

修士論文

深層学習による
骨腫瘍悪性度自動推定

令和4年度修了

三重大学大学院 工学研究科 情報工学専攻
ヒューマンコンピュータインタラクション研究室

古尾 海音

はじめに

骨腫瘍とは、骨に生じた異常な細胞の増殖物である。骨腫瘍には悪性と良性があり、悪性骨腫瘍には骨肉腫や軟骨肉腫などの原発性骨肉腫や他の臓器に発生したがんが骨に転移する転移性骨腫瘍があるが、いずれも命にかかわる難治性の疾患である。患者数は人口の0.3%と非常に少なく、腫瘍の90%以上が良性であり、良性腫瘍と悪性腫瘍の形状や大きさなどの違いが少ないことから、悪性腫瘍の発見には高い専門性を必要とし、多大な労力と時間を要する。骨の見た目が異常であることや骨の増殖・穴の形成から骨の異常を目視で確認できるため、簡易的な骨腫瘍の診断はX線画像診断により行われている。しかし、専門医でなければ見落としが多いことが早期発見への障壁となるため、予後不良の原因となっている。また、X線画像でそうした異常を見ることはできるものの、通常は腫瘍が良性か悪性かは判断できない。よって、多くの場合、腫瘍専門医が勤めているがんセンターや大学病院を受診し、腫瘍や骨の組織サンプルを採取して顕微鏡で調べる生検が必要である。骨腫瘍を自動で発見し、骨腫瘍の良性・悪性を判断できるコンピュータ支援診断(Computer-aided Diagnosis:CAD)システムを初診のクリニックや大学病院などに導入することができれば、医師の経験と知識を補う大きな力となり、患者にとっても早期発見、早期治療へと繋がる。

近年、CADシステムにおいて深層学習を用いる手法が多く提案され、骨腫瘍分野においても深層学習を用いた骨腫瘍自動抽出の先行研究[1]があるが、抽出の精度が低く骨腫瘍の良性・悪性のクラス分類は行われていない。本研究では深層学習を用いて単純X線画像から骨腫瘍の領域を自動で抽出し、腫瘍の悪性度を推定する手法を提案する。CNNモデルを用いて単純X線画像から骨腫瘍領域を自動で抽出し、抽出された骨腫瘍を小領域毎のパッチ画像に細分化する。細分化したパッチ画像をCNNモデルで学習し、得られた結果を加重平均で統合することで腫瘍の悪性度を計算する。腫瘍の悪性度から良性腫瘍か悪性腫瘍かを推定する。

実験では、三重大学医学部附属病院から提供された医学系研究倫理審査委員会承認済みの被験者34名の単純X線画像と、十分な訓練を受けた整形外科医によって認定された腫瘍領域のマスク画像を用いた。単純X線画像から骨腫瘍領域を自動で抽出する実験にお

いて、IoU が 0.68, Dice 係数が 0.73 の結果が得られた。また、腫瘍の悪性度から良性腫瘍か悪性腫瘍かを推定する実験では、骨腫瘍領域を自動抽出した被験者 34 人に対してクラスを正しく推定することができなかった画像も存在したが、悪性腫瘍においては 1 名の被験者を除いてはクラスを正しく推定することができた。また、正解のマスク画像を用いて実験を行った結果とほとんど差異がなかった。このことから、推定した腫瘍領域の画像を使用してもクラスを正しく分類できることが分かった。

目次

はじめに	i
第 1 章 緒言	1
1.1 研究背景	1
1.2 関連研究	2
1.3 研究目的	2
1.4 研究知識	3
第 2 章 データセット	5
2.1 骨腫瘍画像データ	5
第 3 章 提案手法	6
3.1 概要	6
3.2 骨腫瘍の自動抽出	8
3.3 骨腫瘍のクラス分類	10
3.4 深層学習モデル	12
第 4 章 実験	15
4.1 実験データ	15
4.2 骨腫瘍の自動抽出・クラス分類における交差検証	17
4.3 実験条件	18
4.4 実験結果	20
第 5 章 結言	30
5.1 まとめ	30
5.2 今後の課題	31
付録 A 付録	32
A.1 プログラムの詳細	32

目次	iv
----	----

A.2 データセット	33
----------------------	----

謝辞	36
----	----

第 1 章

緒言

1.1 研究背景

骨腫瘍とは、骨に生じた異常な細胞の増殖物である。骨腫瘍には悪性と良性があり、悪性骨腫瘍には骨肉腫や軟骨肉腫などの原発性骨肉腫や他の臓器に発生したがんが骨に転移する転移性骨腫瘍があるが、いずれも命にかかわる難治性の疾患である。腫瘍は、原因不明の進行性の骨の痛み、腫れ、骨折の傾向などを引き起こすことがあり、悪性腫瘍においては骨が徐々に弱くなる。これにより、日常的な動きをしている際に骨が折れることが非常に起こりやすい。患者数は人口の 0.3% と非常に少なく、腫瘍の 90% 以上が良性であり、良性腫瘍と悪性腫瘍の形状や大きさなどの違いが少ないことから、悪性腫瘍の発見には高い専門性を必要とし、多大な労力と時間を要する。骨の見た目が異常であることや骨の増殖・穴の形成から骨の異常を目視で確認できるため、簡易的な骨腫瘍の診断は X 線画像診断により行われている。しかし、専門医でなければ見落としが多いことが早期発見への障壁となるため、予後不良の原因となっている。また、X 線画像でそうした異常を見ることはできるものの、通常は腫瘍が良性か悪性かは判断できない。よって、多くの場合、腫瘍専門医が勤めているがんセンターや大学病院を受診し、腫瘍や骨の組織サンプルを採取して顕微鏡で調べる生検が必要である。初診での X 線検査から大学病院での生検まで多大な時間を必要とするため、初診のクリニックでいかに早期発見ができるかが重要である。

骨腫瘍を自動で発見し、骨腫瘍の良性・悪性を判断できるコンピュータ支援診断 (Computer-aided Diagnosis: CAD) システムを初診のクリニックや大学病院などに導入することができれば、医師の経験と知識を補う大きな力となり、医師や患者の負担を軽減することができる。

1.2 関連研究

近年、深層学習による医用画像処理・解析の研究が多く提案されている。骨腫瘍分野においても深層学習を用いた骨腫瘍領域自動抽出の先行研究がある [1]。この研究では X 線画像と深層学習を用いて、骨腫瘍領域を自動推定する。また、腫瘍を良性・悪性・正常などのクラスに分類する研究では、多視点アンサンブル学習を用いた研究 [2] や脳腫瘍に対する腫瘍分類を行う研究 [3] がある。しかし、骨腫瘍に対する良悪性の判定については未だ提案されていない。

また、近年ではデータセットとして用いる医用画像も様々である。CT 画像を用いた大腸がん検出と肺疾患予後予測のための CAD に関する研究 [4] や病理画像と CNN を用いた乳がんの確定診断に関する研究 [5] などがある。本研究では、初診のクリニックなどの比較的規模が小さい医療機関での CAD システムの導入を目的としているため、X 線画像を用いて実験を行う。

1.3 研究目的

本研究では骨腫瘍の診断において、単純 X 線画像の腫瘍領域を抽出した後、その腫瘍が良性か悪性かを推定することができるコンピュータ支援診断システム (Computer-aided Diagnosis:CAD) を開発することを目的としている。これにより、良悪性の正確な判断や手術計画の時間削減を実現し、腫瘍専門医・患者の負担軽減を見込める。また、日常診断において誤診の減少や悪性腫瘍の発見率を向上させることで、早期発見や予防医療も期待できる。手法構築のためには、

1. 単純 X 線画像から骨腫瘍領域を自動抽出
2. 骨腫瘍の悪性度を推定

が必要である。本論文では、単純 X 線画像から骨腫瘍を自動でセグメンテーションし、腫瘍の悪性度を推定する手法を提案し、実験によってその有効性を検証する。

1.4 研究知識

1.4.1 深層学習

深層学習とは、ニューラルネットワークを多層に結合して表現・学習能力を高めた機械学習の一手法である。単純に多層にするだけでは、表現力不足や過学習などの問題があったが、Dropout 層や ReLU などの数々の工夫とビッグデータの登場により解決された。本研究では、画像のセグメンテーションに適した CNN モデルと、画像分類に特化した CNN モデルを使用する。

