UNet++を用いたCT画像からの疾病領域検出における多スライス間損失による小領域検出精度の改善

後藤 拓矢 †‡　中野 宏毅 ‡　山下 淳 †　淺間 一 †
† 東京大学　‡ 日本 IBM
E-mail: gototakuya@robot.t.u-tokyo.ac.jp

1 序論

医療画像から疾病領域を検出することは切除手術や治療の経過を把握する上で重要である [1]．複数の剪断画像を組み合わせた Computed Tomography（CT）画像は単一の画像と異なり 3 次元の情報を含むため，単一のスライスでは判別が困難な疾病領域の検出が可能である．しかし，人手によって CT 画像から疾病領域を検出することは多くのスライスを処理する必要があり，より多くの負担を要する [1]．

人手による負担軽減のため，疾病領域の自動検出手法についての研究が多くなされている．近年では Convolutional Neural Network（CNN）により画像認識の性能が飛躍的に向上したことを受け，医療画像からの疾病領域においても CNN の適用が盛んである．CT 画像を対象とした研究においても，2 次元の畳み込みを用いた CNN（2D-CNN）を 3 次元の畳み込みを用いた CNN（3D-CNN）に拡張することで従来の画像処理を用いた手法から大きく精度が向上している [2,3]．医療画像では疾病ではない領域に比べて疾病領域は小さく，学習データには偏りがある．CNN を用いた自動検出では過学習により本来疾病領域である部分が False-Negative として検出されない場合が多く，とりわけ小さな疾病領域の検出率が低いことが課題となっている．また，2D-CNN に比べて 3D-CNN は計算量が多く，必要とするリソースが多いため適用が困難であるという課題がある．

本研究は，単一のスライスからは検出が困難な小さな疾病領域の検出精度を改善することを目的とする．そのためには検出漏れを防ぐ効果を示す，Recall を高める必要がある．目的達成のために，CT 画像からの疾病領域検出において複数のスライスを入力に用いる．また，入力したスライス間の関係性を考慮して学習することで，疾病領域を空間的に捉えることが可能となり，小さな疾病領域の検出が容易になる．提案手法では 3 次元成分の学習のために複数のスライスを用いて正則化を行う Multi-Slice Loss（MS Loss）を導入する．また，領域検出には必要とする計算リソースが少ない 2D-CNN である UNet++ [4] を用いる．

2 関連研究

Çiçek らは MRI からの臓器領域検出において 3D-CNN を適用し 2D-CNN による結果より検出精度が高いことを示した [2]．しかし，3D-CNN は入力サイズが大きくパラメータ数が多いため，計算リソースの制約により複雑な構造を適用することは困難である．ここで，領域検出 CNN における前半の特徴抽出を担う部分は Back Born とよばれる．公開されている 3D-CNN のモデルは少なく，Back Born に事前学習済みのモデルを適用した転移学習による学習時間の短縮ができないため学習に時間がかかるという課題がある．Li らは CT 画像からの肝臓領域と肝腫瘍領域の検出において 3D-CNN と 2D-CNN を組み合わせて共通の特徴抽出に事前学習した 2D-CNN を用いる手法を提案した [3]．しかし，3D-CNN を用いている上に 2D-CNN も併用しているため，必要な計算リソースは増えており，依然として 3D-CNN の学習は困難である．

Chlebus らは 3D-CNN ではなく 2D-CNN による結果を位置や CT 画像における数値などの固定の特徴量によって事後処理を行うことで False-Positive を減らす手法を提案した [5]．しかし，あらかじめ設定した固定の特徴量のみを用いているため，設定していない特徴を持つ疾病には対応できないという課題がある．また，過学習により False-Negative として検出されなかった疾病領域は事後処理の対象とはならないため小さな腫瘍は検出が困難であるという課題がある．Cai らはスライス間の関連性を 2D-CNN によって学習するために予測対象のスライスに加えて隣接するいくつかのスライスを組み合わせて入力に用いる手法を提案した [6]．入力の次元が増えてはいるが，2D-CNN を用いているため必要な計算リソースは 3D-CNN に比べて少ない．しかし，2D-CNN では入力のスライス間の関係性を学習に考慮されないという課題がある．

