

# 不鮮明な疾患領域輪郭に対応した 輪郭誤差を含むアノテーションにおける CT 画像からの疾患領域検出 \*

後藤拓矢 \*\* 中野宏毅 \*\*\* 浅間 一† 山下 淳†

Disease Segmentation Using a Robust Method for Unclear Boundaries in Training with Annotations Including Boundary Errors

Takuya GOTO, Hiroki NAKANO, Hajime ASAMA and Atsushi YAMASHITA

The predictive performance of segmentation highly depends on the quality of annotations. However, in case of medical images, it is difficult to annotate precise region of disease since they have variety of shapes and ambiguous boundaries. Improving segmentation performance of models trained with rough annotation will make it easier to create training data. This paper proposes a novel method to improve auto tumor segmentation from computed tomography (CT) images using noisy annotations. The proposed method uses pre-training to detect rough tumor regions and fine-tuning to refine the roughly detected regions. In order to refine roughly detected regions, Active Contours Dice (ACD) Loss is introduced in fine-tuning. In our experiments using simulated rough annotations, the proposed method improves segmentation performance.

**Key words:** medical image processing, computed tomography, 3D segmentation, rough annotation

## 1. 序 論

腫瘍などの診断において画像診断は疾患の状態をより正確に知るために重要な診断手法である<sup>1)</sup>。画像診断では疾患の有無だけでなく、良性か悪性かの判断や外科手術の計画、治療効果の判定のために疾患の大きさ、輪郭、形状を含む疾患の領域を正確に検出することが重要である<sup>2)</sup>。しかし、医師にとって Computed Tomography (CT) 画像など膨大な量の画像データから不鮮明で微細な異常箇所を検出することは多大な労力を要する。そこで、診断支援として医用画像からの自動疾患領域検出の実現に期待が高まっている。本論文では疾患として肝腫瘍を対象とする。

近年、Deep Convolutional Neural Network (DCNN) による領域検出の精度が向上していることから、医用画像からの疾患領域検出においても DCNN を用いた研究が多くなされている<sup>3)</sup>。DCNN の学習のためには対象となる医用画像とそれに対する検出対象の領域を示す領域アノテーションが必要である。また、高い領域検出のための DCNN を学習するためには学習データの領域アノテーションが疾患領域を正確に示していることが重要である<sup>4)</sup>。

CT 画像の 1 スライスと対応する学習データとして医師によって作成された肝腫瘍領域のアノテーションの例を図 1 に示す。図 1(b) において白く示された部分が疾患領域のアノテーションである。図 1(a) の同じ位置が疾患領域であることを示す。図 1(a) からわかるように CT 画像における疾患領域は臓器や骨などの領域に比べて周辺の正常な組織の画素値から緩やかに変化し、また変化量は小さく輪郭が不鮮明である。複雑な形状の疾患領域に対して正確な領域アノテーションの作成は困

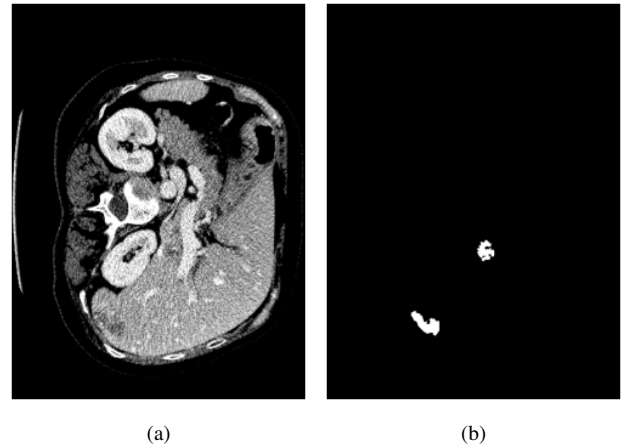


Fig. 1 Samples of input slices (a) and annotations of tumor (b)

難である。したがって、医用画像における疾患領域と領域アノテーションの間には輪郭の誤差が生じやすい<sup>4)</sup>。本論文では、領域アノテーションと実際の疾患領域との誤差を輪郭誤差とする。多くの公開されている学習データでは、できる限り輪郭誤差が小さく信頼度が高い疾患領域のアノテーションを作成される<sup>1)</sup>。図 1(b) は、複数の熟練の医師によって領域アノテーションを作成した後、互いに領域アノテーションに関して精査と修正を繰り返し、医師の間で合意のもと最終的な領域アノテーションが作成された輪郭誤差を含まない例である。しかし、膨大な量の CT 画像のスライスに対して輪郭誤差が含まないように領域アノテーションを作成することには多くの労力を要するため、領域アノテーション作成の簡易化が求められている<sup>5)</sup>。図 1(b) のように疾患領域の輪郭に沿った領域アノテーションを作成することと比べて、楕円やバウンディングボックスなど平易な図形によって疾患を含む領域アノテーションを作成することは簡易的である。しかし、実際には複雑な形状を持った疾患領域に対して平易な図形によって作成された領域アノテ

\* 原稿受付 令和 4 年 5 月 26 日

掲載決定 令和 4 年 8 月 31 日

\*\* 学生会員 東京大学大学院 (千葉県柏市柏の葉 5-1-5)

\*\*\* 正 会 員 日本 IBM (東京都中央区日本橋箱崎町 19-21)

