On the implementation and evaluation of loss functions for robust multiple anatomy segmentation on CT images

Chengyin Li[1], Rafi Ibn Sultan[1], Hassan Bagher-Ebadian[2], Yao Qiang[1], Kundan Thind[2], Dongxiao Zhu[1], Indrin J. Chetty[3]
1 College of Engineering, Department of Computer Science, Wayne State University, Detroit MI, USA
2 Department of Radiation Oncology, Henry Ford Cancer Institute, Detroit MI, USA
3 Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, USA

Abstract

In the realm of CT-based medical image segmentation, the selection of a loss function significantly influences deep neural network training. Loss functions, such as Cross Entropy (CE), Dice, Boundary, and Top-K, excel in specific areas but have inherent limitations for tissue segmentation. Our motivation arises from the understanding that relying solely on a single loss function can introduce bias to models. We have implemented a combination of CE, Dice, Boundary, and Top-K loss functions employing both loss-level linear combination and model-level ensemble on two large CT dataset cohorts to investigate model accuracy for auto-segmentation of normal tissues for different anatomic locations. Through extensive experiments on institutional and public datasets, consistent observations emerged: (1) linear combination of loss functions did not exhibit a clear advantage over single loss methods; (2) ensemble-based methods demonstrated a 2% - 7% average increase in DSC scores, and appreciable reductions in Hausdorff distances (HD's) and Average Surface Distances (ASD's) compared to both linear combination or single loss methods; and (3) ensemble approach with optimized weights generally characterized finer details in predicted masks, as evidenced by qualitative analyses. This study provides quantitative results on the accuracy of different loss functions employed in machine learning networks for automatic medical image segmentation. Enhanced accuracy was achieved using an ensemble approach with optimized model weights, compared to the use of a single loss function or a linear combination of loss functions.

1. Introduction

The training of DL models involves iterative parameter adjustments to minimize training errors, typically employing optimization methods such as stochastic gradient descent [1]. To quantify these errors, a loss function, also referred to as a cost function, is utilized to minimize the error between the algorithm prediction and the benchmark and thereby produce a convergent result. In the domain of medical image segmentation on CT images, CE and/or Dice are used commonly as loss functions [2-4]. CE loss [5] gauges the pixel-wise discordance between predicted class probabilities and ground truth labels. However, it tends to assign equal importance to background and foreground regions during training, resulting in potential class imbalance. Conversely, Dice loss [6] evaluates the overlap between predicted and ground truth labels, demonstrating effectiveness in handling class imbalance by emphasizing mask intersection, which is crucial for unbalanced classes. However, Dice as a sole loss function is subject to errors when dealing with small and irregularly shaped organ volumes. To enhance the precision of medical image segmentation, researchers have also introduced alternative loss functions such as Boundary loss [7] and Top-K loss [8]. Boundary Loss proves effective for smaller, imbalanced datasets but lacks global context and may not robustly address class imbalance. It is typically combined with other loss functions for more comprehensive segmentation tasks. Top-K loss aims to steer networks toward challenging samples during training, offering benefits for such classes or when imbalanced class distributions exist in the training dataset. Medical image segmentation relies on a range of loss functions, each with its unique strengths and weaknesses [9]. As such, segmentation based on a single loss model can introduce bias to the model training process. Improvement in loss-driven algorithms and computation power has made it possible to use these models simultaneously to improve robustness in medical image segmentation.

One method of concurrently utilizing multiple segmentation losses involves creating a unified model through a linear combination of various loss functions [9, 10]. This approach calculates a weighted sum of the individual loss function's contributions, with each loss being assigned a weight reflecting its relative significance within the overall loss. Such an approach permits the prioritization of specific aspects of the segmentation task, which can be tailored to domain knowledge, or the specific challenges posed by the dataset. Advanced options like ensemble learning [11] can be appealing for greater flexibility, especially when dealing with variations in architecture, hyperparameters, and training strategies. The ensemble approach frequently leads to enhanced segmentation performance by harnessing diversity among the models. Ensemble models have been extensively employed in medical image classification tasks, often in diverse dataset splitting scenarios and utilizing the CE loss function. However, there is a paucity of studies assessing the impact of ensemble DL models trained with different loss functions for medical image segmentation.

This study contributes to the existing knowledge as follows: (1) Evaluates the impact of unimodal and multimodal loss functions in multi-anatomy CT-based image segmentation; (2) Proposes a learnable ensemble approach that dynamically combines CE, Dice, Boundary, and Top-k loss-induced feature extraction modules to enhance segmentation performance and robustness; (3) Demonstrates the benefits of learnable ensemble models.

