

令和3年度
修士学位論文

GANを用いた CNNの分類結果の説明性に関する研究

A Study on Explainability of Convolutional Neural
Networks Using Generative Adversarial Networks

1245127 筒井 康行

指導教員 吉田 真一

2022年2月28日

高知工科大学大学院 工学研究科 基盤工学専攻
情報学コース

要旨

GAN を用いた CNN の分類結果の説明性に関する研究

筒井 康行

近年 Computer-Aided Diagnosis(CAD) は重要な研究分野の 1 つとなっており, Convolutional Neural Network(CNN) を用いた画像認識による診断が広く使われる. しかし, CNN を用いた診断では診断過程の説明が困難という課題があり, 患者への説明責任が果たせないため, 現場への導入が進んでいないのが現状である. これまでに CNN の分類過程の説明には学習したモデルから得られる特徴マップやパラメータの重み, 勾配を用いて CNN の分類に寄与する領域を特定する Class Activation Map(CAM) ベースの手法が用いられてきた. しかし結果の説明には分類に寄与する領域だけでなく, 領域内の形状やパターンの違いも知る必要があると考えられ, 分類に寄与する領域と領域内の形状やパターンの違いの両方が得られる手法が必要になると考えられる.

そこで本研究では, 対象のデータセットによく似たデータを生成できるようなデータ分布を学習からの獲得が期待できる Generative Adversarial Network(GAN) を用いて, 分類に寄与する領域と領域内の形状やパターンの違いを取得する手法を提案する. 本研究では GAN のモデルとして Attention-Guided CycleGAN(AG-CycleGAN) を用い, AG-CycleGAN から得られる注意マップから分類に寄与する領域, 相互変換の結果から領域内の形状やパターンの違いの獲得を目指す. また, 変換の妥当性や説明可能性における注意マップの信頼性, 有効性を評価するために, AG-CycleGAN による相互変換結果の妥当性, Grad-CAM++と注意マップの比較, 学習済み CNN における変換前後の識別結果の変化から変換の有効性を評価し, CNN の説明における手法の有効性の検証を行う.

胸部 X 線画像データセットにおける心肥大-検出なし間，気胸-検出なし間，脳構造画像データセットにおける男女間で検証を行った結果，心肥大-検出なし間では症状に沿った分類に寄与する領域と領域内の形状やパターンの違いが得られた．しかし，気胸-検出なし間では説明性が低い結果，男女間では有効な領域，違いが得られなかった．そのため，今後は学習内で変換の抑制をしない損失関数の検討やより説明性の高い結果の表現方法の検討などについて検討していく必要があると考えられる．

キーワード 敵対的生成ネットワーク，畳み込みニューラルネットワーク，CycleGAN，説明可能 AI，医療情報システム

Abstract

A Study on Explainability of Convolutional Neural Networks Using Generative Adversarial Networks

Yasuyuki Tsutsui

In recent years, computer-aided-diagnosis (CAD) has become an important research field, and diagnosis by image recognition using convolutional neural network (CNN) is widely used. However, it is difficult to explain the diagnosis process using CNNs, and the introduction of CNNs into the field has not progressed because of the lack of accountability to patients. Class activation map (CAM)-based methods have been used to explain the classification process of CNNs by using feature maps, parameter weights, and gradients obtained from trained models to identify regions that contribute to CNN classification result. However, in order to explain the results, it is necessary to know not only the regions that contribute to classification but also the differences in shapes and patterns within the regions, and a method that can obtain both the regions that contribute to classification and the differences in shapes and patterns within the regions is required.

In this study, we proposed a method to obtain the regions that contribute to classification and the differences in shapes and patterns within the regions using generative adversarial network (GAN), which is expected to acquire data distributions that can generate data similar to the target dataset from learning. In this study, we used attention-guided CycleGAN (AG-CycleGAN) as a model of GAN, and aimed to obtain the regions that contribute to classification from the attention map obtained from

AG-CycleGAN, and the differences in shapes and patterns within the regions from the results of mutual transformation. In order to evaluate the reliability and effectiveness of the attention map in terms of the validity and explainability of the transformation, we also evaluate the effectiveness of the transformation based on the validity of the mutual transformation results from AG-CycleGAN, the comparison between Grad-CAM++ and the attention map, and the change in the discrimination results before and after the transformation in the trained CNN.

The results of validation between cardiomegaly-undetected in the chest X-ray dataset, pneumothorax-undetected, and between men and women in the structural brain imaging dataset showed that the cardiomegaly-undetected dataset showed regions that contributed to classification along the symptoms and differences in shape and pattern within the regions. However, the results between pneumothorax and no detection were less explanatory, and no valid regions or differences were obtained between men and women. Therefore, it is necessary to study the loss function that does not suppress the transformation in learning and to study how to express the results with higher explainability in the future.

key words Generative adversarial network(GAN), convolution neural network(CNN), CycleGAN, explainable artificial intelligence(XAI), medical information system

