

学 位 論 文

深層学習を用いた胎児超音波動画における
胎児胸壁に対するセグメンテーションを行うための
Model-Agnosticな新手法の確立

富山大学大学院

医学薬学教育部（博士課程） 東西統合医学専攻

産科婦人科学講座

氏 名 生水 貫人

要旨

近年、人工知能(artificial intelligence : AI)はその技術革新とともに医学分野を含め幅広い分野で目覚ましい成果を上げている。強力な機械学習の一つである深層学習を用いて、ピクセル単位でラベルを付与できる画像セグメンテーション手法を医用画像に応用することは、AIを用いた診断支援モデルを構築するための有効な戦略となっている。胎児超音波スクリーニングにおいてもAIを用いた診断支援モデルにより、先天性疾患の検出率の向上が期待できると考える。しかしながら現状として、超音波画像は対象物と背景のコントラストが低く曖昧であるため、解剖学的構造を正確にセグメンテーションすることは依然として難しい課題となっている。また、胎児の超音波診断では、先天性疾患の検出や新生児の予後の評価に重要な要素である心臓と肺を含む胸部の評価、特に四腔断面 (four-chamber-view : 4CV) による評価が不可欠である。そして胸壁は4CVを評価する上で重要な構造であり、検査者は胸壁を確認することで胸郭内の構造物の相対的な向きや大きさを認識することができる。そのため胸壁のセグメンテーション性能を向上させることは診断支援モデルを開発する上で重要と考える。本研究では胎児超音波動画における胸壁のセグメンテーション性能を向上させるために、深層学習技術を用いたmodel-agnosticな(モデルに依存しない)手法であるMulti-Frame + Cylinder法(MFCY)を新しく提案した。Multi-Frame法(MF)は超音波動画の時系列情報を利用し、Cylinder法(CY)は胸壁の形状の特性を利用した手法である。今回の提案手法であるMFCYを適用した際の胎児超音波動画における胸壁セグメンテーション性能について、臨床データを利用して検討した。

MFは、対象となる画像とその前後フレームに対して各々胸壁のセグメンテーションを行い、それらの予測ラベルを統合することにより予測結果を得る。これは高フレームレートの超音波動画を構成する連続した時系列画像の中で、前後するフレーム間の類似性を利用することにより予測結果を補完することが期待できる。また、CYは、胸壁に加え、胸腔と胸郭全体の3つの予測ラベルを独立した学習済みモデルから取得しこれらを統合することにより予測結果を得る。胸壁は常に円筒形であるという事前知識を利用し、統合時には胸郭全体のラベルから胸腔ラベルを差し引きし、胸壁ラベルを足し合わせるにより、お互いの情報を補完することが期待できる。そして、MFとCYを組み合わせた手法がMFCYとなる。本研究では提案手法の性能評価を臨床データで行うために、2018年4月から2019年5月の間に、昭和大学の4病院で妊娠中期に胎児超音波スクリーニングを受けた正常胎児を妊娠する妊婦を対象とし、超音波動画を取得した。結果、妊娠18週から28週(中央値:20週)の妊婦256例から538枚の4CV画像を得ることができた。この4CV画像に対して胸壁のアノテーションラベルを付与し、これを利用してMFCYに必要な教師あり学習を行った。その際には、畳み込みニューラルネットワーク(convolutional neural network : CNN)としてU-netとDeepLabv3+を使用した。今回、セグメンテーション性能を評価するために5-foldクロスバリデーションを行った。MFC

Yを使用すると、平均 intersection over union (mIoU) は0.448 (U-net) vs. 0.493 (U-net + MFCY) , 0.417 (DeepLabv3+) vs. 0.470 (DeepLabv3+ + MFCY) であり、既存モデルよりも優れた結果を得ることができた。また、平均再現率でも、0.568 (U-net) vs. 0.738 (U-net+MFCY) , 0.525 (DeepLabv3+) vs. 0.729 (DeepLabv3++MFCY) と、既存モデルよりも高かった。

本研究では、2つのmodel-agnosticな手法を組み合わせることで、胎児超音波動画における新規のセグメンテーション手法であるMFCYを提案した。MFCYは、2つの手法の独立した予測結果のアンサンブル学習に基づいており、超音波動画の時系列情報と胸壁の形状情報を相互に補完することができる。また、臨床データを用いて行った胸壁セグメンテーションにおいて、MFCYは既存モデルよりも優れた性能を示した。MFCYは、既存のCNNの予測結果をそのまま利用し、各動画のターゲット画像とその前後フレームの予測結果を統合することで機能する。データセットが限られた中で、ネットワークに修正を加えることなく、またより多くの追加のアノテーションを行わずとも、セグメンテーション性能を向上させることができた。MFCYはこのようにネットワーク構造に依存しないため、他のネットワークにも広く適用できる。このことは、model-agnosticな手法の大きな利点である。4CVの評価を支援するAIベースのモデルを構築するためには、胎児の胸壁の正確なセグメンテーションが不可欠である。MFCYが、胎児超音波の自動診断支援技術の開発につながることに期待する。

Key word: 深層学習, 胎児超音波検査, 出生前診断, 胸壁セグメンテーション, model-agnostic, アンサンブル学習

1. 背景

近年、人工知能 (artificial intelligence : AI) に基づくモデルは、医療をはじめとする幅広い分野の問題に応用され、目覚ましい成果を上げている[1,2]。特に、深層学習における対象の各画素にラベルを付与する画像処理手法である画像セグメンテーションは[3]、優れた性能を発揮している[4]。セグメンテーションを医用画像に応用することで、臨床的診断のための新たな支援システムを構築できる可能性がある[5,6]。胎児超音波検査においても、画像内の構造物を認識することが診断に重要な要素であることから、セグメンテーションを活用することはAIを用いた診断支援モデルを構築するための有効な戦略となっている[7,8]。

胎児の超音波診断では、先天性疾患の検出や新生児の予後予測に重要な要素である心臓と肺を含む胸部の評価、特に四腔断面 (four-chamber-view : 4CV) による評価が不可欠である[9]. そして胸腔を取り囲み、胸椎、肋骨、胸骨、筋肉などの複数の構成要素からなる胸壁は[10], 4CV を評価する上で重要な構造物である (図 1). 検査者は、胸壁を参照することで、得られた 4CV が適切な断面であることを確認し[11], さらに胸郭内の構造物の相対的な向きや大きさを確認することで、異常所見を検出するための評価が可能となる[12,13]. このため、胸壁のセグメンテーションは、検査者が胸壁の所在や向きを認識するのに役立ち、4CV の評価を支援するための実現可能な AI ベースのモデルを開発することにつながる.

しかしながら現状として、医療で扱われる超音波画像では、対象物と背景のコントラストが低く曖昧であるため、解剖学的構造を正確にセグメンテーションすることは依然として困難な課題となっている[8]. 実際、私が行った心臓、肺、胸壁のセグメンテーションの予備実験では、既存の畳み込みニューラルネットワーク (convolutional neural network : CNN) モデルを使用してセグメンテーションを行ったところ、心臓と肺の結果と比較し、胸壁ではさらに難しいことが示された (付録表 1).

本研究では、胸壁のセグメンテーション性能を向上させるために、深層学習技術を用いた Multi-Frame + Cylinder 法 (MFCY) という新手法を提案した. MFCY は、超音波動画の時系列情報を利用した Multi-Frame 法 (MF) と、胸壁の形状を利用した Cylinder 法 (CY) で構成されている. MF と CY はいずれもオリジナルに考案したものである. MFCY は、任意のニューラルネットワークモデルからアンサンブル学習によって得られた予測結果のみを統合したものである. したがって、MFCY は model-agnostic な手法であり、どのようなニューラルネットワークモデルに対しても、ネットワークの構造を改変することなく適用することができる. 今回、臨床データを用いて、MFCY を実際に使用した際の胎児超音波動画における胸壁セグメンテーション性能について検討した.

1.1. 関連研究

近年、画像のセグメンテーションの主流となっている強力な CNN として U-net と DeepLabv3+ が挙げられる. Ronneberger らは、完全畳み込みネットワークをベースにした U-net を開発し、他の手法と比較してより少ない学習データでより正確なセグメンテーションを実現した[14]. U-net は、バイオメディカル画像のセグメンテーションに特に適しており、U-net に基づくモデルを用いた優れたセグメンテーション性能について示したいくつかの研究報告がなされている[15]. また、Chen らは、ピラミッド型プーリングモジュールとエンコーダ・デコーダモデルを組み合わせる DeepLabv3+ を開発し、都市景観の画像で最先端の性能を実証した[16]. 本研究では、提案手法の性能

を比較するために、U-net と DeepLabv3+を用いてセグメンテーションを行った。以下、U-net と DeepLabv3+を単独で用いたものを既存モデルと呼ぶこととする。

画像のセグメンテーションは、ピクセル単位で対象物の位置や大きさを自動的に認識できるため(図2)、様々な分野の画像解析において重要な役割を果たしている[17,18]。例えば、都市景観のセグメンテーションは、視覚的なシーンの理解に不可欠な技術であり、自動運転技術にも貢献している[19]。また、医用画像解析では、コンピュータ断層撮影、磁気共鳴画像、X線、病理画像、超音波画像などの様々なモダリティにおいて、臓器や腫瘍などのセグメンテーションがコンピュータ支援検出/診断に貢献している[20]。胎児の超音波診断にフォーカスを当てると、セグメンテーション手法を用いたAIベースのモデルは検査者を支援する大きな可能性を秘めている。異なる深層学習ベースの手法を用いて、胎児超音波画像における頭部、心臓、肺、胸郭全体、胎盤、羊水、胎児全体などの胎児構造のセグメンテーションに関する研究が複数報告されている[21-25]。Arnaoutらは、U-netを用いて胸郭全体、心臓、心房、心室、脊椎のセグメンテーションを行い、心胸郭断面積比、心臓軸、心室面積変化率を算出した[22]。Burgos-Artizzuらは、胎児の肺のセグメンテーションを利用して、新生児の呼吸器疾患リスクを推定するソフトウェアの性能を向上させた[23]。本研究では、私の知る限り、初めて胎児超音波における胸壁のセグメンテーションに着目した。

胎児超音波のセグメンテーション性能を向上させるために、動画の時系列情報を採用した研究がいくつかある。Dozenらは、前後フレームの情報を取り込んだキャリブレーションモジュールを構築することにより、CNNで算出した心室中隔のセグメンテーション結果を改善させた[26]。Yuらは、胎児の左心室のセグメンテーションのために、連続したフレームで微調整されたダイナミックCNNを提案した[27]。しかし、MFは、アンサンブル学習の要領で、前後フレームからの胸壁セグメンテーションの出力を統合する点で、前述の手法とは異なる。

アンサンブル学習は、学習済みの複数のモデルから得られた予測結果を組み合わせることで、集約的な性能を向上させる強力な手法である[28]。複数の研究者が、超音波画像のセグメンテーションのために、さまざまなアンサンブル学習に基づく手法を提案している。Kusunoseらは、心臓超音波画像の局所的な壁運動異常を識別するために、10個のCNNでの多数決アンサンブルを用いて確率スコアを計算した[29]。本研究でも、胎児胸壁のセグメンテーション性能を向上させるために、MFとCYの両方でアンサンブル学習を採用している。