3 提案手法

図 1 の上段に入力するスライスの例，下段に医師によって作成された疾病領域のアノテーションの例を示

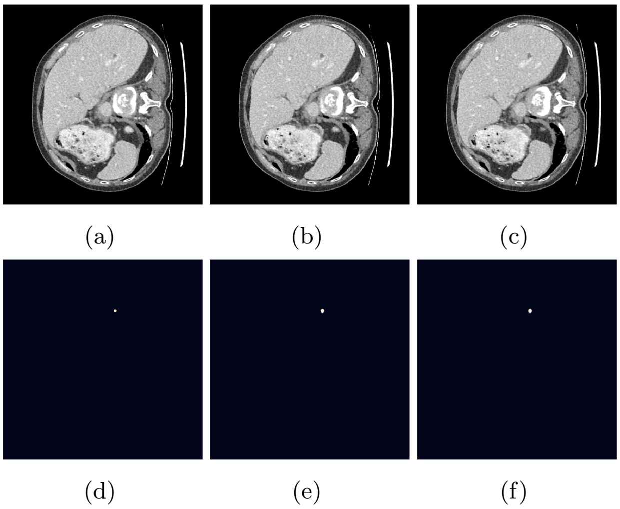

図 1: 入力スライスの例（上段）と疾病領域のアノテーションの例（下段）

す．疾病領域は肝腫瘍領域である．図 1(a)，(b)，(c) は 1 つの CT 画像における連続する 3 枚のスライスである．図 1(d)，(e)，(f) はそれぞれ図 1(a)，(b)，(c) のスライスに対するアノーションである．図 1(a)，(b)，(c) に示したような含まれる疾病領域が小さい場合，単一のスライスから疾病領域の検出は非常に困難である．また，スライスには位置情報が含まれていない．CT 画像は撮影方法によってスライス間の距離が異なるため，連続したスライスであっても，スライス間の関係性は CT 画像ごとによって異なる．複数のスライスを入力し，入力したスライス間がどの程度近いかという，スライスの関係性を学習に考慮することで，複数のスライスに含まれる疾病領域を空間的に捉えることが可能となり疾病領域の検出が容易になる．

CNN では入力および中間から最終までの出力における画像平面に対して鉛直方向の次元をチャネルと呼ぶ．2D-CNN では畳み込みの際に同一フィルタによって出力先の 1 つのチャネルに集約される．したがって，入力スライス間の関係性は学習に考慮されない．そこで本手法では 2D-CNN に複数のスライスを入力するだけではなく，学習時の損失関数において複数のスライス間の関係性を加味して複数のアノテーションから計算される Multi-Slice Loss（MS Loss）を導入することで複数のスライス間の関係性を考慮する．また，MS Loss を一般的な交差エントロピー損失（BCE Loss）に加えて用いることで，MS Loss が正則化項として過学習を防ぐ効果を期待できる．

本研究では，領域検出ネットワークとして 2D-CNN である UNet++ [4] を用いる．UNet++は 2D-CNN で構成されるため 3D-CNN に比べて必要な計算リソースは少ない．また，Back Born 部分に事前学習した任意のネットワークを用いることができるため学習時間の短縮が期待できる．

一般的な学習手法では予測対象となる単一のスライスを入力し，出力される疾病領域の予測と予測対象のスライスに対する教師データにおけるアノテーションを比較することで損失を計算する．一方，提案手法では予測対象のスライスと隣接するスライスの複数のスライスを入力し，対象のスライスに対する疾病領域の予測と対象のスライスおよび隣接するスライスに対する複数の教師データにおけるアノテーションを比較することで損失を計算する．

Multi-Slice Loss を式 (1)，(2)，(3) によって定義する．

$$\mathcal{L}_{ms} = \sum_t w(X_i, X_{i+t}) \mathrm{bce}(\hat{Y}_i, Y_{i+t}), \tag{1}$$