2. Materials and Methods

2.2.1. Network architecture

With the introduction of Fully Convolutional Networks (FCNs), the Attention U-Net (AttUNet) [12] stands out as an extension of the U-Net [5] architecture. AttUNet retains the U-Net's encoder-decoder architecture, capturing hierarchical features through generation of a segmentation map. AttUNet uses soft attention gates within skip connections to actively suppress irrelevant regions of the feature maps [12], thereby reducing redundant features and enabling characterization of fine details through the attention mechanism [12]. Recently, vision transformer approaches have emerged with the property of self-attention, which enables encoding of long-range dependencies and highly effective feature representations. Among these techniques, the shifted window UNet Transformer (SwinUNETR [4]) has shown superior accuracy for medical image segmentation. SwinUNETR exhibits several key design features: (a) adopts the Swin Transformer [13] architecture which captures long-range dependencies and contextual information through multi-layer self-attention mechanisms; (b) employs an encoder-decoder structure with the Swin Transformer to produce segmentation maps; (c) incorporates multi-scale processing via multiple encoder levels to handle objects of various sizes effectively; (d) integrates U-Net-style skip connections with residual operations to preserve spatial resolution and fine details; (e) employs patch-based image processing for efficient computations of large image datasets.

Here we employ AttUNet and SwinUNETR as two representative 3D segmentation architectures to investigate the impact of single and multiple loss functions on segmentation accuracy.

2.2.1. CE loss

Cross Entropy (CE) is derived from the Kullback-Leibler (KL) Divergence [14], which measures dissimilarity between two probability distributions, typically denoted as *P* and *Q*.

2.2.2. Dice loss

Dice loss directly optimizes the Dice Similarity Coefficient (DSC), a widely employed metric for segmentation evaluation. In general, two variants of the Dice loss are recognized (Isensee et al. 2021). One of them includes squared terms in the denominator (Milletari et al. 2016)

In the following experiments, the linear (non-squared version) as the default function.

2.2.3. Boundary loss

Boundary loss was introduced with the aim of minimizing dissimilarities between predicted and ground truth segmentations [7].

Here we use boundary loss in conjunction with the distribution-based CE loss:

$$\mathcal{L}_{CE+BD} = \alpha\mathcal{L}_{CE} + (1-\alpha)\mathcal{L}_{BD},$$

where α is a hyperparameter and can be optimized through empirical studies.

2.2.4. Top-K loss

Top-K loss is a variation of cross entropy designed to prioritize challenging samples during training. It retains the *K* percent worst pixels for loss, irrespective of their loss/probability values [8].

2.3.1 Linear combination

As depicted in Figure 1A, the combined loss or total loss is achieved by a linear combination of multiple losses.

$$\mathcal{L}_{Linear} = \lambda_1\mathcal{L}_{Dice} + \lambda_2\mathcal{L}_{CE} + \lambda_3\mathcal{L}_{CE+BD,}$$

where λ_1, λ_2, and λ_3 are the importance/contribution weights for each loss component. We found that when λ_1, λ_2, and λ_3 are set equally to the same weighting of 1.0, the obtained models after training showed better performance compared to other settings. Therefore, we represented the loss function as $\mathcal{L}_{Linear} = \mathcal{L}_{Dice} + \mathcal{L}_{CE} + \mathcal{L}_{CE+BD}$.

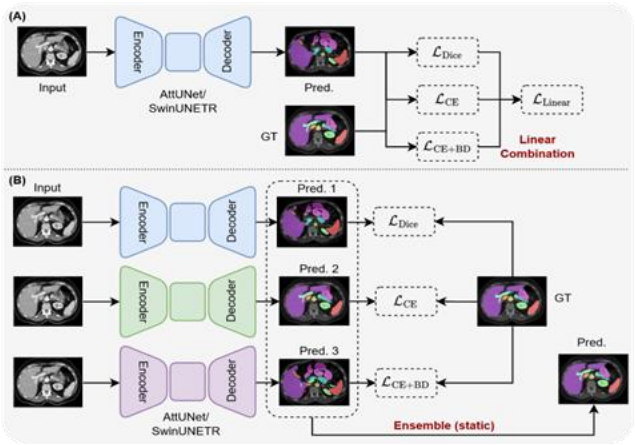

Figure 1: (a) The pipeline for a linear combination of different segmentation loss functions, and (b) the static multiple loss-driven model ensemble model strategy for CT image segmentation.

