaptive t-vMF Dice loss, it is possible to use more compact similarities for easy classes, and wider similarities for difficult classes.

We evaluated our loss function experimentally on four datasets. From the experimental results, although we changed only the loss function, we confirmed that the proposed method was significantly improved in comparison with the Dice loss and conventional loss functions for the image segmentation datasets.

This paper is organized as follows. Section 2 describes related works. Section 3 describes the details of the proposed method. Section 4 shows the experimental results. Finally, we describe our summary and future works in Section 5.

The main contributions of this paper are as follows:

- Through a simple formula transformation, we discovered that the original Dice loss was able to rewrite in the loss function using the cosine similarity, and we confirmed that the cosine similarity had the most influence over the training of networks using Dice loss.
- We proposed two types of novel loss functions, called t-vMF Dice loss and Adaptive t-vMF Dice loss. T-vMF Dice loss is formulated in a more compact similarity loss function than the conventional Dice loss, and Adaptive t-vMF Dice loss is able to use more compact similarities for easy classes and wider similarities for difficult classes.
- Unlike conventional loss functions for medical image segmentation [29,30,34], our novel loss functions can achieve high performance in various types of datasets in spite of setting only one parameter decided by humans.

2. Related works

Dice loss. In recent studies, Dice loss [28] has been the most frequently loss function used in medical image segmentation. Dice loss closes all positive instances predicted by a model to the ground truth, and is a powerful method for achieving a semantic segmentation because it can directly optimize the DSC. Furthermore, various loss functions based on the Dice loss have been proposed [2,29–32,34,36]. Sudre et al. [29] used the class re-balancing properties of the generalized Dice overlap and attempted to achieve balanced learning by using pixel-wise weight in addition to Dice loss, thereby varying the degree of convergence of the learning error for each class. Wang et al. [30] introduced a generalization of Dice loss for multi-class segmentation and the mean absolute error for robustness against noise, then they indicated that it outperforms existing noise-robust loss functions. Abraham et al. [34] used the Tversky similarity index, which is a generalization of the Dice score that can flexibly balance between false positives and false negatives, and then proposed the focal Tversky loss.

However, conventional loss functions based on Dice loss uses the same similarity function for all classes, and it is not sufficient to bridge the learning gap between classes with large and small area. In addition, although methods using the combined loss functions, like both Dice loss and cross entropy loss [37,38], have also been proposed, they are not fundamental solutions to the problem that uses the same similarity function for all classes.

Class-balancing weight. Weights based on the number of pixels/voxels of each class are widely used for imbalanced segmentation [39–42]. However, those methods are difficult to determine optimal weights and can lead to very extreme weights; for instance, the weight using the inverse of the number of pixels of each class would be a very small value, and it would have a negative impact on training. Then, we believe that the solution using the pixel-wise weight is not suitable.

Focal loss. From the perspective of imbalanced semantic segmentation, loss functions based on the focal loss [43] were proposed [44,45]. Focal loss can focus training on a sparse set of hard examples and prevent a large number of easy negatives from overwhelming the network during training. Yeung et al. [45] proposed the Unified Focal loss, which generalizes Dice and cross entropy based loss, and achieved the high DSC and IoU scores across the class imbalanced datasets. Li et al. [44] proposed a focal loss constrained residual network to solve a problem that the model mostly receives optimization signals from easy samples, and an alternative training fashion that trains the model with focal loss and dice loss alternatively. However, the study Yeung et al. [45] shows the results that there is less effectiveness, and the accuracy worsens.

Although various loss functions for semantic segmentation have been proposed, the imbalanced datasets for multi-class segmentation have yet to be sufficiently solved. We believe that conventional solutions for imbalanced segmentation have limited functionality and a novel approach is required. In addition, conventional loss functions need to set a lot of hyperparameters for each dataset. Our proposed loss functions have a more compact similarity and we can change the compactness of the similarity for each class. Then, the high performance is possible in comparison with conventional loss functions just by only setting one hyperparameter even if we work on imbalanced segmentation tasks.

3. Methodology

This section describes the analysis of the original Dice loss and our proposed loss function. In Section 3.1, we analyze the equation of the original Dice loss and present the potential for improving the level of accuracy. In Section 3.2, we present two novel loss functions called t-vMF Dice loss and Adaptive t-vMF Dice loss.

3.1. Analysis of the original Dice loss

Eq. (1) shows the original Dice loss [28].

$$Dice\ loss = \frac{1}{C}\sum_{i=1}^{C}\left(1 - \frac{2\sum_n A_{in}B_{in}}{\sum_n A_{in}^2 + \sum_n B_{in}^2}\right) \tag{1}$$

where C indicates the number of classes, n indicates the number of class samples, A_{in} indicates those vectors containing all positive examples predicted by a specific model, and B_{in} indicates the vectors containing all positive examples of the ground truth in the dataset. Then, when the cosine similarity is defined, Eq. (1) is expanded into Eq. (2).

$$\begin{aligned} Dice\ loss &= \frac{1}{C}\sum_{i=1}^{C}(1 - \cos\theta_i \cdot other_i) \\ \cos\theta_i &= \frac{\sum_n A_{in}B_{in}}{\sqrt{\sum_n A_{in}^2}\sqrt{\sum_n B_{in}^2}} \\ other_i &= \frac{2\sqrt{\sum_n A_{in}^2}\sqrt{\sum_n B_{in}^2}}{\sum_n A_{in}^2 + \sum_n B_{in}^2} \end{aligned} \tag{2}$$

In Eq. (2), we confirm that the original Dice loss contains two components; the cosine similarity ($\cos\theta_i$) and the other ($other_i$). These components are compatible with $\cos\theta_i \in [0,+1]$ and $other_i \in [0,+1]$ because the output used by the Dice loss is normalized using the sigmoid or softmax function.

Now, we evaluated the original Dice loss by intentionally omitting each component to confirm which components are important for the Dice loss. Table 1 shows the evaluation results on four datasets used by experiments of this study by the average DSC. As shown in Table 1, although the Dice loss without $\cos\theta_i$ significantly becomes less accurate than the original Dice loss, the Dice loss without $other_i$ is pretty much equal in accuracy compared to the original Dice loss in all datasets. For these analyses, we consider that the essential function of the Dice loss is to maximize the cosine similarity of all positive examples between predicted by a model and ground truth.

Table 1
Analyzed results of the Dice loss using the average DSC metric. The higher value indicates the predicted region is closer to the ground truth region.

Dataset	CVC-ClinicDB (%)	Kvasir-SEG (%)	ACDC (%)	Synapse (%)
Dice loss	82.02 ± 4.94	91.41 ± 3.93	92.21 ± 1.30	70.57 ± 7.38
w/o $\cos\theta_i$	47.67 ± 0.61	46.69 ± 1.55	24.44 ± 0.08	10.56 ± 0.33
w/o $other_i$	**82.42** ± **4.37**	**90.02** ± **0.71**	**91.98** ± **1.49**	**69.64** ± **7.96**

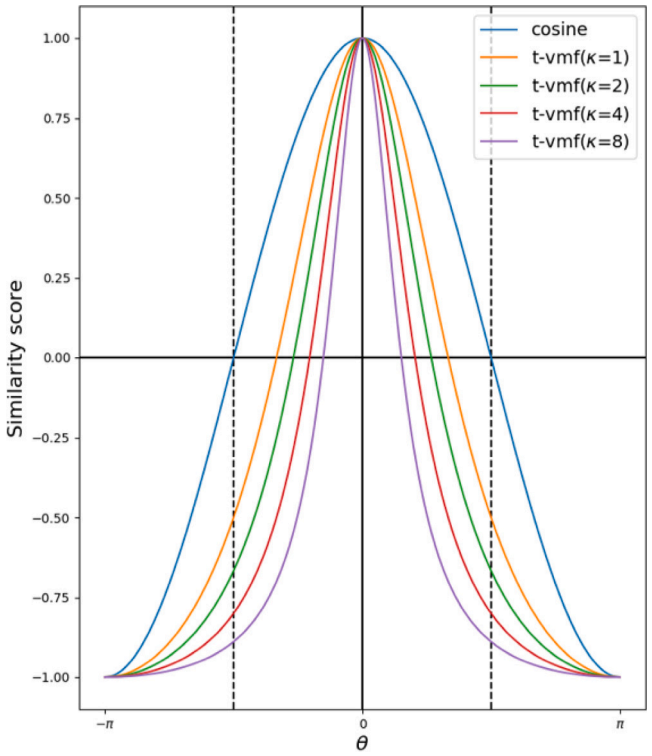


Fig. 1. Comparison of cosine similarity and t-vMF-based similarities.

