

修士論文

不確実性適応型損失関数による 頑健な医用画像セグメンテーション

M243422 廣池 友哉

指導教員 古居 彬 准教授

令和 8 年 2 月 3 日

広島大学 大学院先進理工系科学研究科
情報科学プログラム

目次

1	はじめに	2
2	関連研究	3
2.1	PolyDice Loss	3
2.2	MC Dropout の原理と応用	6
3	提案法	7
3.1	学習プロセスと MC Dropout 推論	10
3.2	画像全体の難易度指標の算出	11
3.3	適応的制御	11
4	実験	13
4.1	実験設定	13
5	結論	17
	謝辞	18
	参考文献	19
	付録 A 何とかの導出	22

1 はじめに

医用画像セグメンテーションは、診断支援や治療計画において不可欠な技術であり、正常組織または異常組織の領域を抽出することが求められる。特に大腸ポリープ [1] や頭頸部がん放射線治療における危険臓器 [2] の検出など、臨床応用が進んでいる。

しかしながら、医用画像セグメンテーションには固有の課題が存在する。特に重要な問題として、クラス不均衡が挙げられる。医用画像では背景領域が大部分を占め、対象となる病変は相対的に小さい領域しか占めないことが多い。この状況下では、従来の分類タスクで広く使われている Cross-Entropy Loss [3] は背景領域の学習に偏り、臨床的に重要な小病変や曖昧な境界部分の正確なセグメンテーションが困難となる。

この課題に対処するため、クラス不均衡に対して頑健な Dice Loss [4] やその多くの拡張手法が提案され、CT 画像 [5, 6] や MRI 画像 [7] において高い性能が報告されている。しかし、これらの損失関数は全画像に対して固定的な形状を持つという制約がある。医用画像の多様性を考慮すると、画像ごとの難易度に応じて損失関数を適応的に調整することが望ましい。このような適応的学習を実現するためには、損失関数の形状を制御する機構と、各画像の難易度を定量化する仕組みの 2 つが必要である。

本研究では、これら 2 つの要素を組み合わせた適応的学習フレームワークの構築を目指す。形状制御には、Dice Loss を多項式展開して得られる PolyDice Loss [8] を用いる。難易度の定量化には、Monte Carlo Dropout [9]（以下、MC Dropout）による不確実性推定を用いる。MC Dropout は推論時に Dropout を有効にすることで、モデルの認識的不確実性を効率的に推定し、モデルが各画像をどの程度「難しい」と感じているかを定量化できる。その後、得られた不確実性指標に基づき、PolyDice Loss の形状パラメータを動的に制御することで、難しい画像には急峻な勾配を、簡単な画像には緩やかな勾配を与える適応的学習が可能である。

2 関連研究

2.1 PolyDice Loss

医用画像セグメンテーションで広く使用される Dice Loss は、クラス不均衡に頑健であるが、全画像に対して固定的な形状を持つという制約がある。本研究では、Dice Loss を多項式展開により拡張した PolyDice Loss [8]、特にその実用的な形式である PolyDice-1 Loss を採用する。PolyDice-1 Loss は、単一のパラメータ ϵ で損失関数の形状を制御でき、画像の難易度に応じて勾配の急峻さを調整することが可能となる。

2.1.1 Dice Loss の定義

画像サイズを $H \times W$ とし、ピクセル位置を (i, j) で表す ($i \in \{1, \dots, H\}, j \in \{1, \dots, W\}$)。セグメンテーションタスクにおいて、モデルの予測確率マップを $\hat{\mathbf{Y}} = \{\hat{y}_{i,j}\}_{i,j} \in \mathbb{R}^{H \times W}$ 、その画像に対する正解マスクを $\mathbf{Y} = \{y_{i,j}\}_{i,j} \in \mathbb{R}^{H \times W}$ とすると、Dice Loss は次式で定義される。

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

2.1.2 幾何学的解釈と多項式展開

予測確率マップ $\hat{\mathbf{Y}}$ と正解マスク \mathbf{Y} をそれぞれ長さ HW のベクトル $\hat{\mathbf{y}}, \mathbf{y}$ として平坦化すると、Dice Loss は以下のように分解できる。

$$\mathcal{L}_{\text{Dice}} = 1 - s \cos \theta \quad (2)$$

ここで、 $s = \frac{2\langle \hat{\mathbf{y}}, \mathbf{y} \rangle}{\|\hat{\mathbf{y}}\|^2 + \|\mathbf{y}\|^2}$ はスケール成分、 $\theta = \frac{\langle \hat{\mathbf{y}}, \mathbf{y} \rangle}{\|\hat{\mathbf{y}}\| \|\mathbf{y}\|}$ は2つのベクトル間の角度を表す。この分解により、Dice Loss はスケール成分 s と $\cos \theta$ の積として理解できる。