2.3.2. Static ensemble model

In this scenario (as shown in Figure 1B), we use an averaging ensemble technique by combining the predictions of multiple individual models trained with different loss functions. For instance, we have *K* different versions of trained models $f_{\theta_k} : k \in 1, ..., K$, where θ denotes the parameters of a model. Each of these versions of the model can rely on a specific loss function, e.g., $\mathcal{L}_{CE}$, $\mathcal{L}_{Dice}$, and $\mathcal{L}_{CE+BD}$.

The ensemble probabilistic prediction usually has a threshold of 0.5 to obtain a binary segmentation for each class for medical images. We refer to this approach as the static ensemble model.

2.3.3. Learnable ensemble model

With the learnable ensemble method, feature map extraction process is achieved by the concatenation of three branches of trained models (e.g., from $\mathcal{L}_{Dice}$, $\mathcal{L}_{CE}$, and $\mathcal{L}_{CE+BD}$). Then a few more iterations of training using the

Top-k loss function ($\mathcal{L}_{\text{Top-K}}$) are applied for the final segmentation prediction. We refer to this approach as the learnable ensemble model.

3.1. Datasets

All experiments were implemented on a server equipped with 8 Nvidia A100 GPUs, each with 40 GB of memory. All experiments are conducted in the PyTorch framework in Python 3.8.13, and each model is trained on a single GPU.

3.1.1. An institutional dataset: pelvic organs segmentation dataset

Institutional review board (IRB) approval was obtained for this study. We retrospectively selected planning CT and structure datasets from 300 patients diagnosed with prostate cancer. These 300 cases were randomly divided into three sets: a training set consisting of 225 cases, a validation set with 30 cases, and a testing set containing 45 cases. The testing datasets were held out from the training/validation steps and are considered "unseen". The models were trained for 200 epochs using the AdamW optimizer with an initial learning rate of 5e−4. An exponential learning rate scheduler with a warmup of 5 epochs was applied to the optimizer. Data augmentation was applied 'on the fly' during training. Random flip, rotation, and intensity scaling were used as augmentation transforms, with probabilities of 0.1, 0.1, and 0.2, respectively. We compared the automatic contours generated by our model against ground-truth .

3.1.2. Public dataset: CT organ segmentation dataset (CT-ORG)

A publicly available dataset has been used for evaluation in the CT-ORG task, as described in the reference [15]. The dataset comprises 140 CT scans, each including manual contours of the liver, lung, bladder, kidney, and bones. To maintain consistency with prior work [15], we adopted the same data split for training and testing. Nineteen CT datasets were reserved for testing (holdout), while the remaining 81 scans were used for training.

3.1.3. Ensemble modeling

In the static ensemble strategy, we directly perform averaging over several individually trained models, each utilizing different a loss function, without undergoing any further training. Conversely, in the learning ensemble strategy, we applied an additional training phase encompassing 20 epochs using the Top-k loss function for generating the final segmentation mask. We kept the other feature map generation modules static throughout this phase. We also applied the same data processing and augmentation techniques, albeit with a relatively smaller learning rate set to 1e−4.

We employed the Dice Similarity Coefficient (DSC), the 95th percentile Hausdorff Distance (HD), and Average Surface Distance (ASD) to assess the accuracy of segmentation.

3. Results

Table 1: DSC, HD, and ASD for the different settings using **AttUNet** on the **institutional pelvic dataset**. Each value represents the mean performance (+/- standard deviation). The most accurate results are shown in bold type. * denotes statistically significant differences (p<0.05) between the ensemble method and other non-ensemble methods.

