情報処理学会研究報告
IPSJ SIG Technical Report

Attention 機構を用いた深層学習による嚥下 4 次元 CT の自動領域分割手法

内田 裕也[1,a)]　大竹 義人[2]　佐藤 嘉伸[2]　菊地 貴博[3]　道脇 幸博[4]　井尻 敬[1]

概要：嚥下とは飲食物を飲み込む動作のことであり，嚥下機能の低下は誤嚥性肺炎など重篤な病気の原因になり得るため，嚥下機能の治療や解析は重要な課題である．また，X 線 CT 撮影を高速に繰り返すことで，人体内部を 4 次元的（時間 + 3 次元空間）に計測できる 4DCT 技術が実用可能となり，この 4DCT を嚥下動作の解析に活用する試みが行われている．本研究では，嚥下動作を撮影した 4DCT より嚥下関連器官を自動的に領域分割する手法の実現を目的とし，U-Net に Attention 機構を組み合わせた深層学習モデルを提案する．提案手法の有用性を確認するため，手作業で領域分割した 5 症例の 4DCT 画像を用意し，単純な U-Net および提案する深層学習モデルを用いて，食塊・軟口蓋・舌など 8 領域の分割を実施した．その結果，提案手法は，Dice 係数において精度向上が確認された．

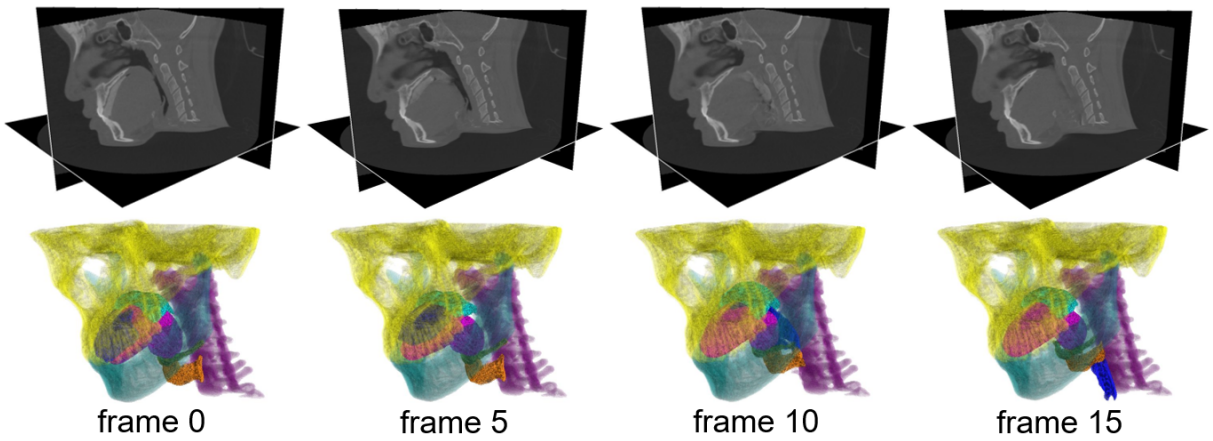

図 1: ある症例の入力 4DCT（上）と，提案手法により自動分割された 8 領域の分割結果（下）．

1. はじめに

嚥下とは，口腔内の飲食物を飲み込み，食道を経て胃へと送る動作のことである．嚥下機能の低下は，誤嚥性肺炎など重篤な病気の原因となり得るため，嚥下動作の解析手法の確立は重要な課題である [1–3]．この嚥下動作を詳細に解析するため，X 線 CT 撮影を高速に繰り返し人体内部を 4 次元的（時間＋ 3 次元空間）に計測する 4DCT 画像を活用する試みが行われている [4,5]．この 4DCT を用いた嚥下動作解析では，4DCT 画像から食塊・軟口蓋・舌・喉頭蓋・喉頭・舌骨・甲状軟骨など，嚥下関連器官を領域分割することが必要となる．しかし，一部の嚥下関連器官の境界は不明瞭であるため，また，4DCT は多数のフレームを含むため，その領域分割作業は多大な時間と手間を要する．

3 次元や 4 次元の CT 画像から関心領域を分割するため様々な手法が提案されている．例えば，輪郭線制約配置による関心領域の分割 [6] や，テンプレート変形を行う手法 [7] が提案されている．しかし，これらの手法では，各フレーム・各領域において，手作業で制約を配置する必要がある．また，深層学習を活用することで全自動で関心領域を分割する手法が提案されている．既存の自動領域分割手法の多くは，U-Net [8,9] に基づくもので，これを嚥下 4DCT の領域分割に適用した報告もある [10]．しかし，分割対象領域の特徴によっては，背景にノイズがのる，予測領域が欠損するなど，精度が不十分となることがある．

本研究では，嚥下動作を撮影した 4DCT 画像より嚥下関連器官を自動的に領域分割する手法の実現を目的とし，

[1] 芝浦工業大学
[2] 奈良先端科学技術大学院大学
[3] 株式会社明治
[4] 株式会社みちわき研究所
[a)] ma23027@shibaura-it.ac.jp

U-Net に Attention 機構を組み込んだ深層学習モデルを提案する．具体的には，nnU-Net [9] のスキップ接続部分の各層に，Attention Gate [11, 12] を並列化して組み込んだモデルを提案する．これにより，複数の分割対象領域に対して，それぞれの領域に特化した異なる Attention を適用できるようになり，分割精度の向上が期待される．

