

2021 年度 卒業論文

論文題目

単一 U-Net モデルでの CT 画像からのリンパ節検出と
転移有無分類

指導教員

舟橋 健司 准教授

名古屋工業大学 工学部 情報工学科

2018 年度入学 30114072 番

鈴木 康介

目次

第 1 章	はじめに	1
第 2 章	背景知識	4
2.1	造影 CT 画像	4
2.2	リンパ節のがん転移有無による特徴の変化	4
2.3	DICOM 画像	6
2.4	ニューラルネットワーク	6
2.5	畳み込みニューラルネットワーク	6
2.5.1	畳み込み層 (Convolution Layer)	7
2.5.2	ReLU(Rectified Linear Units)	8
2.5.3	プーリング層 (Poling Layer)	8
2.5.4	全結合層 (Fully Connected Layer)	9
2.5.5	逆畳み込み (Deconvolution)	9
2.5.6	Attention 機構	11
第 3 章	データセットとモデルの構築	12
3.1	データセットの構築	12
3.1.1	データの概要	12
3.1.2	マスク画像の作成	13
3.1.3	前処理	14
3.1.4	データ拡張	16
3.2	モデルの構築・学習	16
3.2.1	U-Net	17
3.2.2	U-Net++	18
3.2.3	Attention U-Net	18
第 4 章	実験	20
4.1	実験 1	20
4.1.1	手順	20
4.1.2	考察	22
4.2	実験 2	24
4.2.1	手順	25
4.2.2	考察	25
第 5 章	むすび	27

謝辭	28
参考文献	29

第1章 はじめに

がん診断の現場では、X線CTをはじめとする画像診断モダリティが用いられている。X線CTとは、人体の周囲からX線を照射し、X線の透過率の違いを利用して人体の断面を画像化する技術である。CT装置の発展によりCT画像のスライス間隔が小さくなり、一度の検査で大量のデータが得られるようになった。それにより、微小な病変や早期の病変の発見が可能となった。一方で、診断する医師の負担の増加や、従来の診断方法では大量のデータを活用しきれないという問題が発生している。また、画像診断は専門性が高く、医師不足が懸念されている。例えばリンパ節矩形を測定するという診断では、医師の経験量の違いによって、正診率が51-86%と大きくばらつきがある。そのため、コンピュータによる画像診断の支援を行うことで、医師の負担を軽減させることが望まれている。

本研究では、診断の対象として直腸がんのリンパ節転移を取り扱う。直腸がんとは、直腸の組織内に悪性がん細胞が認められる病気である。1975年以降、日本では直腸がんの罹患数・死亡数ともに増加を続けている [1]。そのため、特にコンピュータによる診断支援手法の確立が望まれている病気の1つである。がんはリンパの流れに沿ってリンパ節に転移を生じる可能性があるため、直腸がんが診断された場合はX線CT等での画像診断により、体のほかの部位に転移していないかを調べる病期診断という診断を行う。転移の有無が特に重要となるのが、側方リンパ節と呼ばれる、内腸骨動脈に沿うリンパ節および閉鎖神経と閉鎖動脈周囲のリンパ節である。直腸がんの側方リンパ節転移が陽性の場合の5年生存割合は40%前後であり、転移が陰性の場合の80%と比べて予後不良である。 [2, 3]。直腸がんは治癒切除後でも再発する危険性が高いが、がん周辺のリンパ節だけでなく、側方リンパ節も切除(側方リンパ節郭清)することで骨盤内再発のリスクが50%減少し、5年生存割合は8-9%改善すると予測されている。しかし、すべての症例に対して側方リンパ節郭清を行え

ばよいわけではない。側方リンパ節郭清には排尿機能障害や性機能障害などの後遺症が残る可能性があるためである。そこで、側方リンパ節への転移が陰性である場合には、再発のリスクよりも側方リンパ節郭清による後遺症のリスクを考慮し、側方リンパ節郭清を省略してもよいとされている [4]。したがって、手術前に側方リンパ節への転移有無を正確に診断し、側方リンパ節郭清が必要であるかを検討することが重要となる。しかし、側方リンパ節には確立した診断手法がなく、新たな診断手法の研究が望まれている分野である。

ところで、がんのリンパ節への転移有無の診断の研究は2つの段階に分けることができる。モダリティ機器から得られたデータからリンパ節領域を検出するタスク（以下、検出タスクと呼ぶ）と、すでに検出されたリンパ節のがん転移の有無を分類するタスク（以下、分類タスクと呼ぶ）である。検出タスクの研究は特に盛んに行われており、Odaらは胸部CT画像からの縦隔リンパ節検出に対して、3D U-Netでリンパ節以外の解剖学的構造（肺、気道、大動脈弓、および肺動脈）も合わせて学習させることでクラスの偏りを緩和し、過検出を削減できることを示した [5]。また、JayantらはU-Netを用いた候補生成ステージでVOIを生成した後、3D CNNを用いて偽陽性削減を行うことで、高い感度で検出できることを示した [6]。分類タスクの研究も行われており、Liらはリンパ節のMRI画像から大腸がんの転移分類に有効な手法を調べるために、機械学習、深層学習、深層転移学習を比較し、深層転移学習が最も有効であることを示した [7]。また、岡田はリンパ節のCT画像からがん転移分類を行う際に、3D CNNを用いて3次元的に学習を行うことで悪性の分類精度が向上することを示した。さらに、入力データにリンパ節領域だけでなく周辺の情報を含めることで精度が向上することも示した [8]。

前述の先行研究の課題として、検出タスクと分類タスクが別々であることが挙げられる。実際のがん診断においてはリンパ節の検出とがん転移分類はセットで行いたいという需要があるが、筆者の知る限りセットで行う手法は提案されていない。したがって、別々の手法を組み合わせて用いるしかなく、労力がかかるという現状がある。そこで本研究では検出タスクと分類タスクという2つのタスクを単一のニューラルネットワークモデルで行う手法を提案する。2つのタスクを単一のモデルで行うことにより、実際の診断に使用する際の労力の削減が可能である。さらに2つの

タスクを同時に行うように学習したモデルは，タスクごとに別々に学習したモデルとは異なり、統合的に情報を獲得することが可能である．本論文ではこの2つのタスクを同時に実現するモデルを用いて分析を行うこととした．

第2章 背景知識

本章ではCT画像とリンパ節に関する背景知識として、造影CT画像、リンパ節のがん転移有無による特徴の変化、DICOM画像について説明する。また、ニューラルネットワークの概要について説明する。

2.1 造影CT画像

CT画像を撮影する際に、医師が診断をしやすくするために造影剤と呼ばれる薬剤を使用する場合がある。造影剤は、撮影前に静脈に注射することでCT画像の血管や臓器の部分にコントラストが生じるため、そのままではわかりにくい小さな病変も見つけやすくする効果がある。造影剤を使わずに撮影する方法を単純CTと呼び、造影剤を使用して撮影する方法を造影CTと呼ぶ。造影剤を使うことによるCT画像の違いを図2.1に示す。造影剤を使用していなければリンパ節を見つけ出すことは困難とされており、本研究で使用するデータセットも造影CTで撮影されたものである。

2.2 リンパ節のがん転移有無による特徴の変化

リンパ節にがん細胞が転移しているときとしていないときで変化する特徴がある。その特徴は全部で3つあり、模様・形状・大きさである。通常のリンパ節は一様な模様で、アーモンドのような楕円体の形状をしており、基本的に10mm以下の大きさである。それに対し、転移したリンパ節はまばらな模様となり、腫れ上がることで球形になり、10mm以上の大きさになるとされている[9, 10]。模様の違いは造影剤によるものであり、転移したがん細胞が染まる際の障害となってまばらな模様となる。図2.2と図2.3にそれぞれの転移状態におけるリンパ節領域のCT画像を示す。

