

卒業論文

深層学習を用いた耳下腺腫瘍の良悪性診断
における周波数領域でのデータ拡張

9520391k 正木 亮太郎

指導教員 講師 木下 浩二
講師 一色 正晴

2023 年 2 月 9 日

愛媛大学 工学部 工学科 コンピュータ科学コース

目次

第 1 章	序論	1
第 2 章	畳み込みニューラルネットワークを用いた画像分類	3
2.1	緒言	3
2.2	CNN モデル	3
2.3	データ拡張	6
2.4	評価指標	10
2.5	結言	13
第 3 章	周波数領域での拡張手法と評価実験	14
3.1	緒言	14
3.2	周波数領域でのデータ拡張	14
3.3	評価実験	21
3.4	結言	33
第 4 章	結論	34
	謝辞	36
	参考文献	37

第 1 章

序論

耳下腺は耳の前から下に存在する唾液腺であり、唾液を産生、分泌する臓器である。耳下腺から発生する腫瘍は耳下腺腫瘍であり、良性腫瘍と悪性腫瘍があり、悪性の場合は耳下腺の切除を行う必要があるが、耳下腺内部には顔の表情をつかさどる顔面神経が存在しているため、良性の場合は温存を行うことが基本的である [1]。したがって、良悪性の診断を正しく行うことが重要である。現在、畳み込みニューラルネットワーク (Convolutional Neural Network ; CNN) を用いた画像診断が行われているが、CNN は、畳み込み層、プーリング層、全結合層からなり、その訓練には大量のデータセットが必要である。しかし、医療用画像のデータ数は少なく、本論文でもデータ数が良性 300 枚、悪性 200 枚程度と初出の CNN の訓練に十分でない。このようにデータ数が少ないとき、訓練用データのみに対応してしまう過学習に陥り、新しいデータに対する汎化するようなモデルを訓練することができない。こうした問題を解決する方法で考えられるものとして、訓練用データ数を増やすことが挙げられる。そこで、本論文ではデータ拡張に着目し、既存の訓練用データから何かしらの変換を行うことで、もとの画像の性質を保持したまま、元の画像とは異なる新たな画像を生成することで、訓練用データ数を増やした。さらに、データ拡張を行った後、大規模データを用いて学習した VGG16 を用いた学習済みモデルに、畳み込み層の一部、全結合層を腫瘍画像で訓練するファインチューニングを行う。

空間領域でのデータ拡張手法では、明るさ変換、回転、シフト処理など空間領域での拡張を用いているが、その種類や手法には限界があるため本論文では、周波数領域でのデータ拡張を提案する。周波数領域でのデータ拡張は、ハイパスフィルタ、ローパスフィルタ、同一クラスでの振幅特性・位相特性の交換、高周波数成分・低周波数成分の交換、mixup を用いた振幅特性・位相特性の交換、高周波数成分・低周波数成分の交換を行った。これらの手法では、まず画像に対してフーリエ変換を施し、画像を空間領域から空間周波数領域で扱えるようにする。ハイパスフィルタでは低周波数成分の切り取り、ローパスフィルタでは高周波数成分の切り取りを行う。高周波数成分・低周波数成分の交換では、2 枚の画像に対してそれぞれハイパスフィルタ、ローパスフィルタをかけることで、高周波数成分、低周波数成分の抽出を行い、それらの交換し、合成することで新たな画像を生成する [2]。画像に対してフーリエ変換を行ったとき、その画像の振幅特性・位相特性を抽出することができるので、それを 2 枚の画像に対

2 第 1 章 序論

して適用し、それらを交換することで新たな画像を生成する [3]. mixup を用いた手法では、2 枚の画像の割合をランダムで指定し、その割合で、高周波数成分・低周波数成分、振幅特性・位相特性の交換を行う。これらの手法を用いることにより、腫瘍の特徴を残しつつ、新たな画像を生成し、訓練データ数を増加させた。本論文では、拡張しない場合、空間領域での拡張手法、提案する拡張手法で比較実験を行い、どのデータ拡張方法が耳下腺腫瘍の両悪性診断において有効であるのかの検証を行った。

本稿では、第 2 章で、CNN を用いた画像診断の流れ、空間領域でのデータ拡張手法、転移学習について述べる。第 3 章では本論文での提案するデータ拡張手法、データ拡張の手法ごとの評価実験について述べる。第 4 章では結論を述べる。

第 2 章

畳み込みニューラルネットワークを用いた画像分類

2.1 緒言

本章では、始めに、本論文で用いる CNN モデルについて述べる。次に、一般的に使われている空間領域でのデータ拡張手法について述べる。最後に、本論文で用いた評価指標について述べる。

2.2 CNN モデル

2.2.1 CNN とは

CNN は、コンピュータビジョンのアプリケーションにおいてほぼ例外なく用いられるディープラーニングモデルである。これは本論文のような画像分類問題にも用いられる。CNN は、畳み込み層、プーリング層、全結合層からなり、それらの層を重ねることで、画像の特徴を抽出し、それらの特徴から画像の分類を行う。画像を入力すると、まず、畳み込み層において、局所的なパターンを抽出する。次に、プーリング層では、ダウンサンプリングを行い、特徴マップの解像度を減らし、連続する畳み込みが調べるウィンドウを徐々に大きくする。全結合層は、畳み込み層、プーリング層で抽出された特徴量を受け取り、特徴量で分類を行うために利用される。全結合層で出力された結果が、クラス所属確率となる。

2.2.2 ファインチューニングとは

本論文では、データ数が少ないため、学習済みネットワーク VGG16[4] におけるファインチューニングを用いる。VGG16 は、図 2.1 のようになる。

Francois は「学習済みネットワークは、大規模なデータセットで訓練された後、保存されたネットワークであり、一般に画像分類タスクで用いられる。訓練に使用されたデータセットの大きさや汎用性が十分である場合、学習済みネットワークによって特徴量の空間階層は事実上、視覚の世界の汎用モデルになるはずである」と述べている [5]。したがって、学習済みネットワークの特徴量は、腫瘍画像の良悪性判定においても利用できることが考えられる。

しかし、VGG16 は、ImageNet という大規模データセットから作られたモデルであるが、ImageNet は、犬などの動物、ビルなどの建物といった様々な種類のデータから構成されており、そのまま VGG16 に対してそのまま耳下腺腫瘍の画像を入力しても良悪性判定に十分な性能を発揮することが難しいことが考えられる。

そこで、本論文では、VGG16 に対してファインチューニングを適用する。ファインチューニングとは、訓練データを用いて、転移学習に加えてモデルの任意の部分を訓練するものである。本論文では、図 2.1 における block5, classifier を訓練、それより前の層をすべて凍結する。ファインチューニングを行うと、既存のモデルによって学習された表現の一部を腫瘍の良悪性判定に適合させることができ、データ数が少ない場合でも十分な訓練を行うことができる。

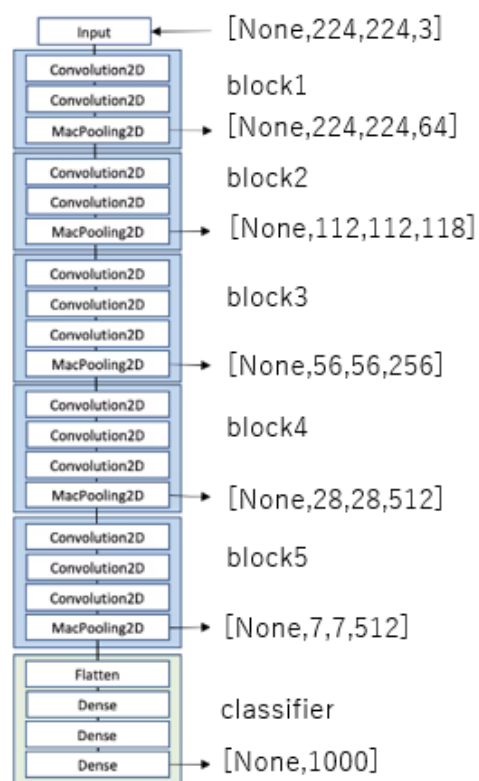


図 2.1: VGG16 の詳細 [6]

