

修士論文

CT 画像を用いた下顎骨および咀嚼筋の 自動セグメンテーションおよび統計解析

森谷 友香

奈良先端科学技術大学院大学

先端科学技術研究科

情報理工学プログラム

主指導教員: 佐藤 嘉伸 教授

生体医用画像研究室（情報科学領域）

令和 2 年 3 月 17 日提出

本論文は奈良先端科学技術大学院大学先端科学技術研究科に
修士 (工学) 授与の要件として提出した修士論文である。

森谷 友香

審査委員：

佐藤 嘉伸 教授	(主指導教員，情報科学領域)
加藤 博一 教授	(副指導教員，情報科学領域)
大竹 義人 准教授	(副指導教員，情報科学領域)
Mazen Soufi 助教	(副指導教員，情報科学領域)
日朝 祐太 助教	(副指導教員，情報科学領域)

CT 画像を用いた下顎骨および咀嚼筋の 自動セグメンテーションおよび統計解析*

森谷 友香

内容梗概

顎口腔機能異常の画像検査では、筋骨格の位置や形状について、変形後の確認および変形前の予測が必要とされる。また、現在の臨床現場では大規模な臨床データに基づく定量的な治療指標が少なく、新たな定量的指標を検討することが求められている。そのため、患者個別の筋骨格モデルの構築を行うことが臨床的に有用である。下顎骨や咬筋のセグメンテーションや、下顎骨上のランドマーク検出といった研究は既に進められているが、歯科充填物等の金属物質の影響により CT 画像上に重度の金属アーチファクトが生じるとその精度は低下する。一方、金属アーチファクト低減の手法は既に提案されてきたが、金属領域を正確に識別できない場合、性能が著しく低下してしまう。本研究では、実臨床 CT 画像データベース中の CT 画像を用いた患者個別の筋骨格モデル構築およびその統計解析を行うため、従来の金属アーチファクト低減手法を改良した提案手法とランドマーク自動検出を前処理として用い、筋骨格セグメンテーションを行う手法を提案する。この結果、実臨床 CT 画像データベース (742 症例) において顎頭間距離と筋骨格体積の計測ができ、性別により有意差のある検証結果を得た。したがって、提案手法による対象領域の自動セグメンテーションとその統計解析の可能性を示すことができた。

キーワード

下顎骨, 咀嚼筋, CT 画像, 金属アーチファクト低減, 自動セグメンテーション, 畳み込みニューラルネットワーク, 3D u-net, 統計解析

*奈良先端科学技術大学院大学 先端科学技術研究科 修士論文, 令和 2 年 3 月 17 日.

Automatic segmentation and statistical analysis of mandible and masticatory muscles in CT images*

Yuka Moritani

Abstract

The verification and prediction of the position and shape of mandible and masticatory muscles in computed tomography (CT) images is essential for the diagnosis and treatment planning of maxillofacial disorders. However, no quantitative measures for characterizing these structures based on large populations have been established. For this, the segmentation of the maxillofacial musculoskeletal structures and detection of their landmarks in CT images are of a clinical importance. Several approaches were proposed for those tasks; however, the presence of a metal prosthesis, such as a dental filling, leads to metal artifacts in the CT images that affect the segmentation accuracy. In this research, I aimed at the construction of a patient-specific musculoskeletal model based on the segmentation of real clinical cases. In addition, I proposed an improvement of the conventional metal-artifact-reduction method. Finally, the condylar distance and volume of the muscles and mandible bone were estimated on real clinical cases (n=742). As a result, statistically significant differences between male and female was obtained. Thus, the proposed modeling approach has shown feasibility for segmentation and analysis of maxillofacial CT images.

Keywords:

mandible, masticatory muscles, CT images, metal artifact reduction, automatic segmentation, Convolutional Neural Network, 3D u-net, statistical analysis

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, March 17, 2020.

目次

図目次	v
第 1 章 はじめに	1
第 2 章 手法	3
2.1. 概要	3
2.2. 金属アーチファクト低減	6
2.3. ランドマーク自動検出	8
2.3.1. 概要	8
2.3.2. 学習フェーズ	8
2.3.3. 推論フェーズ	9
2.4. 下顎骨および咀嚼筋の自動セグメンテーション	10
2.4.1. 概要	10
2.4.2. 学習フェーズ	14
2.4.3. 推論フェーズ	14
2.5. 顎頭間距離および筋骨格体積計測	15
第 3 章 実験	16
3.1. 使用データ	16
3.2. 金属アーチファクトシミュレーションにて考慮するパラメータ . . .	18
3.3. 精度評価方法	19
3.3.1. 金属アーチファクト低減	19
3.3.2. ランドマーク自動検出	20
3.3.3. 下顎骨および咀嚼筋の自動セグメンテーション	20
第 4 章 実験結果	21
4.1. 金属アーチファクト低減	21
4.1.1. 予備実験：金属アーチファクトシミュレーションの適用結果	21

4.1.2.	実臨床 CT 画像データベースへの適用結果	23
4.2.	ランドマーク自動検出	24
4.2.1.	予備実験：交差検証による学習モデルの精度評価	24
4.2.2.	実臨床 CT 画像データベースへの適用結果	29
4.3.	下顎骨および咀嚼筋の自動セグメンテーション	31
4.3.1.	予備実験 1：交差検証による学習モデルの精度評価比較 . . .	31
4.3.2.	予備実験 2：金属アーチファクトシミュレーションによる 学習モデルの精度評価比較	36
4.3.3.	実臨床 CT 画像データベースへの適用結果	38
4.4.	下顎骨および咀嚼筋の統計解析	41
4.4.1.	予備実験：金属アーチファクトのない CT 画像での計測結果	41
4.4.2.	実臨床 CT 画像データベースへの適用結果	43
第 5 章	考察	50
5.1.	金属アーチファクト低減	50
5.2.	下顎骨および咀嚼筋の自動セグメンテーション	51
5.3.	下顎骨および咀嚼筋の統計解析	51
5.3.1.	予備実験の結果について	51
5.3.2.	本実験の結果について	52
第 6 章	総括	57
	謝辞	58
	参考文献	59
	発表リスト	62

図目次

2.1	本研究にて対象とする筋骨格およびランドマークと顎頭間距離 . . .	4
2.2	提案手法の流れ	5
2.3	NMAR 従来法適用時の失敗例および従来法と提案法の比較	6
2.4	ランドマーク自動検出のためのネットワーク構成	8
2.5	各アーキテクチャのネットワーク構成図	11
3.1	実臨床 CT 画像データベース内の CT 画像の z 方向の大きさの分布	18
3.2	シミュレーションにて考慮するパラメータとそのとき生成される CT 画像の例	19
4.1	シミュレーションにより mild artifact を生じさせた場合の低減結 果 (一例)	22
4.2	シミュレーションにより severe artifact を生じさせた場合の低減 結果 (一例)	22
4.3	金属アーチファクトのある CT 実画像の低減前後	23
4.4	入力 CT 画像の WL, WW を変更して学習を行ったときのランド マーク自動検出結果 (5-fold)	25
4.5	入力 CT 画像を WL = 300, WW = 2000 で学習を行ったときの 5-fold 交差検証によるランドマーク自動検出結果一覧	26
4.6	実臨床 CT 画像データベースにおけるランドマーク自動検出結果 (一部)	30
4.7	6 種類の CNN アーキテクチャによる筋骨格自動セグメンテーショ ン精度比較 (4-fold)	32
4.8	アーキテクチャ 1 で学習を行ったときの 4-fold 交差検証による筋 骨格自動セグメンテーション結果一覧	33
4.9	シミュレーションによるアーチファクト低減前後画像による筋骨 格セグメンテーション結果	37

4.10	実臨床 CT 画像データベースにおける筋骨格自動セグメンテーション結果およびランドマーク自動検出結果（一部）	39
4.11	金属アーチファクトのない CT 画像における顎頭間距離および各筋骨格体積の計測結果	42
4.12	実臨床 CT 画像データベースにおける顎頭間距離の計測結果	44
4.13	実臨床 CT 画像データベースにおける各筋骨格の体積の計測結果ヒストグラム（男女合計）	45
4.14	実臨床 CT 画像データベースにおける各筋骨格の男女別の体積計測結果	46
5.1	金属アーチファクトシミュレーション画像に提案法を適用し低減に失敗する例	50
5.3	関連研究 [20] と本研究の計測位置の違い	52
5.4	金属アーチファクトの残存による筋骨格自動セグメンテーションの失敗例	54
5.5	実臨床 CT 画像データベースにおける性別・年齢別の咬筋の体積計測結果	55
5.6	実臨床 CT 画像データベースにおける年齢と咬筋の体積の相関関係	56

第 1 章 はじめに

顎口腔系は、咬合、顎関節、咀嚼筋等により構成されている [1]。これらの構成要素において、種々の要因によりその形態や位置関係に異常が生じると、顎変形症や顎関節症といった顎口腔機能異常が引き起こされる [2]。顎口腔機能異常の診断・治療では、画像検査として、従来から単純 X 線やパノラマ X 線を用いることが主であったが、CT (Computed Tomography) や MRI (Magnetic Resonance Imaging) の普及に伴い、これらを日常診療に利用する一般医科・歯科開業医も急増している [3]。このような臨床現場において、コンピュータを用いて自動で画像処理を行い、その結果を医師に提示することによって医師の診断を支援するコンピュータ支援診断 (computer-aided diagnosis, CAD) システムの開発が進んでいる [4]。

顎口腔機能異常の画像検査では、機能異常により変形した筋骨格の位置や形状を確認することに加え、異常の生じる以前の咬合や歯列の状態を予測することが必要とされる。この際に考慮されるのは、下顎頭の形態変化と、それに伴う咬合平面 (左右大白歯の咬頭と中切歯の midpoint を結ぶ平面)、ボンウィル三角 (切歯点と下顎両側顆頭 (左右の下顎頭上面の中央部頂点: 下顎頭頂) を結ぶ三角)、バルクウィル角 (ボンウィル三角と咬合平面の成す角度) の変化などである。しかし、現在の臨床現場ではいくつかの治療指標はある一方で定量的な指標が少なく、大規模な実臨床 CT 画像を用いて新たな定量的指標を検討することが求められている。この課題解決のため、患者個別の筋骨格モデルの構築を自動で行うことが臨床的に有用であると考えられる。

筋骨格モデルの構築に関して、下顎骨のセグメンテーション [5][6] や咬筋のセグメンテーション [7]、下顎骨上のランドマーク検出 [8] といった研究が進められてきた。しかし、顎口腔領域では、金属製の歯科充填物等がある場合、CT 画像上に重度の金属アーチファクトが生じることにより筋骨格セグメンテーションの精度が低下するという課題がある [9]。金属アーチファクト低減の手法として、Meyer らの手法 [10] (nomalized metal artifact reduction, 以降 NMAR) 等が既に提案されており、金属アーチファクトの程度によっては良好な結果を示している。しかし、

金属領域を正確に識別できない場合，このような金属アーチファクト低減は性能が著しく低下してしまう．特に歯科の領域では，補綴材料の X 線の減衰が類似している歯のエナメル層や，金属領域の周囲に生じた金属アーチファクト部分を金属領域と誤認識してしまうことから，しばしば金属アーチファクトの性能が低下してしまう．このため，歯科や顎口腔の領域において金属アーチファクト低減手法を適用するためには，金属領域を正しく認識するための工夫が必要である．このような背景から，畳み込みニューラルネットワーク（Convolutional Neural Network，以降 CNN）を用いて金属領域を自動抽出する手法 [11] 等が提案されている．一方，筋骨格セグメンテーションでは，現在，CNN を用いた手法である u-net[12] および 3D u-net[13] が医用画像の分野で広く用いられている．また，ランドマーク検出では，CNN を用いた手法として，正解ランドマークの位置から作成した三次元ガウス分布のヒートマップを学習の正解データとする手法が提案されている [14][15]．

これらの背景より，本研究では，顎口腔機能異常の治療計画や術後経過の予測のための患者個別の筋骨格モデル構築を目的として，実臨床 CT 画像データベースにおける筋骨格自動セグメンテーションおよびランドマーク自動検出を提案し，提案手法により得られた結果の統計解析を行う．本研究で用いる実臨床 CT 画像データベースは，金属アーチファクトが生じているという上記の課題に加え，様々な疾患・症状のデータベースであることから症例によって CT 画像の撮影範囲が異なり，本研究にて対象としている筋骨格領域全てを含んでいるとは限らないという注意点がある．このため，本研究ではまず金属アーチファクト低減と撮影範囲の統一を行うことで，対象とする筋骨格の自動セグメンテーション精度の向上を目指した．

本稿では，第 2 章にて提案手法について述べる．第 3 章では実験に使用した 2 つのデータセットの詳細と金属アーチファクトシミュレーションのパラメータ，および各実験の精度評価方法について述べる．第 4 章では提案手法の精度検証結果および実臨床 CT 画像データベースに適用した結果について述べ，第 5 章にてそれを考察し，総括する．

第 2 章 手法

2.1. 概要

本研究にて焦点を当てている顎口腔機能異常の診断・治療では、機能異常により変形した筋骨格の位置や形状の確認が必要不可欠である。特に重要視されるのは、下顎頭の形態変化とそれに伴う歯列関連の平面と基準点の変化である。歯列関連の平面の一つとしてボンウィル三角が定義され、これを構成する 3 点のうち 2 点は左右の下顎頭頂である。この 2 点間の距離を顎頭間距離といい、平均的な距離は 10cm とされている [17] が、年齢や性別による相違については知られておらず、この統計解析が臨床的に有用である。また同時に、下顎骨と咀嚼筋の形態について、個々の患者での位置や形状の確認だけでなく、異常の生じる前後および年齢や性別による相違を統計解析により検証することが求められている。

提案手法では、金属アーチファクトのある CT 画像を入力とし、下顎骨上のランドマーク自動検出と、下顎骨および咀嚼筋の自動セグメンテーションを行う。また、その結果から、筋骨格の体積やランドマーク間距離を計測し、これらを出力とする。本研究にて対象とした筋骨格とランドマークについて、図 2.1 に示す。なお、咀嚼筋とは一般的に 4 種の筋肉を指すが、本研究ではこのうち金属アーチファクトによる影響を受ける位置に存在する咬筋と内側翼突筋の 2 種を対象としている。

提案手法の流れを図 2.2 に示す。まず、金属アーチファクトのある CT 画像において、金属アーチファクトを低減する。次に、この低減後画像を入力画像とし、CNN を用いて下顎骨上のランドマーク検出を行う。学習フェーズでは、金属アーチファクトのない CT 画像を入力画像とし、正解ランドマークの位置から作成した三次元ガウス分布のヒートマップを正解データとして学習する。このランドマーク自動検出の結果を用いて撮影範囲の統一を行った CT 画像を入力画像とし、CNN を用いて下顎骨および咀嚼筋の自動セグメンテーションを行う。この学習フェーズでは、入力画像としてランドマーク自動検出と同様に金属アーチファクトのない CT 画像を用い、正解ラベル画像を正解データとして学習する。最後に、以上によって得たランドマーク位置と筋骨格セグメンテーション結果を用い、顎頭間距離

および各筋骨格の体積を求める．次節以降，各ステップの詳細を説明する．

また，金属アーチファクトとその低減が筋骨格セグメンテーションに与える影響について検証するため，本研究では予備実験として金属アーチファクトのない CT 実画像上に金属アーチファクトを生じさせる金属アーチファクトシミュレーションを文献 [16] を参考にして行っている．このパラメータについては次章にて説明する．

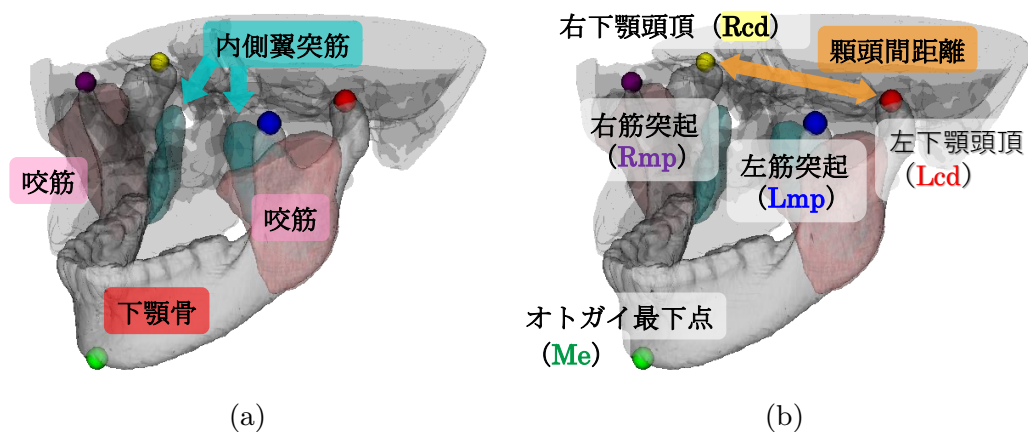


図 2.1: 本研究にて対象とする筋骨格およびランドマークと顎頭間距離．(a) 筋骨格：下顎骨，咀嚼筋（咬筋，内側翼突筋）．(b) ランドマーク 5 点：下顎頭頂（左右），筋突起（左右），オトガイ最下点．顎頭間距離：左右の下顎頭頂間の距離．

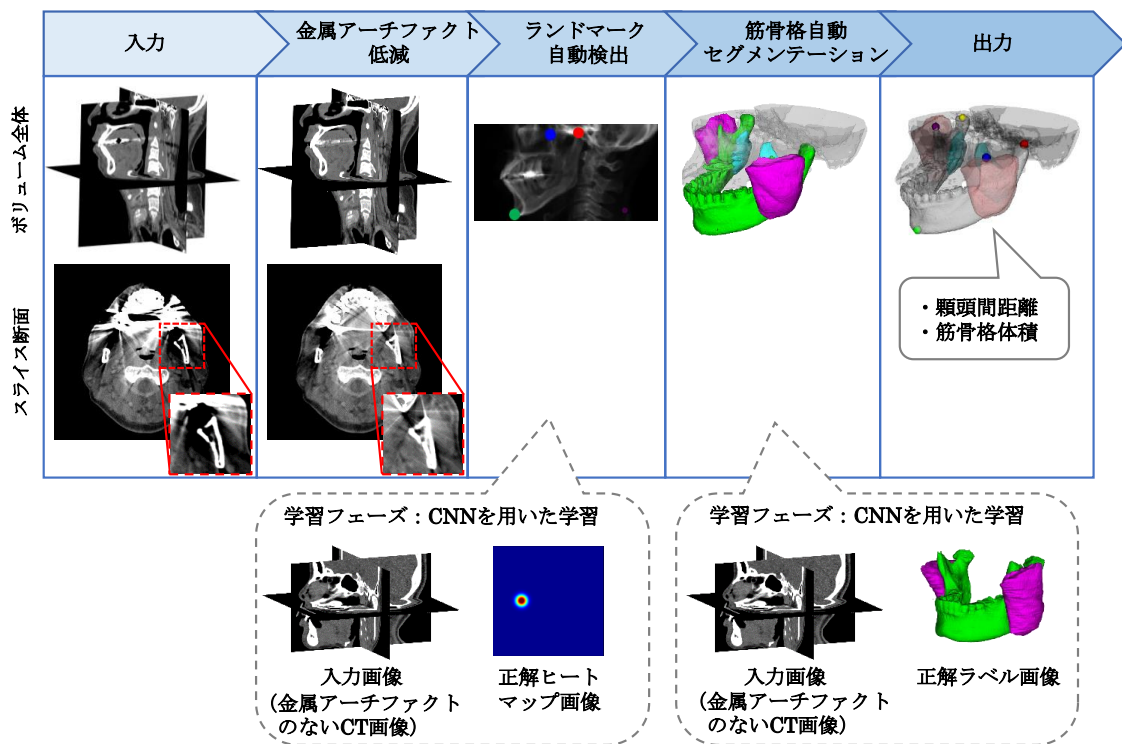


図 2.2: 提案手法の流れ

2.2. 金属アーチファクト低減

本研究では，金属アーチファクト低減手法の一つである NMAR[10] を一部変更した手法を提案し，これを CT 実画像に対して適用する．NMAR では金属領域の指定が必要であるが，撮影対象者の口腔内に存在する金属製の歯科充填物等の位置や形状の情報は既知でないため，金属領域はしきい値（4000HU）処理により推定する他ない．しかし，この推定によって得られた金属領域を NMAR にて使用すると，症例によっては図 2.3 に示すように NMAR に失敗してしまう場合があった．これは，しきい値により抽出した金属領域の一部と NMAR 適用時に生成される prior の空気領域の一部が図 2.3 の水色矢印で示すように重なった場合， $p^{norm} = p^{uncollected} / p^{prior}$ により算出した， p^{norm} の値が大きくなり，金属領域における p^{norm} の線形補間に影響が生じるためである．