After that, we normalize two vectors of A_{in} and B_{in} for each class to simplify expressions of Eq. (2) because the Dice loss only needs to maximize the cosine similarity in the result of Table 1. When we define $\sqrt{\sum_n A_{in}^2} = \sqrt{\sum_n B_{in}^2} = 1$ using L2 normalization, Eq. (2) can be rewritten into a new equation using the cosine similarity in Eq. (3).

$$Dice\ loss_{norm} = \frac{1}{C}\sum_{i=1}^{C}(1-\cos\theta_i) \tag{3}$$
$$\cos\theta_i = \sum_n A_{in}B_{in}$$

where C is the number of classes. This is compatible with $\cos\theta_i \in [0,+1]$ because the output used by the Dice loss is normalized using the sigmoid or softmax function. The original Dice loss is then a loss function that only maintains the cosine similarity between the set of values predicted by the model and the set of ground truth values at close to 1.

3.2. Adaptive t-vMF Dice loss

In Section 3.1, we confirmed that the original Dice loss is able to rewrite the equation using only the cosine similarity. Therefore, we adopt a novel approach applying the t-vMF similarity to the Dice loss. The t-vMF similarity, which is an extension of the cosine similarity was proposed in [35]. Eq. (4) shows the t-vMF similarity, and Fig. 1 shows a comparison of the cosine similarity and the t-vMF based similarities.

Algorithm 1 Adaptive t-vMF Dice loss

Require: The number of epoch E, Training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, Validation dataset $\hat{\mathcal{D}} = \{(\hat{x_j}, \hat{y_j})\}_{j=1}^{l}$, a parameterized model f_θ, the parameter $\mathcal{K} = \{\kappa_1, \kappa_2, ..., \kappa_c\}_{c=1}$, the parameter λ.

```
Initialize the model parameters θ randomly
Initialize 𝒦 to zero
for e = 1 to E do
    A mini-batch of m examples
    for t = 1 to T do
        ℬ ← SampleMiniBatch(𝒟, m)
        ℒ(f_θ) ← (1/m) Σ_{(x,y)∈ℬ} ℒ(x, y, 𝒦)
        f_θ ← f_θ − α∇_θ ℒ(f_θ)
    end for
    DSC_c ← DiceScoreCoefficient(D̂)
    𝒦 ← DSC_c × λ
end for
```

$$\phi_t(\cos\theta;\kappa) = \frac{1+\cos\theta}{1+\kappa(1-\cos\theta)} - 1 \tag{4}$$

By changing the parameter κ, the t-vMF similarity can freely change the range of similarities, and achieve a more compact similarity by using a large κ than the cosine similarity. Here, $\kappa = 0$ is exactly the same as the original cosine similarity, $\phi_t(\cos\theta;\kappa=0) = \cos\theta$ without treatment about practical computation. With this compact similarity and Eq. (4), we present a novel t-vMF Dice loss in Eq. (5).

$$t-vMF\ Dice\ loss = \frac{1}{C}\sum_{i=1}^{C}(1-\phi_t(\cos\theta_i;\kappa))^2 \tag{5}$$

When $\cos\theta \in [0,+1]$ and $\phi_t(\cos\theta;\kappa=1) \in [-0.5,+1]$, because the similarity takes a negative value, we use the mean squared error (MSE) loss to place the similarity closer to 1 according to the original Dice loss. From Eq. (5), the novel Dice loss has a more compact similarity than the conventional Dice loss.

However, the parameter κ is a fixed value in all classes, and we should use a suitable κ for each class. Therefore, we consider an effective algorithm for the t-vMF Dice loss to adaptively determine κ, and present Adaptive t-vMF Dice loss.

$$Adaptive\ t-vMF\ Dice\ loss = \frac{1}{C}\sum_{i=1}^{C}(1-\phi_t(\cos\theta_i;\kappa_i))^2 \tag{6}$$

As a measure for determining the parameter κ, we use the DSC for the validation set. Fig. 2 and Algorithm 1 show the flow of training and validation in the Adaptive t-vMF Dice loss. In Adaptive t-vMF Dice loss, κ is prepared for the number of classes, and they have initialized the value to zero at the first epoch. After we evaluate the DSC of each class during the validation step, we multiply the class DSC by λ to obtain the κ_c in the next epoch. Here, λ is a hyperparameter that determines the upper limit of κ, which we chose based on [35]. Using Adaptive t-vMF Dice loss, it is possible to dynamically choose the most suitable κ_c for each class based on DSC for the validation dataset, and achieve more compact similarities for an easy class and wider similarities for difficult classes.

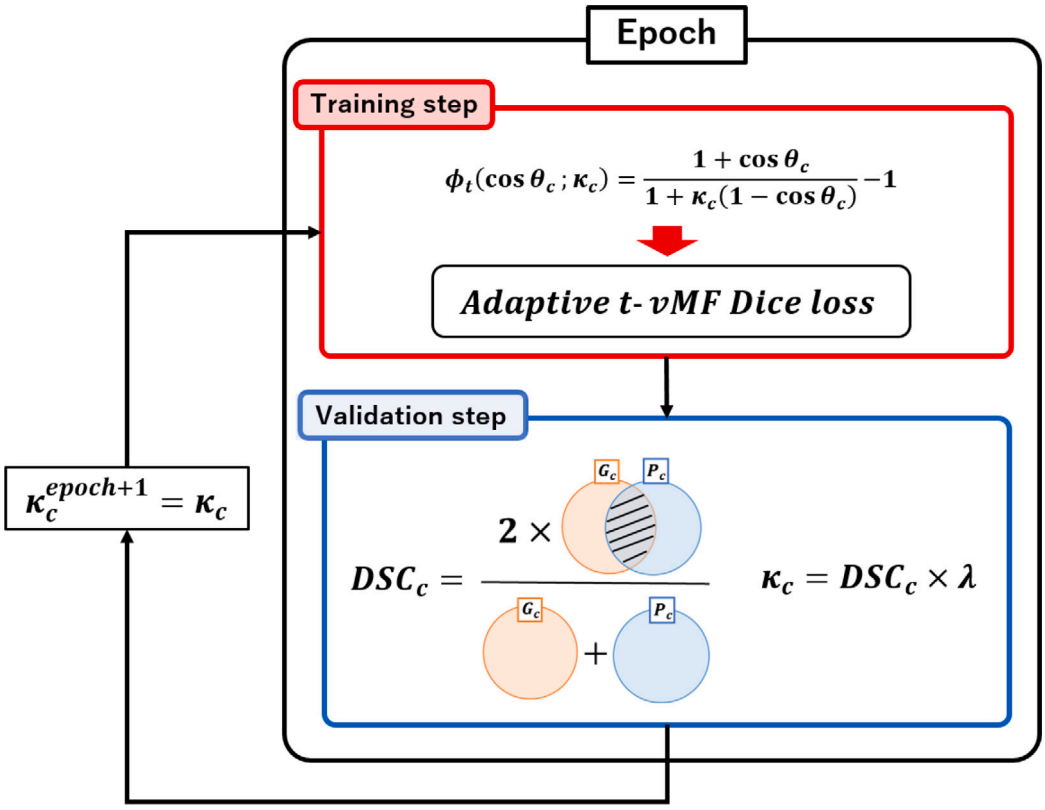


Fig. 2. Overview of training and validation flows using Adaptive t-vMF Dice loss.

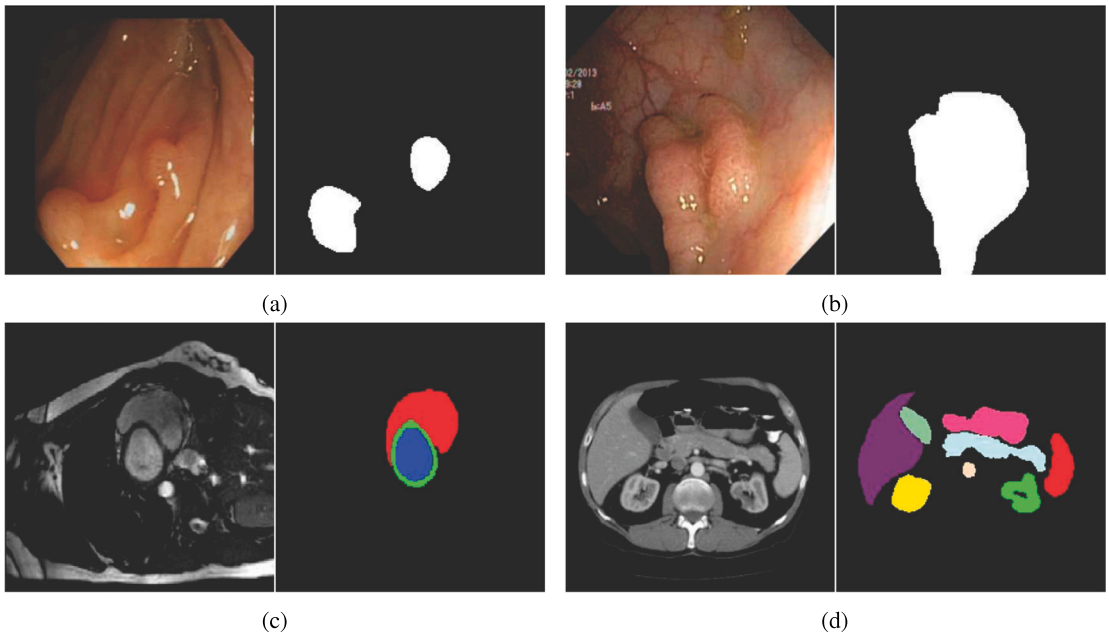


Fig. 3. Examples of datasets. (a) The CVC-ClinicDB dataset, (b) The Kvasir-SEG dataset, (c) ACDC dataset, and (d) Synapse multi-organ dataset.

