

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

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 M243422 廣池 友哉

1 はじめに

医用画像セグメンテーションは、医用画像中の全てのピクセルに対し物体のクラスのラベルを推定することで、正常組織または異常組織の領域を抽出するタスクである。特に近年では、ポリープ [1] や頭頸部がん (HNC) 放射線治療における危険臓器 [2] などに用いられている。

医用画像セグメンテーションは、多様な症例に対して頑健に機能することが求められるが、対象領域のサイズや形状が画像ごとに大きく異なるため、頑健な検出が困難であるという課題がある。この課題に対処するために、損失関数を工夫する手法が提案されてきた。例えば分類タスクで広く使われている Cross-Entropy Loss [3] と比較して、Dice Index と呼ばれる類似度指標に基づいて定義されている Dice Loss [4] は、少数クラスの誤差も全体の平均損失に大きく寄与するため、クラス不均衡下でも検出性能を向上させる効果がある。また、Dice Loss の多くの拡張手法も提案されており、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、これらのアプローチでは検出難易度に関わらず誤差関数の形状が固定されており、検出難易度が高い画像に対する学習が不十分になり、頑健な検出が困難であるという課題がある。

そこで本研究では、検出難易度を画像毎に算出したものを学習に取り入れ、誤差関数の形状を動的に変化させる手法を提案する。具体的には、画像毎に形状を変更できる誤差関数である PolyDice-1 Loss を用いる。学習中には一定 epoch 毎に推論フェーズを挿入し、Monte Carlo Dropout [8] (以下、MC Dropout) により各画像毎に複数枚の予測画像を得た後、ピクセル単位で不確実性指標を計算し、そこから画像全体の不確実性指標を集約したものを難易度指標として算出する。この難易度指標を PolyDice-1 Loss の学習に取り入れることで、検出難易度を考慮した頑健な検出を実現することができる。

2 提案法

2.1 概要

本研究の最終目標は、画像ごとのセグメンテーション難易度を学習中に動的に評価し、その難易度に応じて損失関数の形状を適応的に変化させるシステムの構築である。図に提案法の全体構成を示す。提案法では、学習中各画像に対して 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。将来的には、この不確実性情報を画像単位に集約し、PolyDice-1 Loss のハ

イパーパラメータ ϵ 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を実現する。

本稿では、最終目標に向けた基礎的検証として、ピクセル単位の不確実性指標が学習過程でどのように推移するかを分析し、画像難易度の指標としての妥当性を検証する。

2.2 PolyDice Loss による損失形状制御

2.2.1 Dice Loss の定式化

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 $\hat{\mathbf{y}} \in \mathbb{R}^{H \times W}$ 、その画像に対する正解マスクを $\mathbf{y} \in \mathbb{R}^{H \times W}$ とすると、Dice Loss は次式で定義される。

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

2.2.2 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る) 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。

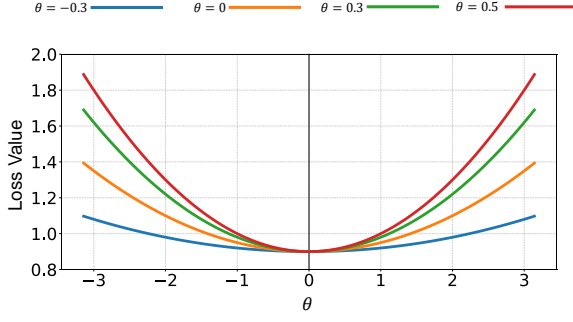


図 1: Plot of PolyDice-1 Loss ($s = 0.1$)

2.2.3 PolyDice-1 Loss

実用的な観点から、[9] 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唯一のハイパーパラメータである。図 1 に $s = 0.1$ として PolyDice-1 Loss の ϵ を変化させた際の値の推移を示す。 $\epsilon > 0$ 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。本稿では、まず $\epsilon = 0$ で固定して実験を行い、将来的にはこの ϵ を不確実性に基づいて動的に調整する。

2.3 MC Dropout による不確実性推定

2.3.1 MC Dropout の理論的背景

Dropout は、ニューラルネットワークの学習時に、一部のユニットを確率 p でランダムに不活性化することで過学習を抑制し、モデルの汎化性能を向上させる手法である。