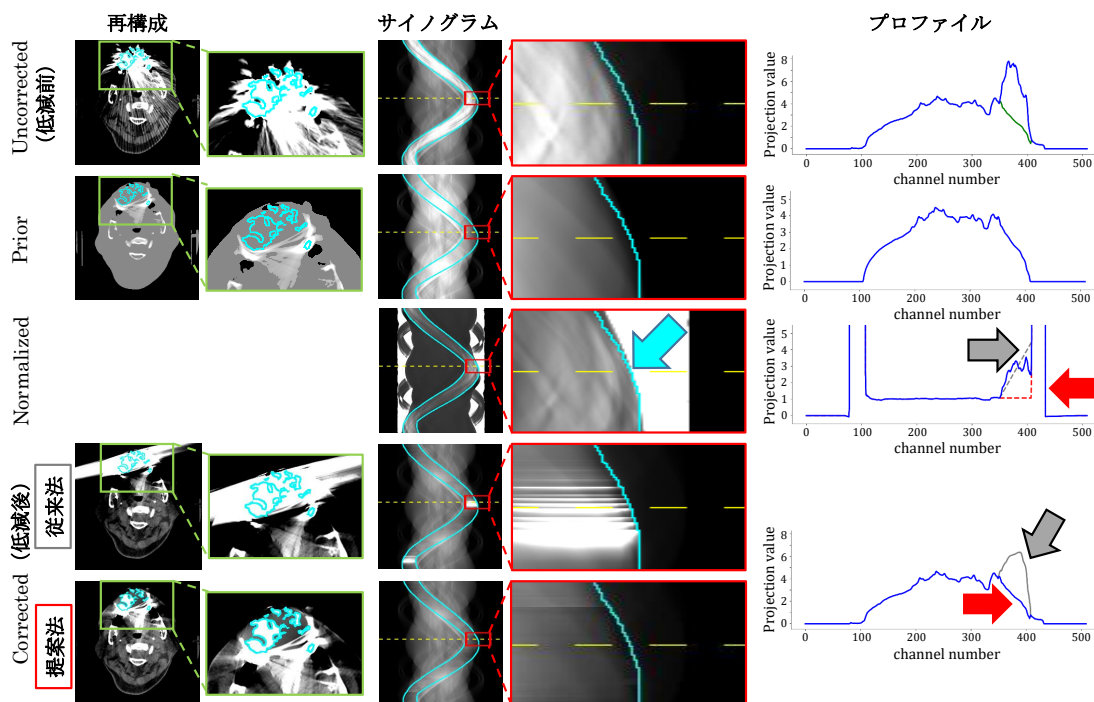


図 2.3: NMAR 従来法適用時の失敗例および従来法と提案法の比較．水色線で囲まれた領域：金属領域，灰色矢印：従来法適用時，赤色矢印：提案法適用時．

このため、NMAR 適用時に図 2.3 に示す問題が生じた場合、 p^{norm} における補間の方法とその範囲を変更する次のような処理を追加し、これを提案法とする。

1. Normalized サイノグラム上の値 p^{norm} について、 $p_c^{norm} > 3$ を満たす c が存在するかどうかを判定する。存在する場合、下記 2 に進む。存在しない場合、以下の処理は行わない。
2. サイノグラム上の金属領域（従来法での線形補間領域） $c_a \leq c \leq c_b (a < b)$ において、サイノグラム上の値 $p_{c_a}^{norm}, p_{c_b}^{norm}$ がそれぞれ $p_{c_a}^{norm} > 2$ あるいは $p_{c_b}^{norm} > 2$ を満たすかどうか判定する。 $p_{c_a}^{norm} > 2$ を満たす場合、下記 3(a) に、 $p_{c_b}^{norm} > 2$ を満たす場合、下記 3(b) に、それぞれ進む。
3. (a) 上記 2. で $p_{c_a}^{norm} > 2$ を満たす場合、 $c_a < c < c_b$ の範囲内で p_c^{norm} が局所的最小値をとる c_{min} をすべて求め、その中で c_a に最も近い位置を $c_{a'}$ 、 $p_{c_{a'}-1}^{norm}$ を $p_{c_{a'}-1}^{norm} = \min \{p_{c_b+1}^{norm}, 2\}$ とする。補間後の p_c^{norm} は次式で表せる。

$$p_c^{norm} = \begin{cases} p_c^{norm}, & c < c_{a'} - 1 \\ p_{c_{a'}-1}^{norm} + \frac{p_{c_b+1}^{norm} - p_{c_{a'}-1}^{norm}}{c_b - c_{a'} + 2}, & c_{a'} - 1 \leq c \leq c_b + 1 \\ p_c^{norm}, & c > c_b + 1 \end{cases} \quad (2.2.1)$$

- (b) 上記 2. で $p_{c_b}^{norm} > 2$ を満たす場合、 $c_a < c < c_b$ の範囲内で p_c^{norm} が局所的最小値をとる c_{min} をすべて求め、その中で c_b に最も近い位置を $c_{b'}$ 、 $p_{c_{b'}+1}^{norm}$ を $p_{c_{b'}+1}^{norm} = \min \{p_{c_a-1}^{norm}, 2\}$ とする。補間後の p_c^{norm} は次式で表せる。

$$p_c^{norm} = \begin{cases} p_c^{norm}, & c < c_a - 1 \\ p_{c_a-1}^{norm} + \frac{p_{c_{b'}+1}^{norm} - p_{c_a-1}^{norm}}{c_{b'} - c_a + 2}, & c_a - 1 \leq c \leq c_{b'} + 1 \\ p_c^{norm}, & c > c_{b'} + 1 \end{cases} \quad (2.2.2)$$

2.3. ランドマーク自動検出

2.3.1. 概要

ランドマーク自動検出の学習は、深さ 2 の Residual 3D u-net を用いる。本研究のランドマーク自動検出において最適化手法は Adabound[18]，損失関数は Focal Loss[19] を用いる。ネットワーク構造を図 2.4 に示す。

関連研究 [14] では、学習用 CT 画像データのうちいくつかの症例においてそのサイズが z 方向に大きいとき、 z 方向にランダムに画像をクロップして学習に用いている。本研究においても、撮影範囲が不揃いである実臨床 CT 画像データベースに適用することを踏まえ、パッチベースの学習・推論を行う。

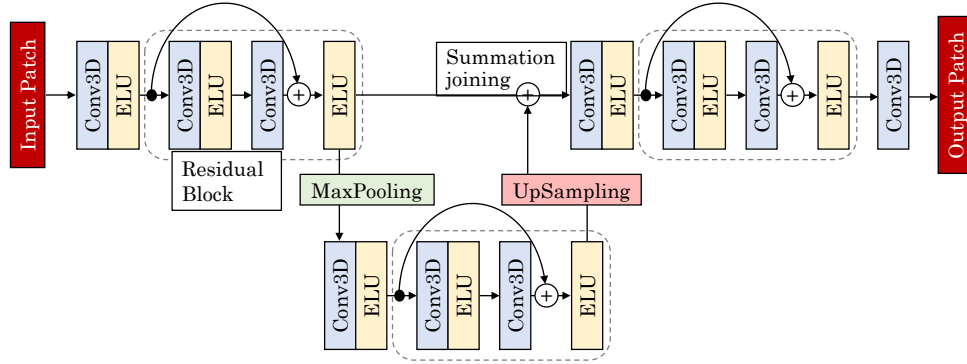


図 2.4: ランドマーク自動検出のためのネットワーク構成

2.3.2. 学習フェーズ

本件研究では、 $512 \times 512 \times z$ voxel (z は症例により異なる) の CT 画像データを用いる。このような 1 ボリュームについて、1voxel あたりのサイズ $s_x \times s_y \times s_z$ mm^3 を $4s_x \times 4s_x \times 4s_x$ mm^3 にリサンプリングすると、1 ボリュームのサイズは $128 \times 128 \times (zs_z/4s_x)$ voxel となる。このボリュームを、 $128 \times 128 \times 32$ voxel の大きさになるよう、 z 方向に 32voxel ずつ重なりのないようパッチを取得していく。 $z \bmod 32 \neq 0$ の場合、逆側方向からも同様にクロップする。このように作成した

大きさ $128 \times 128 \times 32$ voxel のパッチを入力画像とし、学習を行う。

正解データについては、次のように作成する。まず、手動で指定したランドマーク $L_i (1 \leq i \leq 5)$ の正解座標 $\mathbf{x}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$ を μ とし、 σ を 7 として、各ランドマークの三次元ガウス分布（ヒートマップ）を作成する。これを正解データとして入力画像と同様にリサンプリングおよびクロップを行い、学習に用いる。

2.3.3. 推論フェーズ

入力画像は、まず学習フェーズと同様のリサンプリングを行い、その後、 $128 \times 128 \times 32$ voxel の大きさのパッチを生成する。ここで、パッチの取り方は z 方向に 4voxel ずつずらし、 $128 \times 128 \times 28$ voxel 分は直前のパッチと被るようにする。これは、本研究にて学習フェーズで使用するデータセットは撮影範囲が概ね統一されているが、推論フェーズにて使用する実臨床 CT 画像データベースは撮影範囲が不揃いであることから学習フェーズと同様のパッチの取り方では推論精度が低下する可能性があるためである。

推論後、得られたヒートマップからランドマーク位置座標を次のように求める。1 症例あたりの入力画像のパッチ数を n とすると、推論時の出力では、1 症例 1 ランドマークあたり n 個のヒートマップを得ることとなる。この n 個のヒートマップのうちの 1 つを $h_k (1 \leq k \leq n)$ と表したとき、1 症例 1 ランドマークあたりのヒートマップの最高値 $value_i$ は次のように求められる。

$$value_i = \arg \max_k \left\{ \arg \max_{x,y,z} r(x,y,z) \in h_k \right\} \quad (2.3.1)$$

$value_i$ が存在するパッチについて、1CT 画像全体 $128 \times 128 \times (zs_z/4s_x)$ voxel 中のどの位置であるかを合わせて周りをゼロで埋めてから、元のサイズ $512 \times 512 \times z$ voxel にアップサンプリングする。このボリュームにおける最大値を confidence value $value_{\mathbf{x}_i}^c$ と定義し、 $value_{\mathbf{x}_i}^c$ が存在する座標をランドマーク自動検出座標 $\hat{\mathbf{x}}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i) \in \mathbb{R}^3$ とする。

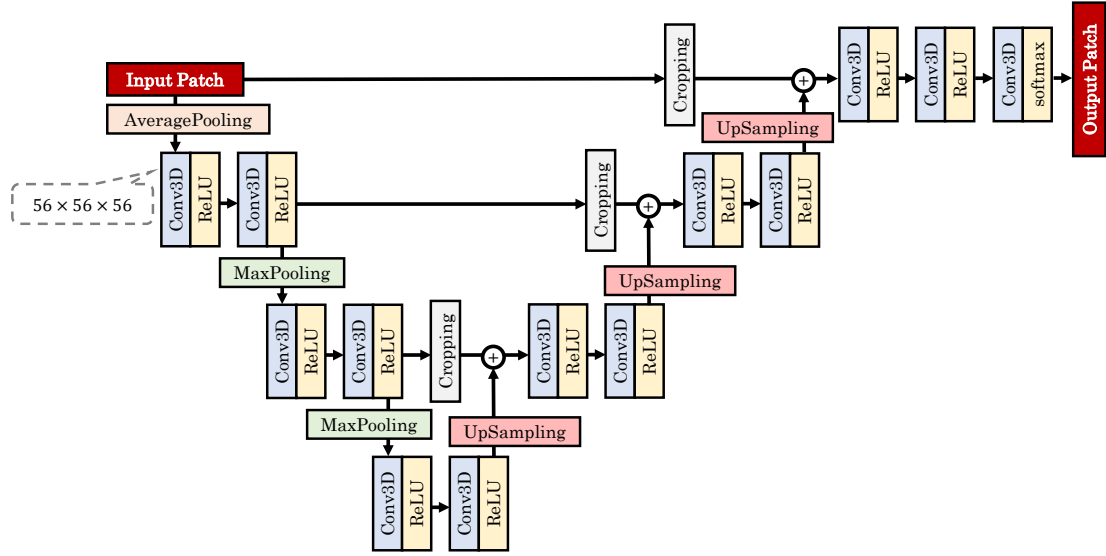
2.4. 下顎骨および咀嚼筋の自動セグメンテーション

2.4.1. 概要

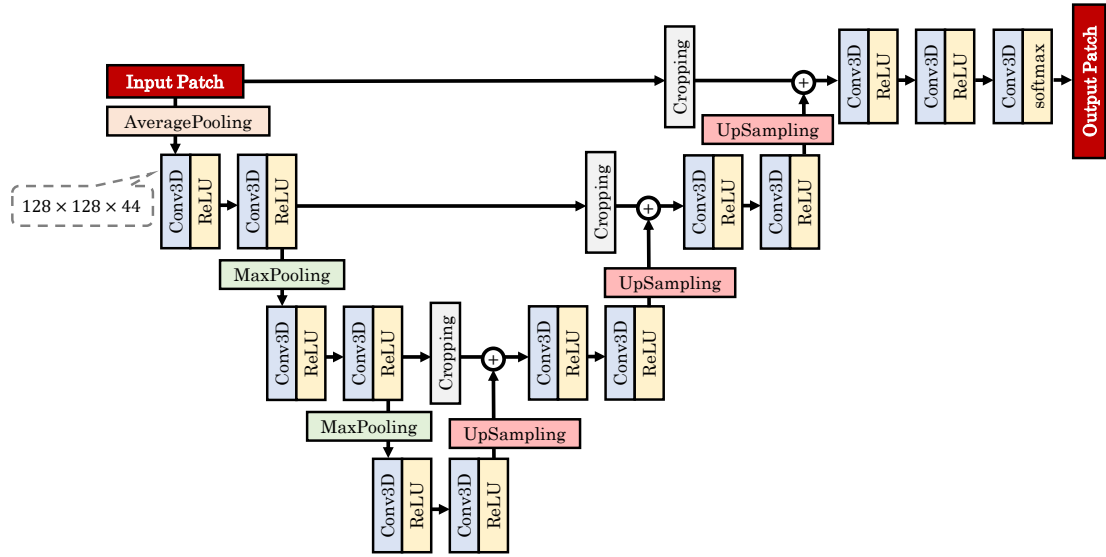
筋骨格自動セグメンテーションでは，CNN を用いた u-net の構造の 6 種のモデルアーキテクチャを構成し，それぞれで学習・認識を行うことによりその精度を検証する．いずれのモデルにおいても，最適化手法は Adam，損失関数は sparse categorical cross entropy を用いる．6 種のモデルアーキテクチャの概要を表 2.1 に示す．バッチサイズは GPU のメモリ最大限使用によりそれぞれこのように決定した．また，表 2.1 に示す入力パッチは，すべて元のサイズ $512 \times 512 \times z$ voxel からクロップしたものであり，出力パッチを x, y, z 方向に並べ結合すると元のサイズ $512 \times 512 \times z$ voxel を得ることができる．各アーキテクチャのネットワーク構成図を図 2.5 に示す．

表 2.1: CNN を用いた 6 種のモデルアーキテクチャの概要一覧

モデルアーキテクチャ	入力パッチサイズ	出力パッチサイズ	バッチサイズ
1: 3D u-net isotropic	$168 \times 168 \times 168$	$44 \times 44 \times 44$	32
2: 3D u-net 512	$512 \times 512 \times 176$	$512 \times 512 \times 12$	2
3: 3D 深さ 2 + 2D 深さ 1	$512 \times 512 \times 64$	$512 \times 512 \times 1$	16
4: 3D 深さ 2 + 2D 深さ 2	$512 \times 512 \times 64$	$512 \times 512 \times 1$	4
5: 2D u-net 深さ 4	512×512	512×512	8
6: 2D u-net 深さ 5	512×512	512×512	8

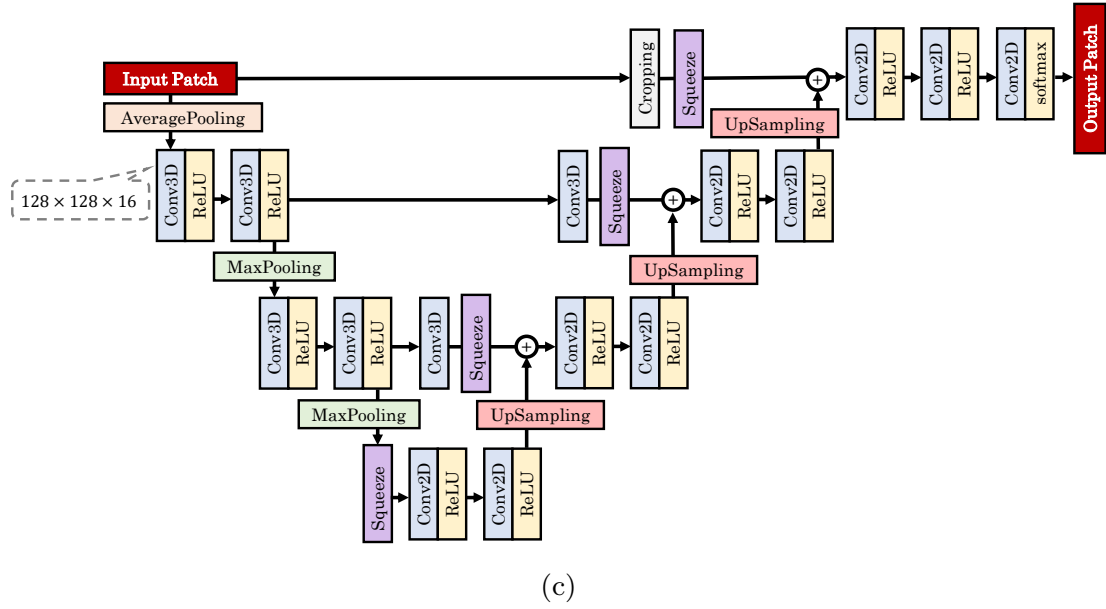


(a)

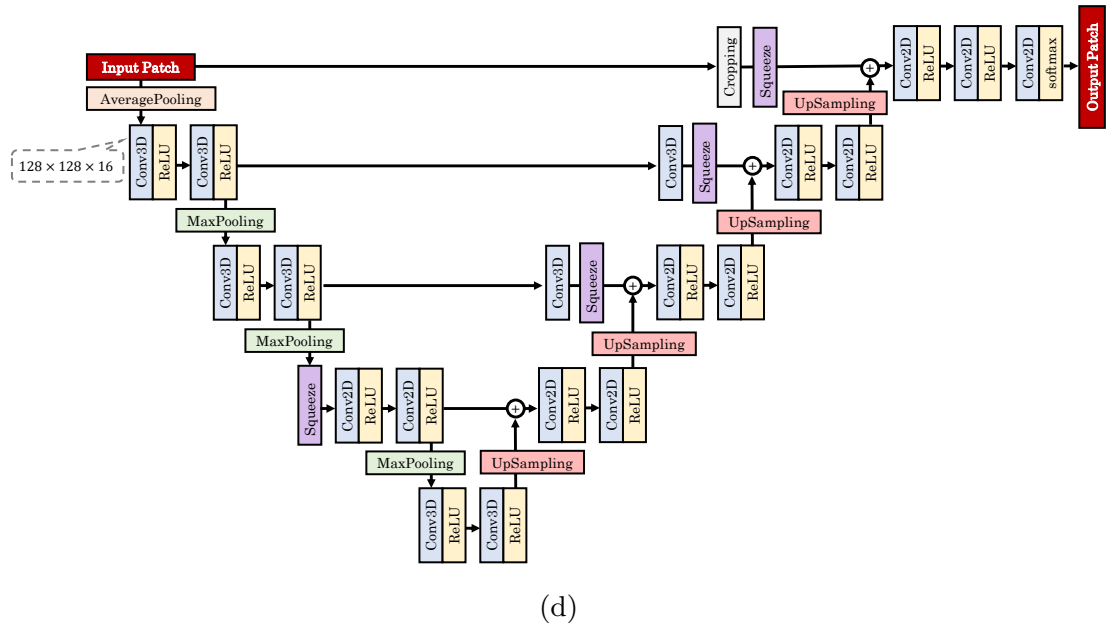


(b)

図 2.5: 各モデルのネットワーク構成図 (次項へ続く). (a) アーキテクチャ 1, (b) アーキテクチャ 2.