4. Experiments

4.1. Datasets

We used the CVC-ClinicDB [46] and Kvasir-SEG [47] datasets for binary segmentation in Fig. 3(a) and (b). The CVC-ClinicDB dataset contains 612 RGB colonoscopy images with labeled polyps from MICCAI 2015 having a pixel resolution of 288 × 384. The Kvasir-SEG dataset contains 1000 polyp images with a pixel resolution ranging from 332 × 487 to 1920 × 1072 and their corresponding ground truth. Following Benčević et al. [48], we used an 80%, 10%, and 10% split for training, validation, and testing.

For multi-class segmentation, we used the Automated Cardiac Diagnosis Challenge (ACDC) [49] and Synapse multi-organ segmentation (Synapse) datasets in Fig. 3(c) and (d). The ACDC collects examination results from different patients, acquired from MRI scanners, found in

Table 2
Comparison results on CVC-ClinicDB using the DSC and IoU metrics. The higher value indicates the predicted region is closer to the ground truth region.

Loss function	UNet		TransUNet		FCBFormer	
	DSC (%) ↑	IoU (%) ↑	DSC (%) ↑	IoU (%) ↑	DSC (%) ↑	IoU (%) ↑
CE loss	81.83 ± 3.54	72.29 ± 4.18	90.64 ± 3.61	84.12 ± 5.24	90.84 ± 0.42	82.13 ± 7.41
Dice loss	82.02 ± 4.94	72.72 ± 5.92	91.41 ± 3.93	85.34 ± 5.87	90.15 ± 4.71	83.54 ± 6.62
CE loss + Dice loss	83.45 ± 4.06	74.35 ± 4.97	91.03 ± 4.17	84.80 ± 6.09	90.37 ± 4.62	83.87 ± 6.72
Focal loss + Dice loss	82.25 ± 3.98	72.87 ± 4.71	91.54 ± 3.88	85.52 ± 5.80	90.29 ± 4.31	83.68 ± 6.15
Generalised Dice loss	83.37 ± 2.53	74.00 ± 3.10	91.01 ± 4.37	84.78 ± 6.31	89.63 ± 5.43	82.90 ± 7.56
Noise-robust Dice loss	83.14 ± 4.71	74.00 ± 5.59	**91.91 ± 3.57**	**86.04 ± 5.43**	89.31 ± 5.02	82.36 ± 7.04
WS Dice loss	82.41 ± 3.10	72.89 ± 3.77	90.49 ± 4.82	84.07 ± 6.75	89.17 ± 5.94	82.31 ± 8.07
Focal Dice loss	84.38 ± 4.34	71.06 ± 3.10	91.13 ± 3.94	84.90 ± 5.86	90.35 ± 4.36	83.78 ± 6.29
$\beta FL - log(Dice)$ loss	86.57 ± 3.39	78.27 ± 4.38	89.22 ± 5.48	82.37 ± 7.61	89.66 ± 4.06	82.72 ± 6.02
Focal Tversky loss	83.39 ± 4.12	74.26 ± 4.87	91.81 ± 3.48	85.88 ± 5.29	90.02 ± 3.57	83.18 ± 5.23
Dice loss (with normalized)	82.44 ± 2.36	72.87 ± 2.79	90.43 ± 4.21	85.41 ± 6.07	89.89 ± 4.79	83.17 ± 6.75
t-vMF Dice loss ($\kappa = 2$)	83.05 ± 3.44	73.70 ± 4.35	90.51 ± 4.49	84.06 ± 6.51	87.91 ± 4.18	80.22 ± 5.94
t-vMF Dice loss ($\kappa = 32$)	86.77 ± 5.10	78.84 ± 6.77	90.46 ± 3.76	83.85 ± 5.39	87.84 ± 4.14	80.12 ± 5.87
t-vMF Dice loss ($\kappa = 128$)	88.16 ± 2.43	80.35 ± 3.37	90.76 ± 4.39	83.85 ± 5.39	87.45 ± 4.52	79.65 ± 6.31
t-vMF Dice loss ($\kappa = 512$)	87.57 ± 3.57	79.62 ± 4.82	90.50 ± 5.36	84.22 ± 7.48	88.67 ± 4.36	81.32 ± 6.10
Adaptive t-vMF Dice loss ($\lambda = 2$)	82.73 ± 3.59	73.37 ± 4.41	91.32 ± 3.76	85.16 ± 5.59	88.50 ± 5.77	81.31 ± 7.95
Adaptive t-vMF Dice loss ($\lambda = 32$)	87.39 ± 4.56	79.59 ± 6.13	90.71 ± 4.93	84.44 ± 7.19	**91.09 ± 4.36**	**84.90 ± 6.36**
Adaptive t-vMF Dice loss ($\lambda = 128$)	**88.68 ± 3.87**	**81.21 ± 5.41**	90.34 ± 4.21	83.76 ± 6.04	90.00 ± 4.34	83.27 ± 6.15
Adaptive t-vMF Dice loss ($\lambda = 512$)	87.68 ± 5.27	80.08 ± 6.89	90.53 ± 3.86	83.99 ± 5.66	90.67 ± 4.56	84.31 ± 6.73

the MICCAI 2017 dataset. Each patient scan is manually annotated for the left ventricle (LV), right ventricle (RV), and myocardium (MYO). Following Chen et al. [21], we split the data into 70 training cases, 10 validation cases, and 20 test cases. The Synapse multi-organ segmentation dataset is from the Multi-Atlas Abdomen Labeling Challenge with a total of 3779 axial contrast-enhanced abdominal clinical CT images from the MICCAI 2015 dataset. We used 30 abdominal CT volumes, where each CT volume consists of 85 to 198 slices having a pixel resolution of 512 × 512. Following Chen et al. [21], we evaluated 8 abdominal organs (aorta, gallbladder, spleen, left kidney, right kidney, liver, pancreas, spleen, and stomach) with a split of 18 training, 6 validation, and 6 test cases.

4.2. Dataset availability

All original datasets used by our experiments are available at.[1234] The code to produce our experiments including the preprocessed datasets itself has been made freely available at the GitHub repositories cited in the work.[5]

4.3. Training conditions and evaluation metrics

For comparison, we used Dice loss [28], Generalized Dice loss [29], Noise-robust Dice loss [30], WS Dice loss [31], Focal Dice loss [32], $\beta FL - log(Dice)$ loss [33], and Focal Tversky loss [34]. In addition, UNet [39] with full-scratch training, TransUNet [21], and FCBFormer [50] were applied as the networks. In TransUNet, we used the model combined ResNet50 [51] and ViT [52] as an encoder following Chen et al. [21]. ViT and ResNet50 were trained on ImageNet [53]. In FCBFormer, we used the pyramid vision transformer v2 [54] for the image encoder, which is pre-trained on ImageNet [53], following Sanderson and Matuszewski [50].

The batch size was set to 24, the number of training iterations was set to 4000, and the optimizer was a stochastic gradient descent (SGD) with a momentum of 0.9 and weight decay of 2×10^{-4}. We aligned a learning schedule using the rule described in [21]. The initial learning rate was 0.01, and we decayed by $lr \times (1 - iteration/iteration_{max})^{0.9}$, where lr is a learning rate and $iteration_{max}$ is the maximum number of iterations. For the data pre-processing, training samples were resized to 224 × 224, flipped horizontally, rotated with an angle randomly selected within $\theta = -90°$ to 90°, and normalized from zero to one. For the inference, the images were resized to 224 × 224 and normalized from zero to one.

We applied the Dice Score Coefficient (DSC) and the Interactive over Union (IoU) metrics to evaluate the accuracy of segmentation in our experiments. DSC and IoU are measures of overlap that are widely used to assess the segmentation performance and are defined by Eqs. (7) and (8).

$$DSC(G, P) = \frac{2\sum_{i=1}^{I} G_i P_i}{\sum_{i=1}^{I} G_i + \sum_{i=1}^{I} P_i} \tag{7}$$

$$IoU(G, P) = \frac{\sum_{i=1}^{I} G_i P_i}{\sum_{i=1}^{I} G_i + \sum_{i=1}^{I} P_i - \sum_{i=1}^{I} G_i P_i} \tag{8}$$

where G_i and P_i denote the ground truth and prediction masks for pixel i. When the predicted and ground truth regions completely overlap, the maximum value is one, and when there is no overlap at all, the minimum value is zero. All experiments were conducted using five-fold cross-validation and the average DSC and IoU metrics of five validations were used for evaluation. We utilized a single Nvidia RTX Quadro 8000 GPU as a calculator.