方向成分 $\cos \theta$ に対して Taylor 展開を適用することで、PolyDice Loss の多項式表現を導出する。つまり $\theta \approx 0$ (予測と正解が大きく異ならない) と仮定し、 $\cos \theta$ を

$\theta = 0$ まわりでテイラー展開すると以下のように近似できる.

$$\cos \theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \dots \quad (3)$$

これを Dice Loss に代入し, 整理すると PolyDice の一般形が得られる:

$$\mathcal{L}_{\text{PolyDice}} = (1 - s) + s \sum_{k=1}^{\infty} \alpha_k \theta^{2k} \quad (4)$$

ここで, $\alpha_k = \frac{(-1)^{k-1}}{(2k)!}$ は各 Taylor 項の符号係数である.

2.1.3 PolyDice-1 Loss

実用的な観点から, [10] のアプローチに従い, 第 1 項のみを調整可能とする PolyDice-1 Loss を採用する:

$$\mathcal{L}_{\text{PolyDice-1}} = (1 - s) + s \left(\frac{1}{2} + \epsilon \right) \theta^2 \quad (5)$$

ここで, $\epsilon \in \mathbb{R}$ は損失関数の形状を制御するハイパーパラメータである. 図 1 に, ϵ に応じた PolyDice-1 Loss の形状変化を示す. $\epsilon > 0$ では予測誤差に対するペナルティが強化され, $\epsilon < 0$ では緩和される.

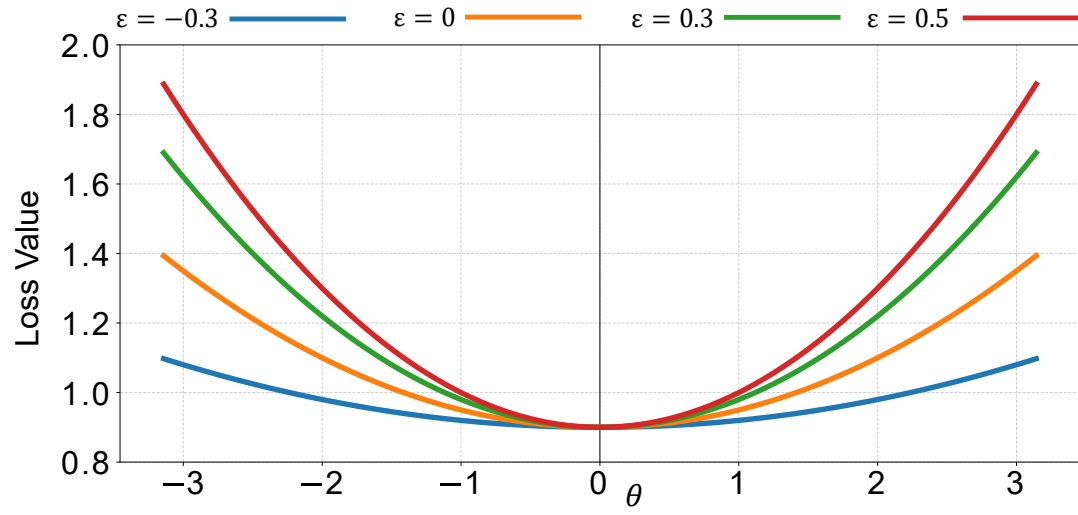


Fig. 1. Plot of PolyDice-1 Loss($s = 0.1$)

2.2 MC Dropout の原理と応用

Dropout は元々、ニューラルネットワークの過学習を防ぐ正則化手法として提案された [11]。訓練時に各層のニューロンを確率 p でランダムに不活性化することでモデルの汎化性能を向上させる。通常、推論時には Dropout は無効化され、全ニューロンが活性化される。

MC Dropout は、学習時のみだけでなく、推論時にも Dropout を有効にすることで、モデルの認識的不確実性を推定する手法である。通常、推論時の出力は決定論的であるが、Dropout を有効にすることで、異なるサブネットワークが形成され、確率的な出力が得られる。同一入力に対してこの確率的な推論を複数回実行して得られる予測結果の分布は、Bayesian Neural Network における予測事後分布近似とみなせ、変分推論の一種として解釈できる。