目次

第 1 章	序論	1
第 2 章	関連研究	3
2.1	Generative Adversarial Networks(GAN) モデル	3
2.1.1	Generative Adversarial Networks(GAN)	3
2.1.2	CycleGAN	4
2.1.3	Attention-Guided CycleGAN(AG-CycleGAN)	5
2.2	Class Activation Mapping(CAM) ベースの手法	6
2.2.1	GradCAM++	8
2.3	CNN モデル構造	10
2.3.1	ResNet	10
2.3.2	U-Net	10
2.4	GAN の学習の安定化	11
2.4.1	スペクトル正規化 (Spectral Normalization)	12
2.5	ハイパーパラメータ探索	13
2.5.1	グリッドサーチ (Grid Search)	13
2.5.2	LR range test	13
第 3 章	提案手法	14
第 4 章	実験内容	15
4.1	データセット	15
4.1.1	胸部 X 線画像 NIHCC データセット	15
4.1.2	脳構造画像 IXI データセット	16
4.2	モデル構造	18

目次

4.2.1	CNN モデル (診断モデル・男女診断モデル)	18
4.2.2	生成モデル・注意モデルの構造	19
4.2.3	識別モデル	19
4.3	対象となる CNN モデル (診断モデル, 男女分類モデル)	20
4.3.1	診断モデル (NIHCC データセット: 心肥大-検出なし)	21
4.3.2	診断モデル (NIHCC データセット: 気胸-検出なし)	22
4.3.3	男女分類モデル (IXI データセット: 男性-女性)	23
4.4	実験手順	24
第 5 章	実験結果	26
5.1	AG-CycleGAN による相互変換の学習	26
5.1.1	NIHCC データセット: 心肥大-検出なし	26
5.1.2	NIHCC データセット: 気胸-検出なし	27
5.1.3	IXI データセット: 男性-女性	28
5.2	GradCAM++ と注意マップの比較	29
5.2.1	NIHCC データセット: 心肥大-検出なし	30
5.2.2	NIHCC データセット: 気胸-検出なし	30
5.2.3	IXI データセット: 男性-女性	31
5.3	変換前後での CNN の分類結果の違い	33
5.3.1	NIHCC データセット: 心肥大-検出なし	33
5.3.2	NIHCC データセット: 気胸-検出なし	33
5.3.3	IXI データセット: 男性-女性	34
第 6 章	追加実験	35
6.1	追加実験内容	35
6.1.1	データセット	35
6.1.2	識別モデル	35

目次

6.1.3	対象となる CNN	35
6.2	実験手順	37
6.3	追加実験結果	37
6.3.1	AG-CycleGAN による相互変換の学習	37
6.3.2	GradCAM++と注意マップの比較	39
6.3.3	変換前後での CNN の分類結果の違い	39
第 7 章	考察	41
7.1	AG-CycleGAN による相互変換	41
7.1.1	NIHCC データセット: 心肥大-検出なし	41
7.1.2	NIHCC データセット: 気胸-検出なし	41
7.1.3	IXI データセット: 男性-女性	43
7.1.4	IXI データセット: 男性-女性 追加実験	43
7.2	AG-CycleGAN の学習	44
7.3	注意マップと Grad-CAM++の比較	46
7.4	変換前後での CNN の分類結果の違い	47
第 8 章	結論	49
	謝辞	50
	参考文献	51
付録 A	ハイパーパラメータ探索の結果	54

目次

2.1	GAN 上でのデータ遷移	4
2.2	CycleGAN 上でのデータ遷移	5
2.3	Attention-Guided CycleGAN(AG-CycleGAN) 上でのデータ遷移	7
2.4	AG-CycleGAN で用いる変換モジュールの構造	7
2.5	Resnet の残差ブロック	10
2.6	Unet の構造	11
4.1	NIHCC データセット内の各ラベルのデータ分布	17
4.2	Resnet の図示に用いるモジュール	18
4.3	CNN モデルの構造 (Resnet18	19
4.4	生成モデル・注意モデルの構造 (U-net)	20
4.5	識別モデルの構造	20
5.1	心肥大-検出なし間の相互変換の結果	27
5.2	気胸-検出なし間の相互変換の結果	28
5.3	男性-女性間の相互変換の結果	29
5.4	GradCAM++と注意マップの比較 (心肥大-検出なし)	30
5.5	GradCAM++と注意マップの比較 (気胸-検出なし)	31
5.6	GradCAM++と注意マップの比較 (男性-女性)	32
6.1	変更後の識別モデルの構造	36
6.2	男性-女性間の相互変換の結果	38
6.3	GradCAM++と注意マップの比較 (男性-女性)	40
7.1	変換前後の気胸による胸腔と肺の境界と見られる領域	42

図目次

7.2	識別モデルの損失 IXI データセット: 男性-女性	43
7.3	識別モデルの損失 IXI データセット: 男性-女性 追加実験	45
A.1	LR Range Test + Grid Search の結果 (NIHCC: 心肥大-検出なし)	54
A.2	LR Range Test + Grid Search の結果 (NIHCC: 胸部肥厚-検出なし)	55
A.3	LR Range Test + Grid Search の結果 (IXI: 男性-女性)	56
A.4	LR Range Test + Grid Search の結果 (IXI: 男性-女性) 追加実験	57

