

簡易アノテーションを用いた CT 画像からの

輪郭強調による疾病領域検出

○後藤 拓矢[‡], 中野 宏毅[‡], 山下 淳[†], 浅間 一[†]

†: 東京大学大学院工学系研究科精密工学専攻

‡: 日本 IBM

gototakuya@robot.t.u-tokyo.ac.jp

概要: 医用画像から疾病領域を検出するモデルの学習には領域アノテーションが正しく疾病を示し品質が高いことが重要である。しかし、医用画像における疾病領域の輪郭は不鮮明であり、様々な形をしているため高品質な領域アノテーションを作成することは困難である。本研究では領域アノテーション作成を簡易化するために誤りを含む品質の低い領域アノテーションを用いた場合における、疾病領域検出手法を提案する。提案手法では CT 画像から 3D Unet を用いて誤りを許容して検出した 1 次推定領域と輪郭強調加工した CT 画像から事後処理によって推定領域を精緻化する。実験により 1 次推定領域を精緻化し、高品質な領域アノテーションに近づけることが可能であることを示した。

<キーワード> 簡易アノテーション, 疾病領域検出, 3D UNet

1. はじめに

近年, Deep Convolutional Neural Network (DCNN) を用いた手法により医用画像から疾病領域を検出する精度は大きく向上している[1].

精度が高い 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病領域を正確に示していることが重要である[2]. 近年では, 機械学習のための学習データとして様々な臓器の領域アノテーションを含むデータを公開している事例が増えている[3]. しかし, 疾病領域のアノテーションがついているデータは少なく, 疾病領域を検出するためには個別にデータを作成する必要がある. 図 1 に 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を示す. 図 1(a)のように CT 画像における疾病領域の輪郭は臓器や骨などの領域に比べて不鮮明であり, 複雑な形状の疾病領域に対して領域アノテーションの作成は困難である. また, 経験の浅い医師の場合, 作成したアノテーションと実際の疾病領域との誤差は大きい[2]. 通常, より正確な疾病領域のアノテーションを作成するために, 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, 正しい領域アノテーションに関して議論と修正を繰り返し, 同意のもと最終的な領域アノテーションが作成される[2]. ここで, 複数の熟練の医師によって作成され, 医師間による同意プロセスを経て修正された教師データを高品質データ, 高品質データに対して領域アノテ

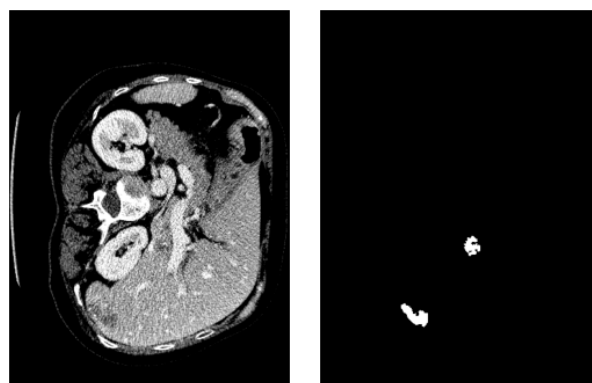


図 1 (a)CT 画像のスライスと(b)熟練の医師が作成した領域アノテーション

ーションと実際の領域との誤差が大きいものを低品質データと呼ぶ. 高品質データを作成するためには非常に多くの時間と労働力を要するという課題がある[2]. したがって, 本研究では領域アノテーション作成の簡易化のため, 誤りを含む低品質な領域アノテーションを用いた場合における, 疾病領域検出手法の提案を目的とする.

低品質データを用いた疾病領域検出モデルの学習手法として, Wang らはノイズロバスト損失関数を導入し, 領域アノテーションの輪郭の誤差に対してロバストな学習手法を提案した[2]. Wang らは簡易領域アノテーションとして, 熟練の医師によって作成された複数の医師間による同意プロセスを省略した領域アノテーションを対象としている.

