

第4回 とめ研究所若手研究者懸賞論文

セマンティックセグメンテーション手法を用いた
X線画像からの椎骨検出

Vertebrae Detection from X-ray Images
Using Semantic Segmentation Method

森川 大翔

Yamato MORIKAWA

和歌山大学大学院システム工学研究科
Graduate School of Systems Engineering, Wakayama University

概要

脊椎圧迫骨折は圧潰が進むと難治性の腰痛となるため、早期の発見が必要である。そのため、簡便かつ早期に診断できる単純 X 線撮影を用いた検査が多く用いられている。CT 装置や MRI 装置に比べて単純 X 線装置は安価であることから、小規模の病院にも設置されており撮影が手軽であることも、単純 X 線撮影を用いた検査が多く用いられる理由の一つである。しかし、X 線画像は不明瞭な場合が多いため、医師が常に正確に読影できるとは限らない。誤った診断は誤った方向へ治療が進んでしまうことに繋がる。脊椎疾患領域の画像診断は専門性が高く、信頼に足る画像診断ができるようになるためには多くの時間と経験が必要とされるため、Artificial Intelligence(AI)を用いた画像診断分野での活用が期待されている。

また、椎骨の X 線画像読影における医師の人数が不足している。医用画像の撮影装置は日々進歩し撮影効率も向上しているにも関わらず医師の人数が少ないため、画像診断の需給バランスが逼迫し診断の質が低下していることから、誤診や見逃しが起きる危険性が高まっている。そのため、診療現場における画像診断の効率化、支援が必要とされている。

撮影する際の体位は状況によって異なるため、体位が様々な X 線画像に適した CNN (畳み込みニューラルネットワーク) モデルを開発することで、どのような体位の X 線画像においても、迅速かつ正確な読影の支援が可能となる。しかし、これまでに X 線画像を使用した椎骨検出のための CNN モデルは様々な提案されてきたが、患者の体位が立位である X 線画像を使用している場合が多く、体位が様々な X 線画像を使用した研究は行われていない。そこで本研究では、撮影する際の患者の体位が様々な X 線画像における、CNN を用いたセマンティックセグメンテーションによる椎骨の検出手法を提案する。実際に病院で撮影された体位が様々な (立位・坐位・仰臥位) な X 線画像を使用し、セマンティックセグメンテーション用に開発された 4 種類の CNN モデルにおける椎骨の検出精度、および予測画像を比較する。またその際、CNN モデルにおけるハイパーパラメータを調整することで、より高い検出精度を目指した。

本研究では、体位が様々な X 線画像における椎骨検出のための CNN モデルの精度評価実験を行った。精度評価実験の結果より、以下の知見を得た。

1. 体位が様々な X 線画像においても、既存の CNN モデルでセマンティックセグメンテーションを行うことができる。
2. 今回のデータセットでは、U-Net が最も椎骨のセマンティックセグメンテーションに適している。
3. 一部の予測画像では、正解画像を超えた、不明瞭な部分の椎骨まで検出している可能性がある。

目次

第1章 序論	1
第2章 関連研究	3
2.1 椎骨の検出手法	3
2.2 深層学習を用いた椎骨検出	3
2.3 X線画像からの椎骨検出	4
2.4 本研究の位置づけ	5
第3章 学習モデル	7
3.1 セグメンテーション	7
3.2 使用モデル	8
3.2.1 U-Net	8
3.2.2 Residual U-Net	8
3.2.3 Dense U-Net	9
3.2.4 SegNet	9
第4章 精度評価実験	13
4.1 本手法の概要	13
4.2 使用データの概要	13
4.3 正解データ	13
4.4 評価指標	14
4.5 椎骨検出精度の評価結果と考察	15
4.5.1 ハイパーパラメータ	15
4.5.2 評価結果と考察	16
4.5.3 予測画像	18
第5章 結論	27
参考文献	28

第1章 序論

背骨の前方にある椎体が壊れて変形してしまう骨折を脊椎圧迫骨折という。骨粗鬆症による骨脆弱性（骨が脆くなる）を背景として主に高齢者にみられる外傷であり、身体を曲げたり動かしたりしたときの腰や背中の痛み（動作時痛）が主な症状である。特に、脊椎圧迫骨折は圧潰が進むと難治性の腰痛となるため、早期の発見が必要である。そのため、簡便かつ早期に診断できる単純 X 線撮影を用いた検査が多く用いられている。CT 装置や MRI 装置に比べて単純 X 線装置は安価であることから、小規模の病院にも設置されており撮影が手軽であることも、単純 X 線撮影を用いた検査が多く用いられる理由の一つである。しかし、X 線画像はコントラストが低く不明瞭な場合が多いため、医師が常に正確に読影できるとは限らない。誤った診断は誤った方向へ治療が進んでしまうことに繋がる。脊椎疾患領域の画像診断は専門性が高く、信頼に足りる画像診断ができるようになるためには多くの時間と経験が必要とされるため、Artificial Intelligence(AI)を用いた画像診断分野での活用が期待されている。Kim らの研究 [1] によると、高齢化社会の到来により骨粗鬆症の症例が増加しているため、椎体圧迫骨折の迅速かつ正確な診断が必要であり、それに対応した自動椎骨セグメンテーションモデルの開発が不可欠であるとしている。

また、椎骨の X 線画像読影における医師の人数が不足している。脊椎脊髄疾患の治療において深い知識や経験、及び指導力を有しており、脊椎脊髄分野のエキスパートである日本脊椎脊髄病学会認定脊椎脊髄外科指導医 [2] は、日本国内に約 1,670 人 [3] しかいない。また、医用画像から患者さんの病態や状況を判断し、その所見を元に病気を診断する専門家である放射線診断専門医 [4] は、日本国内に約 5,600 人 [4] しかいない。医用画像の撮影装置は日々進歩し撮影効率は向上しているにも関わらず医師の人数が少ないため、画像診断の需給バランスが逼迫し診断の質が低下していることから、誤診や見逃しが起きる危険性が高まっている。そのため、診療現場における画像診断の効率化、支援が必要とされている。

撮影する際の体位は状況によって異なるため、体位が様々な X 線画像に適した CNN¹モデルを開発することで、どのような体位の X 線画像においても、迅速かつ正確な読影の支援が可能となる。しかし、これまでに X 線画像を使用した椎骨検出のための CNN モデルは様々提案されてきたが、患者の体位が立位である X 線画像を使用している場合が多く、体位が様々な X 線画像を使用した研究は行われていない。そこで本研究では、撮影する際の患者の体位が様々である X 線画像における、CNN を用いたセマンティックセグメンテーションによる椎骨の検出手法を提案する。実際に病院で撮影された体位が様々（立位・坐位・仰臥位）な X 線画像を使用し、セマンティックセグメンテーション用に開発された 4 種類の CNN モデルにおける椎骨の検出精度、および予測画像を比較する。またその際、CNN モデルにおけるハイパーパラメータを調整することで、より高い検出精度を目指した。本論文では、使用する CNN モデルおよび手法の概要と、X 線画像における椎骨検出のための CNN モデルの精度評価実験と、その結果に対する考察について述べる。2

¹CNN : Convolutional Neural Network (畳み込みニューラルネットワーク)

章では関連研究について述べる．3章では使用する CNN モデルについて詳細な説明を行う．4章では学習手法の詳細および体位が様々な X 線画像における椎骨検出のための CNN モデルの精度評価実験について述べる．5章では本研究の結論について述べる．

第2章 関連研究

本章では、関連研究について述べる。2.1節では、深層学習が注目される以前の椎骨の検出手法に関する研究を示す。2.2節では、深層学習を用いた椎骨の検出手法に関する研究を示す。2.3節では、セグメンテーション手法を用いたX線画像からの椎骨の検出手法に関する研究を示す。2.4節では、本研究の位置づけを明らかにする。

2.1 椎骨の検出手法

Benjelloun らは、医師が選択した頸椎領域から、エッジ検出法およびコーナー検出法を用いて椎骨頂点の候補を算出し、椎体の形状と椎体同士の位置関係を用いて誤検出点を削除し、頸椎の椎体位置を特定する手法を提案した [5]。Sa らは、HOG(Histograms of Oriented Gradients) 特徴量を用いて、腰椎椎体部を4分割して学習するSVM(Support Vector Machine) と、腰椎椎体部全体を検出するSVMを学習させ、GVF(Gradient Vector Flow) ベースのスネーク法を使用して椎骨の輪郭を検出する手法を提案した [6]。

このように椎骨の検出手法は以前から提案されているが、ノイズやコントラストの変化に弱いなどの問題があった。ところが、近年臨床研究において深層学習が注目されるようになり、深層学習を用いた手法が以前の手法を圧倒する性能を示したことから、深層学習による椎骨のセグメンテーションに関する研究が増加している。本研究ではCNNを用いることで、深層学習による椎骨の検出手法を提案する。

2.2 深層学習を用いた椎骨検出

Lessmann らは、完全畳み込みネットワークを用いてCT画像から胸腰部の椎骨を検出する自動セグメンテーションモデルを提案した。[7]。既に分割された椎骨に関する情報を保持するメモリを組み合わせることで、椎骨は常に隣り合っているという事前知識を含みながら、画像とメモリの両方からの情報を用いて画像を繰り返し解析し、椎骨を探索する。これにより、高い精度で椎骨が画像内で完全に見えるかどうかを予測し、不完全に見える椎骨を除外できた。Lehnen らは、CNNを用いてMRI画像から腰椎のセグメンテーションを行い、複数の異なる症例を検出する手法を提案した [8]。具体的には、椎骨と椎間板の検出とラベリング、ならびに椎間板ヘルニア、椎間板膨隆、脊柱管狭窄症、神経根圧迫、脊椎分離症の検出において中程度から高い診断精度を示した。Cheng らは、2段階のDense U-Netを用いたCT画像における椎骨の自動位置特定とセグメンテーションのための深層学習によるアプローチを提案した [9]。第1段階では、2D-Dense U-Netを用いて、高密度なラベルと2Dスライスで椎骨の中心を検出することにより椎骨をローカライズする。第2段階では、3D-Dense U-Netを用いて中心点に基づいて特定された関心領域(ROI)内の

