

# 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、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。本稿では、この最終目標に向けた第一段階として、 픽셀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、画像難易度の定量化手法としての妥当性を評価する。

## 2 提案法

### 2.1 概要

本研究では、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を目指す。図 1 に、構想を含む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

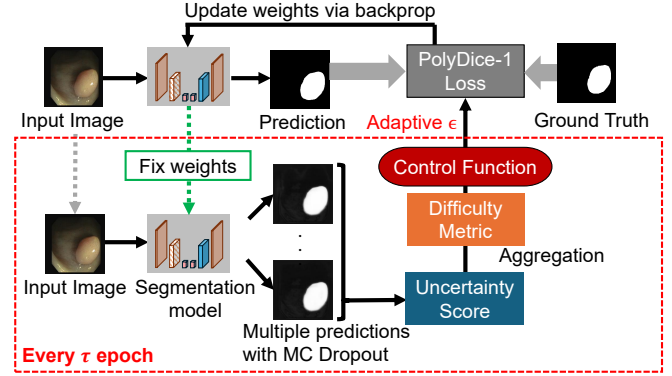


図 1: Overview of the proposed adaptive learning framework

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 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適応的学習を実現する。

### 2.2 PolyDice Loss [8] による損失形状制御

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  $\epsilon$  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。

#### 2.2.1 Dice Loss

画像サイズを  $H \times W$  とし、 픽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。

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

#### 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 は以下のように分解できる。

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

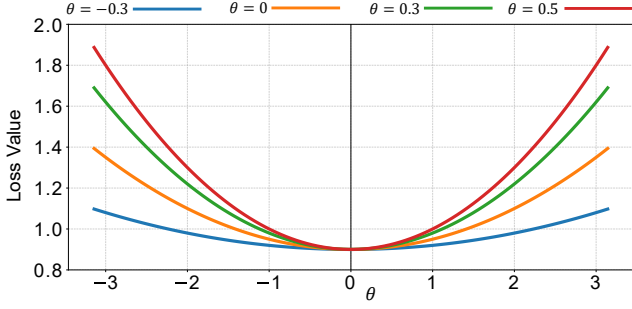


図 2: Plot of PolyDice-1 Loss ( $s = 0.1$ )

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。

### 2.2.3 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 2 に、 $\epsilon$  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。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  $\epsilon$  を不確実性に基づいて動的に調整する。

## 2.3 不確実性による難易度評価

### 2.3.1 MC Dropout の原理と適用

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 を学過程の各段階で適用する。具体的には、 $\tau$  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。

### 2.3.2 学習中の MC Dropout 推論

学習エポック  $e \in \{\tau, 2\tau, \dots, E\}$  において、次の手順で不確実性評価を実施する。セグメンテーションモデル  $f_{\theta} : \mathcal{X} \rightarrow [0, 1]^{H \times W}$  に対し、エポック  $e$  時点のパラメータ  $\theta^{(e)}$  を固定する。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い、予測集合  $\{\hat{\mathbf{Y}}^{(n)}\}_{n=1}^N$  を得る：

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

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)、各  $\hat{\mathbf{Y}}^{(n)} = \{\hat{y}_{i,j}^{(n)} \in [0, 1]\}_{i,j}$  は  $n$  回目の推論における予測確率マップである。この確率的推論により、同一画像に対して複数の予測マップが得られ、そのばらつきから「モデルがどれだけ迷っているか」を定量化可能である。

### 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：

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

ここで、 $\bar{y}_{i,j} = \frac{1}{N} \sum_{n=1}^N \hat{y}_{i,j}^{(n)}$  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：

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

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：

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

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  $\mathbf{U} = \{u_{i,j}\}_{i,j}$  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  $\epsilon$  を動的に制御する：

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

ここで、 $g$  は制御関数、 $\theta_g$  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  $\epsilon$  を割り当てて急峻な勾配を与え、簡単な画像には小さな  $\epsilon$  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

医用画像セグメンテーションにおける画像難易度の指標としての不確実性の妥当性を検証するため、CVC-ClinicDB [12] データセットを用いて実験を行った。本データセットは、612 枚の大腸内視鏡画像 ( $384 \times 288$  pixel) 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。データを 8:2 の比率で訓練・テストに分割し、訓練データの 10% をランダムに抽出して検証データとした。なお、本稿では予備的な解析として、このうち訓練データのみを使用して不確実性指標の推移を分析した。

#### 3.1.2 実装の詳細

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 [13] を採用した。学習には Adam optimizer [14] 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い、ピクセル単位の不確実性指標を計算した。