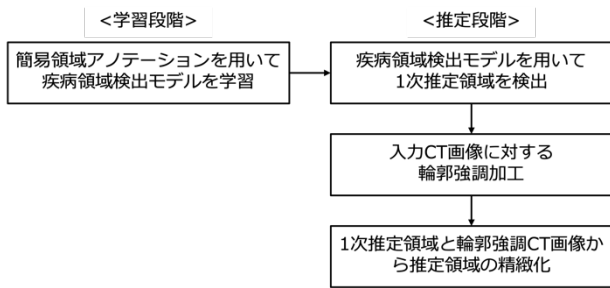


図2 提案手法の概要

しかし、領域アノテーションの作成は依然として困難である。そこで、本研究では医師によって領域アノテーションの作成自体を簡易化するために、CT 画像のスライスごとに疾病領域に近い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、肝臓の腫瘍を対象としており、スライスでは肝腫瘍は楕円に近いことから楕円によって作成された領域アノテーションを対象とする。また、経験の浅い医師によって作成された領域アノテーションも対象とするために、領域アノテーションと疾病領域には大きな誤差があると仮定した。提案手法では簡易領域アノテーションを学習した 3D UNet[3]によって推定領域を検出し、事後処理によって、医用画像の疾病領域の輪郭にフィッティングし精緻化を行う。本研究では、一般に公開されているデータ[3]を用いた肝腫瘍の領域検出を対象とする。

2. 提案手法

2.1. 提案手法の手順

提案手法の概要を図2に示す。まず、誤差を含む領域アノテーションを教師データとして 3D UNet[4]で領域検出モデルを学習する。推定時は、学習した領域検出モデルを用いて CT 画像から疾病領域と思われる領域を検出する。検出された領域を 1 次推定領域とする。次に、入力した CT 画像に対して、輪郭強調加工を施し、輪郭を鮮明化する。最後に、1 次推定領域を初期値とし、輪郭強調加工した CT 画像の輪郭にフィッティングすることによって 1 次推定領域の精緻化を行い、最終的な検出領域とする。

Wang らは誤りを含む領域アノテーションを用いて学習する際、領域アノテーションの中には正しい輪郭であるものも含まれているという前提のもと、輪郭の誤差にロバストな学習手法を提案している。しかし、様々な形状を持つ疾病領域に対して楕円という単一の形状のアノテーションを作

成した場合、輪郭の誤差が大きくロバストな学習では対応できないと考えられる。したがって、1 次推定領域には実際の疾病領域との誤差が含まれていると考えられるため、事後処理による推定領域の精緻化が有効であると考えられる。図 1(a)に示したように医用画像における疾病領域の輪郭は不鮮明である一方で、臓器や骨などの輪郭は疾病領域に比べて鮮明である。したがって、CT 画像のまま 1 次推定領域を輪郭に対してフィッティングを行うと、臓器や骨の領域に強く作用すると考えられる。特定の疾病領域を検出する場合、対象とする臓器は一種であり、対象とする臓器外に疾病領域を検出した場合はその検出領域は誤りである。そこで、提案手法では対象臓器内の輪郭の強調加工を施すことで、1 次推定領域を精緻化する際に疾病領域として検出される範囲を対象臓器内に制限する。

2.2. 疾病領域検出モデルと学習手法

疾病領域検出のための 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。大まかなネットワークの構造は 2 次元データの医用画像に多く用いられる UNet[5]と同様である。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。

本研究では、3D UNet の構成に Project MONAI[6]において公開されているフレームワークを用いる。Project MONAI は DCNN を中心とした機械学習用フレームワーク PyTorch[7]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。Project Monai で構成した 3D UNet にはエンコーダ部、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 5 つのブロックを使用した 3D UNet モデルを構成する。

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

表 1 肝臓領域検出モデルの精度

Dice	Precision	Recall	IoU
0.950	0.950	0.953	0.91

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. また, 各パッチに対してランダムに 3 次元の移動, 回転, 反転, ガウシアンフィルタによるノイズ, スムージングを適用する. 損失関数として Dice Loss[8]を用いて, 500 エポック学習を行う.

テストの際には 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.

本研究は, 3D UNet モデルの学習およびテストを NVIDIA TITAN RTX 上で行う.