2.3 データ拡張

訓練サンプル数が少ない場合は、新しいデータに汎化するようにモデルを訓練することは難しい。もし、データが無限に存在する場合には、モデルはデータの分布をあらゆる角度から学習することになり、過学習に陥ることがなくなるがこれは現実的ではない。そこで、データ拡張で、既存の少ない訓練サンプルからさらに訓練データを生成するというアプローチをとり、同じデータから新たな特徴量を得られるようにする。原画像に対して変換を実行し、元の画像の性質を保存した新たな画像を生成することで、データの水増しを行う。データ拡張では、モデルが全く同じ画像を学習しないこと、元の画像の性質を保存することを目標としている。

2.3.1 空間領域でのデータ拡張

データ拡張の一般的な手法として、空間領域でのデータ拡張を導入する。本論文では、フィルタ処理、明るさ変換、ノイズ付加、シフト処理、せん断処理、回転処理によるデータ拡張を行った。以下、各データ拡張手法について述べる。

フィルタ処理 フィルタ処理を行うために、画像にノイズを付加し、その画像に対してそれぞれのフィルタを適用することで新たな画像を生成することができる。

平均化フィルタ 平均化フィルタでは、画像に対して、すべて同じ値を持つカーネル係数の積和をとることにより、注目画素の周辺画素の平均値をとることができ、その動作を全画素に対して適用する。これにより、図 2.2(b) のように原画像と比べて隣接画素同士の画素値の差が小さくなり、滑らかな画像を生成することができる。

メディアンフィルタ メディアンフィルタでは、画像とカーネルの重なった部分の中央値を変換後画素値とする。それにより、図 2.2(c) のようにエッジ成分を保ちつつノイズ成分を効果的に除去できる。

ガウシアンフィルタ ガウシアンフィルタでは、フィルタの重みをガウス分布関数 (2.1) に従って計算する。平均化フィルタとは異なり、中心部分の注目画素の重みが大きくなり、外側の画素ほど重みが小さくなる。これにより、図 2.2(d) のように画素付近の情報をより残したまま画像をぼかすことが可能になる。

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2.1)$$

バイラテラルフィルタ バイラテラルフィルタは、ガウシアンフィルタを用いてノイズを取り除こうとしたとき、輪郭がぼけるという欠点を解決するために開発されたものである。処理前の画像データの配列を $f(i, j)$ 、処理後の画像データの配列を $g(i, j)$ とすると (2.2) のようになる。ただし、 w はカーネルサイズ、 σ_1 はガウシアンフィルタを制御、 σ_2 は輝度値を制御している。バイラテラルフィルタを適用すると図 2.2 のようになる。

$$g(i, j) = \frac{\sum_{n=-w}^{n=w} \sum_{m=-w}^{m=w} f(i+m, j+n) \exp\left(-\frac{m^2+n^2}{2\sigma_1^2}\right) \exp\left(-\frac{(f(i, j)-f(i+m, j+n))^2}{2\sigma_2^2}\right)}{\sum_{n=-w}^{n=w} \sum_{m=-w}^{m=w} \exp\left(-\frac{m^2+n^2}{2\sigma_1^2}\right) \exp\left(-\frac{(f(i, j)-f(i+m, j+n))^2}{2\sigma_2^2}\right)} \quad (2.2)$$

明るさ変換 画像の位置 (x, y) の画素値を $\text{src}(x, y)$ としたとき、(2.3) で明るさ、図 2.2(f) のようにコントラストの変換を行う。 α 、 β は定数で、 α はコントラスト、 β はバイアスという。本論文では、 $0.75 \leq \alpha \leq 0.9$ 、 $\beta = 0$ としている。

$$\text{dst}(x, y) = \alpha * \text{src}(x, y) + \beta \quad (2.3)$$

ノイズ付加 ノイズ付加では、図 2.2(g) のようにガウシアンノイズを付加する。ガウシアンノイズは、照明が不十分な時に撮影をしたときに発生するものであり、(2.4) のような確率密度分布をとる。

$$F(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2.4)$$

回転処理, シフト処理, セン断処理 回転処理は, 図 2.2(h) のように画像を任意の角度で回転させる処理である. 本論文では, 画像を -10° から -30° の範囲, もしくは 10° から 30° の範囲でランダムに回転させる.

シフト処理は, 図 2.2(i) のように画像を任意の範囲で移動させる処理である. 本論文では, 水平方向, 垂直方向に -10 から 10 ピクセル移動させる.

セン断処理は, 図 2.2(j) のように長方形の画像を平行四辺形に変換する処理である. 本論文では, 画像を -3.5° から -7.5° の範囲, もしくは 3.5° から 7.5° の範囲でランダムにセン断させる.

2.3.2 mixup

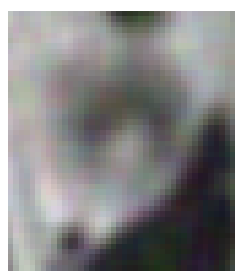
mixup は, データ拡張手法の一つで, 異なるクラスも含むランダムに選んだ 2 つの学習データを組み合わせ, 新たな画像を生成する処理である. (2.5), (2.6) のように画像データ a とラベル b のペア (a_1, b_1) , (a_2, b_2) から, 新たな画像 a と正解ラベル b を作成する.

$$a = \lambda * a_1 + (1 - \lambda) * a_2 \quad (2.5)$$

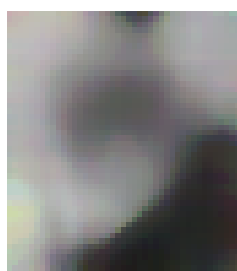
$$b = \lambda * b_1 + (1 - \lambda) * b_2 \quad (2.6)$$



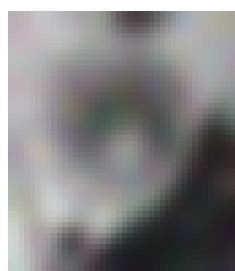
(a) 元画像



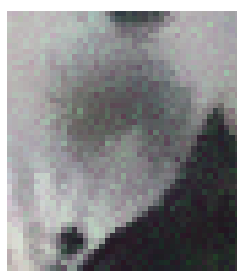
(b) 平均化フィルタ



(c) メディアンフィルタ



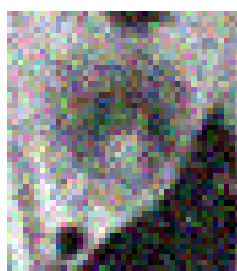
(d) ガウシアンフィルタ



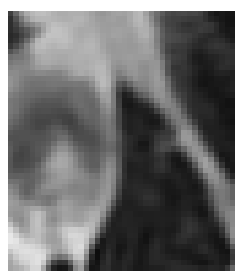
(e) バイラテラルフィルタ



(f) 明るさ変換



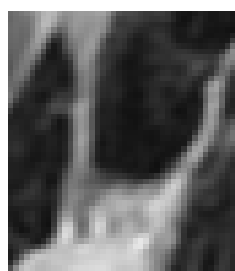
(g) ノイズ付加



(h) 回転処理



(i) シフト処理



(j) セン断処理

図 2.2: 画像一覧

2.4 評価指標

2.4.1 混同行列

混同行列は、クラス分類問題の結果を、「実際のクラス」、「予測したクラス」を軸にしてまとめたものである。2クラス分類においては、実際のクラスと予測したクラスの組み合わせによって、結果を4種類に分けることである。

