Master's Thesis

Robust Medical Image Segmentation using Uncertainty-Adaptive Loss Functions

M243422 Tomoya Hiroike

Supervisor Assoc. Prof. Akira Furui

February 3, 2026

Informatics and Data Science Program
Graduate School of Advanced Science and Engineering
Hiroshima University

Contents

1 Introduction

This is a LaTeX format template intended for use in Master's theses at the Graduate School of Advanced Science and Engineering, Information Sciences Program, Hiroshima University. The chapter and section structure is tentative; please modify it according to your content. Do not modify font sizes or line spacing as these changes may disrupt the layout.

When citing previous literature in the text, use the `\cite` command. Use the citation key listed in the reference list, writing it as:
"`...a method has been proposed based on~\cite{Furui2019-bz}.`" (The `~` prevents the citation symbol from being separated to the next line).

By labeling figures and equations using the `\label{label_name}` command, you can easily retrieve figure or equation numbers in the text by writing `\ref{label}`. For example, a figure labeled with `\label{fig:method}` can be cited as
"`The proposed method is shown in Fig.~\ref{fig:method}`", and an equation labeled with `\label{eq:model}` can be cited as
"`As shown in equation (\ref{eq:model})...`". Avoid manually typing figure or equation numbers.

Various other LaTeX command examples are provided in the following chapters for reference. Note that figures (illustrations and graphs) should be prepared as PDF files. The method for inserting figure files is described in Chapter 3.

In the following, Chapter 2 describes a methodology and various closely related existing methods. Chapter 3 explains the proposed XX, and Chapter 4 describes the experimental method, results, and discussion. Finally, Chapter 5 presents the conclusions and future work.

2 Related Work

This chapter organizes related research concerning XX from various perspectives.

2.1 Methods Based on XX

Traditionally, various approaches have been taken to achieve XX. In particular, the XX method [1] deals with YY and can be expressed by the following equation:

$$\hat{\theta} = \arg\min_{\theta \in \mathcal{S}} \mathcal{J}(\theta). \tag{1}$$

However, as shown in equation (1), this approach has limitations in terms of ZZ. The proposed method resolves this issue by taking an approach based on AA.

2.2 Methods Based on YY

Meanwhile, approaches based on BB have also been proposed as methods to improve CC. Among these, the XX method [2] enables DD.

However, this method also has limitations due to EE and faces constraints in FF. In contrast, the proposed method enables GG by implementing HH.

3 Proposed Method

Fig. 1 shows an overview of the proposed method. In the proposed method, MC Dropout is used during training to perform multiple inferences for each image, and uncertainty is quantified on a per-pixel basis from the variance of the predictions. This uncertainty information reflects how "difficult" the model finds the segmentation of that image. Subsequently, synchronized with the timing of this uncertainty assessment, this uncertainty information is aggregated on a per-image basis to dynamically control the shape of the PolyDice-1 Loss, thereby achieving adaptive learning which applies a steep gradient to difficult images and a gentle gradient to easy images.