$$t \in \{-d, -d+1, \cdots, -1, 1, \cdots, d-1, d\}, \tag{2}$$

$$w(X_i, X_{i+t}) = \exp(-\lambda \mathrm{dis}(X_i, X_{i+t})), \tag{3}$$

ここで X_i は CT 画像における最初のスライスを 0 とし，順に数えた際の i 番目のスライスを示す．$w(X_i, X_{i+t})$ は，X_i と X_{i+t} から式 (3) によって定義されるスライス間の関係性を表す重みである．$\hat{Y}_i$ はスライス X_j に対する疾病領域の予測，Y_i はスライス X_i に対応する教師データのアノテーションを示す．Y_i は，疾病領域であれば 1 それ以外であれば 0 で定義される．$\mathrm{bce}(\hat{Y}_i, Y_{i+t})$ はスライス画像 X_i に対する $\hat{Y}_i$ と Y_{i+t} から計算される交差エントロピー損失関数を示す．ここで $\mathrm{dis}(X_i, X_{i+t})$ はスライス X_i，X_{i+t} の類似度を示す．CT 画像の場合，スライスの距離が離れるほど含まれる臓器などの位置は変化せず，大きさが変化することが多い．臓器の大小によって類似度を変化させるために，類似度の計算にはスライスを Oriented FAST and Rotated BRIEF（ORB）[7] を用いてベクトル化し，ハミルトニアン距離を用いる．λ は距離の正規化パラメータとする．学習時には式 (4) で示した MS Loss に予測対象のスライス X_i に対する予測 $\hat{Y}_i$ と，X_i に対応する教師データのアノテーション Y_i との交差エントロピーを加えた損失関数を最小化するように 2D-CNN の重みを更新する．

$$\mathcal{L} = \mathrm{bce}(\hat{Y}_i, Y_i) + \mathcal{L}_{ms}. \tag{4}$$

4 評価実験と結果

4.1 評価実験の設定

データセットは LiTS Challenge [1] で公開されている LiTS データセットを用いる．LiTS データセットは腹部 CT 画像と肝臓および肝腫瘍の領域のアノテーションを含む．それぞれのスライスのボクセル数は 512×512 で解像度は $0.56 - 1$mm である．スライス間の解像度は

表 1: 実験結果

手法	MAP	MAR	MAF	mDice
2D input w/ BCE Loss	**0.822**	0.578	0.553	0.553
3D input w/ MS Loss（提案手法）	0.762	**0.669**	**0.604**	**0.604**

0.45 – 6.0mm である．本研究では LiTS データセットにおける肝腫瘍を含む 118 セットの肝臓部分のみを用いる．さらに含まれる肝腫瘍のボクセル数の分布が均一になるようにトレーニング，バリデーション，テストの目的に分割して用いる．本研究では LiTS データセットにおける肝腫瘍のアノテーションのみを用いる．

実験における U-Net++ [4] の Back Born には ImageNet データセットによって事前学習した EfficientNet B4 [8] を用いる．学習はバッチサイズを 8 として，50 エポック行う．

入力には予測対象のスライスに隣接する前後のスライスを加えた 3 スライスを入力する．提案手法の式 (2)，(3)，(4) におけるパラメータは $d = 2$，$\lambda = 0.5$ とする．

評価時には予測値 $\hat{Y}_i$ における閾値 p 以上である領域を予測領域とする．評価指標にはスライスごとの Precision，Recall，F 値および Dice 係数を平均した，MAP，MAR，MAF，mDice を用いる．評価時における閾値 p は学習した 2D-CNN にバリデーションデータセットを入力し，評価した mDice が最も大きくなる p を選択する．

4.2 実験結果

実験結果を表 1 に示す．2D input は単一スライス入力を示し，3D input は予測対象のスライスとその隣接する前後 1 スライスを加えた 3 スライスを入力することを示す．BCE Loss は学習時に損失関数に交差エントロピーを用いたことを示し．MS Loss は提案手法における式 (4) を用いたことを示す．2D input w/ BCE Loss は一般的な 2D-CNN における手法のため，従来手法である．

MAP 以外の評価手法において提案手法による結果が高くなっていることがわかる．また．従来手法の 2D input w/ BCE Loss と提案手法の 3D input w/ MS Loss における MAR を比較すると 2D input と BCE Loss を用いて学習した場合と比べて提案手法では 9%ほど向上している．Recall が高いということは検出漏れを防ぐ性能が高いということである．医療画像における疾病領域の検出の場合には予測が間違えないことよりも漏れなく検出する方が重要であると考えられる．従来手法の 2D input w/ BCE Loss と提案手法の 3D input w/ MS Loss における mDice を比較しても提案手法で従来手法より高い結果を得られた．