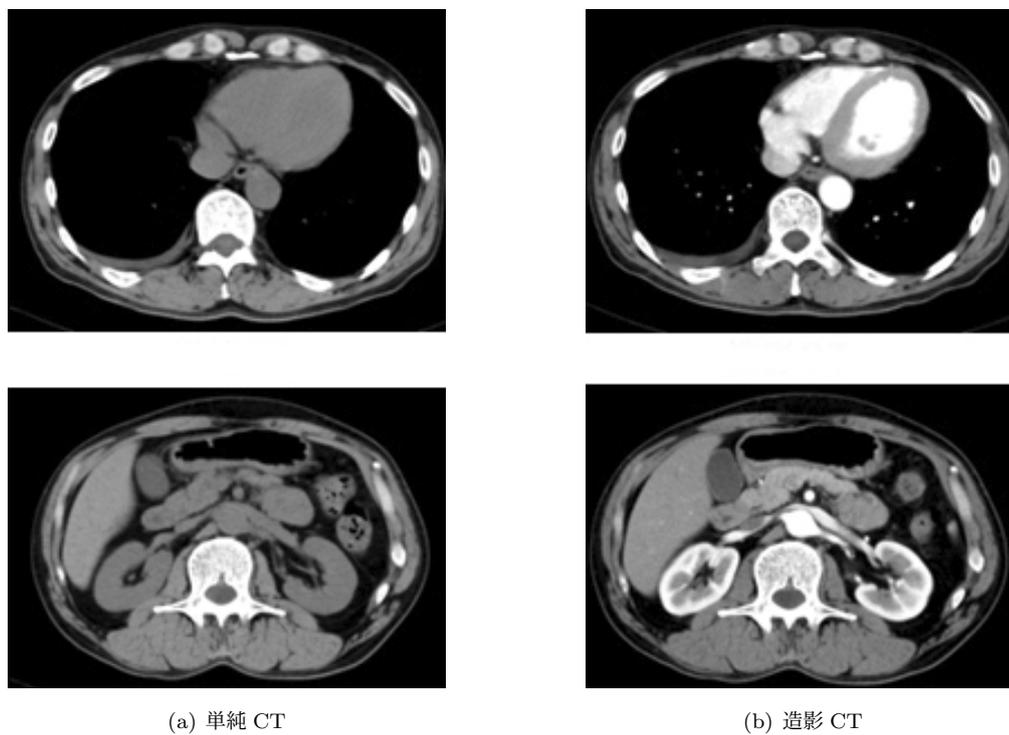


図 2.1: 造影剤による CT 画像の違い

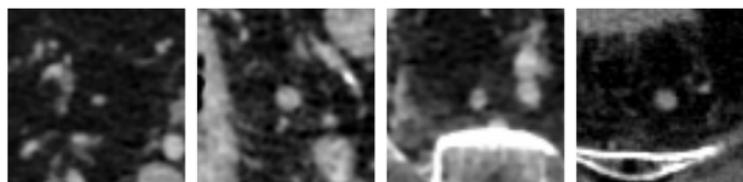


図 2.2: 転移なし (良性) のリンパ節の例 (画像中央)

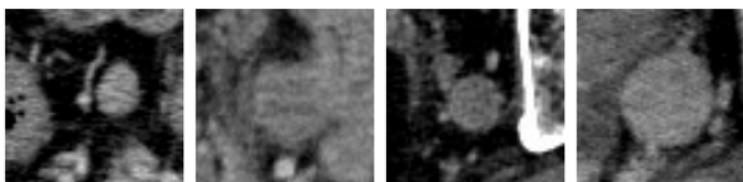


図 2.3: 転移あり (悪性) のリンパ節の例 (画像中央)

2.3 DICOM 画像

DICOM とは医用画像で一般的に使用されている形式である。デジタル画像の形式の多くが 8 ビットの階調で表現するのに対し、DICOM は 10 ビット以上の階調を表現することができる。医師が画像診断を行う際には、元の DICOM 画像の階調から、ウィンドウレベル変換という処理により 8 ビットの階調に切り出してから表示を行う。ウィンドウレベル変換で切り出す階調の中心値と幅をそれぞれウィンドウレベルとウィンドウ幅という。ウィンドウレベルとウィンドウ幅を適切に変更することで見たい組織のみを表示することができ、診断がしやすくなる。

2.4 ニューラルネットワーク

ニューラルネットワークとは人間の脳の神経細胞であるニューロンのはたらきを模して作られた数理モデルである。ニューラルネットワークは入力層、隠れ層、出力層から構成される。さらに各層は複数のノードから構成される。層と層の間にはニューロン同士の繋がりや強さを示す重みがある。特に、隠れ層が 2 層以上存在するものをディープニューラルネットワーク (DNN) と呼ぶ。ニューラルネットワークにおける学習とは、出力層の出力が正しい答えに近づくようにパラメータ (重みとバイアス) を調整する作業を指す。この学習を繰り返すことで精度を向上させ、高精度な予想を可能にする。ニューラルネットワークの基本的な概要図を図 2.4 に示す。

2.5 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク (CNN) は画像認識や音声認識などで多く使われているニューラルネットワークである。一般的な順伝搬型のニューラルネットワークとは異なり、畳み込み層 (Convolution Layer) やプーリング層 (Pooling Layer) から構成されているという特徴がある。CNN の入力には RGB の要素を持ったカラー画像やグレースケール画像である。その画像に対して $n_x \times n_y$ のフィルタを畳み込むことで特徴マップを得る。その際に得られた値に対し活性化関数を適用することで次の層の入力となる。そしてプーリング層によって特徴マップのダウンサンプリングを行う。このような処理がいくつ重なるとネットワークが深くなっていき、

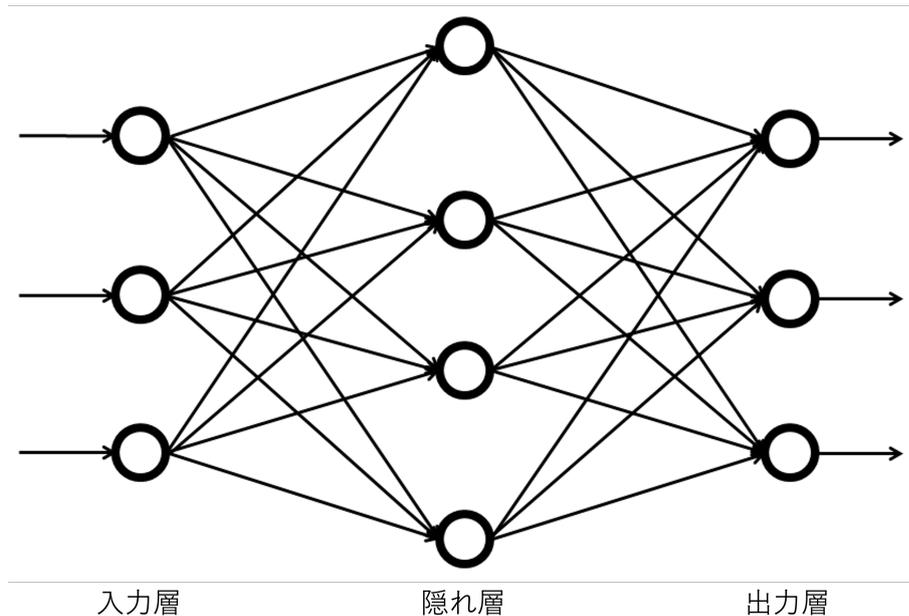


図 2.4: ニューラルネットワークの概要図

画像から特徴を抽出することが出来る。そしていくつか処理を重ねた後にそれまでの特徴マップを1次元にし全結合層への入力とする。最終的な出力層のノード数により最終的な結果が変わり、2クラス分類や多クラス分類など様々な問題に対応することが出来る。CNNをベースとした代表的なアーキテクチャには AlexNet[11] や LeNet[12] などが存在する。CNNの基本的な概要図を図2.5に示す。

2.5.1 畳み込み層 (Convolution Layer)

畳み込み層は、入力をフィルタを用いて畳み込み演算した結果である特徴マップを出力する。一つの層には複数のフィルタが存在し、フィルタの数だけ特徴マップを出力するため、出力のチャンネル数はフィルタ数と同じとなる。フィルタは画像全体に対して適用されるため画像の特徴がどこにあっても抽出することが出来る。これを位置不変性という。このような処理を行うことにより、画像から全体的な漠然とした特徴を抽出することが出来る。この出力に対し次に示す活性化関数を適用することにより最終的な出力とし、次の層への入力となる。