提案手法の有用性を確認するため，手作業で領域分割した 5 症例の嚥下 4DCT 画像を用意し，単純な U-Net および提案する深層学習モデルを用いて，食塊・軟口蓋・舌・頭蓋骨・下顎骨・頸椎・舌骨・甲状軟骨の 8 領域の分割を実施した．その結果，提案手法は，一部領域の平均 Dice 係数において精度向上が確認された．図 1 に，ある症例の入力 4DCT 画像と，提案手法により自動分割された 8 領域（食塊・軟口蓋・舌・頭蓋骨・下顎骨・頸椎・舌骨・甲状軟骨）の分割結果を示す．また，Attention Gate を並列化したモデルと，Attention Gate を 1 つのみ用いたモデルにて分割精度を比較した．その結果，両モデルとの間に大きな精度の差は見られなかった．

2. 関連研究

2.1 深層学習による画像の自動領域分割

画像の自動領域分割は，自動運転やロボティクスなど，幅広い分野において必要不可欠な技術であり，深層学習に基づく Fully Convolutional Networks [13] や SegNet [14] などが提案され，高い精度を達成している．特に，Ronneberger らにより医用画像分割のために設計された U-Net [8] は，少量のデータセットであっても高い精度での自動領域分割が可能である．このことから，U-Net は，大規模なデータセットを集めるのが難しい医用画像において広く用いられており，様々な派生モデルの基盤となる．例えば，U-Net を 3 次元畳み込みに拡張した 3D U-Net [15] や V-Net [16] などが知られる．

本研究にて利用した nnU-Net [9] も U-Net をベースとしたネットワークモデルである．nnU-Net は，従来手動であったネットワークパラメータを，入力されたデータセットの特性や使用するハードウェアの条件から自動で決定し，最適な自動領域分割のパイプラインを提供する手法であり，高い汎用性と精度を誇ることから多くの自動領域分割のタスクで用いられている深層学習モデルである．nnU-Net は，2 次元畳み込みを行うモデル（2D-nnUNet），3 次元畳み込みを行うモデル（3D-nnUNet），入力データの解像度を下げて 3 次元畳み込みを行うモデルの，3 つの基本アーキテクチャがあり，ユーザがいずれかのアーキテクチャを選択することができる．

この nnU-Net を嚥下 4DCT の領域分割に適用した結果が報告されている．中谷ら [10] は，嚥下 4DCT 画像から嚥下関連器官を自動領域分割することを目的とし，nnU-Net をベースとして，2 次元畳み込みモデル，3 次元畳み込みモデル，3 次元データの連続したフレームを入力として 3 次元畳み込みを適用したモデル（3.5D 法）といった複数のモデルで嚥下関連器官の自動領域分割を実施した．しかし，分割対象領域の特徴によっては，背景にノイズがのる，予測領域が欠損するなど，精度が不十分となることがある．

2.2 Attention 機構

Attention 機構は，ニューラルネットワークにおいて，入力された時系列データの重要な部分に重み付けする機構である．この機構は，自然言語の機械翻訳タスクのため，Bahdanau らによって提案された [17]．当時の Recurrent Neural Network をベースとした Seq2seq [18] などの機械翻訳のモデルには，入力時系列データが長くなると精度が低下するという課題があった．この課題に対して Attention 機構は，入力時系列データに動的に重み付けをして重要部分を強調することで，翻訳精度の向上に大きく寄与した．また，Vaswani らが提案した Transformer [19] は，Attention 機構を中心に用いることで，計算量の減少と翻訳精度の向上に寄与した．

この Attention 機構は，自然言語処理だけでなく，画像処理にも応用される．特に，画像サイズに比べて分割対象領域が小さい場合，背景が過剰に抽出される偏ったモデルが生成されることがある [11, 16]．そのため，Attention 機構を用い，空間的に重要な部分に重み付けを行うことにより精度の向上が見込まれている．実際，Oktay らが提案した Attention U-Net [11] は，U-Net に Attention 機構を組み合わせることで，領域分割精度の向上に寄与した．

3. 提案手法

本研究では，嚥下動作を撮影した 4DCT 画像より嚥下関連器官を自動的に領域分割する手法の実現を目的とし，nnU-Net [9] をベースに，Attention U-Net [11] で用いられた Attention Gate [12] を組み込んだ深層学習モデルを提案する．このとき，nnU-Net のスキップ接続部分において，Attention Gate を並列化して組み込む．これにより，複数の分割対象領域に対して，それぞれの領域に特化した異なる部分に注目した学習，推論が可能になると考えられる．提案する深層学習モデルを図 2 に示す．本研究では，4DCT における各フレームの 3D 画像を 2D スライス画像に分割し，2 次元畳み込みにより処理を行う *2D-Att-nnUNet* と，4DCT における各フレームの 3D 画像を，3 次元畳み込みにより処理を行う *3D-Att-nnUNet* を提案する．

我々が利用する Attention Gate を図 3a に示す．この Attention Gate は，エンコーダからスキップ接続によって渡される特徴マップと，1 層深いデコーダから渡される特徴マップを入力とする．Attention Gate で計算される Attention は 1 つのチャンネルを持つ特徴マップとして生成される．この Attention マップと，エンコーダから