1.2. 本研究の貢献内容

この研究のモチベーションは、予測ラベルが不連続になることが多い円筒形の胸壁のセグメンテーション性能を向上させることにある。そのため、胸壁セグメンテーションに

関連する複数の予測ラベルを用いてアンサンブル学習を行った。予測ラベルには、超音波動画は前後フレームが非常に似ている一連の画像であるという性質と、胸壁は常に円筒形であるという性質の2つを利用した。これらの性質を利用して得られた予測ラベルは、アンサンブル学習によって追加情報が与えられ、胸壁のセグメンテーション性能を向上させることが期待できる。

今回提案する MF, CY, MFCY の貢献内容は以下の通りである。

MF は、対象となる画像とその前後フレームに対応する予測ラベルを、1つの予測ラベルに統合する。これは高フレームレートの超音波動画を構成する連続した時系列画像の中で、前後するフレーム間の類似性を利用している。隣り合う超音波画像の間には、わずかな断面の違いでも構造物のアピランスに大きな影響を与えることがある。この違いは胸壁では特に顕著であり、単一のセグメンテーションでは認識できなかった胸壁領域を、他のセグメンテーションで認識した場合に補足する情報を得ることができる。このように、同じ動画に属する連続した画像の複数の予測ラベルを統合することで、お互いに情報を補完することができる。

CY は、胸壁のセグメンテーションのために、胸壁、胸腔、胸郭全体の3つの予測ラベルを統合する。この3つの予測ラベルは、それぞれのアノテーションラベルによって学習された3つの独立したモデルから得られる。CY は、胸壁は常に円筒形であるという事前知識を利用している。つまり、胸壁のラベルは、胸郭全体のラベルから胸腔のラベルをくり抜くことで簡単に取得することができる。CY で出力された胸壁の予測ラベルは、胸郭全体のラベルで外側の境界を、胸腔のラベルで内側の境界を、よりしっかりと区別することができる。このように、3つの独立したモデルによる3つの予測ラベルは、お互いに情報を補完することができる。

前述の2つの手法を組み合わせた MFCY は、使用するモデルを問わない model-agnostic な手法である。つまり、任意の CNN モデルから得られた予測結果を利用することで機能する。そのため、MFCY はその時代の最先端の深層学習技術を用いて使用することができ、MFCY に合わせてネットワークを改変する必要がない。

本学位論文は、Biomolecules に公表した原著論文を基に作成した[30]。

2. 方法

2.1.1. 本研究で提案するオリジナルな手法である MFCY の解説

本研究では、4CVの胸壁セグメンテーションを支援するために、MFCYを提案した。MFは、対象画像とその前後フレームを利用し、それらの予測ラベルを統合することで、より良いセグメンテーション性能を実現する。CYでは、胸壁、胸腔、胸郭全体の3つの予測結果を利用する。本研究では、胸郭の外側を「胸壁」、胸壁に囲まれた部分を「胸腔」、胸壁と胸腔の結合部分を「胸郭全体」と定義する。求める予測結果は胸壁の予測ラベルと胸郭全体の予測ラベルの和から胸腔の予測ラベルを除いた領域である。

MFCYの概要を図3に示す。まず、胎児超音波スクリーニング動画から抽出した4CV画像に対して胸壁のアノテーションを行った(図3a)。そして、胸腔、胸郭全体のラベルは、胸壁のラベルをもとに、その定義に基づいて自動的に生成した。次に、胸壁、胸腔、胸郭全体をそれぞれセグメンテーションする独立した学習済みモデルを作成した(図3b)。CNNにはU-netとDeepLabv3+を採用した。最後に、学習した3つのモデルを用いてセグメンテーションを行い、MFとCYを順に適用して各予測ラベルを統合することにより最終的な予測結果を取得した(図3c)。このフェーズでは、胎児超音波動画から抽出した対象の4CV画像とその前後フレームを、学習済みCNNモデル(胸壁、胸腔、胸郭全体)に入力することによりそれぞれ複数の予測ラベルを得る。ここで得た予測ラベルはMFによって胸壁、胸腔、胸郭全体のそれぞれの対象ごとに単一のラベルに統合され、CYがこの3種類の単一ラベルを統合する(図4)。

MFCYの性能を検証するために、CNNには既存モデルとしてU-netとDeepLabv3+を使用し、MF単独、CY単独、MFCYを用いてそれぞれの胸壁セグメンテーションを行い(図5)、既存モデルと3つの手法(MF, CY, MFCY)の性能を比較した。さらに、MFとCYについては、個別の予測結果を用いてその有効性を評価した。また、補足実験として、既存モデルとMFを用いて心臓と肺のセグメンテーションを行い、その性能を比較した。

2.1.2. MFとCYについて

MFは、超音波動画中の各ターゲット超音波画像の前後フレームを利用している。ここで $X_{t=0}$ を対象の超音波画像とし、 $X_{t=0}$ の各要素は0と1の間の実数とする。 $X_{t=t \in T}$ はその前後フレームであり、ここで T は一連の前後画像のセットを表す。さらに、 a と k をそれぞれフレーム間隔とフレーム番号の係数とすると、 $T(a, k) = \{0, \pm a, \pm 2a, \dots, \pm ak\}$ と表すことができ、 $X_{t=0}$ の前後フレームの各要素も0から1までの実数とする。その上で以下のように要素ごとのステップ関数を閾値関数として定義する。

$$Y = H_\phi(X) \Leftrightarrow y_{ij} = \begin{cases} 0 & (x_{ij} < \phi) \\ 1 & (x_{ij} \geq \phi) \end{cases}, y_{ij} \in Y, x_{ij} \in X. \quad (1)$$

ここで、ニューラルネットワークを N 、学習パラメータを θ とし、正規化されたネットワーク $\bar{N}(X; \theta) = H_{\phi=0.5}(N(X, \theta))$ とすると、 N の出力の各要素は0から1の実数である。すると、MFは次のように与えることができる。

$$\text{MF}(X_{t=t \in T}; \bar{N}, \theta, \phi) = H_\phi \left(\sum_{t \in T} \bar{N}(X_t; \theta) \right), \quad (2)$$

ここで ϕ はMFの閾値を示し、 θ は $X_{t=0}$ と対応するラベル $Y_{t=0}$ のセットに基づいて得られたものである。

CYは、セグメンテーション対象が壁に厚みのある円筒状のトポロジー形状であるという性質を利用している。ここで X を超音波画像とする。ニューラルネットワークを N とし、学習パラメータを θ とし、次の3つの学習パラメータを定義する。 θ_{TW} 、 θ_{TC} 、 θ_{WT} はそれぞれ胸壁のラベル Y_{TW} 、胸腔のラベル Y_{TC} 、胸郭全体のラベル Y_{WT} を用いて得られたものである。定義から $Y_{\text{TW}} = Y_{\text{WT}} - Y_{\text{TC}}$ と表すことができるため、 Y_{WT} と Y_{TC} は一意的に Y_{TW} を決定づけることができる。また前述の式では、要素ごとの減算が行われている。ここで、学習パラメータである $\Theta = \{\theta_{\text{TW}}, \theta_{\text{TC}}, \theta_{\text{WT}}\}$ とそのしきい値について考慮する。もし $\Phi = \{\phi_1, \phi_2\}$ とすると、CYは次のように与えられる。

$$\text{CY}(X; \bar{N}, \Theta, \Phi) = H_{\phi_1} \left(\bar{N}(X; \theta_{\text{TW}}) + H_{\phi_2} (\bar{N}(X; \theta_{\text{WT}}) - \bar{N}(X; \theta_{\text{TC}})) \right), \quad (3)$$

ここで“+”は要素ごとの加算を表す。

最後に、MF、CY、MFCYの組み合わせは次のように与えられる。

$$\begin{aligned} \text{MFCY}(X_{t=t \in T}; \bar{N}, \Theta, \Phi) \\ = H_{\phi_1} \left(\text{MF}(X_{t=t \in T}; \bar{N}, \theta_{\text{TW}}, \phi_3) \right. \\ \left. + H_{\phi_2} (\text{MF}(X_{t=t \in T}; \bar{N}, \theta_{\text{WT}}, \phi_3) - \text{MF}(X_{t=t \in T}; \bar{N}, \theta_{\text{TC}}, \phi_3)) \right), \end{aligned} \quad (4)$$

ここで $\Phi = \{\phi_1, \phi_2, \phi_3\}$ はしきい値を表す。

2.2. ネットワーク構成

U-net は、入力に近い層と出力に近い層を結合するために、標準的なエンコーダ・デコーダ構造に skip-connection を追加して、U 字型の構造になっている [14]. 本研究では pix2pix の一部として提案された U-net の改良版を採用した [31]. この改良版では、エンコーダは畳み込み層で構成され、デコーダは畳み込み層と 1 つのアップサンプリング層で構成されている. 入力と出力のサイズは 256×256 ピクセルで、活性化関数は正規化線形ユニットを用いた. 各入力要素は -1 から 1 の実数、各出力は 0 から 1 の実数とし、損失関数には Dice 損失を採用した. 詳細については、[図 6](#) を参照されたい.

DeepLab は Atrous 畳み込みを利用しており、高性能なセグメンテーション手法のひとつである [32]. DeepLabv3+ はその最新バージョンで、エンコーダとデコーダを組み合わせた構造を採用している [16]. 本研究では、エンコーダのバックボーンネットワークとして、ImageNet を用いて事前学習した Resnet101 を用い、損失関数としてクロスエントロピー損失を採用した. 入力と出力のサイズは、 513×513 ピクセルである. 各入力要素は -1 ~ 1 の実数、各出力は 0 ~ 1 の実数とした. 詳細は参考文献 [16] を参照されたい.

2.3. 臨床データの取得方法とラベルの準備

本研究のプロトコルは、理化学研究所、富士通株式会社、昭和大学および国立がん研究センターの Institutional Review Board の承認を得ている (承認 ID : Wako1 29-4). また、ヘルシンキ宣言および人を対象とする医学研究の倫理原則に基づいて実施した. 本研究の対象は、2018 年 4 月から 2019 年 5 月の間に、昭和大学の 4 病院 (東京および横浜) で妊娠中期に胎児超音波スクリーニングを受けた単胎妊娠で正常胎児を妊娠する妊婦とした. 対象の胎児は出生後 1 カ月以上経過した時点で、小児科医による診察を行い、その新生児に先天性疾患がないことを確認している. 胎児超音波動画の取得は、熟練度が異なる産科医が、2~6MHz の経腹用コンベックス型トランスデューサを搭載した Voluson® E8 または E10 (GE Healthcare, Chicago, IL, USA) を用い胎児心臓モードで行った. 動画は、胃の高さから 4CV を経て弓部大動脈に至るまでの、一連の断面図で構成されるように取得した. これらのビデオのすべてのフレームは、胎児の胸部が画面の少なくとも 2 分の 1 から 3 分の 2 を占めるように拡大された. このような条件で、1 人の妊婦から 1 つ以上の動画を取得した. 得られた動画から、4CV の評価に必要な情報を認識できるクオリティの 4CV 画像を抽出した. 以上のプロセスを経て、胎児胸壁セグメンテーションのためのデータセットは、256 例の 280 本の動画に含まれる 538 枚の 4CV 画像となった. 256 例の妊娠期間の 18 週から 28 週で中央値は 20 週であった.