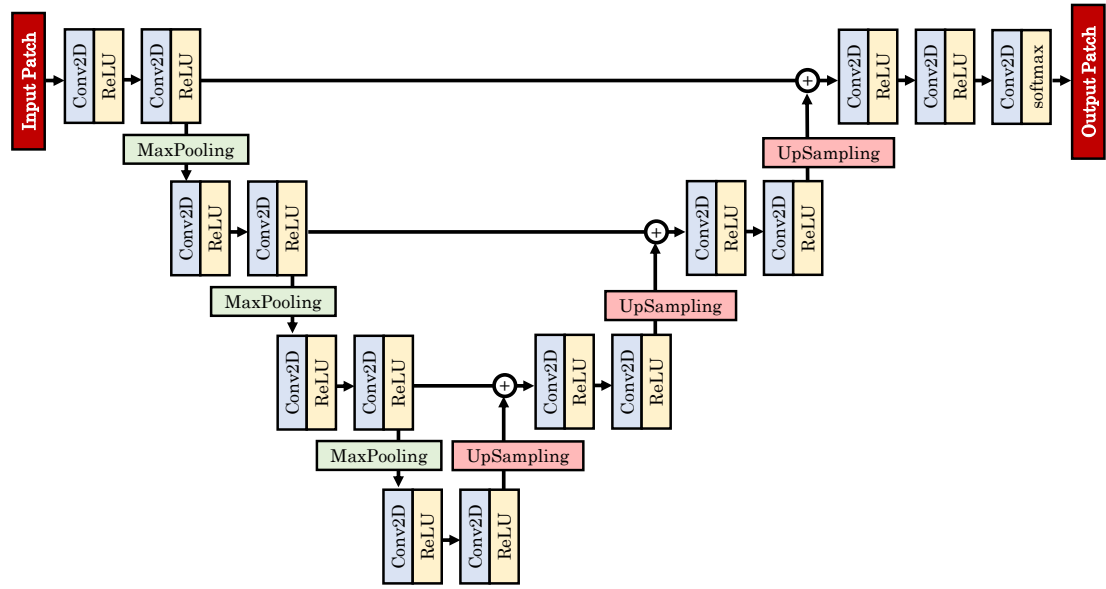


(c)

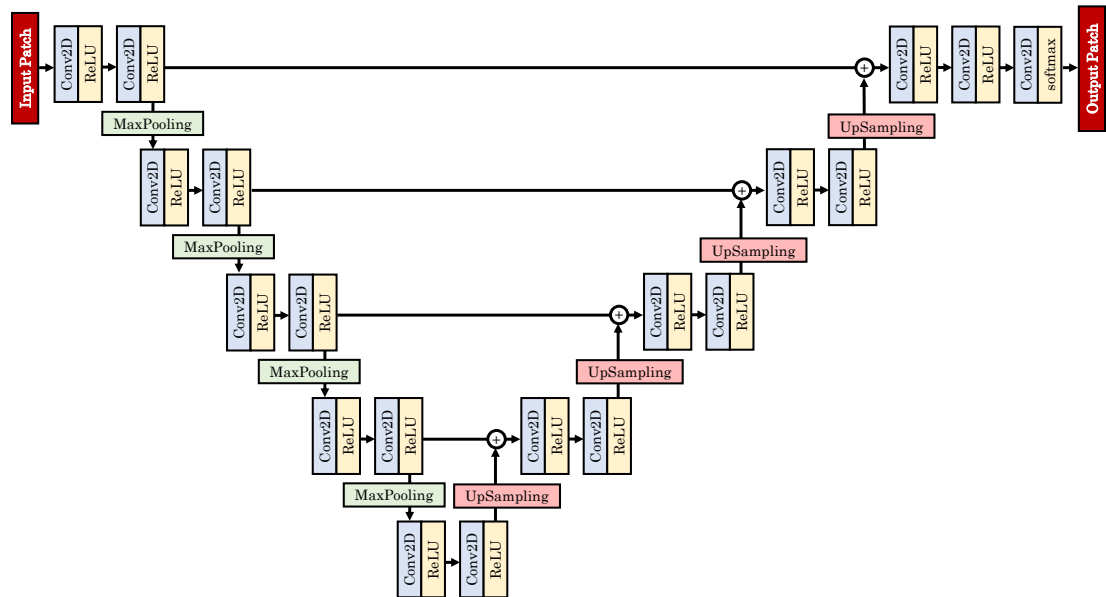


(d)

図 2.5: 各アーキテクチャのネットワーク構成図（前項の続き，次項へ続く）. (c) アーキテクチャ 3, (d) アーキテクチャ 4.



(e)



(f)

図 2.5: 各アーキテクチャのネットワーク構成図 (前項の続き). (e) アーキテクチャ 5, (f) アーキテクチャ 6.

2.4.2. 学習フェーズ

筋骨格自動セグメンテーションの学習フェーズにおける前処理として、次の2つの処理を行う。1点目は体表外に存在する器具等をCT画像内から除去するための体表自動セグメンテーション、2点目は筋骨格ラベルの学習・推論のためのCT画像のリサンプリングである。

まず、体表の自動セグメンテーションを行う。ここでは深さ4の2D u-netを用い、体表外をラベル0、体表内をラベル1とした学習・推論を行い、出力として2値のラベル画像を得る。このラベル画像をCT画像に重ね、体表外と推定された領域のCT値をすべて -1000 HUとすることで、体表外に存在する器具等をCT画像内から除去する。

さらに、 $512 \times 512 \times z$ voxel (z は症例により異なる)のCT画像データについて、1voxelあたりのサイズ $s_x \times s_y \times s_z$ mm³を $s_x \times s_x \times s_x$ mm³にリサンプリングする。この処理により1ボリュームのサイズは $512 \times 512 \times (zs_z/s_x)$ voxelとなり、このボリュームを筋骨格自動セグメンテーションの学習における入力画像とする。正解データにはマニュアルトレースにより作成した筋骨格ラベル画像を使用する。

2.4.3. 推論フェーズ

推論フェーズにおいても、学習フェーズと同様に、体表の自動セグメンテーションおよびリサンプリングを前処理として行う。また、実臨床CT画像を使用する際は、体表セグメンテーションを行う前に、症例間で異なる撮影範囲を統一するためにCT画像から必要領域のみをクロップする次のような処理を追加する。ここでは、ランドマーク自動検出によって得た $value_{\hat{\mathbf{x}}_i}^c$ とその存在位置座標 $\hat{\mathbf{x}}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)$ を用い、次の3つの判定をする。左下顎頭頂、右下顎頭頂、オトガイ最下点、左筋突起、右筋突起を順に $i = 1, 2, 3, 4, 5$ とすると、

1. 左右の下顎頭頂およびオトガイ最下点の3点すべてのconfidence valueについて、 $value_{\hat{\mathbf{x}}_i}^c > 0.8$ ($1 \leq i \leq 3$)をみたす。

2. 左右の筋突起のうち少なくとも 1 つの confidence value について,
 $value_{\hat{\mathbf{x}}_{i=4}}^c > 0.8$ または $value_{\hat{\mathbf{x}}_{i=5}}^c > 0.8$ をみたす.
3. オトガイ最下点の推定座標 $\hat{\mathbf{x}}_{i=3}$ の z 座標 $\hat{z}_{i=3}$ が, 左右の下顎頭頂の推定座標 $\hat{\mathbf{x}}_{i=1,2}$ の z 座標 $\hat{z}_{i=1,2}$ より下部に位置する. すなわち $\hat{z}_{i=3} < \hat{z}_{i=1}$ かつ $\hat{z}_{i=3} < \hat{z}_{i=2}$ をみたす.

上記 3 項目すべてを満たしている症例について, 5 点のランドマークの推定座標のうち, 最下部に位置するランドマーク $\hat{\mathbf{x}}_\alpha$ と最上部に位置するランドマーク $\hat{\mathbf{x}}_\beta$ をそれぞれ次のように求める.

$$\hat{\mathbf{x}}_\alpha = \hat{\mathbf{x}}_3 \quad (2.4.1)$$

$$\hat{\mathbf{x}}_\beta = \hat{\mathbf{x}}_i \left(i = \arg \max_{i \in \{1,2,\dots,5\}} \hat{z}_i \right) \quad (2.4.2)$$

このようにして得た $\hat{\mathbf{x}}_\alpha$ と $\hat{\mathbf{x}}_\beta$ を用いて, 元の $512 \times 512 \times z$ voxel (z は $-z/2 \leq z_{index} \leq z/2$ mm をみたす z_{index} の総数) の大きさのボリュームを $512 \times 512 \times z'$ voxel (z' は $\hat{z}_\alpha - 5 \leq z'_{index} \leq \hat{z}_\beta + 5$ mm をみたす z'_{index} の総数) にクロップする. なお, 上記 3 項目のうち 1 項目でも満たしていない場合, その CT 画像の撮影範囲内には本研究で対象とする領域すべてが含まれていないと判断し, 本研究では筋骨格自動セグメンテーションを行わない.

2.5. 顎頭間距離および筋骨格体積計測

顎頭間距離 d の計測は, 推論により得た右下顎頭頂 L_{Rcd} の座標 $\hat{\mathbf{x}}_{Rcd} = (\hat{x}_{Rcd}, \hat{y}_{Rcd}, \hat{z}_{Rcd})$ と左下顎頭頂 L_{Lcd} の座標 $\hat{\mathbf{x}}_{Lcd} = (\hat{x}_{Lcd}, \hat{y}_{Lcd}, \hat{z}_{Lcd})$ を用いて次式より求められる.

$$d = \sqrt{|\hat{x}_{Rcd} - \hat{x}_{Lcd}|^2 + |\hat{y}_{Rcd} - \hat{y}_{Lcd}|^2 + |\hat{z}_{Rcd} - \hat{z}_{Lcd}|^2} \quad (2.5.1)$$

また, 各筋骨格 (下顎骨および咬筋, 内側翼突筋) の体側 v の計測は, 1voxel あたりのサイズ $s_x \times s_y \times s_z$ mm³ と各筋骨格の voxel 数 n_v を用いて次式より求められる.

$$v = s_x \times s_y \times s_z \times n_v \quad (2.5.2)$$

第 3 章 実験

3.1. 使用データ

本研究では，顎口腔領域を対象とした CT 実画像を使用した．これらの CT 画像は鶴見大学歯学部にて撮影されたものである．本研究では，次の 2 つのデータセットを用いて実験を行った．

1 つ目は，金属アーチファクトのない CT 画像のみで構成されたデータセットである．詳細を表 3.1 に示す．このデータセットは，ランドマーク自動検出の学習および筋骨格自動セグメンテーションの学習において使用する．また，予備実験である金属アーチファクトのシミュレーション実験においても，本データセット内の 1 症例を用いる．

2 つ目は実臨床 CT 画像データベースである．このデータベースには，CT 画像だけでなく，文字情報も含まれている．ここで述べる文字情報とは，患者年齢，性別，検査日，検査部位，医師による画像診断結果の 5 項目を指し，.xlsx 形式で保存されている．CT 画像データと文字情報は，患者 ID によって紐づけられている．本研究で用いる実臨床 CT 画像データベースは，様々な症状・疾患のデータベースであるため，症例によって CT 画像の撮影範囲や撮影に用いられるフィルタが図 3.1 のように異なり，本研究にて対象とする筋骨格領域すべてを含んでいるとは限らない．このデータセットの詳細を表 3.2 に示す．本研究では，この実臨床 CT 画像データベースに対して一連の提案手法を適用する．

表 3.1: 金属アーチファクトのない CT 画像データセット

項目	
対象領域	顎口腔領域
画像形式	CT 画像
症例数	金属アーチファクトなし 36 症例
画像サイズ	$512 \times 512 \times 77 \sim 301$ voxel
1voxel あたりのサイズ	$0.41 \times 0.41 \times 0.5 \sim 1.25\text{mm}^3$ (1 症例のみ $0.35 \times 0.35 \times 0.5\text{mm}^3$)
共同研究機関	鶴見大学歯学部
撮影機器	日立メディコ社製 SUPRIA

表 3.2: 実臨床 CT 画像データベース

項目	
対象領域	顎口腔領域
画像形式	CT 画像
症例数	1196 (CT 画像データ数 2387)
画像サイズ	$512 \times 512 \times 38 \sim 723$ voxel
1voxel あたりのサイズ	$[0.39 \sim 0.69] \times [0.39 \sim 0.69] \times [0.16 \sim 5.00]\text{mm}^3$ (よく見られるのは $0.41 \times 0.41 \times 0.31 \sim 1.25\text{mm}^3$)
共同研究機関	鶴見大学歯学部
撮影機器	日立メディコ社製 SUPRIA

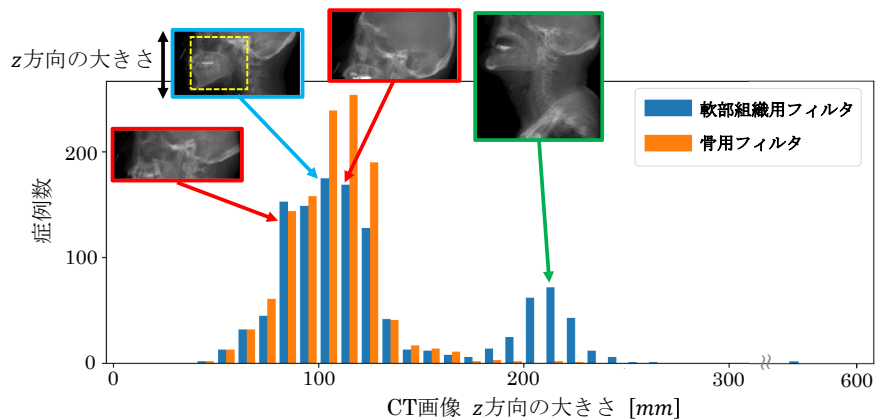


図 3.1: 実臨床 CT 画像データベース内の CT 画像の z 方向の大きさの分布. Digitally Reconstructed Radiographs (DRR) 画像について, 赤色枠の画像は下顎骨全体を含まない症例, 水色枠の画像は下顎骨全体 (黄色点線枠内) を含む症例, 緑色枠の画像は下顎骨全体を含むが撮影範囲が広すぎる症例をそれぞれ指す.

3.2. 金属アーチファクトシミュレーションにて考慮するパラメータ

金属アーチファクトのシミュレーション [16] において, 次のパラメータを仮定し, 様々な CT 画像撮影条件を想定したシミュレーションを行った. その中で金属アーチファクトの程度を目視で確認し, 図 3.2 のように, Mild Artifact と Severe Artifact の 2 グループに分けた.

- 金属の位置・形状
 - 2 種類
- 金属の種類
 - チタン, 鉄, パラジウム合金の 3 種類
- 金属の密度
 - 5.0, 6.0, 8.0, 11.6, 13.0g/cm³ の 5 種類
- CT 画像上に生じるポワソンノイズの程度 (照射 X 線強度)
 - 3 種類

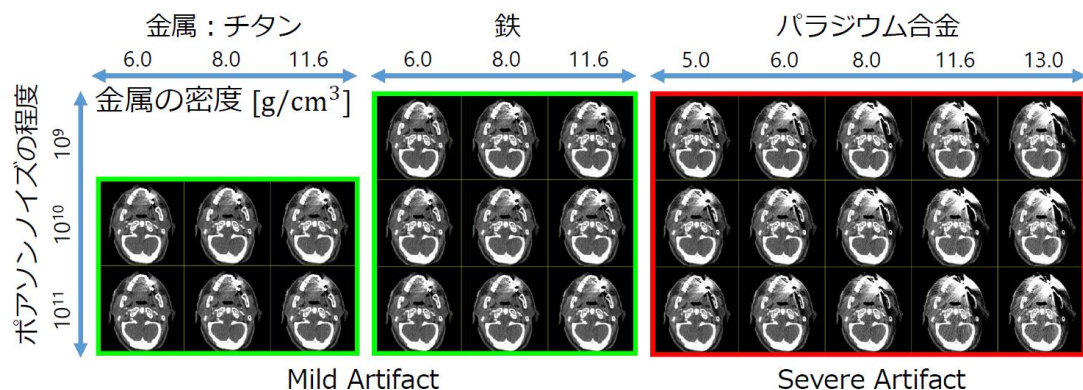


図 3.2: シミュレーションにて考慮するパラメータとそのとき生成される CT 画像の例 (WL=100, WW=500)。本研究では、緑線で囲んでいるものを Mild Artifact, 赤線で囲んでいるものを Severe Artifact とする。

3.3. 精度評価方法

金属アーチファクト低減の精度検証には、金属アーチファクトシミュレーションにて疑似的にアーチファクトを起こした CT 画像 1 症例 60 パターンを用いる。ランドマーク自動検出および、下顎骨および咀嚼筋の自動セグメンテーションの精度検証には、金属アーチファクトのない CT データセットを用いる。

3.3.1. 金属アーチファクト低減

金属アーチファクトシミュレーションを用い、本研究にて提案する金属アーチファクト低減法の精度検証を行う。生成画像の画質精度の評価には、PSNR (Peak Signal-to-Noise Ratio) および相互情報量 (Mutual Information:MI) を用いる。なお、PSNR の算出に用いる画像は CT 画像を WW=500, WL=100 で正規化したものであり、また、MI の算出には、CT 画像を-1000HU から 4000HU の範囲でビン数 255 の度数分布にしたものを用いる。

3.3.2. ランドマーク自動検出

金属アーチファクトのない CT 画像データセット内の CT 画像 36 症例のうち、本研究にて対象とするランドマーク 5 点すべてが CT 画像中存在する 35 症例において、3D Slicer を用いて手動で設定したランドマーク L_i の位置を正解座標とし、5-fold の交差検証により学習・推論を行う。ランドマーク L_i において、ランドマークの正解座標 $\mathbf{x}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$ と自動検出座標 $\hat{\mathbf{x}}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i) \in \mathbb{R}^3$ のユークリッド距離誤差 R_i は次式で表すことができ、これを、精度評価座標とする。

$$R_i = \sqrt{|x_i - \hat{x}_i|^2 + |y_i - \hat{y}_i|^2 + |z_i - \hat{z}_i|^2} \quad (3.3.1)$$

3.3.3. 下顎骨および咀嚼筋の自動セグメンテーション

金属アーチファクトのない CT 画像データセット内の CT 画像 36 症例すべてにおいて、VINCENT を用いて手動でトレースし作成した 4 つの筋骨格ラベルを正解領域とし、4-fold の交差検証により学習・推論を行う。正解領域と自動セグメンテーション領域の比較について、Dice Coefficient (DC), 平均表面間距離誤差 (Average Symmetric Surface Distance: ASD) を精度評価指標とする。正解領域を A , 自動セグメンテーション領域を B とすると、各評価指標は次のように表せる。なお、ASD の算出において、 M_A, M_B はそれぞれ A, B の表面上の点群を指し、 x, y はそれぞれ M_A, M_B のうちの 1 点を表す。 d は括弧内の 2 点間のユークリッド距離である。

$$DC = \frac{2 * |A \cap B|}{|A| + |B|} \quad (3.3.2)$$

$$ASD = \frac{1}{|M_A| + |M_B|} \left(\sum_{x \in M_A} \min_{y \in M_B} d(x, y) + \sum_{y \in M_B} \min_{x \in M_A} d(y, x) \right) \quad (3.3.3)$$