1.4.2 Convolutional Neural Network(CNN)

最初に、Neural Network(NN) とは、人間の脳内にある神経細胞と神経回路網を簡略化し、人工ニューロンという数式モデルで表現したものである。図 1.1 に NN の概要を示す。NN は、複数のニューロンが様々な形で結合されるが、無秩序に結合されるのではなく、複数のニューロンが集まった層を構造に持つ。最初の層は入力層と呼ばれ、データが入力され、最後の層は出力層と呼ばれ結果を出力する。また、入力層と出力層の間の層は隠れ層と呼ばれ、自由度の高いデザインから複数のデータ学習を可能にする。一般に、隠れ層が複数存在する NN を深層学習と呼ぶ。CNN は、畳み込みニューラルネットワークである。何段もの深い層を持つ NN で、畳み込み層やプーリング層などのいくつかの機能を持った層を積み上げることで構成される。畳み込み層で特徴抽出し、プーリング層ではロバスト性を与える(抽出された特徴が平行移動などでも影響を受けないようにする)。

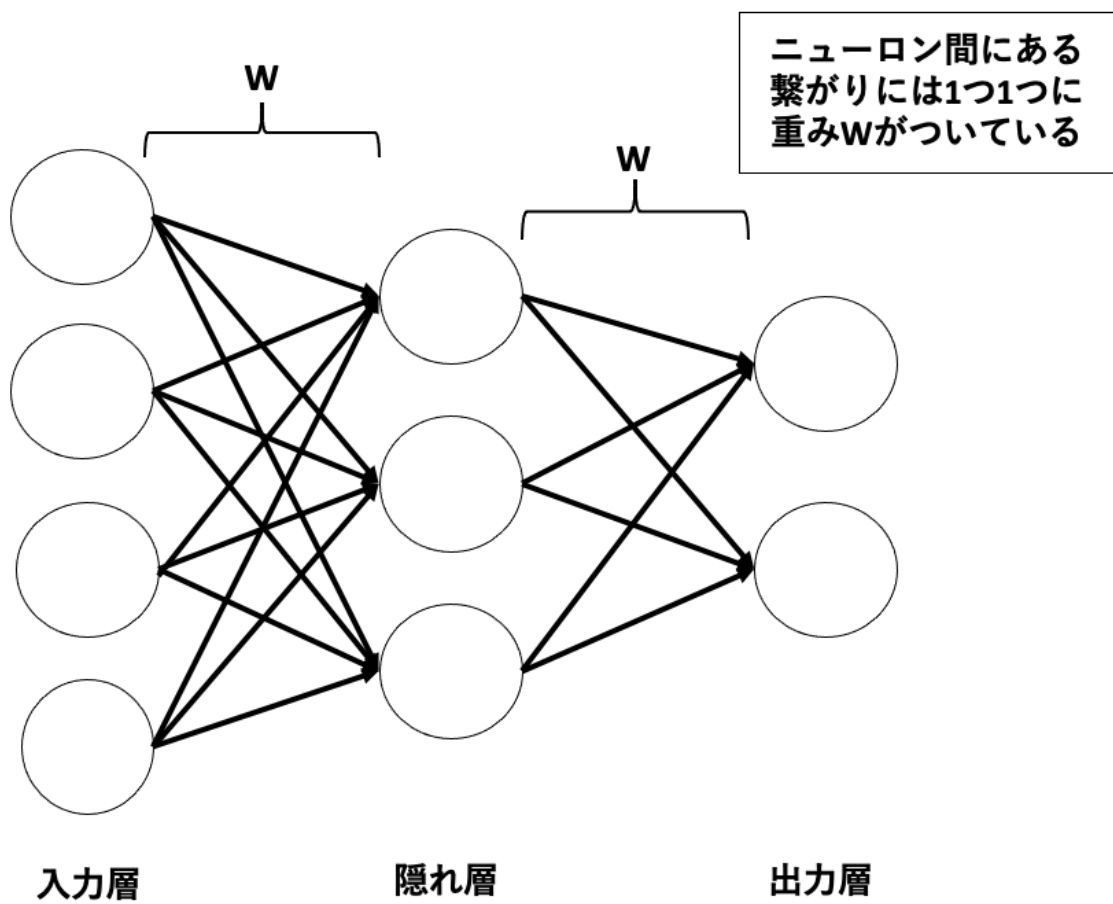


図 1.1: NN の概要

第 2 章

データセット

2.1 骨腫瘍画像データ

本研究で扱う骨腫瘍画像データは，三重大学医学部附属病院から提供された 34 人の被験者の X 線画像である．画像枚数は 96 枚で，画像サイズは 1693×2033 ピクセル， 2505×3015 ピクセル， 2336×2836 ピクセルなど画像ごとに異なり，正面や側面のみでなく，斜め方向など様々な方向や角度から撮影されている．骨腫瘍のマスク画像は専門医の指導により作成した．なお，本研究の実施は三重大学の医学系研究倫理審査委員会で承認済みである．図 2.1 は骨腫瘍患者の膝の X 線画像，図 2.2 は十分な訓練を受けた整形外科医によって認定された腫瘍領域のマスク画像を示す．マスク画像では，白色の前景 (腫瘍) 領域の画素値は 1 で，背景領域の画素値は 0 である．



図 2.1: 膝の骨腫瘍 X 線画像

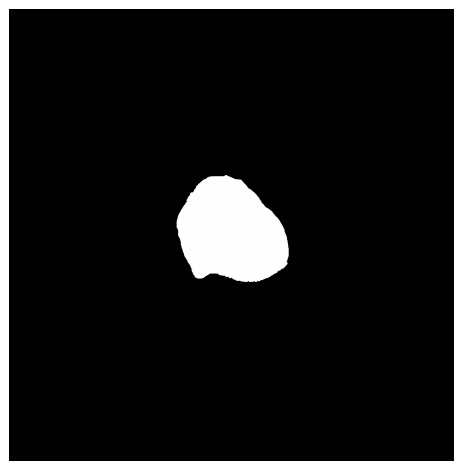


図 2.2: 骨腫瘍マスク画像

第 3 章

提案手法

3.1 概要

本章では、単純 X 線画像から骨腫瘍領域を抽出し、腫瘍の悪性度を推定する手法について述べる。本研究では、骨腫瘍領域の抽出を行うセグメンテーション手法と、腫瘍の悪性度から良性・悪性のクラス分類を行う手法を提案する。セグメンテーション手法では、CNN のモデルである Mask R-CNN [6] を用いて骨腫瘍領域を自動抽出する。クラス分類を行う手法では、単純 X 線画像をパッチに分割し、パッチ画像を CNN モデルである VGG-16 [7] または ResNet152[9] に学習させ、得られた学習済みモデルを用いて腫瘍の悪性度を計算し、良性・悪性を推定する。以下の手順で行う。