提案法では、この MC Dropout を学習過程の各段階で適用する。具体的には、 τ エポックごとに推論フェーズを挿入し、その時点でのモデルが各セグメンテーションをどの程度「難しい」と感じているかを定量化する。

3 提案法

図 2 に提案手法の概要を，アルゴリズム 1 に詳細なアルゴリズムを示す．提案法では，学習中各画像に対して τ エポックごとに MC Dropout を用いて画像毎に複数枚推論を行い，予測のばらつきから不確実性を定量化する．この不確実性情報は，モデルがその画像のセグメンテーションをどの程度「難しい」と感じているかを反映する．その後，この不確実性情報を画像単位に集約し，PolyDice Loss の形状を動的に制御することで，難しい画像には急峻な勾配を，簡単な画像には緩やかな勾配を与える．更新された ϵ を次の τ エポック間の学習に適用することで，学習の進行度に応じて最適化の重み付けを動的に変化させる適応的学習を実現する．

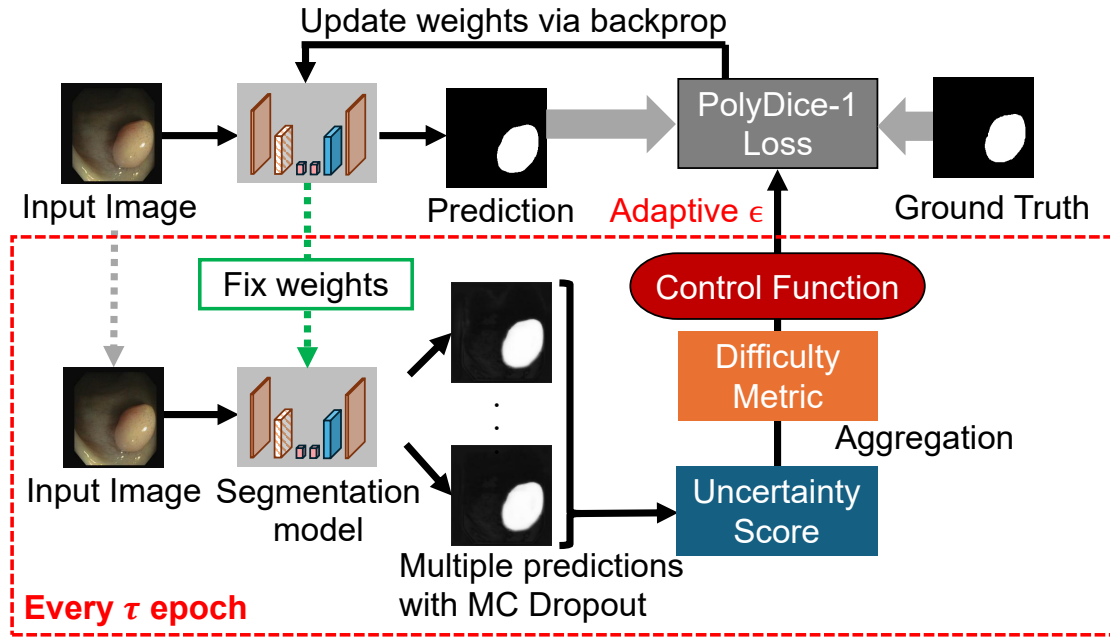


Fig. 2. Overview of the proposed adaptive learning framework

Algorithm 1 Uncertainty-based Adaptive PolyDice-1 Loss Learning

Require: Training dataset \mathcal{D} , Max epochs E , Model f_θ

Require: Hyperparameters: Adaptive start epoch E_0 , Interval τ , MC iterations N , Slope k , Range $[\epsilon_{\min}, \epsilon_{\max}]$

```
1: Initialize model parameters  $\theta$ 
2: Initialize loss parameter  $\epsilon_i \leftarrow 0$  for all images  $x_i \in \mathcal{D}$ 
3: for  $e = 1$  to  $E$  do                                 $\triangleright$  Step 1: Adaptive Control Phase
4:   if  $e \geq E_0$  and  $(e - E_0) \pmod{\tau} = 0$  then
5:      $List_D \leftarrow \emptyset$ 
6:     for each image  $x_i \in \mathcal{D}$  do
7:       Perform  $N$  stochastic inferences:  $\{\hat{y}^{(n)}\}_{n=1}^N \leftarrow \text{MC\_Dropout}(x_i)$ 
8:       Update  $\epsilon_i \leftarrow \dots$ 
9:     end for
10:  else
11:     $\triangleright$  Keep  $\epsilon_i$  fixed (Initial Training / Interval)
12:  end if
13:     $\triangleright$  Step 2: Training Phase
14:  for each batch  $(\mathbf{x}, \mathbf{y}) \in \mathcal{D}$  do
15:    Update parameters:  $\theta \leftarrow \theta - \eta \nabla_{\theta} \mathcal{L}$ 
16:  end for
17: end for
18: return Trained parameters  $\theta$ 
```

3.1 学習プロセスと MC Dropout 推論