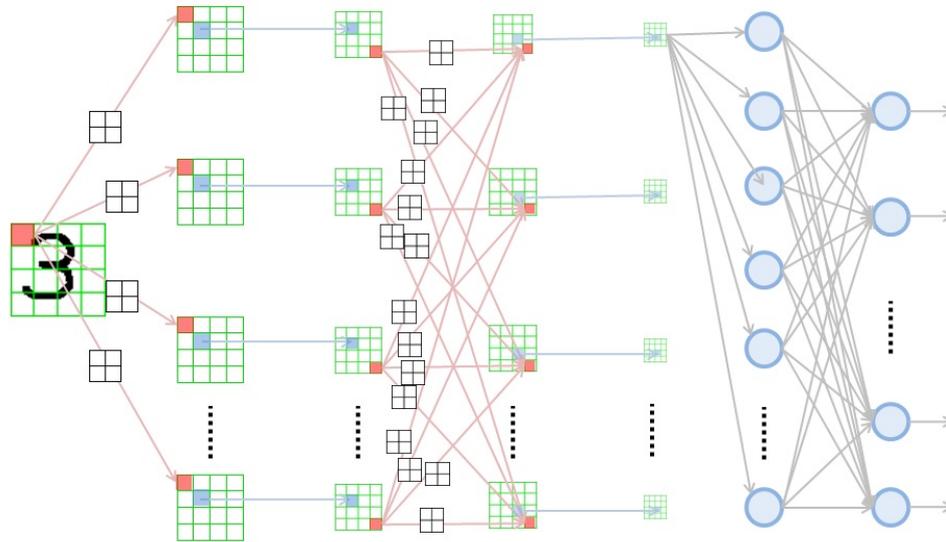


図 2.5: CNN の概要図

2.5.2 ReLU(Rectified Linear Units)

ある層の出力を次の層への入力とするために、活性化関数を適用する。畳み込み層においては活性化関数として一般的に Rectified Linear Units(ReLU) 関数が用いられている。ReLU 関数の式を式 (2.1) に示す。ReLU 関数は 0 以上の入力に対してはそのままの入力を返し、0 より小さい場合は 0 を返す。これにより勾配消失問題を回避することが可能なため、よく用いられている。ReLU 関数を図 2.6 に示す。

$$a_{ij} = \text{ReLU}(x_{ij}) = \max(0, x_{ij}) \quad (2.1)$$

2.5.3 プーリング層 (Pooling Layer)

プーリング層は畳み込み層の後に適用されることが多く、入力されたデータに対しダウンサンプリングを行う。プーリング層ダウンサンプリングにより画像サイズが小さくなり情報が圧縮される。それにより微小な変化に頑健になったり計算コストを減少させたりすることが可能になる。プーリング層の処理にはいくつか種類が存在し、ここでは Max Pooling と Average Pooling について記述する。Max Pooling

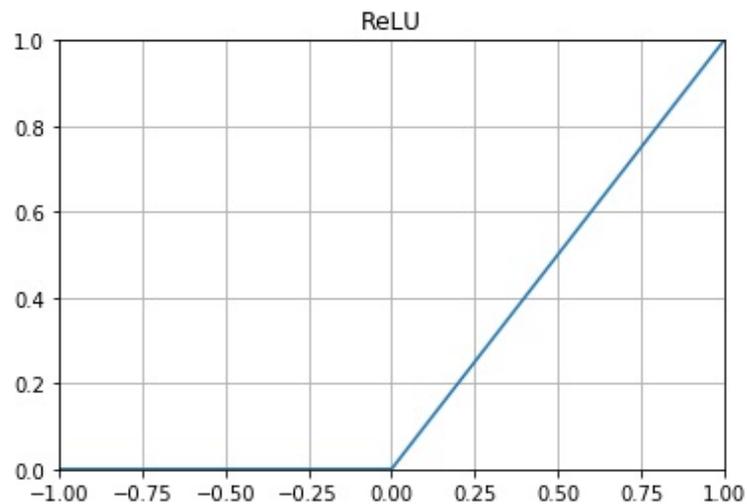


図 2.6: ReLU 関数

は指定された領域の中の最大値を結果とし、Average Pooling は指定領域中の平均値を結果とする手法である。それぞれの流れを図 2.7 に示す。

2.5.4 全結合層 (Fully Connected Layer)

これまでで得られた特徴を抽出した画像データを 1 次元のベクトルに変換し一つのノードに結合する。さらに活性化関数によって変換された特徴変数を出力とする。ノードの数が増えると各領域を特徴付ける特徴変数の数が増えることになる。

2.5.5 逆畳み込み (Deconvolution)

逆畳み込みは転置畳み込み (Transposed Convolution) とも言われ、元の特徴マップを拡大した後に畳み込みを行う処理である。図 2.8 のように 3×3 の特徴マップの間や周囲へゼロパディング処理をしたあとフィルタを適用することで 5×5 の特徴マップが得られていることが分かる。この処理を行うことで局所的な特徴をグローバルなものへと変化させることが可能になる。

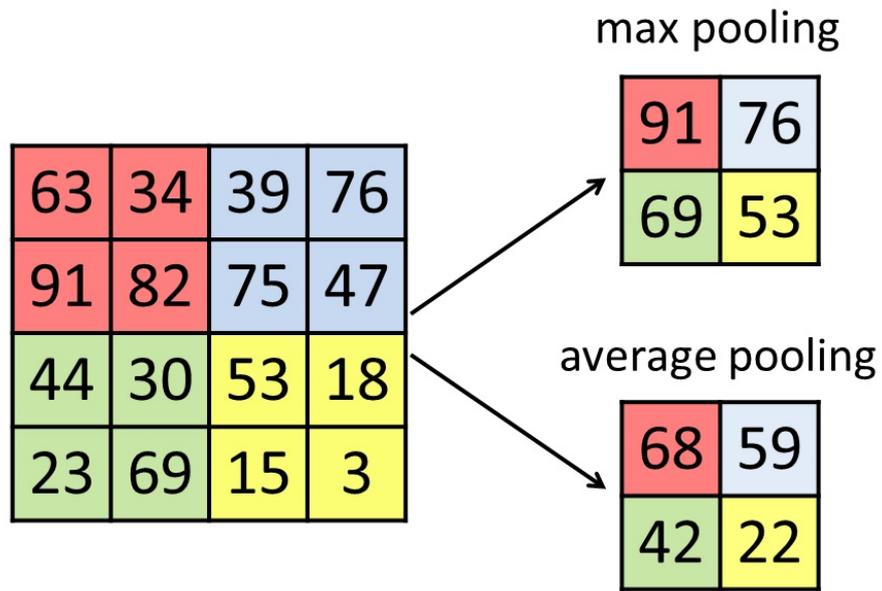


図 2.7: Max Pooling, Average Pooling

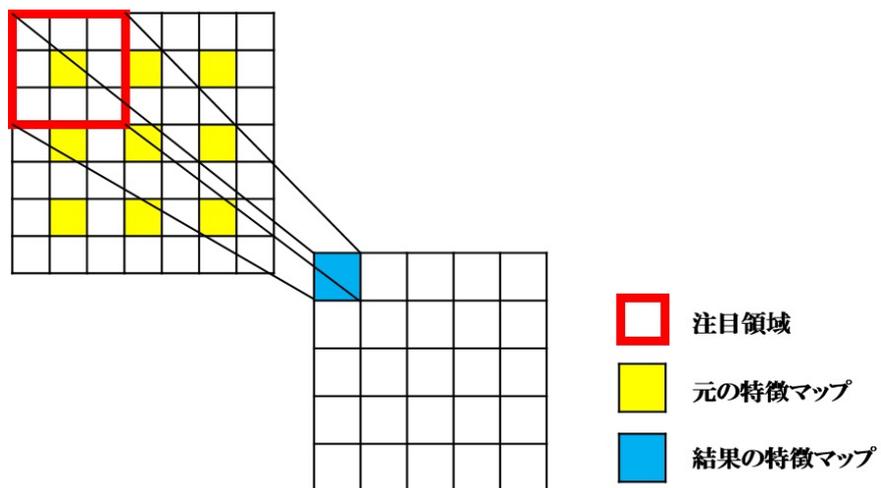


図 2.8: 逆畳み込み

2.5.6 Attention 機構

Attention 機構は入力のどこに注意すべきかを学習する機構である。例えば、畳み込み層から出力された特徴マップに対して 1×1 畳み込みなどを行い、幅と高さは元の入力と同じでチャンネル数が1である Attention マスクを出力する層を作る。Attention マスクは入力の注意すべき部分は1に、注意しなくてよい部分は0に近い値となるマスクである。この Attention マスクを元の特徴マップと掛け合わせることで、特徴マップの注意すべき部分を際立たせた重み付き特徴マップが得られる (図 2.9)。この一連の流れを行う部分を Attention 機構という。