- 真陽性 (TP : True Positive) : 実際のクラスが陽性で予測も陽性である。
- 真陰性 (TN : True Negative) : 実際のクラスが陰性で予測も陰性である。
- 偽陽性 (FP : False Positive) : 実際のクラスが陰性で予測が陽性である。
- 偽陰性 (FN : False Negative) : 実際のクラスが陽性で予測が陰性である。

これを行列にしたものが混同行列である。混同行列は表 2.1 のようになる。混同行列の値を用いて正解率、適合率、再現率、F1 値を求めることができる。

表 2.1: 混同行列

		予測結果	
		Negative	Positive
実際のクラス	Negative	TN	FP
	Positive	FN	TP

正解率

正解率は、すべてのサンプルのうち、正解したサンプルの割合である。(2.7) のように算出できる。

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.7)$$

適合率

適合率は、陽性と予測されたサンプルうち正解したサンプルの割合である。(2.8) のように算出できる。FP が大きくなると適合率が小さくなるため、偽陽性を避けたい場合に指標として用いる。

$$\text{precision} = \frac{TP}{TP + FP} \quad (2.8)$$

再現率

再現率は、実際のサンプルのうち正解したサンプルの割合である。(2.9) のように算出できる。FN が大きくなると再現率は小さくなる。偽陽性を避けるために指標として用いる。

$$\text{recall} = \frac{TP}{TP + FN} \quad (2.9)$$

F1 値

F1 値は適合率と再現率の調和平均である。(2.10) のように算出できる。

$$\text{F1-measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2 * TP}{2 * TP + FP + FN} \quad (2.10)$$

2.4.2 ROC_AUC_Score

ROC_AUC_Score は, 機械学習の分類問題などの評価指標として用いられ, Roc は, Receiver operating character(受信者操作特性), Auc は, Area under the curve の略で, Roc-Auc を Area under an Roc curve(ROC 曲線下の面積) と呼ぶ.

Roc 曲線は, 正解クラスと予測クラスが与えられたとき, 各予測スコアを閾値としたときの FPR(偽陽性率) を横軸, TPR(真陽性率) を縦軸にプロットしたものである.

FPR, TPR の定義は以下のようになる.

- FPR : False Positive Rate
 - 陰性を誤って陽性と判定した割合
 - $FPR = FP / (FP + TN)$
- TPR : True Positive Rate
 - 陽性を誤って陽性と判定した割合
 - 再現率と同じである.
 - $TPR = TP / (TP + FN)$

これらのことから, ROC_AUC_Score は, 図 2.3 の曲線の下面積が大きいほど性能が良いといえることができる.

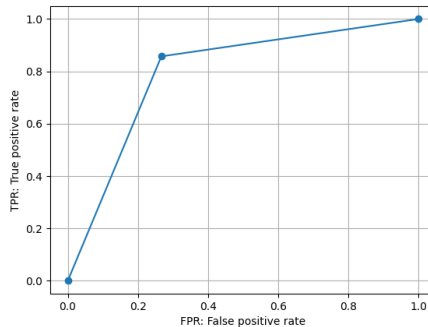


図 2.3: ROC_AUC_Score の例

2.5 結言

本章では，CNN モデルについて，空間領域でデータ拡張の手法，評価指標について述べた．

次章では，提案するデータ拡張手法，評価実験について述べる．提案するデータ拡張手法は，本章のような空間領域でのデータ拡張ではなく，フーリエ変換を行うことによる周波数領域でのデータ拡張を行う．さらに，評価実験における CNN モデルは本章では述べた VGG16，評価指標は正解率，適合率，再現率，F1 値，ROC_AUC_Score を用いている．

第3章

周波数領域での拡張手法と評価実験

3.1 緒言

本章では，周波数領域でのデータ拡張を提案する．周波数領域でのデータ拡張を行うことで，空間領域でのデータ拡張とは異なる性質を持たせることができる．そして，次に，評価実験について述べる．

3.2 周波数領域でのデータ拡張

空間領域におけるデータ拡張手法だけでは，その種類や組み合わせに限界がある．そのため，データ拡張の目的であるあらゆる，角度からデータ分布を学習するというものを達成できない．そこで，本論文では周波数領域でのデータ拡張手法を提案する．空間領域にある画像に対して，フーリエ変換を行うと，ある波形を異なる振幅や周波数，位相を持つ正弦波に分解できる．それにより，周波数領域で扱えるようになり，周波数領域で処理を行うことができ，周波数領域での信号に対して逆フーリエ変換を行うことで，画像が元に戻る．

3.2.1 ローパスフィルタ

ローパスフィルタを用いると、高周波成分を取り除き、低周波成分のみを含む画像を生成することができる。一連の処理を図 3.1 に示す。周波数領域での画像に対して、中心が 1、それ以外が 0 であるローパスフィルタを掛け合わせることで、中心部分のみが残る周波数領域での画像が得られる。中心部分のみの画像に対してフーリエ逆変換を行うことにより、高周波成分が除去された新たな画像が得られている。

実際の画像にローパスフィルタを適用した結果を図 3.2 に示す。中心部分の明るさの急激な変化がある部分は変化がなだらかになっていることがわかる。

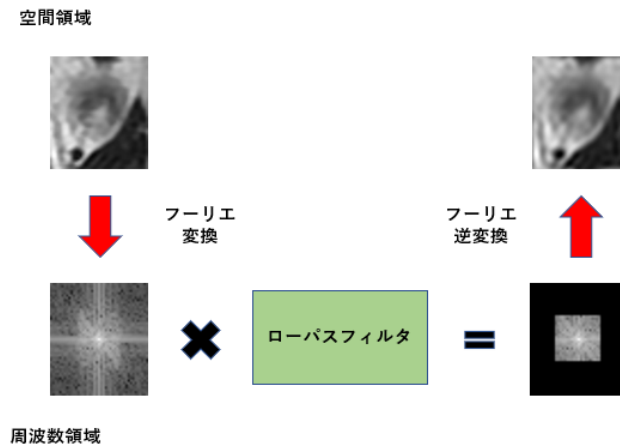


図 3.1: ローパスフィルタの処理の流れ

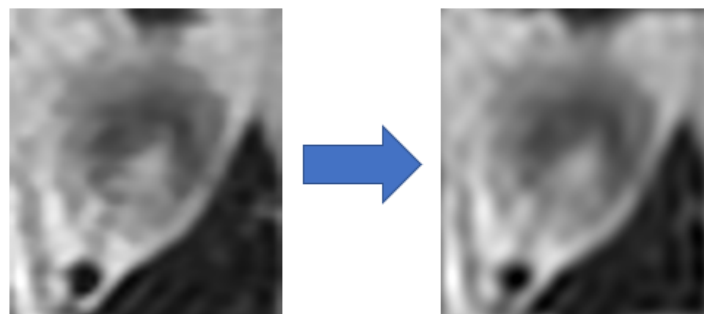


図 3.2: ローパスフィルタの実例

3.2.2 ハイパスフィルタ

ハイパスフィルタを用いると、低周波成分を取り除き、高周波成分のみを含む画像を生成することができる。一連の処理を図 3.3 に示す。原画像に対してフーリエ変換を行うことにより、画像を周波数領域で扱えるようになる。周波数領域での画像に対して、中心が 0、それ以外が 1 であるハイパスフィルタを掛け合わせることで、中心部分が除かれた周波数領域での画像が得られる。中心部分が除かれた画像に対してフーリエ逆変換を行うことにより、低周波成分が除去された新たな画像が得られている。