Organ	Method	DSC ↑	HD (mm) ↓	ASD (mm) ↓
Prostate	$\mathcal{L}_{\text{Dice}}$	0.87 ± 0.03	4.66 ± 2.58	1.55 ± 0.53
	$\mathcal{L}_{\text{CE}}$	0.87 ± 0.02	4.48 ± 1.73	1.56 ± 0.42
	$\mathcal{L}_{\text{CE+BD}}$	0.87 ± 0.03	4.55 ± 1.78	1.63 ± 0.46
	$\mathcal{L}_{\text{Linear}}$	0.87 ± 0.03	4.68 ± 1.83	1.61 ± 0.42
	Ensemble (static)	0.88 ± 0.02*	4.39 ± 1.68	0.87 ± 0.03*
	Ensemble (learnable)	**0.89±0.02***	**4.12±1.63***	**0.79±0.06***
Bladder	$\mathcal{L}_{\text{Dice}}$	0.94 ± 0.02	2.62 ± 2.20	0.84 ± 0.45
	$\mathcal{L}_{\text{CE}}$	0.94 ± 0.02	2.70 ± 3.19	0.84 ± 0.63
	$\mathcal{L}_{\text{CE+BD}}$	0.94 ± 0.03	2.81 ± 3.19	0.93 ± 0.64
	$\mathcal{L}_{\text{Linear}}$	0.94 ± 0.02	2.70 ± 3.15	0.88 ± 0.62
	Ensemble (static)	0.95 ± 0.02*	**2.57±2.13***	0.82 ± 0.61*
	Ensemble (learnable)	**0.95±0.02***	2.65 ± 2.15*	**0.79±0.06***
Rectum	$\mathcal{L}_{\text{Dice}}$	0.87 ± 0.03	6.59 ± 5.42	1.66 ± 0.78
	$\mathcal{L}_{\text{CE}}$	0.87 ± 0.03	6.08 ± 4.80	1.47 ± 0.62
	$\mathcal{L}_{\text{CE+BD}}$	0.87 ± 0.03	6.70 ± 5.41	1.67 ± 0.75
	$\mathcal{L}_{\text{Linear}}$	0.87 ± 0.03	6.86 ± 6.08	1.62 ± 0.81
	Ensemble (static)	0.88 ± 0.03*	6.27 ± 5.19*	1.57 ± 0.73
	Ensemble (learnable)	**0.88±0.02***	**6.13±4.65***	**1.42±0.58***

Table 2: DSC, HD, and ASD for the different settings using **SwinUNETR** on the **institutional pelvic dataset**. Each value represents the mean performance (+/- standard deviation). The most accurate results are shown in bold type. * denotes statistically significant differences (p<0.05) between the ensemble method and other non-ensemble methods.

Organ	Method	DSC ↑	HD (mm) ↓	ASD (mm) ↓
Prostate	$\mathcal{L}_{\text{Dice}}$	0.86 ± 0.03	4.60 ± 1.42	1.61 ± 0.45
	$\mathcal{L}_{\text{CE}}$	0.85 ± 0.03	4.89 ± 1.82	1.67 ± 0.48
	$\mathcal{L}_{\text{CE+BD}}$	0.86 ± 0.03	4.81 ± 1.58	1.73 ± 0.43
	$\mathcal{L}_{\text{Linear}}$	0.86 ± 0.03	4.61 ± 1.22	1.62 ± 0.39
	Ensemble (static)	0.87 ± 0.02*	4.55 ± 1.23	1.59 ± 0.41*
	Ensemble (learnable)	**0.88±0.02***	**4.23±1.13***	**1.49±0.40***
Bladder	$\mathcal{L}_{\text{Dice}}$	0.94 ± 0.02	2.71 ± 2.17	0.82 ± 0.23
	$\mathcal{L}_{\text{CE}}$	0.93 ± 0.02	2.68 ± 1.69	0.83 ± 0.21
	$\mathcal{L}_{\text{CE+BD}}$	0.93 ± 0.03	2.97 ± 2.64	0.91 ± 0.31
	$\mathcal{L}_{\text{Linear}}$	0.93 ± 0.03	3.10 ± 3.01	0.89 ± 0.41
	Ensemble (static)	0.94 ± 0.02*	2.60 ± 1.53*	**0.81±0.20***
	Ensemble (learnable)	**0.95±0.02***	**2.53±1.54***	0.78 ± 0.17*
Rectum	$\mathcal{L}_{\text{Dice}}$	0.85 ± 0.05	7.47 ± 6.19	1.78 ± 0.97
	$\mathcal{L}_{\text{CE}}$	0.84 ± 0.04	8.14 ± 6.11	1.76 ± 0.87
	$\mathcal{L}_{\text{CE+BD}}$	0.84 ± 0.05	8.10 ± 6.37	2.17 ± 1.31
	$\mathcal{L}_{\text{Linear}}$	0.86 ± 0.04	8.23 ± 6.84	2.00 ± 1.11
	Ensemble (static)	0.86 ± 0.03*	7.11 ± 5.92*	1.77 ± 0.91
	Ensemble (learnable)	**0.88±0.02***	**6.04±4.21***	**1.36±0.61***

Table 3: DSC, HD, and ASD for the different settings using **SwinUNETR** on the **CT-ORG dataset**. Each value represents the mean performance (+/- standard deviation). The most accurate results are shown in bold type. * denotes statistically significant differences (p<0.05) between the ensemble method and other non-ensemble methods.