第 4 章 実験結果

4.1. 金属アーチファクト低減

4.1.1. 予備実験：金属アーチファクトシミュレーションの適用結果

金属アーチファクトのない CT 実画像に金属部分を追加しシミュレーションおよび金属アーチファクト低減を行った結果を図 4.1 および図 4.2 に示す．図 4.1 はチタン，図 4.2 はパラジウム合金どちらも密度 $6.0\text{g}/\text{cm}^3$ を仮定し，どちらもノイズの大きさは想定した 3 種のうち中程度ものを仮定した．PSNR および MI は，いずれも金属アーチファクトなし CT 画像と各生成画像の 2 画像により算出した．

図 4.1 より，Mild Artifact の場合，従来手法・提案手法どちらを使用した場合も，低減前より PSNR の値が高くなり，MI は同程度であった．また，金属領域が既知の場合を想定した低減との差も大きくないことが読み取れる．

図 4.2 より，Severe Artifact の場合，金属領域が既知の場合を想定した低減と提案手法を適用した低減について，低減前と比較して PSNR と MI の値が高くなった．一方，従来手法を適用した場合，図 2.3 と同様に，低減に失敗している．

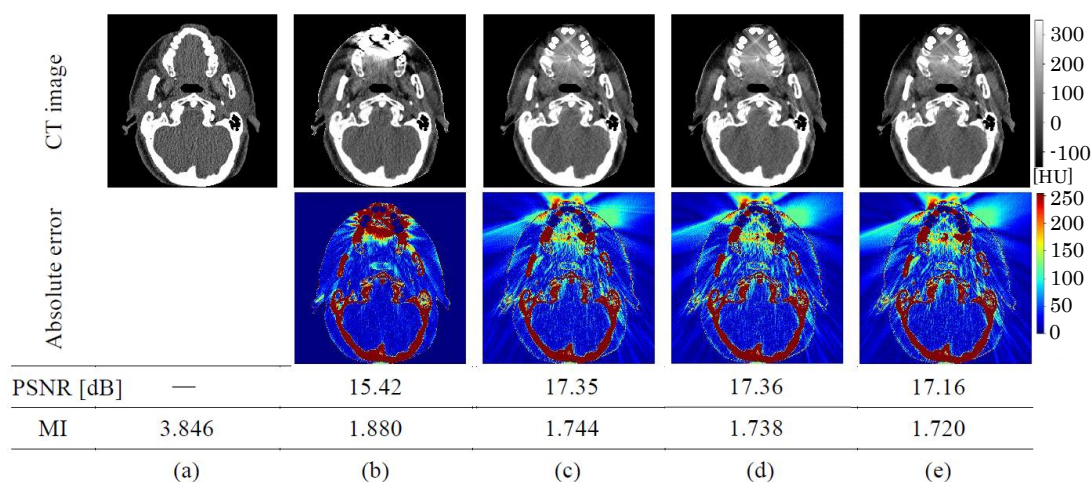


図 4.1: シミュレーションにより mild artifact を生じさせた場合の低減結果（一例）. (a) 金属アーチファクトなし CT 実画像, (b) 低減前画像, (c) 金属領域既知の場合の NMAR による低減後画像, (d) 金属領域が既知でない場合の NMAR による低減後画像（従来手法）, (e) 提案手法適用による低減後画像.

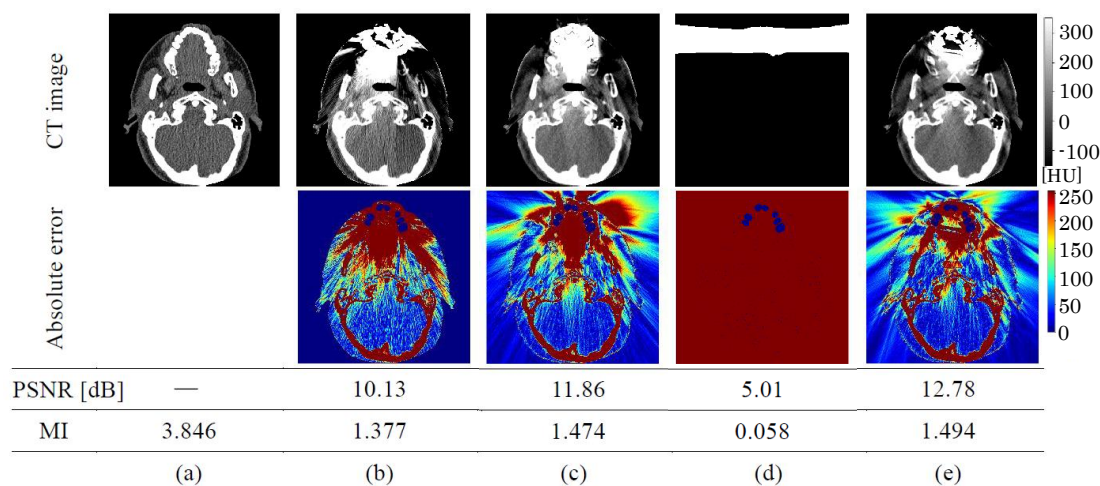


図 4.2: シミュレーションにより severe artifact を生じさせた場合の低減結果（一例）. (a) 金属アーチファクトなし CT 実画像, (b) 低減前画像, (c) 金属領域既知の場合の NMAR による低減後画像, (d) 金属領域が既知でない場合の NMAR による低減後画像（従来手法）, (e) 提案手法適用による低減後画像.

4.1.2. 実臨床 CT 画像データベースへの適用結果

本提案手法を実臨床 CT 画像データベースに適用した際の一例を図 4.3 に示す。このような CT 画像においては、金属アーチファクトのない正解画像が存在しないため定性的な評価となるが、従来手法適用後では確認できなかった画像上の情報が、提案手法適用後では図 4.3 中矢印で示すように確認することができた。

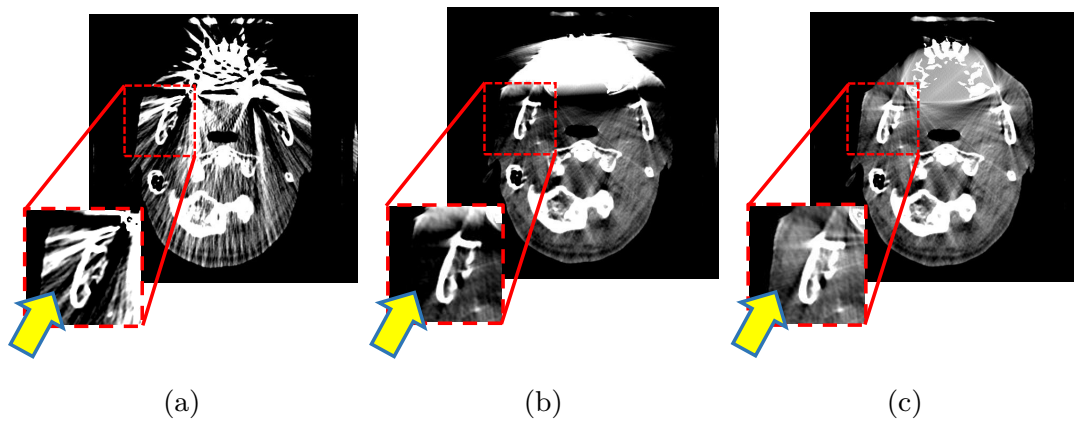


図 4.3: 金属アーチファクトのある CT 実画像の低減前後 (WL=100, WW=500).
(a)CT 実画像 (低減前), (b) 従来手法適用後, (c) 提案手法適用後

4.2. ランドマーク自動検出

4.2.1. 予備実験：交差検証による学習モデルの精度評価

金属アーチファクトのない CT 画像データセット内の CT 画像 36 症例のうち、本研究にて対象とするランドマーク 5 点すべてが CT 画像中に存在する 35 症例を用い、5-fold による交差検証を行った。ここで、入力画像のウィンドウレベル WL およびウィンドウ幅 WW について $[WL, WW] = [100, 500]$, $[300, 2000]$ の 2 種類のネットワークを作成し、それぞれ学習・推論を行ってその精度比較を行った。結果を図 4.4 に示す。図 4.4 より、2 種類のネットワークの検出精度の違いはランドマークによって異なることが読み取れる。

本研究では顎頭間距離の構成要素である左右の下顎頭頂の精度について注目し、2 種類のモデルのうち左右の下顎頭頂の距離誤差の 35 症例での中央値がより低く精度が高いと考えられる $[WL, WW] = [300, 2000]$ の場合のモデルを、実臨床 CT 画像データベースに適用する。入力 CT 画像を $[WL, WW] = [300, 2000]$ で学習を行ったときの 5-fold 交差検証によるランドマーク自動検出結果一覧を図 4.5 に示す。

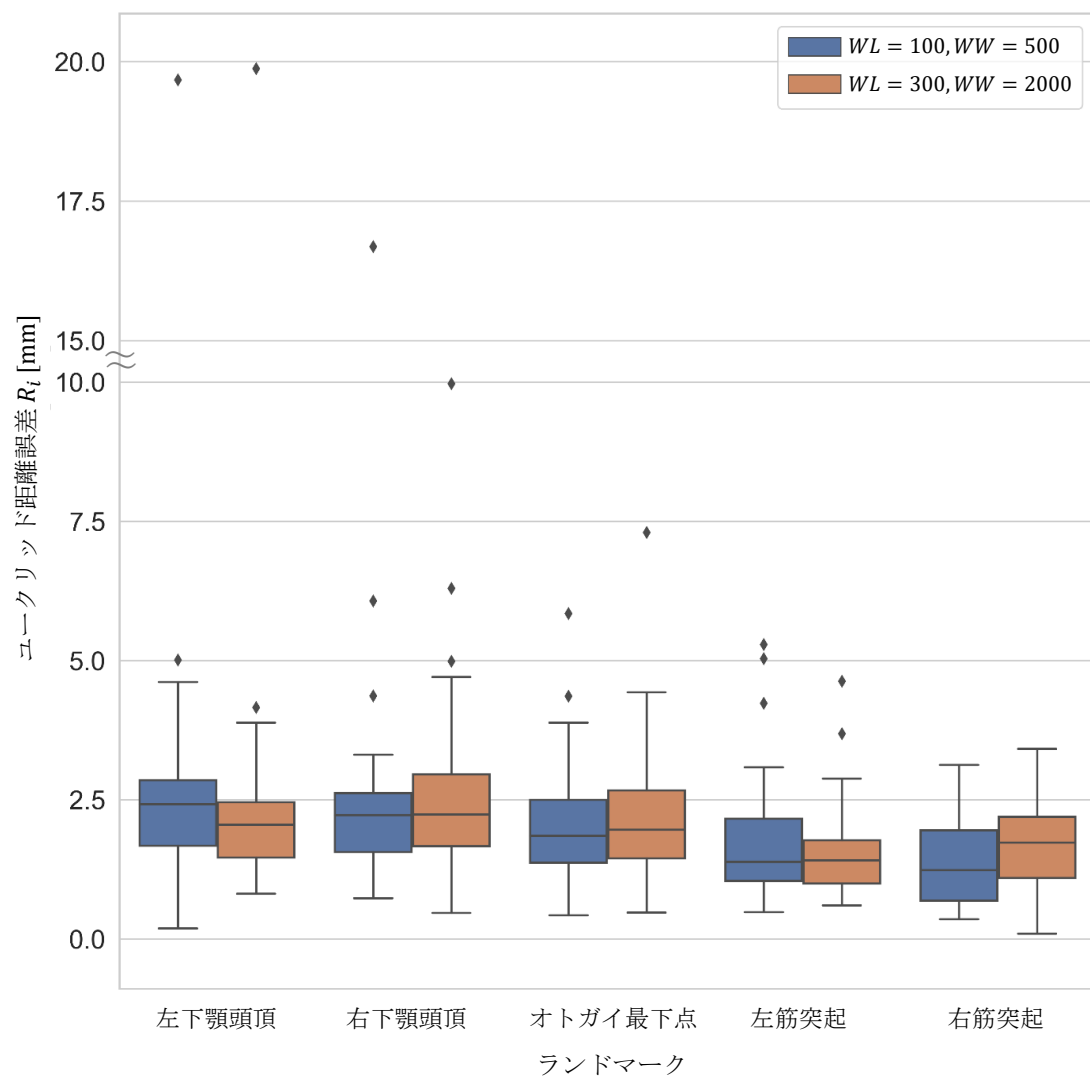


図 4.4: 入力 CT 画像の WL, WW を変更して学習を行ったときのランドマーク自動検出結果 (5-fold)

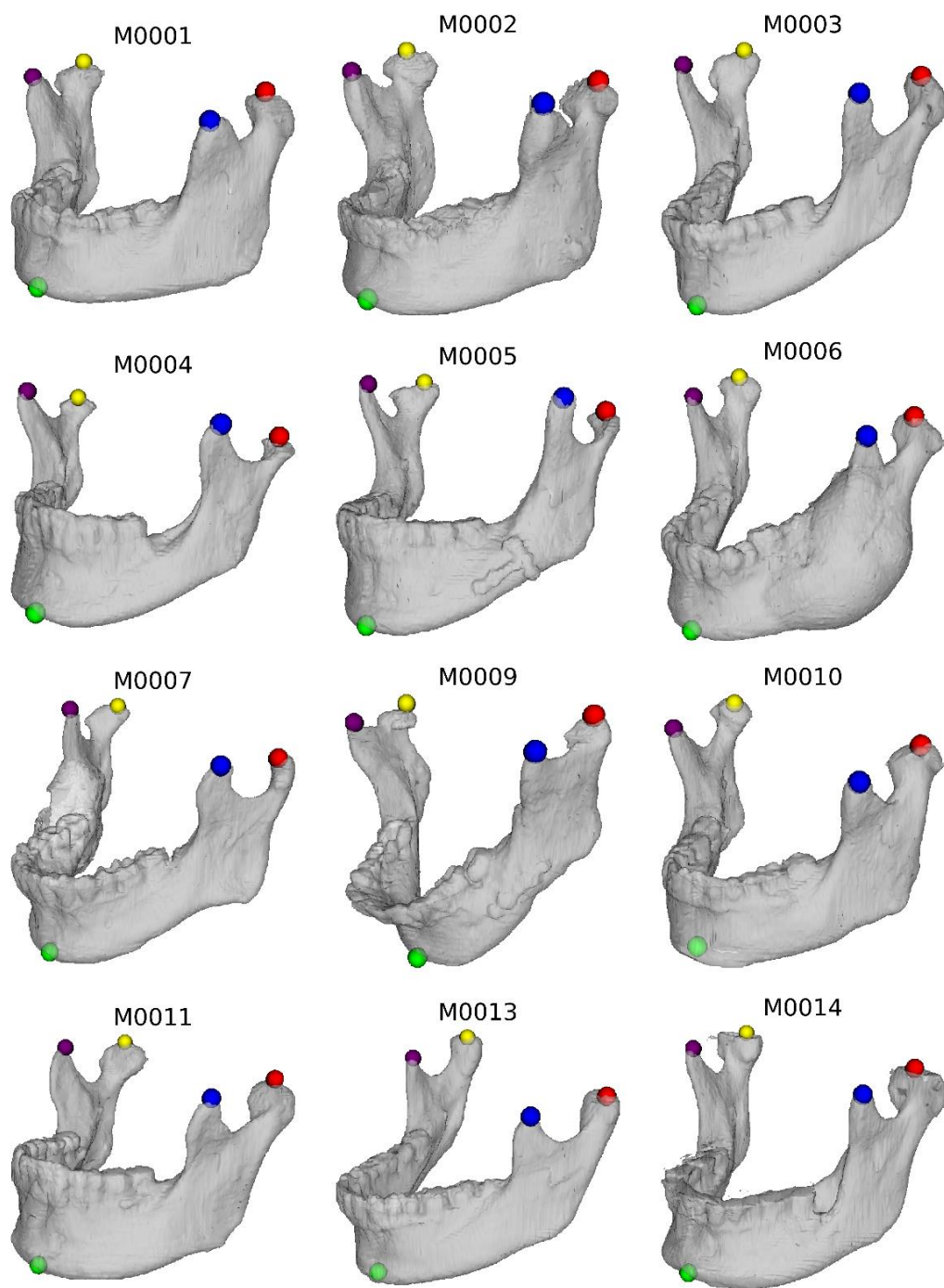


図 4.5: 入力 CT 画像を $WL = 300, WW = 2000$ で学習を行ったときの 5-fold 交差検証によるランドマーク自動検出結果一覧（次項へ続く）

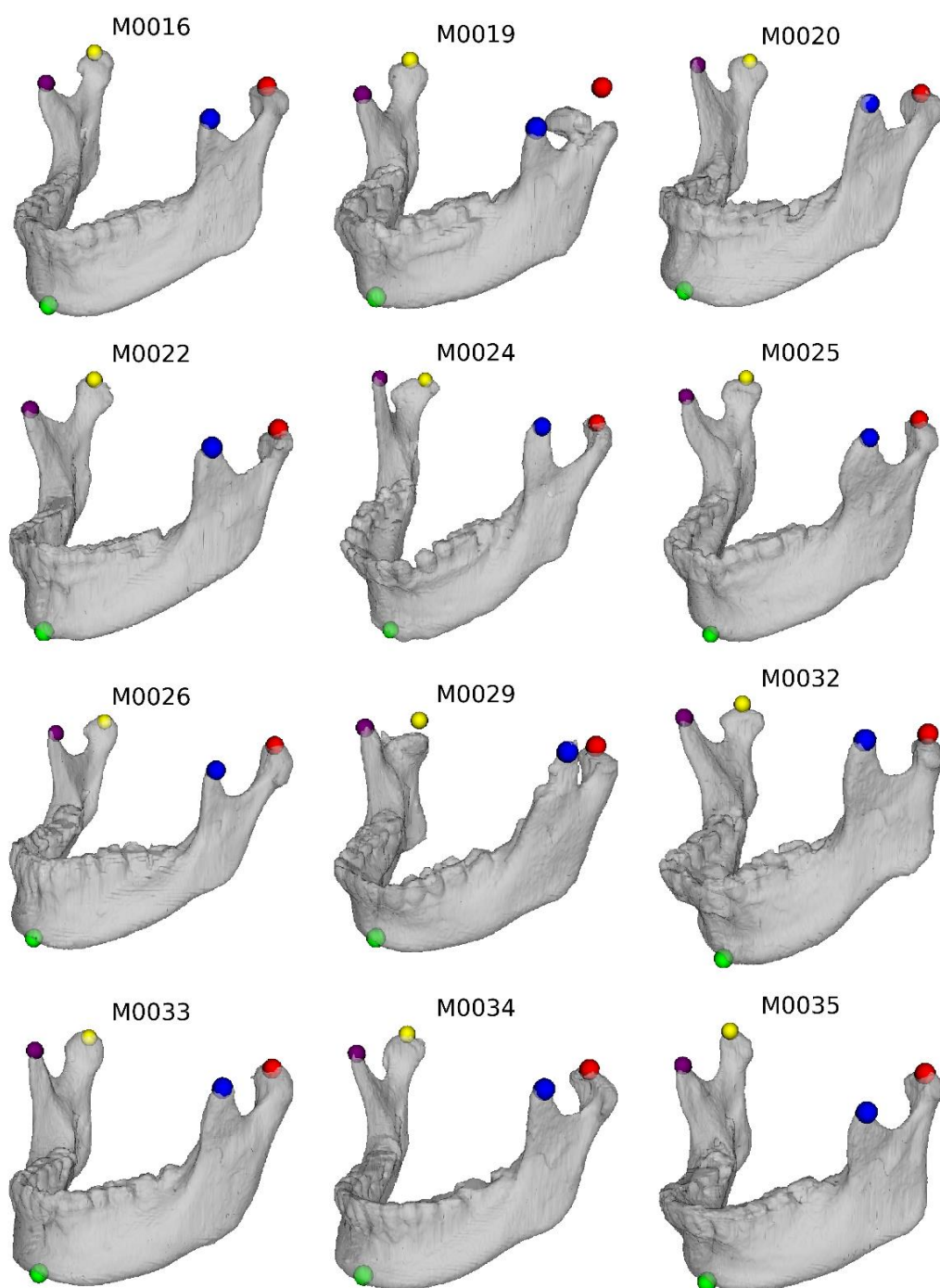


図 4.5: 入力 CT 画像を $WL = 300, WW = 2000$ で学習を行ったときの 5-fold 交差検証によるランドマーク自動検出結果一覧（前項の続き，次項へ続く）