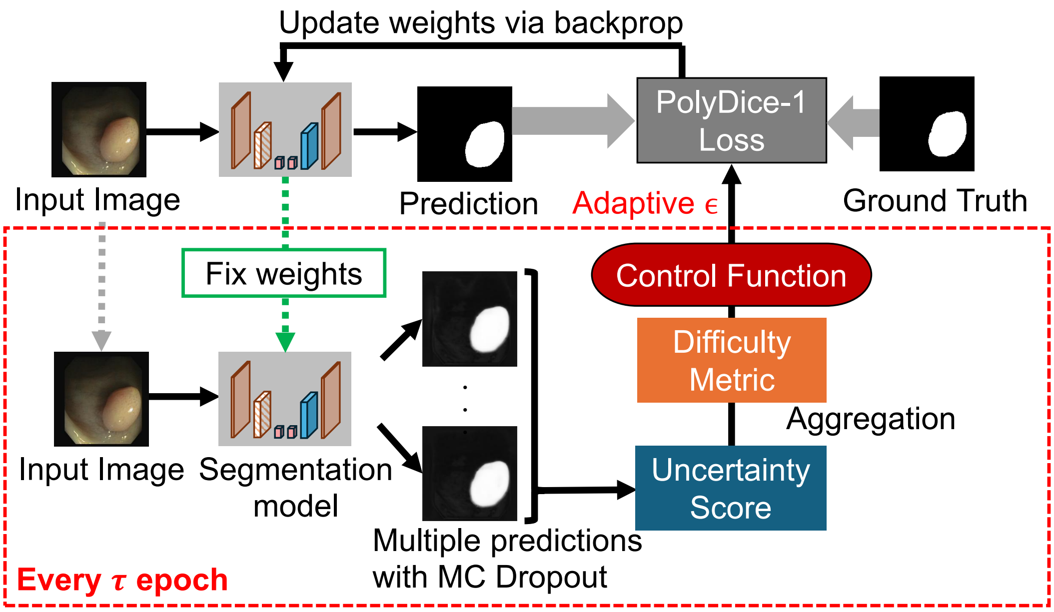

Fig. 1. Overview of the prposed method

3.1 Controlling Loss Shape with PolyDice-1 Loss [3]

Dice Loss, which is widely used in medical image segmentation, is robust to class imbalance but has the constraint of having a fixed shape for all images. In this study, we adopt PolyDice Loss, an extension of Dice Loss achieved through polynomial expansion [3], specifically its practical form, PolyDice-1 Loss. PolyDice-1 Loss can control the shape of the loss function with a single parameter ϵ, making it possible to adjust the steepness of the gradient according to the difficulty of the image.

3.1.1 Dice Loss

Let the image size be $H \times W$, and let the pixel position be denoted by (i, j), where $(i \in \{1, \ldots, H\}, j \in \{1, \ldots, W\})$. In a segmentation task, if $\hat{\mathbf{Y}} = \{\hat{y}_{i,j}\}_{i,j} \in \mathbb{R}^{H \times W}$ is the model's predicted probability map and $\mathbf{Y} = \{y_{i,j}\}_{i,j} \in \mathbb{R}^{H \times W}$ is the ground truth mask for that image, the Dice Loss is defined by the following equation:

$$\mathcal{L}_{\text{Dice}}(\hat{\mathbf{Y}}, \mathbf{Y}) = 1 - \frac{2 \sum_{i=1}^{W} \sum_{j=1}^{H} \hat{y}_{i,j} y_{i,j}}{\sum_{i=1}^{W} \sum_{j=1}^{H} (\hat{y}_{i,j}^2 + y_{i,j}^2)} \tag{2}$$

3.1.2 Geometrical Interpretation and Polynomial Expansion

By flattening the predicted probability map $\hat{\mathbf{Y}}$ and the ground truth mask $\mathbf{Y}$ into vectors $\hat{\mathbf{y}}$ and $\mathbf{y}$ of length HW, respectively, the Dice Loss can be decomposed as follows:

$$\mathcal{L}_{\text{Dice}} = 1 - s \cos\theta \tag{3}$$

Here, $s = \frac{2\langle \hat{\mathbf{y}}, \mathbf{y} \rangle}{\|\hat{\mathbf{y}}\|^2 + \|\mathbf{y}\|^2}$ is the scale component, and $\cos\theta = \frac{\langle \hat{\mathbf{y}}, \mathbf{y} \rangle}{\|\hat{\mathbf{y}}\| \|\mathbf{y}\|}$ is the cosine of the angle between the two vectors. This decomposition allows the Dice Loss to be understood as the product of the scale component s and $\cos\theta$.

Applying the Taylor expansion to the directional component $\cos\theta$ allows us to derive the polynomial representation of the PolyDice Loss. Specifically, assuming $\theta \approx 0$ (the prediction and ground truth are not largely different), $\cos\theta$ can be approximated by the Taylor expansion around $\theta = 0$ as follows:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots \tag{4}$$

Substituting this into the Dice Loss and rearranging yields the general form of PolyDice:

$$\mathcal{L}_{\text{PolyDice}} = (1 - s) + s \sum_{k=1}^{\infty} \alpha_k \theta^{2k} \tag{5}$$

Here, $\alpha_k = \frac{(-1)^{k-1}}{(2k)!}$ is the sign coefficient for each Taylor term.