表目次

4.1	NIHCC データセットの分割内容	16
4.2	IXI データセットの分割内容	16
4.3	データセット調整結果 (NIHCC: 心肥大-検出なし)	21
4.4	ハイパーパラメータ (NIHCC: 心肥大-検出なし)	21
4.5	モデルの性能 (NIHCC: 心肥大-検出なし)	22
4.6	データセット調整結果 (NIHCC: 気胸-検出なし)	22
4.7	ハイパーパラメータ (NIHCC: 気胸-検出なし)	23
4.8	モデルの性能 (NIHCC: 気胸-検出なし)	23
4.9	ハイパーパラメータ (IXI: 男性-女性)	23
4.10	モデルの性能 (IXI: 男性-女性)	24
4.11	各データセットごとのハイパーパラメータ (AG-CycleGAN)	25
5.1	変換結果の各画像と意味の対応	26
5.2	変換前後の混同行列 (心肥大-検出なし)	33
5.3	変換前後の混同行列 (気胸-検出なし)	33
5.4	変換前後の混同行列 (男性-女性)	34
6.1	ハイパーパラメータ (IXI: 男性-女性, 追加実験)	36
6.2	モデルの性能 (IXI: 男性-女性, 追加実験)	36
6.3	IXI データセットの追加実験に用いるハイパーパラメータ (AG-CycleGAN)	37
6.4	変換前後の混同行列 (男性-女性, 追加実験)	39
7.1	GradCAM++と AG-CycleGAN の比較	47

第 1 章

序論

近年、医療現場では人手不足による医師の負担の増加が問題となっている。そのため、近年 Computer-Aided Diagnosis(CAD) は重要な研究分野の 1 つとなっており、Convolutional Neural Network(CNN) を用いた画像認識による診断が広く使われる [1]。CNN による高精度な診断の例として、Manickavasagamr らの研究では肺結節の診断で 98.88%[2]、Alkurdi らの研究では乳がんの診断では 98%[3] の精度で診断可能なモデルが提案されている。しかし、CNN を用いた診断では診断過程の説明が困難という課題があり、患者への説明責任が果たせないため、現場への導入が進んでいないのが現状である。これまでに CNN の分類過程の説明には学習したモデルから得られる特徴マップやパラメータの重み、勾配を用いて CNN の分類に寄与する領域を特定する Class Activation Map(CAM)[4] ベースの手法が用いられてきた。しかし結果の説明には分類に寄与する領域だけでなく、領域内の形状やパターンの違いも知る必要があると考えられ、分類に寄与する領域と領域内の形状やパターンの違いの両方が得られる手法が必要になると考えられる。

そこで本研究では、対象のデータセットによく似たデータを生成できるようなデータ分布を学習から獲得することが期待できる Generative Adversarial Network(GAN)[5] を用いて、分類に寄与する領域と領域内の形状やパターンの違いを取得する手法を提案する。本研究では GAN のモデルとして CycleGAN[6] に注意モデルを導入した Attention-Guided CycleGAN(AG-CycleGAN)[7] を用いる。AG-CycleGAN ではデータドメイン間の関係性に基づいて 2 つのデータセット間の相互変換を教師なしで学習する。この学習過程で注意モデルは生成モデルによる変換の領域を制限する $[0 \sim 1]$ の値を持つ注意マップを出力し、生成モデルとともに識別モデルが元のクラスのデータと変換されたデータが区別できない

様に学習する。本研究では、注意モデルから得られる注意マップがクラス間の違いを表す領域を選択することに注目し、注意マップから分類に寄与する領域、相互変換の結果から領域内の形状やパターンの違いの獲得を目指す。また、変換の妥当性や説明可能性における注意マップの信頼性、有効性を評価するために、AG-CycleGAN による相互変換結果の妥当性、CAM ベースの手法と注意マップの比較、学習済み CNN における変換前後の識別結果の変化から変換の有効性を評価し、CNN の説明における手法の有効性の検証を行う。本研究では、これらの検証の為に肺の疾患ラベルが付けられた 2 次元胸部 X 線画像、MRI で撮像された 3 次元脳構造画像を対象に二値分類を学習させた CNN を用いて手法の検証を行う。

第 2 章

関連研究

本研究で行った CNN, GAN の学習において用いた技術や関連する技術について記述する。

2.1 Generative Adversarial Networks(GAN) モデル

本研究で用いた AG-CycleGAN の前提となる GAN, CycleGAN の説明と AG-CycleGAN について記述する。

2.1.1 Generative Adversarial Networks(GAN)

GAN[5] は Goodfellow らが提案した機械学習フレームワークであり, 図 2.1 のように生成モデル G , 識別モデル D の 2 つのモデルから構成され, 生成モデル G は入力ノイズ z に基づいて元のデータ x と区別がつかないデータの生成を学習する。また, Goodfellow らが提案した GAN の損失関数は Zhu らの研究 [6] では Adversarial Loss と呼ばれ, 式 2.1 のように表される。ここで式 2.1 の第 1 項は D が元データ x を正しく識別するために用いられ, 第 2 項は G によって生成されたデータと元のデータを識別できるように学習するために用いられる。これらの G と D の関係は通貨の偽造者と偽装通貨を取り締まる警察との関係に例えられており, 偽装者は警察を欺ける偽札を作ろうと努力し, 警察はその偽札と本物のお札を見分けようと努力する。これらの偽造者と警察間の敵対的な関係によって通貨の偽造者が製造する通貨が徐々に本物の通貨に近づく。同様に GAN においても G , D が敵対的に学習することで G は元データに近いデータの生成を学習している。また図 2.1 は GAN