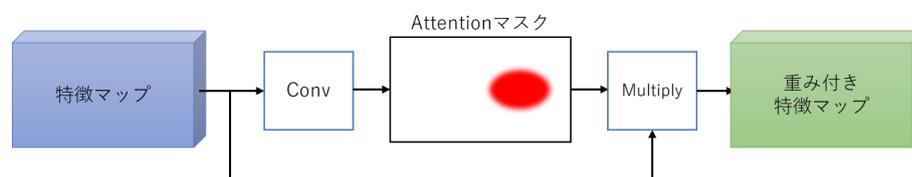


図 2.9: Attention 機構

第3章 データセットとモデルの構築

本研究の目的は、腹部の造影 CT 画像を対象として単一のモデルで検出タスクと分類タスクを同時に実行することである。単一のモデルで検出タスクと分類タスクを行うことは、マルチクラスセグメンテーションを行うことと等しい。本章ではリンパ節のマルチクラスセグメンテーションのためのデータセットとモデルの構築方法について記述する。

3.1 データセットの構築

本研究では学習および評価のためのデータとして、愛知県がんセンターから提供された患者 50 人分の CT 画像データを使用する。このデータからデータセットを構築する手順と、学習の効率化のための前処理およびデータ拡張の手順について説明する。

3.1.1 データの概要

提供されたデータは、患者ごとに約 500～800 スライスの DICOM 形式画像と、その何スライス目のどの位置にがん転移なしリンパ節（以下、G1 と呼ぶ）とがん転移ありリンパ節（以下、G3 と呼ぶ）が存在するかを医師がアノテーションした PowerPoint プレゼンテーションファイル（以下、pptx ファイルと呼ぶ）のセットとなっている。1つのリンパ節が複数スライスに写っている場合は、そのうち最も鮮明に写ったスライスのみがアノテーションされている。この pptx ファイルの例を図 3.1 に示す。合計でリンパ節 390 個分のアノテーションがされており、内訳は G1 が 349 個、G3 が 41 個である。



図 3.1: 医師によるアノテーション pptx ファイル

3.1.2 マスク画像の作成

3.1.1 節で説明した pptx ファイルのデータのままで学習に使用できないため、pptx ファイルを元にマスク画像の作成を行う。マスク画像作成の手順は以下の通りである。

Step 1 CT 画像を DICOM 形式から、TIFF 形式に変換する。

Step 2 Step 1 で変換した画像を、画像編集ソフトを用いて pptx ファイルでアノテーションされているリンパ節を G1 は $(R,G,B)=(0,255,0)$ 、G3 は $(R,G,B)=(255,0,0)$ で塗りつぶす。それ以外の部分は $(R,G,B)=(0,0,0)$ で塗りつぶす。

1つのマスク画像に対して、対応する CT スライスのみをセットにしたデータセット

と、対応する CT スライスとその前後合わせて 3 スライスをセットにしたデータセットの 2 種類のデータセットを作成する。データセットの例を図 3.2, 図 3.3 に示す。

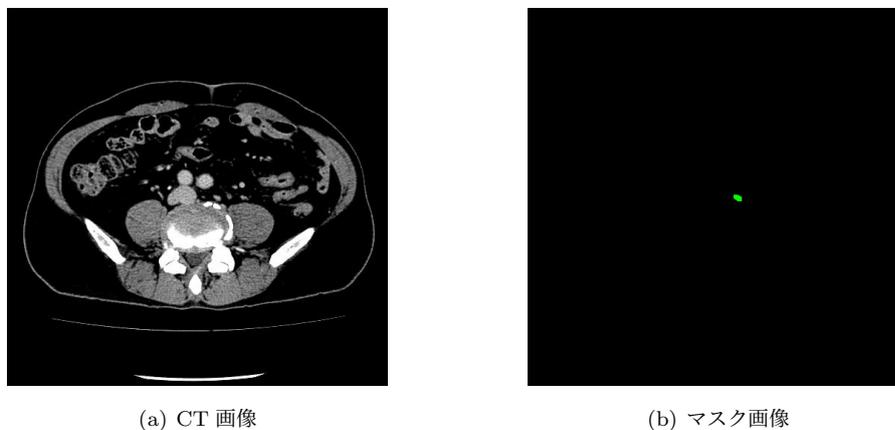


図 3.2: G1 の例

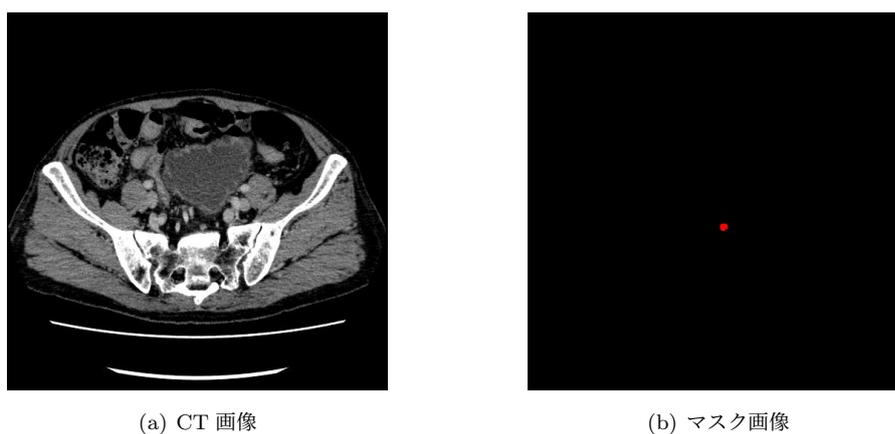


図 3.3: G3 の例

3.1.3 前処理

学習の効率化のため、データセットに対して 2 つの前処理を行う。

- (1) 2.3 節で述べた通り、医師が画像診断をするときには、リンパ節が見やすくなるようにウィンドウレベル変換を行う。この処理がディープラーニングの学習の効率をよくすることにも効果があると考え、元の DICOM 画像を、医師がリンパ節を診断するとき用いる値であるウィンドウレベル 75, ウィンドウ幅 280

でウィンドウレベル変換を行う。すなわち、CT 値-65 から 215 を画素値 0 から 255 にマッピングしてグレースケール画像に変換する。

- (2) データセットに含まれる CT 画像のサイズは 512×512 ピクセルである。U-Net では入力画像のサイズが大きくなるほどパラメータ数が多くなり、計算リソースの制約により複雑なネットワークを学習することが困難となるため、入力画像のサイズはより小さいほうが望ましい。そのため、元の画像から中央部分 256×256 ピクセルに切り出して使用する。本研究の対象とするリンパ節は画像の中央部分に位置しており、かつ画像の端の部分は人体が写っていない背景部分が多くを占めているため、画像の端の部分は学習に不要であると判断した。