2.3. 輪郭強調加工

輪郭強調加工について述べる. 輪郭強調処理にはソベルフィルタなども用いられるが, 画像中のノイズの影響を強く受けるという欠点がある. 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デリバティブガウシアンフィルタを用いる. また, フィルタを適用後, 0 から 1 の範囲に正規化し, 逆数をとる. したがって, 鮮明な輪郭ほど 0 に近く, 平坦な領域は 1 に近くなる.

1 次推定領域は対象とする臓器外の領域を含むことも考えられる. そこで本研究では, フィッティングされる輪郭を対象とする臓器領域内に限定するために, フィルタによる輪郭強調処理後の画像に対して臓器領域外をマスクする. 臓器領域を得るためには別途学習した臓器領域検出モデルを用いる. 臓器領域検出モデルの学習のために, 臓器領域アノテーションが必要である. しかし, 臓器領域データは近年多くの公開データがあり, 学習済みのモデルも多く公開されている[3]. また, 疾病領域に比べて臓器領域の輪郭は鮮明であり, 個人によって大きな差はないため, 臓器の位置を知るものであれば容易にデータの作成ができる. 領域検出のモデルも少量のデータで高い精度のモデルを学習できることから, 本研究の課題として高品質なアノテーションを得ることが困難であるデータの対象外とする. 今回, 提案手法では Liver Tumor Segmentation (LiTS) データセット[3]に含まれる肝臓の領域アノテーションによって学

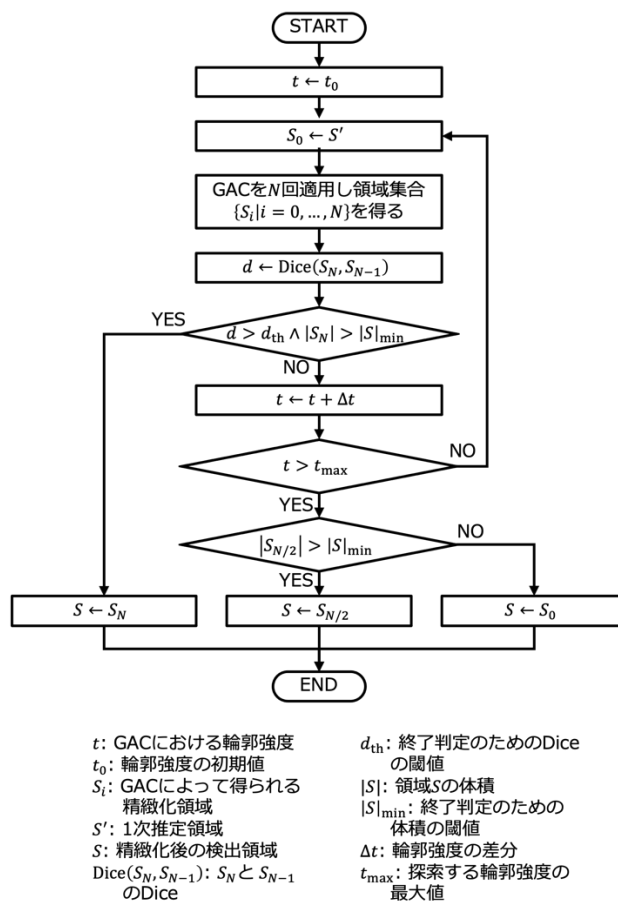


図 3 精緻化処理の手順

習した 3D UNet による肝臓領域検出モデルを用いる. 学習した 3D UNet は 2.1 節で述べた疾病領域検出モデルと同じものを用いる. 学習データとして実験におけるトレーニングデータセットと同じものを用いる. 肝臓領域検出モデルの精度を表 1 に示す. Dice, Precision, Recall いずれも高く, 非常に高精度で肝臓領域の検出が行うことができる. マスク処理ではフィルタ処理前の CT 画像に対して検出した肝臓領域以外の画素値を 0 とする.

2.4. 一次推定領域からの精緻化

一次推定領域からの精緻化手法について述べる. 精緻化には一次推定領域を初期値とした輪郭強調画像に対する Geodesic Active Contours (GAC) [9]を用いる. GAC も他の動的輪郭手法と比べて画像中のノイズの影響を受けにくいという特徴がある.