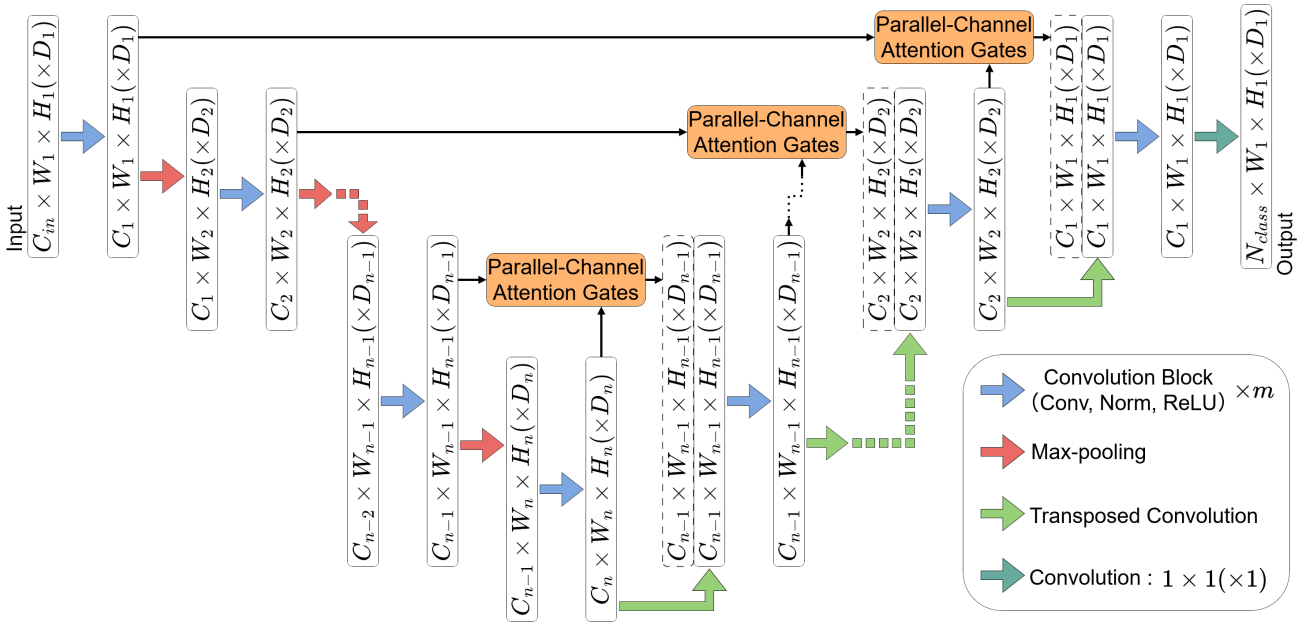

図 2: 提案する深層学習モデルのアーキテクチャ．Attention Gate は各スキップ接続に並列に配置 (Parallel-Channel Attention gates) され，スキップ接続を介して連結される特徴マップに対して重み付けを行う．なお，レイヤの深さ，Convolution Block の数，各畳み込みや Max-Pooling のパラメータ等は，nnU-Net の概念に基づき，データセットの特性や使用するハードウェアの条件により自動で決定される．図中の特徴マップサイズを表す $C \times W \times H(\times D)$ という表記は，2D-Att-nnUNet では $C \times W \times H$ であり，3D-Att-nnUNet では $C \times W \times H \times D$ であることを示す．

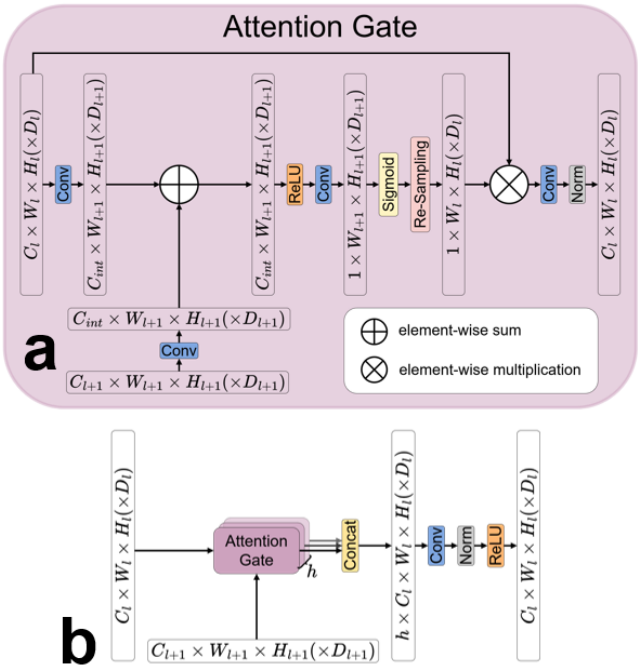

図 3: (a) Attention Gate のアーキテクチャ．(b) Parallel-Channel Attention gates のアーキテクチャ．nnU-Net に組み合わせる際には，複数の Attention Gate を並列に配置し，処理を行う．

スキップ接続によって渡される特徴マップの要素ごとの乗算（図中の element-wise multiplication）を行うことで，Attention マップにより重み付けされた特徴マップを出力する．