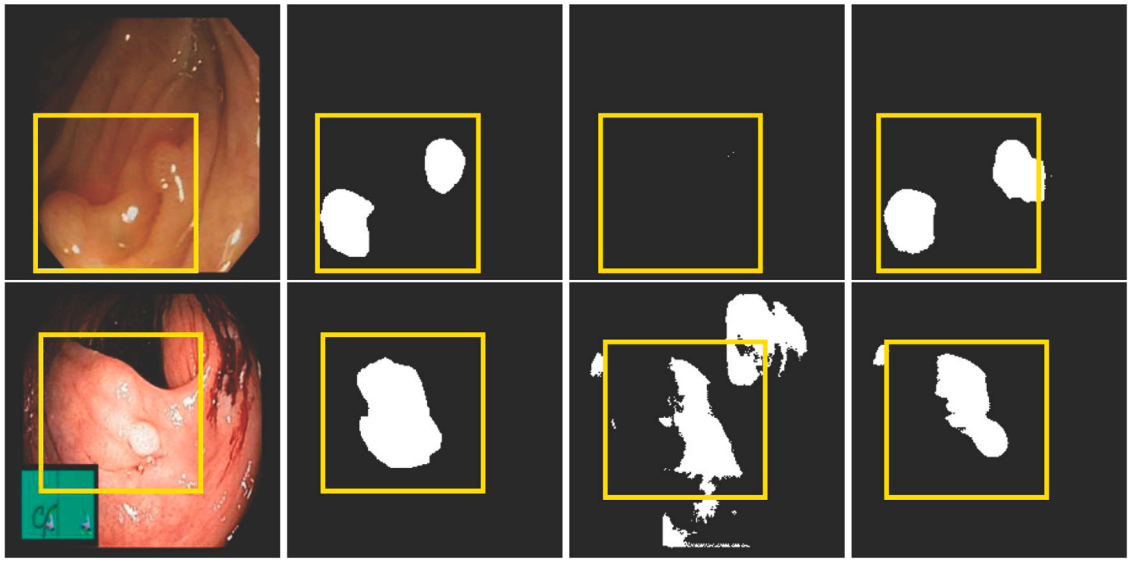
4.4. Results of methods

Tables 2 and 3 show the evaluation results on binary segmentations using the DSC and IoU metrics, and Fig. 4 presents a qualitative comparison of different approaches through a visualization of the CVC-ClinicDB and Kvasir-SEG datasets. The bold letters show the best score. For the CVC-ClinicDB dataset, in the average DSC metric, t-vMF Dice loss with $\kappa = 128$ was improved by over 6.14% in comparison with the original Dice loss when we used UNet, and Adaptive t-vMF Dice loss with $\lambda = 128$ was improved by over 6.66%. When we used TransUNet, t-vMF Dice loss with $\kappa = 2$ was the second highest DSC, and when we used FCBFormer, Adaptive t-vMF Dice loss with $\kappa = 32$ was the highest DSC and IoU scores. In the Kvasir dataset, the DSC metric was improved by over 2.11% when we used t-vMF Dice loss with $\kappa = 128$ for UNet in comparison with the original Dice loss, and over 0.05% when we used Adaptive t-vMF Dice loss with $\lambda = 128$ for TransUNet. When we used FCBFormer, t-vMF Dice loss with $\kappa = 2048$ was the highest DSC and IoU. In TransUNet and FCBFormer, the difference in

[1] https://polyp.grand-challenge.org/CVCClinicDB/.
[2] https://datasets.simula.no/kvasir-seg/.
[3] https://www.creatis.insa-lyon.fr/Challenge/acdc/.
[4] https://doi.org/10.7303/syn3193805.
[5] https://github.com/usagisukisuki/Adaptive_t-vMF_Dice_loss.

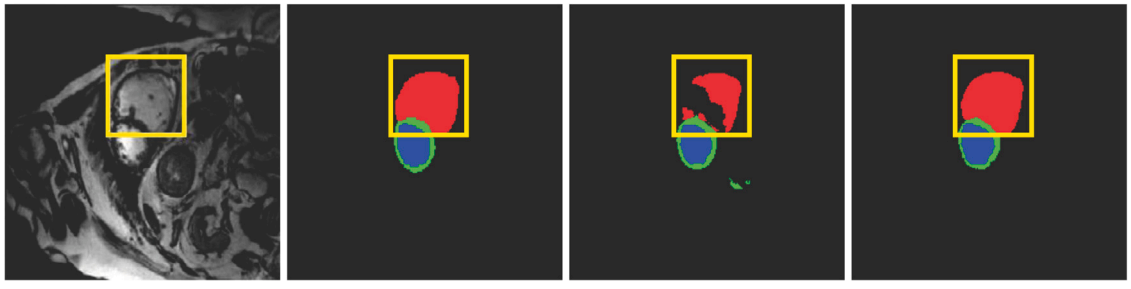
Table 3
Comparison results on Kvasir-SEG using the DSC and IoU metrics. The higher value indicates the predicted region is closer to the ground truth region.

Loss function	UNet		TransUNet		FCBFormer	
	DSC (%) ↑	IoU (%) ↑	DSC (%) ↑	IoU (%) ↑	DSC (%) ↑	IoU (%) ↑
CE loss	89.10 ± 5.36	83.86 ± 0.64	93.68 ± 0.37	88.44 ± 0.61	94.21 ± 0.38	89.33 ± 0.64
Dice loss	90.34 ± 0.71	83.09 ± 1.07	93.50 ± 0.49	88.14 ± 0.81	94.08 ± 0.53	89.11 ± 0.88
CE loss + Dice loss	91.25 ± 0.62	84.50 ± 0.97	93.48 ± 0.51	88.11 ± 0.86	94.27 ± 0.43	89.44 ± 0.73
Focal loss + Dice loss	90.19 ± 0.43	82.85 ± 0.67	**93.86 ± 0.29**	**88.74 ± 0.48**	94.02 ± 0.43	89.01 ± 0.71
Generalised Dice loss	88.70 ± 0.51	80.59 ± 0.74	93.21 ± 0.51	87.65 ± 0.84	93.62 ± 0.44	88.33 ± 0.73
Noise-robust Dice loss	90.34 ± 0.58	83.09 ± 0.91	93.53 ± 0.49	88.19 ± 0.81	94.19 ± 0.34	89.29 ± 0.57
WS Dice loss	90.30 ± 0.54	83.00 ± 0.81	92.67 ± 0.79	86.77 ± 1.27	93.90 ± 0.36	88.81 ± 0.59
Focal Dice loss	90.98 ± 0.51	84.07 ± 0.80	93.50 ± 0.54	88.15 ± 0.89	94.24 ± 0.46	89.38 ± 0.77
$\beta FL - log(Dice)$ loss	91.80 ± 0.54	85.39 ± 0.85	92.22 ± 0.57	86.05 ± 0.92	94.03 ± 0.59	89.04 ± 0.99
Focal Tversky loss	90.61 ± 0.20	83.50 ± 0.32	93.42 ± 0.37	88.00 ± 0.60	94.35 ± 0.39	89.58 ± 0.67
Dice loss (with normalized)	90.06 ± 0.41	82.65 ± 0.62	92.90 ± 0.60	88.47 ± 0.64	93.67 ± 0.52	88.42 ± 0.85
t-vMF Dice loss ($\kappa = 2$)	91.01 ± 0.42	84.13 ± 0.64	93.55 ± 0.30	87.40 ± 0.78	94.04 ± 0.33	89.05 ± 0.57
t-vMF Dice loss ($\kappa = 32$)	92.31 ± 0.30	86.18 ± 0.48	93.39 ± 0.36	86.06 ± 1.47	93.77 ± 0.66	88.59 ± 1.11
t-vMF Dice loss ($\kappa = 128$)	**92.45 ± 0.10**	**86.41 ± 0.14**	93.39 ± 0.36	86.61 ± 1.09	93.93 ± 0.53	88.87 ± 0.89
t-vMF Dice loss ($\kappa = 2048$)	90.71 ± 0.42	83.66 ± 0.67	93.67 ± 0.68	88.43 ± 1.13	**94.45 ± 0.40**	**89.74 ± 0.68**
Adaptive t-vMF Dice loss ($\lambda = 2$)	90.66 ± 0.57	83.58 ± 0.88	93.55 ± 0.30	88.22 ± 0.49	93.87 ± 0.30	88.76 ± 0.50
Adaptive t-vMF Dice loss ($\lambda = 32$)	92.18 ± 0.30	85.97 ± 0.46	92.66 ± 0.57	86.76 ± 0.91	93.97 ± 0.53	88.92 ± 0.88
Adaptive t-vMF Dice loss ($\lambda = 128$)	92.24 ± 0.32	86.07 ± 0.51	92.60 ± 0.32	86.65 ± 0.51	93.65 ± 0.72	88.40 ± 1.19
Adaptive t-vMF Dice loss ($\lambda = 2048$)	92.24 ± 0.32	84.16 ± 0.74	92.60 ± 0.32	88.58 ± 1.16	94.27 ± 0.51	89.34 ± 1.00



(a) (b) (c) (d)

Fig. 4. Qualitative comparison based on the visualization for binary segmentations. The top row is for the CVC-ClinicDB and the bottom row is for the Kvasir-SEG. (a) Input image, (b) Ground truth, (c) Dice loss, and (d) Adaptive t-vMF Dice loss ($\lambda = 128$).