特定の椎骨に対してセグメンテーションを行う。最終的にセグメンテーションされた各椎骨を元の解像度に再サンプリングする。結果として、椎骨の中心点を検出することで椎骨の位置特定に成功し、脊椎全体における椎骨のインスタンス分割を可能にした。Malinda らは、CT 画像から脊椎をセグメンテーションするための完全自動セグメンテーション手法を開発し、その結果を他の CNN モデルにおけるセグメンテーション手法と比較した [10]。この手法では、CNN と FCN(Fully Convolutional Network) を組み合わせることで、クラスの冗長性を柔軟な制約として利用し、セグメンテーション結果の大幅な向上を目指した。結果として、セグメンテーション結果の精度とシステムの処理時間を大幅に向上させた。Roman らは、転移や椎体圧迫を有する患者の脊椎における形状に変異が起きている可能性のある CT 画像において、椎体の位置と識別(ラベリング)を目的とした学習アプローチに基づく完全自動システムを提案した [11]。3つの CNN を含むいくつかの高度な手法を組み合わせることで、変異が起きている脊椎や形状が歪んだ症例を含む画像でも、中程度の精度で完全に自動で動作させることができた。Siemionow らは、腰椎の CT 画像を利用した CNN における自動計測の精度を評価し、解剖学的にセグメンテーションを行った [12]。セグメンテーションされた解剖学的ランドマークは、椎弓根、椎体、棘突起、横突起、椎間関節、椎弓板などであった。手動計測と自動計測を比較し、結果として高い精度でセグメンテーションを行い、椎体の解剖学的構造を自動的に識別できた。Hallinan らは、腰椎 MRI 画像における腰椎中心管、側方陥凹、および神経孔狭窄の自動検出および分類のための CNN モデルを開発した [13]。2つの CNN を用意し、1つ目の CNN をトレーニングして関心領域(ROI)を検出し、2つ目の CNN を分類用を使用した。放射線科医によってラベル付けされた分類結果と自動検出結果とを比較し、中心管狭窄と外側陥凹狭窄の検出と分類については放射線科医と同等の一致を示し、神経孔狭窄についてはやや低い一致を示した。

このように CT 画像、MRI 画像のような高精細な画像を用いた研究は多く行われているが、それらをコントラストの低い X 線画像へ適用することは困難である。そこで本研究では、CNN を用いて X 線画像から椎骨を検出する手法を提案する。

2.3 X 線画像からの椎骨検出

Horng らは、3種類の CNN を用いて X 線画像から椎骨の検出を行い、出力したセグメンテーション画像をもとに脊椎彎曲率を自動で測定する手法を提案した [14]。Zhang らは、マスク領域ベースの CNN を用いて X 線画像における脊椎の位置を特定しセグメンテーションを行う手法を提案した [15]。Lee らは、骨粗鬆症の高リスク集団に注意を促すために、深層学習によって抽出された脊椎 X 線画像の特徴分析を通じて、骨密度異常のある個人を特定するための機械学習モデルを評価した [16]。骨密度正常者と骨密度異常者を連続的に抽出し、X 線画像の関心領域から、CNN を用いて画像特徴を生成した。その後、機械学習の分類アルゴリズムで学習させた画像特徴を用いて、骨密度異常の予測モデルを設計し、優れたパフォーマンスを発揮した。Arif らは、X 線画像における頸椎のセグメンテーションのための深層学習ベースの完全自動フレームワークを提案した [17]。順序としては、まず FCN を用いて画像中の脊椎領域をローカライズする。次に、深層確率的空間回帰ネットワークを用いて椎骨の中心をローカライズする。最後に、形状認識型深層セグメンテーションネットワークを用いて、画像中の椎骨にセグメンテーションを行う。結果として、中程度の精度で椎骨の自動セグメンテーション結果を生成することができた。Kuok らは、

前方視 (AP) 全脊椎 X 線画像から胸椎と腰椎をセグメンテーションするためのハイブリッド手法を提案した [18]. 画像処理技術を適用して椎骨領域を検出し, CNN を用いて椎骨のセグメンテーションを行う. 提案手法によるセグメンテーション性能はかなり高い結果を示した. Cinal らは, 深層学習を用いて胸腰椎の矢状 X 線画像における椎骨の角の座標を自動的に算出し, それらのランドマークから, L1-L5, L1-S1 前弯, 仙骨傾斜などの放射線学的パラメータを算出することを提案した [19]. まず, 2つの CNN を学習させ, 画像中の各椎骨を識別し, ランドマーク座標をそれぞれ計算させる. その後, 1つの椎骨を切り出した画像を別の CNN の入力として用い, 算出した椎骨の角を元の画像に対応させることで定位処理を精緻化した. 結果として, 椎骨の角の座標を高い精度で計算することができ, 臨床作業における測定の信頼性と再現性を向上させる大きな可能性を示した.

このように X 線画像を使用した椎骨の検出手法は提案されているが, 画像を撮影する際の体位が立位である場合が多く, 坐位や臥位などの画像を使用した研究は行われていない. 本研究では, 撮影の際の患者の体位が様々である X 線画像を使用して椎骨を検出する.

2.4 本研究の位置づけ

図 2.1 に本研究の位置づけを示す. 深層学習が注目される以前にも, X 線画像からの椎骨の検出手法はいくつか提案されてきたが, ノイズやコントラストの変化に弱いなどの問題があった. 近年, 深層学習による手法が注目されてからは, 深層学習による椎骨のセグメンテーションに関する研究が増加している. しかし, 体位が立位である X 線画像を使用している研究しか行われていない.

よって, 本研究と関連研究との大きな違いは, 体位が様々な X 線画像に着目して椎骨の検出を行ったという点である.

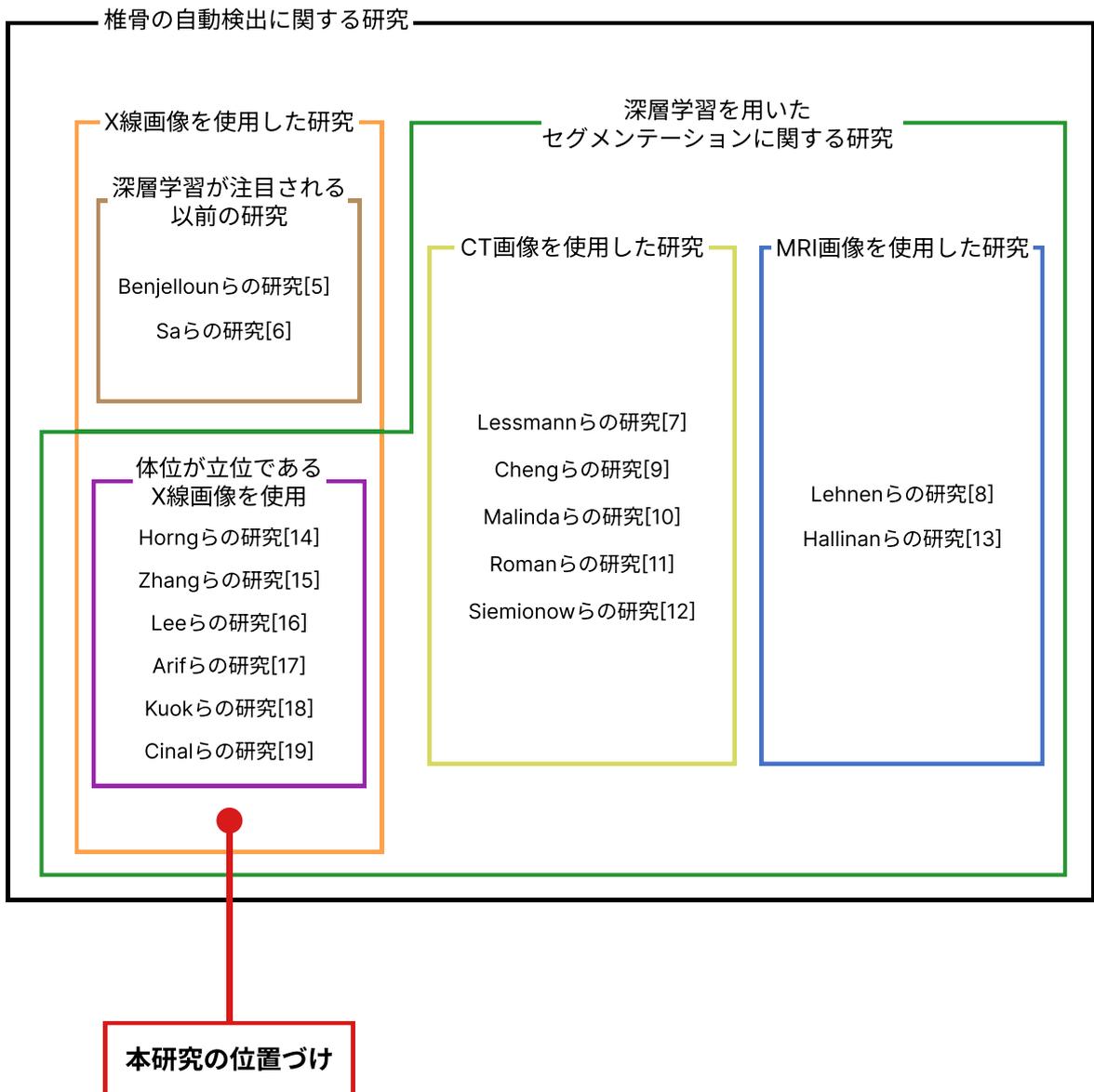


図 2.1: 本研究の位置付け

第3章 学習モデル

本章では、セグメンテーション手法の詳細について述べる。また、使用した学習モデルについて述べる。3.1節ではセグメンテーション、3.2節では使用したモデルについて述べる。

3.1 セグメンテーション

セグメンテーションとは、特徴などに関して注目している部分領域に画像を分割する処理のことである。近年の画像処理では、画像内の対象とする物体と背景の領域を分割してピクセル単位で認識するとともに、画像内のどこにどのような物体が写っているかというクラスの分類を行う。つまり、領域とクラスを同時に認識するタスクを実質的に意味する。セグメンテーションの手法については、古くから種々の処理手法が提案されてきた。しかし、近年CNNが注目を集め始め、セグメンテーションでもCNNにより得られた特徴マップを用いた手法が提案されたことから、それ以降は深層学習による手法が主流となっている。

セグメンテーションには、セマンティックセグメンテーション、インスタンスセグメンテーション、パノプティックセグメンテーションの3つの手法が存在する。

(1) セマンティックセグメンテーション

セマンティックセグメンテーションは、画像全体や画像の一部の検出ではなく、ピクセル（画素）一つひとつに対して示す意味をラベル付けし、領域を分割する手法である。画像のピクセルがどのクラスに属するかを分類し、何が写っているかのラベル付けやカテゴリ関連付けを行う。物体をクラスごとに領域分割するため、物体が重なっているときにはそれぞれの区別ができないが、空、道路、標識など不定形の領域を検出することが可能である。

セマンティックセグメンテーションは特に、建物の劣化における損傷箇所や、医用画像における疾患箇所など、細かな領域の抽出・特定が必要となる場合のモデルなどで活用されている。