2.1. GENERATIVE ADVERSARIAL NETWORKS(GAN) モデル

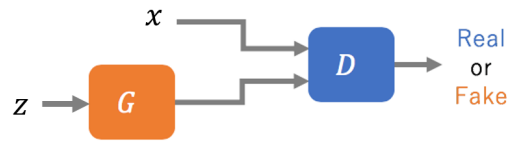


図 2.1: GAN 上でのデータ遷移

上でのデータ遷移を表したものである。

$$L(G, D) = \mathbb{E}_{x \sim P_{\text{data}}}[\log(D(x))] + \mathbb{E}_{z \sim P_Z}[1 - \log(D(z))] \quad (2.1)$$

2.1.2 CycleGAN

CycleGAN[6] は Zhu らが提案したデータドメイン間での相互変換を教師無しで学習するフレームワークであり、データドメイン X, Y に対応する生成モデル G_x, G_y 、識別モデル D_x, D_y の 4 モデルで構成されている。CycleGAN では教師無しでの相互変換: $Y : G_x : Y \rightarrow X, G_y : X \rightarrow Y$ を学習することから学習時のデータセットへの制約が緩和され、男性から女性の変換、患者から健常者への変換といった現実には解が得られない変換やデータの獲得が困難な変換もデータドメイン間の分布の違いから学習できる。また Zhu らの研究では CycleGAN における損失関数を式 2.4 のように定義されている。ここで式 2.4 における L_{GAN} は式 2, L_{cyc} は式 3 で表される。 L_{GAN} は GAN の損失関数である Adversarial Loss を示しており、一方のドメインのデータをもう一方のドメインのデータに変換することを意図した 2 つの L_{GAN} によりドメイン間での相互変換を学習する。 L_{cyc} は変換の一貫性を保持する Cycle Consistency Loss を示し、 L_{cyc} により生成モデルが一方のドメインに変換後、元のドメインに戻すように変換をした際に元のデータに戻る様な一貫性をもった変換を学習する。また図 2.2 は CycleGAN 上でのデータの遷移を表したものである。

2.1. GENERATIVE ADVERSARIAL NETWORKS(GAN) モデル

$$L(G_x, G_y, D_x, D_y) = L_{\text{GAN}}(G_x, D_x, X, Y) + L_{\text{GAN}}(G_y, D_y, Y, X) + \lambda L_{\text{cyc}}(G_x, G_y) \quad (2.2)$$

$$L_{\text{GAN}}(G_x, D_x, X, Y) = \mathbb{E}_{x \sim p_{\text{data}}(X)} [\log(D_x(x))] + \mathbb{E}_{y \sim p_{\text{data}}(Y)} [\log(D_x(1 - G_x(y)))] \quad (2.3)$$

$$L_{\text{cyc}}(G_x, G_y) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|G_x(G_y(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G_y(G_x(y)) - y\|_1] \quad (2.4)$$

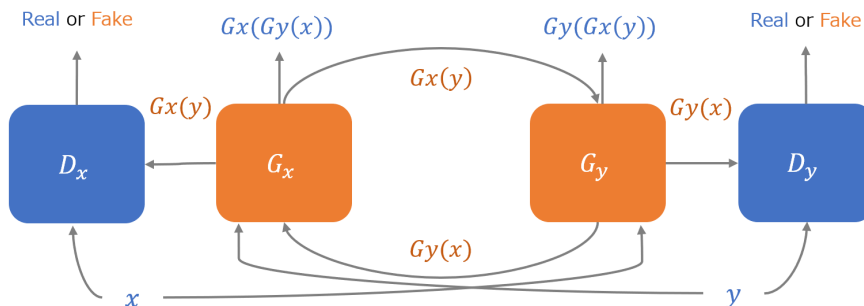


図 2.2: CycleGAN 上でのデータ遷移

2.1.3 Attention-Guided CycleGAN(AG-CycleGAN)

AG-CycleGAN[7] は Mejjati らが提案した CycleGAN に注意機構を導入したモデルであり、生成モデル G_x, G_y 、識別モデル D_x, D_y 、注意モデル A_x, A_y の計 6 つのモデルから構成されている。また、CycleGAN と同様に AG-CycleGAN でもデータドメイン X, Y 間において教師なし学習で相互変換 $G_x : Y \rightarrow X, G_y : X \rightarrow Y$ を学習する。変換の中で注意モデルは生成モデルによる変換の領域を限定する $[0 \sim 1]$ の注意マップを出力し、生成モデルとともに識別モデルが区別できないデータの生成をするように学習する。本研究では変換領