実際の画像にハイパスフィルタを適用した結果を図 3.4 に示す。明るさの急激な変化のある部分が抽出されていることがわかる。

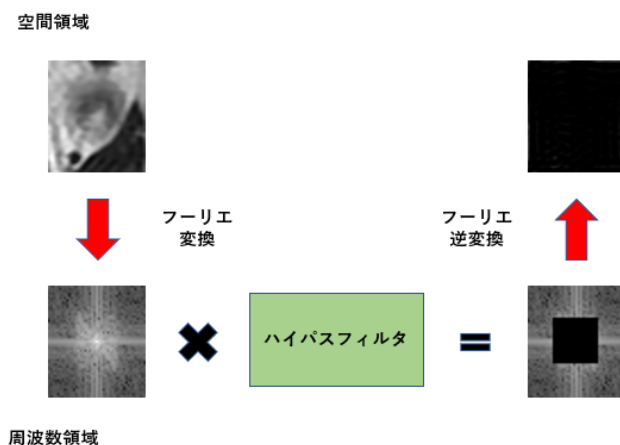


図 3.3: ハイパスフィルタの処理の流れ

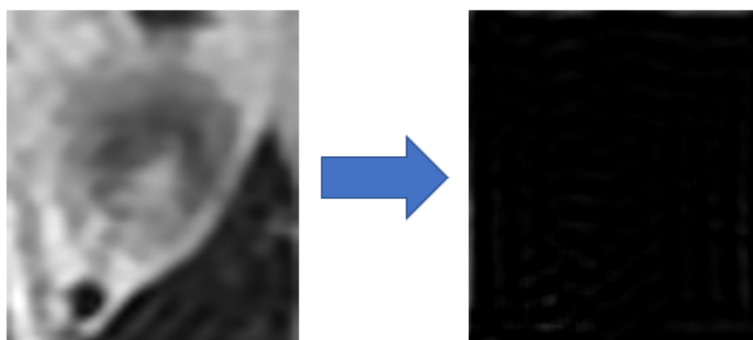


図 3.4: ハイパスフィルタの実例

3.2.3 同一クラス内での振幅特性と位相特性の交換

画像 $f(x, y)$ に対するフーリエ変換は (3.1) で定義される. $F(u, v)$ は, 複素数であるため, 振幅特性と位相特性がそれぞれ (3.2) と (3.3) で定義される.

$$F(u, v) = \int f(x, y) \exp(2\pi j(ux + vy)) dx dy \quad (3.1)$$

$$A(u, v) = |F(u, v)| \quad (3.2)$$

$$\phi(u, v) = \tan^{-1} \frac{\text{Im}F(u, v)}{\text{Re}F(u, v)} \quad (3.3)$$

img1 の振幅特性と位相特性を $A_1(u, v)$, $\phi_1(u, v)$ で, img2 の振幅特性と位相特性を $A_2(u, v)$ と定義する. このとき, 振幅特性 $A_1(u, v)$ と位相特性 $\phi_2(u, v)$ を組み合わせ, 振幅特性 $A_2(u, v)$ と位相特性 $\phi_1(u, v)$ を組み合わせ, 逆フーリエ変換を行うことで新たな画像を 2 枚生成することができる. このとき, 同一クラス内のみでのデータの振幅特性・位相特性を交換することで, データのラベル付けをそのクラスのラベルとすることができる.

実際に 2 枚の画像の振幅特性・位相特性を交換した結果を図 3.5 に示す.

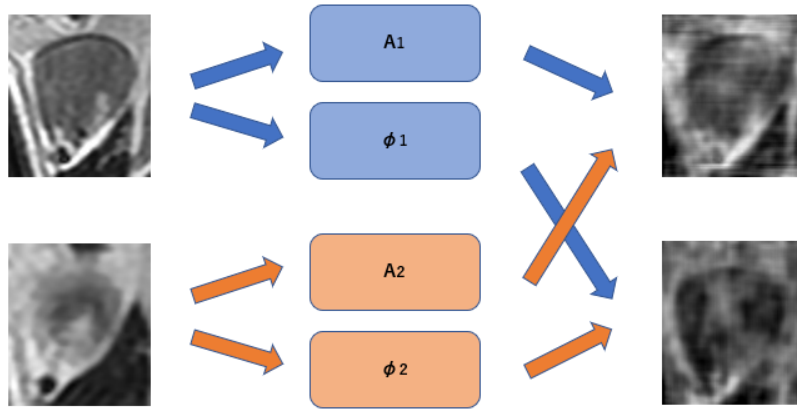


図 3.5: 振幅特性と位相特性の交換の実例

3.2.4 同一クラス内での高周波数成分と低周波数成分の交換

img1, img2 に対して, フーリエ変換を行い, ハイパスフィルタ, ローパスフィルタをかけることで, img1 の高周波数成分と低周波数成分を $high_1$, low_1 で, img2 の高周波数成分と低周波数成分を $high_2$, low_2 とおく. 高周波数成分 $high_1$ と低周波数成分 low_2 を組み合わせ, 高周波数成分 $high_2$ と低周波数成分 low_1 を組み合わせ, 逆フーリエ変換を行うことで新たな画像を生成することができる.

実際に2枚の画像の高周波数成分・低周波数成分を交換した結果を図3.6に示す.

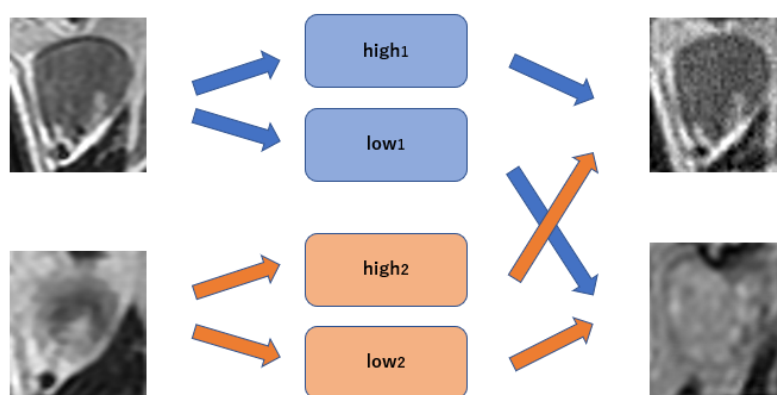


図 3.6: 高周波数成分と低周波数成分の交換の実例

3.2.5 周波数領域における mixup を用いた振幅特性と位相特性の交換

mixup を用いた手法では、3.2.3 における組み合わせの割合を変更させることでクラスの異なるデータを用いた拡張を行うことができる．図 3.7 のように、2 枚の画像に対して、(2.5) と同じく、 $0 \leq \lambda \leq 1$ を設定し、1 枚目の画像の振幅特性、位相特性に λ をかけ、2 枚目の画像の振幅特性、位相特性に $1 - \lambda$ をかけ、それらを組み合わせることで新たな画像を生成する．1 枚目の振幅特性と 2 枚目の位相特性を組み合わせたもの、2 枚目の振幅特性と 1 枚目の位相特性を組み合わせたものを新たな画像とする．クラスは式 2.6 のようにして求めることができる．本論文では、クラスのラベルを閾値より大きい場合に良性、小さい場合を悪性としている．