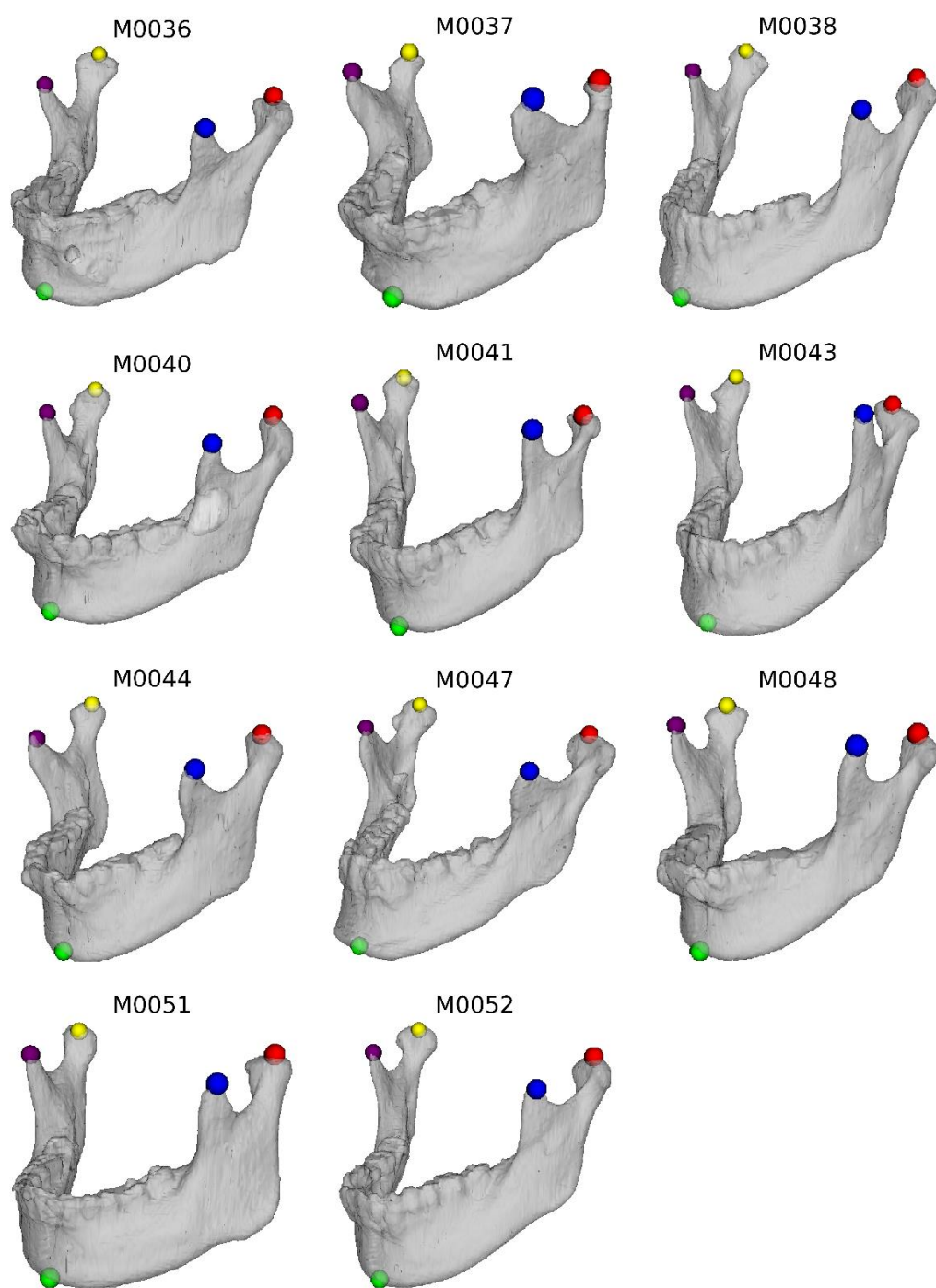
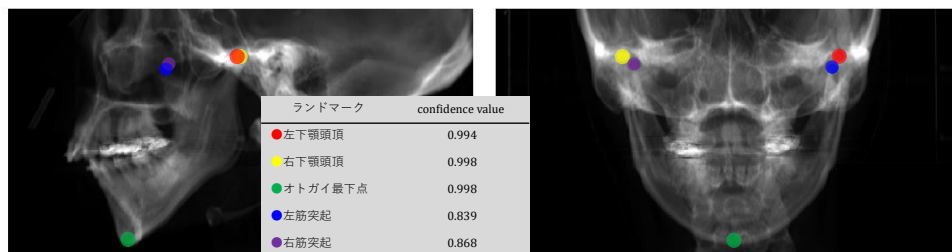


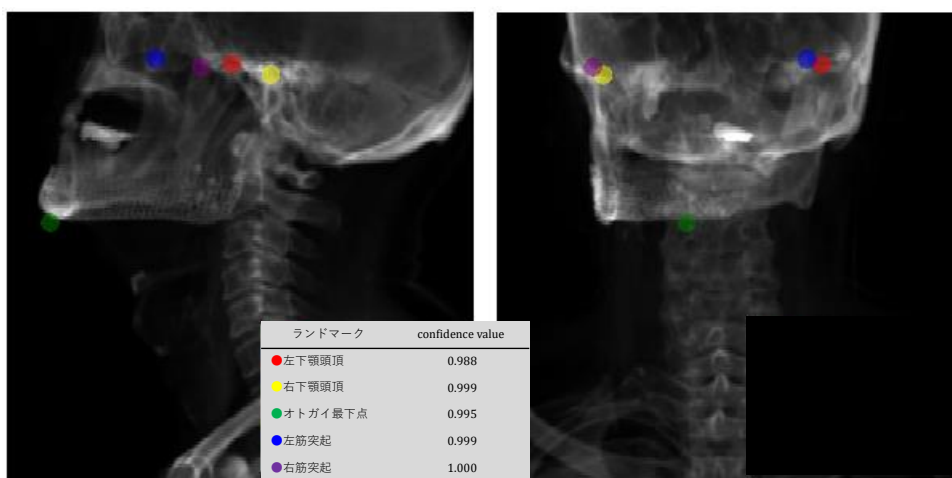
図 4.5: 入力 CT 画像を $WL = 300, WW = 2000$ で学習を行ったときの 5-fold 交差検証によるランドマーク自動検出結果一覧（前項の続き）

4.2.2. 実臨床 CT 画像データベースへの適用結果

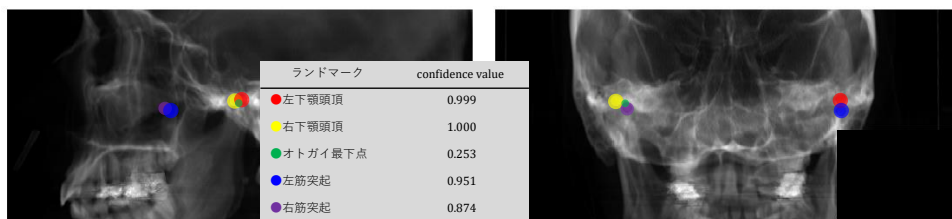
前述の 5-fold 交差検証によって得られた学習済みモデルのうちの一つを用い、実臨床 CT 画像データベース内の金属アーチファクト低減後の CT 画像でランドマーク自動検出の推論を行った。この実験結果の一部を図 4.6 に示す。画像内のラベルには、ランドマーク名とともに、ランドマーク検出にて推定された confidence value $value_{\hat{\mathbf{x}}_i}^c$ の値を示している。このステップでは、まだ筋骨格自動セグメンテーションを行っていないため、入力 CT 画像から Digitally Reconstructed Radiographs (DRR) を作成しその上に自動検出した 5 点のランドマークを表示させている。図 4.6 に示すとおり、実臨床 CT 画像データベース内には様々な大きさの CT 画像が存在するが、CT 画像内にランドマークが存在する場合は confidence value $value_{\hat{\mathbf{x}}_i}^c$ の値は 1 に近い値をとっており、また、CT 画像内にランドマークが存在しない場合はこの値が低くなっていることが読み取れる。



(a)



(b)



(c)

図 4.6: 実臨床 CT 画像データベースにおけるランドマーク自動検出結果 (一部). (a) 下顎骨全体を含んでいる CT 画像, (b) 下顎骨全体を含んでるが, 撮影範囲が大きすぎるためクロップが必要な CT 画像, (c) 下顎骨の一部が撮影範囲の対象に入っていない CT 画像.

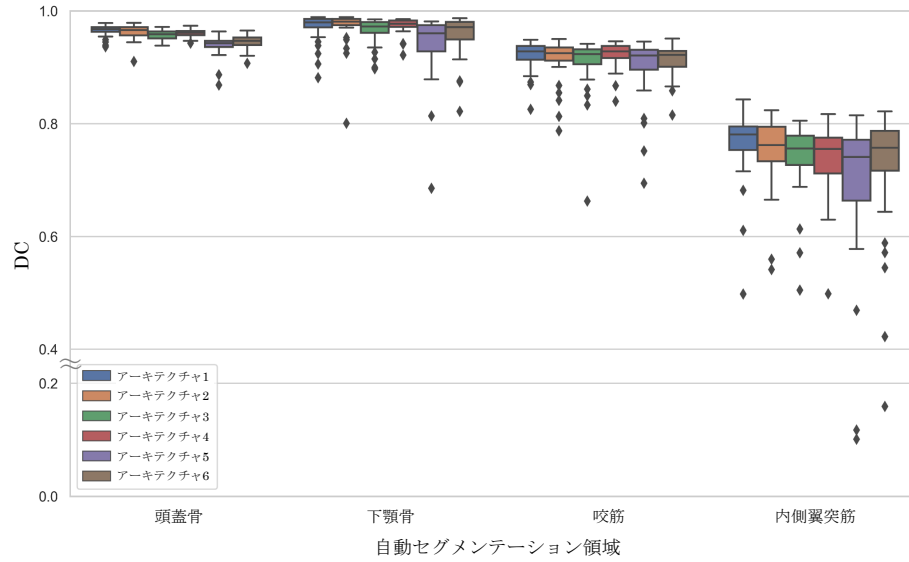
4.3. 下顎骨および咀嚼筋の自動セグメンテーション

4.3.1. 予備実験 1：交差検証による学習モデルの精度評価比較

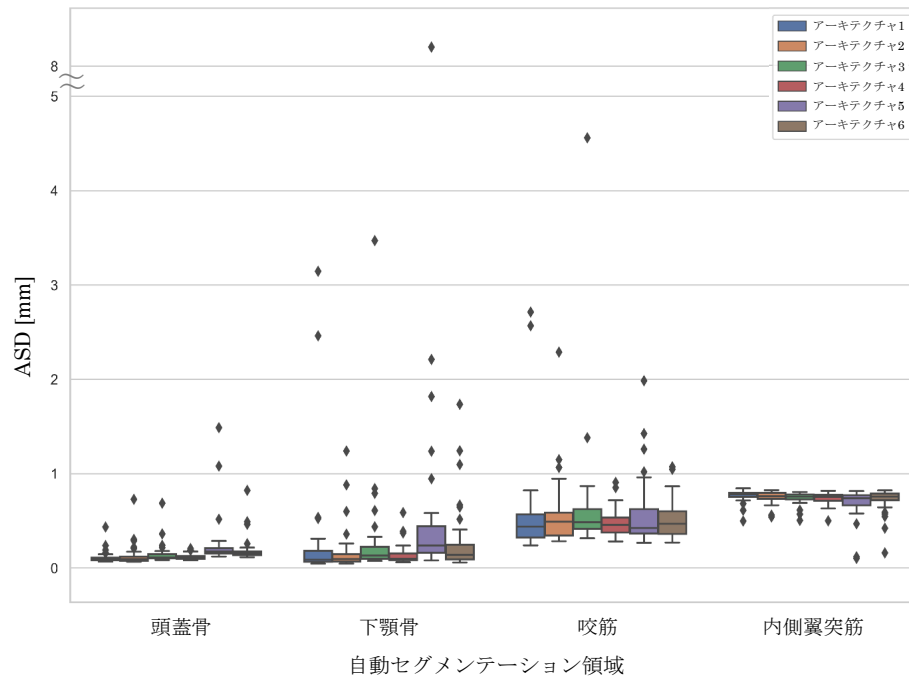
筋骨格自動セグメンテーションを行うにあたり，まず，手法の章にて示した CNN を用いた 6 種類のアーキテクチャを構築し，金属アーチファクトのない CT 画像データセット内の CT 画像 36 症例を用いて 4-fold の交差検証を行い，セグメンテーション精度検証を行った．結果を図 4.7 に示す．この図より，アーキテクチャ 1 による学習・推論の精度が最も高いように見える．ここで，筋骨格自動セグメンテーションの精度 DC について，アーキテクチャ 1 と他のアーキテクチャ間で Wilcoxon の符号順位検定を行い p 値を求めると，次のことが分かった．

- アーキテクチャ 1 とアーキテクチャ 4 の間では，内側翼突筋のセグメンテーション結果のみにおいて p 値が 0.01 以下となり，有意差があるといえる．
- アーキテクチャ 1 とアーキテクチャ 6 の間では，内側翼突筋のセグメンテーション結果のみにおいて p 値が 0.01 以下となり，有意差があるといえる．
- アーキテクチャ 1 と上記以外の 3 アーキテクチャの間では，4 種類全ての筋骨格のセグメンテーション結果において p 値が 0.01 以下または 0.05 以下となり，有意差があるといえる．

以上より，実臨床 CT 画像データベースに適用するモデルは，アーキテクチャ 1 を用いて学習を行ったモデルとする．金属アーチファクトのない CT 画像データセットについてアーキテクチャ 1 で学習を行ったときの 4-fold 交差検証による筋骨格自動セグメンテーション結果一覧を図 4.8 に示す．



(a)



(b)

図 4.7: 6 種類の CNN アーキテクチャによる筋骨格自動セグメンテーション精度比較 (4-fold). (a)DC, (b)ASD.