GAC にはハイパーパラメータとして輪郭のみならず輪郭強度を指定する必要がある. 疾病領域の画素値は個々の疾病によって異なるため, 疾病領域の輪郭強度も一定ではない. したがって, 本研究では CT 画像と 1 次推定領域の 1 セットごとに適当な輪郭強度を探索する処理を行った. 図 3 に探索の手順を示す. まず輪郭強度 t の初期値として

表 2 低品質データにおける領域アノテーションの精度

Dice	Precision	Recall	IoU
0.374	0.233	0.944	0.233

t_0 を設定する. 次に 1 次推定領域 S' を領域の初期値として GAC を N 回適用し精緻化した領域の集合 $\{S_i\}$ を得る. 次に, 最終的に得られた領域 S_N とその1 回前の精緻化領域 S_{N-1} の Dice を計算する. 次に, GAC によっていずれかの輪郭にフィットし精緻化が収束しているかを判定する. ここでは前の手順で計算した Dice が閾値 d_{th} より大きく, また領域の体積が閾値 $|S|_{min}$ より大きいかを基準とする. Dice による判定は GAC の処理が収束しているかを判定し, 領域体積による判定は GAC によって輪郭が検出されず推定領域が消失してしまっているかを判定する. ここで, 条件を満たす場合, GAC によって得られた S_N を最終的な検出領域とする.

医療画像の場合, 見落としを防ぐことが重要である. 検出された一次推定領域には何らかの疑わしい領域が含まれていると考えられる. したがって, 指定した輪郭強度によって輪郭が検出されなかった場合は, 輪郭とみなす輪郭強度を大きくし再度 GAC によって一次推定領域を精緻化する. ここで, Δt は探索する輪郭強度の差分を示す. 探索する輪郭強度の最大値 t_{max} まで GAC によって条件を満たす検出領域が得られなかった場合, 最後に行った GAC によって得られた精緻化領域の中間結果を最終的な検出領域とする. ただし, 中間結果の推定領域の体積が閾値 $|S|_{min}$ 以下だった場合には, 1 次推定領域を最終的な検出領域とする. 本研究では, t_0 を 0.5, t_{max} を 0.8, Δt を 0.1, d_{th} を 0.95, $|S|_{min}$ を 1000mm^3 , N を 10 回とする.

3. 実験

3.1. 実験手法

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[3]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高品質のものである. 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簡易的に作成された低品質な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

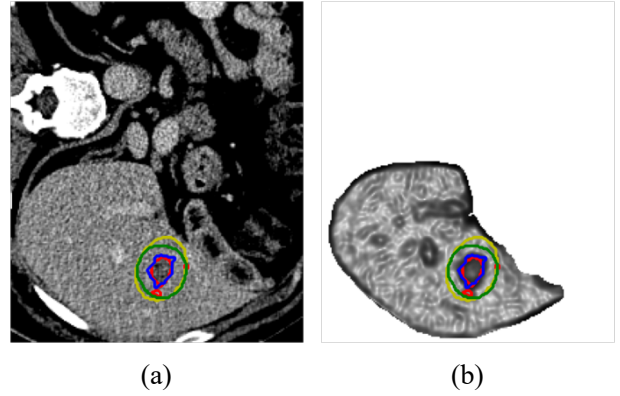


図 4 (a)CT 画像のスライス, (b)輪郭強調加工後の CT 画像のスライス (赤:精緻化後の検出領域, 緑: 1 次推定領域, 青: 高品質な領域アノテーション, 黄: 擬似的に作成した低品質な領域アノテーション)

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%, 楕円軸角度に対して $\pm 0.1\text{rad}$, 長軸と短軸に対して 1.8~2.2 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. 擬似的に作成した低品質な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高品質な領域アノテーションを Ground Truth として評価した結果を表 2 に示す. Recall は高いが Precision および Dice は低く, 実際に低品質なデータであるといえる.

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.

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, Precision, Recall, Intersection over Union (IoU) を CT 画像ごとに平均したものをを用いる.

提案手法の実験結果を, 高品質データを学習し, 精緻化処理を行わない場合の結果, 低品質データを学習し, 精緻化処理を行わない場合の結果と比較する. ここで, 高品質データとは変更を加えていない 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である.