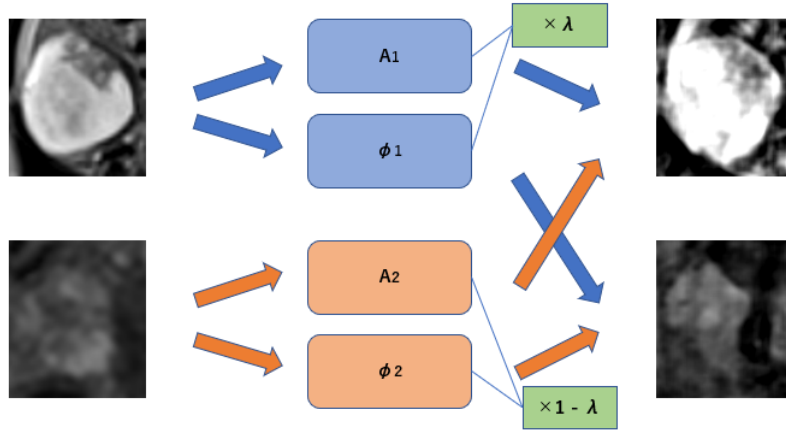


図 3.7: mixup を用いた振幅特性と位相特性の交換の実例

3.2.6 周波数領域における mixup を用いた高周波数成分と低周波数成分の交換

mixup を用いた手法では、3.2.4 における組み合わせの割合を変更させることでクラスの異なるデータを用いた拡張を行うことができる。図 3.8 のように、3.2.5 と同じ手法で高周波数成分と低周波数成分を組み合わせることで、クラスの異なるデータを用いたデータ拡張を行うことができる。クラスの決め方も 3.2.5 と同様である。

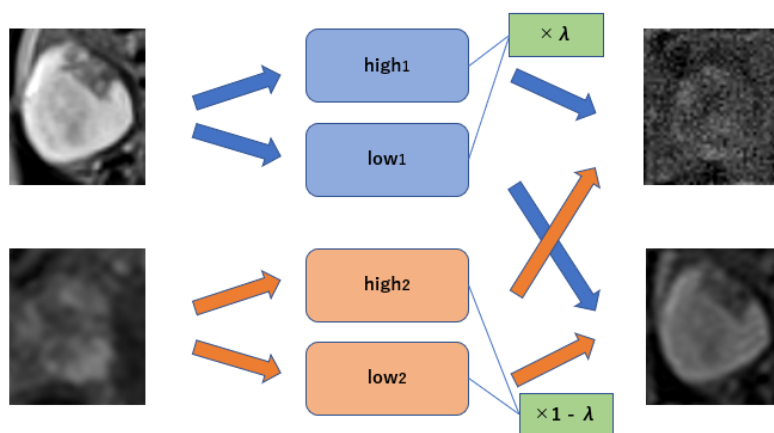


図 3.8: mixup を用いた高周波数成分と低周波数成分の交換の実例

3.3 評価実験

3.3.1 データセット

本論文では, MRIT2 画像を用い, 151 人分 (良性: 88 人, 悪性: 63 人) の MRI 画像を用いて実験を行った. 1 人あたり 1~5 枚のスライス画像があり, 合計 652 枚 (良性: 385 枚, 悪性: 267 枚) である. まず, 151 人分のデータに対して, 患者 ID ごとに訓練用, 検証用データ, テスト用のデータに分割する.

図 3.9 のような MRIT2 画像全体の赤い部分を図 3.10 のように腫瘍部分を切り取ることで, CNN に入力する画像を生成する.



図 3.9: MRIT2 画像



図 3.10: 切り抜き画像

訓練用データでは、スライス画像の枚数をそろえるため、図 3.11 のように、1～4 枚のスライス画像を持つ画像に対して、ガウシアンノイズを標準偏差 20～40 においてランダムに付加することで 5 枚にそろえる。図 3.11 では、上から順にスライス画像の数が 1, 2, 3, 4, 5 枚の場合で、同じ色のものは同じデータである。また、色が同じで A が入っているものは元データに対してランダムにノイズを付加したものである。スライス画像の枚数が 2 枚の場合、1 枚目の画像を 2 回ランダムにノイズを付加し、2 枚目の画像を 1 回ランダムにノイズを付加することで、ノイズを付加していない画像を含めて 5 枚のスライス画像が生成できたことになる。この操作を行わなかった場合、1 枚のスライス画像をもつ患者の画像に対して 6 種類のデータ拡張手法を用いると、6 枚の訓練用データが生成されるのに対して、4 枚のスライス画像をもつ患者の画像に対して 6 種類のデータ拡張手法を用いると、24 枚の訓練用データが生成されることになる。このとき、スライス画像の枚数が少ない患者の腫瘍の特徴が CNN モデルに反映されにくいという問題が発生する。そこで、ノイズの付加により、スライス画像の枚数を 5 枚にそろえることで、すべての患者の特徴を反映させる CNN モデルを作成することができる。その後、各データ拡張手法を用いて訓練用データを作成する。

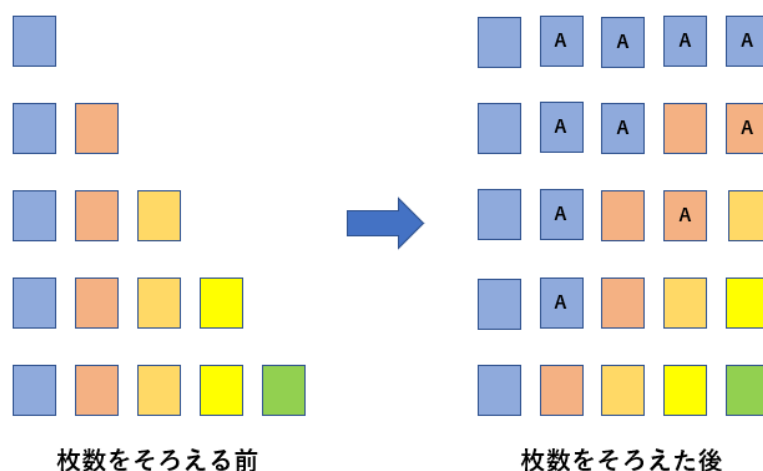


図 3.11: データ枚数をそろえる方法

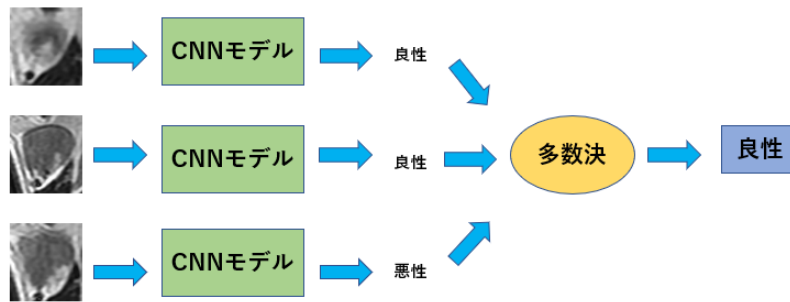
3.3.2 良悪性の判定方法

良悪性の判定は、テスト用データを用いる。テスト用データでは、各患者ごとに1～5枚のスライス画像からなる。そこで、スライス画像1枚ずつ判定を行い、多数決を行うことで、患者一人の判定となる。なお、良性、悪性の判定数が同じ場合は、悪性とする。

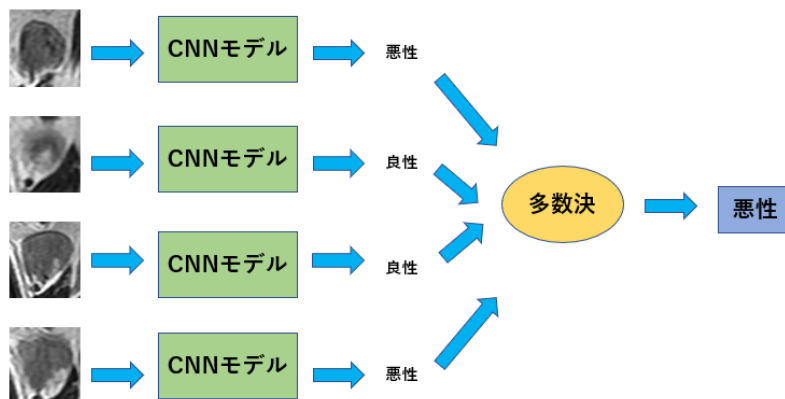
良悪性判定の例は図 3.12(a)、図 3.12(b) のようになる。

図 3.12(a) の場合、ある患者の3枚のスライス画像に対して CNN モデルを用いて判定を行った結果、良性が2枚、悪性が1枚と予測されている。ここでは、良性の枚数のほうが悪性の枚数より多いため、その患者の判定は良性となる。