MC Dropout は、学習時のみだけでなく、推論時にも Dropout を用いることで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、学習段階で M (epoch) 毎に推論フェーズを挿入して Dropout を確率 p で MC Dropout を N 回実行することで、各学習段階で N 枚の複数の予測画像を得る。また Dropout 層はモデルの最も深いエンコーダおよびデコーダに配置する。浅い層と比較してより高次の特徴を抽出する深い層で Dropout を実行することで、高次の意味的判断におけるモデルの確信度がより直接的に反映されることが期待される。

2.3.2 MC Dropout の定式化

学習済みモデル $f_\theta : \mathcal{X} \rightarrow [0, 1]^{H \times W}$ に対し、Dropout 率 $p \in (0, 1)$ で MC Dropout を適用する。同一入力画像 $x \in \mathcal{X}$ に対して N 回の確率的推論を行い、予測集合 $\{\hat{Y}^{(n)}\}_{n=1}^N$ を

得る：

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

ここで、 $\mathbf{z}^{(n)}$ は n 回目の推論における Dropout マスク（各ニューロンの活性/非活性を決定）である。各 $\hat{Y}^{(n)} = \{\hat{y}_{i,j}^{(n)}\} \in [0, 1]_{i,j}$ は n 回目の推論における予測確率マップである。

2.3.3 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、ピクセル単位で不確実性指標を計算する。不確実性マップから、以下の 3 つの不確実指標をピクセル単位で計算する。

● 予測分散：

$$\text{Var}(\hat{y}_{ij}) = \frac{1}{N} \sum_{n=1}^N (\hat{y}_{ij}^{(n)} - \bar{y}_{i,j})^2 \quad (7)$$

ここで、 $\bar{y}_{i,j} = \frac{1}{N} \sum_{k=1}^N \hat{y}_{ij}^{(k)}$ は予測の平均である。この指標は、複数回の予測がどの程度ばらついているかを表し、値が大きいほどモデルの予測が不安定であることを示す。

● 予測エントロピー：

$$H(\bar{y}_{i,j}) = -\mu_{ij} \log \bar{y}_{i,j} - (1 - \bar{y}_{i,j}) \log (1 - \bar{y}_{i,j}) \quad (8)$$

これは平均予測の不確実性を表す。エントロピーは予測確率が 0.5 に近いほど高くなり、モデルが前景・背景の判断に迷っていることを示す。確率が 0 または 1 に近い場合は低い値となる。

● 相互情報量：

$$I(\hat{y}_{ij}) = H(\bar{y}_{i,j}) - \frac{1}{N} \sum_{n=1}^N H(\hat{y}_{ij}^{(n)}) \quad (9)$$

これはモデルパラメータの不確実性（認識的不確実性）を表す。全体の不確実性から偶然的な不確実性を引いたもので、より多くのデータで学習することで減少する性質を持つ。この値が高い領域は、モデルが十分に学習できていない可能性を示唆する。

2.4 画像全体の難易度への集約と適応的制御（将来構想）

2.4.1 画像単位への集約

ピクセル単位の不確実性マップ $U = \{u_{i,j}\}_{i,j}$ から画像全体の難易度スコア $D \in \mathbb{R}^+$ を算出する必要がある。セグメンテーションタスクの特性を考慮すると、境界領域の不確実性に着目する方法や、予測誤差と相関の高い領域を重視する方法などが考えられる。また、不確実性の空間的分布パターンや、前景・背景それぞれの不確実性の違いも重要な情報となる可能性がある。

2.4.2 適応的パラメータ制御

画像難易度スコア D に基づいて、Poly Dice Loss のパラメータ ϵ を動的に制御する：

$$\epsilon = g(D; \theta_g) \quad (10)$$

ここで、 g は制御関数、 θ_g はその学習可能なパラメータである。難易度が高い画像には大きな ϵ を割り当てて急峻な勾配を与え、簡単な画像には小さな ϵ で緩やかな勾配を与えることで、効率的な学習を実現する。

3 実験

3.1 実験設定

3.1.1 データセット

画像難易度の指標としての妥当性を検証するために、医用画像セグメンテーションデータを用いて実験を行った。データセットは、 384×288 pixel の 612 枚の大腸内視鏡画像とポリープマスクが含まれる CVC-ClinicDB [10] を用い、訓練データとテストデータの割合は 8:2 とし、訓練データのうち 10% をランダムに抽出し、検証データとした。