データセットとして集めた 4CV 画像には、1 人の産科医が胸壁のアノテーションを手作業で行った. ここで胸壁とは、肋骨と脊柱を含み、皮膚や外側の筋肉を含まない外側

の境界と、胸膜に一致する内側の境界の2つの境界の間にある円筒状構造物と定義した。この定義は、日本胎児心臓病研究会および日本小児循環器学会の胎児心エコー検査ガイドライン作成委員会が発表した「胎児心エコー検査ガイドライン」を参考にした[13]。補足実験のために心臓および肺のアノテーションも同様に行った。

CYでは、胸壁、胸腔、胸郭全体の3種類のアノテーションラベルで学習したそれぞれ3種類のモデルが必要となる。胸壁のアノテーションラベルは手動で作成したが、これを利用することにより、残りの胸腔と胸郭全体のラベルは自動的に、すなわちコストフリーで生成することができる。そしてMFで使用する各ターゲット画像の前後フレームをコンフィギュレーションに従って抽出した。入力データとして使用する画像やラベルは、U-netの場合は 256×256 ピクセル、DeepLabv3+の場合は 513×513 ピクセルにリサイズした。

2.4. 評価指標とクロスバリデーション

今回、セグメンテーションモデルの性能を評価するために、intersection over union (IoU)、Dice 係数 (Dice)、適合率 (Precision)、再現率 (Recall) という4つの指標を用いた[33]。真陽性を TP、偽陽性を FP、偽陰性を FN、偽陽性を FP とすると、IoU、Dice、Precision、Recall は以下のように定義できる。

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (5)$$

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

Jaccard index とも呼ばれる IoU と Dice は、セグメンテーション手法の評価に最もよく使われる指標である。また、推論の詳細な評価を行うために、Precision と Recall も一般的に使用されている。前述の4つの評価指標はすべて0から1の間の値をとり、1に近い値ほど良い予測結果を示す。

既存モデルと MF、CY、MFCY の手法の性能を一般化するために、5-fold クロスバリデーションを行った。この目的のために、データセットは一様に5つの独立したサブセットに分割し、それらのサブセットはほぼ等しいカーディナリティとした。5-fold クロスバリデーションでは、1つのサブセットをテストデータセットとし、残りの4つのサブセットをトレーニングデータセットとして順次割り当てるというフォールドを繰り返す。

返す。各テストデータセットについて、平均 IoU (mIoU), 平均 Dice (mDice), 平均 Precision (mPrecision), 平均 Recall (mRecall)を算出した。最後に、mIoU, mDice, mPrecision, mRecall の 5-fold クロスバリデーションでの平均値と母集団の標準偏差を計算した。

また、MF や CY の性能を詳細に評価するために、個々のテスト画像の IoU を算出した。IoU_{U-net}, IoU_{DeepLabv3+}, IoU_{MF}, IoU_{CY}を、それぞれ U-net, DeepLabv3+, MF, CY による胸壁のセグメンテーションの IoU とする。それらの IoU 値の違いを評価した ($\Delta \text{IoU} = \text{IoU}_{\text{MF}} - \text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$, または $\Delta \text{IoU} = \text{IoU}_{\text{CY}} - \text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$) ΔIoU は、5-fold クロスバリデーションで得たすべてのテスト画像の結果で計算した。CNN モデルは、U-net と DeepLabv3+を使用した。

2.5. 実験に使用したコンフィギュレーション

U-net と DeepLabv3+を使用する際には、両手法で動作条件を同じにするために、セグメンテーションの過程で、データ拡張やファインチューニングは行わなかった。U-net ではバッチサイズ 12, エポックサイズ 40, DeepLabv3+では 30,000 回のイテレーションとバッチサイズ 8 を設定した。U-net では Adam optimizer を採用し、学習率は 0.001 とした。また、DeepLabv3+では、Nesterov momentum を 0.9, 初期学習率を 0.007, 2000 回のイテレーションごとの減衰率を 0.9 とする確率的勾配降下法を採用した。CY については、 $T(a, k) = T(3, 3)$ とした。また、式(1)では $\phi = 0.5$, 式(2)では $\phi = 3/7$, 式(3)では $\Phi\{\phi_1, \phi_2\} = \{0.5, 0.5\}$, 式(4)では $\Phi\{\phi_1, \phi_2, \phi_3\} = \{0.5, 0.5, 3/7\}$ と設定した。U-net には Keras ver2.2 (TensorFlow ver1.9 backend) を、DeepLabv3+には TensorFlow ver1.10 を使用した。すべての実験は、CentOS Linux リリース 7.2, CUDA バージョン: 10 のコンピュータクラスタ上で行われ、各ノードのハードウェアは、Intel(R) Xeon(R) CPU E5-2690 v4 (2.60GHz), GeForce GTX 1080 Ti を使用した。本研究で提案した手法のソースコードは、GitHub (<https://github.com/rafcc/2020-circle-segmentation>) で公開している。

3. 結果

3.1. 既存モデルと提案手法の性能の比較

U-net, DeepLabv3+, MF, CY, MFCY を用いて得られた胸壁の予測結果を、mIoU, mDice, mPrecision, mRecall のそれぞれで比較した (表 1)。各モデルを介して得られた予測結果を、対応する正解ラベルとともに 図 7 に示す。MFCY を使用すると、mIoU で 0.448 (U-net) vs. 0.493 (U-net + MFCY), 0.417 (DeepLabv3+) vs.

0.470 (DeepLabv3+ + MFCY) となり、既存モデルよりも優れた結果を得ることができた。mRecall でも、0.568 (U-net) vs. 0.738 (U-net+MFCY) , 0.525

(DeepLabv3+) vs. 0.729 (DeepLabv3++MFCY) と、既存モデルよりも高かった。しかし、mPrecision においては、0.679 (U-net) vs. 0.596 (U-net + MFCY) , 0.675 (DeepLabv3+) vs. 0.566 (DeepLabv3+ + MFCY) と、MFCY を使用すると既存モデルよりも低い値を示した。5-fold クロスバリデーションで得たそれぞれ得たすべての結果は [図 8](#) に示す。MF を用いて得られた胸腔および胸郭全体の予測結果を [表 2](#) に示した。また、補足実験として心臓と肺のセグメンテーションに対応する MF の性能を評価した ([表 3](#))。その結果、MF を使用すると、胸壁、胸腔、胸郭全体を含め、心臓や肺についても mIoU が高くなり、より正確なセグメンテーションを行うのに有効であった。

3.2. MF と CY の個別効果の分析

次に、MF または CY を用いた胸壁セグメンテーションの改善度を分析し、それぞれの貢献度を評価した ([図 9](#), [10](#))。セグメンテーション結果の改善度は、画像毎で異なっていた ([図 9a](#), [10a](#))。 [図 9b](#) のケース 5 と 6, [図 10b](#) のケース 8 と 9 は、典型的な改善例である。これらのケースでは、MF と CY の両方が、既存モデルの単独使用では認識することができなかったセグメンテーション領域 (以下、失敗領域) を補うことができた。 [図 9b](#) のケース 7 と [図 10b](#) のケース 10 は、典型的な改悪した例である。ケース 7 では、MF の前後フレームの予測結果が悪かったため、統合予測が改悪した。また、ケース 10 では、胸部以外の物体による不適切なセグメンテーション結果により、CY での予測結果が改悪した。

4. 考察

胸壁は、先天性疾患を検出・推定するための胎児超音波検査における重要な構造である。第一に、検査者は左右の肋骨の長さが揃った対称的な胸壁を得ることで、スキャンビューが適切な 4CV であることを確認する[11]。第二に、胸壁の形状の変化は、先天性疾患を示唆する。特に、胸郭が狭い骨格形成不全の症例では胸郭周囲が小さくなり[34]、漏斗胸の症例では胸壁自体が変形する[35]。第三に、先天性疾患をスクリーニングする際や疾患の予後予測をするためのある指標は、胸壁の測定によって算出される。例えば、4CV における心臓と胸郭の面積の比である心胸郭断面積比は、先天性心疾患や胎児心不全の代表的な指標である[36,37]。また、4CV における対側の肺の面積と胸郭の面積の比である肺・胸郭横断面積比は、先天性横隔膜ヘルニアの予後指標となる[38,39]。そして、胸壁に対する心臓の向きと位置によって決まる心臓軸も先天性疾患を検出するきっかけの指標となる[12,13]。このことから、胸壁のセグメンテーション

を採用した AI ベースのモデルは、胎児超音波検査における 4CV の評価や先天性疾患の診断をサポートする可能性がある。この点において、MFCY は、胸壁の正確なセグメンテーションを行うことで、胎児超音波動画における AI ベースの診断支援モデルの進展に貢献することができる。

本研究では、2つの model-agnostic な手法を組み合わせることで、胎児超音波動画における新規のセグメンテーション手法である MFCY を提案した。そして MFCY を用いた胸壁のセグメンテーションの mIoU と mDice は、既存モデルを用いたセグメンテーションよりも高く、既存モデルよりも優れた性能を示した。MFCY は、既存のニューラルネットワークの予測結果をそのまま適用し、各動画のターゲット画像とその前後フレームの予測結果を統合することで機能する。驚くべきことに、データセットが限られた中で、ネットワークに修正を加えることなく、またより多くの追加のアノテーションを行わずとも、セグメンテーション性能を向上させることができた。MFCY が他のネットワークアーキテクチャにも適用でき、他のセグメンテーション手法にも広く適用できることは、model-agnostic な手法の利点である。

次に、今回の結果をもとに、MF や CY が胸壁のセグメンテーションの改善にどのように貢献しているかを詳細に評価した。その結果、MF や CY を単独で使用した場合でも、胸壁セグメンテーションの性能は既存モデルよりも向上した。詳細な検討では MF、CY、MFCY の mRecall は高くなった一方で、mPrecision は既存モデルよりも低い結果となった。これは、MF では、画像を少しずつずらして使用しているため、カバーする範囲が広いことが原因と考えられる。さらに、CY では統合予測結果を求めるために数学的な演算 (式 (3)) が必要のため、カバーする領域は対応する正解ラベルよりも広がる可能性が考えられた。一つ一つのケースに着目すると、MF、CY は多くのケースで、失敗領域を補完し、より円筒形に近づけることで、セグメンテーション結果を改善させていた。これは、MF と CY の独立した予測結果のアンサンブル学習により、補正が行われているためと考えられた。MF では、各ターゲット画像に対応する前後フレームを利用することで、必ずしもすべての前後フレームに現れるわけではない失敗領域を補うことで、セグメンテーション性能を向上させた (図 9b のケース 5, 6 など)。胸壁のアピランスの時間的・空間的变化がセグメンテーション性能に影響を与えるが、MF ではターゲット画像の失敗領域を、前後フレームの情報で補うことができる。また、CY では、3つの異なるモデルの出力を統合することで、単一のモデルを実装するよりも優れた性能を発揮した。胸壁が円筒形を呈しているという予備知識に基づき、胸壁のラベルから胸腔と胸郭全体のラベルを生成することができる。逆に、胸腔と胸郭全体の予測ラベルから胸壁の予測ラベルを決めることができる (例えば、図 10b のケース 8 と 9)。今回の研究ではデータセットのサイズは限られていたが、これらの定性的な分析から、提案手法は有効であると考えられた。