† 正 会 員 東京大学大学院

ションには必ず輪郭誤差が含まれる。したがって、輪郭誤差を含む領域アノテーションを用いた場合においても、自動疾患検出によって輪郭誤差を小さく検出することが求められる。そこで、本研究では腫瘍に対して平易な図形によって作成された領域アノテーションを対象とする。本研究では、輪郭誤差を含む領域アノテーションを用いた場合における、疾患領域検出精度の改善を目的とする。

本研究は、医用画像として腹部 CT 画像を対象とする。また、対象とする疾患として肝腫瘍を用いる。使用する CT 画像は対象臓器の検査を目的として撮影された CT 画像である。本研究は、一般に公開されているデータを用いて行われており、倫理委員会等の承認は不要である。

## 2. 関連研究

従来の CT 画像を用いた領域検出では、Active Contours<sup>6)</sup> や Graph Cut<sup>7)8)</sup> などの画像処理技術を利用した手法が医用画像に適用されてきた。しかし、画像処理技術を利用した手法では、対象とする疾患ごとの特徴量を設計する必要があるため適用範囲が限られてきた。

近年では、特徴量の抽出を含めて学習する DCNN を用いた手法の精度が高く主流となっている。Ronneberger らが提案した UNet<sup>9)</sup> は医用画像を対象とした領域検出に多く用いられている DCNN モデルである。UNet には 3D UNet<sup>10)</sup>, ResUNet<sup>11)</sup> など多くの派生モデルが提案されている。しかし、これらの DCNN モデルは学習データとして実際の疾患領域に対して正確な領域アノテーションを対象としており、輪郭誤差を含む領域アノテーションに対応できない。

輪郭誤差を含む領域アノテーションを用いた疾患領域検出モデルの学習手法として、Wang らは Noise-Robust Loss を導入し、領域アノテーションの輪郭誤差に対してロバストな学習手法を提案した<sup>5)</sup>。Wang らは経験の浅い医師によって作成された輪郭誤差を含む領域アノテーションを対象とした。Wang らの手法では、学習データの領域アノテーションや検出した領域に含まれる輪郭誤差を補正することが考慮されていない。その結果、輪郭誤差を含む領域アノテーションを用いて学習されたモデルによって検出された領域には輪郭誤差が含まれる。したがって、輪郭誤差に対してロバストな学習手法では、全ての領域アノテーションが輪郭誤差を含む場合に検出精度が低い。

Acuna らは路上などの風景画像に含まれる一般物体を対象とした領域検出モデルの学習において多角形によって簡易的に作成した領域アノテーションに対応する手法を提案した<sup>12)</sup>。Active Contours の 1 つである Morphological Geodesic Active Contours (GAC)<sup>13)</sup> によって領域アノテーションの輪郭誤差を補正した。領域の輪郭において画素値が不連続に変化する一般物体などと比べて疾患領域の輪郭は進行度合いによって画素値が緩やかに変化し、不鮮明な場合がある。したがって、Active Contours などの画像処理手法によって領域アノテーションの輪郭誤差を補正する手法では期待通りに疾患領域の輪郭に沿った領域に補正されるとは限らない。Zhao らは細胞領域の検出において、領域内の任意の点によるポイントアノテーションを初期アノテーションとし、複数のモデルを用いて互いに学習した出力結果によってアノテーションを更新していくブートストラップ法による手法を提案した<sup>14)</sup>。しかし、輪郭が鮮明で検出が容易な細胞を対象としており、不鮮明な輪郭に対応できない。したがって、疾患領域の不鮮明な輪郭への対応

が課題である。

以上、疾患を対象として輪郭誤差を含む領域アノテーションを用いた従来研究では、全ての領域アノテーションが輪郭誤差を含む場合に検出精度が低いという課題がある。また、一般物体領域や臓器領域を対象とした輪郭誤差を含む領域アノテーションを用いた従来手法では、疾患の不鮮明な輪郭への対応が困難であるという課題がある。本研究では、領域アノテーションに含まれる輪郭誤差に対応し学習データより高い検出精度を実現するため、学習において領域アノテーションの輪郭誤差を補正する Active Contours Dice Loss 新たに設計する。また、不鮮明な輪郭に対応するため、輪郭誤差を補正する ACD Loss の効果を領域アノテーションの内部に止めるように制御する学習手法を提案する。

## 3. 提案手法

DCNN は学習データにおける入力画像と領域アノテーションから領域アノテーション部分における入力画像の共通する特徴を抽出し、その特徴に合致する領域を検出するように学習する。DCNN ではある特定の画素についてその周辺の情報を加味して特徴抽出を行う。したがって、学習データにおいて疾患領域の輪郭と領域アノテーションの輪郭が一致しているものも含まれる場合には、正しい輪郭画素の周辺の疾患領域の情報から対象領域の特徴を抽出し、検出ができると考えられる。Wang ら<sup>5)</sup> の輪郭誤差に対してロバストな学習を行うことができる Noise-Robust Loss を用いた学習では DCNN の周辺の情報を加味して特徴抽出を行う性質を用いて、輪郭誤差を含む領域アノテーションに対する過学習を抑制することによって輪郭誤差に対応していると考えられる。しかし、平易な図形によって簡易的に作成された領域アノテーションを用いた場合、実際の疾患領域の輪郭と領域アノテーションの輪郭は一致しておらず、輪郭誤差が必ず含まれる。その結果、輪郭誤差部分の周辺に疾患領域がないため疾患領域の特徴抽出が学習できず疾患領域を検出できないか、領域アノテーションの形状を過学習することによって輪郭誤差部分の特徴抽出も学習され、検出領域に輪郭誤差が含まれると考えられる。学習データの精度より高い検出精度を実現するためには、領域アノテーションまたは検出領域に含まれる輪郭誤差を補正する必要がある。そこで、本研究では輪郭誤差を含めて領域を検出するモデルを学習し、検出された領域の輪郭誤差を補正する 2 段階の手法によって検出精度を高める。