本研究では，この Attention Gate を，並列化して利用する（図 3b）．この構造を *Parallel-Channel Attention gates* と呼ぶこととする．並列数 h は，分割対象領域と同じ数（本研究では $h = 8$）とする．本研究では，食塊・軟口蓋・舌など，複数の分割対象領域を同時に扱うため，複数の Attention Gate を設けることで，それぞれの領域に特化した異なる Attention を適用できるようになると期待される．また，Attention Gate 内で使用する全ての正規化について，nnU-Net と同様に Instance Normalization [9, 20] を使用する．この Instance Normalization は，バッチ単位でなく，入力された個々のデータごとの平均と分散を用いて正規化する手法であり，バッチサイズに依存しない安定した学習が可能となる．特にバッチサイズが小さくなりがちな大規模なモデルや 3D データを扱う際に利用されることが多い正規化手法である．

4. 評価実験

4.1 実験内容

提案手法の精度評価のため，5 症例の嚥下動作を撮影し

た 4DCT 画像を用意した．各症例には 22 ∼ 31 フレームの 3DCT 画像が含まれており，5 症例のフレーム数の合計は 129 である．各フレームの解像度は 512 × 512 × 320 であり，ボクセル間隔 [mm] は 0.545 × 0.545 × 0.500，0.468 × 0.468 × 0.500，0.625 × 0.625 × 0.500 の 3 種類のいずれかである．また，全症例において，食塊・軟口蓋・舌・頭蓋骨・下顎骨・頸椎・舌骨・甲状軟骨の 8 領域について，専門家の手作業による領域分割が行われている．この領域分割マスクを正解データとして，学習および評価に利用する．

本研究では，leave-one-patient-out 交差検証 [10] を行う．つまり，5 症例のうち 4 症例を学習データ，残りの 1 症例をテストデータとする精度評価を，5 症例分回繰り返す．この評価方法を行う理由は，同一症例の異なるフレームは類似度が高く，同一症例のデータを学習データとテストデータの両方に入れると実際より高い精度が出てしまうためである．また，精度評価には，Dice 係数 $DSC_r(P_r, G_r)$ を用いる，

$$DSC_r(P_r, G_r) = \frac{2\,|P_r \cap G_r|}{|P_r| + |G_r|}. \qquad (1)$$

ただし，P_r は対象領域 r に対する予測領域，G_r は正解領域，$|P_r|$ は領域 P_r の体積を示す．

本研究では，上記の評価を 2D-nnUNet，2D-Att-nnUNet，3D-nnUNet，3D-Att-nnUNet という 4 通りのモデルで実施する．2D のモデルでは，4DCT 画像における各フレームの 3D 画像を 2D スライス画像に分割し，2 次元畳み込みにより処理を行う．一方，3D のモデルでは，4DCT 画像における各フレームの 3D 画像を，3 次元畳み込みにより処理を行う．また，2D-nnUNet/3D-nnUNet では，nnU-Net [9] をそのまま利用し，2D-Att-nnUNet/3D-Att-nnUNet では，提案手法である，Attention Gate を並列に組み込んだモデルを利用する．

我々は，AMD Ryzen 9 7950X，128GB RAM，NVIDIA GeForce RTX 3090 を搭載した計算機を利用して本実験を実施する．ネットワークアーキテクチャと学習条件は，nnU-Net の概念に基づき，学習データセットの特性や使用するハードウェアの条件から自動で決定される．表 1 に，本研究でのネットワークアーキテクチャと学習条件を示す．また，学習エポック数は 1000，学習率には初期学習率 0.01 の Polynomial Learning Rate [21] を，損失関数には Dice Loss [16] と Cross-entropy Loss の組み合わせを，最適化手法には momentum=0.99 の確率的勾配降下法を用いる．

表 1: ネットワークアーキテクチャと学習条件．C_1 は初期チャンネル数，C_n は最大チャンネル数，n はレイヤの深さを示す．

モデル	2D モデル	3D モデル
C_1	32	32
C_n	512	320
n	8	6
バッチサイズ	12	2
パッチサイズ	512 × 512	160 × 160 × 96

4.2 結果

交差検証を実施した結果を，各モデルおよび各領域における Dice 係数を箱ひげ図にまとめたものを図 4 に示す．また，各モデルおよび各領域における Dice 係数の平均値と標準偏差を表 2 に示す．2D モデルの場合は，軟口蓋・舌・下顎骨において，3D モデルの場合は，舌・下顎骨・頸椎において，提案手法が既存手法よりも平均値が高く，標準偏差が少ないため，安定した高い精度を示した．

2D モデルにおいて，提案手法の方が高い精度を示した軟口蓋と舌の予測結果の代表的なフレームを図 5 に示す．提案手法の方がより正しい形状を出力したことが確認できる．また，食塊と甲状軟骨において提案手法の Dice 係数の平均値は既存手法を下回ったが，これらの領域の分割精度はフレームによって大きく変化することが確認された．図 6 に，ある症例における 3 フレーム分の食塊と甲状軟骨の分割結果を示す．この症例の食塊領域において，2 フレームでは提案手法が高い精度を示した（図 6 緑円）．一方で，12 フレームにおいては既存手法のほうが高い精度を示した（図 6 赤円）．また，21 フレームの甲状軟骨領域では，提案手法が高い精度を示した（図 6 緑円）．