Organ	Method	DSC ↑	HD (mm) ↓	ASD (mm) ↓
Lung	$\mathcal{L}_{\text{Dice}}$	0.97 ± 0.02	10.37 ± 12.42	0.15 ± 0.14
	$\mathcal{L}_{\text{CE}}$	0.98 ± 0.01	6.17 ± 11.22	0.14 ± 0.13
	$\mathcal{L}_{\text{CE+BD}}$	0.98 ± 0.01	5.56 ± 7.78	0.15 ± 0.13
	$\mathcal{L}_{\text{Linear}}$	0.98 ± 0.01	7.29 ± 13.99	0.15 ± 0.13
	Ensemble (static)	0.98 ± 0.01*	**4.97±7.61***	0.14 ± 0.13*
	Ensemble (learnable)	**0.98±0.01***	5.23 ± 6.25*	**0.12±0.11***
Liver	$\mathcal{L}_{\text{Dice}}$	0.94 ± 0.02	3.90 ± 3.36	0.65 ± 0.37
	$\mathcal{L}_{\text{CE}}$	0.94 ± 0.01	3.03 ± 3.49	0.59 ± 0.17
	$\mathcal{L}_{\text{CE+BD}}$	0.94 ± 0.02	3.10 ± 1.46	0.72 ± 0.29
	$\mathcal{L}_{\text{Linear}}$	0.95 ± 0.02	2.91 ± 1.83	0.58 ± 0.21
	Ensemble (static)	0.96 ± 0.01*	2.55 ± 1.34*	0.56 ± 0.15*
	Ensemble (learnable)	**0.96±0.01***	**2.12±0.93***	**0.35±0.10***
Kidney	$\mathcal{L}_{\text{Dice}}$	0.92 ± 0.03	3.26 ± 1.11	0.61 ± 0.21
	$\mathcal{L}_{\text{CE}}$	0.91 ± 0.03	6.29 ± 14.08	0.58 ± 0.19
	$\mathcal{L}_{\text{CE+BD}}$	0.91 ± 0.03	6.97 ± 13.76	0.69 ± 0.22
	$\mathcal{L}_{\text{Linear}}$	0.92 ± 0.03	3.03 ± 1.76	0.52 ± 0.17
	Ensemble (static)	0.93 ± 0.03*	3.25 ± 0.96	0.56 ± 0.21*
	Ensemble (learnable)	**0.94±0.02***	**2.21±0.82***	**0.46±0.22***
Bladder	$\mathcal{L}_{\text{Dice}}$	0.86 ± 0.08	9.97 ± 18.31	0.77 ± 0.48
	$\mathcal{L}_{\text{CE}}$	0.83 ± 0.13	7.29 ± 12.28	0.90 ± 0.80
	$\mathcal{L}_{\text{CE+BD}}$	0.83 ± 0.13	5.71 ± 10.73	1.01 ± 0.71
	$\mathcal{L}_{\text{Linear}}$	0.84 ± 0.13	5.86 ± 10.54	0.89 ± 0.67
	Ensemble (static)	0.87 ± 0.09*	4.71 ± 8.90*	0.81 ± 0.61
	Ensemble (learnable)	**0.87±0.08***	**3.56±6.38***	**0.68±0.49***
Bone	$\mathcal{L}_{\text{Dice}}$	0.88 ± 0.04	4.91 ± 3.52	0.76 ± 0.35
	$\mathcal{L}_{\text{CE}}$	0.89 ± 0.05	5.50 ± 4.25	0.98 ± 0.37
	$\mathcal{L}_{\text{CE+BD}}$	0.88 ± 0.05	5.84 ± 3.87	1.20 ± 0.48
	$\mathcal{L}_{\text{Linear}}$	0.89 ± 0.05	4.88 ± 2.82	1.04 ± 0.28
	Ensemble (static)	0.90 ± 0.04*	4.85 ± 3.28*	0.98 ± 0.46

4. Discussion

The learnable ensemble model achieved the most accurate segmentation results across nearly all metrics for the institutional pelvic dataset and the public CT-ORG dataset across multiple organs.