2.2. CLASS ACTIVATION MAPPING(CAM) ベースの手法

域の制限のために出力される注意マップがドメイン間の違いとなる領域を選択していると捉えられることに注目し、注意マップを変換に寄与する領域の分析に用い、生成モデルによる変換結果を領域内の形状やパターンの違いとして用いる。式 2.5 は Mejjati らが提案した AG-CycleGAN の損失関数である。Mejjati らの研究では良い注意領域が得られた後に注意領域以外を学習する方法も提案されていたが、学習が進むにつれて識別モデルが一方的に優位になり生成モデルが同じ様な画像しか生成しなくなるモード崩壊に陥る傾向が見られたので本研究では用いていない。また、式 2.5 における L_{adv}^x は CycleGAN と同様にもう一方のドメインのデータに近づける Adversarial Loss を表し、 L_{cyc}^x は一貫性のある変換を学習させる Cycle Consistency Loss を表す。また式 2.6, 式 2.7 で用いられる M_x は式 2.8 によって注意マップで制限された領域を変換するモジュールを表している。 $L_{\text{adv}}^y, L_{\text{cyc}}^y, M_y$ についても同様である。また図 2.3 は AG-CycleGAN 上でのデータ遷移を表したものであり、図 2.4 は変換モジュール M_x, M_y のデータ遷移を表したものである。

$$L(G_x, G_y, D_x, D_y, A_x, A_y) = L_{\text{adv}}^x + L_{\text{adv}}^y + \lambda_{\text{cyc}}(L_{\text{cyc}}^x + L_{\text{cyc}}^y) \quad (2.5)$$

$$L_{\text{adv}}^x(G_x, A_x, D_x) = \mathbb{E}_{x \sim p_X(x)}[\log(D_x(x))] + \mathbb{E}_{y \sim p_Y(y)}[\log(D_x(1 - M_x(y)))] \quad (2.6)$$

$$L_{\text{cyc}}^x(x, M_x(M_y(x))) = \|x - M_x(M_y(x))\|_1 \quad (2.7)$$

$$M_x(y) = A_x(y) \odot G_x(y) + (1 - A_x(y)) \odot y \quad (2.8)$$

2.2 Class Activation Mapping(CAM) ベースの手法

CAM ベースの手法では、学習済みモデルにおける最終の畳み込み層の特徴マップと出力層の重みや勾配から分類に寄与する領域を表す顕著性マップを算出する。しかし、CNN モデルでは物体の位置に対するロバスト性を確保するためにモデル内部で特徴マップの縮小が