別の観点から、胎児超音波で胸壁のセグメンテーションが困難な理由を考える。U-net や DeepLabv3+ を単独で用いた既存手法で胸壁のセグメンテーションがうまくいかなかったケースでは、胸壁の予測ラベルが途切れていることが多かった (図 7, 9b, 10b) 。その理由として考えられるのは、同じ胎児超音波動画の中でも、胸壁のアピランスが時間的・空間的に変化していることである。胎児超音波検査では、胸壁は多様な明るさの構造物が混在しているため、アピランスが異なり、胸壁自体や周囲の構造物に起因するアーチファクトが頻繁に現れる。また、胸壁は単連結空間ではないため、その外側の領域と内側の穴あき部分を識別する必要があり、セグメンテーションの工程を複雑化している。これらの理由から、U-net や DeepLabv3+ を単純に使用するだけでは、最適な予測結果が得られなかった可能性が高いと考えられる。

4.1. 本研究の限界

本研究にはいくつかの限界がある。まず、今回提案する MFCY の性能は、代表的なモデルである U-net と DeepLabv3+ との組み合わせでのみ検証した。MFCY は model-agnostic な手法であるため、形式的にはすべての深層学習アルゴリズムに適用可能である。しかし、MFCY を他の深層学習アルゴリズムに適応することで、セグメンテーション性能がどの程度向上するかは不明である。第 2 に、入力データは、ひとつの企業の 2 種類の超音波診断装置からのみ取得したものであり、ロバスト性を担保することができない。したがって、MFCY が様々な機器から取得したデータセットに応用できるかどうかは、さらなる実験での検証が必要となる。第三に、今回のデータセットには正常な胎児の超音波動画しか含まれていないため、先天性心疾患などの胎児異常がある場合の MFCY の有効性については検証していない。一般的な AI ベースのモデルを確立するには、正常なケースと異常なケースの両方を含める必要がある。第四に、今回の研究では胸壁セグメンテーションに焦点を当てており、今回のデータセットを使って他の構造物に対する MFCY の有効性を検証することはできない。MFCY の重要性を示し、応用を拡大するためには、胎児の頭蓋骨のような円筒形をした他の器官のセグメンテーションをテストする必要がある。最後に、今回の研究は臨床現場での検証は行っていない。最大の mIoU (U-net+MFCY) でも 0.493 であり、これは臨床的な実用性を考えると、まだ比較的低い値である。トレーニングデータの数を増やすことによって、より正確なセグメンテーションが可能になると期待される。したがって、MFCY の臨床レベルでの導入と検証は、今後の課題である。

5. 結論

本研究では、model-agnostic な手法である MFCY によって、胎児の超音波動画における胸壁のセグメンテーション性能が向上することを示した。MFCY は、2 つの手法の独

立した予測結果のアンサンブル学習に基づいており、超音波動画の時系列情報と胸壁の形状情報を相互に補完する。4CV の評価を支援する AI ベースのモデルを構築するためには、胎児胸壁の正確なセグメンテーションが不可欠である。MFCY が、胎児超音波の自動診断支援技術の開発に貢献できることに期待する。

6. 謝辞

国立がん研究センター研究所・医療 AI 研究開発分野の浜本隆二分野長、小松正明先生には研究を進めるにあたり多大なるご指導をいただき、深謝いたします。富士通株式会社・人工知能研究所の酒井彬先生には AI 研究の手法の開発、解析方法、結果の解釈など詳細なご助言をいただき、大変感謝申し上げます。国立がん研究センター研究所・医療 AI 研究開発分野の金子修三先生、浅田健先生、町野英徳先生、同前愛先生、富士通株式会社・人工知能研究所の安富優先生には日頃より、データ解析のご協力をいただきました。昭和大学医学部産婦人科学講座の関沢明彦教授、松岡隆准教授、新垣達也講師、小松玲奈助教らには研究データの取得について多大なるご尽力をいただきました。富士通株式会社の佐野久幸様、ジャパンシステム株式会社吉田裕之様には AI 研究のご助言をいただきました。富山大学産科婦人科学教室の中島彰俊教授には実験を進める環境について手厚いサポートをいただきました。上記の先生方を含め、本研究に携わっていただきましたすべての関係者の皆様、貴重なデータを提供していただきました被験者の皆様に対して、この場を借りて深く御礼申し上げます。

7. 付録

本研究の予備実験として既存モデルによる心臓、肺、胸壁のセグメンテーション性能を評価した。データセットは、本文で述べたものと同様で、256名の単胎妊娠で正常胎児を妊娠する妊婦から取得した4CV画像538枚からなる。これらの画像の心臓、肺、胸壁のラベルは、産科医が手動でアノテーションしたものである。既存モデルとして、U-netとDeepLabv3+を使用した。U-netではバッチサイズ12、エポックサイズ40、DeepLabv3+では30,000回の反復、バッチサイズ8を採用した。U-netのアーキテクチャは、本文で説明したものと同じである(図6)。DeepLabv3+の詳細については、[16]を参照されたい。セグメンテーションは5-foldクロスバリデーションで行った。付録表1に実験の結果を示す。心臓、肺、胸壁のセグメンテーションのmIoUは、U-netではそれぞれ0.686、0.712、0.448、DeepLabv3+ではそれぞれ0.734、0.684、0.417であった。既存CNNモデルによる胸壁のセグメンテーション性能は、心臓や肺に比べて低かった。

8. 参考文献

1. Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2017**, *2*, 230–243, doi:10.1136/svn-2017-000101.
2. Hamamoto, R.; Komatsu, M.; Takasawa, K.; Asada, K.; Kaneko, S. Epigenetics analysis and integrated analysis of multiomics data, including epigenetic data, using artificial intelligence in the era of precision medicine. *Biomolecules* **2020**, *10*, doi:10.3390/biom10010062.
3. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A review of semantic segmentation using deep neural networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93, doi:10.1007/s13735-017-0141-z.
4. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv* **2017**, 1–23.
5. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput. J.* **2018**, *70*, 41–65, doi:10.1016/j.asoc.2018.05.018.
6. Hesamian, M.H.; Jia, W.; He, X.; Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J. Digit. Imaging* **2019**, *32*, 582–596, doi:10.1007/s10278-019-00227-x.
7. Garcia-Canadilla, P.; Sanchez-Martinez, S.; Crispi, F.; Bijmens, B. Machine Learning in Fetal Cardiology: What to Expect. *Fetal Diagn. Ther.* **2020**, *47*, 363–372, doi:10.1159/000505021.
8. Liu, S.; Wang, Y.; Yang, X.; Lei, B.; Liu, L.; Li, S.X.; Ni, D.; Wang, T. Deep Learning in Medical Ultrasound Analysis: A Review. *Engineering* **2019**, *5*, 261–275, doi:10.1016/j.eng.2018.11.020.
9. Salomon, L.J.; Alfirevic, Z.; Berghella, V.; Bilardo, C.; Hernandez-Andrade, E.; Johnsen, S.L.; Kalache, K.; Leung, K.Y.; Malinger, G.; Munoz, H.; et al. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound Obstet. Gynecol.* **2011**, *37*, 116–126, doi:10.1002/uog.8831.
10. Roberts, K.P.; Weinhaus, A.J. Anatomy of the Thoracic Wall, Pulmonary Cavities, and Mediastinum. In *Handbook of Cardiac Anatomy, Physiology, and Devices*; Humana Press: Totowa, NJ, 2559; Vol. 11, pp. 25–50 ISBN 9786162833052.

11. Bethune, M.; Alibrahim, E.; Davies, B.; Yong, E. A pictorial guide for the second trimester ultrasound. *Australas. J. ultrasound Med.* **2013**, *16*, 98–113, doi:10.1002/j.2205-0140.2013.tb00106.x.
12. The International Society of Ultrasound in Obstetrics ISUOG Practice Guidelines (updated): sonographic screening examination of the fetal heart. *Ultrasound Obstet. Gynecol.* **2013**, *41*, 348–359, doi:10.1002/uog.12403.
13. Satomi, G. Guidelines for fetal echocardiography. *Pediatr. Int.* **2015**, *57*, 1–21, doi:10.1111/ped.12467.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2015**, *9351*, 234–241, doi:10.1007/978-3-319-24574-4_28.
15. Liu, L.; Cheng, J.; Quan, Q.; Wu, F.X.; Wang, Y.P.; Wang, J. A survey on U-shaped networks in medical image segmentations. *Neurocomputing* **2020**, *409*, 244–258, doi:10.1016/j.neucom.2020.05.070.
16. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2018**, *11211 LNCS*, 833–851, doi:10.1007/978-3-030-01234-2_49.
17. Hamamoto, R.; Suvarna, K.; Yamada, M.; Kobayashi, K.; Shinkai, N.; Miyake, M.; Takahashi, M.; Jinnai, S.; Shimoyama, R.; Sakai, A.; et al. Application of artificial intelligence technology in oncology: Towards the establishment of precision medicine. *Cancers (Basel)*. **2020**, *12*, 1–32, doi:10.3390/cancers12123532.
18. Liu, X.; Deng, Z.; Yang, Y. Recent progress in semantic image segmentation. *Artif. Intell. Rev.* **2019**, *52*, 1089–1106, doi:10.1007/s10462-018-9641-3.
19. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016**, *2016-Decem*, 3213–3223, doi:10.1109/CVPR.2016.350.
20. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical Image Analysis using Convolutional Neural Networks: A Review. *J. Med. Syst.* **2018**, *42*, 226, doi:10.1007/s10916-018-1088-1.
21. Heuvel, T.L.A. van den; Bruijn, D. de; Korte, C.L. de; Ginneken, B. van Automated measurement of fetal head circumference. *PLoS One* **2018**, *4*, 1–20, doi:10.5281/ZENODO.1322001.
22. Arnaout, R.; Curran, L.; Zhao, Y.; Levine, J.; Chinn, E.; Moon-Grady, A. Expert-level prenatal detection of complex congenital heart disease from screening ultrasound using deep learning. *medRxiv* **2020**, 2020.06.22.20137786, doi:10.1101/2020.06.22.20137786.

23. Burgos-Artizzu, X.P.; Perez-Moreno, Á.; Coronado-Gutierrez, D.; Gratacos, E.; Palacio, M. Evaluation of an improved tool for non-invasive prediction of neonatal respiratory morbidity based on fully automated fetal lung ultrasound analysis. *Sci. Rep.* **2019**, *9*, 1–7, doi:10.1038/s41598-019-38576-w.
24. Looney, P.; Stevenson, G.N.; Nicolaidis, K.H.; Plasencia, W.; Molloholli, M.; Natsis, S.; Collins, S.L. Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning. *JCI insight* **2018**, *3*, 1–9, doi:10.1172/jci.insight.120178.
25. Li, Y.; Xu, R.; Ohya, J.; Iwata, H. Automatic fetal body and amniotic fluid segmentation from fetal ultrasound images by encoder-decoder network with inner layers. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* **2017**, 1485–1488, doi:10.1109/EMBC.2017.8037116.
26. Dozen, A.; Komatsu, M.; Sakai, A.; Komatsu, R.; Shozu, K.; Machino, H.; Yasutomi, S.; Arakaki, T.; Asada, K.; Kaneko, S.; et al. Image Segmentation of the Ventricular Septum in Fetal Cardiac Ultrasound Videos Based on Deep Learning Using Time-Series Information. *Biomolecules* **2020**, *10*, 1526, doi:10.3390/biom10111526.
27. Yu, L.; Guo, Y.; Wang, Y.; Yu, J.; Chen, P. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1886–1895, doi:10.1109/TBME.2016.2628401.
28. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, 1–18, doi:10.1002/widm.1249.
29. Kusunose, K.; Abe, T.; Haga, A.; Fukuda, D.; Yamada, H.; Harada, M.; Sata, M. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc. Imaging* **2020**, *13*, 374–381, doi:10.1016/j.jcmg.2019.02.024.
30. Shozu, K.; Komatsu, M.; Sakai, A.; Komatsu, R.; Dozen, A.; Machino, H.; Yasutomi, S.; Arakaki, T.; Asada, K.; Kaneko, S.; et al. Model-agnostic method for thoracic wall segmentation in fetal ultrasound videos. *Biomolecules* **2020**, *10*, 1–16, doi:10.3390/biom10121691.
31. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* **2017**, *2017-Janua*, 5967–5976, doi:10.1109/CVPR.2017.632.
32. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848, doi:10.1109/TPAMI.2017.2699184.