検出領域の輪郭誤差を補正する手法として、検出領域を初期値として Active Contours を適用して輪郭誤差を補正する手法が考えられる。しかし、疾患領域輪郭は不鮮明であるため、より輪郭が鮮明な臓器や骨の領域が検出される可能性がある。したがって、領域アノテーション内部の検出と検出領域の輪郭誤差の補正を並行して行うことが有効である。そこで、Active Contours による検出領域における輪郭誤差の補正を学習によって実現し、領域アノテーションの内部を検出する学習と組み合わせることで解決する。しかし、既存の Active Contours は DCNN の学習に組み込むことができないため、検出領域の輪郭誤差の補正を行うための損失関数である Active Contours Dice (ACD) Loss を新たに設計する。

提案手法の概要を図 2 に示す。まず、輪郭誤差を含む領域アノテーションを教師データとして DCNN を学習する。ここで学習されるモデルは輪郭誤差も含む領域が検出される。この

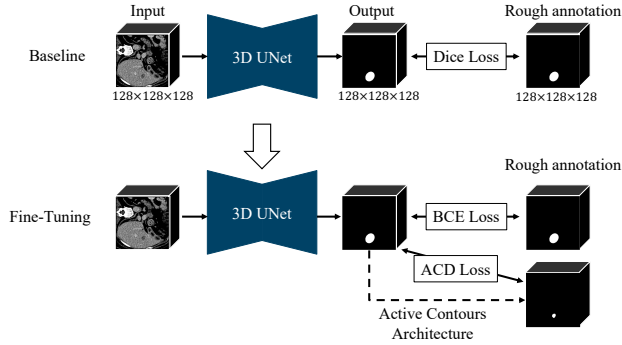


Fig. 2 Training process of proposed method

モデルをベースラインとする。次に、DCNN モデルによる検出領域の輪郭誤差を補正し、より実際の疾患領域に近い領域を検出するように fine-tuning を行う。fine-tuning では DCNN の出力部に DCNN の出力の輪郭誤差を検出するための Active Contours Architecture を追加する。新たに設計した ACD Loss では DCNN の出力と Active Contours Architecture の出力の差である輪郭誤差を損失関数として評価する。また、ACD Loss と輪郭誤差を含む領域アノテーションと比較して領域アノテーションを検出するように学習する Binary Cross Entropy (BCE) Loss を共に最小化する。

### 3.1 Active Contours Dice Loss の設計

検出領域の輪郭誤差を補正するために新たに設計した、Active Contours Dice (ACD) Loss について述べる。ACD Loss の目的はベースラインのモデルにおいて輪郭誤差を含めて疾患領域を検出された領域の輪郭誤差を補正し実際の疾患領域に近づけることである。

検出領域に Active Contours を適用すると、疾患領域の不鮮明な輪郭のため、より輪郭が鮮明な近くの臓器や骨などの領域に収束する恐れがある。医師が介在し、輪郭の閾値や更新回数を調節しながら実行し適度な領域で停止するには、疾患ごとに適当な輪郭の閾値と更新回数が異なるため、全ての疾患に対して個別に行う必要がある。また、CT 画像の場合、3次元画像として目視によって調節することは非常に困難であるため、調節はスライスごとに行わざるを得ない。したがって、膨大な数の調節を行う必要があるため人手によって調節を行うことは非常に困難である。また、機械的な Active Contours のパラメータ調節を行うためには適度に領域更新が行われたかどうかを定量的に評価する必要がある。しかし、実際の疾患領域が未知であるため判断基準を定量的に評価することができない。

そこで、本研究は Active Contours を発展させた損失関数を新たに設計し、付近の輪郭に収束することで輪郭誤差を補正する効果を、損失関数を最小化する学習によって実現する。また、補正する輪郭誤差を損失関数として設計することで、領域アノテーションの内部を検出するように学習する通常の損失関数と組み合わせる学習し、疾患領域の不鮮明な輪郭のために領域アノテーションの外部や検出領域が消失してしまうことを領域内部の検出するように制御することが可能となる。

Active Contours による輪郭誤差を補正する効果を損失関数の最小化として得るため、DCNN の出力における輪郭誤差を検出し、その輪郭誤差を微分可能な関数によって評価する Active Contours Dice (ACD) Loss を新たに設計する。DCNN の出力