3.2 PolyDice-1 Loss

From a practical perspective, following the approach of [4], we adopt the PolyDice-1 Loss, which allows only the first term to be adjustable:

$$\mathcal{L}_{\text{PolyDice-1}} = (1 - s) + s \left(\frac{1}{2} + \epsilon \right) \theta^2 \tag{6}$$

Here, $\epsilon \in \mathbb{R}$ is the hyperparameter that controls the shape of the loss function. Fig. 2 illustrates the change in the PolyDice-1 Loss shape depending on ϵ. When $\epsilon > 0$, the penalty for the prediction error is strengthened, and when $\epsilon < 0$, it is alleviated.

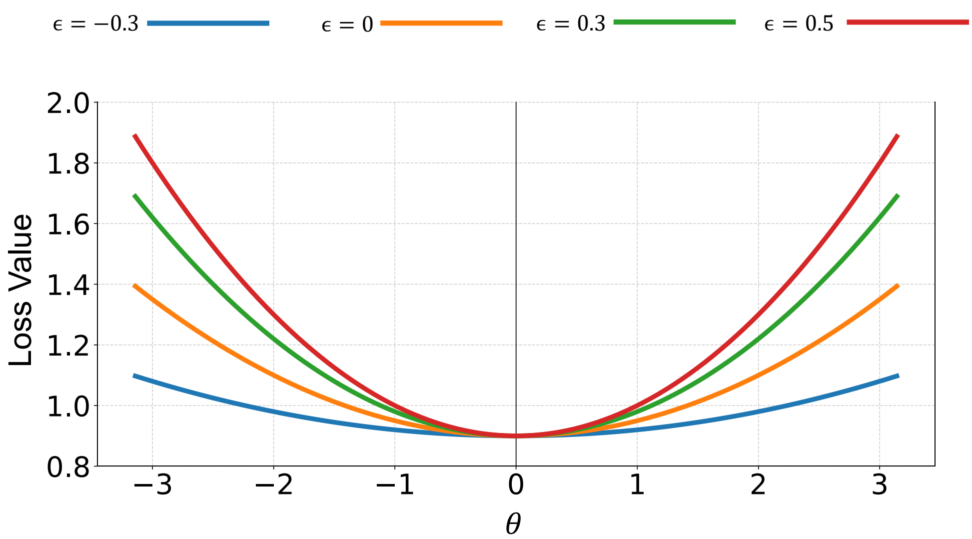

Fig. 2. Plot of PolyDice-1 Loss ($s = 0.1$)

3.3 Difficulty Assessment via Uncertainty

3.3.1 Principle and Application of MC Dropout

Dropout was originally proposed as a regularization technique to prevent overfitting in neural networks [5]. During training, it improves the model's generalization performance by randomly deactivating neurons in each layer with a probability p. Typically, Dropout is disabled during inference, and all neurons are active.

MC Dropout is a technique for estimating the model's epistemic uncertainty by enabling Dropout during inference as well as during training. While the output during conventional inference is deterministic, enabling Dropout creates different subnetworks, resulting in stochastic outputs. The distribution of prediction results obtained by running this stochastic inference multiple times on the same input can be viewed as an approximation of the predictive posterior distribution in a Bayesian Neural Network, and thus can be interpreted as a form of variational inference.In the proposed method, this MC Dropout is applied at each stage of the learning process. Specifically, an inference phase is inserted every τ epochs to quantify how "difficult" the model finds each segmentation at that point in time. Then, based on the uncertainty information obtained from this inference phase, the PolyDice-1 Loss parameter ϵ to be used for the next τ epochs of training is collectively determined and updated.