33. Cai, S.; Tian, Y.; Lui, H.; Zeng, H.; Wu, Y.; Chen, G. Dense-UNet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quant. Imaging Med. Surg.* **2020**, *10*, 1275–1285, doi:10.21037/QIMS-19-1090.
34. Ruano, R.; Molho, M.; Roume, J.; Ville, Y. Prenatal diagnosis of fetal skeletal dysplasias by combining two-dimensional and three-dimensional ultrasound and intrauterine three-dimensional helical computer tomography. *Ultrasound Obstet. Gynecol.* **2004**, *24*, 134–140, doi:10.1002/uog.1113.
35. Çetin, C.; Büyükkurt, S.; Sucu, M.; Özsürmeli, M.; Demir, C. Prenatal diagnosis of pectus excavatum. *Turkish J. Obstet. Gynecol.* **2016**, *13*, 158–160, doi:10.4274/tjod.54514.
36. Paladini, D.; Chita, S.K.; Allan, L.D. Prenatal measurement of cardiothoracic ratio in evaluation of heart disease. *Arch. Dis. Child.* **1990**, *65*, 20–23, doi:10.1136/adc.65.1_Spec_No.20.
37. Awadh, A.M.A.; Prefumo, F.; Bland, J.M.; Carvalho, J.S. Assessment of the intraobserver variability in the measurement of fetal cardiothoracic ratio using ellipse and diameter methods. *Ultrasound Obstet. Gynecol.* **2006**, *28*, 53–6, doi:10.1002/uog.2813.
38. Usui, N.; Kitano, Y.; Okuyama, H.; Saito, M.; Morikawa, N.; Takayasu, H.; Nakamura, T.; Hayashi, S.; Kawataki, M.; Ishikawa, H.; et al. Reliability of the lung to thorax transverse area ratio as a predictive parameter in fetuses with congenital diaphragmatic hernia. *Pediatr. Surg. Int.* **2011**, *27*, 39–45, doi:10.1007/s00383-010-2725-z.
39. Hidaka, N.; Murata, M.; Sasahara, J.; Ishii, K.; Mitsuda, N. Correlation between lung to thorax transverse area ratio and observed/expected lung area to head circumference ratio in fetuses with left-sided diaphragmatic hernia. *Congenit. Anom. (Kyoto)*. **2015**, *55*, 81–84, doi:10.1111/cga.12092.

表 1. MF, CY, MFCY を適用した時の胸壁セグメンテーション性能結果.

| CNN | MF | CY | mIoU | | mDice | | mPrecision | | mRecall | |
|------------|----|----|-------|-------|-------|-------|------------|-------|---------|-------|
| | | | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 |
| U-net | | | 0.448 | 0.017 | 0.610 | 0.016 | 0.679 | 0.033 | 0.568 | 0.032 |
| | ✓ | | 0.461 | 0.017 | 0.623 | 0.016 | 0.659 | 0.035 | 0.606 | 0.032 |
| | | ✓ | 0.486 | 0.006 | 0.647 | 0.005 | 0.599 | 0.014 | 0.718 | 0.025 |
| | ✓ | ✓ | 0.493 | 0.006 | 0.654 | 0.005 | 0.596 | 0.016 | 0.738 | 0.023 |
| DeepLabv3+ | | | 0.417 | 0.007 | 0.582 | 0.007 | 0.675 | 0.009 | 0.525 | 0.013 |
| | ✓ | | 0.443 | 0.006 | 0.608 | 0.006 | 0.663 | 0.009 | 0.575 | 0.014 |
| | | ✓ | 0.462 | 0.005 | 0.626 | 0.005 | 0.567 | 0.004 | 0.709 | 0.012 |
| | ✓ | ✓ | 0.470 | 0.004 | 0.633 | 0.004 | 0.566 | 0.007 | 0.729 | 0.010 |

mIoU : 平均 IoU, mDice : 平均 Dice 係数, mPrecision : 平均適合率, mRecall : 平均再現率.

表 2. MF を適用した時の胸壁, 胸腔, 胸郭全体のセグメンテーション性能結果.

| CNN | 対象 | MF | mIoU | | mDice | | mPrecision | | mRecall | |
|------------|------|----|-------|-------|-------|-------|------------|-------|---------|-------|
| | | | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 |
| CNN | 胸壁 | | 0.448 | 0.017 | 0.610 | 0.016 | 0.679 | 0.033 | 0.568 | 0.032 |
| | | ✓ | 0.461 | 0.017 | 0.623 | 0.016 | 0.659 | 0.035 | 0.606 | 0.032 |
| U-net | 胸腔 | | 0.856 | 0.004 | 0.922 | 0.003 | 0.943 | 0.012 | 0.905 | 0.008 |
| | | ✓ | 0.864 | 0.006 | 0.926 | 0.004 | 0.936 | 0.013 | 0.920 | 0.007 |
| U-net | 胸郭全体 | | 0.843 | 0.029 | 0.912 | 0.019 | 0.954 | 0.005 | 0.880 | 0.036 |
| | | ✓ | 0.852 | 0.022 | 0.918 | 0.014 | 0.950 | 0.005 | 0.894 | 0.028 |
| DeepLabv3+ | 胸壁 | | 0.417 | 0.007 | 0.582 | 0.007 | 0.675 | 0.009 | 0.525 | 0.013 |
| | | ✓ | 0.443 | 0.006 | 0.608 | 0.006 | 0.663 | 0.009 | 0.575 | 0.014 |
| DeepLabv3+ | 胸腔 | | 0.844 | 0.003 | 0.914 | 0.002 | 0.915 | 0.004 | 0.918 | 0.004 |
| | | ✓ | 0.848 | 0.004 | 0.916 | 0.002 | 0.906 | 0.005 | 0.932 | 0.003 |
| DeepLabv3+ | 胸郭全体 | | 0.857 | 0.005 | 0.922 | 0.003 | 0.921 | 0.007 | 0.928 | 0.003 |
| | | ✓ | 0.861 | 0.005 | 0.924 | 0.003 | 0.914 | 0.007 | 0.939 | 0.002 |

mIoU : 平均 IoU, mDice : 平均 Dice 係数, mPrecision : 平均適合率, mRecall : 平均再現率.

表 3. MF を適用した時の心臓, 肺のセグメンテーション性能結果.

| CNN | 対象 | MF | mIoU | | mDice | | mPrecision | | mRecall | |
|------------|----|----|-------|-------|-------|-------|------------|-------|---------|-------|
| | | | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 |
| U-net | 心臓 | | 0.686 | 0.042 | 0.798 | 0.036 | 0.849 | 0.028 | 0.784 | 0.069 |
| | | ✓ | 0.697 | 0.040 | 0.806 | 0.034 | 0.833 | 0.029 | 0.810 | 0.067 |
| U-net | 肺 | | 0.712 | 0.015 | 0.828 | 0.011 | 0.832 | 0.020 | 0.832 | 0.014 |
| | | ✓ | 0.713 | 0.015 | 0.829 | 0.011 | 0.808 | 0.021 | 0.859 | 0.012 |
| DeepLabv3+ | 心臓 | | 0.734 | 0.012 | 0.842 | 0.009 | 0.864 | 0.008 | 0.833 | 0.008 |
| | | ✓ | 0.744 | 0.011 | 0.849 | 0.008 | 0.848 | 0.009 | 0.862 | 0.007 |
| DeepLabv3+ | 肺 | | 0.684 | 0.010 | 0.808 | 0.008 | 0.833 | 0.008 | 0.793 | 0.014 |
| | | ✓ | 0.697 | 0.008 | 0.818 | 0.006 | 0.814 | 0.009 | 0.830 | 0.011 |

mIoU : 平均 IoU, mDice : 平均 Dice 係数, mPrecision : 平均適合率, mRecall : 平均再現率.

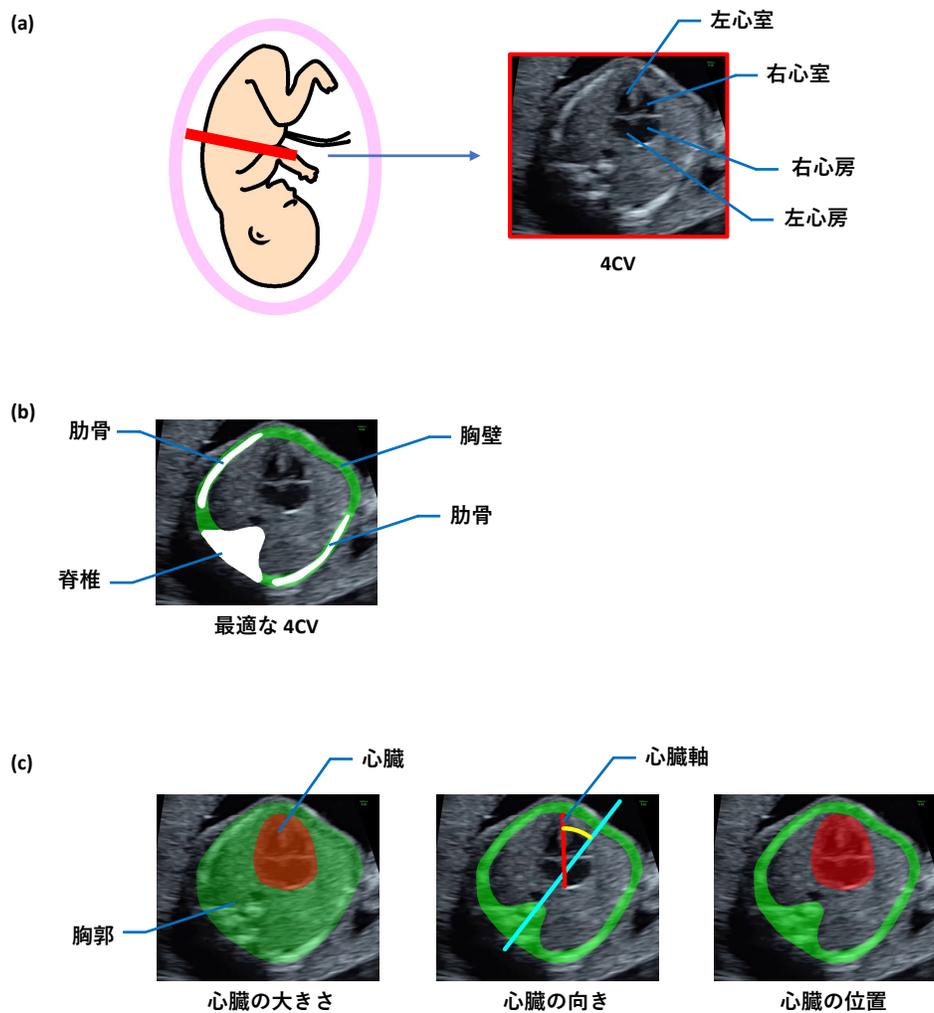


図1. 胎児超音波検査における四腔断面（four-chamber-view：4CV）での評価方法.