表 1: Overview of the U-Net Architecture

Layer	Output Size
— Encoder —	
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)	$14 \times 14 \times 512$
— Decoder —	
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)	$224 \times 224 \times 2$

## 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画像群から代表例を 1 つ選択し (200epoch 時点で Dice 係数 0.981, 以下「簡単な例」), 学習後期においても精度が低い画像群から代表例を 1 つ選択し (200epoch 時点で Dice 係数 0, 以下「難しい例」), それぞれの不確実指標の推移を分析した。図 3 に簡単な例, 図 4 に難しい例の 50, 100, 150, 200 epoch 時点での (b) 予測分散, (c) 予測エントロピー, (d) 相互情報量を示す。参考として, MC Dropout による平均予測マップを (a) に併記している。カラースケールは両例間で統一し, 全ピクセル・全エポックにおける最大値で正規化した。

簡単な例では, 予測平均 (図 3(a)) が 50 エポック時点で既にポリープ領域を明確に捉えており, 学習の進行とともに予測がより鮮明になった。予測分散 (図 3(b)) と相互情報量 (図 3(d)) は学習の進行とともに急速に減少し, 特に相互情報量は 100 エポック以降でほぼ消失した。これはモデルが早期に安定した特徴表現を獲得したことを示している。一方, 予測エントロピー (図 1(c)) はポリープ境界部で高い値を示し続け, 境界の曖昧性が残存していることが分かる。

難しい例では, 予測平均 (図 4(a)) がポリープの正確な位置を捉えられず, 200 エポック時点でも高い予測スコアが得られていないことが観察された。予測分散 (図 4(b)) と相互情報量 (図 4(d)) は, 学習初期には低い値であったが, 100 エポック以降はポリープ領域に集中して高い値を示した。特に相互情報量は 200 エポック時点でもポリープ中心部で高い値を維持しており, モデルが一貫した特徴表現を獲得できていないことを示している。予測エントロピー (図 4(c)) は 50 エポック時点で画像全体に広がっていたが, 学習の進行とともにポリープ領域に集約される傾向を示した。

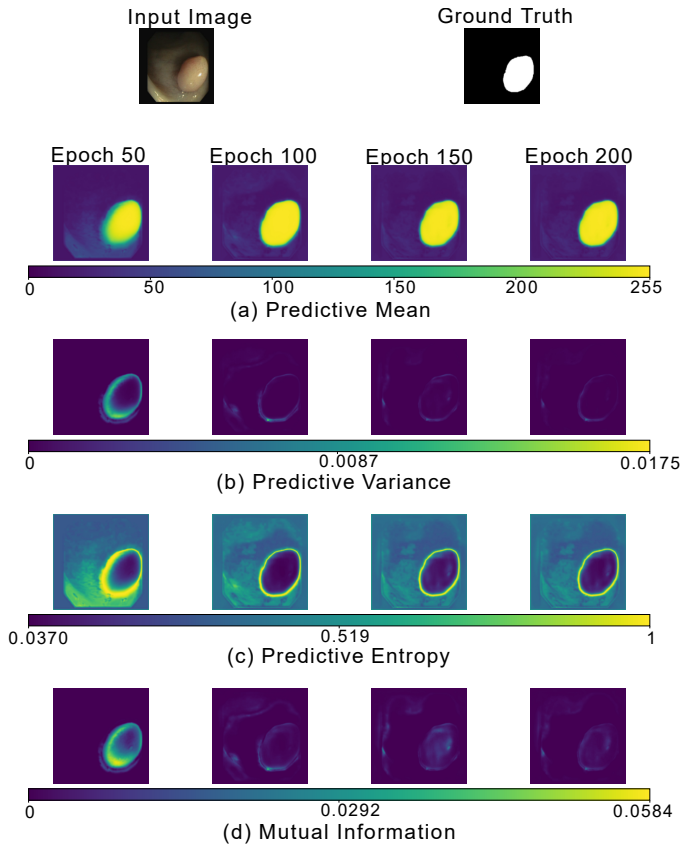


図 3: Temporal evolution of uncertainty metrics for an easy case (Dice = 0.981 at epoch 200).