提案法では、学習プロセスを初期学習期間と適応的学習期間の2段階に分割する。エポック 1 から $E_0 - 1$ までの期間は初期学習期間とし、損失形状パラメータ ϵ を 0（または所定の固定値）に固定して学習を行う。これにより、不確実性推定を行う前に、モデルにセグメンテーションタスクに必要な基礎的な表現力を獲得させる。

適応的学習期間 ($e \geq E_0$) においては、周期 τ ごとに不確実性の再評価と ϵ の更新を行う。すなわち、更新は $e \in \{E_0, E_0 + \tau, E_0 + 2\tau, \dots\}$ を満たすエポックの学習開始前に実行される。更新の際は、その時点のモデルパラメータ θ を固定し、訓練データの各画像 $x \in \mathcal{X}$ に対して Dropout 率 $p \in (0, 1)$ で N 回の確率的推論を行う。得られる予測集合を $\{\hat{\mathbf{Y}}^{(n)}\}_{n=1}^N$ とする：

$$\hat{\mathbf{Y}}^{(n)} = f_{\theta}(x; \mathbf{z}^{(n)}), \quad \mathbf{z}^{(n)} \sim \text{Bernoulli}(1 - p) \quad (6)$$

ここで、 $\mathbf{z}^{(n)}$ は n 回目の推論における Dropout マスクであり、 $\hat{\mathbf{Y}}^{(n)}$ はその予測確率マップである。この確率的推論により得られる予測のばらつきから、モデルの認識的不確実性を定量化する。

3.1.1 ピクセル単位の不確実性指標の計算

得られた N 枚の予測画像に対し、ピクセル単位の不確実性指標として相互情報量 $I_{i,j}$ を計算する。

$$I_{i,j} = H(\bar{y}_{i,j}) - \frac{1}{N} \sum_{n=1}^N H(\hat{y}_{i,j}^{(n)}) \quad (7)$$

ここで $H(\cdot)$ はエントロピー関数、 $\bar{y}_{i,j}$ は予測の平均を表す。この相互情報量はモデルパラメータの不確実性を反映しており、値が高い領域はモデルが十分に学習できていないことを示唆する。

3.2 画像全体の難易度指標の算出

3.2.1 画像単位への集約

医用画像のクラス不均衡性を考慮し，陽性領域に限定した相互情報量の平均を画像全体の難易度スコアとして用いる．画像領域全体を Ω ，正解マスクにおける陽性領域の画素集合を $\mathcal{P} = \{(i, j) \in \Omega \mid y_{i,j} = 1\}$ とすると，画像全体の難易度スコア D は次式で算出される．

$$D = \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} I_{i,j} \quad (8)$$

ここで， $|\mathcal{P}|$ は陽性領域の総画素数を表す．なお，陽性領域が存在しない画像については， $D = 0$ （または画像全体の平均）とする．

次に，各サンプルの相対的な難易度を決定するため，データセット全体の難易度スコア分布に基づいて正規化を行う．算出された全画像のスコア集合における p パーセンタイル値を D_p ，標準偏差を σ_D とすると，正規化されたスコア D_{norm} は次のように計算される．

$$D_{\text{norm}} = \frac{D - D_p}{\sigma_D + \delta} \quad (9)$$

ここで， δ は数値的安定性のための微小定数である．

3.3 適応的制御

得られた難易度指標 D_{norm} に基づき，PolyDice-1 Loss の形状パラメータ ϵ を動的に更新する．更新式には以下のシグモイドベースの制御関数を用いる．

$$f(x) = \frac{1}{1 + \exp(-kx)} \quad (10)$$

$$\epsilon = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min})f(D_{\text{norm}}) \quad (11)$$