(2) インスタンスセグメンテーション

インスタンスセグメンテーションは、画像の中にある物体の領域を特定し個体ごとに領域分割することで、物体の種類を認識できる手法である。RoI (region of interest/関心領域) に対してセグメンテーションを行うため、画像全てのピクセルに対してラベルを振らない。また、不定形の領域は扱えないが、隣接した同種類の物体は区別できる。

(3) パノプティックセグメンテーション

パノプティックセグメンテーションは、セマンティックセグメンテーション、インスタンスセグメンテーションを組み合わせた手法である。全てのピクセルにラベルが振られ、数えられる物体に関しては個別で認識した結果が返される。

X線画像における椎骨検出では、椎骨ごとに異なるクラスとして検出したい場合は、インスタンスセグメンテーションによる手法を用いるべきだが、本研究は椎骨そのものを検出することを目的としており、椎骨を異なる物体として個別に認識する必要はない。また、椎骨の形状など細かな領域の抽出が必要となるため、今回はセマンティックセグメンテーション手法を用いて検出する。

3.2 使用モデル

本研究では、U-Net[20]、Residual U-Net[21]、Dense U-Net[22]、SegNet[23]の4種類のCNNを使用し、椎骨のセグメンテーションを行った。これらは全て、セマンティックセグメンテーション用に開発されたモデルである。

3.2.1 U-Net

FCNが応用されており、生物医学のために開発されたセマンティックセグメンテーション用のモデルである。FCNとは、全結合層を使用せず、ネットワークが全て畳み込み層のみで構成されているCNNの一種である。U-Netの構造を図3.1に示す。U-Netの各層は、アルファベットの「U」の形をしたエンコーダ-デコーダ形式の、対称的な構造を成している。エンコーダ側では、畳み込みを4回実施する。また、最大値プーリング後に、畳み込み層のチャンネル数を、64, 128, 256, 512と増加させていき、各特徴マップをデコーダの同一空間サイズの特徴マップにスキップ接続でコピーする。デコーダ側では、逆畳み込みと呼ばれる通常の畳み込みと逆の処理を4回実施する。また、逆畳み込みごとに、畳み込みチャンネル数を512, 256, 128, 64と減少させていく。さらに、スキップ接続でエンコーダ側からコピーされた特徴マップを、後方チャンネルに結合することにより、エンコーダ側の特徴マップの位置情報をデコーダ側へショートカットして受け渡し、デコーダ側でも使用する。このスキップ接続の処理により、エンコーダ側の大きな特徴マップの情報がデコーダ側に伝わるようになり、逆畳み込み時に物体の位置情報を捉えやすくなっている。

3.2.2 Residual U-Net

基本的な構造はU-Netと同じだが、U-Netにおける畳み込み処理層をResidualブロックに置き換えることによりパフォーマンスを向上させたモデルである。Residualブロックの構造を図3.2に示す。Residualブロックとは、ResNetにおける基本構造の一つであり、並列にスキップ接続を挿入することにより、2分岐のどちらかを選んだ多様な経路を信号・誤差が伝搬可能となる。この構造により、奥の層まで直接接続されるようになるため、勾配消失問題や劣化問題を防ぎながら深い層でも学習が行えるようになった。ResNetはCNNモデルの一種であり、Residualブロックの導入により、深い層数のCNNでも劣化問題無しに高精度な学習を可能にした。また、図3.2におけるBNとはBatch Normalizationのことであり、重みのパラメータを新たな値で置き直し、内部の変数の分布が大きく変わるのを防ぐことによって、勾配消失・爆発を防ぐための手法である。

3.2.3 Dense U-Net

基本的な構造は U-Net と同じだが、U-Net におけるプーリング層を Dense ブロックに、畳み込み層を Transition ブロックに置き換えることで、より密なマージを行うモデルである。Dense ブロックの構造を図 3.3 に、den_con_layer の処理層を図 3.4 に示す。Dense ブロックは 4 つの den_con_layer 層で構成されており、den_con_layer 層ごとに 2 つの畳み込み層を持っている。また、Dense ブロックにおける各層をそれ以前の全ての層と接続することにより、U-Net で問題となっていた解像度の低下を防止する。Dense U-Net には合計で 89 個の畳み込み層が存在しているため、より深い学習が可能となり、先行研究によると一部のデータセットではパフォーマンスが向上した。ここで、den_con_layer 層の 7 層目における Dropout とは、特定のレイヤーの出力を学習時にランダムで 0 に落とすことで、一部のデータが欠損していても正しく認識ができるようにする手法である。これにより、画像の一部分における局所的な特徴が過剰に評価されてしまうのを防ぎ、モデルのロバストさを向上させることができる。

3.2.4 SegNet

FCN が応用されており、ピクセル単位でセマンティックセグメンテーションを行うために考案されたモデルである。SegNet の構造を図 3.5 に示す。エンコーダ部分は一般的な CNN モデルが採用されており、畳み込み層、プーリング層、Batch Normalization 層から構成されている。畳み込み層で画像の局所特徴を抽出し、プーリング層でダウンサンプリングをし、抽出した特徴を後ろの層に伝播する。Batch Normalization 層は、訓練データの分布を正規化することで勾配消失などを防ぐ。デコーダ部分では、エンコーダによって得られた物体の種類と大まかな位置情報を持つ低解像度の特徴マップを、アップサンプリングし、逆畳み込み処理を行うことによって、元の解像度マップに対応付けを行う。これにより、小範囲の特徴における位置情報を捉えることが可能となった。

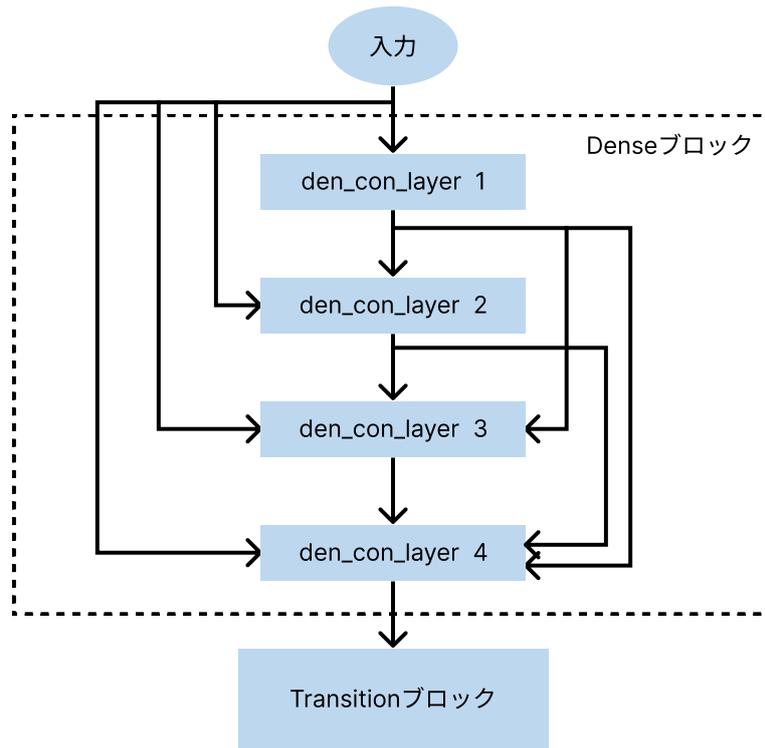


図 3.3: Dense ブロックの構造 (出典 : Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network[22] を参考に作成)



図 3.4: den_con_layer の処理層 (出典 : Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network[22] を参考に作成)

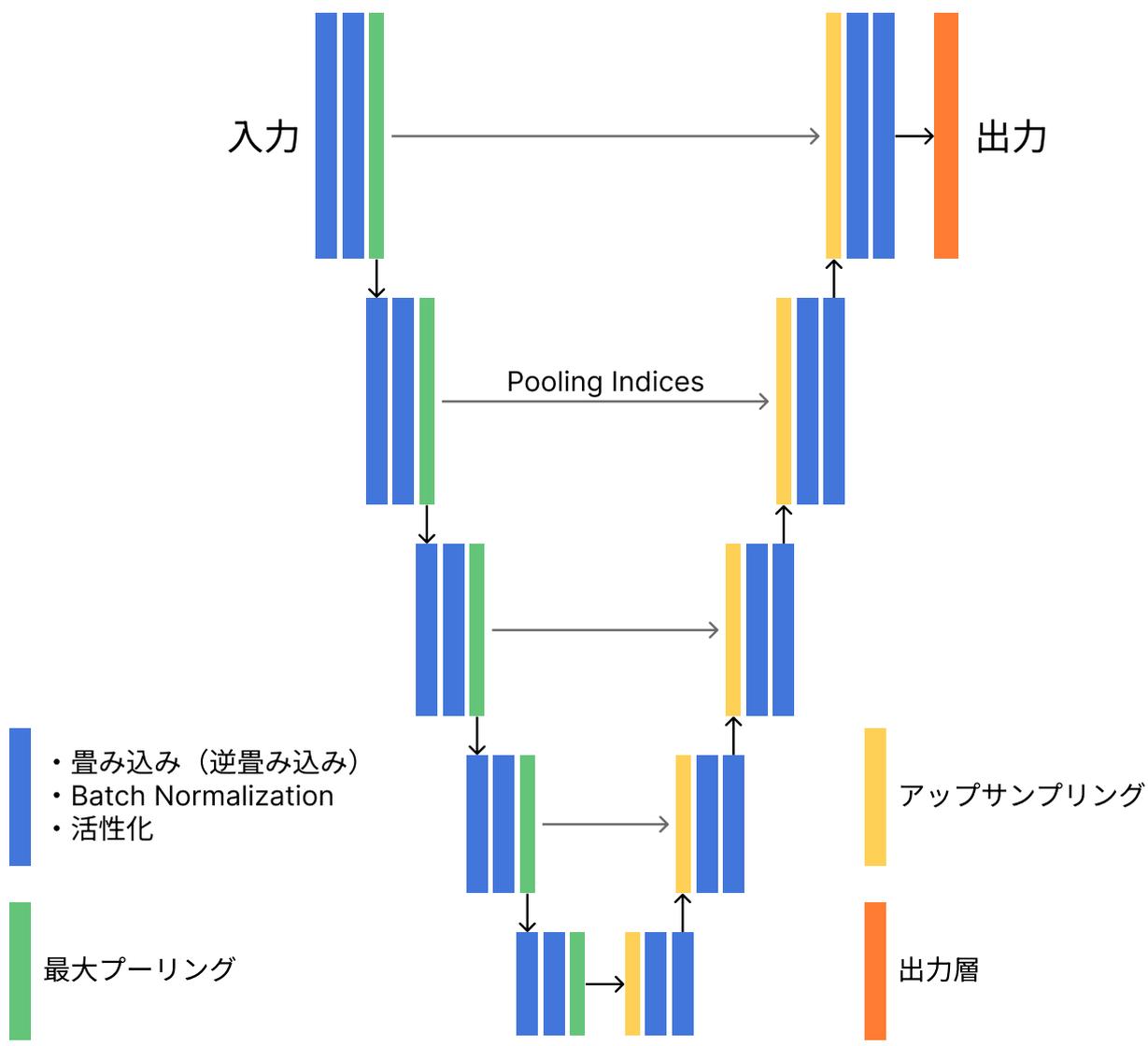


図 3.5: SegNet の構造 (出典: Segnet: A deep convolutional encoder-decoder architecture for image segmentation[23] を参考に作成)