における輪郭誤差を検出するため、DCNN の出力部に Active Contour Architecture を追加する。Active Contours Architecture では、Morphological GAC<sup>13)</sup> をベースとした以下の手順によって行う。まず、画素ごとの疾患であるかの確率を示す DCNN の出力を、0.5 を閾値として領域ラベルに変換したものを領域の初期値とする。CT 画像において疾患領域の輪郭は不鮮明であるため、輪郭強調処理を施す。輪郭強調処理では Gaussian Derivatives フィルタを適用後、逆数をとる。Active Contours における輪郭強調処理として Sobel フィルタも用いられるが、Gaussian Derivatives フィルタは画像中のノイズの影響を受けにくいという特徴がある。図 1(a) に示したように CT 画像には隣接する画素同士の連続性が低くノイズが多い。したがって、フィルタサイズによってノイズの影響を軽減できる Gaussian Derivatives フィルタを用いる。また、逆数をとることにより、鮮明な輪郭ほど 0 に近く、平坦な領域は 1 に近くなる。学習によって検出領域の輪郭誤差を少しずつ補正するために ACD Loss 計算の際に適用する領域更新は 1 回のみ行う。

Active Contours Architecture の処理結果と DCNN の出力の差が DCNN の出力結果における輪郭誤差と見なせる。したがって、ACD Loss では Active Contours Architecture の処理結果と DCNN の出力結果の差を Dice によって評価する。ACD Loss は式 (1) で示される。

$$\mathcal{L}_{ACD} = \mathcal{L}_{Dice}(\hat{Y}, \tilde{Y}), \quad (1)$$

ここで、 $\hat{Y}$  は DCNN の出力、 $\tilde{Y}$  は Active Contours Architecture によって更新された領域ラベルを示す。 $\mathcal{L}_{Dice}(\cdot)$  は式 (2) で示される Dice Loss である。

$$\mathcal{L}_{Dice}(\hat{Y}, \tilde{Y}) = 1 - \frac{2 \sum_i \hat{y}_i \tilde{y}_i}{\sum_i \hat{y}_i^2 + \sum_i \tilde{y}_i^2}, \quad (2)$$

ここで、 $\hat{y}_i$  は DCNN の出力における画素  $i$  に対応するクラス確率を示す。 $\tilde{y}_i$  は画素  $i$  に対するラベルを示す。 $\tilde{y}_i$  は  $\tilde{Y}$  において画素  $i$  が疾患領域であれば 1、そうでなければ 0 を示す。ACD Loss が 0 になることは領域が Active Contours Architecture によって更新されなくなったことを示し、領域更新が収束したことに相当する。fine-tuning では式 (3) を最小化する。

$$\mathcal{L} = (1 - \lambda) \mathcal{L}_{region} + \lambda \mathcal{L}_{ACD}, \quad (3)$$

ここで、 $\mathcal{L}_{region}$  は領域アノテーションと DCNN の出力によって評価する損失関数、 $\lambda$  は学習における  $\mathcal{L}_{region}$  と  $\mathcal{L}_{ACD}$  の影響を調整するための重みパラメータである。 $\mathcal{L}_{region}$  の目的は領域アノテーション内部の学習であり、一般的な領域検出モデルの学習に用いられる損失関数が適当である。Binary Cross-Entropy (BCE) Loss<sup>9)</sup> は Dice Loss に比べて領域アノテーションの輪郭誤差の影響を受けにくい特徴がある<sup>5)</sup>。ベースラインでは領域アノテーションと比較する損失関数による学習が輪郭誤差を含めて検出することを目的としていたが、fine-tuning では ACD Loss によって輪郭誤差を補正しようと付近の輪郭に収束しようとする学習において領域アノテーション内部を検出するように弱い制約を与えることを目的としているため、 $\mathcal{L}_{region}$  には輪郭誤差の影響を受けにくい BCE Loss を用いる。 $\mathcal{L}_{region}$  を式 (4) に示す。

$$\mathcal{L}_{region} = - \sum_i y_i \log \hat{y}_i, \quad (4)$$

ここで、 $y_i$  は領域アノテーション  $Y$  の画素  $i$  におけるラベルを示す。

ACD Loss を用いた fine-tuning では、輪郭誤差を含む領域アノテーションを用いて領域誤差が含まない疾患領域を検出することを目的としている。したがって、学習の進捗度合いをバリデーションデータセットにおける検出精度によって検証することができない。fine-tuning では、トレーニングデータセットにおける過学習を抑制するために、バリデーションデータセットにおいて式 (3) によって評価される損失が最小のモデルを選択する。

### 3.2 ベースラインの学習

ベースラインの学習は、輪郭誤差を含む領域アノテーションを用いて学習し、できる限り見落とされる疾患を減らし ACD Loss による fine-tuning によって補正される検出領域を増やすことを目的としている。Wang ら<sup>5)</sup> は輪郭誤差に対してロバストな学習手法として Noise-Robust Loss を提案した。Noise-Robust Loss を式 (5) に示す。

$$\mathcal{L}_{\text{NR}} = \frac{\sum_i |\hat{y}_i - y_i|^\gamma}{\sum_i \hat{y}_i^2 + \sum_i y_i^2}, \quad (5)$$