ここで、 k は関数の感度を制御するハイパーパラメータであり、 $\epsilon_{\min}, \epsilon_{\max}$ は ϵ の変動範囲である。この制御により、モデルが「難しい」と感じる画像には大きな ϵ が割り当てられ、損失関数の勾配が急峻になる。更新された ϵ は次の τ エポック間の学習に適用され、これによりモデルは困難な画像の学習を重点的に行うことが可能となる。

4 実験

提案法の有効性を検証するために、医用画像データセットを用いた実験を行った。

4.1 実験設定

4.1.1 データセットおよび実装の詳細

CVC-ClinicDB データセット [12] データセットおよび Kvasir-SEG データセット [13] を用いて実験を行った。CVC-ClinicDB データセットは 612 枚の大腸内視鏡画像 (384×288 pixel), Kvasir-SEG データセットは 1000 枚の大腸内視鏡画像で、いずれも大腸内視鏡画像とそれに対応するポリープの正解マスクから構成される。

セグメンテーションモデルには表 1 に示される構造の U-Net [14] を採用した。学習には Adam optimizer [15] を使用し、バッチサイズ 32, 学習率 10^{-3} に設定した。損失関数には PolyDice-1 Loss ($\epsilon = 0$) を用いた。前処理として全画像を $W = 224$ pixel, $H = 224$ pixel にリサイズし、訓練時には 50% の確率で上下左右反転および明度・コントラストの変更を適用した。最大エポック数 E は 200 に設定した。

MC Dropout による不確実評価は、 $\tau = 10$ エポックごとに実施した。Dropout 層はエンコーダの最終ブロックとデコーダの最終ブロックに配置し、各評価時には、 $p = 0.5$ で $N = 10$ 回の確率的推論を行った。またバッチ内の難易度指標の正規化時のパーセンタイルは $p = 25$ に設定し、適応的学習の開始エポックは $E_{\text{start}} = 0$ とし、 $\epsilon_{\text{min}} = 0$, $\epsilon_{\text{max}} = 0.5$, $k = 2$ とした。

4.1.2 比較条件および比較指標

適応的学習の有効性を評価するため、複数の条件下で比較実験を行った。まず、医用画像セグメンテーションにおける標準的なベースラインとして Dice Loss を検討し、併せてクラス不均衡問題に対処する代表的手法である Focal Loss [16] と比較を行った。また易しいサンプルの重みを下げる割合を調整するパラメータは $\gamma = 2$ と設定し

た．次に，PolyDice-1 Loss におけるパラメータ ϵ を動的に制御することの有効性を検証するため， ϵ を固定値とした場合との比較を行った．これには，Dice Loss を近似した $\epsilon = 0$ の設定に加え，予備実験により探索されたデータセットごとの最適な固定値を用いた PolyDice-1 Loss が含まれる．これらの比較手法と，不確実性指標に基づいて ϵ を適応的に更新する提案法のセグメンテーション精度を比較することで，提案手法の有効性を検証する．

またセグメンテーションの性能の評価には，Dice 係数，IoU，Precision，Recall を用いた．

Table 1. Overview of the U-Net Architecture

Layer	Output Size
<i>— Encoder —</i>	
Input	$224 \times 224 \times 3$
inc (DoubleConv)	$224 \times 224 \times 64$
down1 (MaxPool + DoubleConv)	$112 \times 112 \times 128$
down2 (MaxPool + DoubleConv)	$56 \times 56 \times 256$
down3 (MaxPool + DoubleConv)	$28 \times 28 \times 512$
down4 (MaxPool + DoubleConv)	$14 \times 14 \times 512$
<i>— Decoder —</i>	
up1 (Upsample + DoubleConv)	$28 \times 28 \times 256$
up2 (Upsample + DoubleConv)	$56 \times 56 \times 128$
up3 (Upsample + DoubleConv)	$112 \times 112 \times 64$
up4 (Upsample + DoubleConv)	$224 \times 224 \times 64$
outc (Conv2d)	$224 \times 224 \times 2$

Table 2. Performance Comparison with Existing Loss Functions on Kvasir-SEG Dataset