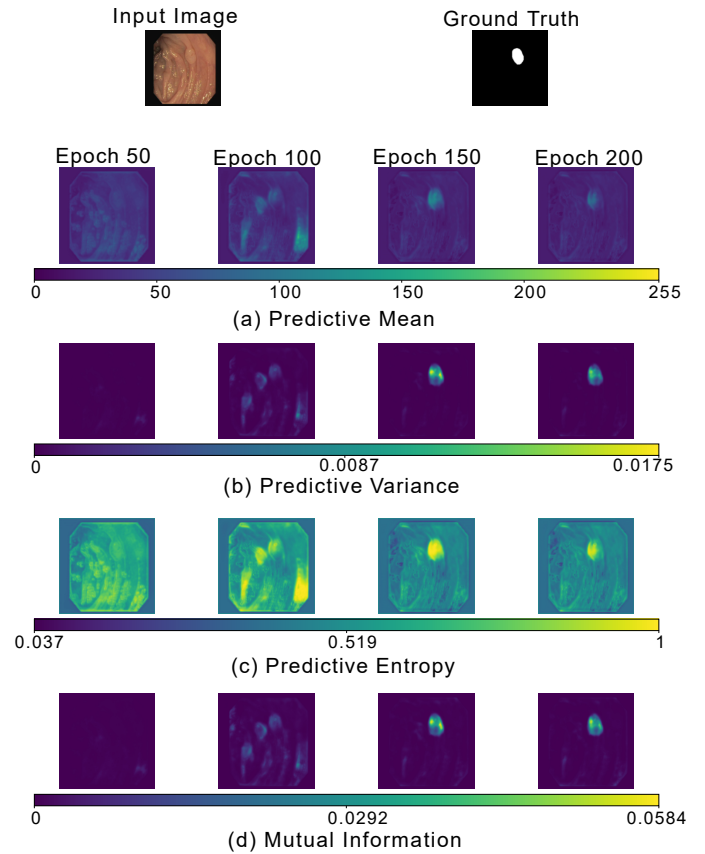


図 4: Temporal evolution of uncertainty metrics for a difficult case (Dice = 0.0 at epoch 200).

### 3.2.2 不確実性指標の比較分析

今回の検証から、各不確実性指標が捉える「難しさ」の側面について、2つの代表例の結果を比較分析した。相互情報量は、簡単な例では100エポック以降でほぼ0に収束したのに対し、難しい例ではポリープ領域で高い値を維持し続けた。この差異は、相互情報量がモデルの知識不足に起因する認識的不確実性を的確に捉えていることを示している。

一方、予測エントロピーは、両例において境界部分で高い値を示し続けた。これは、学習が十分に進んだ後も学習が残存する「データ固有の曖昧さ」に起因する偶然的不確実性を反映していると考えられる。予測分散は相互情報量と類似した傾向を示したが、値のダイナミックレンジが小さく、症例間の差異が相互情報量ほど明確ではなかった。

### 3.2.3 適応的損失関数への示唆

本検証により、相互情報量が画像難易度の指標として最も妥当である可能性が示された。簡単な例では相互情報量が収束し、難しい症例では学習後期でも高い値が残存するという対応関係が確認されたためである。この知見に基づき、相互情報量を集約し難易度指標  $D$  が高い画像に対しては PolyDice-1 Loss のパラメータ  $\epsilon$  を増加させ、急峻な勾配を与える制御が有効であると考えられる。

実装上の課題として、ピクセル単位の相互情報量を画像単位に集約する方法の設計が挙げられる。本実験で観察された空間的分布、特に難しい例では不確実性がポリープ内

に広範囲に分布する傾向を考慮し、正解マスク領域に着目した集約方法が有望である。これにより、病変検出の精度向上に直結する適応的学習の実現が期待される。

## 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ける画像難易度に応じた適応的学習の実現に向けた基礎検証を行った。提案法の構想では、Poly Dice Loss のパラメータ  $\epsilon$  を画像難易度に基づいて動的に制御することで、難しい画像に対する学習の重点化を目指す。本稿では、その第一段階として、MC Dropout によるピクセル単位の不確実性の指標（予測分散、予測エントロピー、相互情報量）が画像難易度の定量化に有効であるかを検証した。

CVC-ClinicDB データセットを用いた実験の結果、学習が簡単な画像では相互情報量が早期に収束するのに対し、難しい画像では学習後期でもポリープ領域で高い相互情報量が維持されることが確認された。この明確な対応関係から、相互情報量が認識的不確実性を的確に捉え、画像難易度の指標として妥当であることが示された。今後は、ピクセル単位の相互情報量を画像単位に集約し、その値に基づいて  $\epsilon$  を動的に制御する手法をする。また、適応的制御による精度向上効果の定量的検証を実施する予定である。



## 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] H. Aizawa, “Polynomial dice loss for medical image segmentation(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, 20–22 Jun 2016, pp. 1050–1059.
- [10] Z. Leng *et al.*, “Polyloss: A polynomial expansion perspective of classification loss functions,” in *International Conference on Learning Representations*, 2022.
- [11] N. Srivastava *et al.*, “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 *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.
- [13] 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.
- [14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.