図 3.12(b) の場合、別の患者の4枚のスライス画像に対して CNN モデルを用いて判定を行った結果、良性が2枚、悪性が2枚と予測されている。ここでは、良性と悪性の枚数が同数であるため、その患者の判定は悪性となる。



(a) CNN を用いた判定の例 1



(b) CNN を用いた判定の例 2

図 3.12: CNN を用いた判定の例

3.3.3 実験設定

データ拡張手法一覧

本論文では、データ拡張の手法を以下の8種類の手法で検証する。

- データセット1：拡張なし
- データセット2：空間領域での拡張
- データセット3：空間領域での拡張 + ローパスフィルタ
- データセット4：空間領域での拡張 + ハイパスフィルタ
- データセット5：空間領域での拡張 + 同一クラスでの振幅特性と位相特性の交換
- データセット6：空間領域での拡張 + 同一クラスでの高周波数成分と低周波数成分の交換
- データセット7：
空間領域での拡張 + 周波数領域における mixup を用いた振幅特性と位相特性の交換
- データセット8：
空間領域での拡張 + 周波数領域における mixup を用いた高周波数成分と低周波数成分の交換

ここで、3.3.3 ~ 3.3.3 の手法において、空間領域での拡張を用いている理由は、ハイパスフィルタ、ローパスフィルタを用いたデータ拡張で、3.2.3 ~ 3.2.6 の拡張手法と比べて、データ拡張により増やせるデータ数が少ない。そこで、空間領域での拡張に加えてほぼ同じ数だけ周波数領域でのデータ拡張を追加することで、データ数が十分である場合の CNN モデルを訓練できる。

実験データ

実験 1 実験 1 における訓練データ数は表 3.1 のようになる．データセットの番号は，データ拡張手法におけるデータ拡張手法の番号に対応しており，空間領域での拡張手法はそれぞれ一定のノイズ付加やシフトなどを 1 回行う．また，ハイパスフィルタ，ローパスフィルタのサイズを，中心から 10px の領域としてそれぞれのデータ拡張手法を 1 回ずつ用いた場合である．同一クラス内，mixup を用いた振幅特性と位相特性の交換，高周波数成分と低周波数成分の交換では，それぞれ 2000 回データ拡張を行う．

表 3.1: 実験 1 の訓練用データ

	訓練用データ		検証用データ		テスト用データ	
	良性	悪性	良性	悪性	良性	悪性
データセット 1	310 枚	200 枚	50 枚	56 枚	64 枚	30 枚
データセット 2	3410 枚	2310 枚	50 枚	56 枚	64 枚	30 枚
データセット 3	4030 枚	2730 枚	50 枚	56 枚	64 枚	30 枚
データセット 4	3720 枚	2520 枚	50 枚	56 枚	64 枚	30 枚
データセット 5	4650 枚	2730 枚	50 枚	56 枚	64 枚	30 枚
データセット 6	4390 枚	2634 枚	50 枚	56 枚	64 枚	30 枚
データセット 7	4772 枚	3794 枚	50 枚	56 枚	64 枚	30 枚
データセット 8	4765 枚	4965 枚	50 枚	56 枚	64 枚	30 枚

実験2 実験2における訓練用データのデータ数は表3.2のようになる。データセットの番号は、データ拡張手法におけるデータ拡張手法の番号に対応しており、空間領域での拡張手法は、それぞれ指定した範囲でランダムに指定した値で5回ノイズ付加やシフトなどを行っている。ハイパスフィルタ、ローパスフィルタのサイズを、中心から3～15pxのランダムな領域としてそれぞれのデータ拡張手法を5回ずつ用いた場合である。同一クラス内、mixupを用いた振幅特性と位相特性の交換、高周波数成分と低周波数成分の交換では、同一クラス内での画像を用いて10000回データ拡張を行う。

表3.1、表3.2のようにデータ拡張を行うと、各種法ごとの訓練用データの枚数が異なるため、同じ条件での比較が難しくなる。そこで、実験1の場合に、データ数を良性・悪性ともに2000までダウンサンプリング、実験2の場合に、データ数を良性・悪性ともに10000までダウンサンプリングする。これにより、各種法ごとの訓練用データの枚数を揃えることができ、さらに、良性・悪性の枚数が同じであるため、良性・悪性の判定する能力に差が出にくくなることが期待できる。

表3.2: 実験2 訓練用のデータ

	訓練用データ		検証用データ		テスト用データ	
	良性	悪性	良性	悪性	良性	悪性
データセット1	310枚	200枚	50枚	56枚	64枚	30枚
データセット2	16764枚	11550枚	50枚	56枚	64枚	30枚
データセット3	18314枚	12600枚	50枚	56枚	64枚	30枚
データセット4	18314枚	12600枚	50枚	56枚	64枚	30枚
データセット5	18664枚	12150枚	50枚	56枚	64枚	30枚
データセット6	18664枚	12150枚	50枚	56枚	64枚	30枚
データセット7	22535枚	17643枚	50枚	56枚	64枚	30枚
データセット8	24607枚	19396枚	50枚	56枚	64枚	30枚

パラメータ

各実験におけるデータのパラメータは表 3.3 のようになる。

表 3.3: 訓練におけるパラメータ

パラメータ	詳細
損失関数	binary_cross_entropy
最後の層の活性化関数	sigmoid
最適化関数	RMSProp
学習率	10^{-5}
評価指標	acc
最大エポック数	100
訓練ステップ数	25
検証ステップ数	5
保存するモデル	検証用データのスコアの最も高いもの
評価指標	正解率, 適合率, 再現率, F1 値, Roc_Auc_Score

2 クラス分類であるため、損失関数を `binary_cross_entropy`、最後の層の活性化関数を `sigmoid` を用いる。また、エポックを回して、検証用データのスコアが最も高いものを保存することで、未知のデータに対して一番対応できるモデルを保存している。さらに、全結合層は、Flatten、活性化関数 ReLU、0.5 のドロップアウト、活性化関数 `sigmoid` を用いたものである。

モデルの評価には、2.4 で述べた正解率、適合率、再現率、F1 値、Roc_Auc_Score を用いる。

評価の際、表 2.1 の Negative が良性、Positive が悪性となるように設定している。これにより、適合率は、悪性と予測されたサンプルのうち、実際に悪性である割合を示す。また、再現率は、実際に悪性のサンプルのうち、正解した割合が求められる。

3.3.4 実験結果

実験 1

良性、悪性のデータ数 2000 枚の場合の実験結果は表 3.4 のようになる。結果より、ローパスフィルタ (データセット 3)、ハイパスフィルタ (データセット 4)、mixup を用いた振幅特性と位相特性の交換 (データセット 7) を用いた手法が有効であることが分かった。ハイパスフィルタ、ローパスフィルタでは、同じ画像の拡張を固定の閾値で 1 回行うため、この場合には新たな性質を持たせられていることが考えられる。

周波数領域における mixup を用いた振幅特性と位相特性の交換では、2 枚の画像の振幅特性と位相特性を任意の割合で混合するが、その時にクラスの性質を十分に保存できることが考えられる。逆に、周波数領域における mixup を用いた高周波数成分と低周波数成分の交換では、拡張しない場合や空間領域での拡張手法のみの場合よりもかなり低い結果となった。これは、ランダムに良悪性を判定した場合である正解率 0.5 とほとんど同じ結果であり高周波数成分と低周波数成分の交換に mixup を用いることで画像の性質が変わってしまったため、実質ランダムな判定器が生成されたが考えられる。