1. 単純 X 線画像から骨腫瘍領域を自動抽出（セグメンテーション）
2. 腫瘍の悪性度を推定（クラス分類）

図 3.1 に実験の概要を示す。

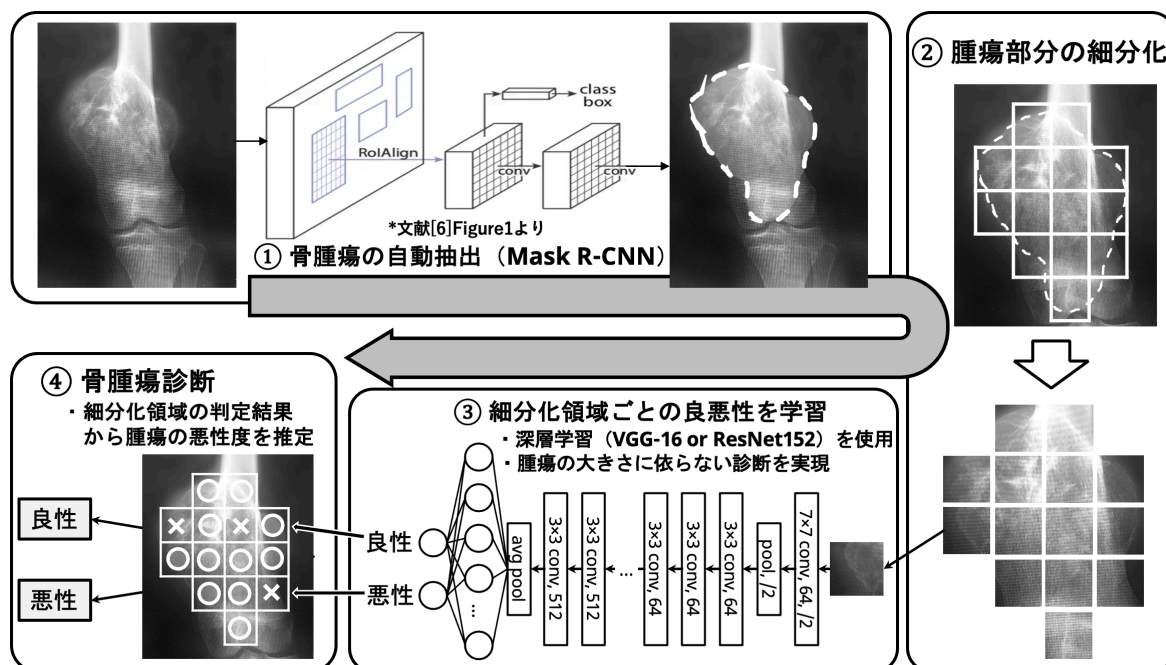


図 3.1: 実験の概要

3.2 骨腫瘍の自動抽出

3.2.1 予備研究

予備研究では，データセットとして単純 X 線画像を使用した場合と単純 X 線画像から腫瘍領域のみを手動でクロッピングした画像を使用した場合の実験を行い，精度を比較する．これは初期段階の腫瘍のサイズが単純 X 線画像に対して小さい場合があるため，セグメンテーションの精度が十分でない場合が考えられるためである．腫瘍領域のバウンディングボックスは，前景領域の最小及び最大座標値を計算することによって決定する．ただし，背景領域の腫瘍隣接ピクセルには，腫瘍または非腫瘍領域に対応するピクセルを決定するための重要な情報があると予測する．したがって，元の単純 X 線画像から検出されたバウンディングボックスの周囲に 50 ピクセルの幅のパディングを追加する．また，セグメンテーションモデルとしては U-Net[10]，DeepLab v3 [11, 12] を用いる．図 3.2 に骨腫瘍の単純 X 線画像，図 3.3 に腫瘍領域をクロッピングした画像を示す．



図 3.2: 骨腫瘍の単純 X 線画像



図 3.3: 腫瘍領域をクロッピングした画像

3.2.2 骨腫瘍の自動抽出手法

提案する骨腫瘍自動抽出手法では、Mask R-CNN を用いる。初期の腫瘍の大きさは、被験者によりばらつきがあり小さいものも存在するため、まず単純 X 線画像から腫瘍領域の画像を切り出し、切り出した腫瘍領域の画像を用いて学習を行うことで精度を向上させることができると考え、前述の予備研究を行った。予備研究の手法では手作業で単純 X 線画像から腫瘍領域のクロッピングが必要であり、手作業での腫瘍領域のクロッピングには多くの時間を必要とする。そのため、自動的に腫瘍領域をクロッピングし、クロッピングした腫瘍領域のみの画像を学習し、腫瘍領域のセグメンテーションを行うことのできる Mask R-CNN を使用する。Mask R-CNN によるセグメンテーションの流れを図 3.4 に示す。図 3.4 では、最初に入力画像として元の単純 X 線画像を入力する。次に、CNN において入力画像から特徴マップを出力し、特徴マップを region proposal 層に入力することで、腫瘍があると予測される領域のみを抽出した特徴マップを出力する。次に、得られた特徴マップのサイズを Rol Align 層で正規化した後に、deconvolution 層を用いて物体用のマスク画像を作成する。

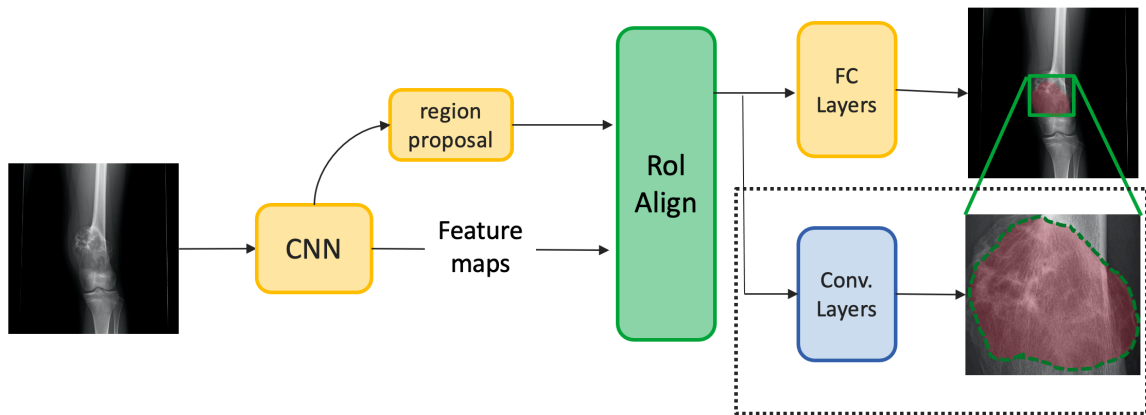


図 3.4: Mask R-CNN による自動抽出の流れ

3.3 骨腫瘍のクラス分類

骨腫瘍のクラス分類では，腫瘍が悪性である確率を悪性度として推定し，閾値を用いて良性・悪性のクラスに分類する手法を提案する．本実験では，CNN モデルである VGG-16 と ResNet152 を用いる．

3.3.1 前処理

腫瘍の大きさは被験者によって異なるため，本研究では単純 X 線画像から切り出した 64×64 ピクセルの小さなパッチ画像を用いて CNN モデルを学習させる． 64×64 ピクセルのウィンドウを 16 ピクセルずつずらしながら，パッチ画像の 90% 以上に腫瘍ピクセルが含まれるパッチ画像のみを抽出する．腫瘍のラベル情報に基づいて各パッチ画像にラベルを付与する．CNN モデルには 32×32 ピクセルの RGB 画像を入力する必要があるため，切り出したパッチ画像をリサイズし，RGB の各チャンネルに同じ画像を入力する．前処理から CNN モデルに画像を学習させるまでの過程を図 3.5 に示す．

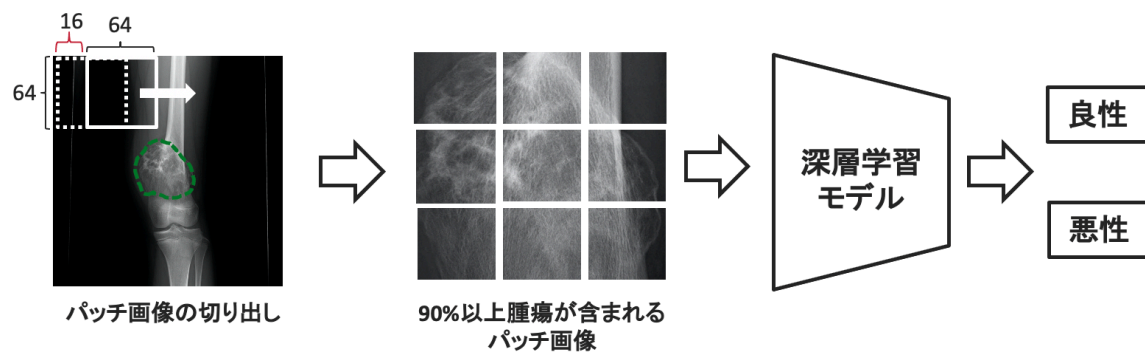


図 3.5: 前処理から学習までの流れ

3.3.2 クラス分類手法

学習した CNN モデルはパッチ画像に対してのみ腫瘍のクラス分類を行うため、クラス分類手法ではこれらの結果を統合して腫瘍のクラス分類を行う。CNN モデルは腫瘍の種類を 2 値で推定し、0 が良性腫瘍、1 が悪性腫瘍を示す。本研究では、パッチ画像の推定結果を単純に加重平均して統合することにより、悪性度 p_m を式 (3.1) により算出する。

$$p_m = \frac{\sum_{w \in \mathbf{W}, c \in \mathbf{C}} wc}{\sum_{w \in \mathbf{W}} w} \quad (3.1)$$