3.1.2 実装の詳細

セグメンテーションモデルは 3 層の U-Net [11] を用い、学習には、バッチサイズ 32、学習率 10^{-3} の Adaptive moment estimation (以下, Adam) [12] を用いた。損失関数は PolyDice-1 Loss において $\epsilon = 0$ としたものをを用いた。各画像は $W = 224$ pixel, $H = 224$ pixel でリサイズを行い、過学習抑制のため、訓練データに対する空間的データ拡張として 50% の確率で上下左右反転及び画像の明るさ・コントラストの変更を施した。最大エポック数は 200 に設定し、検証データに対する損失が最小のエポックにおけるモデルインスタンスを、最終的なテストデータの評価に用いた。また、提案法における MC Dropout のパラメータとして、Dropout を行う確率 $p = 0.5$ 、推論回数 $N = 10$ 、推論フェーズの挿入間隔 $M = 10$ とした。

3.2 結果と考察

3.2.1 学習過程における不確実性の推移

学習初期に高い精度で予測できるようになった症例（以下症例 1）と、学習後期においても精度が低い症例（以下症例 2）について、50, 100, 150, 200 epoch のピクセル単位の不確実性指標の推移を示す。(a) は予測平均, (b) は予測分散, (c) は予測エントロピー, (d) は相互情報量である。なお各不確実性指標において、カラースケールは両症例間で統一し、全ピクセル・全エポックにおける最大値で正規化した。

図 2 に症例 1 の不確実性指標の推移を示す。図 2 (b), (d) から、学習の進行に伴いポリープ内外で減少していることがわかる。一方、両指標ともにポリープの境界部分では、学習後期においても微小な不確実性が残存している。

図 3 に症例 2 の不確実性指標の推移を示す。図 3 (b), (d) から、症例 1 と異なり、学習の進行に伴い、ポリープの輪郭および内部において不確実性が増大していることがわかる。

今回の検証から、各不確実性指標が捉える「難しさ」の側面が明らかになった。学習に伴い、相互情報量は症例 1 では減少し、症例 2 では増大したことから、これらはモデルの知識不足に起因するエピステミック不確実性を主に捉える指標と言える。一方、予測エントロピーは、学習が十分に進んで病変の特徴を捉えた後でも、画像の境界領域で高い値を示した。これは、「データ固有の曖昧さ」に起因するアレアトリック不確実性を捉える指標と言える。

検証の結果、学習の進行に伴い、高精度な症例では相互

情報量が収束したのに対し、精度が低い症例では学習後期でも高い相互情報量が残存することが確認された。本研究の目的は、損失関数を動的に制御して精度を向上させることにあるため、改善の余地がある「モデルの知識不足」を特定することが重要となる。したがって、モデルが「迷っている」状態を最も直接的に示す相互情報量が、本研究における難易度指標として妥当であると考えられる。検証の結果から、集約された難易度指標の値が高く難しい画像には PolyDice-1 Loss の係数 ϵ を増加させ、急峻な勾配を課す適応的な学習を行うことで、特徴を重点的に学習させることができると考えられる。本稿の分析で、困難な症例では不確実性がポリープの内部に広範囲に分布する傾向が確認された。この情報をより効果的に活用するため、病変の正解マスクに着目して相互情報量を集約することで、難しい症例に対して頑健な検出ができる指標の設計を目指す。

4 まとめと今後の課題

本稿では、医用画像セグメンテーションの誤差関数において、検出難易度が高い画像に対する学習が不十分になり、頑健な検出が困難であるという課題に対処するため、検出難易度を画像毎に算出したものを学習に取り入れ、誤差関数の形状を動的に変化させる手法を提案した。提案法では、学習中各画像に対して 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した。医用画像データセットを用いた実験の結果、学習後期においても精度が低い症例に関して、ピクセル単位の相互情報量の分散と相互情報量が増大していくことが明らかになった。今後は、上記の不確実性指標を領域周辺に着目して集約する方法を検討し、難易度が高い画像には急峻な勾配を与えることで、精度が低い症例でも頑健な検出を実現できる制御関数を実装し、提案法の有効性を検証する予定である。