(a) (b) (c) (d)

Fig. 5. Qualitative comparison based on the visualization for ACDC dataset. (a) Input image, (b) Ground truth, (c) Dice loss, and (d) Adaptive t-vMF Dice loss ($\lambda = 32$).

Table 4
Comparison results on ACDC dataset using the average DSC and average IoU metrics. The higher value indicates the predicted region is closer to the ground truth region.

Loss function	UNet		TransUNet		FCBFormer	
	mDSC (%) ↑	mIoU (%) ↑	mDSC (%) ↑	mIoU (%) ↑	mDSC (%) ↑	mIoU (%) ↑
CE loss	90.74 ± 1.74	82.29 ± 2.49	90.80 ± 1.63	83.96 ± 2.54	91.55 ± 1.18	85.06 ± 1.91
Dice loss	92.21 ± 1.30	85.63 ± 2.10	92.72 ± 0.80	86.94 ± 1.35	92.73 ± 0.84	86.93 ± 1.40
CE loss + Dice loss	92.48 ± 1.16	86.54 ± 1.92	92.77 ± 0.66	87.02 ± 1.12	92.71 ± 0.90	86.90 ± 1.50
Focal loss + Dice loss	92.32 ± 1.36	86.29 ± 2.22	92.86 ± 0.66	87.16 ± 1.13	92.74 ± 0.91	86.94 ± 1.52
Generalised Dice loss	92.42 ± 1.46	86.46 ± 2.39	93.01 ± 0.61	87.41 ± 1.05	92.93 ± 0.83	87.25 ± 1.46
Noise-robust Dice loss	92.40 ± 1.30	86.42 ± 2.14	92.84 ± 0.73	87.15 ± 1.25	92.93 ± 0.88	87.25 ± 1.39
WS Dice loss	90.57 ± 1.24	83.59 ± 1.97	91.03 ± 0.95	84.31 ± 1.47	90.84 ± 0.91	83.99 ± 1.44
Focal Dice loss	92.82 ± 1.29	87.10 ± 2.15	92.46 ± 0.70	86.52 ± 1.18	93.00 ± 1.02	87.37 ± 1.45
$\beta FL - log(Dice)$ loss	92.53 ± 1.37	86.63 ± 2.26	92.94 ± 0.63	87.30 ± 1.08	92.90 ± 0.91	87.20 ± 1.47
Focal Tversky loss	92.44 ± 1.47	86.53 ± 2.37	92.74 ± 0.47	86.97 ± 0.80	92.89 ± 0.92	87.20 ± 1.70
Dice loss (with normalized)	92.38 ± 1.34	86.39 ± 2.21	92.78 ± 0.73	87.04 ± 1.23	92.90 ± 0.92	87.21 ± 1.53
t-vMF Dice loss ($\kappa = 2$)	92.75 ± 1.37	86.98 ± 2.27	92.72 ± 0.78	86.95 ± 1.32	93.31 ± 0.73	87.87 ± 1.24
t-vMF Dice loss ($\kappa = 32$)	93.27 ± 1.44	87.87 ± 2.36	93.30 ± 0.60	87.90 ± 1.01	93.63 ± 0.85	88.40 ± 1.43
t-vMF Dice loss ($\kappa = 128$)	93.62 ± 0.71	88.40 ± 1.22	55.93 ± 10.07	49.43 ± 9.11	93.69 ± 0.60	88.51 ± 1.04
Adaptive t-vMF Dice loss ($\lambda = 2$)	92.79 ± 1.36	87.05 ± 2.26	92.79 ± 0.55	87.05 ± 0.95	93.09 ± 0.93	87.53 ± 1.58
Adaptive t-vMF Dice loss ($\lambda = 32$)	**93.68 ± 0.93**	**88.51 ± 1.59**	93.29 ± 0.78	87.88 ± 1.33	**93.71 ± 0.66**	**88.55 ± 1.13**
Adaptive t-vMF Dice loss ($\lambda = 128$)	93.50 ± 1.08	88.22 ± 1.84	**93.43 ± 0.66**	**88.11 ± 1.13**	93.63 ± 0.74	88.41 ± 1.25

Table 5
Comparison results on Synapse multi-organ segmentation dataset using the average DSC and average IoU metrics. The higher value indicates the predicted region is closer to the ground truth region.

Loss function	UNet		TransUNet		FCBFormer	
	mDSC (%) ↑	mIoU (%) ↑	mDSC (%) ↑	mIoU (%) ↑	mDSC (%) ↑	mIoU (%) ↑
CE loss	50.59 ± 3.95	40.04 ± 3.83	48.48 ± 5.88	40.94 ± 5.42	70.77 ± 5.12	58.05 ± 5.90
Dice loss	70.57 ± 7.38	57.62 ± 7.75	76.46 ± 6.72	64.79 ± 8.02	77.33 ± 5.87	65.57 ± 7.31
CE loss + Dice loss	71.49 ± 7.64	58.76 ± 8.23	76.49 ± 6.11	64.74 ± 7.38	77.36 ± 6.42	65.63 ± 8.11
Focal loss + Dice loss	70.55 ± 7.37	57.66 ± 7.85	76.84 ± 6.24	65.06 ± 7.62	77.17 ± 6.22	65.47 ± 7.66
Generalised Dice loss	66.53 ± 8.80	54.65 ± 8.66	63.42 ± 3.36	54.70 ± 4.89	69.10 ± 7.76	59.62 ± 8.40
Noise-robust Dice loss	64.74 ± 6.45	53.12 ± 6.63	60.42 ± 5.32	51.84 ± 6.57	71.56 ± 10.64	61.30 ± 10.57
Focal Dice loss	68.97 ± 6.86	55.57 ± 7.13	76.01 ± 5.89	64.16 ± 7.13	76.58 ± 6.40	64.60 ± 7.95
$\beta FL - log(Dice)$ loss	67.63 ± 7.97	55.42 ± 8.31	76.23 ± 6.22	65.28 ± 7.65	77.32 ± 7.26	65.70 ± 9.04
Focal Tversky loss	64.41 ± 5.80	52.62 ± 6.05	60.81 ± 4.83	48.18 ± 4.61	68.23 ± 6.65	56.12 ± 7.23
Dice loss (with normalized)	69.79 ± 7.66	57.07 ± 7.96	76.42 ± 6.46	64.57 ± 7.90	76.42 ± 6.46	65.55 ± 8.15
t-vMF Dice loss ($\kappa = 32$)	49.76 ± 3.66	40.10 ± 4.24	63.73 ± 3.53	53.77 ± 3.98	79.34 ± 6.78	68.49 ± 8.68
t-vMF Dice loss ($\kappa = 256$)	17.75 ± 1.00	13.43 ± 1.30	24.22 ± 1.87	19.63 ± 1.54	35.13 ± 11.04	29.22 ± 9.33
Adaptive t-vMF Dice loss ($\lambda = 32$)	73.81 ± 6.13	61.31 ± 6.80	**78.27 ± 5.54**	**67.14 ± 7.18**	**80.26 ± 5.91**	**69.43 ± 7.70**
Adaptive t-vMF Dice loss ($\lambda = 256$)	**74.22 ± 7.93**	**62.19 ± 8.78**	77.10 ± 5.06	66.29 ± 6.58	79.50 ± 6.35	68.52 ± 8.32

DSC and IoU was low because the encoder of networks was pre-trained on ImageNet, and the model performances are maxed out. In Fig. 4, the proposed loss function could segment tumors that conventional Dice loss could not segment well. Although the proposed losses are a simple expansion of the Dice loss, they are effective in comparison with conventional loss functions for semantic segmentation. Please see detailed ablation studies for hyperparameters κ and λ in Appendix A of the supplementary material.