続いて，推論時の一番浅い層における Attention マップをヒートマップとして可視化した結果を図 7 に示す．提案手法は，分割対象領域の個数に応じて，8 個の Attention Gate を並列化して組み込むため，推論時には 8 種の Attention マップが得られる．図 7 は，ある症例におけるフレームの 2D-Att-nnUNet と 3D-Att-nnUNet の Attention マップである．いくつかの Attention マップ（図 7adfm など）は，頸椎・食塊・舌・舌骨など，別々の分割対象領域に対応する位置に強い重みを持つことがわかる．一方，背景や分割対象領域ではない部分に対して，重みを持つよう学習された Attention マップも確認される（図 7bhjl など）．

4.3 考察

単純な nnU-Net と，Attention 機構を組み込んだ nnU-Net の Dice 係数の結果を比較すると，軟口蓋・舌・下顎骨・頸椎といった領域において，Attention 機構の追加により精度の向上が確認された．一方，これ以外の領域においては，精度がほぼ変わらないか，かえって低下することが確認された．精度の向上が認められた，軟口蓋・舌・下顎骨・頸椎は，嚥下動作を通して形状や位置が大きく変化しないという特徴と，比較的大きく境界が明瞭であるという特徴を持つ．これより，Attention 機構は，動きが小さく境界が比較的明瞭な領域に対して有効に働く可能性が考

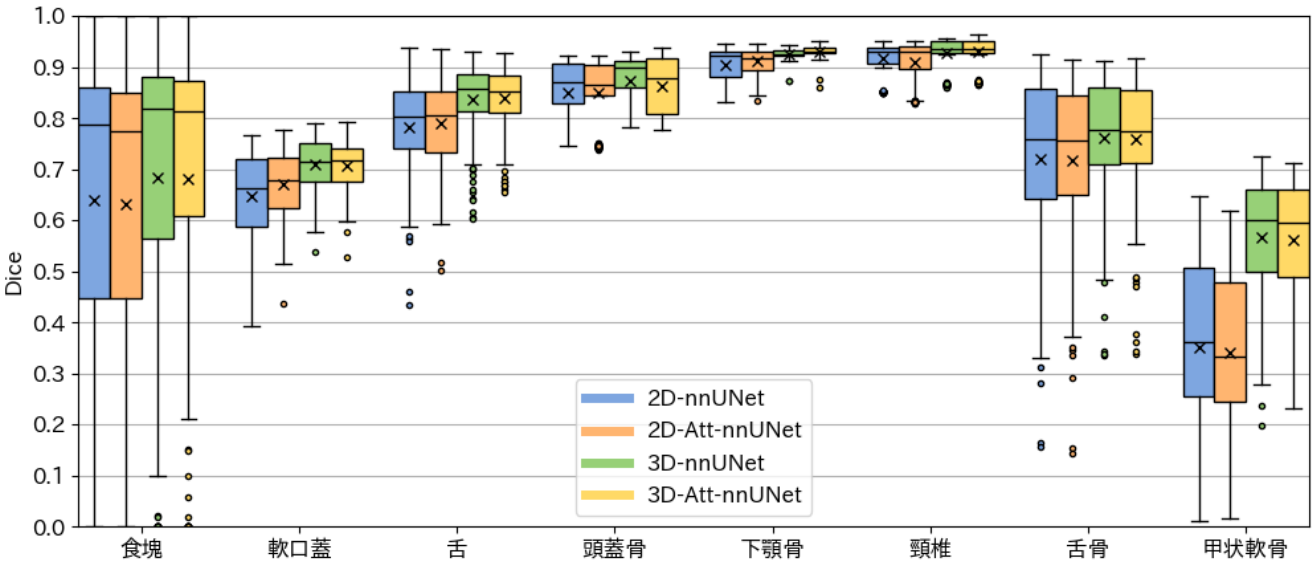

図 4: 各モデルおよび各領域ごとの Dice 係数の箱ひげ図.

表 2: 各モデルおよび各領域ごとの Dice 係数の平均と標準偏差．2D-nnUNet と 2D-Att-nnUNet，および 3D-nnUNet と 3D-Att-nnUNet において，精度が高い方を太字で示す．

モデル	2D-nnUNet	2D-Att-nnUNet	3D-nnUNet	3D-Att-nnUNet
食塊	**0.640 ± 0.299**	0.632 ± 0.303	**0.684 ± 0.286**	0.682 ± 0.288
軟口蓋	0.647 ± 0.089	**0.671 ± 0.060**	**0.710 ± 0.052**	0.707 ± 0.049
舌	0.783 ± 0.099	**0.789 ± 0.089**	0.836 ± 0.074	**0.838 ± 0.061**
頭蓋骨	**0.850 ± 0.060**	0.849 ± 0.062	**0.873 ± 0.049**	0.861 ± 0.053
下顎骨	0.904 ± 0.036	**0.912 ± 0.024**	0.926 ± 0.010	**0.931 ± 0.011**
頸椎	**0.916 ± 0.032**	0.910 ± 0.039	0.928 ± 0.030	**0.929 ± 0.029**
舌骨	**0.720 ± 0.165**	0.717 ± 0.164	**0.762 ± 0.127**	0.760 ± 0.126
甲状軟骨	**0.351 ± 0.178**	0.340 ± 0.155	**0.567 ± 0.118**	0.560 ± 0.119