ここで、 $\hat{y}_i$  は画素  $i$  に対する領域アノテーションのラベルを示す。 $\gamma$  は輪郭誤差に対する感度を調節するパラメータで大きいほど輪郭誤差に対する感度は高くなる。 $\gamma$  が 2 の時は Dice Loss<sup>15)</sup>、1 のときは Mean Absoluted Error (MAE) Loss と同等になる。Wang ら<sup>5)</sup> は実験により  $\gamma = 1.5$  とした時が輪郭誤差に対して中程度の感度となることを示した。式 (5) を改めて解釈すると、 $\gamma$  が小さくなるほど Noise-Robust Loss による勾配は小さくなることから、Noise-Robust Loss は領域アノテーションに対して徐々に学習することで輪郭誤差に対する過学習を抑制していると考えられる。したがって、全ての領域アノテーションに輪郭誤差が含まれる場合、領域アノテーションの学習が十分に進まず Noise-Robust Loss を用いた学習では対応できないと考えられる。

本研究では ACD Loss によって検出領域の輪郭誤差を補正するため、ベースラインでは輪郭誤差を含めて疾患の見落としを防ぐように Dice Loss によって学習を行う。

CT 画像はアーティファクトによるノイズが生じやすい特性がある。可能な限り、入力における CT 画像のにおいて輪郭を保持したままノイズを除去するためにメディアンフィルタを適用する。

疾患領域検出のため DCNN モデルを 3D UNet<sup>10)</sup> と ResUNet<sup>11)</sup> をもとに構成する。3D UNet は医用画像、とりわけ CT 画像や MRI などの 3 次元データにおける領域検出のために多く用いられる DCNN モデルである。3D UNet は Convolution 層と Pooling 層などによって構成されるエンコーダ部、Convolution 層と Transposed Convolution 層などから構成されるデコーダ部の 2 部構造を持ち、エンコーダ部とデコーダ部の間にスキップコネクションを持つという特徴がある。3D UNet は各 Convolution 層において通常の Convolution に用いられるフィルタより 1 次元多い 3D Convolution 層を用いるという特徴がある。ResUNet はエンコーダ部、デコーダ部において residual ブロックが使われており、ブロック同士がスキップコネクションで接続されている。本研究ではエンコーダ部、デコーダ部においてそれぞれ 4 つ、その中間に 1 つの residual ブロックを使用した 3D UNet モデルを構成する。

Table 1 Quality of training dataset

| Dice  | IoU   |
|-------|-------|
| 0.429 | 0.289 |

3D UNet は通常の 2 次元の Convolution 層を用いた UNet と比較してモデルのパラメータ数が多くなる傾向がある。また、入力画像も 3 次元データのためデータ量が大きくなり、計算機のリソースの制約を受ける。本研究では、各ブロックにおけるチャンネル数を 16, 32, 64, 128, 256 とし、学習の際には入力された CT 画像からランダムに  $128 \times 128 \times 128$  のパッチを作り 3D UNet の入力とする。なお、テスト及びバリデーションの際には 50% の重複領域を取りながら  $128 \times 128 \times 128$  の領域をスライドさせてパッチを作成し入力とする。

## 4. 評価実験と結果

### 4.1 評価実験の設定

提案手法の有用性を検証するため、実際の CT 画像を用いて実験を行う。

実験データとして、腹部 CT 画像と肝臓および肝腫瘍の領域アノテーションを含む LiTS データセット<sup>1)</sup> を用いる。LiTS データセットに含まれる肝腫瘍の領域アノテーションは熟練の医師によって作成された輪郭誤差を含まないものである。したがって、実験のために LiTS データセットをもとに本研究の対象である輪郭誤差が含まれた領域アノテーションを擬似的に作成する。擬似データの作成手法は次の通りである。まず、LiTS データセットの肝腫瘍の領域アノテーションをスライスごとに楕円近似する。次に、中心位置に対して軸方向に軸の長さの 5% 以下、楕円軸角度に対して  $\pm 0.1$  rad 以下、長軸と短軸に対して 1.8 倍から 2.2 倍の誤差を含める。実際の領域より大きいアノテーションを作成する方が簡易的であると考え、軸の長さを実際の領域アノテーションより大きく作成した。擬似的に作成した輪郭誤差を含む領域アノテーションの精度を LiTS データセットに含まれる元の領域アノテーションを Ground Truth として評価した結果を表 1 に示す。

LiTS データセットの学習データをランダムに分割し、CT 画像と肝腫瘍の領域アノテーションの 74 セットを学習データセット、19 セットをバリデーションデータセット、18 セットをテストデータセットとして用いる。

検出精度を測る評価指標として、LiTS データセットにおける肝腫瘍領域アノテーションを Ground Truth とした Dice、Intersection over Union (IoU) を CT 画像ごとに平均したものをを用いる。Dice、IoU はどちらも領域検出の精度として用いられる評価指標である。Dice は式 (6) で示される。

$$\text{Dice} = \frac{2|G \cap P|}{|G| + |P|}, \quad (6)$$

ここで、 $G$  は Ground Truth における疾患領域の画素の集合、 $P$  は検出結果における疾患領域の画素の集合、 $|\cdot|$  は集合の要素数を示す。また、IoU は式 (7) で示される。

$$\text{IoU} = \frac{|G \cap P|}{|G \cup P|}, \quad (7)$$

IoU は要素数の差による影響が強く、要素数が異なると著しく値が小さくなる評価指標である。一方で、Dice は分母を平均要素数とすることで要素数の差による影響を軽減した、より共通

**Table 2** Results of an experiment

| Method                        | Dice         | IoU          |
|-------------------------------|--------------|--------------|
| Previous method <sup>5)</sup> | 0.322        | 0.224        |
| Baseline                      | 0.376        | 0.265        |
| Baseline & Active Contours    | 0.368        | 0.285        |
| Baseline & fine-tuning (Ours) | <b>0.430</b> | <b>0.322</b> |