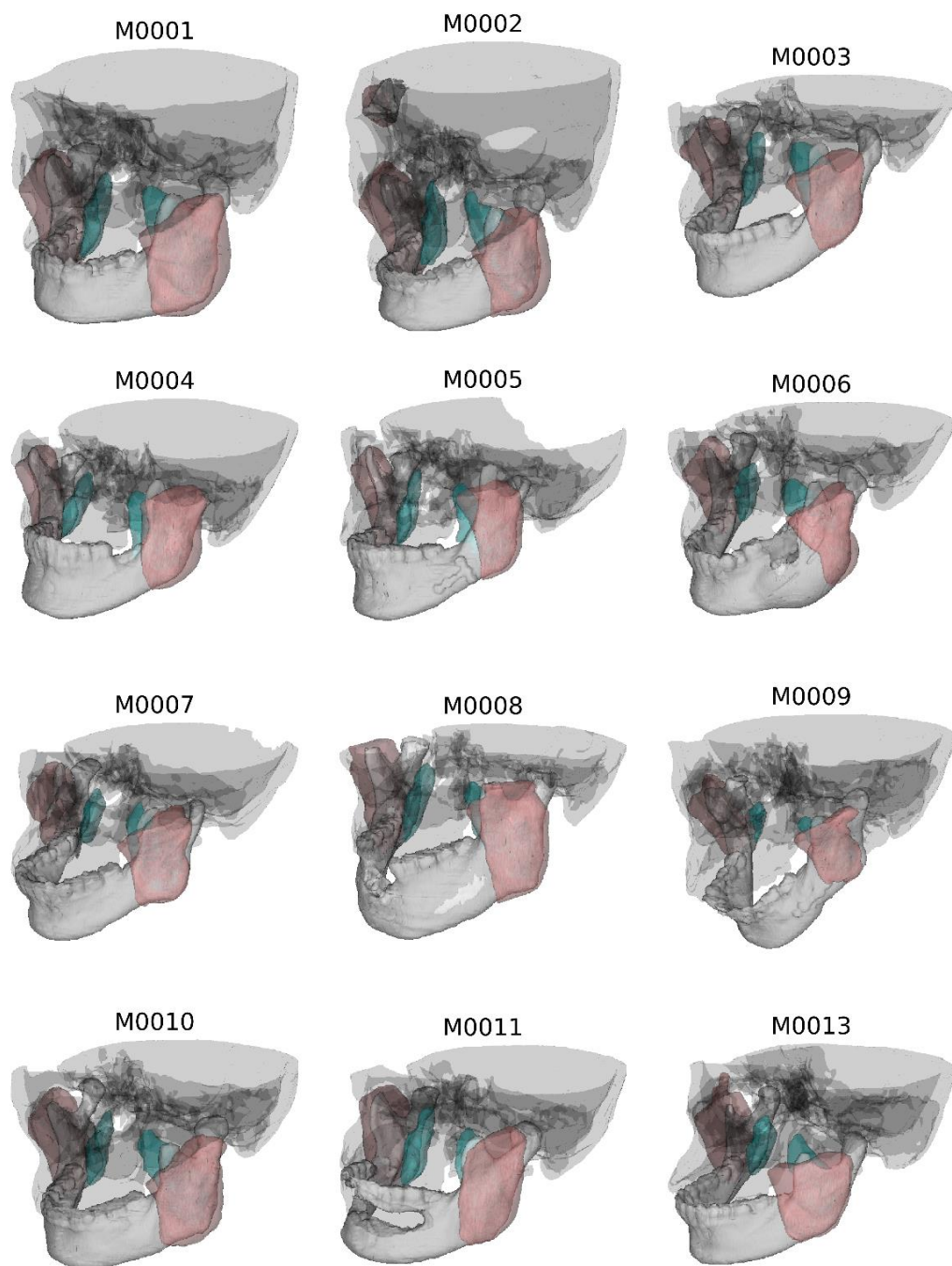


図 4.8: アーキテクチャ 1 で学習を行ったときの 4-fold 交差検証による筋骨格自動セグメンテーション結果一覧（次項へ続く）

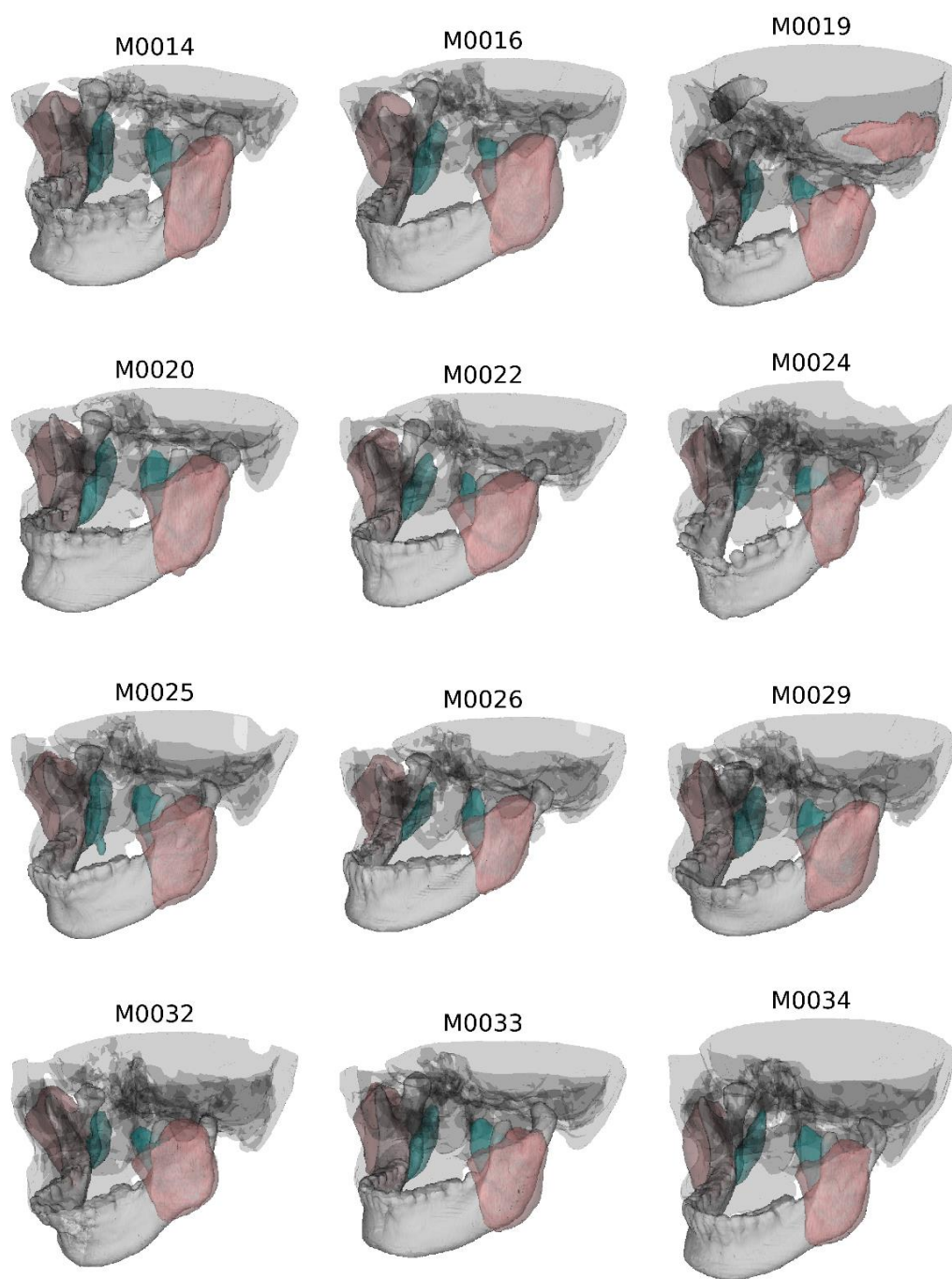


図 4.8: アーキテクチャ 1 で学習を行ったときの 4-fold 交差検証による筋骨格自動セグメンテーション結果一覧（前項の続き，次項へ続く）

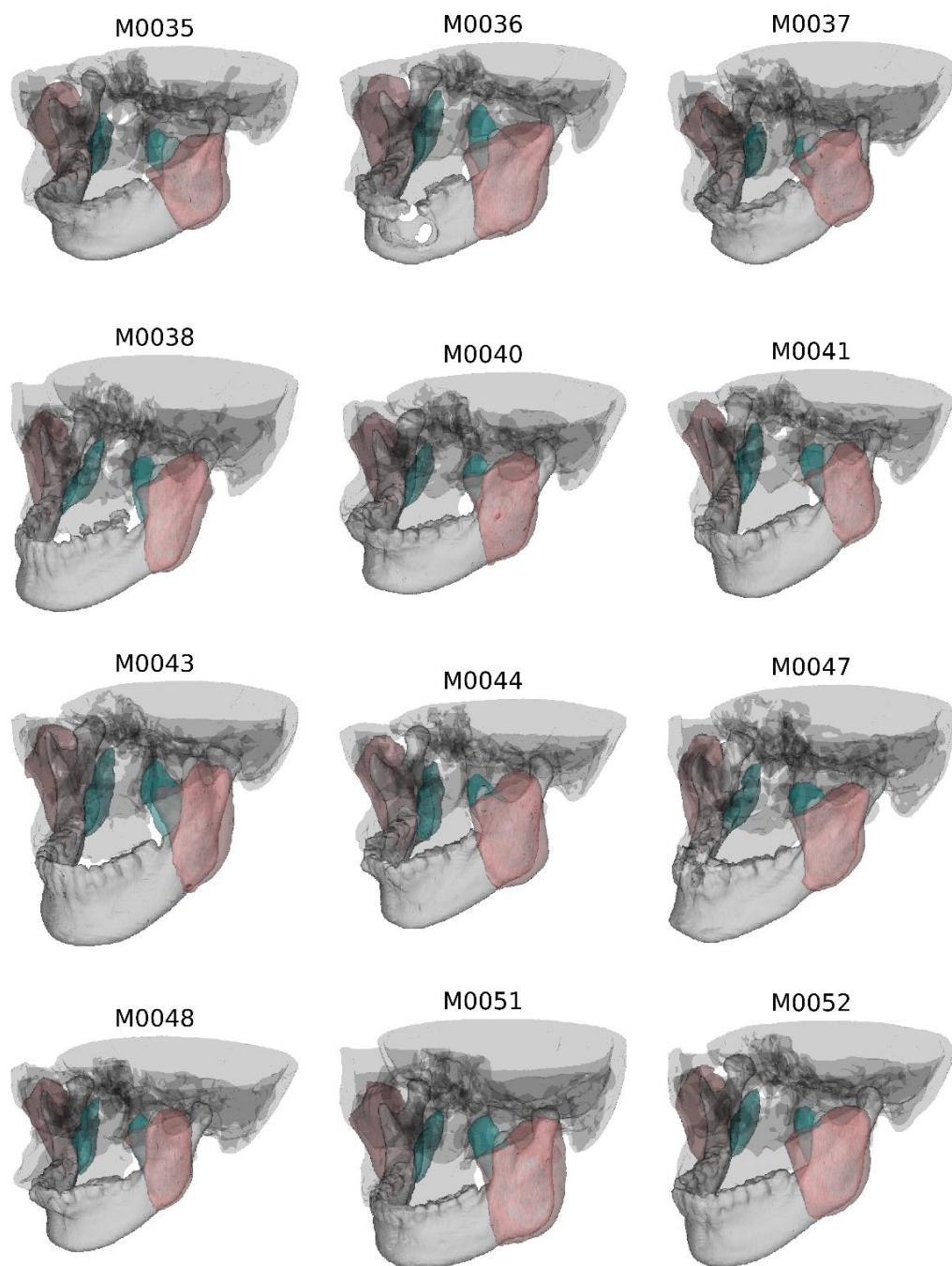


図 4.8: アーキテクチャ 1 で学習を行ったときの 4-fold 交差検証による筋骨格自動セグメンテーション結果一覧（前項の続き）

4.3.2. 予備実験 2：金属アーチファクトシミュレーションによる学習モデルの精度評価比較

金属アーチファクトシミュレーションにより得た金属アーチファクト CT 画像およびその低減度画像を用い，前小節にて最も高精度であったアーキテクチャ 1 のモデルにて筋骨格セグメンテーションの推論を行った結果を図 4.9 に示す．この図において，ヒートマップの縦方向は，金属の種類，密度，照射 X 線強度といったシミュレーションのパラメータの違いを指し，横方向は，その金属アーチファクトシミュレーション画像および金属アーチファクト低減後画像（金属既知，従来手法，提案手法）を左から順に指している．このヒートマップにて各画像の筋骨格セグメンテーション精度を示している．この結果について，金属アーチファクト低減前後および低減手法の違いによる筋骨格自動セグメンテーションの精度 DC に着目し，Wilcoxon の符号順位検定により p 値を求めると，特に severe artifact の結果において次の結果を得た．

- 4 種類すべての筋骨格について，低減前 - 金属領域既知の場合の低減後間の p 値は 0.01 以下となり，有意差があるといえる．
- 咬筋については，金属領域既知の場合の低減後のみならず，低減前 - 提案手法による低減後間の p 値も 0.01 以下となり，また，低減前 - 従来手法による低減後間の p 値は 0.05 以下となり，これらに有意差があるといえる．
- 内側翼突筋については，低減前 - 提案手法による低減後間の p 値は 0.05 以下となり有意差があるといえる．また，従来手法による低減後 - 提案手法による低減後間 p 値は 0.01 以下となり有意差が認められた．

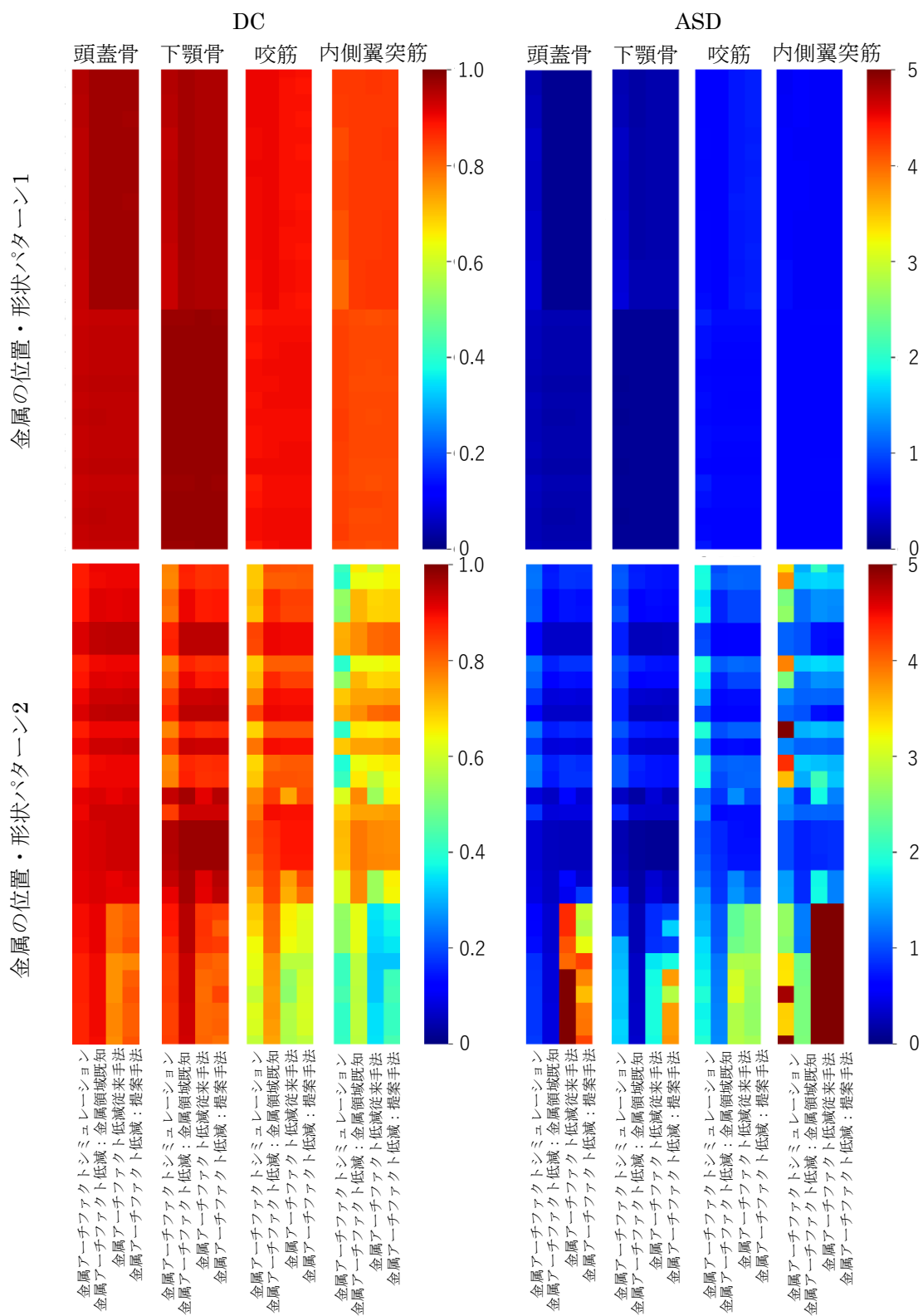


図 4.9: シミュレーションによるアーチファクト低減前後画像による筋骨格セグメンテーション結果

4.3.3. 実臨床 CT 画像データベースへの適用結果

前述のアーキテクチャ 1 による 4-fold 交差検証によって得られた学習済みモデルのうちの一つを用い，実臨床 CT 画像データベース内の CT 画像で筋骨格自動セグメンテーションの推論を行った．本ステップの筋骨格自動セグメンテーション結果と，前ステップのランドマーク自動抽出結果を同時に表示させると，図 4.10 のように筋骨格領域とランドマーク位置を同時に確認することができる．

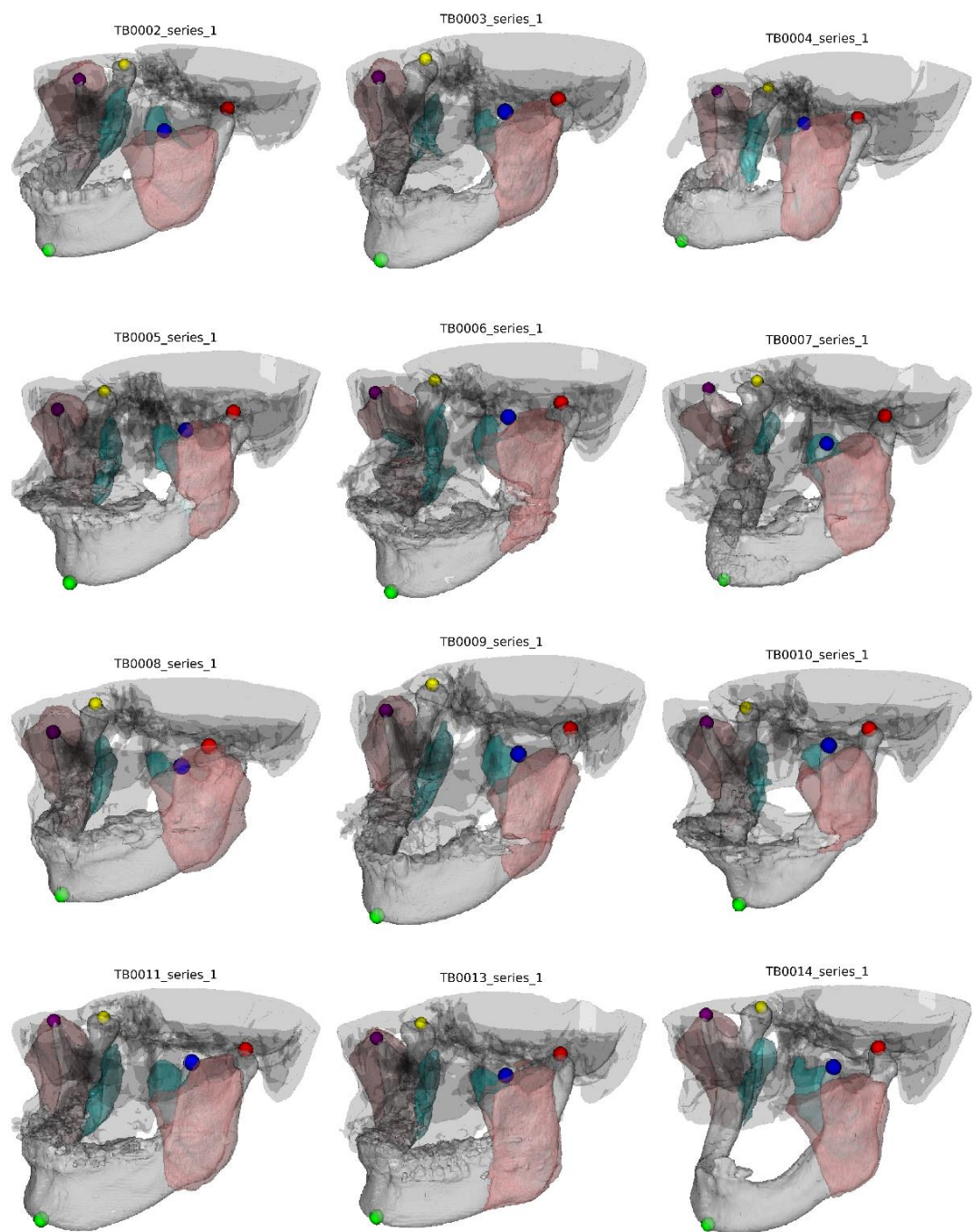


図 4.10: 実臨床 CT 画像データベースにおける筋骨格自動セグメンテーション結果
およびランドマーク自動検出結果（一部）（次項へ続く）

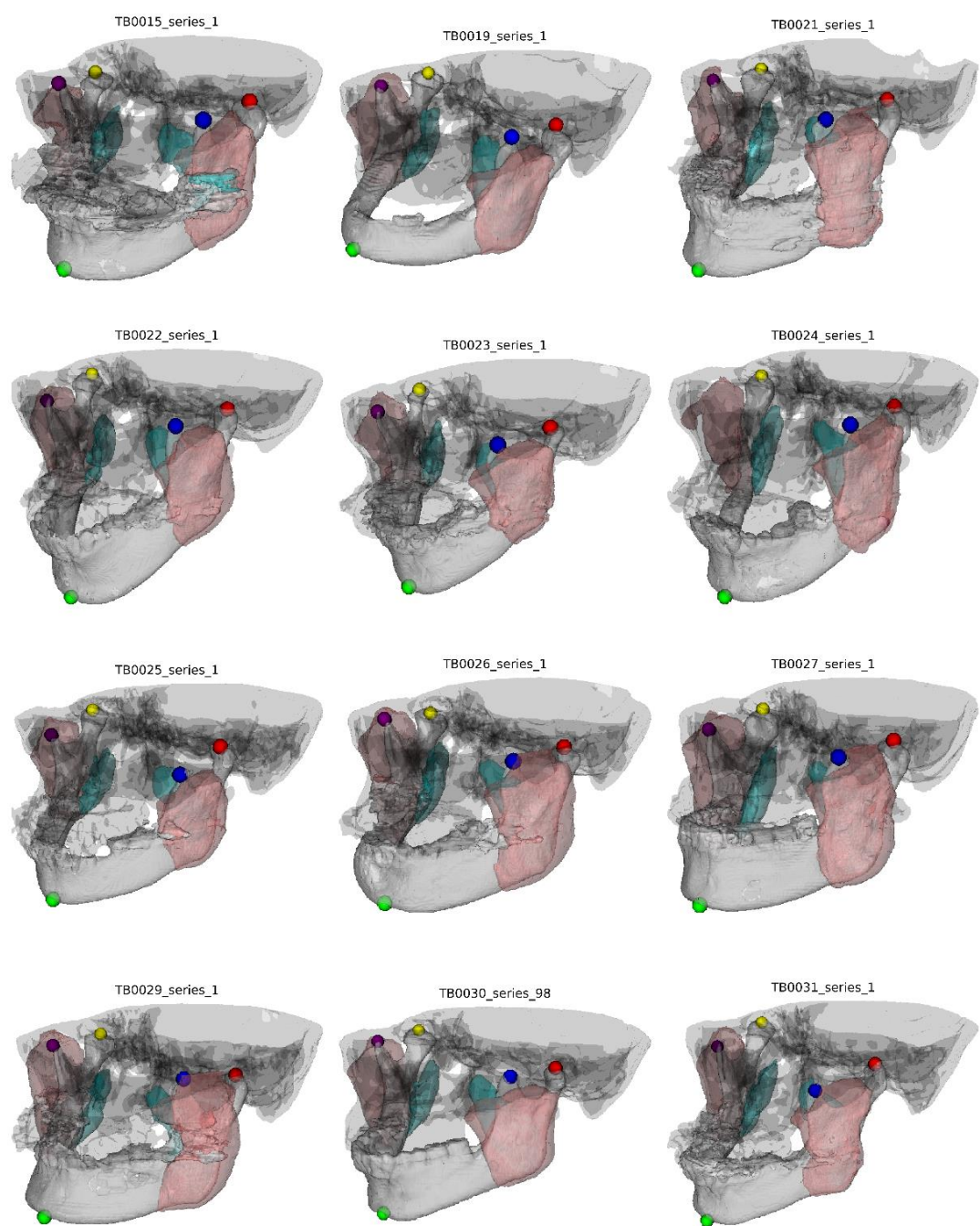


図 4.10: 実臨床 CT 画像データベースにおける筋骨格自動セグメンテーション結果およびランドマーク自動検出結果（一部）（前項の続き）

4.4. 下顎骨および咀嚼筋の統計解析

4.4.1. 予備実験：金属アーチファクトのない CT 画像での計測結果

金属アーチファクトのない CT 画像について、35 症例での顎頭間距離の計測および 36 症例での各筋骨格体積の計測を行った結果を表 4.1 および図 4.11 に示す。ここで、正解座標およびラベルを用いて算出した距離や体積と推論結果を用いて算出した距離や体積の間で Wilcoxon の符号順位検定により p 値を求めると、次のことが分かった。