Table 4 shows the evaluation results on the ACDC dataset with the DSC and IoU metrics, and Fig. 5 presents a qualitative comparison of different approaches through a visualization of the ACDC dataset. The mDSC shows a mean DSC and the mIoU shows a mean IoU. Although conventional loss functions have a lower average DSC and IoU than the original Dice loss, our proposed loss function has the highest average DSC and average IoU for all network architectures. Especially, Adaptive t-vMF Dice loss achieved the highest average DSC and IoU with $\lambda = 32$ for UNet and FCBFormer and $\lambda = 128$ for TransUNet. Compared to the Dice loss, we confirmed an improvement in average IoU over 2.88% for UNet, over 1.17% for TransUNet, and 1.62% for FCBFormer. In Fig. 5, Adaptive t-vMF Dice loss could segment the red class (Right ventricle) could not segment well. In the case of t-vMF Dice loss, a large κ gave a lower average DSC and IoU metrics. However, Adaptive t-vMF Dice loss further improves the average DSC and IoU, even if the parameter λ is larger and the loss function has a higher compact similarity. These results demonstrate that there is an appropriate κ for each class and Adaptive t-vMF Dice loss is effective for multi-class segmentation. Please see detailed ablation studies for the accuracy of each class and hyperparameters κ and λ in Appendix B of the supplementary material.

Table 5 shows the evaluation results on the Synapse multi-organ segmentation dataset using the average DSC and IoU metrics, and Fig. 6 presents a qualitative comparison of different approaches through a visualization. As shown in Table 5, conventional loss functions based on the original Dice loss, such as Generalised DIe loss and Focal Tversky loss, had lower accuracy than the Dice loss, and we can confirm that it is not effective for datasets with a large number of classes. One of the causes is setting too many hyperparameters, for instance, Focal Tversky loss has three hyperparameters. On the other hand, all you need to set in our proposed losses is only one hyperparameter. This makes training the model very easy, and Adaptive t-vMF Dice loss achieved the highest average DSC with $\lambda = 256$ for UNet and $\lambda = 32$ for TransUNet and FCBFormer. Compared to the original Dice loss, we confirmed an improvement in average DSC over 3.65% for UNet, over 1.81% for TransUNet, and 2.93% for FCBFormer. In the IoU metric, we confirmed an improvement of over 4.57% for UNet, over 2.35% for TransUNet, and 3.86% for FCBFormer. Furthermore, our loss functions achieved higher performance than conventional loss functions based on the Dice loss and were confirmed to have the most effective loss function for multi-class segmentation. In Fig. 6, the incorrect prediction of the green class (left kidney) as the yellow class (right kidney) and the red class (Spleen) as the purple class (Liver) was improved, and a correct recognition was achieved. We confirmed that our loss function predicted fewer false positives and maintained finer information. Fig. 7

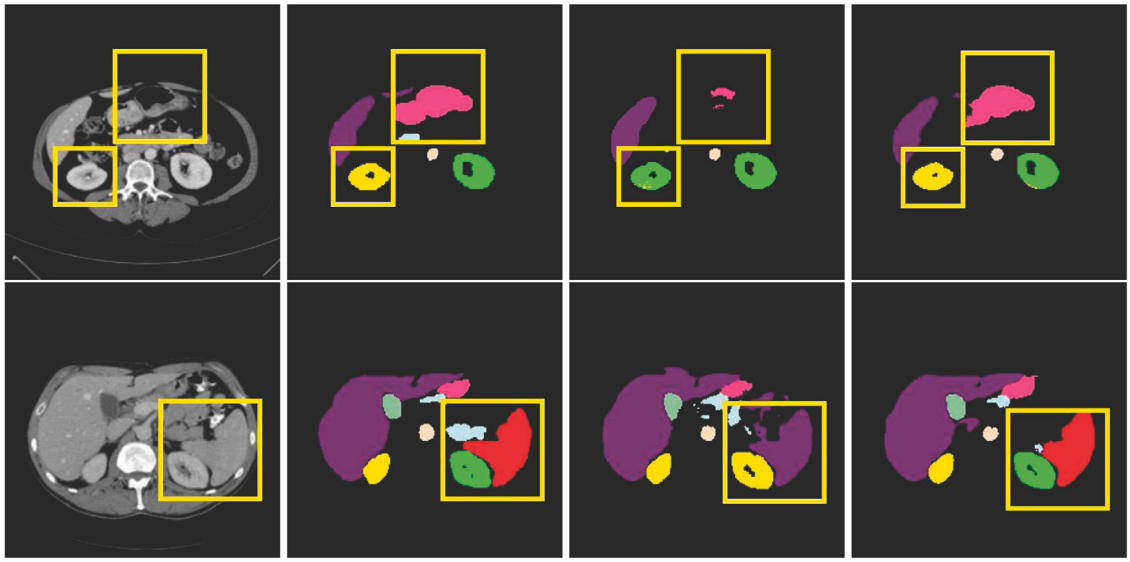


Fig. 6. Qualitative comparison based on the visualization for Synapse multi-organ segmentation dataset. (a) Input image, (b) ground truth, (c) Dice loss, and (d) Adaptive t-vMF Dice loss ($\lambda = 256$).

Table 6
Comparison results of the latest and SOTA methods. The higher value indicates the predicted region is closer to the ground truth region.

Dataset	Methods	Loss	mDSC (%) ↑	mIoU (%) ↑
CVC-ClinicDB	DUCK-Net (2023)	Baseline	96.30	92.87
		Ours	**96.84**	**93.43**
	Meta-Polyp (2023)	Baseline	92.93	86.80
		Ours	**93.73**	**88.19**
Kvasir-SEG	DUCK-Net (2023)	Baseline	91.95	85.10
		Ours	**92.20**	**85.52**
	Meta-Polyp (2023)	Baseline	90.86	82.79
		Ours	**93.70**	**88.22**
ACDC	MERIT (2023)	Baseline	91.83	85.30
		Ours	**92.01**	**85.55**
	FCT (2023)	Baseline	92.36	86.51
		Ours	**94.26**	**89.63**
Synapse	MERIT (2023)	Baseline	83.90	74.70
		Ours	**84.90**	**76.48**
	FCT (2023)	Baseline	72.79	63.24
		Ours	**76.66**	**65.14**

presents the DSC of each category. The red bar indicates UNet, the blue bar indicates TransUNet and the green bar indicates FCBFormer and the average DSC metric of each class is compared when using Dice loss and when using Adaptive t-vMF Dice loss ($\lambda = 32$). As shown in Fig. 7, in all categories, we can confirm that the class DSC metric is improved when the proposed loss function is used. Especially, although the classes with small regions, for instance, Gallbladder and Pancreas classes, have lower class DSC than that of the other classes even when we used Dice loss, it is possible to improve accuracy even for classes with small regions by using Adaptive t-vMF Dice loss. Please see detailed ablation studies for the accuracy of each class and hyperparameters κ and λ in Appendix C of the supplementary material.

Table 6 shows the comparison results of the latest and the state-of-the-art (SOTA) methods. For CVC-ClinicDB and Kvasir-SEG datasets, we evaluated DUCK-Net [25] and Meta-Polyp [26]. For ACDC and the Synapse multi-organ segmentation datasets, we evaluated Multiscale hiERarchical vIsion Transformer (MERIT) [22] and Fully Convolutional Transformer (FCT) [27]. DUCK-Net achieved SOTA results for polyp segmentation in the original paper. Trinh [26] proposed a fusion of MetaFormer [55] with UNet [39], and demonstrated that Meta-Polyp has a competitive performance and obtains the top result in the SOTA on the Kvasir-SEG dataset. Rahman and Marculescu [22] evaluated two widely used medical image segmentation benchmarks (i.e., Synapse Multi-organ dataset, ACDC dataset) and demonstrated the superior performance of MERIT over SOTA methods. Tragakis et al. [27] evaluated the ACDC Post-2017-MICCAI-Challenge online test set and demonstrated FCT is a new SOTA on unseen MRI test cases outperforming large ensemble models as well as nnUNet with considerably fewer parameters. The source code and preprocessed dataset used in the experiment were those published in the GitHub repository of each method. We also compared the loss function used in the original experiment with our proposed loss functions. The t-vMF Dice loss was used in experiments for CVC-ClinicDB and Kvasir-SEG datasets, and adaptive t-vMF Dice loss was used in experiments for ACDC and the Synapse multi-organ segmentation datasets. As shown in Table 6, we confirm the effectiveness of t-vMF Dice loss and Adaptive t-vMF Dice loss for all datasets and all networks. In particular, DUCK-Net with t-vMF Dice loss achieved the best accuracy for CVC-ClinicDB dataset, and Meta-Polyp with t-vMF Dice loss achieved the best accuracy for Kvasir-SEG dataset. Further in the ACDC datasets, we confirmed an improvement in mean IoU over 3.12% for FCT, and in the Synaps dataset, mean IoU was improved 1.78% for MERIT. We confirmed that our proposed loss functions are effective even with SOTA methods and can further update the SOTA results.