部分に着目した評価指標である。疾患検出では疾患領域を漏れなく検出することが重要であるため、評価指標として Dice が多く用いられる。本研究では、加えて輪郭の正確さも重視しているため IoU も評価する。

提案手法の実験結果を、従来手法として Wang らの式 (5) の  $\gamma = 1.5$  とした Noise-Robust Loss を用いて学習した結果、fine-tuning を行わないベースラインの結果、ベースラインの検出結果に対して Active Contours を適用した結果と比較する。Active Contours では領域更新によって変化しなくなるまで適用する。提案手法における式 (3) の  $\lambda$  と ACD Loss を適用する際の Morphological GAC に与える輪郭閾値に関してはハイパーパラメータチューニングにより決定する。 $\lambda$  に関しては、0.7 から 0.3 までの範囲で 0.1 ずつ設定し、エポック数を減らした 100 エポックだけの学習を行い、テストデータセットにおける高い値を選択する。輪郭閾値に関しては、ランダムにサンプリングした CT 画像 10 件に対して疑似的に作成した領域アノテーションを初期値として Morphological GAC を適用し、疾患領域が正しく検出されているかを目視で判断し、より多くの CT 画像において検出されている閾値であった値を設定した。以上のハイパーパラメータチューニングの結果、提案手法における式 (3) の  $\lambda$  は 0.4 とし、ACD Loss を適用する際の Morphological GAC に与える輪郭閾値は 0.8 とする。

実験において、3D UNet モデルの学習およびテストを Project MONAI\* フレームワークを用いて行う。Project MONAI は DCNN を中心とした機械学習用オープンソースフレームワーク PyTorch\*\* 上で動く、医用画像に特化したライブラリを含むフレームワークである。実験は、NVIDIA TITAN RTX および NVIDIA GeForce RTX 3090 上で行う。

## 4.2 実験結果

実験結果における各種評価指標の結果を表 2 に示す。Dice では Noise-Robust Loss を用いた結果において 0.322、ベースラインの結果において 0.376、ベースラインの検出結果に Active Contours を適用した結果において 0.368、ACD Loss を用いて fine-tuning を行った提案手法の結果において 0.430 と、提案手法による結果が最も高い検出精度が得られた。Dice はベースライン比で約 14% 向上した。IoU でも、Noise-Robust Loss を用いた結果において 0.224、ベースラインの結果において 0.265、ベースラインの検出結果に Active Contours を適用した結果において 0.285、ACD Loss を用いて fine-tuning を行った提案手法の結果において 0.322 と、提案手法による結果が最も高い検出精度が得られた。IoU はベースライン比で約 22% 向上した。テストデータに対する疾患領域検出の結果の一部を図 3 に示す。入力した CT 画像のパッチのスライスを図 3(a)、図 3(a) に対応する Ground Truth を図 3(b)、疑似的に作成した輪

郭誤差を含む領域アノテーションを図 3(c)、従来手法である Wang らによる Noise-Robust Loss を用いて学習したモデルによる検出結果を図 3(d)、fine-tuning を行っていないベースラインによる検出結果を図 3(e)、ベースラインによる検出結果に Active Contours を適用した結果を図 3(f)、ACD Loss を用いた fine-tuning を適用した提案手法による検出結果を図 3(g) に示す。従来手法では上側の疾患が正しく検出できていないことがわかる。疑似的に作成した領域アノテーションとベースラインによる検出結果を比べるとベースラインは学習した疑似的に作成した領域アノテーションに近い領域を検出できており、輪郭誤差を含めて Ground Truth の疾患領域を検出できている。ベースラインによる検出結果に Active Contours を適用した結果では臓器の輪郭部分が誤検出され、疾患が正しく検出できていないことがわかる。また、提案手法による検出結果では Ground Truth としての領域アノテーションに近い領域が検出されており、ベースラインによる検出結果に含まれる輪郭誤差が補正されていることがわかる。

## 5. 考察

実験結果について考察する。Noise-Robust Loss を用いた場合において、Dice, IoU は最も低い結果となり、Dice Loss を用いたベースラインより低い検出精度であった。したがって、全ての領域アノテーションに輪郭誤差を含む場合において、Noise-Robust Loss によって精度を改善することが困難であることがわかった。また、ベースラインや Noise-Robust Loss を用いた結果では学習データである疑似的に作成した領域アノテーションにおける Dice の 0.429 や IoU の 0.289 より低い。一方で、ACD Loss を用いて fine-tuning を行った提案手法では Dice, IoU は学習データより高い結果が得られた。したがって、提案手法では全ての領域アノテーションに輪郭誤差を含む場合において、学習データ以上の検出精度によって疾患検出を行うことが可能であることが示唆された。

ベースラインによる検出結果に対して Active Contours を適用した結果では Dice, IoU 共に低下している。したがって、検出結果に対して Active Contours によって輪郭誤差を補正することは困難であることが示された。Active Contours では不鮮明な疾患領域輪郭のために輪郭誤差が補正されなかったと考えられる。ベースラインの検出結果に対して Active Contours を適用した結果より提案手法により fine-tuning を行った結果の方が Dice, IoU 共に高い結果が得られた。したがって、ACD Loss を用いて fine-tuning を行う提案手法は疾患領域の不鮮明な輪郭に対応する効果があることが示された。