第4章 精度評価実験

本章では、学習手法の詳細および体位が様々な X 線画像における椎骨検出のための CNN モデルの精度評価実験について述べる。4.1 節では、本手法の概要について述べる。4.2 節では、使用データの概要について述べる。4.3 節では、正解データの作成方法について述べる。4.4 節では、椎骨検出における評価指標について述べる。4.5 節では、椎骨検出精度の評価結果と考察について述べる。

4.1 本手法の概要

本手法の概要を図 4.1 に示す。使用する元データを JPEG 形式に変換し、その後正規化する。正規化したデータと作成した正解データを入力として、使用するそれぞれの CNN に学習させる。そうして出力される予測画像、および様々な評価指標による検出精度を元に CNN の性能を比較し、今回使用するデータセットに最も適した CNN を探索する。

4.2 使用データの概要

本研究で使用するデータは、2017 年 8 月から 2020 年 10 月の期間に、和歌山県立医科大学附属病院紀北分院で脊椎疾患患者 56 人から撮影された 120 枚の単純 X 線画像であり、胸腰椎を撮影した側面像である。使用する X 線画像の例を図 4.2 に示す。これらは、12bit の DICOM 形式の画像で保存されているため、8bit の JPEG 形式に変換し使用する。また、形式変換の際に階調処理を行い、画素値の幅を 0 から 255 に変換する。今回のデータセットは、学習データ 80 枚、検証データ 20 枚、テストデータ 20 枚の合計 120 枚の X 線画像データで構成されている。本研究は和歌山県立医科大学の倫理委員会の承認を得て行っている。

4.3 正解データ

本研究では、MIT が開発した画像アノテーションツール LabelMe[24] を用いて、X 線画像に手でセグメンテーションを行い、正解データを作成する。出力データは JSON 形式の座標データとして保存されるため、PNG 形式のマスク画像に変換して使用する。作成した正解データは、専門医師 1 名の目によって、椎骨の手動セグメンテーションを正確に実施できていることを確認済みである。図 4.3 に、正解データの作成手順の概要を示し、以下にその手順の詳細を示す。

1. X 線画像 (図 4.3 画像 (a)) 内の椎骨の輪郭に沿って正確にポリゴン (頂点) を作成し、椎骨を囲う (図 4.3 画像 (b)).

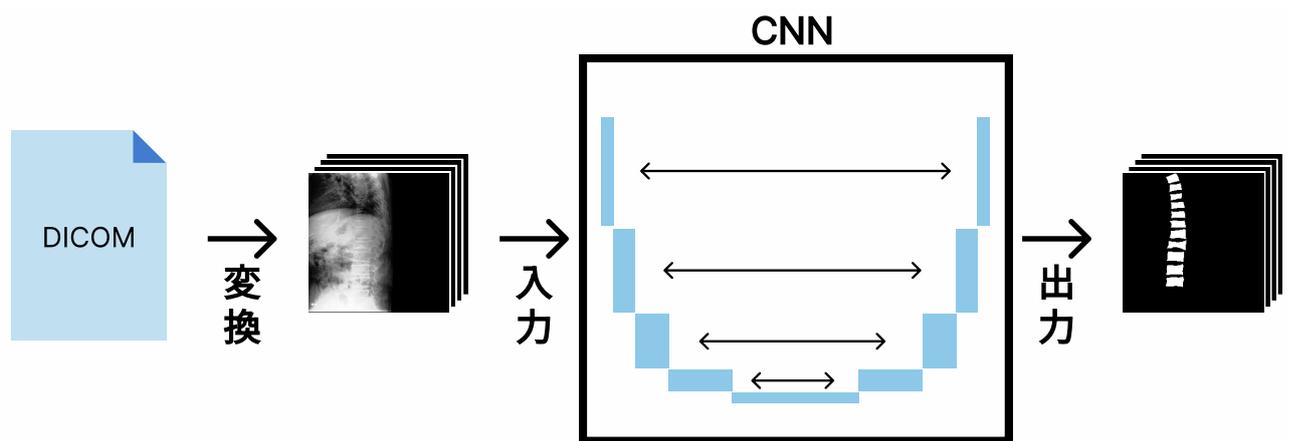


図 4.1: 本手法の概要



図 4.2: 使用する X 線画像の例 (左: 立位, 中央: 坐位, 右: 仰臥位)

2. 作成したポリゴンの座標情報を json ファイルに保存する.
3. 保存された json ファイルを元に PNG 形式のマスク画像 (図 4.3 画像 (c)) を作成する.

4.4 評価指標

本研究で学習したモデルは, Accuracy, Precision, Recall, meanIoU(mean Intersection over Union) を用いて評価を行った. meanIoU とは, IoU のクラス平均の値であり, ここでの IoU とは, 正解データとモデルによる予測結果のオーバーラップ率を表している. meanIoU は物体検出の分野において広く利用されている評価指標であり, 本研究でも特に meanIoU を重視して評価を行う.

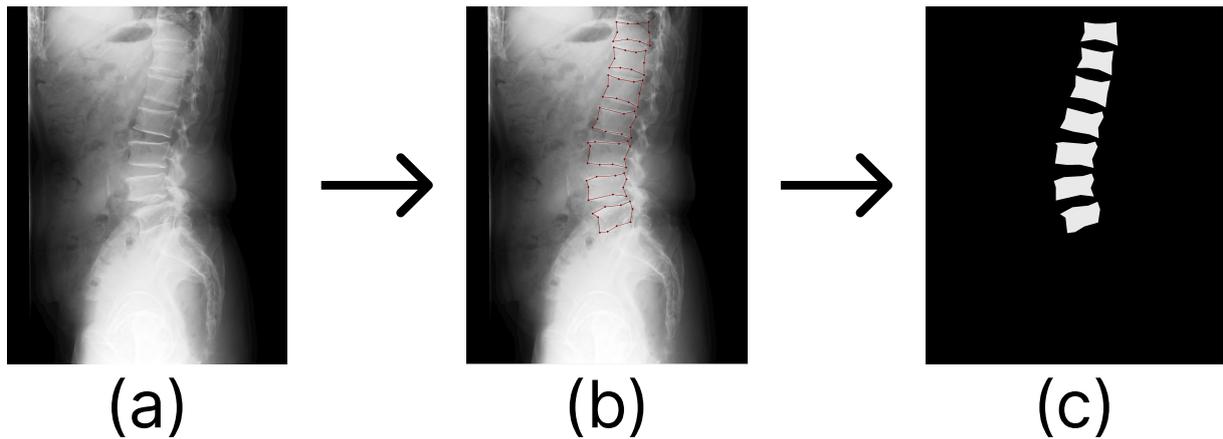


図 4.3: 正解データの作成手順の概要

((a): 元の X 線画像, (b): ポリゴンを作成し椎骨を囲う, (c): 作成した正解データ)

TP 値, TN 値, FP 値, FN 値は各画素に対して評価を行い算出している. ここで, TP は真陽性 (true positive) で, 椎骨として正確にセグメンテーションされた画素数, TN は真陰性 (true negative) で, 背景 (黒) として正確にセグメンテーションされた画素数, FP は偽陽性 (false positive) で, 椎骨として誤ってセグメンテーションされた画素数, FN は偽陰性 (false negative) で, 椎骨が見逃された画素数を表す.

Accuracy, Precision, Recall, meanIoU は以下の式で算出される. meanIoU の算出式における C はクラス数を表す.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{meanIoU} = \frac{1}{C} \sum_{C=1}^C \frac{TP_C}{TP_C + FP_C + FN_C}$$

4.5 椎骨検出精度の評価結果と考察

4.5.1 ハイパーパラメータ

本手法におけるハイパーパラメータを表 4.1 に示す. 今回は学習回数を 3000 に設定して学習を行った. 3000 に設定した理由としては, 損失関数の値が全体的にほぼ収束していたためである.

表 4.1: ハイパーパラメータ

パラメータの種類	パラメータの設定
学習回数	3000epochs
損失関数	Binary Cross Entropy
バッチサイズ	8
学習率	0.0001
オプティマイザ	Adam

表 4.2: 各 CNN の評価結果

モデル	Accuracy	Precision	Recall	meanIoU
U-Net	98.0%	91.9%	94.2%	86.5%
Residual U-Net	97.8%	96.1%	94.1%	78.1%
Dense U-Net	97.8%	95.4%	93.5%	71.5%
SegNet	97.0%	94.1%	89.3%	76.4%

また、損失関数には分類タスクで一般的に用いられる損失関数である Cross Entropy（交差エントロピー）を使用する。Cross Entropy にはマルチクラス分類用と二値分類用の関数が存在するため、今回は Binary Cross Entropy（二値分類用の交差エントロピー）を採用した。今回は学習データが 80 枚と少なく、また慣習的に 2 の n 乗が利用されることから、バッチサイズは 8 とした。

学習率については、今回は 0.0001 に設定した。学習率は別名ステップサイズとも呼ばれ、パラメータの 1 回の更新（ステップ）の大きさを表す値であり、0 から 1 の間で設定される。学習率が大きすぎるとステップが大きくなりすぎて損失の最小値を乗り越えてしまい、逆に学習率が小さすぎると 1 回の更新で少ししか動かなくなり、収束までにかかり時間がかかってしまうため、丁度良い学習率を設定する必要がある。

オプティマイザには Adam を採用した。オプティマイザとは、損失関数の最小値を求めるアルゴリズムであり、機械学習を行う際に伴う損失を限りなく小さくする手法のことである。損失を小さくすることによって、効率よく機械学習が行えるようになるため、性能の高いモデルの開発が可能となる。オプティマイザには様々な種類が存在するが、今回は様々なモデルに広く使用されているデファクトスタンダードな最適化アルゴリズムである Adam を採用した。