2.2. CLASS ACTIVATION MAPPING(CAM) ベースの手法

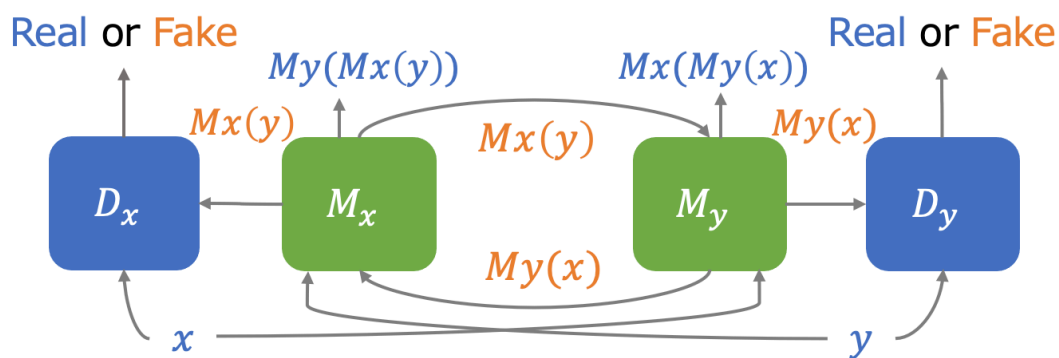


図 2.3: Attention-Guided CycleGAN(AG-CycleGAN) 上でのデータ遷移

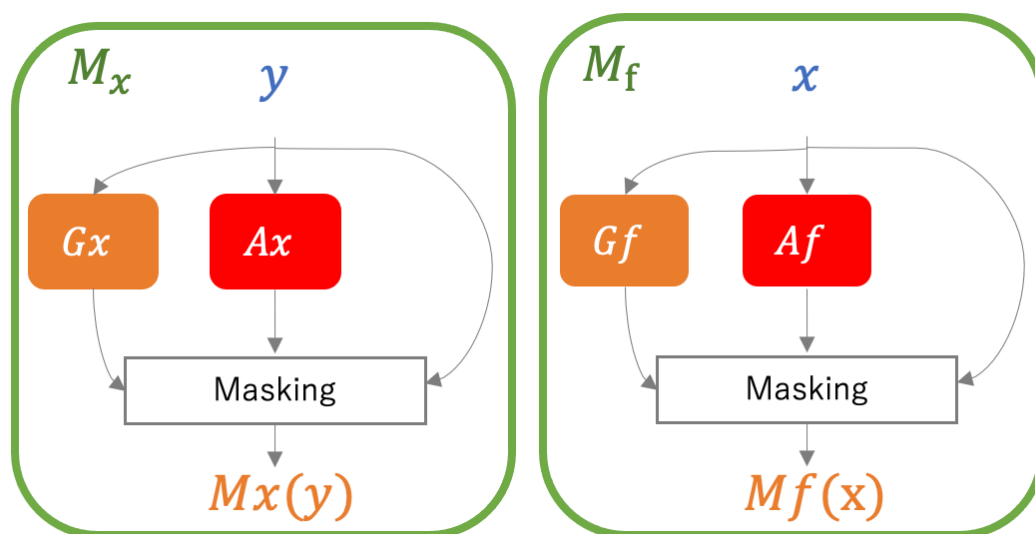


図 2.4: AG-CycleGAN で用いる変換モジュールの構造

行われ、高次の層では特徴が持つ空間情報が失われる。そのため、CAM ベースの手法から得られる顕著性マップは詳細な領域を選択することができない。また、得られる情報が領域のみでモデルが認識する具体的なクラス間の違いを得ることが困難であり、それらが CNN の説明可能性における問題として挙げられる。次に本研究で CAM ベースの手法との比較に用いる Grad-CAM++ について説明する。

2.2. CLASS ACTIVATION MAPPING(CAM) ベースの手法

2.2.1 GradCAM++

Grad-CAM++[8] は Chattopadhyay らによって Grad-CAM[9] に基づいて提案された CNN の分類に寄与する領域の分析手法であり、2 階微分による正の勾配 Grad-CAM を一般化したものと捉えることもできるとされている。式 (2.9) は Grad-CAM++ が領域の重要度として表すヒートマップの各画素値 L_{ij}^c の計算式を示したものであり、分類されたクラス c における分類結果 Y^c と、各空間 (i, j) における最終畳み込み層で得られた k 個の特徴マップの値 $A_{i,j}^k$ 、各特徴マップに対する重み w_k^c が用いられる。関数 relu は入力 x に対して $\text{relu}(x) = \max(0, x)$ を計算する関数である。重み w_k^c は式 (2.10) と定義される。ここで分類結果 Y^c を求める式 (2.11) に対して式 (2.10) を代入し、ReLU 関数の機能を考慮して relu を削除し $A_{i,j}^k$ に対して 1 階編微分を行った結果として式 (2.12)、続けて 2 階編微分を行った結果として式 (2.13) が得られ、式変形によって a_{ij}^{kc} の算出に用いる式 (2.14) が得られる。さらに、得られた式 (2.14) を式 (2.10) に代入すると重み w_k^c の算出に用いる式 2.15 が得られる。加えて、式 (2.10)、式 (2.15) は関数 relu を用いており、式 (2.15) 中の微分部分 $\text{relu}(\partial Y^c / \partial A_{i,j}^k)$ では、 relu に従った微分が行われるため特徴マップ $A_{i,j}^k$ における各空間座標 (i, j) 毎にクラス c の分類に有効な特徴が検出されれば 1、そうでなければ 0 が出力される。これにより、各空間座標 (i, j) 毎にクラス c の分類に有効な特徴における勾配が得られ、その総和から Grad-CAM++ が算出されている。

Grad-CAM[9] のヒートマップは式 (2.16) で求められ、重み w_k^c は式 (2.17) で求められる。 Z は特徴マップの画素数である。ここで Grad-CAM で重みの算出に用いる式 (2.17) と Grad-CAM++ で重みの算出に用いる式 (2.15) を比較すると、Grad-CAM では特徴マップごとに有効な特徴が占める領域の割合を算出していると考えことができ、入力に占める領域が小さい特徴が描画されなくなる可能性がある。一方、Grad-CAM++ では、特徴マップごとに有効な特徴に勾配を考慮した加重平均を求めヒートマップの計算に用いる。これにより、出力値を増加させる特徴量と重みをもとに描画することができ、Grad-CAM が苦手とする小さな特徴量も可視化することが可能になる。

2.2. CLASS ACTIVATION MAPPING(CAM) ベースの手法

$$L_{i,j}^c = \text{relu}\left(\sum_k w_k^c \cdot A_{ij}^k\right) \quad (2.9)$$

$$w_k^c = \sum_i \sum_j a_{ij}^{kc} \cdot \text{relu}\left(\frac{\partial Y^c}{\partial A_{ij}^k}\right) \quad (2.10)$$

$$\begin{aligned} Y^c &= \sum_i \sum_j L_{i,j}^c = \sum_k w_k^c \cdot \sum_i \sum_j A_{ij}^k \\ &= \sum_k \left\{ \sum_a \sum_b a_{ab}^{kc} \cdot \text{relu}\left(\frac{\partial Y^c}{\partial A_{ij}^k}\right) \right\} \cdot \left[\sum_i \sum_j A_{ij}^k \right] \end{aligned} \quad (2.11)$$

$$\frac{\partial Y^c}{\partial A_{ij}^k} = \sum_a \sum_b a_{ab}^{kc} \cdot \frac{\partial Y^c}{\partial A_{ab}^k} + \sum_a \sum_b A_{ab}^k \left\{ a_{ij}^{kc} \cdot \frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} \right\} \quad (2.12)$$