(a) 4CVのシェーマ。4CVは心臓の断面の1つで、右心房、右心室、左心房、左心室を観察することができる。(b) 検査者は、胸壁が円形に近く、左右の肋骨の長さが揃っていることを確認することで得られた4CVが適切な断面であることを確認する。(c) 4CV検査の評価指標の例。心臓の大きさは、通常、胸郭面積の1/3以下である。また、正常な心臓は胸郭の左側に位置し、その心臓軸は約 $45 \pm 20^\circ$ 左に向いている。これらの指標は、先天性疾患を発見するきっかけとなる。



図 2. 深層学習技術を用いた代表的な画像処理手法.

胸壁（thoracic wall：TW）を画像処理手法の対象とした例を示す．画像分類では，画像が何であるかを分類する．物体検出は，画像のどのあたりに何があるかを特定する．一方，セグメンテーションでは，画像内の個々のピクセルが何であるかを分類する．したがって，セグメンテーションを行うことにより画像に対する物体の向きや大きさ，境界線などの詳細な情報を得ることができる．

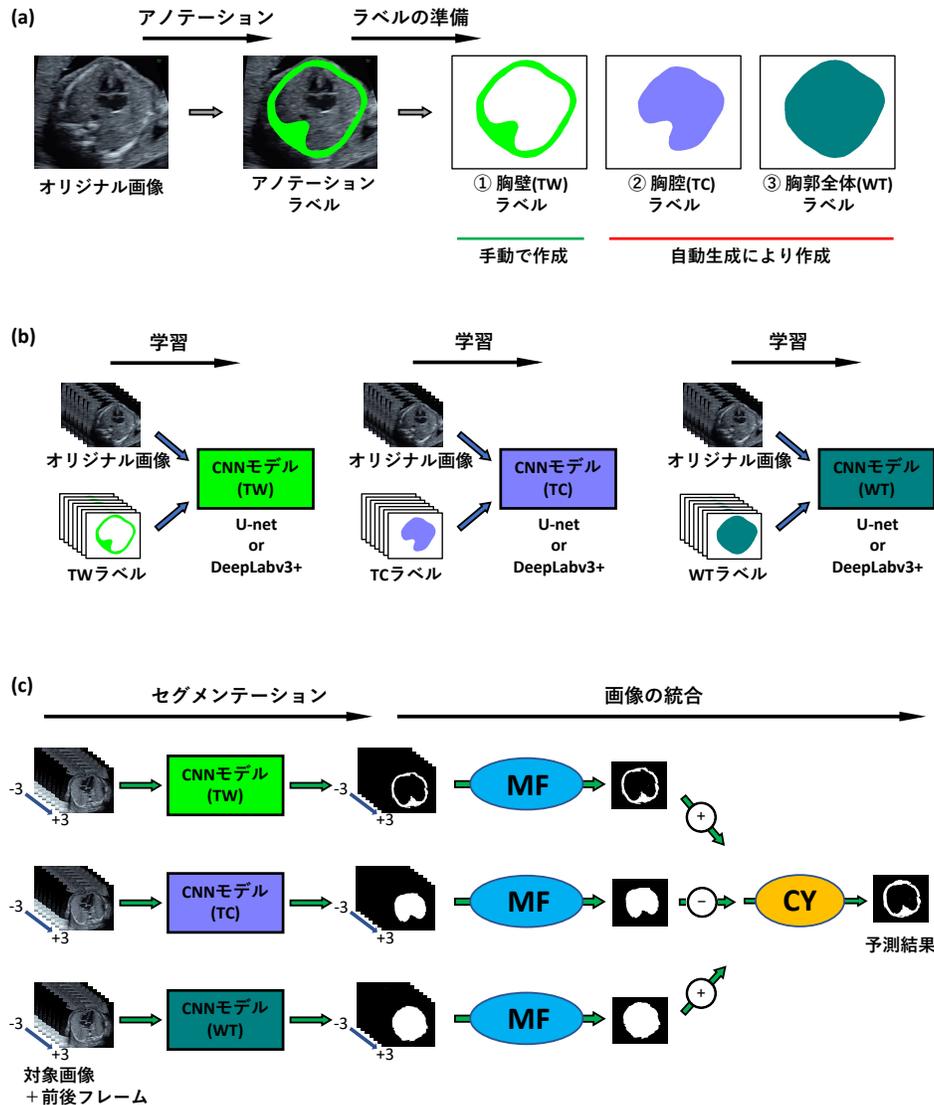


図3. 提案手法である Multi-Frame + Cylinder 法 (MFCY) の概要.

(a) アノテーションラベルの作成フェーズ. 胎児の超音波動画から抽出した 4CV 画像に, 胸壁 (thoracic wall : TW) のラベルをアノテーションした. 胸腔 (thoracic cavity : TC) と胸部全体 (whole thorax : WT) のラベルは, TW のラベルに基づいて自動的に生成した. (b) 教師あり学習での学習フェーズ. オリジナル画像とそれぞれのラベルを用いて, TW, TC, WT のセグメンテーションを行う 3 種類の畳み込みニューラルネットワーク (convolutional neural network : CNN) モデルを作成した. CNN には, U-net と DeepLabv3+ を使用した. (c) 推論フェーズ. このフェーズでは, 胎児の超音波動画から対象となる 4CV 画像とその前後フレームを抽出した. これらの画像を学習済みの CNN モデル (TW, TC, WT) に入力し, 各画像に対応する予測ラベルを取得した. TW, TC, WT の予測ラベルは, MF によってそれぞれ 1 つの予測ラベルに統合した. また, CY を用いて 3 つの独立したラベルを統合し得られた予測ラベルを, MFCY を用いた予測結果とした.

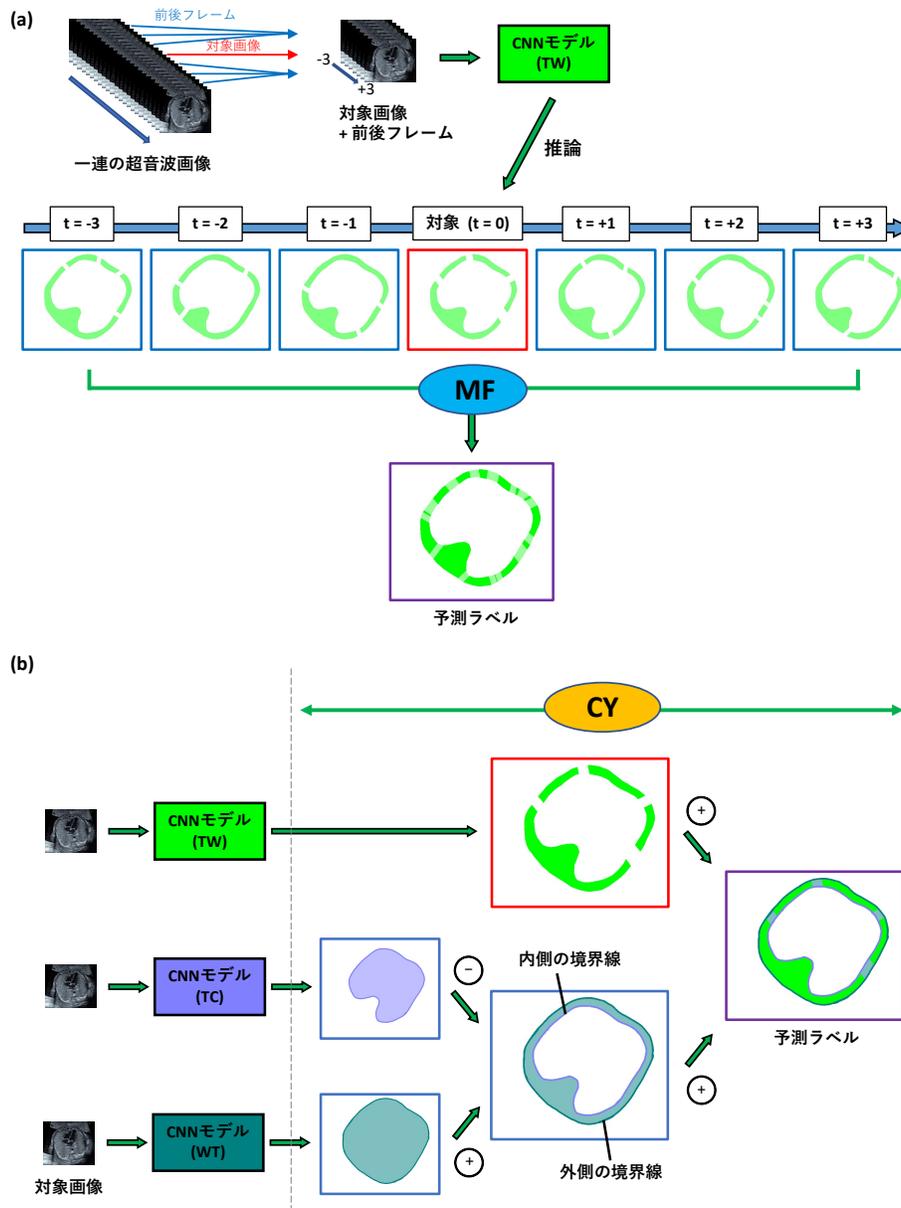


図4. Multi-Frame法(MF)とCylinder法(CY)の手順のダイアグラム。

(a) MFの手順. MFでは、一連の超音波動画の中から対象となる4CV画像とその前後フレームを抽出した。次に、対象画像とその前後フレームを学習済みCNNモデルに入力し、時系列に応じた複数の予測ラベルを取得した。MFはこれらの予測ラベルを統合して、単一の予測ラベルを生成した。時系列の情報を統合することにより、セグメンテーションの性能は向上する。(b) CYの手順. 対象画像をそれぞれの胸壁、胸腔、胸郭全体のCNNモデルに入力し、それぞれの予測ラベルを作成した。CYはこの3つの予測ラベルを統合して、胸壁の出力ラベルを1つ生成する。この統合では、胸郭全体の予測ラベルから胸腔の予測ラベルを差し引き、胸壁の予測ラベルと足し合わせる。これにより、胸壁セグメンテーションの外側の境界線と内側の境界線がより明確となる。