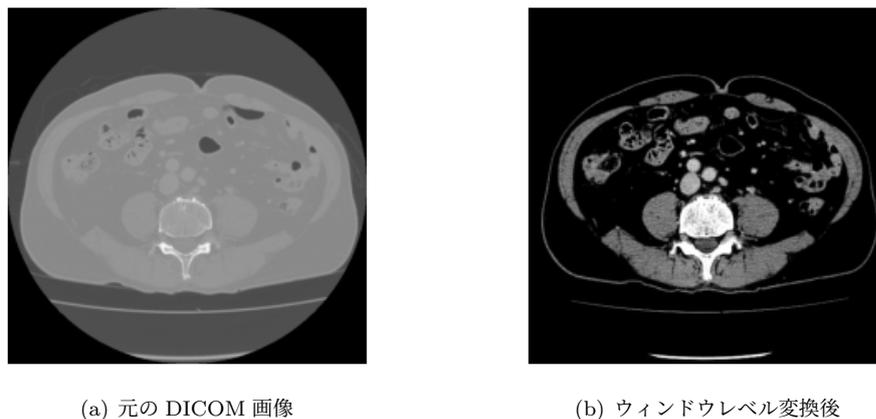


図 3.4: 前処理 (1) の例

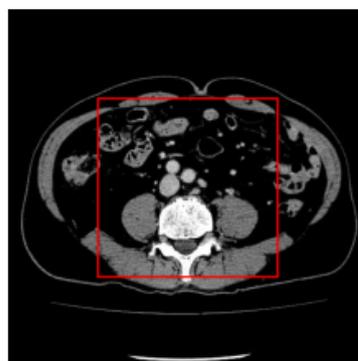


図 3.5: 前処理 (2) で切り取る範囲

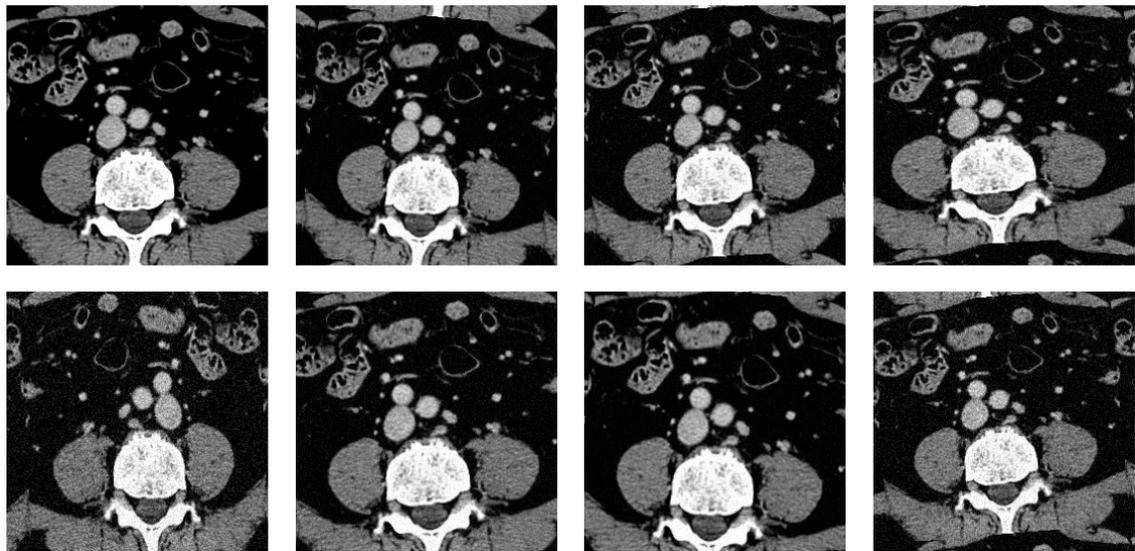


図 3.6: データ拡張の例 (左上がデータ拡張前, その他が拡張後)

3.1.4 データ拡張

ディープラーニングで十分な精度を得るには、1クラスあたり 5000 程度のデータが必要であることが経験則として知られている [13]。しかし、本研究で使用するデータセットの数はそれに対して少ない。また、クラスごとのデータ数の差が大きいと、精度が低くなることが知られているが、本研究で使用するデータセットは G1 と G3 のデータ数の差が大きい。これらの問題を解消するために画像処理によるデータ拡張を行う。行う画像処理は、画像の中心を軸とした 5° までの回転、上下左右各 5% までのシフト、シアー強度 0.1 までのシアー変換、10% までのズーム、水平方向のフリップである。データ拡張は学習時にリアルタイムで行うことで、計算リソースを節約する。また、データ不均衡の対応として、G3 のデータは G1 のデータの 5 倍の頻度で拡張されるようにする。データ拡張の例を図 3.6 に示す。

3.2 モデルの構築・学習

マルチクラスセグメンテーションを行うニューラルネットワークモデルを構築する。リンパ節の検出タスクには一般的に U-Net[14] または U-Net から派生した U-Net ファミリーと呼ばれるネットワークが用いられる。先行研究では、これらのネット

ワークを用いてリンパ節のバイナリセグメンテーションを行なっている。すなわち出力テンソルは $H \times W \times 1$ または $H \times W \times 2$ となる。ここで H と W は入力画像の高さと幅である。本研究ではこれらのネットワークの最終層を変更し、出力テンソルを $H \times W \times 3$ とすることで、G1, G3, リンパ節でないの3クラスのマルチクラスセグメンテーションを可能にする。

本研究では、U-Net と、U-Net ファミリーの中でも代表的なネットワークである U-Net++[15] と Attention U-Net[16] を使用する。それぞれのネットワークで、層の深さを4層と5層にしたモデルを構築する。学習には、3.1で作成した入力スライス数の異なる2種類のデータセットを使用する。学習時のパラメータはエポック数を75、バッチサイズを16、最適化関数をAdam、損失関数をFocalLoss[17]+JaccardLossとする。FocalLossとJaccardLossの計算式は式(3.1)、式(3.2)に示す通りである。次節以降で、使用したネットワークの詳細を紹介する。

$$FocalLoss(gt, pr) = -gt \cdot \alpha \cdot (1 - pr)^\gamma \cdot \log(pr) \quad (3.1)$$

$$JaccardLoss(A, B) = 1 - \frac{A \cap B}{A \cup B} \quad (3.2)$$

3.2.1 U-Net

U-Netはマルチクラスセマンティックセグメンテーションやバイナリセマンティックセグメンテーションに多く用いられるFully Convolutional Network(FCN)[18]である。全結合層が存在せず、ピクセル単位でのセグメンテーションが可能である。U-Netの概形は図3.7のようにU字になっており、エンコーダ部(図3.7の左側)とデコーダ部(図3.7の右側)からなるという特徴がある。エンコーダ部では畳み込み層とプーリング層によって局所的な特徴を抽出していくが、位置的な情報が失われていく。その後、デコーダ部で逆畳み込みによって特徴マップをアップサンプリングしていくが、単純にアップサンプリングするだけでは位置情報がうまく復元されない。そこで各層においてエンコーダ部とデコーダ部を連結することにより、位置情報を補完している。

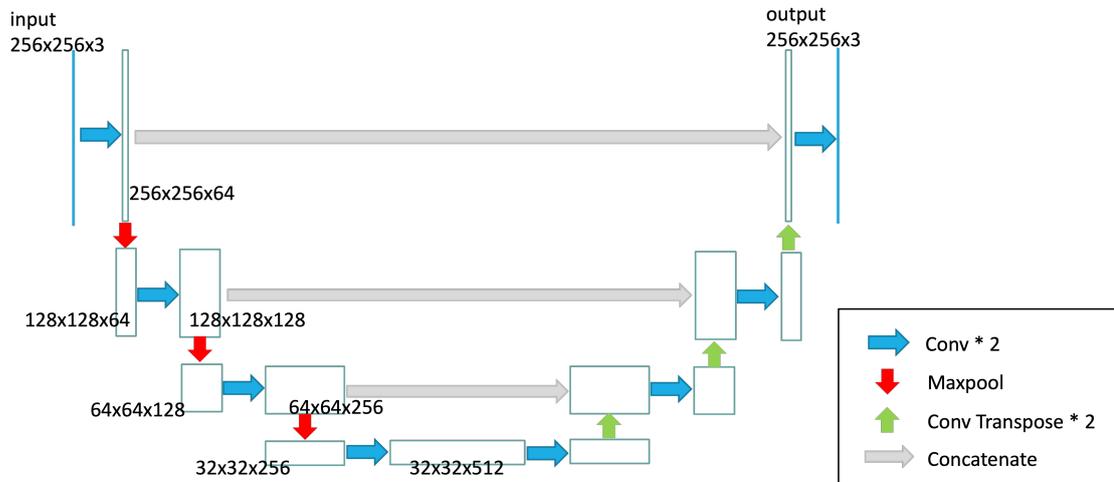


図 3.7: 4層の U-Net の概形

3.2.2 U-Net++

U-Net++はU-Netの派生ネットワークの1種である。U-Net++は図3.8のようになり、深さの違うU-Netがネスト状になっていることからNested U-Netとも呼ばれる。エンコーダとデコーダがスキップコネクションにより高密度で連結されているため、エンコーダとデコーダにおける特徴マップの差を小さくすることができる。これにより、U-Netよりも精度や学習の効率が改善する可能性が示されている。

3.2.3 Attention U-Net

Attention U-Netは図3.9のようにU-Netのエンコーダとデコーダの連結の途中にAttention機構を取り入れたネットワークである。近年、Attention U-Netをはじめ、Trans U-Net[19]やSwin U-Net[20]といったAttention機構を取り入れたU-Netファミリーが多く提案されているため、その代表として使用する。