4.5.2 評価結果と考察

各 CNN モデルの評価結果を表 4.2 に示す。ここでは、評価指標ごとの結果および考察を述べる。

(1) Accuracy

表 4.2 における Accuracy の値に着目する。U-Net は 98.0%, Residual U-Net は 97.8%, Dense U-Net は 97.8%, SegNet は 97.0% と、全体的に高い値となった。また、各 CNN モデルにおける Accuracy の学習曲線を図 4.4 に示す。全てのモデルにおいて、学習が終了するまで高い値を維持できていることが分かる。高くなった理由としては、正解画像における背景（黒）の占める割合が多いため、予測が当たりやすいのだと考えられる。しかし、SegNet の学習曲線では 200epoch 辺りから学習データとテストデータの精度の間に少し差が生まれていることが分かる。このことから、SegNet では少し過学習が起きている可能性が考えられるため、モデルの処理層を改良する余地があると考えられる。

(2) Precision

表 4.2 における Precision の値に着目する。U-Net は 91.9%, Residual U-Net は 96.1%, Dense U-Net は 95.4%, SegNet は 94.1% となり、全てのモデルにおいて 90% を超える結果となったが、U-Net は他の 3 つのモデルと比べて少し低い値となった。理由としては、X 線画像内では不明瞭で見えていない椎骨も予測しようとしているために、正解画像を越えた部分の椎骨を検出しているからだと考えられる。また、各 CNN モデルにおける Precision の学習曲線を図 4.5 に示す。全てのモデルにおいて、高い値を維持できていることが分かる。このことから、椎骨の誤認識・誤検出をかなり抑えられていると言える。

(3) Recall

表 4.2 における Recall の値に着目する。U-Net は 94.2%, Residual U-Net は 94.1%, Dense U-Net は 93.5%, SegNet は 89.3% となり、全てのモデルにおいて高い値となったが、SegNet は他の 3 つのモデルと比べて少し低い値となった。画像診断においては見逃しや取りこぼしは致命的であるため、SegNet は他のモデルと比べてあまり良くない結果だと言える。また、各 CNN モデルにおける Recall の学習曲線を図 4.6 に示す。Precision の学習曲線と同様に、全てのモデルにおいて高い値を維持できていることが分かる。SegNet については、高い値を維持できているが今後より高い値を目指す余地があると考えられる。

(4) meanIoU

表 4.2 における meanIoU の値に着目する。U-Net は 86.5%, Residual U-Net は 78.1%, Dense U-Net は 71.5%, SegNet は 76.4% となり、U-Net の値が最も高い結果となった。このことから、今回のデータセットにおいては U-Net が最も椎骨の検出に適していると言える。しかし、他の 3 つのモデルにおいても値が 70% を超えていることから、全体的に高い精度で椎骨を検出できていると言える。また、各 CNN モデルにおける meanIoU の学習曲線を図 4.7 に示す。U-Net の学習曲線に関しては値が収束に向かっていように見えるのに対して、他の 3 つのモデルに関しては値がまだ収束に向かっていないように見えるため、パラメータの見直しやモデルの処理層の改良を行うことでより学習を進めることができる可能性がある。特に学習回数については、増やせば meanIoU の値が更に上がりそうにも見えるが、学習回数が多すぎると過学習が起る原因となる。そのため、今後は他のパラメータを調整し、よりうまく学習が収束するように改良することを目指したい。

4.5.3 予測画像

(1) U-Net の予測画像

U-Net における椎骨の予測画像を図 4.8 に示す。ほとんどの椎骨の形状および位置をかなり正確に捉えられており、ノイズはほぼ見られない。また、図 4.8 における D, G, N の予測画像では、正解画像を超えて不明瞭な腰椎部分の椎骨まで検出しようとしているように見える。つまり、評価結果の値以上の精度で検出できている可能性があるため、より正確にアノテーションされた正解データが必要である可能性がある。図 4.8 における F, G, I, K, R の予測画像などでは、上手く形状を検出できず崩れてしまっているような椎骨が見られる。原因としては、元画像に白いもやがかかっている部分があるため椎骨の形状が見えにくくなっていることなどが挙げられる。そのため、元画像の前処理段階において改良の余地がある可能性がある。

(2) Residual U-Net の予測画像

Residual U-Net における椎骨の予測画像を図 4.9 に示す。ほとんどの椎骨の形状および位置を正確に捉えられているが、図 4.9 における A, D, I, L, R の予測画像では、U-Net の予測画像と比較して椎骨の形状がより崩れてしまっているように見える。また、U-Net と同様に、図 4.9 における F, G, K の予測画像でも上手く形状を検出できていない椎骨が見られる。図 4.9 における I, T の予測画像では、少しノイズが見られる。原因としては、U-Net と同様に元画像に白いもやがかかっている部分があるため椎骨の形状が見えにくくなっていることの他に、精度が十分ではないことなどが考えられる。

(3) Dense U-Net の予測画像

Dense U-Net における椎骨の予測画像を図 4.10 に示す。ほとんどの椎骨の形状および位置を正確に捉えられており、ノイズもほとんど見られないが、形状が崩れてしまっている椎骨も多くみられる。特に、図 4.10 における A, G, I, J, K, L, M の予測画像では、かなり形状が崩れてしまっている椎骨が見られる。原因としては、Residual U-Net と同様に元画像に白いもやがかかっている部分があるため椎骨の形状が見えにくくなっていることの他に、精度が十分ではないことなどが考えられる。

(4) SegNet の予測画像

SegNet における椎骨の予測画像を図 4.11 に示す。図 4.11 における B, D, E, N, O, S の予測画像では、ほとんどの椎骨の形状および位置を正確に捉えられている。しかし、それ以外の画像では椎骨の形状をうまく検出できず崩れてしまっている椎骨が多く見られ、特に図 4.11 における G, R の予測画像では、ほとんどの椎骨の形状を検出できていない。また、図 4.11 における N, O 以外の予測画像でノイズが見られ、特に、E, F, I, J の予測画像ではノイズが多くみられる。原因としては、4.5.2 節の (1) でも記述した通り過学習が起きている可能性や、精度が十分ではないことなどが考えられる。

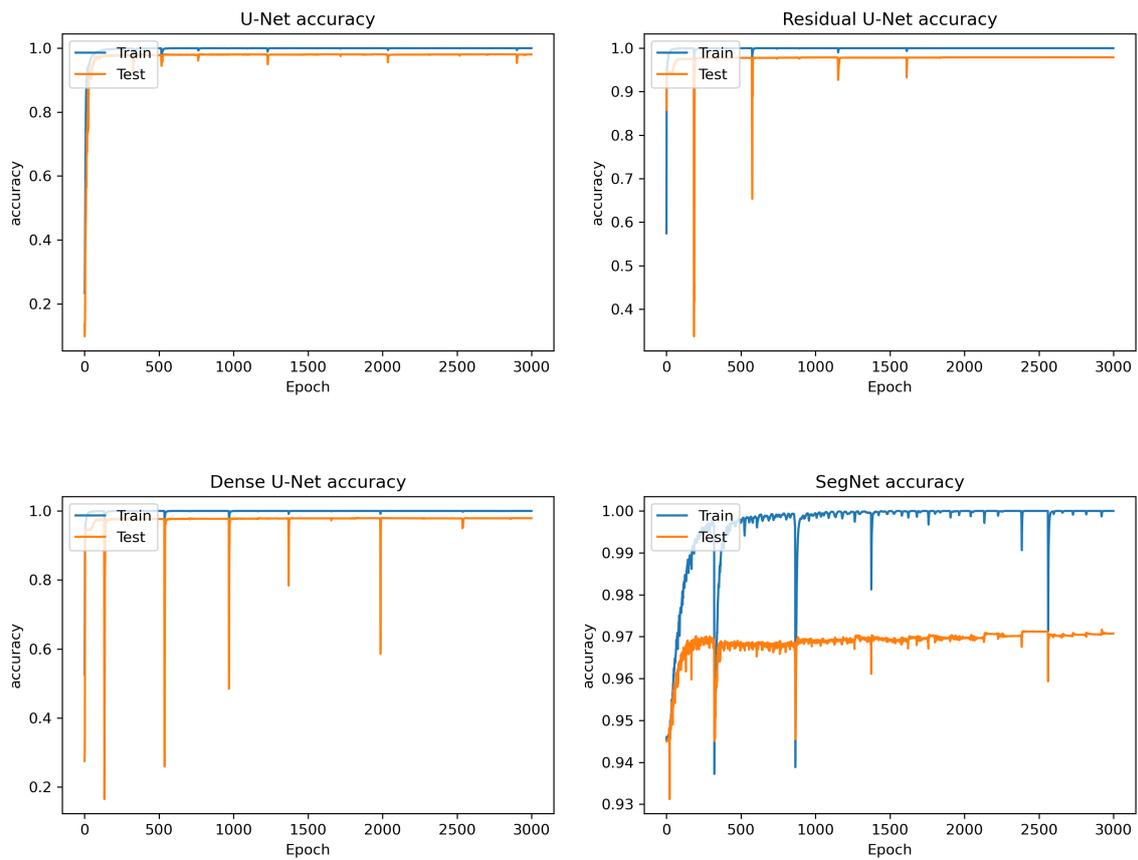


図 4.4: 各 CNN モデルにおける Accuracy の学習曲線
 (左上 : U-Net, 右上 : Residual U-Net, 左下 : Dense U-Net, 右下 : SegNet)