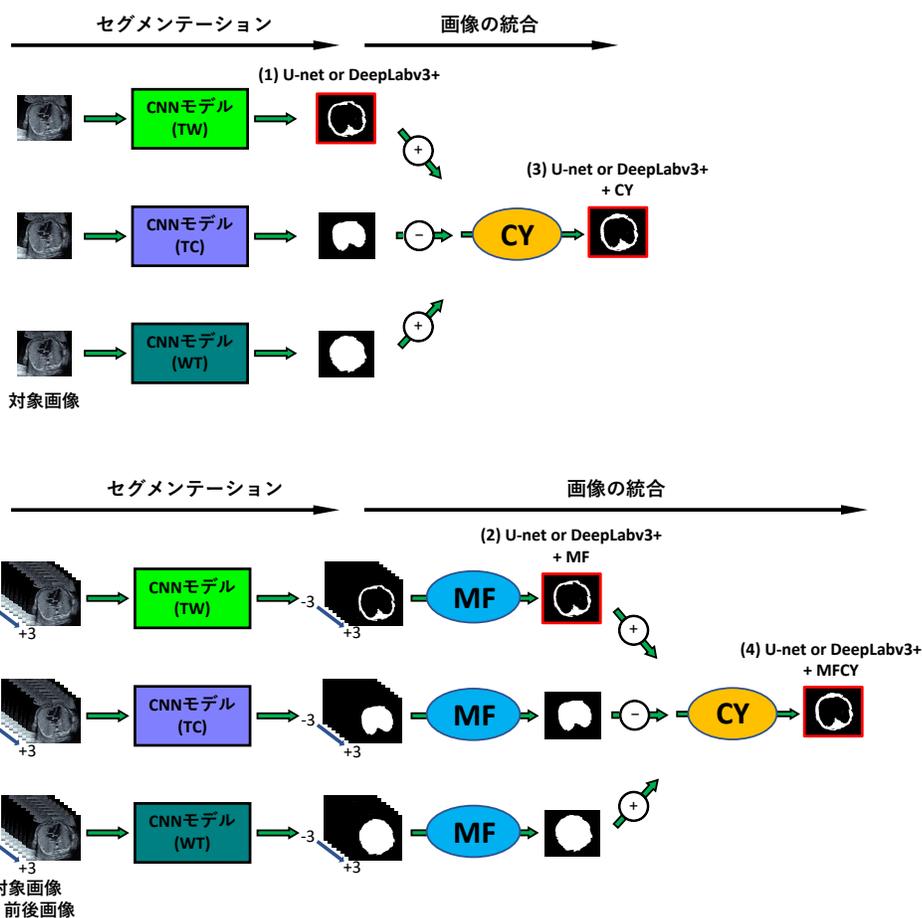


図 5. MF, CY, MFCY の性能評価を行うための実験図.

(1)U-net or DeepLabv3+, (2)U-net or DeepLabv3+ + MF, (3)U-net or DeepLabv3+ + CY, (4)U-net or DeepLabv3+ + MFCY の 4 つの手法のセグメンテーション性能を比較した. (1)の予測ラベルは, U-net or DeepLabv3+の既存モデルを単純に使用することで得た. (2)MF の予測ラベルは, 対象画像とその前後フレームから切り出した胸壁の予測ラベルを統合して得た. (3)CY の予測ラベルは, 胸壁, 胸腔, 胸郭全体からセグメンテーションした 3 種類の予測ラベルを統合して得た. (4)MFCY の予測ラベルは, MF と CY の順に適用させて得た.

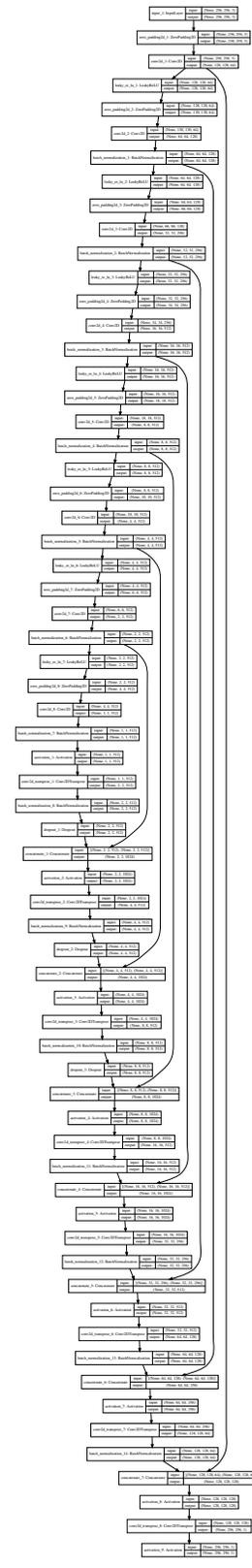


図 6. U-net の構造.

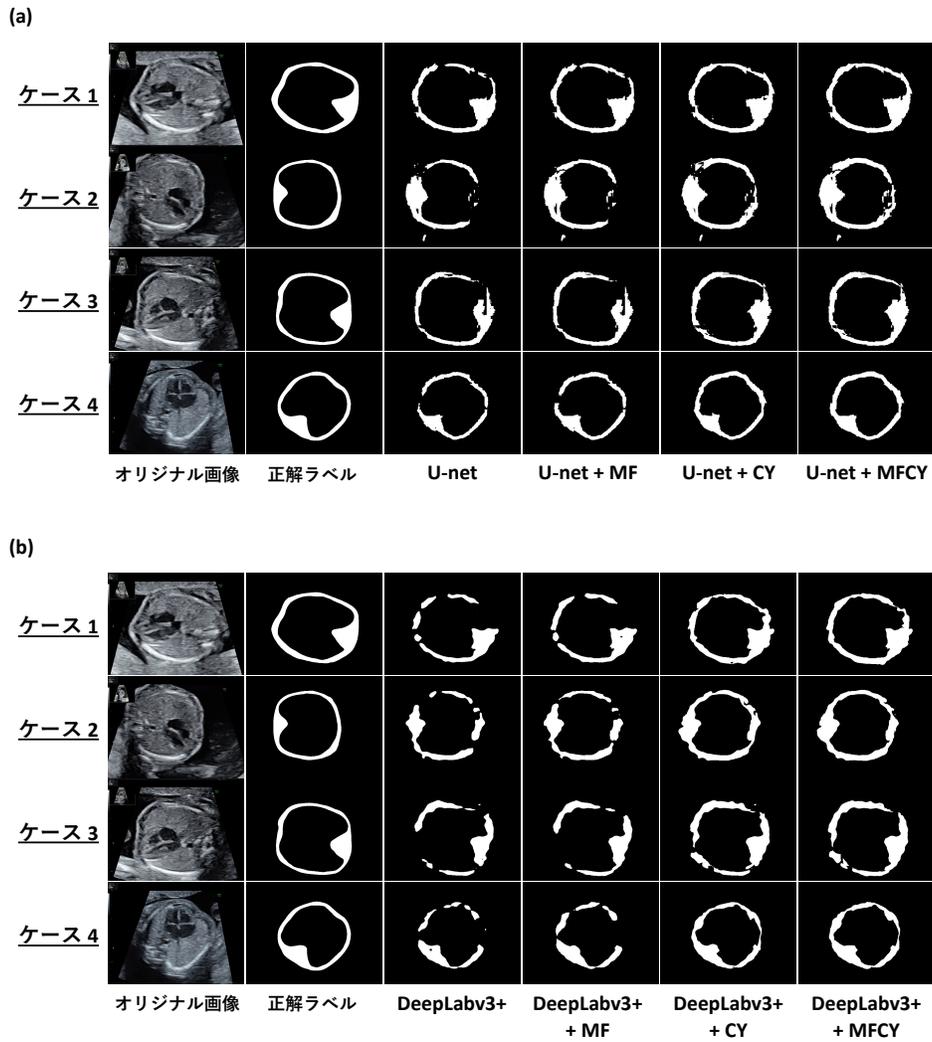


図7. 既存モデルと提案手法による胸壁セグメンテーション結果の代表例.

各行は、特定のケースを示している。各列は、オリジナル画像、正解ラベル、およびU-net, DeepLabv3+, MF, CY, MFCYによる予測に対応している。白塗り部分は胸壁のラベルを表し、(a)はU-net, (b)はDeepLabv3+をそれぞれCNNとして使用した結果である。

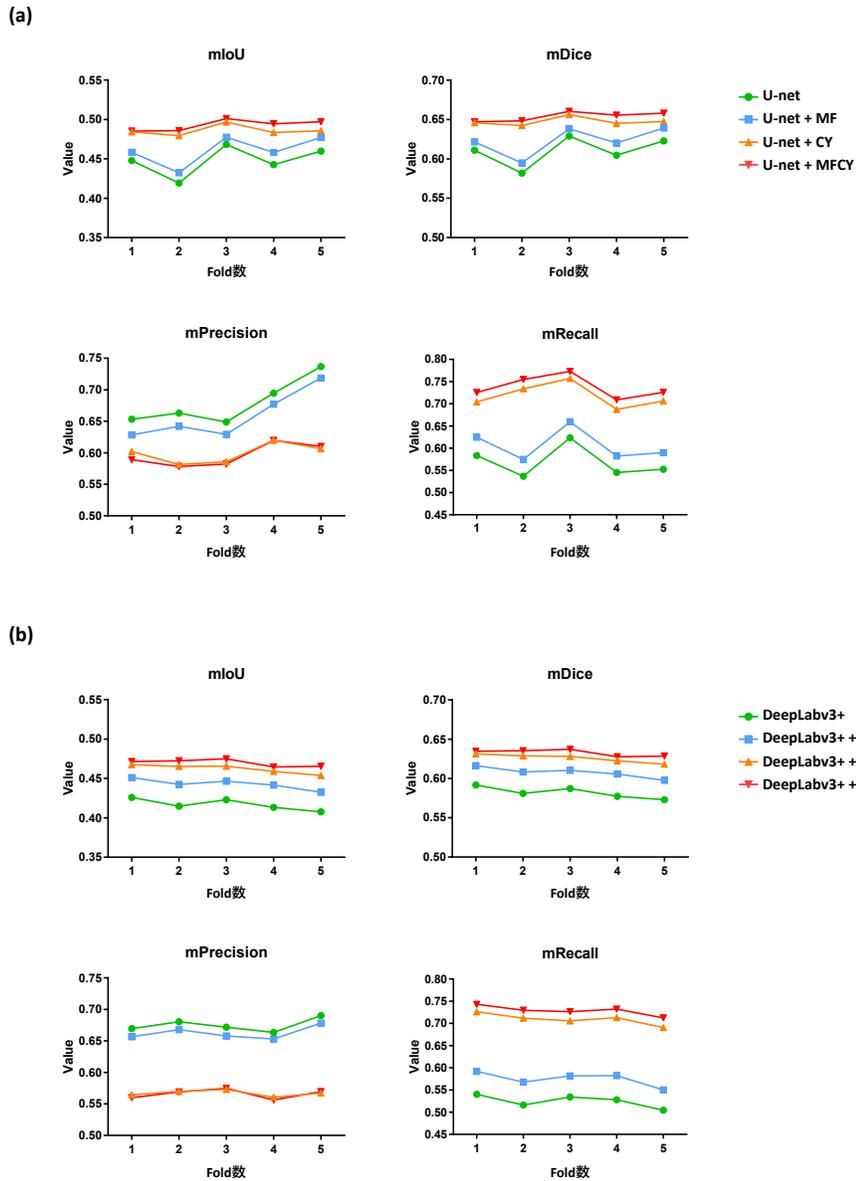
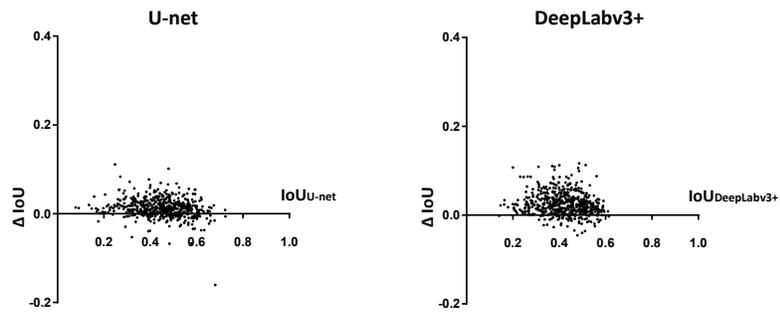


図 8. 各 fold での既存モデルと提案手法による胸壁セグメンテーション結果.