- 顎頭間距離について、正解座標を用いて算出した距離と推論結果を用いて算出した距離の間では $p = 0.077 > 0.05$ となり、有意差はないといえる。
- 下顎骨の体積について、正解ラベルの体積と推論結果の体積の間では $p = 0.232 > 0.05$ となり、有意差はないといえる。
- 咬筋の体積について、正解ラベルの体積と推論結果の体積の間では $0.01 < p = 0.022 < 0.05$ となり、有意差がある。
- 内側翼突筋の体積について、正解ラベルの体積と推論結果の体積の間では $p = 0.140 > 0.05$ となり、有意差はないといえる。

表 4.1: 金属アーチファクトのない CT 画像における顎頭間距離および各筋骨格体積の算出結果。表中 p 値は、正解データを用いて算出した結果群と推論結果を用いて算出した結果群の間で Wilcoxon の符号順位検定により求めた p 値を表す。

項目	正解データ (平均値)	推論結果 (平均値)	p 値
顎頭間距離 [mm]	101.1	100.3	0.077
下顎骨の体積 [mm ³]	88.1×10^3	87.3×10^3	0.232
咬筋の体積 [mm ³]	48.0×10^3	47.0×10^3	0.022
内側翼突筋の体積 [mm ³]	17.9×10^3	17.4×10^3	0.140

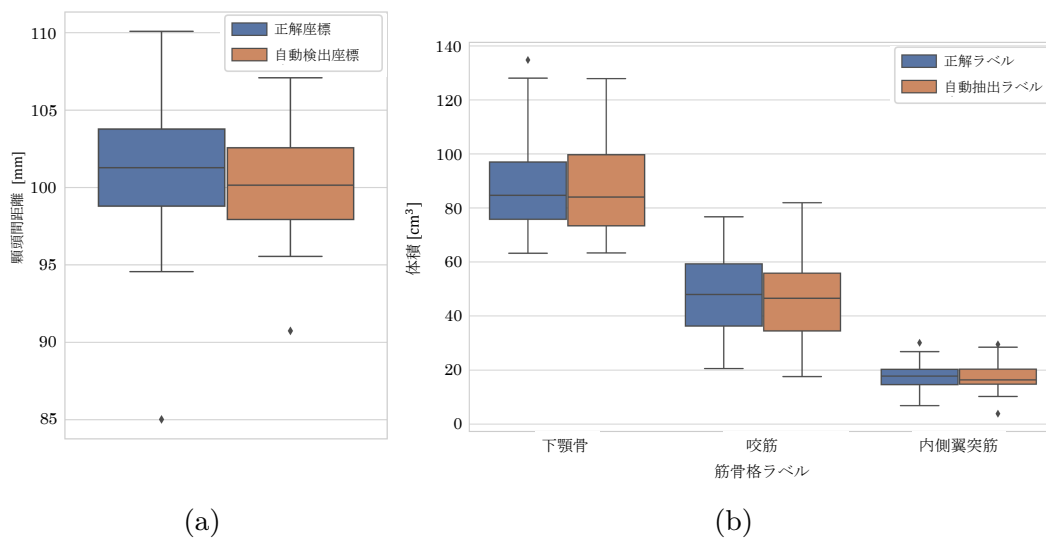
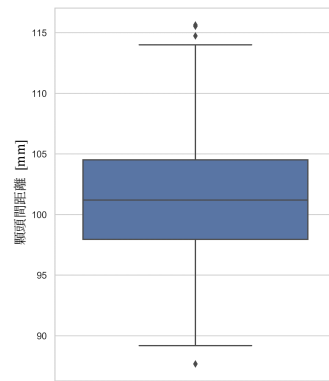


図 4.11: 金属アーチファクトのない CT 画像における顎頭間距離および各筋骨格体積の計測結果. (a) 顎頭間距離, (b) 各筋骨格体積.

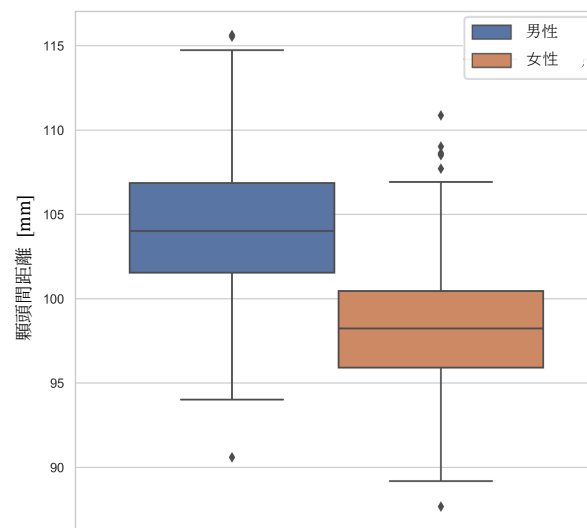
4.4.2. 実臨床 CT 画像データベースへの適用結果

これまでの処理から得られたランドマークおよび筋骨格の情報を用いて、顎頭間距離および各筋骨格の体積の計測を行う。本研究にて使用した実臨床 CT 画像データベース内の、CT 撮影に使用されたフィルタが軟組織強調フィルタである CT 画像全 742 症例（男性：388 名，354 名）について、顎頭間距離および各筋骨格体積の計測結果は図 4.12，図 4.13，図 4.14 に示すとおりである。顎頭間距離の計測結果について、前小節の実験で使用した金属アーチファクトのない CT 画像の正解座標による計測結果および関連研究 [20]，[21] に示されている計測結果を、比較のため本実験結果と並べて表 4.2 に示す。また、咬筋および内側翼突筋の体積の計測結果についても同様に、前小節の実験で使用した金属アーチファクトのない CT 画像の正解ラベルでの計測結果および関連研究 [22] に示されている計測結果を、比較のため本実験結果と並べて表 4.3，表 4.4 に示す。なお、咬筋および内側翼突筋の体積計測について、関連研究では各筋肉左右別で計測を行っているが、本研究では左右合算の計測を行っている。

本実験より、本データベースにおける顎頭間距離は、男性の平均値は 104.1mm，女性の平均値は 98.3mm，全体の平均値は 101.3mm と算出された。また、下顎骨の体積は、男性の平均値は 94.5cm³，女性の平均値は 77.0cm³，全体の平均値は 86.2cm³ と算出され、咬筋および内側翼突筋については、表 4.3，表 4.4 に示すとおりとなった。さらに、図 4.12 および図 4.14 より、顎頭間距離および筋骨格体積には性別間で差があるように見えることから、まず男女別データ群間の等分散性について F 検定により確認し、等分散性の場合は Student の t 検定、不分散性の場合は Welch の t 検定を用いて p 値を算出した。このとき、表 4.5 に示すように、いずれも p 値は 0.01 を大きく下回った。ゆえに、本研究にて得た顎頭間距離および各筋骨格の体積は性別によって有意差があるといえる。

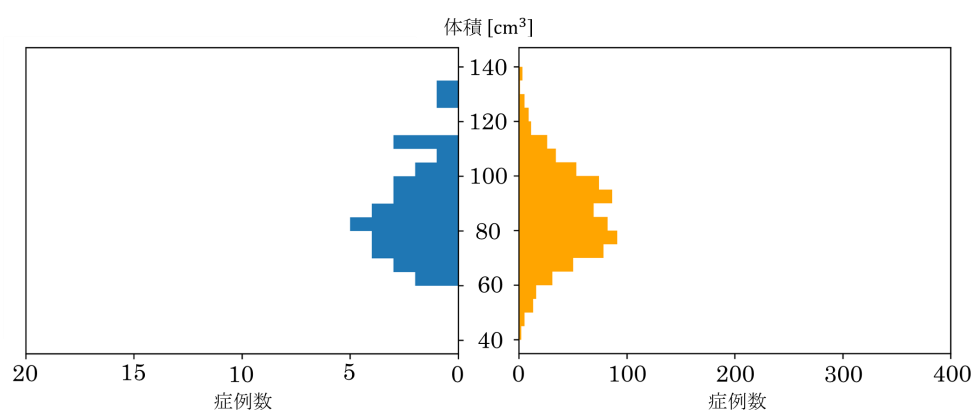


(a)

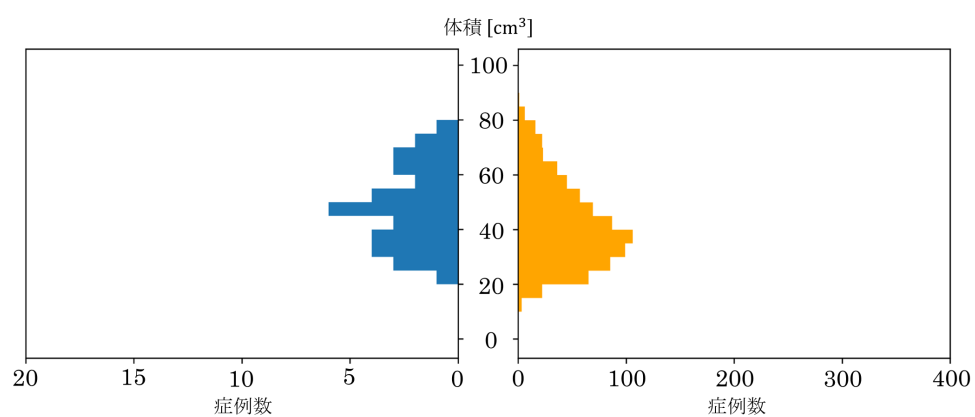


(b)

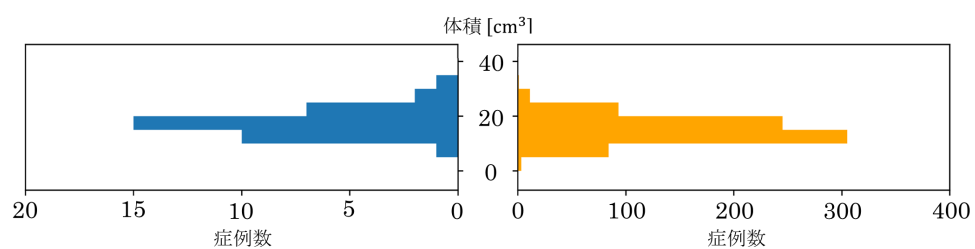
図 4.12: 実臨床 CT 画像データベースにおける顎頭間距離の計測結果. (a) 男女合計, (b) 男女別.



(a)



(b)



(c)

図 4.13: 実臨床 CT 画像データベースにおける各筋骨格の体積の計測結果ヒストグラム（男女合計）。オレンジ色：本データベースでの計測結果，青色：前小節の実験で使用了金属アーチファクトのない CT 画像の正解ラベルでの計測結果．(a) 下顎骨，(b) 咬筋，(c) 内側翼突筋．

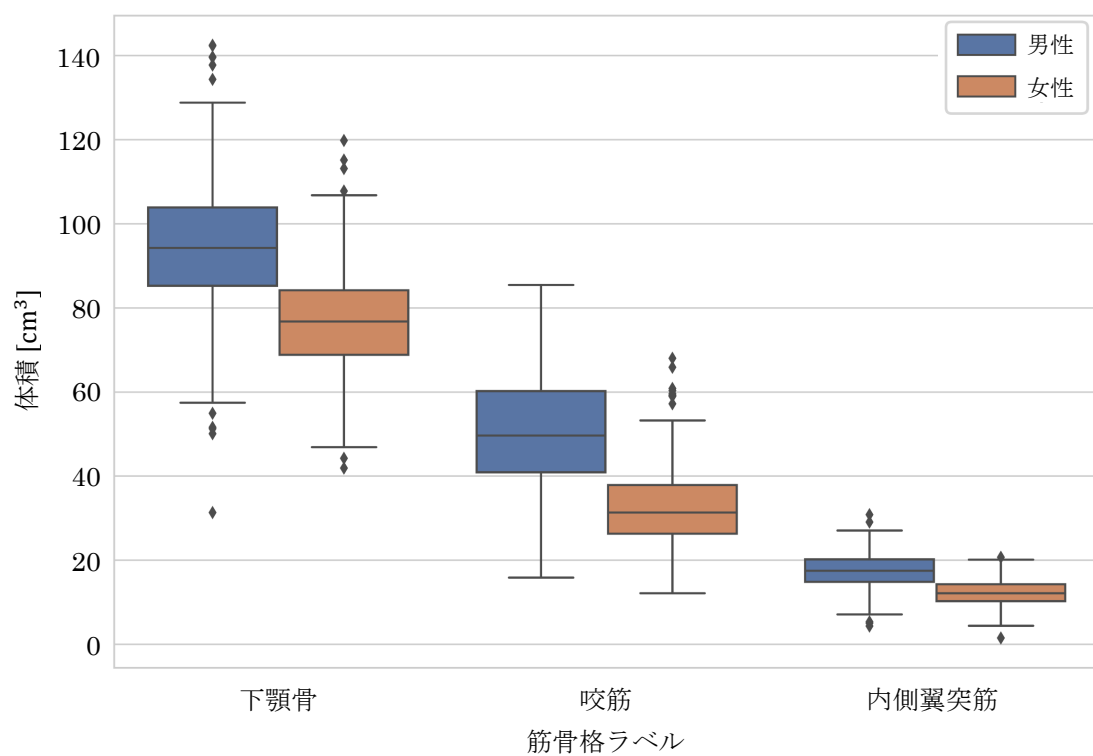


図 4.14: 実臨床 CT 画像データベースにおける各筋骨格の男女別の体積計測結果

表 4.2: 実臨床 CT 画像データベースにおける顎頭間距離の計測結果. 距離の単位は [mm].

項目		計測人数	平均値	標準偏差	最大値	最小値
本研究（左右合計）						
本実験	男性	388	104.1	3.92	115.6	90.6
	女性	354	98.3	3.70	110.9	87.7
	男女計	742	101.3	4.80	115.6	87.7
予備実験	男女計	36	101.1	4.77	110.1	85.0
関連研究 1（本研究とは若干異なる位置） [20]						
	男性	116	128.4	-	141.9	113.1
	女性	106	121.7	-	137.0	105.1
関連研究 2 [21]						
	男性	15	107	5.3	119	96
	女性	15	98	3.0	102	92
	男女計	30	103	6.1	119	92

表 4.3: 実臨床 CT 画像データベースにおける咬筋の体積の計測結果. 体積の単位は $[\text{cm}^3]$.

項目		計測人数	平均値	標準偏差	最大値	最小値
本研究（左右合計）						
本実験	男性	388	50.51	14.24	85.47	15.87
	女性	354	32.73	9.242	68.05	12.15
	男女計	742	42.03	15.02	85.47	12.15
予備実験	男女計	36	47.97	14.74	76.71	20.56
関連研究（左右別）[22]						
	男性	14	36.66	6.34	49.84	26.88
	女性	3	28.90	13.29	44.20	20.12

表 4.4: 実臨床 CT 画像データベースにおける内側翼突筋の体積の計測結果. 体積の単位は $[\text{cm}^3]$.

項目		計測人数	平均値	標準偏差	最大値	最小値
本研究（左右合計）						
本実験	男性	388	17.41	4.02	30.86	4.36
	女性	354	12.27	2.95	20.72	1.54
	男女計	742	14.96	4.38	30.86	1.54
予備実験	男女計	36	17.93	4.57	30.06	6.80
関連研究（左右別）[22]						
	男性	19	11.96	1.54	15.06	9.18
	女性	8	7.98	2.51	13.15	5.88

表 4.5: 男女別データ群間の F 検定および t 検定結果

項目	p 値 (F 検定)	性別データ群間	適用する t 検定	p 値 (t 検定)
顎頭間距離	0.136	等分散性	Student	2.51×10^{-75}
下顎骨の体積	0	不分散性	Welch	4.29×10^{-54}
咬筋の体積	0	不分散性	Welch	7.71×10^{-72}
内側翼突筋の体積	0	不分散性	Welch	8.51×10^{-71}

第 5 章 考察

5.1. 金属アーチファクト低減

実臨床 CT 画像における金属アーチファクト低減に関しては，提案法を適用すると従来法適用時よりその精度の定性的な向上がみられた一方，金属アーチファクトシミュレーションにより疑似的にアーチファクトを生じさせた画像に提案法を適用した場合，提案法を用いても図 5.1 に示すように低減に失敗する場合があった．本提案法は，本研究にて使用した CT 実画像に対しては有効であったが，本研究にて使用したデータセットとは異なる CT 画像に適用する場合については，どの程度有効であるのか検証できていない．また，本研究にて行ったシミュレーションによる金属アーチファクトが，実際に生じる金属アーチファクトとどれほど類似しているのかという検証も必要であると考える．一方，図 5.1 に示すように，従来法による低減では失敗してしまうアキシャル断面であっても，金属領域が既知である場合は NMAR を適用すると金属アーチファクト低減に失敗しなかったことから，関連研究 [11] で述べられているとおり金属アーチファクト低減において金属領域の認識が重要であることが確認できた．

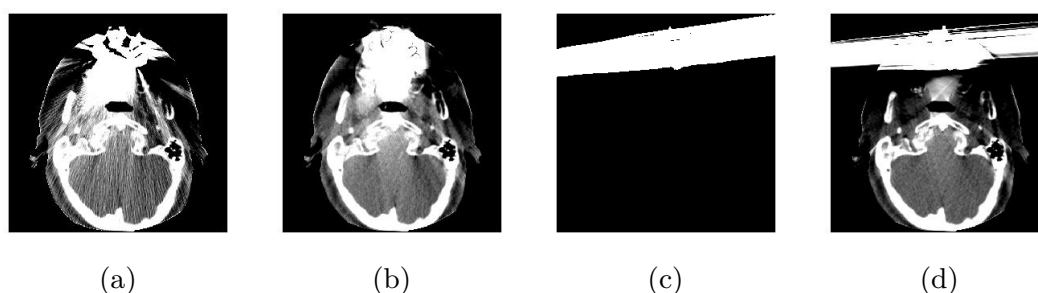


図 5.1: 金属アーチファクトシミュレーション画像に提案法を適用し低減に失敗する例の低減前後 (WL=100, WW=500)．(a) 金属アーチファクトシミュレーション画像 (低減前)，(b) 金属既知の場合の低減後，(c) 従来手法適用後，(d) 提案手法適用後．

5.2. 下顎骨および咀嚼筋の自動セグメンテーション

金属アーチファクトのない CT 画像の筋骨格自動セグメンテーションにおける学習モデルの精度評価にて，アーキテクチャ 1 を用いた場合のセグメンテーション精度が最も高くなった理由の一つは，パッチサイズが他のアーキテクチャと比較し小さいため，より多くのパッチサイズを適用できたからであると考えられる．また，2D u-net のみを用いたアーキテクチャ 5, 6 では，アキシャル断面 1 スライスずつ推論するため連続するスライス間で推論結果にばらつきが生じ，また， z 方向の情報を学習できないことから，特に歯の部分で下顎歯列と上顎歯列の識別に課題があった．一方，畳み込み後のゼロパディングを行わない 3D u-net を用いたアーキテクチャでは，出力パッチ間のアーチファクトが見られないという特徴がある．ゆえに，筋骨格自動セグメンテーションではアーキテクチャ 1 が有用であると考えられる．

5.3. 下顎骨および咀嚼筋の統計解析

5.3.1. 予備実験の結果について

予備実験にて，咬筋の体積についてののみ，正解ラベルの体積と推論結果の体積の間で有意差が認められたが，これは図 5.2 のように学習データの中にもこのように撮影範囲の異なる症例が数症例あり，推論時に撮影範囲が広い CT 画像をそのまま入力すると咬筋の過抽出が生じてしまうことが原因であると考えられる．

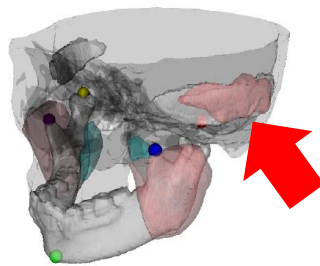


図 5.2: 金属アーチファクトのない CT 画像における筋骨格自動セグメンテーションの失敗例

5.3.2. 本実験の結果について

まず、顎頭間距離について、解剖学的にその距離は 10cm とされている。また、表 4.2 に示しているとおり、関連研究 [20] では、性別による有意差は示されなかったが男性平均 128.4mm，女性平均 121.7mm となり，女性平均より男性平均の方が 6.7mm 長いという結果が示されている。なお，男性平均値および女性平均値が本実験の計測結果と比較して長いのは，この関連研究では本研究と若干異なる位置を計測しているからである（図 5.3 参照）。さらに，同じく表 4.2 に示しているとおり，関連研究 [21] では，性別による有意差についての言及は記載されていなかったが男性平均 107mm，女性平均 98mm となり，女性平均より男性平均の方が 9mm 長いという結果が示されている。これらの結果を踏まえ，本研究にて得た顎頭間距離の男性平均 104.1mm，女性平均 98.3mm，および全体での平均 101.3mm はいずれも妥当であると考えられる。また，性別間で有意差が認められたことも妥当ではないかと考えられる。

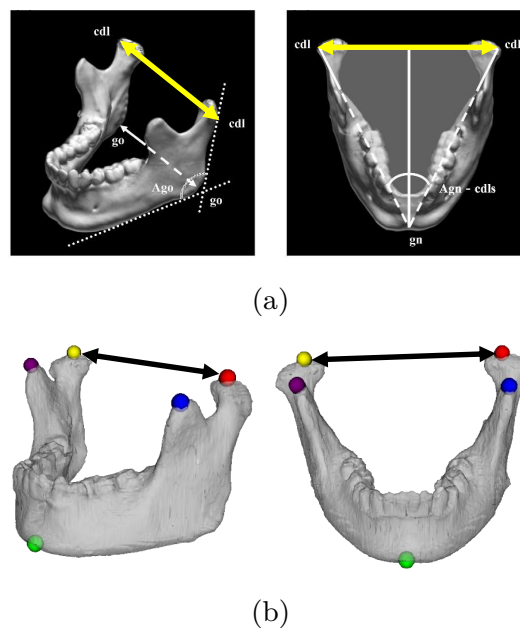


図 5.3: 関連研究 [20] と本研究の計測位置の違い. (a) 関連研究 [20] での計測位置（黄色矢印部分）（文献 [20]p.169 Fig.1 に一部加筆），(b) 本研究での計測位置（黒色矢印部分）.