表 3.4: 実験 1 における結果一覧

	正解率	適合率	再現率	F1 値	Roc_Auc_Score
データセット 1	0.77	0.77	0.93	0.85	0.68
データセット 2	0.77	0.92	0.73	0.81	0.80
データセット 3	0.86	0.83	1.00	0.91	0.79
データセット 4	0.86	0.88	0.93	0.90	0.82
データセット 5	0.68	0.70	0.93	0.80	0.54
データセット 6	0.64	0.73	0.73	0.73	0.58
データセット 7	0.86	1.0	0.8	0.89	0.9
データセット 8	0.55	0.86	0.4	0.55	0.63

実験 2

良性，悪性のデータ数 10000 枚の場合の実験結果は表 3.5 のようになる．結果より，同一クラス内での振幅特性，位相特性の交換（データセット 5），高周波数成分と低周波数成分の交換（データセット 6）が有効であることが分かった．これらの手法では，同じクラス内でデータを交換するため，クラスの性質を保存したまま新たな特徴を持たせることができたため，データ数が増えた際により多くの特徴を持たせることができ，CNN モデルの品質が高くなったことが考えられる．また，ローパスフィルタ（データセット 3）を用いた場合が空間領域での拡張（データセット 2）を用いた場合とほとんど同じ結果となった．この結果から，ローパスフィルタにより拡張した場合のデータセットが空間領域での拡張のみを行ったときのデータセットとほとんど同じであり，新たな性質を持たせられなかったことが考えられる．

表 3.5: 実験 2 における結果一覧

	正解率	適合率	再現率	F1 値	Roc_Auc_Score
データセット 1	0.77	0.77	0.93	0.85	0.68
データセット 2	0.68	0.90	0.60	0.72	0.73
データセット 3	0.73	1.00	0.60	0.75	0.80
データセット 4	0.82	1.00	0.73	0.85	0.87
データセット 5	0.91	0.93	0.93	0.93	0.90
データセット 6	0.91	0.93	0.93	0.93	0.90
データセット 7	0.82	1.00	0.8	0.86	0.83
データセット 8	0.82	1.00	0.73	0.85	0.87

3.3.5 考察

実験 1, 実験 2 より, 評価指標ごとにスコアの一覧は, 図 3.13 ~ 図 3.17 のようになる. データセットの後の () の数字は, 実験 1 のデータセットを (1), 実験 2 のデータセットを (2) とする. これらの図において, スコアの最も高いものは赤色, 2 番目に高いものは橙色, その他のものは青色としている.

これらの結果から, すべての指標において 2 番目以上スコアにデータ数 10000 における同一クラス内での高周波数成分と低周波数成分の交換, 振幅特性と位相特性の交換が入っていることが分かる.

さらに, ローパスフィルタを用いた場合であるのデータセット 3(1), データセット 3(2) の結果から, データ数が多いと CNN モデルの性能が良いとは限らないことがわかる. 例えば, データセット 3(1), データセット 3(2) では, 拡張手法は同じローパスフィルタであるが, 拡張を行う回数が異なるため, データ数が 2000, 10000 と異なっている. しかし, データセット 3(1) を用いた場合の適合率以外の指標がデータセット 3(2) を用いた場合よりも高くなっている. これは, データ拡張を行うことにより, 同じようなデータ, もしくは全く同じデータが生成されたため, データ拡張の目的であるデータの特徴を保存したまま新たな性質をもつ画像の生成ができなかったことが考えられる. ローパスフィルタを用いた際, データセット (1) は中心から 10px のローパスフィルタを用いているが, データセット (2) は中心から 3~15px のランダムな値でローパスフィルタを 5 回用いている. このため, ランダムな値が似たように設定されると, 新たな性質を学習できない.

また, データセット 7(1) とデータセット 7(2), データセット 8(1) とデータセット 8(2) の結果から, 周波数領域での mixup を用いた振幅特性と位相特性の交換, 高周波数成分と低周波数成分の交換を行った場合, すべての指標においてデータセット 8(2) の結果が高くなっていることが分かる. このことから, mixup を用いた場合, データ数が多くなった場合でも新たな性質を学習でき, これらの手法はデータ数をより多くしたときに有効なものであることが考えられる.