3.2. 実験結果

実験結果の一部を図 4 に示す. ある CT 画像に

表 3 実験結果における評価結果

学習データ種別 ／精緻化处理	Dice	Precision	Recall	IoU
高品質データ ／なし	0.524	0.625	0.473	0.420
低品質データ ／なし	0.244	0.163	0.571	0.153
低品質データ ／あり (提案手法)	0.377	0.406	0.403	0.275

対して 1 次推定領域, 最終的な検出領域, Ground Truth とした LiTS データにおける高品質な肝腫瘍の領域アノテーション, 擬似的に作成した低品質な肝腫瘍の領域アノテーションを付加した 1 スライスを図 4(a)に示す. また, 同様の領域を輪郭強調加工後の画像に付加したものを図 4(b)に示す. 図 4 より黄色で示した擬似的に作成した低品質な領域アノテーションと緑色で示した 1 次推定領域を比べると一次推定領域はおおよそ低品質な領域アノテーション領域を検出できているといえる. また, 赤色で示した精緻化後の検出領域は緑色で示した 1 次推定領域から輪郭強調加工を施した CT 画像の輪郭部に近づき, 青色で示した高品質な領域アノテーションに近い領域に精緻化されていることがわかる.

各種評価指標の結果を表 3 に示す. 学習データに高品質データを用いて, 精緻化处理を行っていない手法は従来の領域検出手法を示している. したがって, 高品質データを用いた場合の評価結果は, 誤りを含む低品質な領域アノテーションを用いて学習する本研究が目指すべき精度と考えられる. 低品質データを用いて学習し, 精緻化处理を行っていない評価結果は 1 次推定領域の精度を示している. Precision は 1 次推定領域において 0.163 で, 提案手法において 0.406 であった. 1 次推定領域の Precision より, 提案手法の Precision の方が高いことから, 提案手法によって検出領域の精緻化ができているといえる. また, 1 次推定領域に対して, 提案手法における Dice は約 1.5 倍, IoU は約 1.8 倍である. したがって, 疾病領域は単純に低品質な領域アノテーションを用いて学習した場合より高い精度で検出できることが示された.

一方で, Recall は 1 次推定領域における結果の方が, 提案手法における結果より高い. 提案手法において 1 次推定領域のいずれかの領域は最終的な

表 4 臓器マスクの有無, 輪郭強度探索の有無による評価結果

臓器マスク ／輪郭強度探索	Dice	Precision	Recall	IoU
なし／なし	0.294	0.260	0.497	0.194
あり／なし	0.309	0.310	0.443	0.207
なし／あり	0.332	0.281	0.495	0.232
あり／あり (提案手法)	0.377	0.406	0.403	0.275

検出領域として含まれているため, 完全に見落とされているわけではない. したがって, 部分的に過剰に輪郭に対してフィッティングし, 真の疾病領域が検出領域から漏れてしまったものと考えられる. 実際に Recall が一次推定領域と比べて低下した CT 画像を見ると LiTS データセットにおける肝腫瘍領域のアノテーションがついた領域は目では非常に見えにくいものであった. 輪郭強調処理の際にコントラストを調整するような改善が必要である.

また, 1 次推定領域と比べた場合本研究における目標精度である高品質データを用いた結果と比較すると, 精緻化前と比べると精度の面での改善は見られたものの, いずれの指標においても未だ乖離がある. とりわけ, Precision が提案手法の約 1.5 倍と差が大きい. 精緻化による改善の余地があるといえる.

4. 考察

提案手法の輪郭強調加工における臓器領域マスクと精緻化处理における輪郭強度の探索がどの程度の精度向上に効果があるのかを検証した. 実験データ及び評価手法は 3 章と同様の設定で行った. また, 輪郭強度探索を行わない場合の GAC で指定する輪郭強度は 0.8 とした.