Method	Dice	IoU	Precision	Recall
Dice Loss	0.7895	0.7021	0.8281	0.8154
Focal Loss ($\gamma = 2$)	0.7192	0.6082	0.8634	0.6769
PolyDice-1 ($\epsilon = 0$)	0.8095	0.7198	0.8461	0.8268
PolyDice-1 (Opt. ϵ)	0.8095	0.7198	0.8461	0.8268
Adaptive PolyDice-1	0.8272	0.7440	0.8707	0.8397

5 結論

本論では，何とか法に基づく何とかかんとかについて述べた．特に，何とかかんとかすることで，何とかが可能である．

実験では，何とかかんとかを目的に何とかを行った．その結果，何とかであったことから，何とかである可能性が示された．

今後は，何とかかんとかを改善していく必要がある．

謝辞

下記の謝辞はあくまでも記載例です。下記を参考にするのは結構ですが、必ず自分の言葉で感謝の気持ちを伝えるべき人間に伝えましょう（記載例の通りの文言で謝辞を述べるのは避けましょう）。

本論文は、広島大学大学院先進理工系科学研究科情報科学プログラム知能生体情報学研究室にて行われた研究に基づくものです。

まず、本研究を進めるにあたり、ご多忙のところ熱心にご指導くださいました指導教員の古居彬准教授に心より感謝申し上げます。古居彬准教授には、論文執筆や研究に関する議論など、細部にわたるまで丁寧にご指導いただきました。

本論の副査である〇〇〇〇教授と〇〇〇〇准教授には、〇〇〇〇に関して大変有益な助言を頂き、大変お世話になりました。ここに厚く御礼申し上げます。

〇〇株式会社の〇〇〇〇博士には、共同研究ミーティングを通じ、数多くの有益なご指摘のほか、様々なご支援をいただきました。心より感謝申し上げます。

〇〇大学の〇〇〇〇〇教授には、共同研究を通じ多数のご支援とご助言をいただきました。心より感謝申し上げます。

情報数理研究室の先輩である専攻科〇年の〇〇〇〇氏には、〇〇に関するアドバイス等をはじめ、数々のご支援をいただきました。また、研究室同期の〇〇〇〇氏・〇〇〇〇氏・〇〇〇〇氏、後輩の〇〇〇〇氏・〇〇〇〇氏には、ゼミや日頃のディスカッションのほか、研究室生活を様々な面で支えていただきました。ここに感謝申し上げます。

最後になりますが、現在に至るまで私の学生生活を金銭的に支え、暖かく見守って下さった両親には感謝の念に堪えません。これまで本当にありがとうございました。

参考文献

- [1] G.-P. Ji, G. Xiao, Y.-C. Chou, D.-P. Fan, K. Zhao, G. Chen, and L. Van Gool, “Video polyp segmentation: A deep learning perspective,” *Machine Intelligence Research*, vol. 19, no. 6, pp. 531–549, 2022.
- [2] F. Maleki, W. T. Le, T. Sananmuang, S. Kadoury, and R. Forghani, “Machine learning applications for head and neck imaging,” *Neuroimaging Clinics*, vol. 30, no. 4, pp. 517–529, 2020.
- [3] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.
- [4] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571. Ieee, 2016.
- [5] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie, “AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy,” *Medical Physics*, vol. 46, no. 2, pp. 576–589, 2019.
- [6] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, “A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from ct images,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2653–2663, 2020.
- [7] S. Kato and K. Hotta, “Adaptive t-vMF dice loss: An effective expansion of dice loss for medical image segmentation,” *Computers in Biology and Medicine*, vol. 168, pp. 107695, 2024.
- [8] H. Aizawa, “Polynomial dice loss for medical image segmenta-

tion(submitted),” .

- [9] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” in *Proceedings of The 33rd International Conference on Machine Learning*, vol. 48, pp. 1050–1059, 20–22 Jun 2016.
- [10] 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.
- [11] 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.
- [12] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, “WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians,” *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.
- [13] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. d. Lange, D. Johansen, and H. D. Johansen, “Kvasir-seg: A segmented polyp dataset,” in *Proceedings of the 26th International Conference on Multimedia Modeling (MMM 2020), Part II*, vol. 11962 of *Lecture Notes in Computer Science*, pp. 451–462, Springer. 2020.
- [14] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.
- [15] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the 3rd International Conference on Learning Representations*

(*ICLR*), 2015.

- [16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

付録 A 何とかの導出

提案法における何とかを導出する.