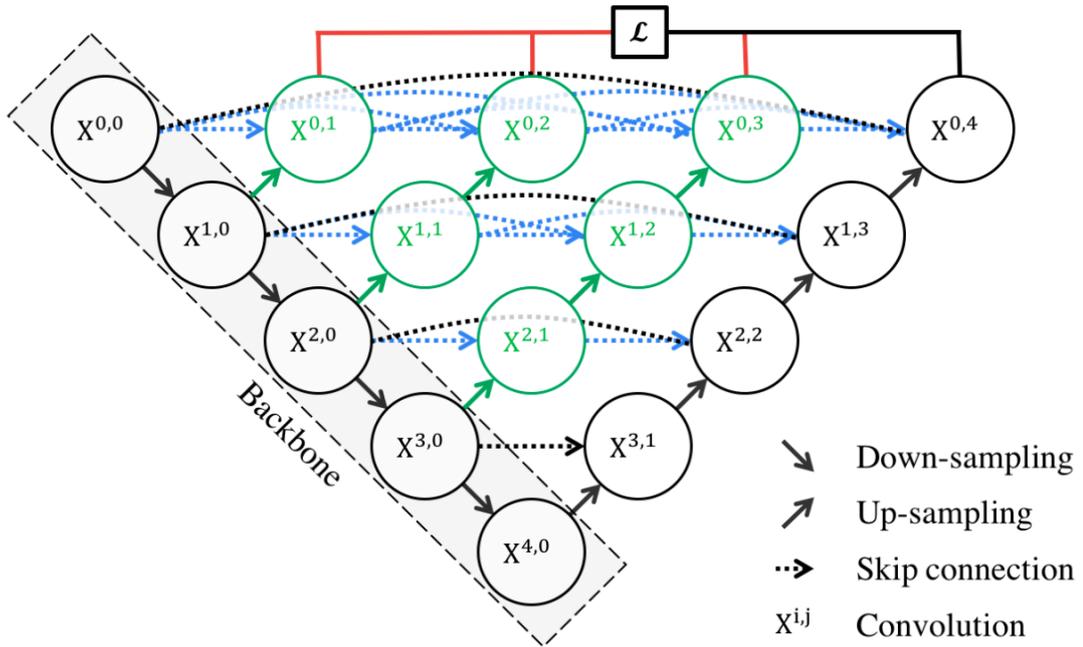


図 3.8: U-Net++の概形 (文献 [15] より引用)

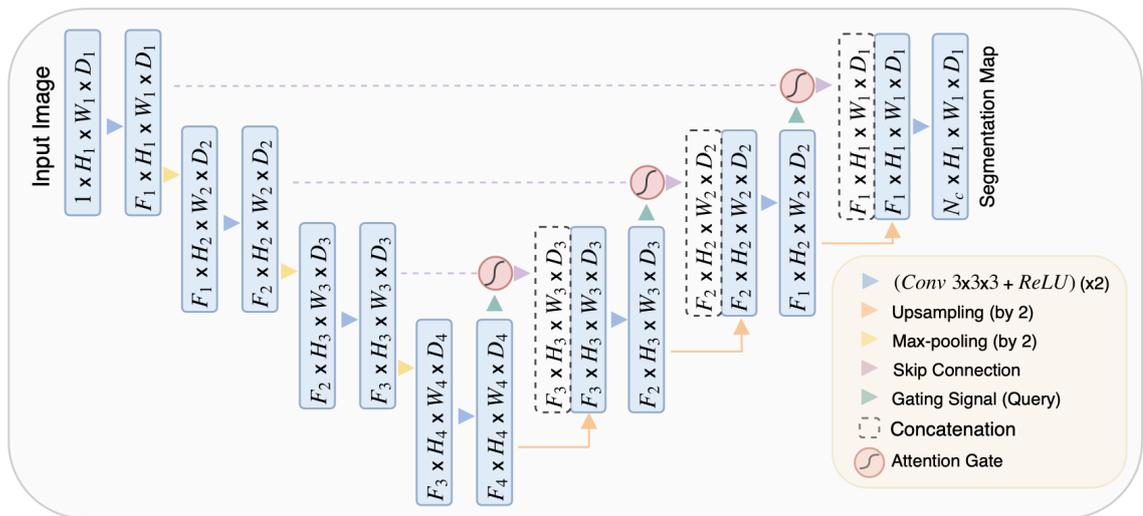


図 3.9: Attention U-Net の概形 (文献 [16] より引用)

第4章 実験

本研究の有効性を示すために実験を行う。まず、3.2節で構築したモデルの中での精度比較実験（実験1）を行う。次に、実験1で決定したモデルを用いて検出タスクと分類タスクを同時に行う場合と、先行研究で提案されている手法を用いて、検出タスクと分類タスクを別々に行なう場合の精度比較実験（実験2）を行う。

4.1 実験1

3.2節で構築したモデルの中で、リンパ節の検出タスクと分類タスクに最も適しているモデルを決定するため、精度比較実験を行う。

4.1.1 手順

提案手法の出力はマルチクラスセグメンテーションされた画像であるため、検出タスクと分類タスクの結果が融合したものであるが、評価は検出タスクと分類タスクで分けて行う。まず、G1とG3に関わらずリンパ節領域を検出できたかどうかで検出タスクの評価を行う。そして、検出タスクでTPとなったリンパ節に対して、分類できたかどうかで分類タスクの評価を行う。3.1節で構築した2種類のデータセットを8:2の割合で学習データとテストデータに分割し、それぞれに対してデータ拡張を行う。検出タスクの評価指標は、

TP リンパ節領域を正しくリンパ節領域と検出

FN リンパ節領域を誤ってリンパ節でない領域と検出

TN リンパ節でない領域を正しくリンパ節でない領域と検出

FP リンパ節でない領域を誤ってリンパ節領域と検出

としたときに式 (4.1) から式 (4.4) に示すものを用いる。分類タスクの評価指標は、

TP G3 を正しく G3 と分類

FN G3 を誤って G1 と分類

TN G1 を正しく G1 と分類

FP G1 を誤って G3 と分類

としたときに式 (4.1) から式 (4.4) に示すものを用いる。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.3)$$

$$F \text{ 値} = \frac{2 * Recall * Precision}{Recall + Precision} \quad (4.4)$$

結果を表 4.1, 表 4.2 に示す。また, 出力の例を図 4.2 から図 4.3 に示す。モデルの出力は各ピクセルについて G1, G3, リンパ節でないの各クラスである確率を示した $256 \times 256 \times 3$ のテンソルとなるため, 各ピクセルの最も確率の高いクラスが G1 なら緑, G3 なら赤, リンパ節でないなら黒とした画像として示す。

表 4.1: 各モデルの検出タスクの結果

ネットワーク	スライス数	深さ	Accuracy	Precision	Recall	F 値
U-Net	1	4	0.9945	0.3755	0.4681	0.4167
U-Net	1	5	0.9920	0.4831	0.3237	0.3876
U-Net	3	4	0.9744	0.4841	0.5115	0.4974
U-Net	3	5	0.9947	0.3793	0.4995	0.4312
U-Net++	1	4	0.9737	0.4206	0.4961	0.4553
U-Net++	1	5	0.9737	0.4113	0.4962	0.4498
U-Net++	3	4	0.9740	0.4235	0.5035	0.4600
U-Net++	3	5	0.9738	0.4209	0.4976	0.4561
Attention U-Net	1	4	0.9737	0.4212	0.4968	0.4559
Attention U-Net	1	5	0.9739	0.4290	0.4999	0.4618
Attention U-Net	3	4	0.9742	0.4193	0.5070	0.4590
Attention U-Net	3	5	0.9711	0.4784	0.4510	0.4643

表 4.2: 各モデルの分類タスクの結果

ネットワーク	スライス数	深さ	Accuracy	Precision	Recall	F 値
U-Net	1	4	1.000	1.000	1.000	1.000
U-Net	1	5	1.000	1.000	1.000	1.000
U-Net	3	4	0.9688	1.000	0.9630	0.9286
U-Net	3	5	1.000	1.000	1.000	1.000
U-Net++	1	4	1.000	1.000	1.000	1.000
U-Net++	1	5	1.000	1.000	1.000	1.000
U-Net++	3	4	1.000	1.000	1.000	1.000
U-Net++	3	5	1.000	1.000	1.000	1.000
Attention U-Net	1	4	1.000	1.000	1.000	1.000
Attention U-Net	1	5	1.000	1.000	1.000	1.000
Attention U-Net	3	4	1.000	1.000	1.000	1.000
Attention U-Net	3	5	0.9167	0.8889	0.9412	0.9143