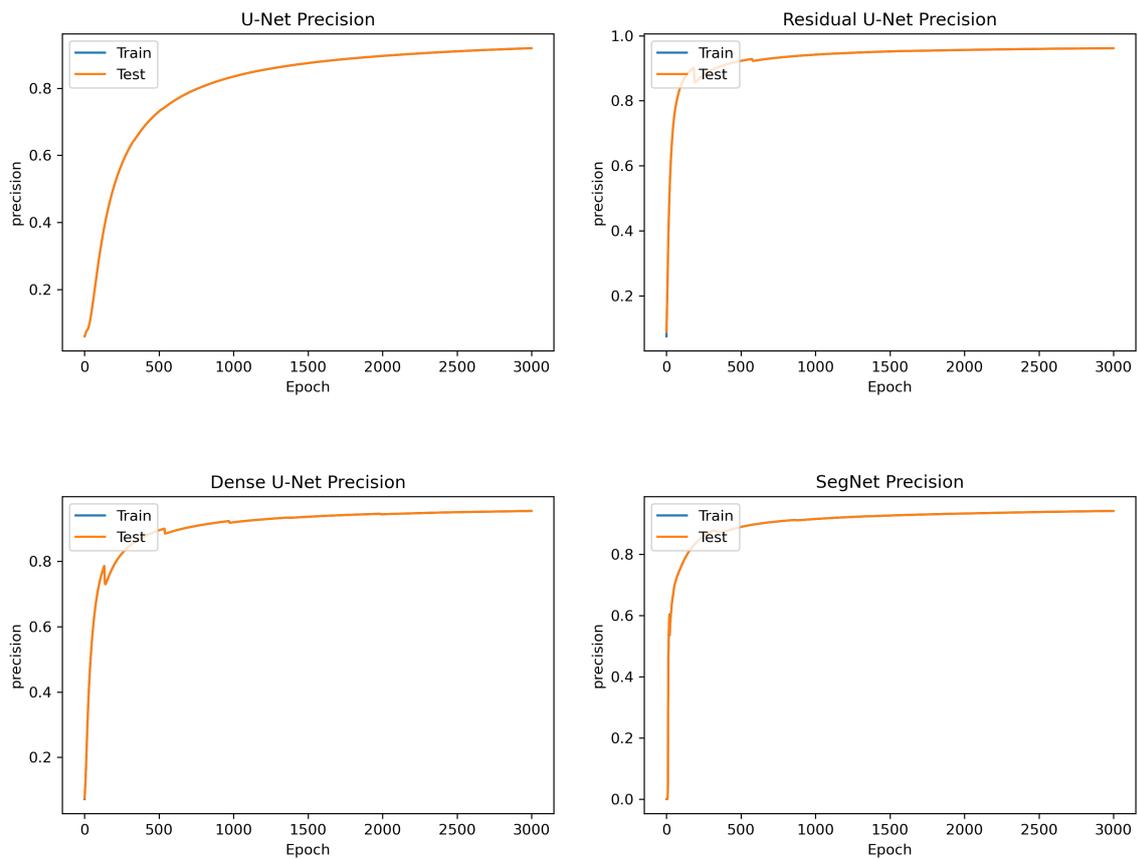


図 4.5: 各 CNN モデルにおける Precision の学習曲線
 (左上: U-Net, 右上: Residual U-Net, 左下: Dense U-Net, 右下: SegNet)

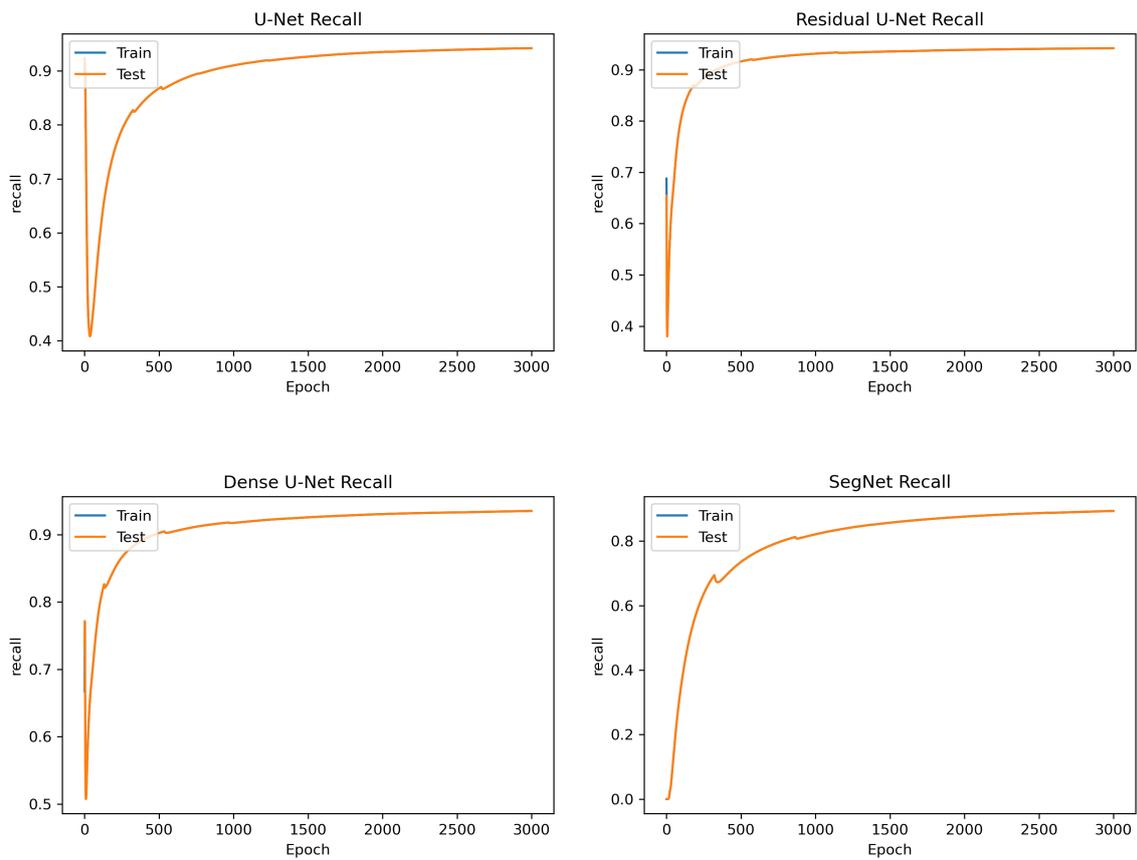


図 4.6: 各 CNN モデルにおける Recall の学習曲線
 (左上: U-Net, 右上: Residual U-Net, 左下: Dense U-Net, 右下: SegNet)

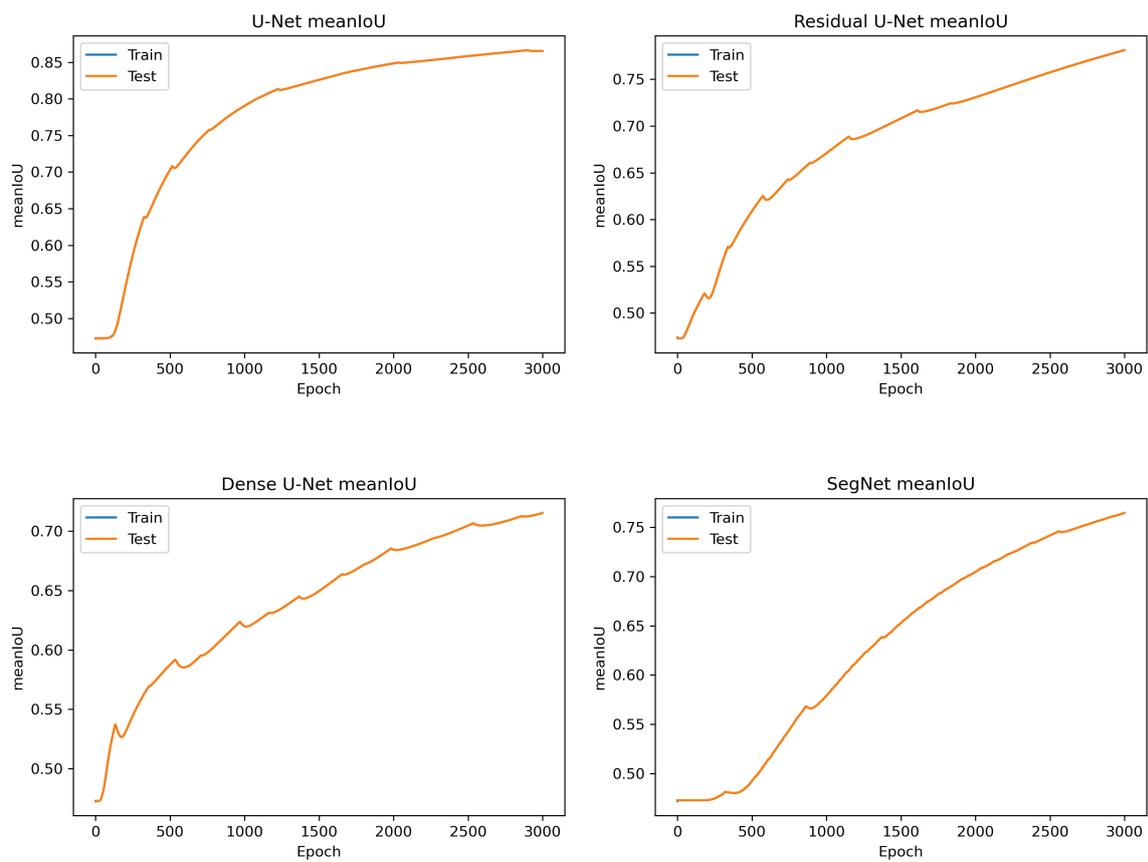


図 4.7: 各 CNN モデルにおける meanIoU の学習曲線
(左上 : U-Net, 右上 : Residual U-Net, 左下 : Dense U-Net, 右下 : SegNet)