ここで、 \mathbf{C} は全ての画像パッチのクラス値 (0 または 1) の集合、 \mathbf{W} は各画像パッチの重みの集合で、その要素 w は次式で定義される。

$$w = \frac{1}{MN} \sum T \quad (3.2)$$

ここで、 M と N はパッチ画像の幅と高さであり、 T はパッチ領域の腫瘍マスクで、2 値である。悪性度 p_m は 0 から 1 の範囲で計算され、腫瘍が良性 (0) か悪性 (1) かを決定する閾値を定義する。実験では、試験的に閾値として 0.5 を用い、 p_m が 0.5 以上の腫瘍を悪性腫瘍、0.5 未満の腫瘍を良性腫瘍に分類した。学習で得られた学習済みモデルを用いて、パッチ画像の推定結果から加重平均を計算し悪性度 p_m を求める流れを図 3.6 に示す。

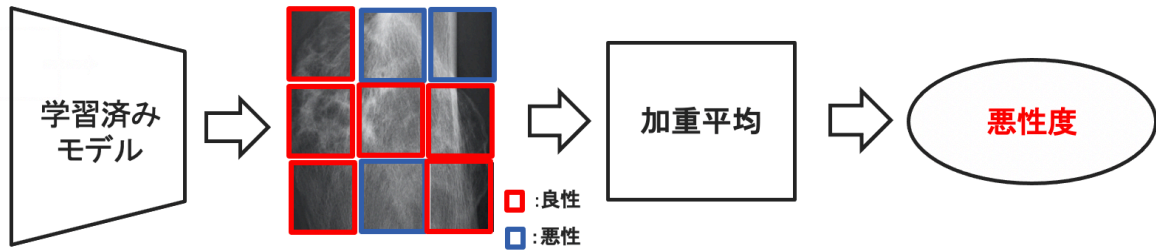


図 3.6: 悪性度を求めるまでの流れ

3.4 深層学習モデル

本研究で使用する深層学習モデルについて説明する。

3.4.1 Mask R-CNN

Mask R-CNN は Microsoft の研究グループが開発したディープラーニングのモデルであり、入力された画像に対して、決められた種類の物体を認識してその位置とクラス名を特定し、なおかつ画像中のピクセル単位で物体の写っている領域を特定するマルチタスクモデルである。Mask R-CNN は、その前身となる Faster R-CNN という手法をベースにしている。Faster R-CNN は、物体検出とクラス分類の機能を有していて、Mask R-CNN はそれにインスタンス・セグメンテーションの機能を付加する形で進化した手法といえる。図 3.7 に Mask R-CNN のアーキテクチャを示す。

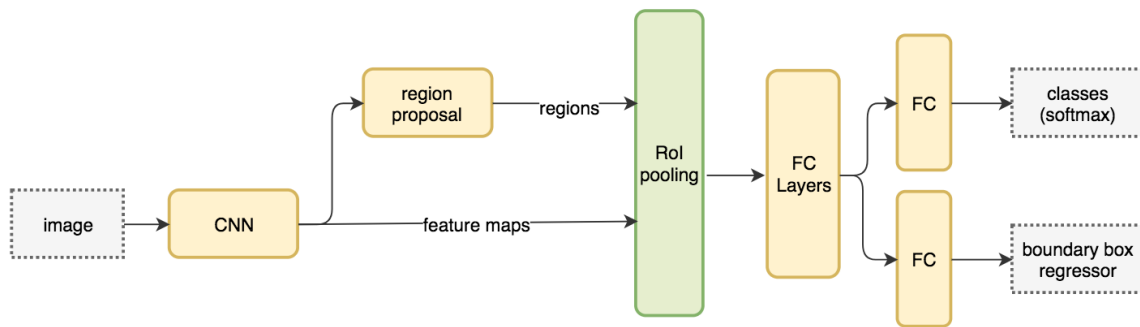


図 3.7: Mask R-CNN アーキテクチャ [6]

3.4.2 VGG16

VGG-16 は, ImageNet [8] と呼ばれる大規模画像データセットで学習された 16 層からなる CNN モデルである. 層の数が 11, 13, 19 層の VGG モデルも存在する. ImageNet データセットとはスタンフォード大学がインターネット上から画像を集めて分類したデータセットである. VGG モデルは構成がシンプルであるため, 深層学習の様々な応用手法のベースとして使用されることが多いモデルである. 図 3.8 に VGG-16 アーキテクチャを示す.

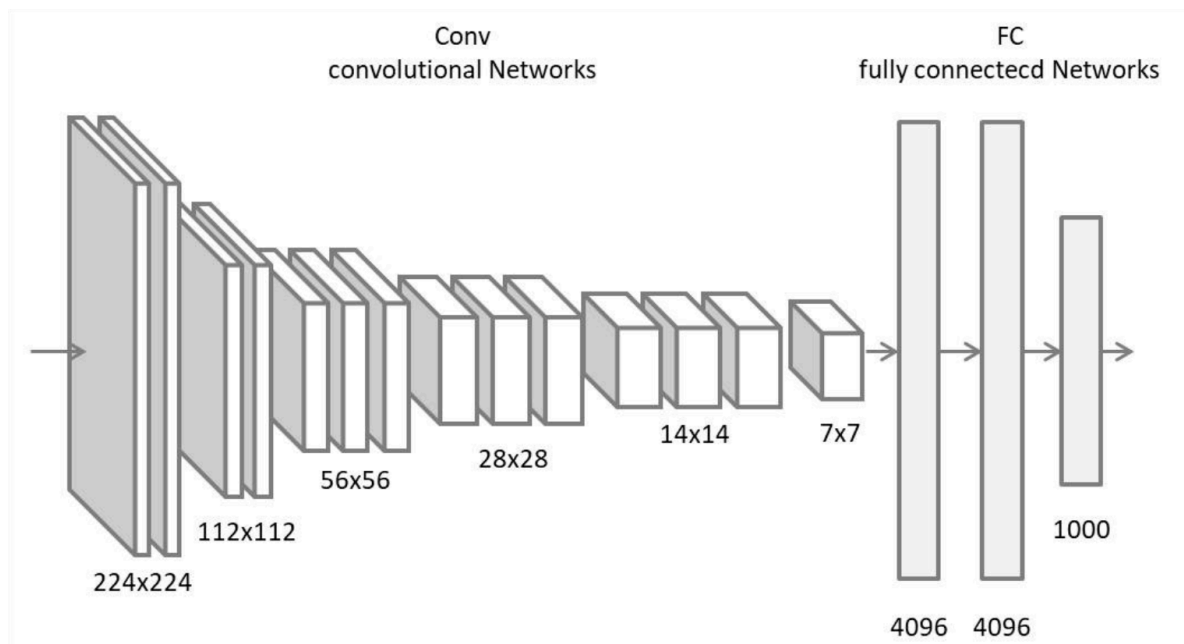


図 3.8: VGG-16 アーキテクチャ [7]

3.4.3 ResNet152

ResNet は、Microsoft Research(現 Meta AI Research) の Kaiming He 氏が 2015 年に考案した NN のモデルで、特徴としては層が非常に深いことと shortcut connection の導入である。shortcut connection の導入により深い層を実現する。精度は良いが、学習時間が長いという欠点がある。層の数が 18, 34, 50, 101 層の ResNet モデルも存在する。図 3.9 に ResNet アーキテクチャを示す。

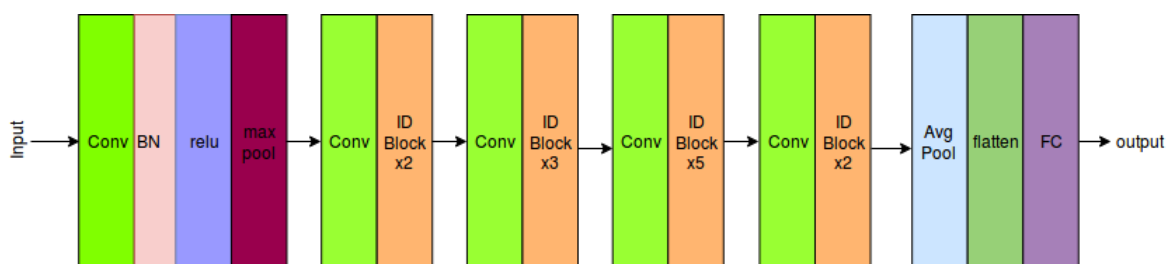


図 3.9: ResNet アーキテクチャ [9]