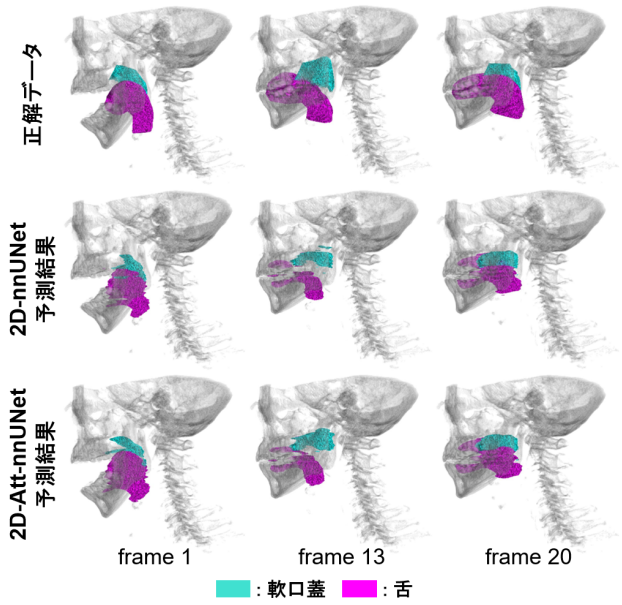

図 5: 軟口蓋と舌における予測結果の可視化．正解データ（上段），2D-nnUNet の予測結果（中段），2D-Att-nnUNet の予測結果（下段）．

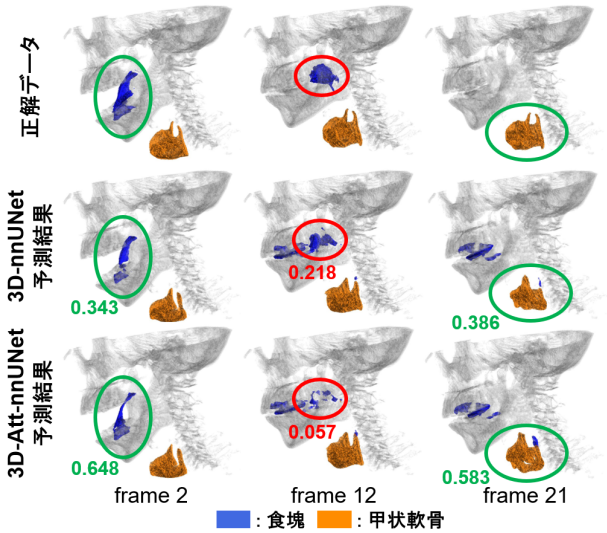

図 6: 食塊と甲状軟骨における予測結果の可視化．正解データ（上段），3D-nnUNet の予測結果（中段），3D-Att-nnUNet の予測結果（下段）．緑円は提案手法の方が Dice 係数が向上した領域，赤円は提案手法の方が Dice 係数が減少した領域を示す．数値は Dice 係数を示す．

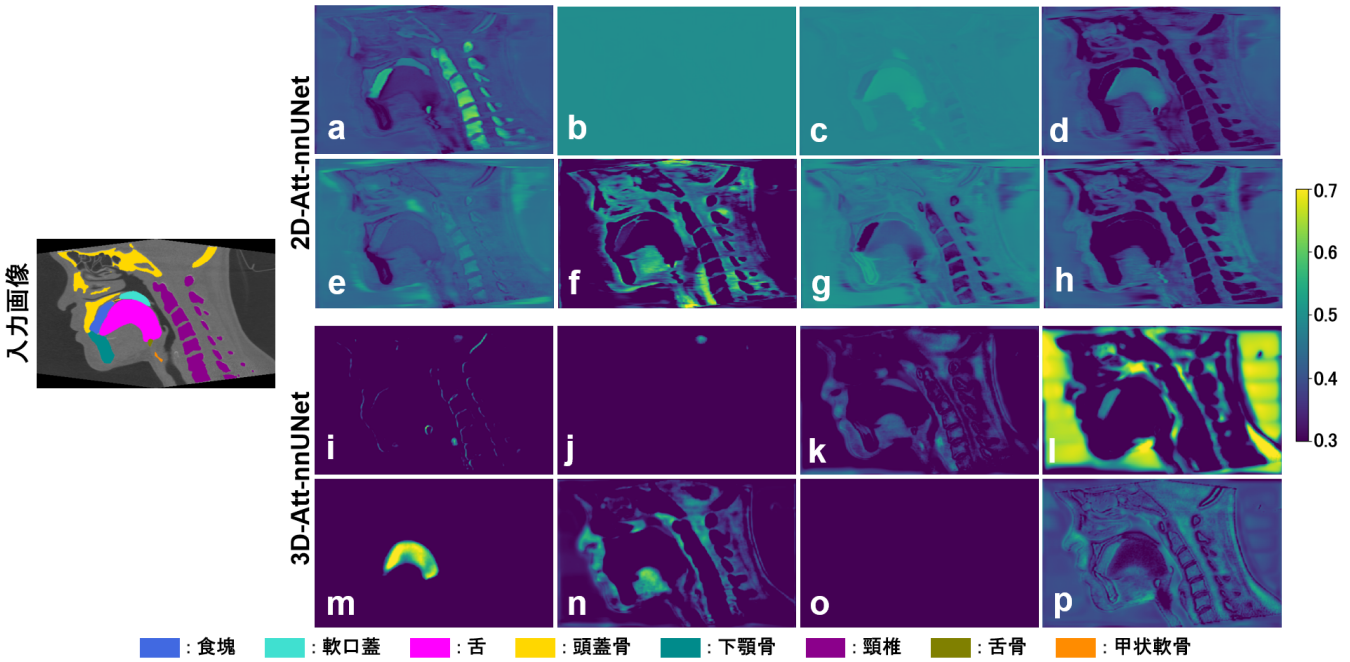

図 7: Attention マップの可視化．(a-h) は 2D-Att-nnUNet，(i-p) は 3D-Att-nnUNet における各 Attention Gate の Attention マップである．