$$\frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} = 2 \cdot a_{ij}^{kc} \cdot \frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \left\{ a_{ij}^{kc} \cdot \frac{\partial^3 Y^c}{(\partial A_{ij}^k)^3} \right\} \quad (2.13)$$

$$a_{ij}^{kc} = \frac{\frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2}}{2 \frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \left\{ \frac{\partial^3 Y^c}{(\partial A_{ij}^k)^3} \right\}} \quad (2.14)$$

$$w_k^c = \sum_i \sum_j \left[\frac{\frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2}}{2 \frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \left\{ \frac{\partial^3 Y^c}{(\partial A_{ij}^k)^3} \right\}} \right] \cdot \text{relu}\left(\frac{\partial Y^c}{\partial A_{ij}^k}\right) \quad (2.15)$$

$$L_{i,j}^c = \sum_k w_k^c \cdot A_{ij}^k \quad (2.16)$$

$$w_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial Y^c}{\partial A_{ij}^k} \quad (2.17)$$

2.3 CNN モデル構造

本研究で用いた CNN モデルである Resnet, U-Net について記述する.

2.3.1 ResNet

He らによって提案された Resnet[10] では, 深層ニューラルネットワークの各層における入力 x から出力 $H(x)$ を得るために, 入力からの残差を学習する層 $F(x) := H(x) - x$ を用い, 入力との和から出力 $H(x) := F(x) + x$ を求める. これらの処理を行うモジュールである残差ブロックの構造は図 2.5 で表される. Resnet 以前の残差学習を行わないモデルでは, 入力から直接最適化された特徴マップを得ようとするため, 層が深くなるにつれて入力の情報を残すことが難しく, 深いモデルでの学習の精度を確保することが困難だった. Resnet では入力を参照した残差学習を行うことで, 各層から得られる特徴マップの最適化が容易となり, Resnet 以前の CNN で問題とされていた 50 層, 100 層を超える深いモデルでの学習においても精度を得ることが可能とされている. 本研究では手法の分析対象となる CNN である診断モデル, 男女識別モデルに 18 層の Resnet である Resnet18 を用いた.

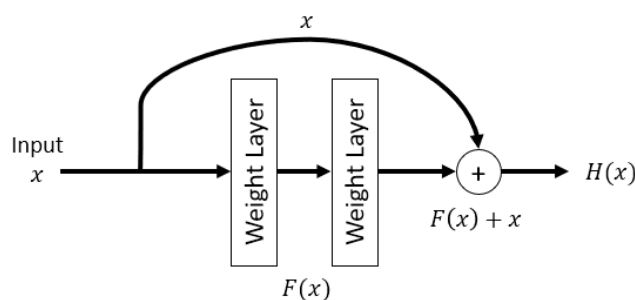


図 2.5: Resnet の残差ブロック

2.3.2 U-Net

Ronneberger らによって提案された U-Net[11] は, 段階的に特徴マップの縮小 (Down sampling) を行いつつ有効な特徴マップの獲得を目指すエンコーダ部分 (特徴抽出部分) と,

2.4. GAN の学習の安定化

段階的に特徴マップの拡大 (Up sampling と同サイズの縮小過程の特徴の連結 (concatenate)) を行った上で目的とする画像への変換を行うデコーダ部分 (出力部) によって構成される CNN のネットワーク構造である。図 2.6 は U-net における入力から出力までの特徴マップの流れを簡単に表したものである。主にセグメンテーションで用いられ、エンコーダから得られる大小幅広い受容野を持つ特徴を用いることが、画像中の様々なサイズの物体に対して高精度なセグメンテーションを可能にしていると考えられる。

本研究では、U-Net が低次の特徴から高次の特徴といった幅広い特徴を考慮した画像の変換を学習することに着目し、エンコーダ部分に分析対象となる CNN の構造と重みを導入することでデコーダ部分が CNN の推論過程の特徴も考慮した画像変換を行うことを期待し、AG-CycleGAN の生成モデルと注意モデルでエンコーダ部分の構造を分析対象の Resnet18 の特徴抽出部と同様にしたモデルを用いた。