第 4 章

実験

4.1 実験データ

本研究では第 2 章データセットで記した通り，34 名の患者の画像 96 枚を使用する．骨腫瘍の領域を抽出するセグメンテーションの実験においては，学習に使用する画像のみ上下反転・左右反転・上下左右反転することでデータ拡張を行い，360 枚を使用した．腫瘍の悪性度からクラス分類を行う実験においては，学習用データの良性のパッチ画像を 7,904 枚，悪性のパッチ画像を 9,604 枚取得し，評価用データの良性のパッチ画像を 3,110 枚，悪性のパッチ画像を 6,080 枚取得した．こちらの実験においても，データ数が少ないためこのデータセットに対してデータ拡張を行う．左右反転を追加し，データセットのパッチ画像数は 2 倍となり，学習用データとして良性 15,808 枚，悪性 19,208 枚を使用し，評価用データとして良性 6,220 枚，悪性 12,160 枚を使用する．表 4.1 にセグメンテーション実験のデータセット内訳を，表 4.2 にクラス分類実験のデータセット内訳を示す．

表 4.1: セグメンテーション実験におけるデータセット

	学習用	評価用	合計
患者数	26	8	34
画像枚数	88	8	96
データ拡張後	352	8	360

表 4.2: クラス分類実験におけるデータセット

	良性	悪性	合計
患者数	23	11	34
切り出したパッチ枚数	11,014	15,684	26,698
データ拡張後パッチ枚数	22,028	31,368	53,396

4.2 骨腫瘍の自動抽出・クラス分類における交差検証

交差検証とは，汎化性能を評価する統計的な手法である．本実験では，データを4グループに分割して，そのうち1つを評価用データに，残りを学習用データとして使用することで，データセット1～4を作成する．これらの4つのデータセットを用いて骨腫瘍の自動抽出，骨腫瘍のクラス分類におけるモデルを学習する．表4.3にデータセット1～4の内訳を示す．

表 4.3: 交差検証におけるデータセット

	学習用	評価用	合計
データセット 1	26	8	34
データセット 2	26	8	34
データセット 3	25	9	34
データセット 4	25	9	34

4.3 実験条件

4.3.1 骨腫瘍領域自動抽出

単純 X 線画像から骨腫瘍を自動で抽出するセグメンテーション手法では, Adam optimizer を用いて, 学習率を 0.001, 損失関数は Binary Cross Entropy (BCE) とし, Mask R-CNN を学習させる. 評価指標として, Dice 係数 (DSC) と IoU を使用する. Dice 係数とは二つの集合がどれだけ類似しているかを表す式で, この値が 1 に近いほど, 形が似ていることになる. また, IoU とは物体検出で利用される評価指標の 1 つで, 画像の重なる割合を表しており, IoU が 1 に近いほど画像が重なっていることを意味する. Binary Cross Entropy を式 (4.1) に, Dice 係数を式 (4.2) に, BCEDice 損失関数を式 (4.3) に, IoU を式 (4.4) に示す.

$$BCE(P, Q) = \sum_{p \in P, q \in Q} p \log q \quad (4.1)$$

$$DSC(P, Q) = \frac{2|P \cap Q|}{|P| + |Q|} \quad (4.2)$$

$$Loss(P, Q) = wBCE(P, Q) + (1 - w)(1 - DSC(P, Q)) \quad (4.3)$$

$$IoU(P, Q) = \frac{|P \cap Q|}{|P \cup Q|} \quad (4.4)$$

ここで, w は重みパラメーター (実験では 0.5 を使用), P は推定腫瘍マスク, Q は正解腫瘍マスクであり, p, q はそれぞれの画素値である.

4.3.2 骨腫瘍悪性度推定

骨腫瘍の悪性度を推定する手法では, 損失関数にはバイナリクロスエントロピーを, 最適化アルゴリズムには SGD を, 深層学習モデルとして VGG-16 と ResNet152 を使用する. バッチサイズは 64, エポック数 1000 で実験を行う. 表 4.4 は学習パラメータをまとめたものである. また, 評価指標として, 学習済みモデルの精度においては, 適合率 (Precision) と再現率 (Recall) の調和平均を表す加重 F 値 (f1-score) を用いる. TP を良性を良性と正しく認識した数, TN を悪性を悪性と正しく認識した数, FP を悪性を誤って良性と誤認識した数, FN を良性を悪性と誤認識した数とする. 以下に適合率, 再現率, F 値の式を式 (4.3)~(4.5) に示す.

表 4.4: 学習パラメータ

	パラメータ 1	パラメータ 2
深層学習モデル	VGG-16	ResNet152
最適化関数	SGD	SGD
学習係数	0.001	0.001
エポック数	1,000	1,000
バッチサイズ	64	64
損失関数	交差エントロピー	交差エントロピー

$$Precision = \frac{TP}{TP + FP} \quad (4.5)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.6)$$

$$f1score = \frac{2Precision * Recall}{Precision + Recall} \quad (4.7)$$

4.4 実験結果

4.4.1 予備研究

表 4.5, 表 4.6 に各実験条件下での, 元の単純 X 線画像による学習, 腫瘍領域をクロッピングした画像による学習の実験結果を示す. 結果から, 最適化関数の比較では Adam, 損失関数の比較では BCEDice 損失を用いた場合に, 高い Dice 係数が得られることが確認できた. また, Adam, BCEDice の組み合わせを用いた実験においては, 元の単純 X 線画像を用いた実験 (表 4.5) では U-Net が Dice 係数 0.67 で, 手動クロッピング画像を用いた実験 (表 4.6) では DeepLab v3 が Dice 係数 0.92 で腫瘍領域の抽出を行えた. 実験結果から, 少ないデータでも検証用データにおいて十分な精度で抽出されたことから過学習は起こっていないことが確認できた. しかし, 単純 X 線画像全体に対する腫瘍領域は比較的小さいため, 元の単純 X 線画像で学習されたモデルが腫瘍領域を特定することは難しく, 元画像を用いた実験では十分な抽出精度が得られなかった. 一方で, 腫瘍領域をクロップすることで入力画像中の腫瘍領域が大きくなり, クロッピング画像で学習したモデルが腫瘍領域の情報を捉えやすくなり, どの実験条件の場合にも元の単純 X 線画像より高い精度となったと考えられる.

実験条件の比較では, 学習モデルや損失関数よりも, 最適化関数による違いで精度が変化することを確認できた. 最適化関数は損失関数が最小になるように, 損失関数を微分した勾配を用いてパラメータを更新していくアルゴリズムである. SGD はパラメータ更新にランダムに選ばれるデータを用いて, 損失を小さくする勾配方向を逐次的に探してその方向にパラメータを調整していく方法である. Adam は勾配の速度を導入してパラメータの振動を減らし学習が速く進むようにする方法と, 勾配の大きさに応じて学習率を下げることでパラメータの振動を抑える方法を組み合わせたものである. SGD では前の勾配を考慮しないため, 最小値でない極小値で余計な振動が起きてしまうが, Adam では振動を抑える方法を導入しているためより早く学習することが可能である. 今回の実験データでは, SGD による学習では Adam と比較して損失関数が最小となるようなパラメータに到達できず, Adam の方が SGD より高い精度になるような結果になったと考えられる.

表 4.5: 元の単純 X 線画像での実験結果

学習モデル	最適化関数	損失関数	IoU	Dice
U-Net	SGD	BCE	0.21	0.36
		BCEDice	0.27	0.47
	Adam	BCE	0.34	0.58
		BCEDice	0.41	0.67
DeepLab v3	SGD	BCE	0.15	0.28
		BCEDice	0.11	0.21
	Adam	BCE	0.38	0.58
		BCEDice	0.40	0.62

表 4.6: 手動クロッピング画像での実験結果

学習モデル	最適化関数	損失関数	IoU	Dice
U-Net	SGD	BCE	0.74	0.85
		BCEDice	0.75	0.86
	Adam	BCE	0.82	0.90
		BCEDice	0.83	0.91
DeepLab v3	SGD	BCE	0.78	0.88
		BCEDice	0.73	0.85
	Adam	BCE	0.84	0.91
		BCEDice	0.84	0.92

4.4.2 骨腫瘍領域自動抽出の実験結果

予備実験の結果では、腫瘍領域をクロッピングした画像で学習したモデルが腫瘍領域の情報を捉えやすくなり、元の単純 X 線画像より高い精度を示した。また、最適化関数では本実験において Adam の方が SGD より優れているという結果となった。しかし、手動で元の単純 X 線画像から腫瘍領域のクロッピングを行う必要があるため、多大な時間を要する。さらに、表 4.6 では腫瘍領域をクロッピングした画像を評価に使用した。実際に医療現場では腫瘍領域が不明な未知の患者の画像を扱うため、事前にクロッピングすることは不可能である。