These findings agree with previous studies showing improved accuracy of ensemble models for image analysis tasks, albeit on different network architectures [16, 17]. By combining multiple models trained on different loss functions, the proposed ensemble approach overcomes limitations of individual loss functions and integrates their complementary strengths. For instance, CE loss optimizes pixel-wise probabilities, enabling precise localization but ignoring class imbalances. Dice loss directly maximizes overlap with ground truth, overcoming imbalance but sacrificing localization accuracy. Boundary loss focuses on contour distances, handling smaller objects effectively. Top-K loss concentrates on challenging samples regardless of loss value. No single loss completely addresses all segmentation challenges simultaneously. The ensemble approach enables harnessing of complementary advantages of different losses.

The learnable ensemble model achieves superior performance by increasing model diversity through an additional trainable module, enabling dynamic weighting instead of fixed averaging to aggregate predictions, and allowing targeted optimization of challenging samples. This combination provides greater segmentation accuracy and robustness compared to the static ensemble or individual models with distinct loss functions, as evidenced by the statistically significant improvements in segmentation accuracy across multiple anatomic locations. However, the ensemble approach also has limitations. Training multiple models increases computational requirements for image analysis. Additionally, increasing the sample size will benefit from understanding the tradeoffs between variance and bias in the training datasets. Although we have utilized data augmentation, the influence of augmentation on the training data bias, variance and ultimately the generalization error is not fully understood. Determining the optimal combination of loss functions and models for a given segmentation task is a challenging task and one that is under further investigation. Studies providing greater insight into designing high-performance ensembles tailored to different imaging modalities and anatomical structures would be of benefit to the field.

5. Conclusion

Comprehensive experiments, conducted on institutional and public datasets, provide evidence of the superior performance achieved through a learnable ensemble model compared to results based solely on a single loss function or a linear combination of loss functions. These findings will inform future efforts to develop automated segmentation techniques for significantly enhancing efficiency and accuracy of segmentation of anatomy on CT images.

Acknowledgement: This work was supported in part by a grant from Varian Medical Systems, Siemens Healthineers.

References

[1] Wang L, Yang Y, Min R, et al. Accelerating deep neural network training with inconsistent stochastic gradient descent. Neural Networks. 2017;93:219-29.

[2] Li C, Bagher-Ebadian H, Goddla V, et al. FocalUNETR: A Focal Transformer for Boundary-aware Segmentation of CT Images. arXiv preprint arXiv:221003189. 2022.

[3] Balagopal A, Kazemifar S, Nguyen D, et al. Fully automated organ segmentation in male pelvic CT images. Physics in Medicine & Biology. 2018;63:245015.

[4] Li C, Bagher-Ebadian H, Sultan RI, et al. A new architecture combining convolutional and transformer-based networks for automatic 3D multi-organ segmentation on CT images. Med Phys 50:6990; 2023

[5] Ronneberger O, Fischer P, Brox T. 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; 2015. p. 234-41.

[6] Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV): IEEE; 2016:565-71.

[7] Kervadec H, Bouchtiba J, Desrosiers C, et al. Boundary loss for highly unbalanced segmentation. International conference on medical imaging with deep learning: PMLR; 2019. p. 285-96.

[8] Wu Z, Shen C, Hengel Avd. Bridging category-level and instance-level semantic image segmentation. arXiv:160506885. 2016.

[9] Ma J, Chen J, Ng M, et al. Loss odyssey in medical image segmentation. Medical Image Analysis. 2021;71:102035.

[10] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. Proceedings of the IEEE conference on computer vision and pattern recognition2018. p. 7482-91.

[11] Dong X, Yu Z, Cao W, et al. A survey on ensemble learning. Frontiers of Computer Science. 2020;14:241-58.

[12] Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018.

[13] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF international conference on computer vision2021. p. 10012-22.

[14] Csiszar I. Divergence Geometry of Probability Distributions and Minimization Problems. The Annals of Probability. 1975;3:146-58, 13.

[31] Rister B, Yi D, Shivakumar K, et al. CT-ORG, a new dataset for multiple organ segmentation in computed tomography. Scientific Data. 2020;7:381.

[32] Rajaraman S, Zamzmi G, Antani SK. Novel loss functions for ensemble-based medical image classification. Plos one. 2021;16:e0261307.

[33] Müller D, Soto-Rey I, Kramer F. An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks. Ieee Access. 2022;10:66467-80.