次に、咀嚼筋の体積の計測結果について考察する。関連研究 [22] では、平均年齢 24 歳 3 ヶ月の男子 43 名、23 歳 2 ヶ月の女子 18 名の対象者の中から表 4.3 および表 4.4 に示す人数に対して各筋肉の体積を計測したところ、表 4.3、表 4.4 に示すとおり咬筋の体積の男性平均は 36.66cm^3 、女性平均 28.90cm^3 、内側翼突筋の体積の男性平均は 11.96cm^3 、女性平均は 7.98cm^3 となり、咬筋、内側翼突筋ともに女性より男性の方が大きいという結果が示されている。このとき、咬筋の体積は左側咬筋と右側咬筋のうち片側のみの体積を指し、内側翼突筋の体積も同様に片側のみの体積を指している。しかし、本研究では咬筋の体積は左側咬筋と右側咬筋の合計を指し、内側翼突筋の体積も同様に左右の合計の体積を指す。ここで、本研究にて得た各咀嚼筋の体積は、咬筋の体積の男性平均が 50.51cm^3 、女性平均 32.73cm^3 、内側翼突筋の体積の男性平均が 17.41cm^3 、女性平均 12.27cm^3 であり、関連研究 [22] にて示された各体積を 2 倍した値と比較すると、いずれの筋肉も関連研究の計測結果より小さいことが読み取れる。この原因として考えられるのは次の 2 点である。1 点目は金属アーチファクトの影響、2 点目は関連研究と本研究では計測の対象とした年齢層が異なることである。

1 点目について、本研究で用いた実臨床 CT 画像には金属アーチファクトが含まれており、低減を行っても金属アーチファクトが残存してしまう場合が多く、筋肉の自動セグメンテーションにおいて図 5.4 のように一部が欠けた結果となる場合があった。このような症例がデータベース内に多く存在する場合、本研究の計測結果が関連研究のものより小さくなる可能性がある。しかし、表 4.3、表 4.4 中の予備実験結果より、金属アーチファクトのない CT 画像データセット 36 症例の咬筋および内側翼突筋の正解ラベルを用いた体積計測結果の平均値はそれぞれ 47.97cm^3 および 17.93cm^3 であり、これらと関連研究 [22] の結果を比較した場合も、いずれの筋肉も関連研究の計測結果より小さいことがわかる。また、図 4.13 より、金属アーチファクトのない CT 画像データセット 36 症例の筋骨格の正解ラベルと、金属アーチファクトのある実臨床 CT 画像データベース 742 症例における筋骨格の自動セグメンテーション結果について、各データセット内での各筋骨格の体積分布を確認すると、実臨床 CT 画像データベース内で筋骨格自動セグメンテーションに明らかに失敗している症例の数が多いわけではないといえる。よって、本研究で得た

筋骨格体積は、金属アーチファクトの影響によって実際の体積よりいくらか低く算出されている可能性はあるが、これだけが原因ではないと考えられる。

2 点目について、本研究での計測の対象は 8 歳から 94 歳までの 742 症例（平均年齢 47.8 歳）である一方、関連研究での対象は前述のとおり男性平均年齢 24 歳 3 ヶ月、女性平均年齢 23 歳 2 ヶ月である。このため、本研究で計測した咬筋の体積が性別や年齢でどのように異なるのか確認すると、図 5.5 および図 5.6 に示すような結果となった。これらの図のみでは咬筋の体積と年齢の間に相関があるかどうか判断できないため、咬筋の体積と年齢の間の相関係数を求めると、その相関は男性では -0.0856 、女性では -0.2605 となった。ゆえに、男性では年齢と咬筋の体積にはほとんど相関がなく、女性では弱い負の相関があるといえる。

よって、本研究で得た筋骨格体積が関連研究での計測結果と比較して小さい原因として、本研究での計測対象者の平均年齢が関連研究での計測対象者の平均年齢より高いことが関係している可能性はあるが、原因の究明にはさらなる検証が必要である。

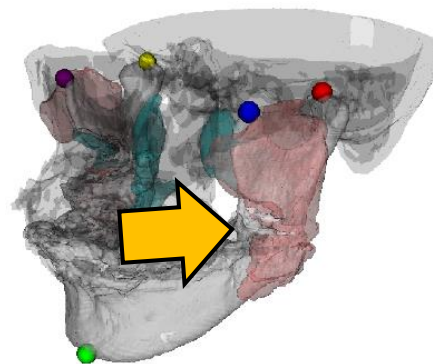


図 5.4: 金属アーチファクトの残存による筋骨格自動セグメンテーションの失敗例。黄色矢印で示した箇所において、本来咬筋が存在すると思われる部分にもかかわらず、一部が欠けたような自動セグメンテーション結果となっている。

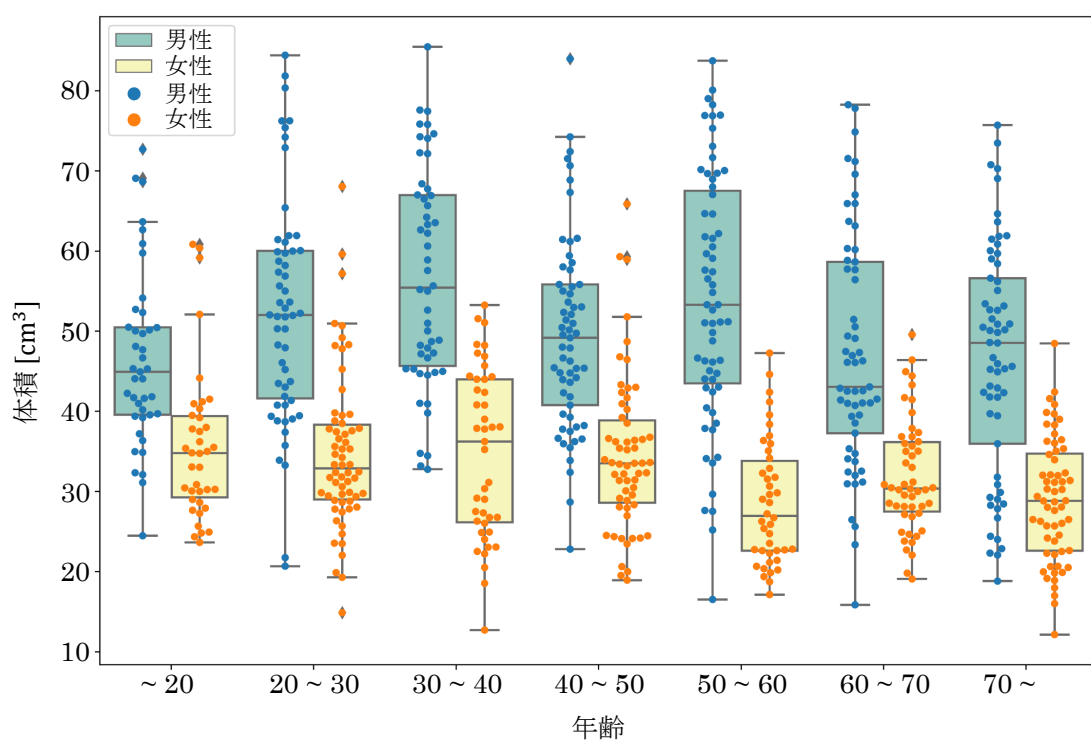
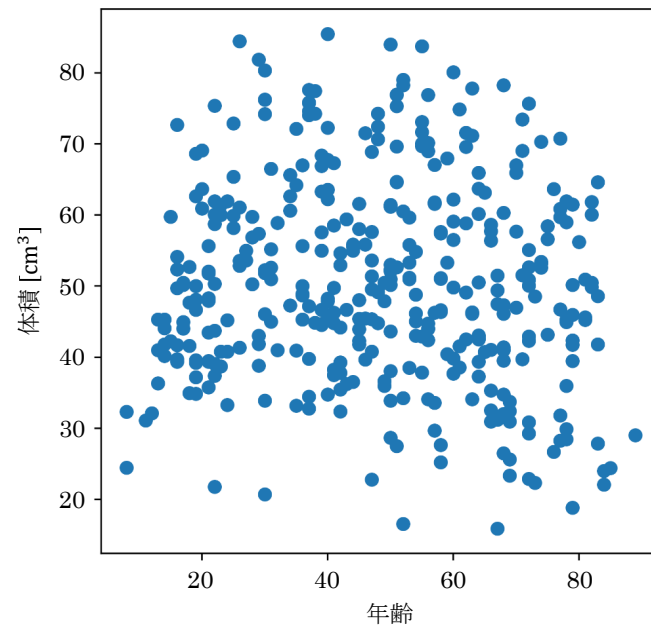
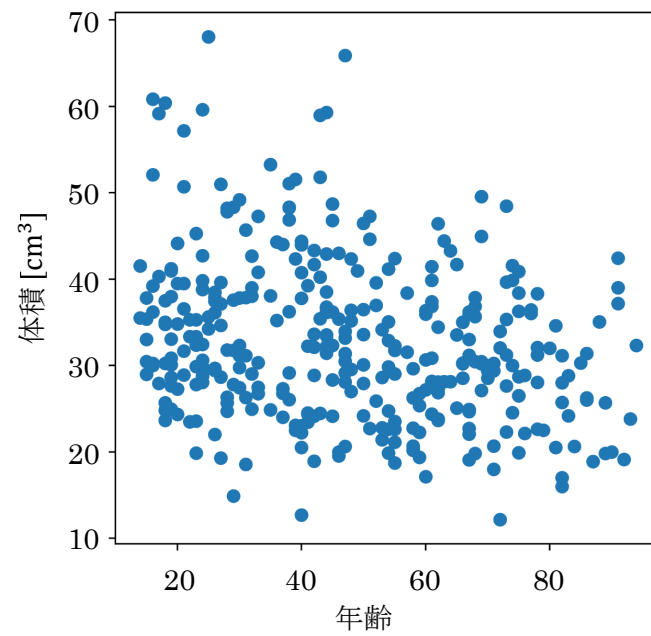


図 5.5: 実臨床 CT 画像データベースにおける性別・年齢別の咬筋の体積計測結果



(a)



(b)

図 5.6: 実臨床 CT 画像データベースにおける年齢と咬筋の体積の相関関係. (a) 男性, (b) 女性.

第 6 章 総括

本稿では，顎口腔機能異常の治療計画や術後経過の予測のための患者個別の筋骨格モデル構築を目的として，実臨床 CT 画像データベースにおける筋骨格自動セグメンテーションおよびランドマーク自動検出を提案し，この提案手法により得られた結果について統計解析を行った．提案手法にて使用する金属アーチファクト低減手法として，従来の金属アーチファクト低減手法中の金属領域の識別処理を改良した手法を提案し，金属アーチファクトシミュレーション画像および CT 実画像において，従来手法より低減精度が向上することを定量的・定性的に確認した．また，この提案法を用いて金属アーチファクト低減を行った画像に対して，ランドマーク自動検出を行い，臨床的に有用なランドマークを自動検出すると同時に症例間で異なる CT 画像の撮影範囲の統一を行った．さらに，筋骨格自動セグメンテーションについて CNN を用いたアーキテクチャを複数検討し，金属アーチファクトのない CT 画像 36 症例での交差検証を行ったところ，最も精度が高いアーキテクチャではその精度 DC が下顎骨 0.971，咬筋 0.922，内側翼突筋 0.867 となった．この学習モデルを用い，撮影範囲統一後の実臨床 CT 画像に対して筋骨格自動セグメンテーションを行った．

結果，実臨床 CT 画像データベース (742 症例) においてランドマーク間距離と各筋骨格の体積の計測を行うことができた．742 症例全体の平均顎頭間距離は 101.3mm と算出され，解剖学的にこの距離は 10cm とされていることや，関連研究での計測結果から妥当であると考えられる計測結果を得ることができたといえる．咬筋，内側翼突筋の体積についてはそれぞれ 42.03cm^3 ， 14.96cm^3 となり，金属アーチファクトのない CT 画像を用いた予備実験の結果と比較して妥当であると考えられる計測結果を得た．また，顎頭間距離の計測結果および咬筋，内側翼突筋の体積計測結果にて，性別により有意差のある検証結果を得た．提案手法により，実臨床 CT 画像データベースにおける筋骨格自動セグメンテーションおよびランドマーク自動検出，さらにそれらを用いた筋骨格体積やランドマーク間距離の計測とその統計解析を可能とすることができた．

謝辞

本研究を進めるにあたり，日頃より有益な御助言と丁寧な御指導御鞭撻を賜りました奈良先端科学技術大学院大学 情報科学領域 生体医用画像研究室 佐藤嘉伸教授に深く感謝申し上げます．また，研究発表の際に貴重な御助言を賜りました奈良先端科学技術大学院大学 情報科学領域 インタラクティブメディア設計学研究室 加藤博一教授に厚く御礼申し上げます．さらに，日頃より多くの有益な御助言と丁寧な御指導を頂きました奈良先端科学技術大学院大学 情報科学領域 生体医用画像研究室 大竹義人准教授，SOUFI Mazen 助教，日朝祐太助教，鶴見大学 歯学部 クラウンブリッジ補綴学講座 井川知子助教に心より感謝申し上げます．CT 画像の御提供ならびに多くの有益な御助言を賜りました鶴見大学歯学部をはじめ，諸先生方に厚く御礼申し上げます．本研究を遂行するにあたり，有益な御助言を頂きました大阪大学 人工知能画像診断学共同研究講座 鈴木裕紀特任助教，University of Alberta Department of Radiology and Diagnostic Imaging PostDoc. Fatemeh Abdolali，奈良先端科学技術大学院大学 情報科学領域 生体医用画像研究室の皆様 に心より感謝致します．最後に，大学院生活全般にわたり様々な面で支えて頂きました田川郁代秘書をはじめ，同研究室の皆様に深く感謝を申し上げ，ここに謝辞とさせていただきます．

参考文献

- [1] 全国歯科技工士教育協議会 [編著], “最新歯科技工士教本 顎口腔機能学,” 医歯薬出版株式会社, 2016.
- [2] 常磐肇, et al. “汎用型顎口腔機能総合解析システムの開発,” 日本顎口腔機能学会雑誌, vol.3, no.1, pp.11-24, 1996.
- [3] 山下康行 [監修], 金田隆, 中山秀樹, 平井俊範, 生嶋一郎 [編著], “画像診断別冊 KEY BOOK シリーズ 知っておきたい顎・歯・口腔の画像診断,” 学研メディカル秀潤社, 2017.
- [4] 勝又明敏, “歯科画像情報の現状と将来展望,” 日本歯科保存学雑誌, vol.62, no.5, pp.238-242, 2019.
- [5] M Yan, et al. “Symmetric convolutional neural network for mandible segmentation. Knowledge-Based Systems,” vol.159, pp.63-71, 2018.
- [6] B Qiu, et al. “Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network,” Physics in Medicine and Biology, vol.64, no.17, 2019.
- [7] H Qin, et al. “Masseter segmentation from computed tomography using feature-enhanced nested residual neural network,” International Workshop on Machine Learning in Medical Imaging, pp.355-362, 2018.
- [8] N Torosdagli, et al. “Deep geodesic learning for segmentation and anatomical landmarking,” IEEE transactions on medical imaging, vol.38, no.4, pp.919-931, 2018.
- [9] AA Kohan, et al. “The impact of orthopedic metal artifact reduction software on interreader variability when delineating areas of interest in the head and neck,” Practical radiation oncology, vol.5, no.4, pp.e309-e315, 2015.
- [10] E Meyer, et al. “Normalized metal artifact reduction (NMAR) in computed tomography,” Medical physics, vol.37, no.10, pp.5482-5493, 2010.

- [11] MAA Hegazy, et al. “U-net based metal segmentation on projection domain for metal artifact reduction in dental CT,” Biomedical engineering letters, vol.9, no.3, pp.375-385, 2019.
- [12] O Ronneberger, et al. “U-net: Convolutional networks for biomedical image segmentation,” International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp.234-241, 2015.
- [13] Ö Çiçek, et al. “3D U-Net: learning dense volumetric segmentation from sparse annotation,” International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp.424-432, 2016.
- [14] C Payer, et al. “Integrating spatial configuration into heatmap regression based CNNs for landmark localization,” Medical Image Analysis, vol.54, pp.207-219, 2019.
- [15] WANG, et al. “A benchmark for comparison of dental radiography analysis algorithms. Medical image analysis,” vol.31, pp.63-76, 2016.
- [16] Y Zhang, et al. “YU, Hengyong. Convolutional neural network based metal artifact reduction in x-ray computed tomography,” IEEE transactions on medical imaging, vol.37, no.6, pp.1370-1381, 2018.
- [17] 関西北陸地区歯科技工士学校連絡協議会 [編著], “要点チェック歯科技工士国家試験対策 3 顎口腔機能学 全国統一国家試験完全対応,” 医歯薬出版株式会社, 2015.
- [18] L Luo, et al. “Adaptive gradient methods with dynamic bound of learning rate,” arXiv preprint arXiv:1902.09843, 2019.
- [19] TY Lin, et al. “Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision,” pp.2980-2988, 2017.
- [20] T Kano, et al. “Postmortem CT morphometry with a proposal of novel parameters for sex discrimination of the mandible using Japanese adult data. Legal Medicine,” vol.17, no.3, pp.167-171, 2015.
- [21] 真山肇, “ヒトの上顎 3 角に関する研究,” 日本顎咬合学会誌, vol.4, no.1, pp.31-38, 1983.

- [22] 松島静吾, “咀嚼筋の体積および筋厚と顎顔面形態との関連性に関する研究,”
岩手医科大学歯学雑誌, vol.23 no.2, pp.106-115., 1998.

発表リスト

[1] 森谷友香, Fatemeh Abdolali, 阪本充輝, 大竹義人, 重田優子, 井川知子, 三島章, 小川匠, 佐藤嘉伸, “顎口腔領域の CT 画像における金属アーチファクト低減を用いた筋骨格セグメンテーション,” 電子情報通信学会 医用画像研究会, 2019.1.

[2] 森谷友香, Fatemeh Abdolali, 阪本充輝, 大竹義人, 重田優子, 井川知子, 三島章, 小川匠, 佐藤嘉伸, “顎口腔領域の CT 画像における金属アーチファクト低減を用いた筋骨格セグメンテーション -金属アーチファクトのシミュレーションによる精度検証-,” 第 38 回日本医用画像工学会大会, 2019.7.