肝腫瘍を含む CT 画像 16 件中、提案手法ではベースラインより 11 件の検出精度が向上した。したがって、テストに用いた CT 画像の約 68.8% において不鮮明な疾患領域輪郭を検出することができた。また、図 3 に着目すると、従来手法の Wang らによる手法の結果では Dice が 0.334, IoU が 0.200、ベースラインの結果では Dice が 0.302, IoU が 0.178、ベースラインの結果を初期値として Active Contours を適用した結果では 0.105, IoU が 0.056、提案手法の結果では Dice が 0.558, IoU が 0.367 であった。したがって、他の手法と比較すると、提案手法では Dice, IoU 共に向上していることから、提案手法では不鮮明な輪郭に対応する性能が高い。

本提案手法では、ベースラインのモデルに疾患の輪郭情報を事前知識として使い fine-tuning している。疾患自体を検出す

\* <https://monai.io/>

\*\* <https://pytorch.org/>

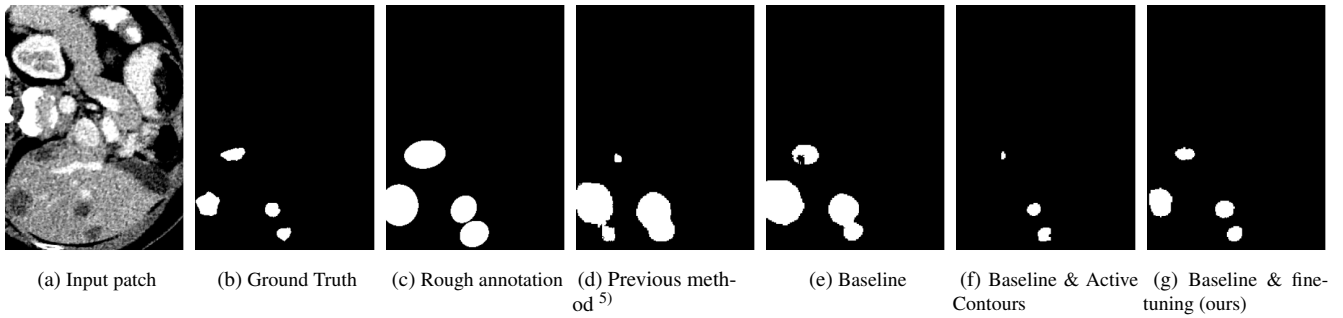


Fig. 3 Results of prediction

る部分はベースラインと共通であるため、ベースラインにおいて検出が困難である画像では提案手法においても精度が低いと考えられる。実験結果においてベースライン、提案手法共に疾患を検出できなかった例を図4に示す。図4(a)における図4(b)の白で示した部分に疾患領域が存在している。しかし、図4(c)、図4(d)をみると、いずれの領域も検出されていないことがわかる。このようにベースライン、提案手法いずれにおいても疾患が検出されないような疾患の検出精度を向上するためには、ベースラインの検出精度を向上する必要がある。本実験では、評価のためLiTSデータセットのトレーニングデータとして公開されているものの一部を学習に使わず、テストデータとした。Liら<sup>16)</sup>は本実験と同じLiTSデータセットにおいてより多くの学習データセットを用いることで精度が向上することを示した。したがって、本研究においても学習データの数を増やすことでベースラインの精度が向上することが期待できる。

また、図3からもわかるように、提案手法では他の手法より相対的に精度が向上しているが、Ground Truthの疾患領域を厳密には検出できていない領域がある。要因として学習データの輪郭誤差が大きいことが考えられる。しかし、実験において提案手法によって学習データの領域アノテーションの精度を向上させることができると示唆された。したがって、提案手法によって学習したモデルを学習データに適用して、更新した学習データを用いて改めて提案手法を適用し、逐次的に学習データを更新していくことによってさらに精度を向上することができると考えられる。

## 6. 結 論

本研究では輪郭誤差を含む領域アノテーションを用いた場合における、疾患領域検出精度を改善した。輪郭誤差を含む領域アノテーションに対応するために、学習において検出領域の輪郭誤差の補正を行う損失関数であるACD Lossを設計し、疾患領域の不鮮明な輪郭に対応するためにACD Lossと通常の損失関数を併用するfine-tuningを行う学習手法を提案した。また、擬似的に作成した輪郭誤差を含む簡易的な領域アノテーションを用いて実験を行い、提案手法により領域アノテーションの輪郭誤差に対応する従来手法やベースラインのモデルより、高い精度で領域検出を行えることを示した。また、提案手法によって輪郭誤差を含む領域アノテーションの学習データ以上の精度による疾患検出が可能であることが示唆された。

輪郭誤差を含む領域アノテーションによる自動疾患領域検出の精度が改善されたことで、より簡易的に作成した領域アノテーションを使用することが可能と考えられる。新たな疾患を含むCT画像に対して簡易的に領域アノテーションを作成する