表 3: 各モデルおよび各領域ごとの Dice 係数の平均と標準偏差．軟組織は軟口蓋と舌，骨は頭蓋骨・下顎骨・舌骨・甲状軟骨を含む．2D-SCAG と 2D-PCAG，および 3D-SCAG と 3D-PCAG において，精度が高い方を太字で示す．

モデル	2D-SCAG	2D-PCAG	3D-SCAG	3D-PCAG
食塊	0.623 ± 0.300	**0.632 ± 0.303**	0.682 ± 0.292	**0.682 ± 0.288**
軟組織（軟口蓋・舌）	**0.735 ± 0.098**	0.730 ± 0.096	0.773 ± 0.089	**0.773 ± 0.086**
骨（頭蓋骨・下顎骨・舌骨・甲状軟骨）	**0.756 ± 0.233**	0.746 ± 0.240	**0.811 ± 0.158**	0.808 ± 0.161

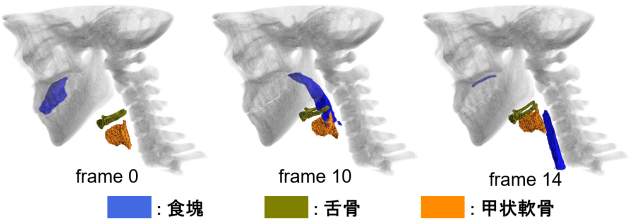

図 8: 食塊・舌骨・甲状軟骨の形状と位置の変化．

えられる．嚥下動作において，舌と軟口蓋は特に重要な役割を持つ組織であり，この 2 領域において精度向上が認められたことは，嚥下動作解析において重要な意味を持つといえる．

一方，食塊・舌骨・甲状軟骨といった領域については，Attention 機構の追加により精度の低下が確認された．これらの領域は，比較的小さな形状を持ち，嚥下動作中に大きく移動する．そのため，Attention 機構は，これらの領域に正しく重みを付けられず，誤ってノイズや他の大きな領域に過剰に反応してしまい，精度低下を起こしたと考えられる．図 8 に食塊・舌骨・甲状軟骨の形状と位置の変化を示す．嚥下中，食塊は口腔内から食道へ移動し，舌骨と甲状軟骨は上下に移動することが分かる．このような領域については，現状，Attention 機構による精度向上は確認できなかった．

頭蓋骨において Attention 機構が逆効果となった理由として，Attention 機構が冗長であった可能性が考えられる．頭蓋骨は非常に明瞭な境界を持ち，ほぼ動かない領域であるため，容易に自動領域分割が可能である．このため，Attention 機構による強調の必要性が低く，Attention 機構の導入により，逆に不要な領域やノイズが強調されてしまったと考えられる．

4.4 Ablation Study

Attention 機構並列化の効果を確認するため追加実験を行った．Attention Gate を 1 つのみ使用した Single-Channel Attention gate (SCAG) モデルと，Attention Gate を並列化した Parallel-Channel Attention Gates (PCAG) モデルについて，分割精度を比較した．結果を表 3 に示す．また，SCAG モデルの，推論時の一番浅い層における Attention マップの一部を図 9 に示す．

2D モデルでは，食塊領域ついては PCAG がわずかに良い精度を示し，軟組織と骨領域については SCAG がわずかに良い精度を示した．また，3D モデルでは，SCAG と

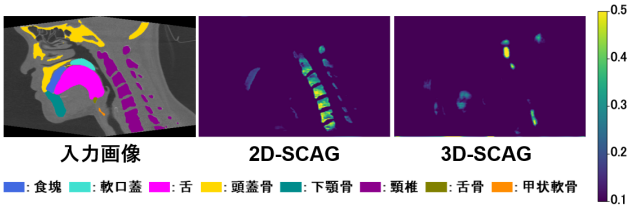

図 9: Attention Gate を 1 つのみを用いたモデル (SCAG) の Attention マップの可視化.

PCAG の間にほぼ差が見られなかった．Attention マップの可視化結果（図 9）を確認すると，2D モデルは頸椎と食塊領域のみに重みづけがなされていることが分かり，3D モデルでは分割対象領域ではない部分に対する誤った重み付けがなされてしまっていることが確認された．

5. まとめと展望

本研究では，嚥下動作を撮影した 4DCT 画像より嚥下関連器官を自動的に領域分割する手法の実現を目的とし，nnU-Net に Attention 機構を並列に組み込んだ深層学習モデルを提案した．実験結果から，移動や動きが比較的小さい，軟口蓋・舌・下顎骨・頸椎といった領域について，提案手法による精度向上が確認された．一方，移動や動きの大きい，食塊・舌骨・甲状軟骨といった領域に対しては，提案手法が逆に精度低下を起こす可能性も確認された．また，Attention マップを可視化した結果，分割対象領域に対して重み付けがなされていることが確認されたが，誤った部分に対して重み付けがなされてしまうケースも確認された．本研究の実験結果より，嚥下動作を撮影した 4DCT 画像の自動領域分割に対する Attention 機構の導入は，分割対象領域の移動や変形に依存し，特定の領域に対しては効果的である一方，領域によっては逆効果となってしまう可能性が示唆された．