3.3.2 MC Dropout Inference During Training

The uncertainty assessment is performed at training epochs e where $e \in \{E_0, E_0 + \tau, E_0 + 2\tau, \ldots, E\}$ using the following procedure. Here, E_0 denotes the starting epoch for the adaptive control mechanism.For the segmentation model $f_\theta : \mathcal{X} \to [0,1]^{H \times W}$, the parameters at epoch e are fixed. For each image $x \in \mathcal{X}$ in the training data, N times of stochastic inference are performed with a Dropout rate

$p \in (0,1)$ to obtain the set of predictions

$$\hat{\mathbf{Y}}^{(n)} = f_{\theta^{(e)}}(x; \mathbf{z}^{(n)}), \quad \mathbf{z}^{(n)} \sim \text{Bernoulli}(1-p) \tag{7}$$

Here, $\mathbf{z}^{(n)}$ is the Dropout mask (which determines the activation/deactivation of each neuron) for the n-th inference, and each $\hat{\mathbf{Y}}^{(n)} = \{\hat{y}_{i,j}^{(n)} \in [0,1]\}_{i,j}$ is the predicted probability map from the n-th inference. This stochastic inference yields multiple prediction maps for the same image, and the variance among them allows for the quantification of how confused the model is.

3.3.3 Calculation of Pixel-wise Uncertainty Metrics

For the obtained N prediction images, the uncertainty metrics are calculated on a per-pixel basis. The following three uncertainty metrics are calculated pixel-wise from the uncertainty map.

- **Prediction Variance**:

$$\text{Var}(\hat{y}_{i,j}) = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_{i,j}^{(n)} - \bar{y}_{i,j})^2 \tag{8}$$

  Where $\bar{y}_{i,j} = \frac{1}{N}\sum_{k=1}^{N}\hat{y}_{i,j}^{(n)}$ is the mean of the predictions. This metric represents the variability of the multiple predictions; a higher value indicates that the model's prediction is unstable.

- **Predictive Entropy**:

$$H(\bar{y}_{i,j}) = -\bar{y}_{i,j}\log\bar{y}_{i,j} - (1-\bar{y}_{i,j})\log(1-\bar{y}_{i,j}) \tag{9}$$

  This represents the uncertainty of the mean prediction. Entropy is higher when the prediction probability is closer to 0.5, indicating that the model is confused about the foreground/background classification. The value is low when the probability is close to 0 or 1.

- **Mutual Information**:

$$I(y_{i,j}) = H(\bar{y}_{i,j}) - \frac{1}{N}\sum_{n=1}^{N}H(\hat{y}_{i,j}^{(n)}) \tag{10}$$

This represents the uncertainty of the model parameters (epistemic uncertainty). It is calculated by subtracting the aleatoric uncertainty (or expected entropy) from the total uncertainty (predictive entropy), and it has the property of decreasing with more training data. Regions with high mutual information suggest that the model may not have been sufficiently trained.

3.4 Aggregation of Image Difficulty and Adaptive Control

3.4.1 Aggregation to Per-Image Unit

Considering the uncertainty in medical images, the average of the uncertainty metric, restricted to the positive region (foreground/segmented area), is used as the overall image difficulty score.

3.4.2 Adaptive Parameter Control

The adaptive parameter control based on uncertainty is initiated only after the model has undergone sufficient initial training to ensure stable uncertainty estimation. Specifically, the PolyDice Loss parameter ϵ is dynamically controlled based on the image difficulty score D only when the current epoch $e \geq E_0$. To achieve adaptive learning, the pixel-wise uncertainty aggregated into the image difficulty score D, which is updated every τ epochs, is utilized to dynamically control the shape parameter ϵ of the PolyDice-1 Loss. This mechanism ensures that images deemed "difficult" by the model are assigned a steeper loss gradient, while easy images receive a gentler gradient, focusing the optimization on challenging samples.

The difficulty score D (based on the aggregated Mutual Information) is first normalized based on the overall statistics of the training batch to determine the relative difficulty of each sample. Let D_p be the p-th percentile and σ_D be the standard deviation of the difficulty scores within the current batch. The normalized score D_{norm} is computed as follows:

$$D_{\text{norm}} = \frac{D - D_p}{\sigma_D + \delta} \tag{11}$$

where δ is a small constant (e.g., 10^{-9}) for numerical stability.