4.1.2 考察

表 4.1 より、検出タスクでは総合的な検出精度を測る指標である F 値について、すべてのネットワークの種類において入力スライス数が 1 スライスの時より 3 スライスの際のほうが高くなっている。これは複数スライスから 3 次元的な形状を学習することで、精度が高くなることを示している。また、一般的には U-Net よりも U-Net++ や Attention U-Net といった U-Net 派生ネットワークのほうが精度が高くなり、層

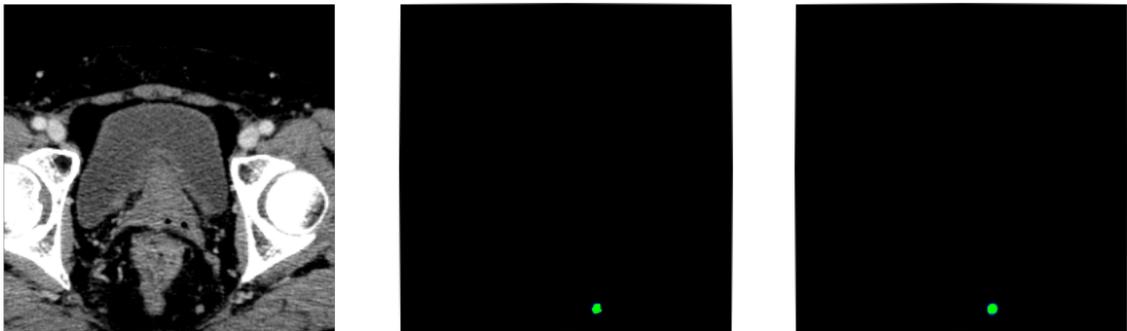


図 4.1: 検出タスクが TP で, 分類タスクが TP となる例 (左から入力, 正解マスク, 出力)

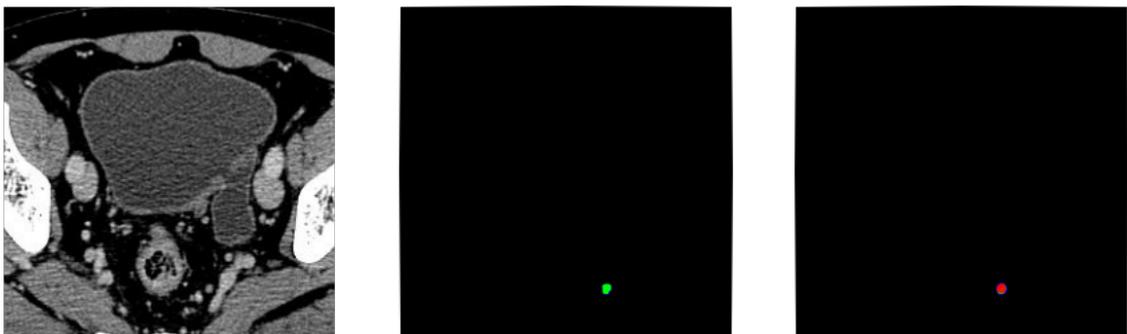


図 4.2: 検出タスクが TP で, 分類タスクが FP となる例 (左から入力, 正解マスク, 出力)

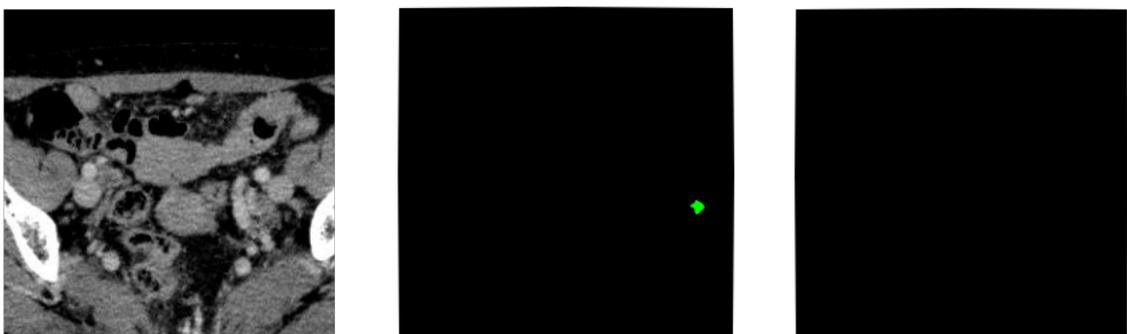


図 4.3: 検出タスクが FN となる例 (左から入力, 正解マスク, 出力)

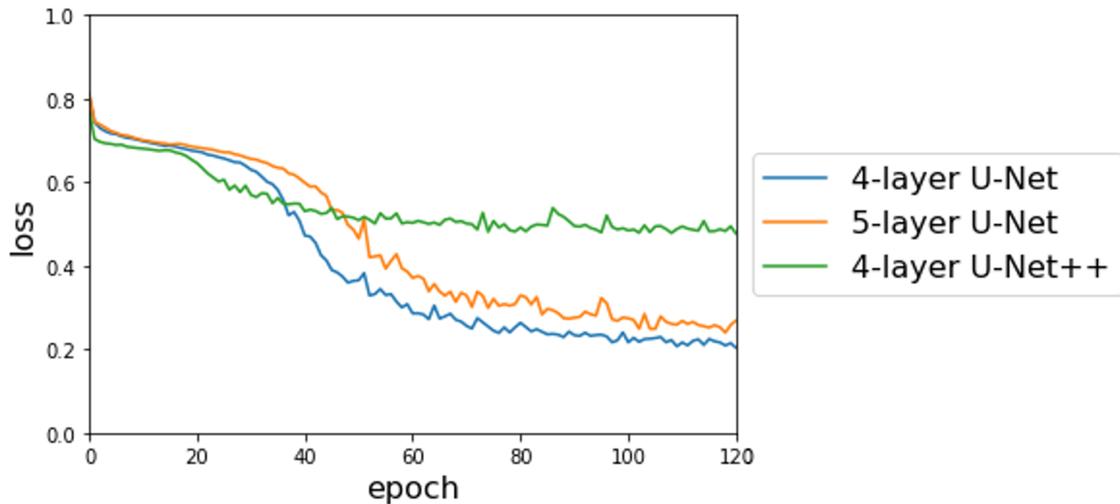


図 4.4: 学習曲線

が深いほど精度が高くなるとされている。しかし、今回の結果では深さ4層のU-Netが最もF値が高くなっている。これは今回用いたデータセットが、検出タスクの学習に十分なものでなかったため、複雑なモデルは十分に学習ができなかったためだと考えられる。深さ4層のU-Netと深さ5層のU-Net、深さ4層のU-Net++の学習曲線を図4.4に示す。深さ4層のU-Netに比べて、他のモデルでは損失関数が下がりがきっていないことから、学習が十分に進んでいないことが確認できる。

表4.2より、分類タスクでは全てのモデルで非常に高い精度で分類できていることが確認できる。理由については実験2で考察する。以上のことから、検出タスクと分類タスクの総合的な精度が最も高いのは、入力スライス3スライス、深さ4層のU-Netであると言える。以降はこのモデルを用いてリンパ節のマルチクラスセグメンテーションを行うことを提案手法と呼ぶ。

4.2 実験2

提案手法を用いて検出タスクと分類タスクを同時に行う場合と、検出タスクと分類タスクを別々の手法を用いて行う場合の精度比較実験を行う。

4.2.1 手順

検出タスクの比較対象には文献 [6] で提案されている, U-Net を用いてリンパ節の候補を検出した後に 3D CNN を用いて候補から偽陽性を削減するという手法を使用する. 分類タスクの比較対象には文献 [7] で提案されている, ImageNet[21] で学習済みの AlexNet[11] を転移学習したモデルを使用する. データセットには, 3.1 節で構築したデータセットを, 比較対象のモデルに合わせて変換したものを使用する. 検出タスクには, 元のデータセットの G1, G3 の 2 種類のマスクを区別せず 1 つのマスクに統合したものを使用する. 分類タスクには, 元のデータセットのマスク画像をもとに, CT 画像のリンパ節とその周辺を 16×16 ピクセルで切り出した画像と, そのリンパ節が G1 か G3 かというラベルのセットにしたものを使用する. 各タスクの評価指標には実験 1 と同じものを使用する. 結果を表 4.3, 表 4.4 に示す.