そこで、本実験では自動で腫瘍領域のクロッピングを行い、セグメンテーションできる Mask R-CNN と最適化関数 Adam を用いて実験を行った。Mask R-CNN におけるセグメンテーションの実験結果は以下の通りである。表 4.7 は得られた Dice 係数と IoU を示す。図 4.1 は Mask R-CNN におけるセグメンテーションの結果の一例である。単純 X 線画像をグレースケールで表示し、腫瘍として正しく抽出された部分を黄色、腫瘍であるが抽出できなかった部分を緑、腫瘍ではないが腫瘍として抽出された部分を赤で表示した。

表 4.7 の結果より、交差検証のために作成した 4 つのデータセットの中で、データセット 1 が IoU 0.676、Dice 係数 0.726 と、一番高い精度を示した。この結果は表 4.6 の手動クロッピング画像での実験の精度と比較すると低くなったが、表 4.5 の元の単純 X 線画像での実験の精度よりは高くなったことから、腫瘍領域の自動クロッピングが有効であることを示した。しかし、データセット 2 やデータセット 4 では他のデータセットより低い精度となり、データによって学習に偏りがある可能性が考えられる。

表 4.7: Mask R-CNN による実験結果（交差検証）

	IoU	Dice 係数
データセット 1	0.68	0.73
データセット 2	0.48	0.49
データセット 3	0.66	0.71
データセット 4	0.51	0.55
平均 ± 標準偏差	0.58±0.088	0.62±0.10

図 4.1 に示す結果では、未抽出領域が少ないことから概ね正しく腫瘍領域を抽出できていることが確認できた。腫瘍領域をクロップすることで画像に対する腫瘍領域の割合が大きくなり、少量のデータでもセグメンテーションの学習ができたと考えられる。元の単純 X 線画像を用いた実験と比較すると、今後の骨腫瘍領域抽出の完全自動化・精度向上には

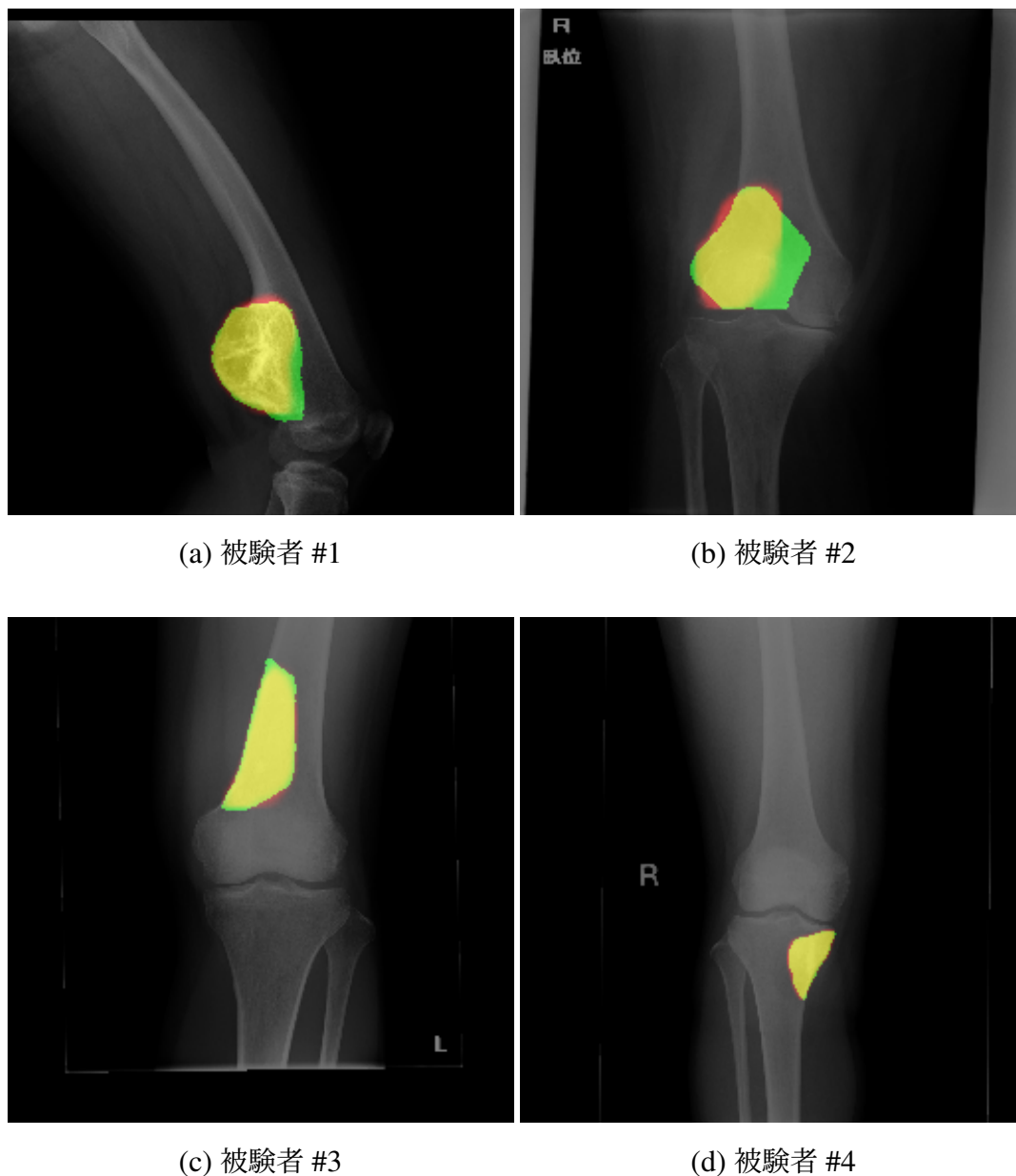


図 4.1: Mask R-CNN によるセグメンテーションの結果の例

外接矩形検出などにより大まかに骨腫瘍領域を自動検出する手法との組み合わせが有効であると考えられる。ただし、今回の実験では、図 4.1 に示した画像以外に未検出や過検出領域が存在する画像がある。この抽出結果から腫瘍の良悪性判別を行うことを考えると、より正確な腫瘍領域の抽出が必要となる。精度を改善するには、深層学習に用いるデータの不足について対策する必要がある。今回、データ数不足を補うため、88 枚の学習データについては垂直方向・水平方向反転の組み合わせによりデータ数を 4 倍にして 352 枚に拡張したが、この枚数では十分な効果が得られなかったため、まだデータ数が不足していると考えられる。データを増やすことでデータ不足を解消できるが、悪性骨腫瘍患者数が非常に

少ないためデータの追加を期待するのは難しい.

4.4.3 深層学習モデルの選定

本実験では、学習済みモデル VGG-16 と ResNet152 の精度比較を行い、学習済みモデルを選択する。パッチ画像を学習させて得られた学習済みモデルの精度を表 4.8、表 4.9 に示す。各モデルでの学習の経過を図 4.2、図 4.3 に示す。横軸はエポック数、縦軸は評価用データに対する F 値を示している。また、パラメータ比較実験の結果を表 4.10 に示す。

パッチ画像を用いた学習済みモデルとしては、F 値 0.790 の性能を得られた VGG16 が適切であると考えられる。また、ResNet152 と VGG16 とでは、ResNet152 が大量のデータを必要とするのに対し、VGG16 は ResNet152 よりモデルが小さく、データ量を必要としないため、その意味でも VGG16 が適すと考えられる。また、表 4.10 の F 値に注目してみると、データ拡張している場合の方が良い値が得られていることから、データ拡張は有効であると考えられ、パッチサイズも 32×32 より 64×64 の方が適していると考えられる。以上のことから、悪性度を使ってクラス分類を行う実験には、F 値 0.790 が得られた VGG-16 のモデルを使用した。

表 4.8: VGG-16 に対する最良のスコア

	Loss	Accuracy	F1Score
学習用データ	0.0671	0.977	0.976
評価用データ	1.07	0.823	0.790

表 4.9: ResNet152 に対する最良のスコア

	Loss	Accuracy	F1Score
学習用データ	0.137	0.924	0.926
評価用データ	1.31	0.824	0.784

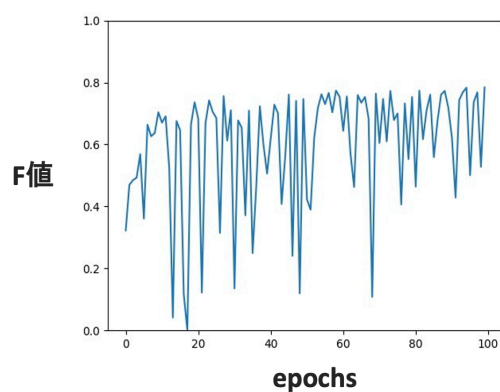


図 4.2: F 値の推移
(ResNet152)