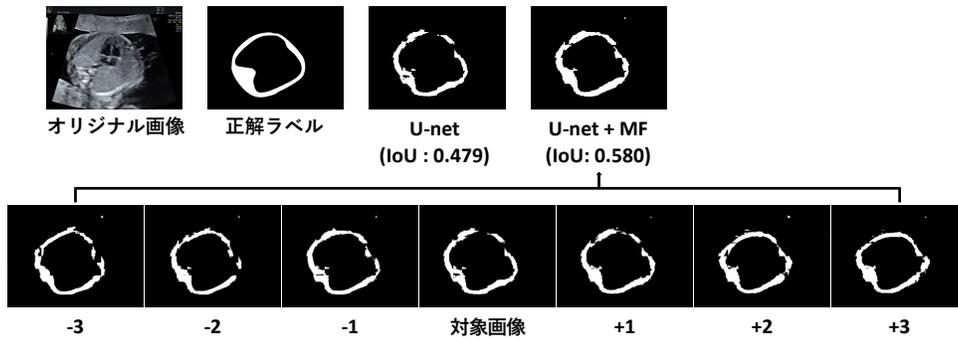
MF, CY および MFCY を使用した際の各 fold の平均 IoU (mIoU), 平均 Dice 係数 (mDice), 平均適合率 (mPrecision), 平均再現率 (mRecall)を示す. 各図では, X 軸が fold 数, Y 軸が指標の値を表している. CNN モデルは, (a)U-net と (b)DeepLabv3+を用いた結果である.

(a)

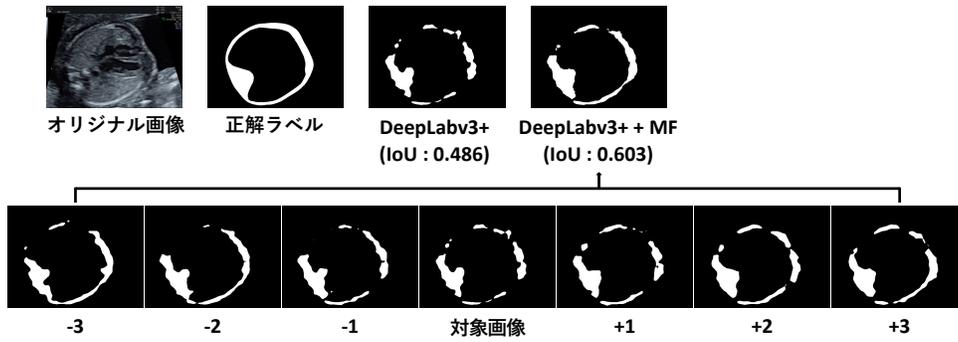


(b)

ケース 5 (U-net)



ケース 6 (DeepLabv3+)



ケース 7 (DeepLabv3+)

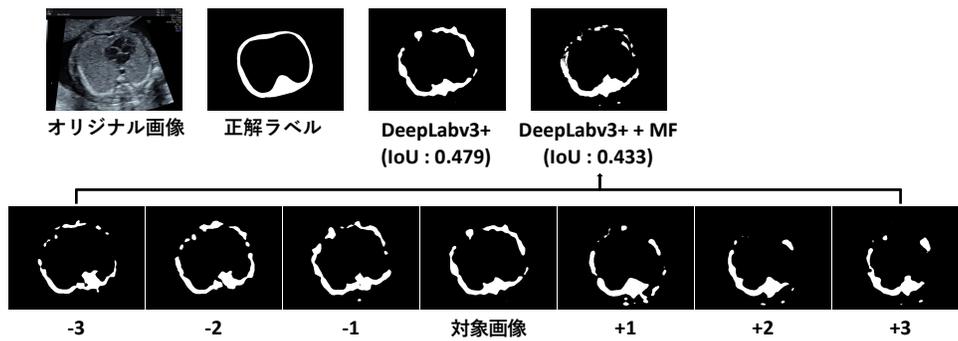
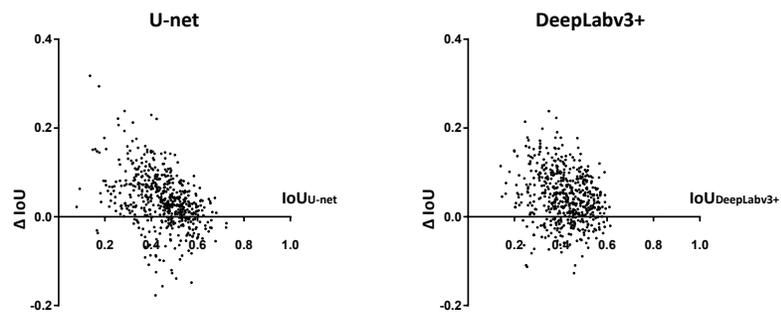


図 9. MF 適用時の IoU 指標を用いた胸壁のセグメンテーションの分析.

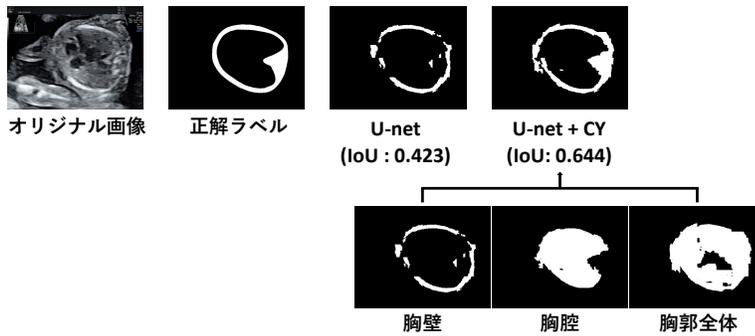
(a) U-net または DeepLabv3+ と MF を用いた全個別画像の IoU 値の差分の分布. U-net または DeepLabv3+ の IoU ($\text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$), MF の IoU (IoU_{MF}), およびそれらの IoU 値の差 ($\Delta \text{IoU} = \text{IoU}_{\text{MF}} - \text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$) を計算した. x 軸は $\text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$, y 軸は ΔIoU を表す. (b) MF を用いた胸壁セグメンテーションの代表例. CNN モデルは, U-net (ケース 5) と DeepLabv3+ (ケース 6, 7) を使用した. 各ケースの上段は, U-net または DeepLabv3+ 単独と MF から得られた予測ラベルを示している. 下段は, MF で統合された対象画像とその前後フレームの予測ラベルを示している. ケース 5 と 6 では, MF によってセグメンテーション性能が向上した. 一方で, ケース 7 では, MF によってセグメンテーション性能が悪化した.

(a)

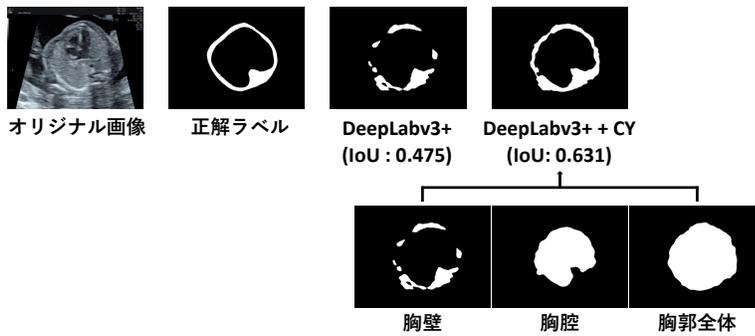


(b)

ケース 8 (U-net)



ケース 9 (DeepLabv3+)



ケース 10 (U-net)

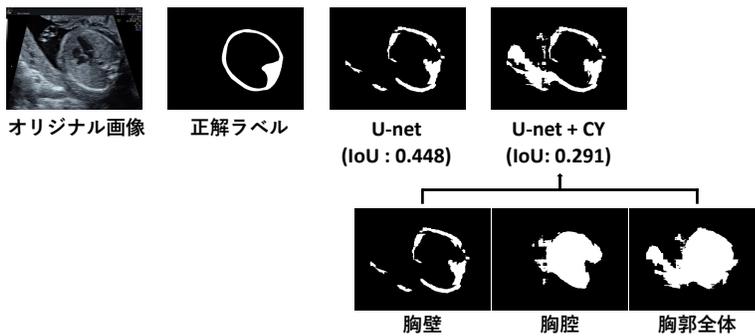


図 10. CY 適用時の IoU 指標を用いた胸壁セグメンテーションの分析.

(a) U-net または DeepLabv3+ と CY を用いた全個別画像の IoU 値の差分の分布. U-net または DeepLabv3+ の IoU ($\text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$), CY の IoU (IoU_{CY}), およびそれらの IoU 値の差 ($\Delta \text{IoU} = \text{IoU}_{\text{CY}} - \text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$) を計算した. x 軸は $\text{IoU}_{\text{U-net}}$ または $\text{IoU}_{\text{DeepLabv3+}}$, y 軸は ΔIoU を表す. (b) CY を用いた胸壁セグメンテーションの代表例. CNN モデルは, U-net (ケース 8, 10) と DeepLabv3+ (ケース 9) を使用した. 各ケースの上段は, U-net または DeepLabv3+ 単独と CY から得られた予測ラベルを示している. 下の行は, CY によって統合された予測ラベルを示している. ケース 8 と 9 では, CY によってセグメンテーション性能が向上した. 一方, ケース 10 では, CY によってセグメンテーション性能が改悪した.

付録表 1. 既存モデルを使用した際の心臓, 肺, 胸壁セグメンテーション性能結果.

| CNN | 対象 | mIoU | | mDice | | mPrecision | | mRecall | |
|------------|----|-------|-------|-------|-------|------------|-------|---------|-------|
| | | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 | 平均 | 標準偏差 |
| U-net | 心臓 | 0.686 | 0.042 | 0.798 | 0.036 | 0.849 | 0.028 | 0.784 | 0.069 |
| | 肺 | 0.712 | 0.015 | 0.828 | 0.011 | 0.832 | 0.020 | 0.832 | 0.014 |
| | 胸壁 | 0.448 | 0.017 | 0.610 | 0.016 | 0.679 | 0.033 | 0.568 | 0.032 |
| DeepLabv3+ | 心臓 | 0.734 | 0.012 | 0.842 | 0.009 | 0.864 | 0.008 | 0.833 | 0.008 |
| | 肺 | 0.684 | 0.010 | 0.808 | 0.008 | 0.833 | 0.008 | 0.793 | 0.014 |
| | 胸壁 | 0.417 | 0.007 | 0.582 | 0.007 | 0.675 | 0.009 | 0.525 | 0.013 |

mIoU : 平均 IoU, mDice : 平均 Dice 係数, mPrecision : 平均適合率, mRecall : 平均再現率.