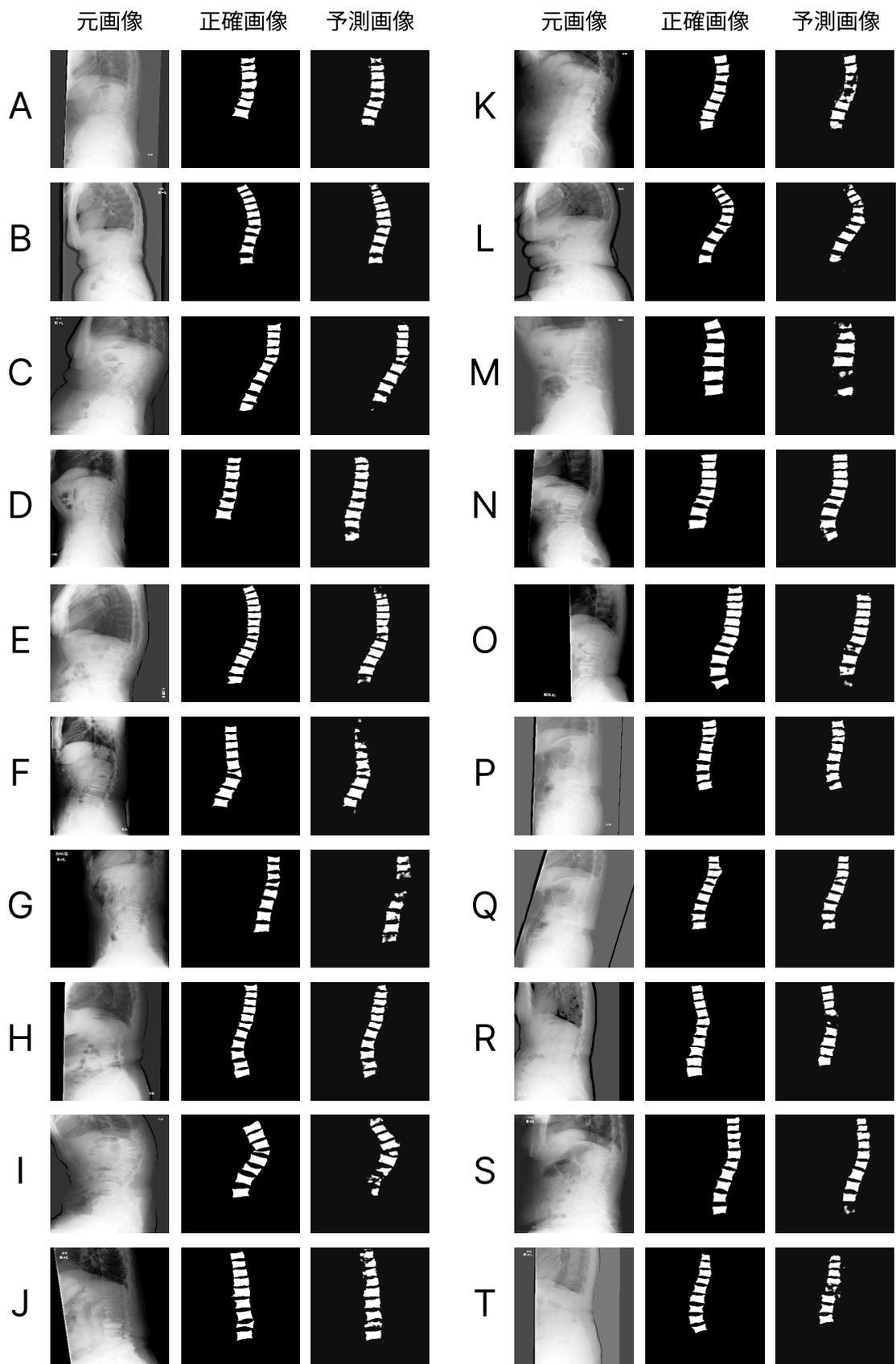


図 4.8: U-Net の予測画像

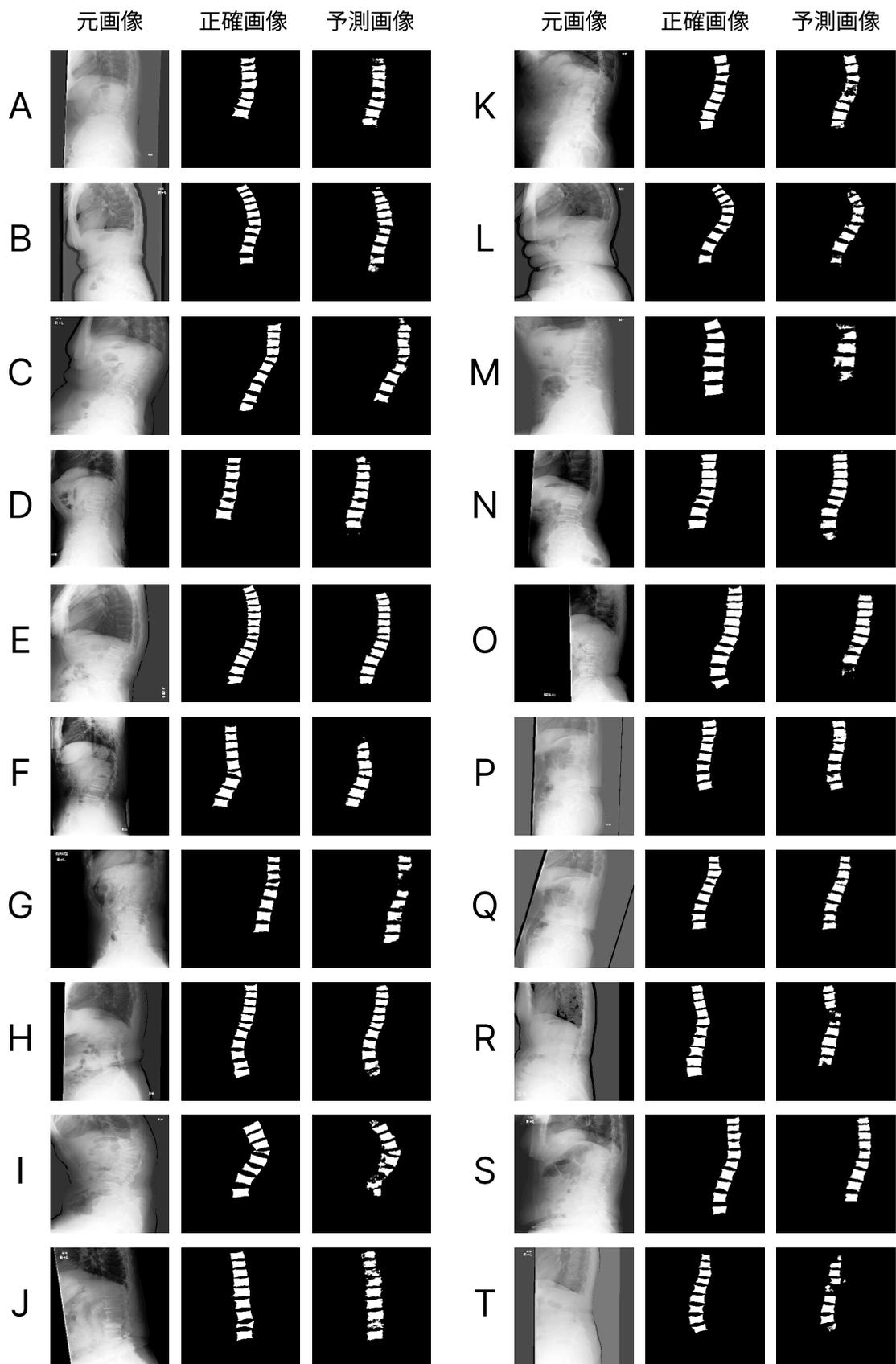


図 4.9: Residual U-Net の予測画像

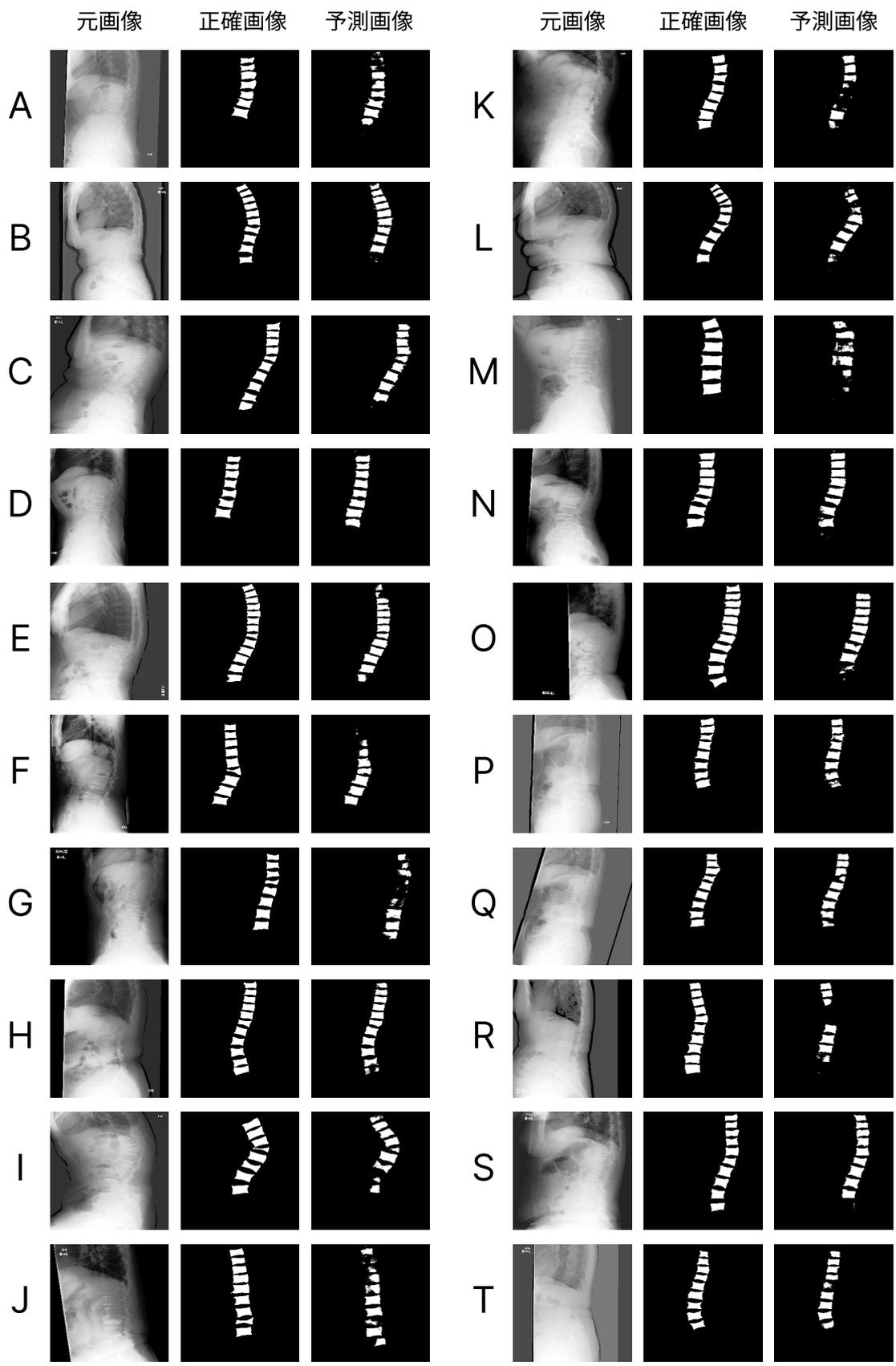


図 4.10: Dense U-Net の予測画像

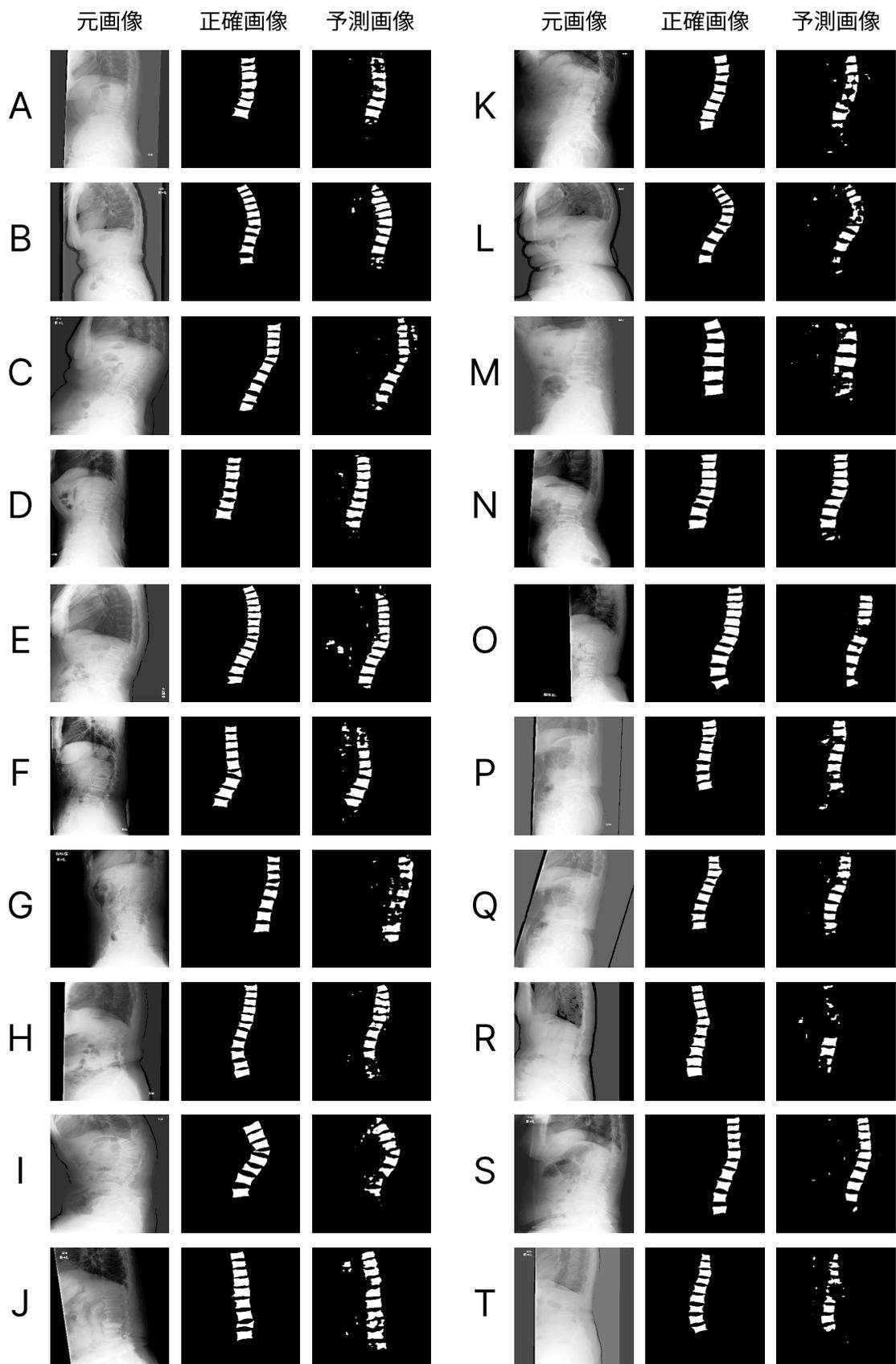


図 4.11: SegNet の予測画像

第5章 結論

椎体骨折は圧潰が進むと難治性の腰痛となるため、早期の発見が必要であり、簡便かつ早期に診断できる単純 X 線撮影を用いた検査が多く用いられている。しかし、X 線画像は不明瞭な場合が多く、正確な読影には多くの経験が必要である。また、撮影する際の体位は状況によって異なるため、体位が様々な X 線画像に適した CNN モデルを開発することで、どのような体位の X 線画像においても、迅速かつ正確な読影の支援が可能となる。

本研究では、撮影する際の患者の体位が様々である X 線画像における、CNN を用いたセマンティックセグメンテーションによる椎骨の検出手法を提案した。

本論文では、使用したモデルの詳細および学習手法の詳細について述べた。また、体位が様々な X 線画像における椎骨検出のための CNN モデルの精度評価実験を行った。精度評価実験の結果より、以下の知見を得た。

1. 体位が様々な X 線画像においても、既存の CNN モデルでセマンティックセグメンテーションを行うことができる。
2. U-Net の meanIoU は 86.5% となり、今回のデータセットでは、U-Net が最も椎骨のセマンティックセグメンテーションに適している。
3. 一部の予測画像では、正解画像を超えた、不明瞭な部分の椎骨まで検出している可能性がある。

今後は、より適したハイパーパラメータの選定や、前処理の見直し、モデルにおけるネットワーク構造の改良・新たな処理層の追加などを行うことで、より高い検出精度を目指す。また、他の CNN モデルを実装し学習精度を比較することで、より性能が高いモデルを探索する。

参考文献

- [1] Kang Cheol Kim, Hyun Cheol Cho, Tae Jun Jang, Jong Mun Choi, Jin Keun Seo: Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation, *Computer Methods and Programs in Biomedicine*, Vol. 200, 105833 (2021).
- [2] 日本脊椎脊髓病学会 脊椎脊髓指導医とは, 入手先< <https://ssl.jssr.gr.jp/student/about/advisor.html> > (参照日 2022 年 2 月 10 日).
- [3] 一般社団法人日本脊椎脊髓病学会 指導医リスト, 入手先< <https://area18.smp.ne.jp/area/table/9662/6cPeec/M?S=1jrhm2pbmjp> > (参照日 2022 年 2 月 10 日).
- [4] 日本放射線科専門医会・医会 放射線科の紹介, 入手先< https://jcr.or.jp/radiologist/radiology_doctor/ > (参照日 2022 年 2 月 10 日).
- [5] Benjelloun Mohammed, Saïd Mahmoudi, Fabian Lecron: A framework of vertebra segmentation using the active shape model-based approach, *International journal of biomedical imaging*, pp. 1–14 (2011).
- [6] Ruhan Sa, William Owens, Raymond Wiegand, Vipin Chaudhary: Fast scale-invariant lateral lumbar vertebrae detection and segmentation in X-ray images, 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 1054–1057 (2016).
- [7] Nikolas Lessmann, Bram van Ginneken, Pim A. de Jong, Ivana Išgum: Iterative fully convolutional neural networks for automatic vertebra segmentation and identification, *Medical image analysis*, Vol. 53, pp. 142–155 (2019).
- [8] Nils Christian Lehnen, Robert Haase, Jennifer Faber, Theodor Rüber, Hartmut Vatter, Alexander Radbruch, Frederic Carsten Schmeel: Detection of Degenerative Changes on MR Images of the Lumbar Spine with a Convolutional Neural Network: A Feasibility Study, *Diagnostics*, Vol. 11.5, No. 902 (2021).
- [9] Pengfei Cheng, Yusheng Yang, Huiqiang Yu, Yongyi He: Automatic vertebrae localization and segmentation in CT with a two-stage Dense-U-Net, *Scientific Reports*, Vol. 11.1, pp. 1–13 (2021).
- [10] Vania Malinda, Dawit Mureja, Deukhee Lee: Automatic spine segmentation from CT images using convolutional neural network via redundant generation of class labels, *Journal of Computational Design and Engineering*, Vol. 6.2, pp. 224–232 (2019).
- [11] Jakubicek Roman, Jiri Chmelik, Jiri Jan, Petr Ourednicek, Lukas Lambert, Giampaolo Gavelli: Learning-based vertebra localization and labeling in 3D CT data of possibly incomplete and pathological spines, *Computer methods and programs in biomedicine*, Vol. 183, No. 105081 (2020).