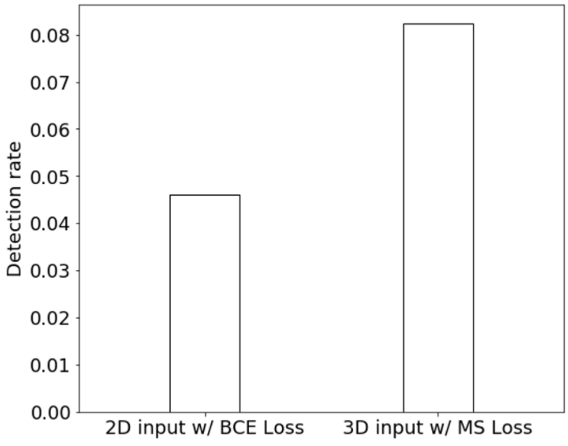

図 2: 1000 ピクセル以下の疾病領域の検出率

5 考察

小さな疾病領域の検出における効果を考察する．スライスにおいて 1000 ピクセル以下の疾病領域に対する検出率を，従来手法と提案手法による結果を元に比較する．対象としている LiTS データセットの平均疾病領域面積は 1.58×10^5 ピクセルであるため，1000 ピクセル以下の疾病領域は小疾病領域といえる．また，スライスの解像度が最大 1 mm であるため 1000 ピクセル以下の疾病領域は $10\,\mathrm{cm}^2$ 以下の疾病領域である．含まれる疾病領域の面積が 1000 ピクセル以下のスライスをテスト対象スライスとする．スライスごとに Ground Truth の疾病領域と予測領域が一致した部分があれば検出したとみなして，テスト対象スライスに対する疾病を検出したスライスの割合を検出率とする．

図 2 にスライスにおける疾病の領域面積が 1000 ピクセル以下の小疾病領域に対する検出率の結果を示す．従来手法を 2D input w/ BCE Loss と，提案手法を 3D input w/ MS Loss と表記している．疾病対象のスライスに対するアノテーションのみから計算する BCE Loss を用いて学習した場合より，複数のスライスを入力し，複数のアノテーション画像から計算する MS Loss を用いて学習した場合の方が面積が 1000 ピクセル以下の疾病の検出率がおよそ 2 倍高い．

以上より，MS Loss を導入することによって単一のスライスからは検出が困難な小さな疾病領域を検出できるようになった．

6 結論

本研究は，単一のスライスからは検出が困難な小さな疾病領域の検出精度を改善した．また，CT 画像からの疾病領域検出において 2D-CNN である UNet++を用いて 3 次元情報を学習する MS Loss を提案した．MS

Loss を導入することで Recall を高くし，検出漏れを防ぐ効果があることが示された．

今回，MS Loss の導入により予測が可能となった小さな疾病領域には，False-Positive となる予測が多いことが課題として残っている．今後，予測領域を精緻化することで False-Positive を減らし，より正確な疾病領域検出を行う手法を検討していく．

参考文献

[1] P. Bilic, P.F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.W. Fu, X. Han, P.A. Heng, J. Hesser et al.: “The Liver Tumor Segmentation Benchmark (LiTS),” arXiv preprint arXiv:1901.04056, 2019.

[2] Ö.Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, O. Ronneberger: “3D U-net: Learning Dense Volumetric Segmentation from Sparse Annotation,” In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp.424–432, 2016.

[3] X. Li, H. Chen, X. Qi, Q. Dou, C.W. Fu, P.A. Heng: “H-DenseUNet: Hybrid Densely Connected Unet for Liver and Tumor Segmentation from CT Volumes,” IEEE Transactions on Medical Imaging, vol.37, no.12, pp.2663–2674, 2018.

[4] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang: “Unet++: A Nested U-net Architecture for Medical Image Segmentation,” In Proceedings of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, vol.11045, pp.3–11, 2018.

[5] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, H. Meine: “Automatic Liver Tumor Segmentation in CT with Fully Convolutional Neural Networks and Object-Based Postprocessing,” Scientific Reports, vol.8, no.15497, 2018.

[6] J. Cai, Y. Tang, L. Lu, A. P. Harrison, K. Yan, J. Xiao, L. Yang, R. M. Summers: “Accurate Weakly Supervised Deep Lesion Segmentation on CT Scans: Self-Paced 3D Mask Generation from RECIST,” arXiv preprint arXiv:1801.08614, 2018.

[7] E. Rublee, V. Rabaud, K. Konolige, G. Bradski: “ORB: An Efficient Alternative to SIFT or SURF,” In Proceedings of the IEEE International Conference on Computer Vision, pp.2564–2571, 2011.

[8] “EfficientNet,” https://github.com/qubvel/efficientnet, 2019.