4.5. Ablation studies

Comparison of the hyper parameters. Fig. 8 shows the comparisons of our proposed t-vMF Dice loss and Adaptive t-vMF Dice loss when we evaluated various κ and λ parameters. Fig. 8(a) shows the results of t-vMF Dice loss and, Fig. 8(b) shows the results of Adaptive t-vMF Dice loss. As shown in Fig. 8(a), the average DSC was little change even if the parameter of κ was larger when we evaluated binary segmentations. However, for the multi-class segmentations, the larger κ was, the lower the average DSC was. These results indicate that the κ used by t-vMF Dice loss is a fixed value for each class, and we consider that classes with small regions may not be well trained due to a too compact similarity function generated by setting κ larger. On the other hand, as shown in Fig. 8(b), we can confirm no decrease in the average

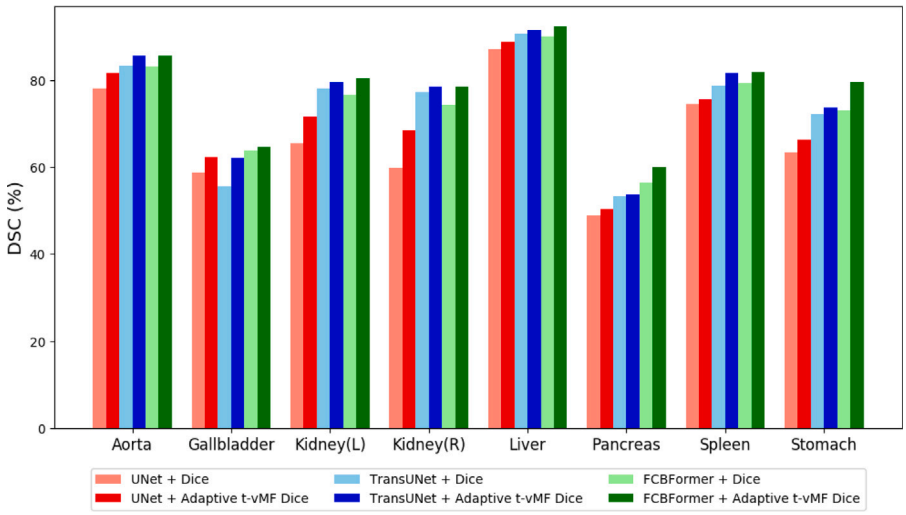
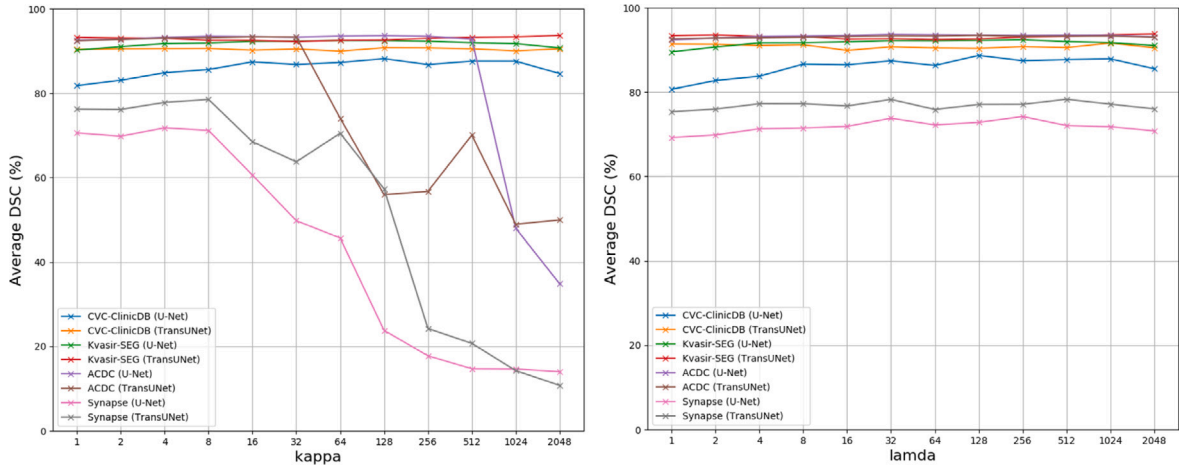


Fig. 7. Comparison among the average DSC metric of each category using Synapse multi-organ dataset. The higher value indicates the predicted region is closer to the ground truth region.



(a) (b)

Fig. 8. Comparisons of our proposed t-vMF Dice loss and Adaptive t-vMF Dice loss when we evaluated various κ and λ parameters using the average DSC metric. The higher value indicates the predicted region is closer to the ground truth region. (a) t-vMF Dice loss. (b) Adaptive t-vMF Dice loss.

Table 7
Comparison results of loss functions using the average DSC metric. The higher value indicates the predicted region is closer to the ground truth region.

Type		Loss	CVC-ClinicDB (%)	Kvasir-SEG (%)	ACDC (%)	Synapse (%)
t-vMF	$\kappa = 32$	Original	86.88 ± 3.79	92.09 ± 0.21	93.03 ± 1.10	26.57 ± 2.12
		MAE	**87.77 ± 3.92**	91.86 ± 0.62	93.27 ± 1.17	28.56 ± 2.87
		MSE	86.77 ± 5.10	**92.31 ± 0.30**	**93.27 ± 1.44**	**49.76 ± 3.66**
	$\kappa = 128$	Original	87.55 ± 2.11	91.90 ± 0.21	91.71 ± 1.79	11.88 ± 1.95
		MAE	87.90 ± 3.35	91.48 ± 0.42	91.69 ± 1.46	12.55 ± 1.70
		MSE	**88.16 ± 2.43**	**92.45 ± 0.10**	**93.62 ± 0.71**	**23.78 ± 2.90**
Adaptive	$\lambda = 32$	Original	**87.67 ± 3.63**	91.80 ± 0.43	93.44 ± 1.08	65.95 ± 9.60
		MAE	86.46 ± 4.93	92.15 ± 0.31	93.30 ± 1.24	66.17 ± 7.27
		MSE	87.39 ± 4.56	**92.18 ± 0.30**	**93.68 ± 0.93**	**73.81 ± 6.13**
	$\lambda = 128$	Original	87.86 ± 3.08	91.96 ± 0.21	93.39 ± 0.83	55.78 ± 7.29
		MAE	86.71 ± 4.41	92.21 ± 0.43	93.08 ± 1.37	53.03 ± 5.63
		MSE	**88.68 ± 3.87**	**92.24 ± 0.32**	**93.50 ± 1.08**	**72.82 ± 7.05**

DSC occurs for binary segmentations and multi-class segmentations. Adaptive t-vMF Dice loss can dynamically choose the most suitable κ for each class even if there is a class with a small region. Then, it is possible to use the more compact similarity function during training even if classes with small regions include, and achieve the higher DSC than when we use t-vMF Dice loss.

Comparison of loss functions to optimize. Table 7 shows the average DSC when we changed the loss functions of our method. In Table 7,

Original indicates $loss = 1 - \phi_t(\cos\theta_i; \kappa)$, MAE indicates the mean absolute error (MAE) loss $loss = |1 - \phi_t(\cos\theta_i; \kappa)|$, and MSE indicates the mean squared error (MSE) loss in Eqs. (5) and (6). As shown in Table 7, MSE outperformed for most of the results than other loss functions. Especially, in the case of the Synapse dataset, we can confirm that the average DSC was greatly reduced in the Original and MAE loss functions. MSE loss can give larger gradients to the model than other loss functions even if the difference between the prediction and the ground truth is smaller. Then, we consider that MSE loss outperformed the best average DSC. These comparisons demonstrated that the MSE loss is effective for t-vMF Dice loss and Adaptive t-vMF Dice loss.

5. Conclusion

In this paper, we showed that the original Dice loss can be rewritten, and introduced a novel t-vMF Dice loss using the t-vMF similarity. Furthermore, we presented Adaptive t-vMF Dice loss that automatically determines the parameter κ based on the validation accuracy, and it was possible to use more compact similarities for easy classes and wider similarities for difficult classes. As demonstrated through experiments conducted on four datasets, t-vMF Dice loss and Adaptive t-vMF Dice loss showed significantly improved accuracy in comparison with conventional loss functions. Additionally, we evaluated the latent SOTA methods using the proposed loss functions and confirmed that our loss functions can further update the SOTA results.

However, the accuracy of the class for small regions remains insufficient, and therefore improving the accuracy of these classes is one of our future studies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP23KJ2065 and JP22H04735.

Appendix A. Supplementary data

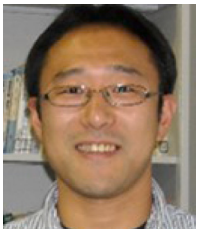
Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.107695.