Next, this normalized score is passed through a sigmoidal control function to map it onto the desired range of the loss parameter $\epsilon \in [\epsilon_{\text{min}}, \epsilon_{\text{max}}]$. This transformation dynamically adjusts ϵ to be proportional to the relative difficulty of the image. The adaptive ϵ is calculated as follows:

$$\epsilon = \epsilon_{\text{min}} + (\epsilon_{\text{max}} - \epsilon_{\text{min}}) \cdot \sigma(\lambda D_{\text{norm}}) \tag{12}$$

Here, $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function, and λ is a sensitivity hyperparameter that controls the steepness of the mapping, allowing fine-tuning of how aggressively the loss gradient changes in response to uncertainty. This framework effectively uses the model's self-assessed epistemic uncertainty to realize a curriculum-based, adaptive training strategy.

4 Experiments

4.1 Experimental Method

We conducted experiments to evaluate XX.

4.2 Results

The experimental results are presented using figures and tables. Table 1 shows an example table.

4.3 Discussion

Here we discuss the obtained results.

Table 1.: Performance evaluation

Model	Accuracy	Sensitivity	Specificity	AUC-ROC
Baseline #1	0.578	XXXX	XXXX	XXXX
Baseline #2	0.622	XXXX	XXXX	XXXX
Ours	0.751	XXXX	XXXX	XXXX

5 Conclusion

In this thesis, we presented XX based on the YY method. In particular, we demonstrated that ZZ is achievable through AA.

In our experiments, we conducted BB to evaluate CC. The results indicated DD, suggesting the potential for EE.

Future work should focus on improving FF.

Acknowledgement

The following acknowledgments are provided as an example only. While you may use this as a reference, you should express your gratitude in your own words to those who deserve recognition (avoid using the exact wording from this example).

This thesis is based on research conducted at the Intelligent Biosignal Informatics Laboratory, Informatics and Data Science Program, Graduate School of Advanced Science and Engineering, Hiroshima University.

First and foremost, I would like to express my deepest gratitude to my supervisor, Associate Professor Akira Furui, for his dedicated guidance despite his busy schedule. Associate Professor Furui provided meticulous supervision in both research discussions and thesis writing.

I am deeply grateful to Professor XX and Associate Professor XX, who served as secondary reviewers of this thesis, for their invaluable advice regarding XX.

I would like to express my sincere appreciation to Dr. XX of XX Corporation for providing numerous valuable suggestions and various support through our collaborative research meetings.

I am deeply grateful to Professor XX of XX University for the numerous support and advice provided through our collaborative research.

I would like to thank Mr./Ms. XX, a senior student in the XX Laboratory, for providing various support, including advice on XX. I am also grateful to my laboratory colleagues, Mr./Ms. XX, Mr./Ms. XX, and Mr./Ms. XX, as well as my junior colleagues, Mr./Ms. XX and Mr./Ms. XX, who supported my laboratory life in various ways, including seminars and daily discussions.

Finally, I cannot express enough gratitude to my parents, who have financially supported my student life and watched over me with warmth until now. Thank

you very much for everything.

References

[1] A. Furui, S. Eto, K. Nakagaki, K. Shimada, G. Nakamura, A. Masuda, T. Chin, and T. Tsuji, "A myoelectric prosthetic hand with muscle synergy-based motion determination and impedance model-based biomimetic control," *Science Robotics*, vol. 4, no. 31, pp. eaaw6339, 2019.

[2] A. Furui, T. Igaue, and T. Tsuji, "EMG pattern recognition via Bayesian inference with scale mixture-based stochastic generative models," *Expert Systems with Applications*, vol. 185, pp. 115644, 2021.

[3] H. Aizawa, "Polynomial dice loss for medical image segmentation(submitted)," .

[4] Z. Leng, M. Tan, C. Liu, E. D. Cubuk, J. Shi, S. Cheng, and D. Anguelov, "Polyloss: A polynomial expansion perspective of classification loss functions," in *International Conference on Learning Representations*, 2022.

[5] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

A Derivation of Something

The derivation of XX in the proposed method is presented.