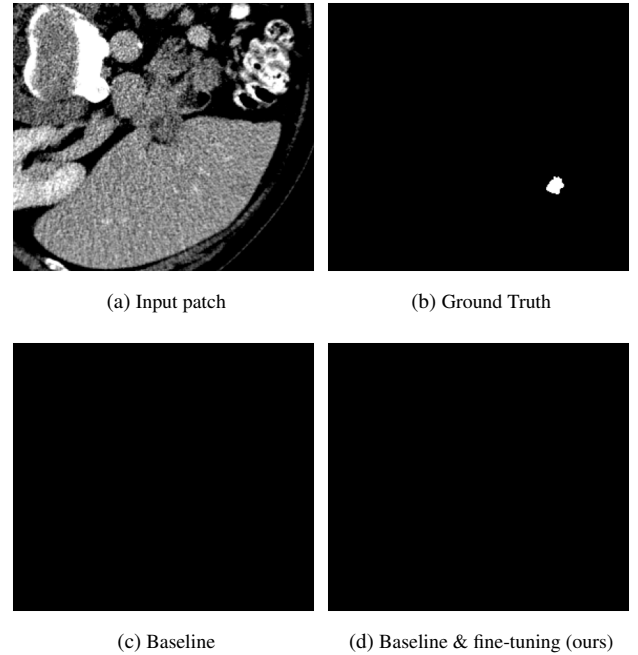


Fig. 4 Example of error cases

ことによって、より多くの学習データが作成されることが期待できる。

## 参 考 文 献

- 1) P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C. Fu, X. Han, P. Heng, J. Hesser, S. Kadoury, T. Konopczynski, M. Le, C. Li, X. Li, J. Lipkova, J. Lowengrub, H. Meine, J. H. Moltz, C. Pal, M. Piraud, X. Qi, J. Qi, M. Rempfler, K. Roth, A. Schenk, A. Sekuboyina, P. Zhou, C. Hülsemeyer, M. Beetz, F. Ettliger, F. Gruen, G. Kaissis, F. Lohöfer, R. Braren, J. Holch, F. Hofmann, W. Sommer, V. Heinemann, C. Jacobs, G. E. H. Mamani, B. van Ginneken, G. Chartrand, A. Tang, M. Drozdal, A. Ben-Cohen, E. Klang, M. M. Amitai, E. Konen, H. Greenspan, J. Moreau, A. Hostettler, L. Soler, R. Vivanti, A. Szeskin, N. Lev-Cohain, J. Sosna, L. Joskowicz and B. H. Menze: The Liver Tumor Segmentation Benchmark (LiTS), arXiv:1901.04056, (2019).
- 2) E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe and J. Verweij: New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1), *European Journal of Cancer*, **45**, 2, (2008) 228.
- 3) 後藤 拓矢, 中野 宏毅, 山下 淳, 浅間 一: CT画像における複数スライスを用いた小疾患領域検出, *精密工学会誌*, **86**, 12, (2020) 1020.
- 4) L. Zhang, R. Tanno, M. Xu, C. Jin, J. Jacob, O. Ciccarelli, F. Barkhof and D. C. Alexander: Disentangling Human Error from the Ground Truth in Segmentation of Medical Images, *Proceedings of the Advances in Neural Information Processing Systems*, (2020) 15750.
- 5) G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, S. Zhang: A Noise-robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions from CT Images, *IEEE*



- Transactions on Medical Imaging, **39**, 8, (2020) 2653.
- 6) 清水 昭伸, 松坂 匡芳, 長谷川 純一, 鳥脇 純一郎, 鈴木 隆一郎: 動的輪郭モデルを用いた輪郭線抽出手順の自動構成と胸部 X 線像上の肺輪郭線抽出への応用, コンピュータ支援画像診断学会論文誌, **1**, 1, (1997) 1.
  - 7) S. Dai, K. Lu, J. Dong, Y. Zhang and Y. Chen: A Novel Approach of Lung Segmentation on Chest CT Images using Graph Cuts, *Neurocomputing*, **168**, (2015) 799.
  - 8) M. Rajchl, M. C. H. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M. A. Rutherford, J. V. Hajnal, B. Kainz and D. Rueckert: DeepCut: Object Segmentation from Bounding Box Annotations using Convolutional Neural Networks, *IEEE Transactions on Medical Imaging*, **36**, 2, (2017) 674.
  - 9) O. Ronneberger, P. Fischer, T. Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation, In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, (2015) 234.
  - 10) 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, (2016) 424.
  - 11) E. Kerfoot, J. Clough, I. Oksuz, J. Lee, A. P. King and J. A. Schnabel: Left-Ventricle Quantification using Residual U-Net, In Proceedings of the International Workshop on Statistical Atlases and Computational Models of the Heart, (2018) 371.
  - 12) D. Acuna, A. Kar and S. Fidler: Devil is in the Edges: Learning Semantic Boundaries from Noisy Annotations, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2019) 11067.
  - 13) P. Marquez-Neila, L. Baumela and L. Alvarez: A Morphological Approach to Curvature-Based Evolution of Curves and Surfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**, 1, (2014) 2.
  - 14) T. Zhao and Z. Yin: Weakly Supervised Cell Segmentation by Point Annotation, *IEEE Transactions on Medical Imaging*, **40**, 10, (2021) 2736.
  - 15) F. Milletari, N. Navab, S.-A. Ahmadi: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, arXiv:1606.04797, (2016).
  - 16) X. Li, H. Chen, X. Qi, Q. Dou, C. Fu and P. Heng: H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes, *IEEE Transactions on Medical Imaging*, **37**, 12, (2018) 2663.