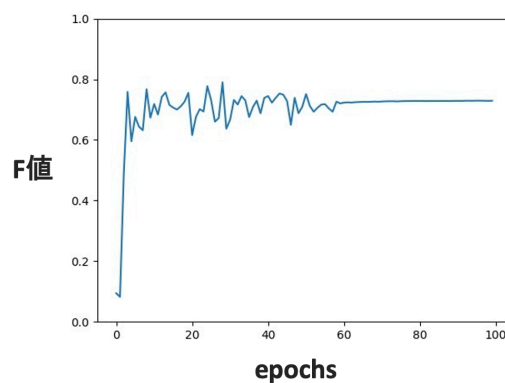


図 4.3: F 値の推移
(VGG16)

表 4.10: パラメータ比較実験

モデル	パッチサイズ	データ拡張	Loss	Accuracy	F1Score
VGG16	32 × 32	×	1.10	0.781	0.714
		○	1.25	0.784	0.722
	64 × 64	×	1.68	0.800	0.758
		○	1.07	0.823	0.790
ResNet152	32 × 32	×	1.80	0.777	0.703
		○	0.800	0.780	0.715
	64 × 64	×	0.975	0.770	0.733
		○	1.31	0.824	0.784

4.4.4 骨腫瘍悪性度推定の実験結果

VGG-16 による交差検証を行った結果を表 4.11 に示す。セグメンテーションの結果で得られた被験者 34 人分の推定マスク画像（黄色の領域のみの画像）と各交差検証で得られた学習済みモデルを用いて、悪性度からクラスを分類する実験と正解マスク画像（緑色と黄色の領域の画像）を用いた実験を行い、結果を比較した。本実験では、悪性度 p_m の閾値として 0.5 を用いる。この値は今回の実験において試験的に使用する。悪性度 p_m が 0.5 以上の腫瘍を悪性腫瘍、0.5 未満の腫瘍を良性腫瘍に分類した。表 4.12 にクラス分類の精度結果を示す。Mask R-CNN が抽出した推定マスクを用いた場合 F 値 0.69 となり、正解マスクを用いた場合の F 値 0.67 を上回った。これは CNN が腫瘍の特徴をより捉えてマスク画像を作成したためであると考ええる。また、悪性検出率である Recall が 0.91 と高く、悪性腫瘍の見逃しが少ない結果となった。図 4.4 に閾値と F 値の推移を表したグラフを示す。横軸は閾値、縦軸はクラス分類精度に対する F 値を示している。図 4.4 より、閾値 0.5 で F 値が一番高いことから、悪性度 p_m の閾値として 0.5 を用いることが適切であることがわかった。

表 4.13 に示すように、骨腫瘍領域を自動抽出した被験者 34 名の内、9 名のクラスを正しく推定することができなかったが、残りの 25 名の被験者のクラスを正しく分類することができた。悪性腫瘍の被験者においては 1 名の被験者を除いては正しくクラスを分類できた。また、正解のマスク画像を用いた実験結果と前述の腫瘍領域を自動抽出した画像を用いた実験結果の精度を比較すると、ほとんど差異がなかった。つまり、Mask R-CNN で推定した腫瘍領域の画像を使用しても、良性、悪性を判別できることがわかった。

表 4.11: VGG-16 による実験結果（交差検証）

	Loss	Accuracy	F 値
データセット 1	1.1	0.82	0.79
データセット 2	0.96	0.65	0.75
データセット 3	0.93	0.75	0.71
データセット 4	1.6	0.64	0.59
平均 ± 標準偏差	1.1±0.27	0.72±0.075	0.71±0.077

表 4.12: クラス分類の精度結果

	Precision	Recall	F 値
正解マスク	0.53	0.91	0.67
Mask R-CNN	0.56	0.91	0.69

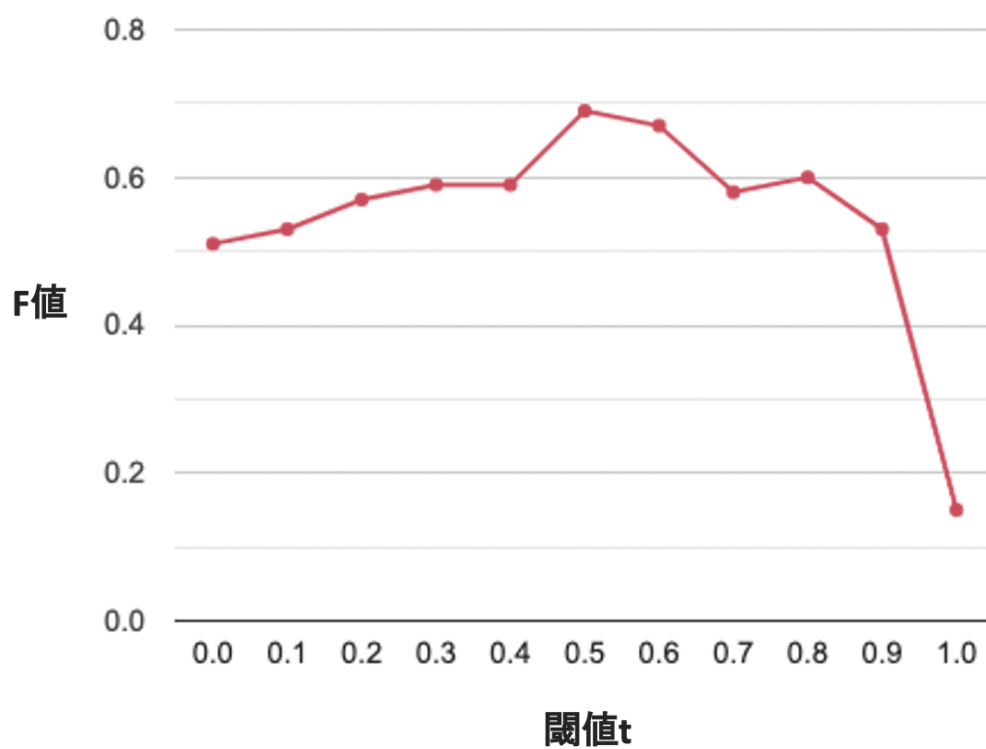


図 4.4: 閾値と F 値の推移

表 4.13: クラス分類の実験結果

正解クラス	被験者	Mask R-CNN		正解マスク	
		悪性度	推定クラス	悪性度	推定クラス
良性	被験者 #1	0.587	悪性	0.574	悪性
良性	被験者 #2	0.00	良性	0.00	良性
良性	被験者 #3	0.0031	良性	0.00216	良性
良性	被験者 #4	0.795	悪性	0.792	悪性
良性	被験者 #5	0.0976	良性	0.0958	良性
良性	被験者 #6	0.641	悪性	0.630	悪性
良性	被験者 #7	0.0409	良性	0.00	良性
良性	被験者 #8	0.472	良性	0.469	良性
良性	被験者 #9	0.795	悪性	0.787	悪性
良性	被験者 #10	0.460	良性	0.459	良性
良性	被験者 #11	0.168	良性	0.185	良性
良性	被験者 #12	0.905	悪性	0.890	悪性
良性	被験者 #13	0.478	良性	0.520	悪性
良性	被験者 #14	0.748	悪性	0.748	悪性
良性	被験者 #15	1.00	悪性	1.00	悪性
良性	被験者 #16	0.164	良性	0.185	良性
良性	被験者 #17	0.910	悪性	0.921	悪性
良性	被験者 #18	0.00	良性	0.00418	良性
良性	被験者 #19	0.202	良性	0.253	良性
良性	被験者 #20	0.401	良性	0.387	良性
良性	被験者 #21	0.409	良性	0.438	良性
良性	被験者 #22	0.178	良性	0.191	良性
良性	被験者 #23	0.00834	良性	0.00883	良性
悪性	被験者 #24	1.00	悪性	1.00	悪性
悪性	被験者 #25	0.997	悪性	1.00	悪性
悪性	被験者 #26	0.957	悪性	0.948	悪性
悪性	被験者 #27	0.996	悪性	1.00	悪性
悪性	被験者 #28	0.883	悪性	0.894	悪性
悪性	被験者 #29	0.639	悪性	0.642	悪性
悪性	被験者 #30	0.0262	良性	0.298	良性
悪性	被験者 #31	0.829	悪性	0.843	悪性
悪性	被験者 #32	0.517	悪性	0.513	悪性
悪性	被験者 #33	0.684	悪性	0.688	悪性
悪性	被験者 #34	0.726	悪性	0.733	悪性