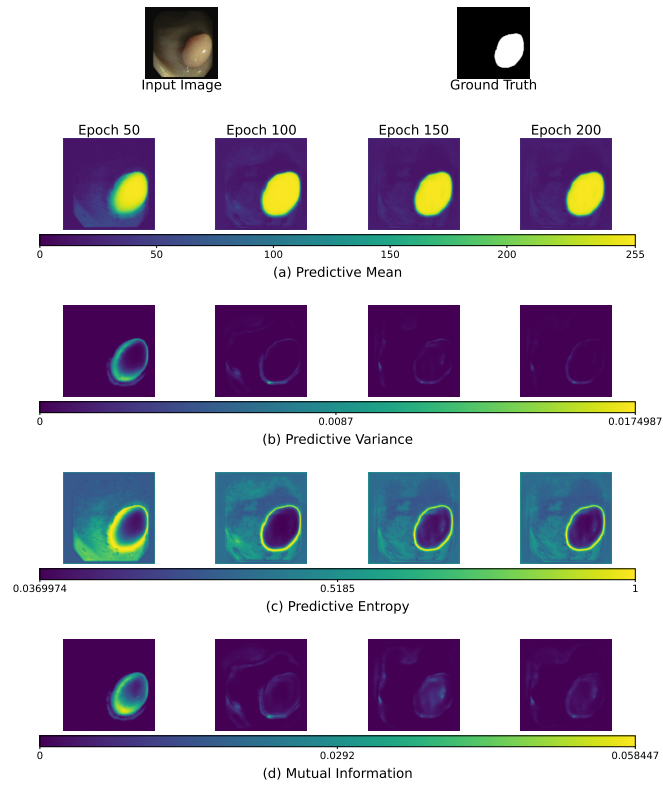


Figure 2: Example of Cases that are learned quickly with high accuracy: (a) Predictive Mean, (b) Predictive Variance, (c) Predictive Entropy, (d) Mutual Information

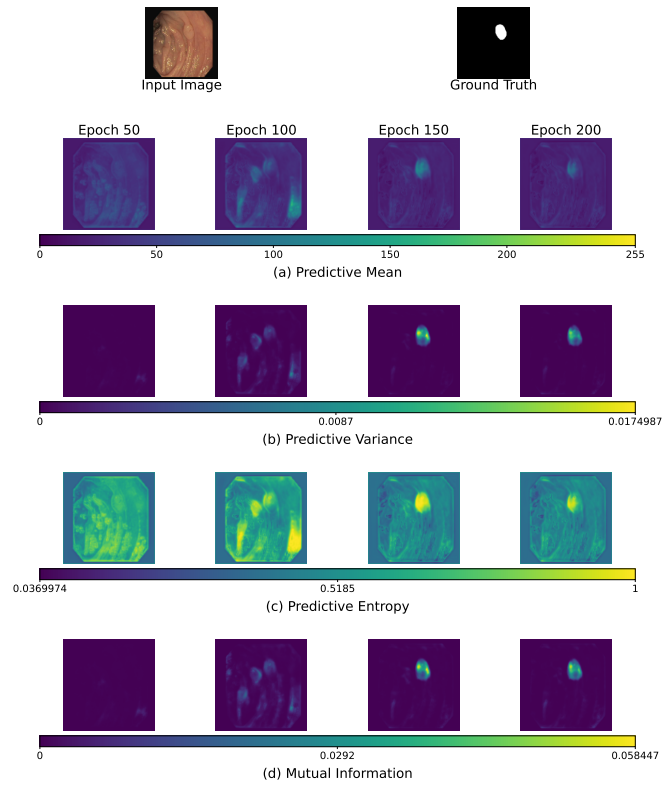


Figure 3: Example of Challenging cases with persistently low performance: (a) Predictive Mean, (b) Predictive Variance, (c) Predictive Entropy, (d) Mutual Information

参考文献

- [1] G.-P. Ji *et al.*, “Video polyp segmentation: A deep learning perspective,” *Machine Intelligence Research*, vol. 19, no. 6, pp. 531–549, 2022.
- [2] F. Maleki *et al.*, “Machine learning applications for head and neck imaging,” *Neuroimaging Clinics*, vol. 30, no. 4, pp. 517–529, 2020.
- [3] J. Long *et al.*, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [4] F. Milletari *et al.*, “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.
- [5] W. Zhu *et al.*, “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 *et al.*, “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, p. 107695, 2024.
- [8] 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*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1050–1059.
- [9] Z. Leng *et al.*, “Polyloss: A polynomial expansion perspective of classification loss functions,” in *International Conference on Learning Representations*, 2022.
- [10] J. Bernal *et al.*, “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.
- [11] O. Ronneberger *et al.*, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [12] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.