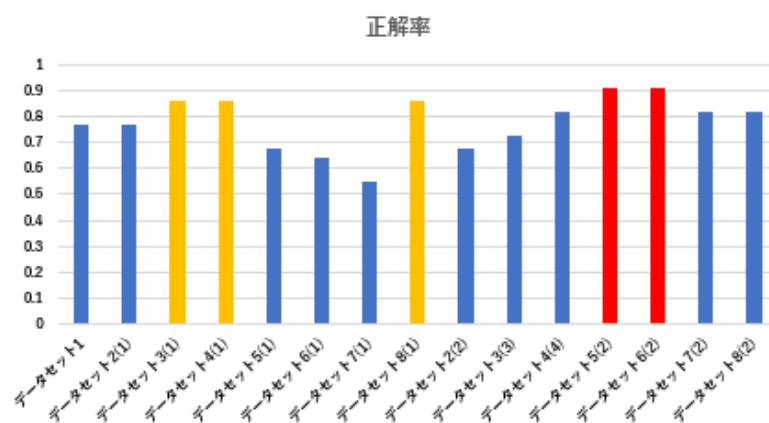


図 3.13: 正解率一覧

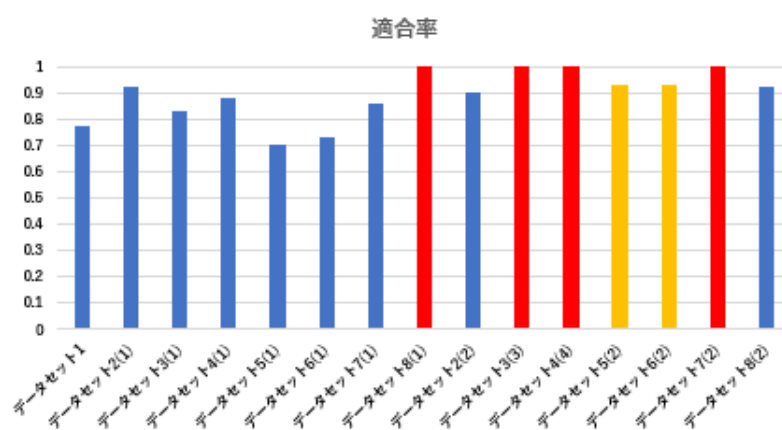


図 3.14: 適合率一覧

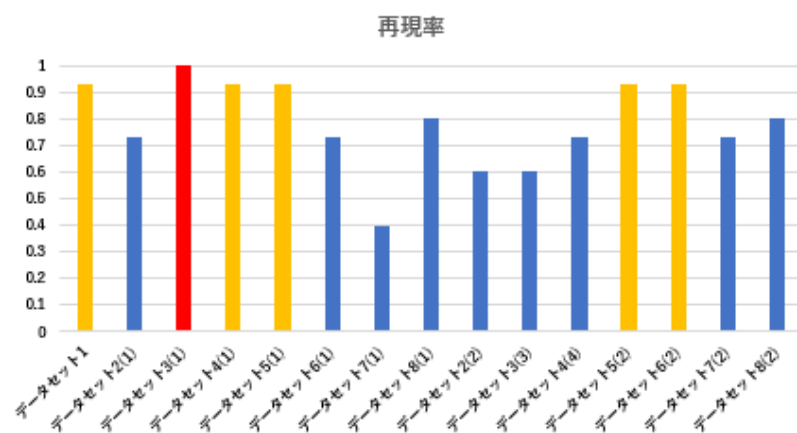


図 3.15: 再現率一覧

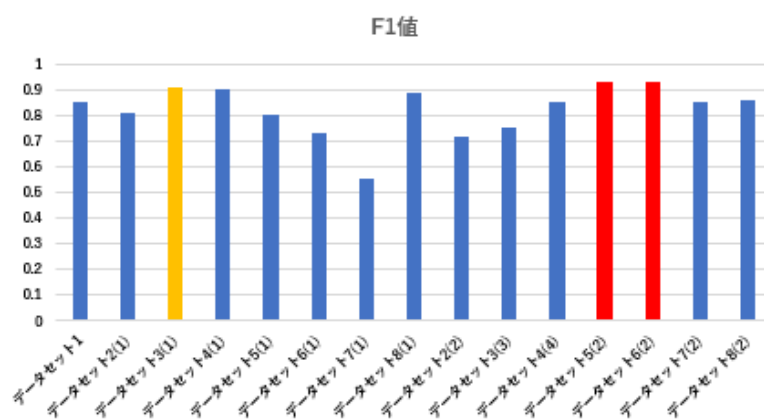


図 3.16: F1 値一覧

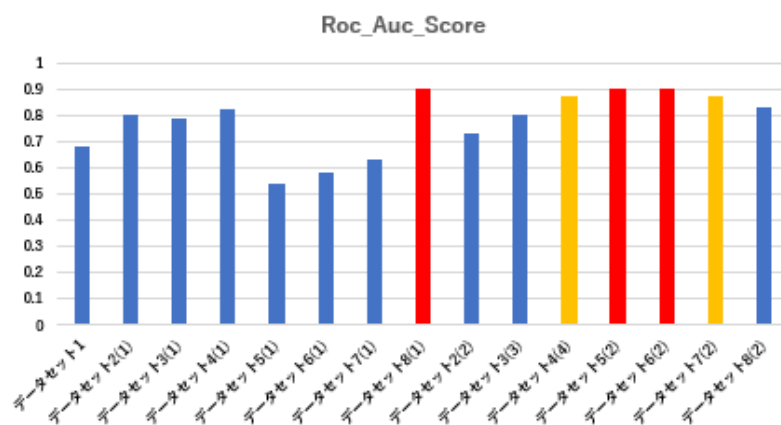


図 3.17: Roc_Auc_Score 一覧

3.4 結言

本章では、データ拡張を行わない場合、空間領域での拡張手法、提案する周波数領域でのデータ拡張を行った。図 3.13～図 3.17 から分かるようにデータ拡張手法ごとに結果が異なることがわかった。さらに、データ拡張によるデータ数と CNN モデルの性能に相関関係がほとんどないことがわかった。

そして、本論文での提案手法はデータ数 10000 における同一クラス内での高周波数成分と低周波数成分の交換、振幅特性と位相特性とする。

第 4 章

結論

耳下腺腫瘍の良悪性判定の CNN モデルの訓練には大量のデータ数が必要である。しかし、与えられているデータセットはそれだけでは CNN モデルの十分な訓練を行うことができず、データ拡張を行うことで、データ数を増加させ、十分な性能を持つ CNN モデルの訓練を行うことができる。そこで、本論文では、どのデータ拡張手法が CNN モデルの性能の向上に良い影響を与えるのかを検討した。

一般的に、データ拡張は元画像に対して回転、明るさ変換などの空間領域での処理を施すことによって行っていた。しかし、空間領域のみのデータ拡張では、データ拡張の手法に限りがある。そこで、本論文では周波数領域でのデータ拡張手法を提案する。周波数領域でのデータ拡張手法では、元画像に対してフーリエ変換を行い、その信号にローパスフィルタなどの様々な処理を行い、逆フーリエ変換を行うことで再び画像に戻す。さらに、一般的に空間領域において行われていた mixup を周波数領域において行うことで、新たな画像を生成した。提案手法は、ローパスフィルタ、ハイパスフィルタ、同一クラス内での振幅特性と位相特性の交換、同一クラス内での高周波数成分と低周波数成分の交換、mixup を用いた振幅特性と位相特性の交換、mixup を用いた高周波数成分と低周波数成分の交換として、データ拡張をしない場合、空間領域のデータ拡張によるデータをデータ数などを変更しながら訓練データに用いた CNN モデルの比較実験を行った。比較結果より、データ数 10000 における同一クラス内での振幅特性と位相特性の交換、高周波数成分と低周波数成分の交換が CNN モデルの性能の向上に有効であるということがわかった。さらに、データ数を増やしたにもかかわらず、CNN モデルの性能が低くなることがあることが実験よりわかった。これは、ノイズ付加、ハイパスフィルタのサイズなどの手法をランダムな量で行っていること、データのダウンサンプリングによる必要なデータの削除が原因であると考えられる。

今後の検討として以下の課題が挙げられる。

- 全結合層が簡単なものであるため，耳下腺腫瘍の診断のために最適なものへの変換
- データ数 10000 のデータセットを作成するのに，ノイズ付加，ハイパスフィルタのサイズなどの手法をランダムな量で行っているが，それを指定した範囲でのランダムな量の設定に変えることで，拡張後の画像の類似性をなくす
- ダウンサンプリングをランダムではなく，指標の悪いものを削除することで，ダウンサンプリング手法による評価に与える影響を抑える
- データ枚数をそろえる際，ノイズを加えるだけでなく，拡大，縮小を加えることで，新たに生成した画像がスライス画像のような性質をもたせることができる。

これらの点を改善することで，CNN モデルの性能をより良くすること，さらに，データ拡張手法による影響をより正確に表すことができると考えられる。

謝辞

本研究を行うにあたり，懇切なご指導並びにご教示いただきました木下 浩二 講師に深く感謝の意を表します。

また，愛媛大学 耳鼻咽喉科・頭頸部外科の三谷 壮平 講師，ならびに愛媛大学 社会連携推進機構の秋丸 國廣 准教授には，論文データの提供や情報共有にご尽力いただきました。深く感謝申し上げます。

また，本論文に際し，ご審査いただいた 木下 浩二 講師，二宮 崇 教授，安藤 和典 准教授に深くお礼申し上げます。

さらに，本論文に関して，有益なご指導頂きました一色 正晴 講師に深く感謝の意を表します。

最後に，様々な助言，支援を下さった本分野の院生，学部生の方々に感謝します。

参考文献

- [1] 一般社団法人 日本形成外科学会：耳下腺腫瘍，https://jsprs.or.jp/general/disease/shuyo/hifu_hika/jikasen.html，(2023 年 1 月閲覧)。
- [2] 向井 皇喜，熊野 創一郎，山崎 俊彦：画像の周波数成分を考慮したデータ拡張による CNN の分布外データへの頑健性向上，情報処理学会第 84 回全国大会講演論文集，pp. 225–226，2022.
- [3] 新田 大悟，庄野 逸：テクスチャ画像識別問題に対するフーリエ変換を用いたデータ拡張の検討，電子情報通信学会技術研究報告，Vol. 118, No. 470, NC2018-87, pp. 233-238, 2019.
- [4] Karen Simonyan and Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition, In 3rd International Conference on Learning Representations, 2015.
- [5] François Chollet, 株式会社クイープ訳，巢籠 悠輔監訳：Python と Keras によるディープラーニング，マイナビ出版，2018.
- [6] cyumizou：Keras で VGG16 のファインチューニングを試してみる，<https://qiita.com/ps010/items/dee9413d3de28de7d2f9>，(2023 年 1 月閲覧)。