第 5 章

結言

5.1 まとめ

本研究では、深層学習を用いて単純 X 線画像から骨腫瘍の領域を自動でセグメンテーションし、腫瘍の悪性度からクラス分類を行う実験を行った。骨腫瘍の領域を自動で抽出するセグメンテーションの実験では、データセット 1 において IoU が 0.68, Dice 係数が 0.73 の結果であった。予備研究では IoU が 0.84, Dice 係数が 0.92 であるが、手動で腫瘍領域のクロッピングを行う必要があるため、多大な時間を要することに加え、腫瘍領域をクロッピングした画像を評価に使用していたため、実際に医療現場には導入できないシステムであった。今回使用した Mask R-CNN では自動で腫瘍領域をクロッピングした画像を学習し、X 線画像から腫瘍領域を自動抽出することができた。腫瘍領域のクロッピングの時間削減を実現し、先行研究の精度に近づけることができた。

また、腫瘍の悪性度からクラス分類を行う実験では、腫瘍領域をパッチ画像毎に細分化し学習することで、腫瘍の大きさに依存しない推定を実現した。学習済みモデルを得る過程では VGG-16 と ResNet152 を比較した。結果は VGG-16 の F 値が 0.790, 正解率が 0.823 と最も高かった。したがって、本実験では F 値と正解率が最も高い VGG-16 が適していると考え、VGG-16 で学習したモデルを使用した。悪性度からクラス分類を行う実験では、正解マスク F 値 0.67 に対して Mask R-CNN の推定マスク F 値 0.69 と上回った。また、悪性検出率である Recall が 0.91 と高く、悪性腫瘍の見逃しが少ない結果となった。骨腫瘍領域を自動抽出した被験者 34 人の画像のクラスを正しく推定することができなかったが、悪性腫瘍においては 1 名の被験者を除いては正しくクラスを推定することができた。また、正解のマスク画像を用いて実験を行った時の結果とほとんど差異がなかった。つまり、Mask R-CNN で推定した腫瘍領域の画像を使用しても、腫瘍の良悪性を自動分類できることがわかった。

5.2 今後の課題

今後の課題としては、骨腫瘍領域の抽出精度が悪性度の推定精度に影響するため、骨腫瘍領域自動抽出精度の更なる向上を目指す。そのために、学習データを増やし、CNN モデルに学習させる必要があると考える。また、今回の実験では Mask R-CNN モデルを用いたが、その他のインスタンス・セグメンテーションモデルの検討も必要である。

また、良悪性クラス分類の実験においては骨腫瘍領域を自動抽出した推定マスク画像と正解マスク画像を用いた実験のどちらにおいても、全ての被験者のクラスを正しく分類することができなかった。学習済みモデルの更なる精度向上が必要である。そのために、データの増加やその他のモデル・手法の検討を行う予定である。

付録 A

付録

A.1 プログラムの詳細

本研究に関するプログラムはすべて以下のディレクトリ

- /home/furuo/Bone-Tumor/mie-Bonetumor/classification
- /home/furuo/Bone-Tumor/mie-Bonetumor/segmentation

に存在する。ファイルの詳細，実行方法は README に記載する。

```
mie-BoneTumor
|   #研究用ディレクトリ
|
|---classification
|           | #クラス分類実験用ディレクトリ
|           |
|           |
|           |---patch_save.py
|           |   #前処理に関するプログラム
|           |
|           |---tumor
|           |   #データセット関連のディレクトリ
|           |
|           |---learning
|               | #学習に関するディレクトリ
|               |
|               |
```

```

|          |---train.py
|          |      #学習評価用プログラム
|          |
|          |---test.py
|          |      #悪性度を求めるプログラム
|          |
|          |---cnn
|          |      | #train.py で扱う関数に関するディレクトリ
|          |      |
|          |      |---utils.py
|          |      |      #パッチ画像にラベルを付与するプログラム
|          |      |
|          |      |---metrics.py
|          |      |      #F 値などの計算を行うプログラム
|          |
|          |---segmentation
|          |      |      #セグメンテーション実験用ディレクトリ
|          |      |
|          |      |---model
|          |      |      |      #モデル用プログラム
|          |      |      |
|          |      |      |---mask_rcnn.py
|          |      |      |      #Mask R-CNN のセグメンテーションプロ

```

グラム

A.2 データセット

本研究に関するデータはすべて以下のディレクトリ

- /home/furuo/Bone-Tumor/mie-Bonetumor/classfication/tumor
- /home/furuo/Bone-Tumor/mie-Bonetumor/classfication/feature
- /home/furuo/Bone-Tumor/mie-Bonetumor/_data

に存在する。細かいデータの体裁などは README に記載する。

```

mie-BoneTumor
|   #研究用データセット
|
|
|
|---classification
|           |   #クラス分類実験用ディレクトリ
|           |
|           |
|           |---tumor
|           |       | #x 線画像
|           |       |
|           |       |---01
|           |       |       | #患者 1 人 1 人の x 線画像
|           |       |       |
|           |       |       |---0177~
|           |       |       |       | #日付毎の x 線画像
|           |       |       |
|           |       |---mask
|           |       |       | #患者毎のマスク画像
|           |       |       |
|           |       |---data.yml
|           |       |       | #x 線画像とマスク画像を
|           |       |       |       | 対応させまとめた yml ファイル
|           |---feature
|           |       |
|           |       |
|           |       |---train
|           |       |       | #学習用データ
|           |       |       |
|           |       |---valid
|           |       |       | #評価用データ
|
|---_data
|
|

```

```
|---segmentation.yml
|           #セグメンテーション用データをまとめた yml ファイル
|
|---classification.yml
|           #クラス分類用データをまとめた yml ファイル
|
|---list.csv
|           #セグメンテーション用データをまとめた csv フィル
```


謝辞

本研究を進めるにあたり，普段の研究活動から多くの適切な助言，御指導をしていただいた若林哲史教授，白井伸宙助教，お忙しい中ディスカッションに参加して下さり，多くのアドバイスをしてくださった盛田健人助教に深く感謝いたします。また，日々の研究活動において様々な場面でお世話をしていただいた吉永みゆき事務員に深く感謝いたします。そして，ディスカッションや研究についてご指導を頂いた研究室の先輩方や後輩，共に切磋琢磨した同期の皆様に感謝します。最後になりましたが，ここまでの大学生活を支えてくれた家族に今一度感謝の意を表して，本論文の結びといたします。

参考文献

- [1] NT. Do, SD. Joo, HJ. Yang, ST. Jung and SH. Kim, “Knee Bone Tumor Segmentation from radiographs using Seg-Unet with Dice Loss,” 25th International Workshop on Frontiers of Computer Vision (IW-FCV2019), 2019.
- [2] Y. Chen, D. Li, X. Zhanga, J. Jina and Y. Shenb, “Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning,” Medical Image Analysis, Vol. 67, 2021.
- [3] K.Padmapriya, S.Divya, P.Ezhumalai, ”A scrutiny on brain tumour classification using deep learning technique”, materialstoday:proceedings, 2021
- [4] T.Uemura, ”Study on CAD Systems for Detection of Colorectal Cancers and Survival Prediction of Patients with Lung Diseases from CT images based on Deep Learning”, <http://doi.org/10.18997/00008312>, 2021
- [5] N.Bayramoglu, J.Kannala and J. Heikkila, ”Deep learning for magnification independent breast cancer histopathology image classification”, 2016 23rd International Conference on Pattern Recognition, Dec.4-8, 2016
- [6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN” , arXiv preprint arXiv:1703.06870, 2018.
- [7] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2014.
- [8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, 2009.
- [9] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2015.
- [10] 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 (MICCAI), pp. 234–241, 2015.

-
- [11] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam, “Re-thinking atrous convolution for semantic image segmentation,” arXiv preprint arXiv:1706.05587, 2017.
 - [12] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” Proceedings of the European conference on computer vision (ECCV), pp. 801–818, 2018.