表 4 に評価結果を示す. 輪郭強度探索を行わない場合のうち, 臓器マスクがある場合と比べて臓器マスクがない場合の Recall が 0.054 高い. このことから精緻化处理の際に臓器の表面付近の疾病領域が過剰にフィッティングしていると考えられる. しかし, 臓器マスクを用いる場合の方が, 輪郭強度探索を行う場合より Precision の向上が大きいことから, Precision の向上には臓器マスクによる効果が大きいといえる. また, 領域検出の精度として多く用いられる Dice と IoU は臓器マスクを用いる場合より輪郭強度探索を行う場合の方が高い. したがって, 領域検出の精度向上のためには Recall

の低下を抑えて Precision を向上させる方が効果は高いと考えられる。さらに、臓器マスクと輪郭強度探索のどちらか一方のみ用いる場合より、両方用いた提案手法の場合の Dice と IoU が最も高い。以上より、臓器マスクと輪郭強度探索をあわせて適用する提案手法が精度向上において有効であると示された。

5. おわりに

本研究では領域アノテーション作成の簡易化のため、誤りを含む低品質な領域アノテーションを用いた場合における、疾病領域検出手法を提案した。提案手法では、輪郭強調加工において臓器領域以外をマスクすることで臓器領域内の輪郭を強調し、精緻化において CT 画像ごとに疾病領域の輪郭として適した輪郭強度を探索した。また、擬似的に作成した低品質な領域アノテーションを用いて実験を行い、提案手法により低品質な領域アノテーションを学習した領域検出モデルによる領域検出より、高い精度で領域検出を行えることを示した。

高品質な領域アノテーションを用いて学習した領域検出モデルによる領域検出と比較すると、提案手法による領域検出の精度に乖離があった。今後、手法の改善を行い、さらなる精度向上を行う。また、今回、低品質な領域アノテーションを高品質な領域アノテーションをもとに擬似的に作成した。想定した低品質な領域アノテーションの作成がどれほどアノテーションの作成を簡易化できるかという定量的な評価を行うことも検討していく。

参考文献

- [1] 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病領域検出. 精密工学会誌, Vol. 86, No. 12, pp. 1020-1025, 2020.
- [2] 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, Vol. 39, No. 8, pp. 2653-2663, 2020.
- [3] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, P. Bilic, P. F. Christ, R. K. G. Do, M. Gollub, J. Golia-Pernicka, S. H. Heckers, W. R. Jarnagin, M. K. McHugo, S. Napel, E. Vorontsov, L. Maier-Hein, M. J. Cardoso: A Large Annotated Medical Image Dataset for the Development and Evaluation of Segmentation Algorithms, arXiv:1902.09063, 2019.
- [4] 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, Vol. 9901, 2016.
- [5] 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, pp. 234-241, 2015.
- [6] Project MONAI, <https://monai.io/>, Accessed on Oct. 21, 2021.
- [7] PyTorch, <https://pytorch.org/>, Accessed on Oct. 21, 2021.
- [8] F. Milletari, N. Navab, S.-A. Ahmadi: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, arXiv:1606.04797, 2016.
- [9] V. Caselles, R. Kimmel, G. Sapiro: Geodesic Active Contours, Vol. 22, No. 1, pp. 61-79, 1997.

後藤 拓矢: 2017 年 3 月 東京大学大学院新領域創成科学研究科人間環境学専攻修士課程卒業。同年 4 月 日本 IBM 入社。2019 年 4 月 東京大学大学院工学系研究科精密工学専攻博士課程入学。現在、業務の傍ら、医用画像を中心とした画像からの領域検出、画像処理に関する研究に従事。

中野 宏毅: 1983 年 3 月 大阪大学基礎工学部卒業。同年 4 月 日本 IBM 入社。以来、画像解析・機械学習に関する研究開発に従事。著書: ウェーブレットによる信号処理と画像処理他。学術博士。

山下 淳: 2001 年 3 月 東京大学大学院工学系研究科精密機械工学専攻博士課程修了, 博士(工学)。静岡大学助手, 助教, 准教授を経て, 2011 年 10 月 東京大学大学院工学系研究科精密工学専攻准教授, 現在に至る。知能ロボット, コンピュータビジョン, 画像処理の研究に従事。

浅間 一: 1984 年 東京大学大学院工学系研究科博士課程修了。1986 年 理化学研究所研究員補。同副主任研究員等を経て, 2002 年 東京大学人工物工学研究センター教授。2009 年 同大学院工学系研究科教授。IEEE フェロー, 日本ロボット学会フェロー, 日本機械学会フェロー。