References

[1] D. Müller, F. Kramer, MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning, BMC Med. Imaging 21 (1) (2021) 1–11.
[2] S. Shit, J.C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J.P. Pluim, U. Bauer, B.H. Menze, clDice-a novel topology-preserving loss function for tubular structure segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16560–16569.
[3] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, A.L. Martel, Loss odyssey in medical image segmentation, Med. Image Anal. 71 (2021) 102035.
[4] J.M.J. Valanarasu, P. Oza, I. Hacihaliloglu, V.M. Patel, Medical transformer: Gated axial-attention for medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 36–46.
[5] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 574–584.
[6] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2018, pp. 3–11.
[7] J. Ma, Y. He, F. Li, L. Han, C. You, B. Wang, Segment anything in medical images, 2023, arXiv preprint arXiv:2304.12306.
[8] E. Shibuya, K. Hotta, Feedback U-Net for cell image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 974–975.
[9] H. Fujii, H. Tanaka, M. Ikeuchi, K. Hotta, X-net with different loss functions for cell image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3793–3800.
[10] F.H. Araújo, R.R. Silva, D.M. Ushizima, M.T. Rezende, C.M. Carneiro, A.G.C. Bianchi, F.N. Medeiros, Deep learning for cell image segmentation and ranking, Comput. Med. Imaging Graph. 72 (2019) 13–21.
[11] M. Majurski, P. Manescu, S. Padi, N. Schaub, N. Hotaling, C. Simon Jr., P. Bajcsy, Cell image segmentation using generative adversarial networks, transfer learning, and augmentations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[12] Y. Lu, X. Qin, H. Fan, T. Lai, Z. Li, WBC-Net: A white blood cell segmentation network based on UNet++ and ResNet, Appl. Soft Comput. 101 (2021) 107006.
[13] T. Tran, O.-H. Kwon, K.-R. Kwon, S.-H. Lee, K.-W. Kang, Blood cell images segmentation using deep learning semantic segmentation, in: 2018 IEEE International Conference on Electronics and Communication Engineering (ICECE), IEEE, 2018, pp. 13–16.
[14] R. Conrad, K. Narayan, Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset, Cell Syst. 14 (1) (2023) 58–71.
[15] A. Wolny, L. Cerrone, A. Vijayan, R. Tofanelli, A.V. Barro, M. Louveaux, C. Wenzl, S. Strauss, D. Wilson-Sánchez, R. Lymbouridou, et al., Accurate and versatile 3D segmentation of plant tissues at cellular resolution, Elife 9 (2020) e57613.
[16] M. Zhao, Q. Liu, A. Jha, R. Deng, T. Yao, A. Mahadevan-Jansen, M.J. Tyska, B.A. Millis, Y. Huo, VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning, in: Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12, Springer, 2021, pp. 437–446.
[17] F. Padovani, B. Mairhörmann, P. Falter-Braun, J. Lengefeld, K.M. Schmoller, Segmentation, tracking and cell cycle analysis of live-cell imaging data with Cell-ACDC, BMC Biol. 20 (1) (2022) 174.
[18] J. Sun, F. Darbehani, M. Zaidi, B. Wang, Saunet: Shape attentive u-net for interpretable medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 797–806.
[19] X. Yan, H. Tang, S. Sun, H. Ma, D. Kong, X. Xie, After-unet: Axial fusion transformer unet for medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3971–3981.
[20] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, J. Wu, Unet 3+: A full-scale connected unet for medical image segmentation, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 1055–1059.
[21] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint arXiv:2102.04306.
[22] M.M. Rahman, R. Marculescu, Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation, in: Medical Imaging with Deep Learning, 2023.
[23] S. Roy, G. Koehler, C. Ulrich, M. Baumgartner, J. Petersen, F. Isensee, P.F. Jaeger, K.H. Maier-Hein, Mednext: transformer-driven scaling of convnets for medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 405–415.
[24] Q. Trinh, Meta-polyp: A baseline for efficient polyp segmentation, in: 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), IEEE Computer Society, Los Alamitos, CA, USA, 2023, pp. 742–747, http://dx.doi.org/10.1109/CBMS58004.2023.00312, URL: https://doi.ieeecomputersociety.org/10.1109/CBMS58004.2023.00312.
[25] R.-G. Dumitru, D. Peteleaza, C. Craciun, Using DUCK-Net for polyp image segmentation, Sci. Rep. 13 (1) (2023) 9803.
[26] Q.-H. Trinh, Meta-Polyp: a baseline for efficient Polyp segmentation, 2023, arXiv preprint arXiv:2305.07848.
[27] A. Tragakis, C. Kaul, R. Murray-Smith, D. Husmeier, The fully convolutional transformer for medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 3660–3669.
[28] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
[29] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 240–248.
[30] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, S. Zhang, A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images, IEEE Trans. Med. Imaging 39 (8) (2020) 2653–2663.
[31] L. Wang, C. Wang, Z. Sun, S. Chen, An improved dice loss for pneumothorax segmentation by mining the information of negative areas, IEEE Access 8 (2020) 167939–167949.

[32] P. Wang, A. Chung, Focal dice loss and image dilation for brain tumor segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2018, pp. 119–127.
[33] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, X. Xie, AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy, Med. Phys. 46 (2) (2019) 576–589.
[34] N. Abraham, N.M. Khan, A novel focal tversky loss function with improved attention u-net for lesion segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, pp. 683–687.
[35] T. Kobayashi, T-vMF similarity for regularizing intra-class feature distribution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6616–6625.
[36] L. Fidon, S. Shit, I. Ezhov, J.C. Paetzold, S. Ourselin, T. Vercauteren, Generalized wasserstein dice loss, test-time augmentation, and transformers for the BraTS 2021 challenge, in: International MICCAI Brainlesion Workshop, Springer, 2022, pp. 187–196.
[37] S.A. Taghanaki, Y. Zheng, S.K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, G. Hamarneh, Combo loss: Handling input and output imbalance in multi-organ segmentation, Comput. Med. Imaging Graph. 75 (2019) 24–33.
[38] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2) (2021) 203–211.
[39] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[40] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.
[41] J. Tian, N.C. Mithun, Z. Seymour, H.-P. Chiu, Z. Kira, Striking the right balance: Recall loss for semantic segmentation, in: 2022 International Conference on Robotics and Automation (ICRA), IEEE, 2022, pp. 5063–5069.
[42] B. Shirokikh, A. Shevtsov, A. Kurmukov, A. Dalechina, E. Krivov, V. Kostjuchenko, A. Golanov, M. Belyaev, Universal loss reweighting to balance lesion size inequality in 3D medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 523–532.
[43] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[44] C. Li, M. Chen, J. Zhang, H. Liu, Cardiac MRI segmentation with focal loss constrained deep residual networks, Phys. Med. Biol. 66 (13) (2021) 135012.
[45] M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation, Comput. Med. Imaging Graph. 95 (2022) 102026.
[46] J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, Comput. Med. Imaging Graph. 43 (2015) 99–111.
[47] D. Jha, P.H. Smedsrud, M.A. Riegler, P. Halvorsen, T.d. Lange, D. Johansen, H.D. Johansen, Kvasir-seg: A segmented polyp dataset, in: International Conference on Multimedia Modeling, Springer, 2020, pp. 451–462.
[48] M. Benčević, I. Galić, M. Habijan, D. Babin, Training on polar image transformations improves biomedical image segmentation, IEEE Access 9 (2021) 133365–133375, http://dx.doi.org/10.1109/ACCESS.2021.3116265.
[49] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M.A.G. Ballester, et al., Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging 37 (11) (2018) 2514–2525.
[50] E. Sanderson, B.J. Matuszewski, FCN-transformer feature fusion for polyp segmentation, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2022, pp. 892–907.
[51] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[52] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.
[53] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2009, pp. 248–255.
[54] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pvt v2: Improved baselines with pyramid vision transformer, Comput. Vis. Media 8 (3) (2022) 415–424.
[55] W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, S. Yan, Metaformer is actually what you need for vision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10819–10829.

Sota Kato received his B.Eng., M.Eng. degrees from Meijo University in 2019 and 2021. From 2021 to 2023, he was a Special Assistant at Meijo University. He is a JSPS research fellow (DC2). His research interests include image recognition methods using deep learning.



Kazuhiro Hotta received his B.Eng., M.Eng., and Dr.Eng. degrees from Saitama University in 1997, 1999, and 2002, respectively. From 1999 to 2002, he was a JSPS research fellow (DC1). From 2002 to 2010, he was an assistant professor at the University of Electro-Communications. From 2010 to 2018, he was an associate professor at Meijo University. Since 2018, he has been a professor at Meijo University. In 2012, he was a visiting scholar at the University of Maryland. His research interests include pattern recognition, computer vision, and machine learning. He is a member of IEEE, IEICE, IPSJ, and JSAI.