今後の展望として，小さく，移動や変形の大きな領域に対しても重み付けがなされるような Attention 機構の開発が挙げられる．また，nnU-Net と同様に，入力されたデータセットの特性に基づいて動的に Attention 機構の構造を最適化する手法の確立も重要な課題である．さらに，より大規模なデータセットを用いて提案手法を訓練し，精度向上を目指すことも重要な将来課題である．

謝辞 本研究は JSPS 科研費 23K25199，23K28478，および R6 藤田研究開発課題（シーズ A 相当）の助成を受けて実施されたものである．また，本研究は，芝浦工業大学生命工学研究倫理審査委員会の承認 (22 − 007) を受けて実施されたものである．

参考文献

[1] 道脇幸博，齋藤真由，丹生かず代，小澤素子，南雲正男，角保徳，本多康聡：四次元 MRI の矢状断画像による嚥下運動の観察，日本口腔科学会雑誌， Vol. 54, No. 3, pp. 309–315 (2005).

[2] Michiwaki, Y., Kamiya, T., Kikuchi, T., Toyama, Y., Takai, M., Hanyu, K., Inoue, M., Yahiro, N. and Koshizuka, S.: Realistic computer simulation of bolus flow during swallowing, *Food Hydrocolloids*, Vol. 108, p. 106040 (2020).

[3] Kikuchi, T., Michiwaki, Y. and Azegami, H.: Identification of muscle activities involved in hyoid bone movement during swallowing using computer simulation, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Vol. 11, No. 5, pp. 1791–1802 (2023).

[4] Shibata, S., Kagaya, H., Inamoto, Y., Saitoh, E., Okada, S., Ota, K. and Kanamori, D.: Swallowing maneuver analysis using 320-row area detector computed tomography (320-ADCT), *Japanese Journal of Comprehensive Rehabilitation Science*, Vol. 2, pp. 54–62 (2011).

[5] Inamoto, Y., Fujii, N., Saitoh, E., Baba, M., Okada, S., Katada, K., Ozeki, Y., Kanamori, D. and Palmer, J. B.: Evaluation of swallowing using 320-detector-row multislice CT. Part II: kinematic analysis of laryngeal closure during normal swallowing, *Dysphagia*, Vol. 26, pp. 209–217 (2011).

[6] Ijiri, T., Yoshizawa, S., Sato, Y., Ito, M. and Yokota, H.: Bilateral Hermite Radial Basis Functions for Contour-based Volume Segmentation, *Computer Graphics Forum*, Vol. 32, No. 2, pp. 123–132 (2013). Proc. of EUROGRAPHICS'13.

[7] Kimura, Y., Ijiri, T., Inamoto, Y., Hashimoto, T. and Michiwaki, Y.: Interactive segmentation with curve-based template deformation for spatiotemporal computed tomography ofswallowing motion, *PLOS ONE*, Vol. 19, No. 10, p. e0309379 (2024).

[8] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional networks for biomedical image segmentation, *Medical image computing and computer-Assisted Intervention–MICCAI 2015*, Springer, pp. 234–241 (2015).

[9] Isensee, F., Jäger, P. F., Kohl, S. A. A., Petersen, J. and Maier-Hein, K. H.: Automated design of deep learning methods for biomedical image segmentation, *arXiv preprint arXiv:1904.08128* (2019).

[10] 中谷亮太，政木勇人，大竹義人，Yi, G.，Soufi, M.，菊地貴博， 井尻敬，道脇幸博，佐藤嘉伸：被験者個別の嚥下動態解析を目的とした 4DCT の自動セグメンテーション，第 32 回日本コンピュータ外科学会大会 (2023).

[11] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. and Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas, *Medical Imaging with Deep Learning (MIDL)* (2018).

[12] Schlemper, J., Oktay, O., Chen, L., Matthew, J., Knight, C., Kainz, B., Glocker, B. and Rueckert, D.: Attention-gated networks for improving ultrasound scan plane detection, *Medical Imaging with Deep Learning (MIDL)* (2018).

[13] Long, J., Shelhamer, E. and Darrell, T.: Fully convolutional networks for semantic segmentation, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440 (2015).

[14] Badrinarayanan, V., Kendall, A. and Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 39, No. 12,

pp. 2481–2495 (2017).
[15] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. and Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation, *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016*, Springer, pp. 424–432 (2016).
[16] Milletari, F., Navab, N. and Ahmadi, S.-A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation, *2016 fourth international conference on 3D vision (3DV)*, IEEE, pp. 565–571 (2016).
[17] Bahdanau, D., Cho, K. and Bengio, Y.: Neural machine translation by jointly learning to align and translate, *3rd International Conference on Learning Representations (ICLR)* (2015).
[18] Sutskever, I., Vinyals, O. and Le, Q. V.: Sequence to Sequence Learning with Neural Networks, *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)*, p. 3104–3112 (2014).
[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.: Attention is all you need, *Advances in Neural Information Processing Systems* (2017).
[20] Ulyanov, D., Vedaldi, A. and Lempitsky, V.: Instance Normalization: The missing ingredient for fast stylization, *arXiv preprint arXiv:1607.08022* (2016).
[21] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 40, No. 4, pp. 834–848 (2017).