表 4.3: 検出タスクの結果

モデル	Accuracy	Precision	Recall	F 値
提案手法	0.9744	0.4841	0.5115	0.4974
先行研究手法 [6]	0.9770	0.4980	0.4932	0.4956

表 4.4: 分類タスクの結果

モデル	Accuracy	Precision	Recall	F 値
提案手法	0.9688	1.000	0.9630	0.9286
先行研究手法 [7]	0.7755	0.8182	0.7200	0.7660

4.2.2 考察

表 4.3 より, 検出タスクでは提案手法と先行研究手法の F 値は同程度であり, 検出精度に差はないと言える. 一方で, 表 4.3 より, 分類タスクでは提案手法のほうが大きく勝っていると言える. 先行研究手法では検出タスクは G1 か G3 かを判定せずに検出を行うため, 分類が難しいリンパ節も検出され分類精度が低くなるが, 提案手法では検出の時点でそれが G1 か G3 かを判定するため, 分類が難しいリンパ節をそもそも検出しないためであると考えられる. このことから, 提案手法は先行研究手法よりも分類タスクの精度が高くなる利点がある一方で, 分類が困難なリンパ節は

検出しづらくなるという欠点があることが示唆される。例として、2.2節で説明したG1とG3の特徴のどちらにも当てはまらないような歪な形状のリンパ節は提案手法では検出できていない(図4.3)。ただし、今回の実験結果では検出精度は同程度で、分類精度は提案手法のほうが高いため、提案手法は、検出タスクと分類タスクを別々の手法で行う場合よりも総合的な精度が高いことが確認できる。また、検出タスクの先行研究手法は、The Cancer Imaging Archive (TCIA) で公開されているCT画像とリンパ節のデータセット [22] に対しては、F値が0.77という高い精度が得られることが示されている [6]。しかし、表4.3より、本実験での先行研究手法のF値は0.4974とTCIAのデータセットに対する精度より低くなっていることから、今回用いたデータセットは検出タスクが難しいデータセットであったと言える。このことから、データセットの規模を大きくすることでさらなる精度向上が期待できる。

第5章 むすび

本研究では、今まで別々の手法を組み合わせる必要があったリンパ節の検出と分類という2つのタスクを、マルチクラスセグメンテーションという1つのタスクに置き換えることにより、単一のU-Netファミリーのモデルで行う手法を提案した。実験により、入力スライス3スライス、深さ3層のU-Netを用いることで、先行研究の手法を用いて検出タスクと分類タスクを別々に行なった場合と比べて、検出精度は同程度を維持しつつ、分類精度が向上することを示した。特に検出タスクにおけるRecallは0.5115となり、医師と同程度の正診率を示した。データセットの質を高めることで、U-Net派生ネットワークや層の深いネットワークが十分に学習できるようになれば、さらに精度が向上する余地があると考えられる。さらに、提案手法には分類が困難なリンパ節は検出しにくいですが、検出できたリンパ節の分類精度は非常に高いという特徴があることが示唆された。このことから、提案手法は検出と分類を同時に行う必要があり、かつ検出精度よりも分類精度を重視したいという、他の問題にも適用できると考えられる。

今後の課題として、データセットの増強や、転移学習などの少ないデータセットでも精度を向上させる手法の導入が挙げられる。

謝辞

本研究を行うにあたり、終始熱心にご指導を頂いた名古屋工業大学 舟橋健司 准教授、伊藤宏隆 助教に深く感謝いたします。また、日頃よりご指導いただきました中部大学 岩堀祐之 教授に感謝いたします。なお、本研究において、CT画像の提供をはじめ、ご協力を頂きました愛知県がんセンター 大内晶 先生、清水 泰博 先生に感謝いたします。さらに、本研究を進めるにあたり多くの助言を頂き、御協力いただいた名古屋工業大学 舟橋研究室の皆様に深く感謝し、厚く御礼申し上げます。

参考文献

- [1] “直腸：[国立がん研究センター がん統計],” https://ganjoho.jp/reg_stat/statistics/stat/cancer/7_rectal.html. (Accessed on 01/31/2022).
- [2] T. Akiyoshi, T. Watanabe, S. Miyata, K. Kotake, T. Muto, and K. Sugihara, “Results of a Japanese Nationwide Multi-Institutional Study on Lateral Pelvic Lymph Node Metastasis in Low Rectal Cancer: Is it regional or distant disease?,” *Annals of surgery*, vol.255, no.6, pp.1129–1134, 2012.
- [3] S. Fujita, S. Yamamoto, T. Akasu, and Y. Moriya, “Prognostic Factors of Rectal Cancer Patients with Lateral Pelvic Lymph Node Eetastasis.,” *Hepato-gastroenterology*, vol.59, no.120, pp.2494–2497, 2012.
- [4] “Jscqr — 大腸癌研究会,” <http://www.jsccr.jp/guideline/2019/cq.html>. (Accessed on 01/12/2022).
- [5] H. Oda, H.R. Roth, K.K. Bhatia, M. Oda, T. Kitasaka, S. Iwano, H. Homma, H. Takabatake, M. Mori, H. Natori, et al., “Dense volumetric detection and segmentation of mediastinal lymph nodes in chest ct images,” *Medical Imaging 2018: Computer-Aided Diagnosis*, vol.10575, p.1057502, 2018.
- [6] Jayant P. Singh, Yuji Iwahori, M. K. Bhuyan, Hiroyasu Usami, Taihei Oshiro, Yasuhiro Shimizu, “Mediastinal Lymph Node Detection using Deep Learning,” *ICPRAM 2020*, pp.159–166, Jan. 2020.
- [7] J. Li, P. Wang, Y. Zhou, H. Liang, and K. Luan, “Different machine learning and deep learning methods for the classification of colorectal cancer lymph node metastasis images,” *Frontiers in Bioengineering and Biotechnology*, p.1521, 2021.

- [8] 岡田佳朗, “CNN を用いた造影 CT 画像からのリンパ節転移の有無分類,” 中部大学修士論文, 2021.
- [9] G. Brown, C.J. Richards, M.W. Bourne, R.G. Newcombe, A.G. Radcliffe, N.S. Dallimore, and G.T. Williams, “Morphologic Predictors of Lymph Node Status in Rectal Cancer with Use of High-Spatial-Resolution MR Imaging with Histopathologic Comparison,” *Radiology*, vol.227, no.2, pp.371–377, 2003.
- [10] T. Akasu, G. Inuma, M. Takawa, S. Yamamoto, Y. Muramatsu, and N. Moriyama, “Accuracy of High-Resolution Magnetic Resonance Imaging in Pre-operative Staging of Rectal Cancer,” *Annals of surgical oncology*, vol.16, no.10, pp.2787–2794, 2009.
- [11] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol.25, pp.1097–1105, 2012.
- [12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol.86, no.11, pp.2278–2324, 1998.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Adaptive Computation and Machine Learning series, MIT Press, 2016.
- [14] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *International Conference on Medical image computing and computer-assisted intervention*, pp.234–241, 2015.
- [15] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp.3–11, Springer, 2018.

- [16] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., “Attention u-net: Learning where to look for the pancreas,” arXiv preprint arXiv:1804.03999, 2018.
- [17] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” Proceedings of the IEEE international conference on computer vision, pp.2980–2988, 2017.
- [18] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp.3431–3440, 2015.
- [19] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, and Y. Zhou, “Transunet: Transformers make strong encoders for medical image segmentation,” arXiv preprint arXiv:2102.04306, 2021.
- [20] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-unet: Unet-like pure transformer for medical image segmentation,” arXiv preprint arXiv:2105.05537, 2021.
- [21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” 2009 IEEE conference on computer vision and pattern recognition, pp.248–255, 2009.
- [22] H.R. Roth, L. Lu, A. Seff, K.M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R.M. Summers, “A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations,” International conference on medical image computing and computer-assisted intervention, pp.520–527, 2014.