- [12] Krzysztof Siemionow, Cristian Luciano, Craig Forsthoefel, Suavi Aydogmus: Autonomous image segmentation and identification of anatomical landmarks from lumbar spine intraoperative computed tomography scans using machine learning: A validation study, *Journal of Craniovertebral Junction & Spine*, Vol. 11.2, pp. 99–103 (2020).
- [13] James Thomas Patrick Decourcy Hallinan, Lei Zhu, Kaiyuan Yang, Andrew Makmur, Diyaa Abdul Rauf Algazwi, Yee Liang Thian, Samuel Lau, Yun Song Choo, Sterling Ellis Eide, Qai Ven Yap, Yiong Huak Chan, Jiong Hao Tan, Naresh Kumar, Beng Chin Ooi, Hiroshi Yoshioka, Swee Tian Quek: Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI, *Radiology*, Vol. 300.1, pp. 130–138 (2021).
- [14] Ming-Huwi Horng, Chan-Pang Kuok, Min-Jun Fu, Chii-Jen Lin, Yung-Nien Sun: Cobb angle measurement of spine from X-ray images using convolutional neural network, *Computational and mathematical methods in medicine*, Vol. 2019, No. 6357171 (2019).
- [15] Le Zhang, Jiabao Zhang, Song Gao: Region-Based Convolutional Neural Network-Based Spine Model Positioning of X-Ray Images, *BioMed Research International*, No. 7512445 (2022).
- [16] Sangwoo Lee, Eun Kyung Choe, Hae Yeon Kang, Ji Won Yoon, Hua Sun Kim: The exploration of feature extraction and machine learning for predicting bone density from simple spine X-ray images in a Korean population, *Skeletal radiology*, Vol. 49, pp. 613–618 (2020).
- [17] S.M.Masudur Rahman Al Arif, Karen Knapp, Greg Slabaugh: Fully automatic cervical vertebrae segmentation framework for X-ray images, *Computer methods and programs in biomedicine*, Vol. 157, pp. 95–111 (2018).
- [18] Chan-Pang Kuok, Min-Jun Fu, Chii-Jen Lin, Ming-Huwi Horng, Yung-Nien Sun: Vertebrae segmentation from X-ray images using convolutional neural network, In: *Proceedings of the 2018 International Conference on Information Hiding and Image Processing*, pp. 57–61 (2018).
- [19] Andrea Cina¹, Tito Bassani, Matteo Panico, Andrea Luca, Youssef Masharawi, Marco Brayda-Bruno, Fabio Galbusera: 2-step deep learning model for landmarks localization in spine radiographs, *Scientific Reports*, Vol. 11.1, pp. 1–12 (2021).
- [20] Ronneberger Olaf, Philipp Fischer, Thomas Brox: U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Vol.2015, pp.234–241 (2015).
- [21] Zhang Zhengxin, Qingjie Liu, Yunhong Wang: Road extraction by deep residual u-net, *IEEE Geoscience and Remote Sensing Letters*, Vol. 15.5, pp. 749–753 (2018).

- [22] Sijing Cai, Yunxian Tian, Harvey Lui, Haishan Zeng, Yi Wu, Guannan Chen: Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network, *Quantitative imaging in medicine and surgery*, Vol. 10.6, No. 1275 (2020).
- [23] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla: Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 39.12, pp. 2481–2495 (2017).
- [24] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, William T. Freeman: LabelMe: a database and web-based tool for image annotation, *International journal of computer vision*, Vol. 77.1, pp. 157–173 (2008).