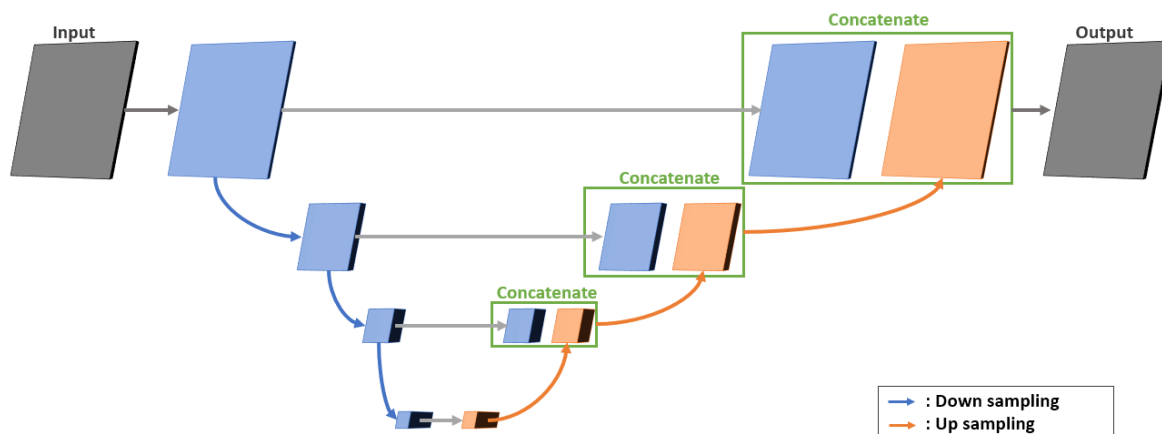


図 2.6: U-net の構造

2.4 GAN の学習の安定化

GAN の学習では、識別モデルと生成モデル間の性能のバランスを適切にすることが良い生成結果を得るために重要であり、両者に制約を設けず学習させた場合に生成モデルよりも識別モデルが強くなる傾向があるため、識別モデルに構造的に弱いモデルを用いるなど制限を行うことで適切なバランスを保つように工夫を行う。ここで識別モデルの強弱は、生成モ

2.4. GAN の学習の安定化

デルが生成する偽のデータと元データとの識別をどのくらい高精度に行えるかを表す。学習がうまく進まない例として、生成・識別モデルの敵対的關係の中で極端に識別モデルが優位に学習された場合、生成モデルが近似する分布と識別モデルが捉える分布が乖離し、生成データと元データを完全に識別可能な識別モデルが学習され、生成モデルが学習できなくなる場合があるとされている [12][13]。その為、GAN では学習の安定化に向けて様々な手法が提案されており、安定化に重要な条件の 1 つに識別モデルがリップシツツ制約を満たすことがある。本研究では識別モデルにリップシツツ制約を満たさせるために、次に説明するスペクトル正規化を用いた。

2.4.1 スペクトル正規化 (Spectral Normalization)

スペクトル正規化 (Spectral Normalization)[12] は、Miyato らによって GAN における学習の安定化に貢献にする正規化手法として提案されたものであり、従来の識別モデル内のバッチ正規化と置き換えることで実装することができる。GAN の安定化に向けた戦略の 1 つとして、識別モデル D にリップシツツ定数 $K = 1$ のリップシツツ連続を満たすリップシツツ制約を設ける方法がある。式 (2.18) はリップシツツ連続である関数 f が満たす式であり、関数 f の任意の 2 点 x, x' を結ぶ傾きの絶対値がリップシツツ定数 K 以下になり、連続関数の中でも急激な変化のない滑らかな関数がリップシツツ制約を満たす関数となる。スペクトル正規化では識別モデルによる識別関数 f の各層のパラメータの重み W に対して特異値分解を行い、得られた最大の特異値で割ることで重み W をのスケーリングを行い、識別関数 f がリップシツツ定数 $K = 1$ のリップシツツ制約を満たすように制御する。(2.19) スペクトル正規化で行われる計算を表したものであり、 $\delta(W)$ は重み w の最大の特異値である。

$$\left| \frac{f(x) - f(x')}{x - x'} \right| \leq K \quad (2.18)$$

$$W_{\text{SN}} = \frac{W}{\delta(W)} \quad (2.19)$$

2.5 ハイパーパラメータ探索

機械学習では学習率, 正則化の強さなどのハイパーパラメータと呼ばれる学習前に設定する変数があり, タスクやモデルなどの学習条件によって最適な学習条件が異なる. ここでは本研究で CNN モデルがタスクに対して適切に学習するためにより適したハイパーパラメータを設定するために用いたハイパーパラメータ探索手法について記述する.

2.5.1 グリッドサーチ (Grid Search)

グリッドサーチはハイパーパラメータ最適化において広く用いられる手法であり, 学習率, 隠れユニット数, 重みの正則化の強さなどの K 個の各ハイパーパラメータに対して探索の候補に用いる複数の値の集合を $L(1), \dots, L(K)$ で表した時, グリッドサーチは可能な限り全ての組み合わせで学習をし, 目的関数が最小となる最適なハイパーパラメータを探索するため, 試行回数 S は $S = \prod_{k=1}^K |L(k)|$ で求められるとされている [14].

2.5.2 LR range test

LR range test[15] は Smith によって提案された学習率 (Learning rate) の探索手法であり, 小さいエポック数で学習する間にステップサイズとして設定した間隔で学習率を設定した最小値から最大値に直線的に増加させ, モデルが収束し精度が上がり始めた値を学習率として決定する手法である. また, Cyclical learning rate を用いる場合は精度が上がり始めた値を Min lr, 精度の上昇が円滑に行われなる直前の値を Max lr として用いる. LR range test を行う上で精度でなく損失を基準に決定する場合もあり, 本研究でも広い値域で学習率の探索を行うために損失を基準にステップサイズを 1 イテレーションに設定し, 損失が収束し始めた値, 損失が増加傾向に転じる直前の値をそれぞれ Min lr, Max lr とした.

第 3 章

提案手法

本研究では、学習済み CNN が認識する分類に寄与する領域と領域内の形状やパターンの違いの獲得を目指して Attention-Guided CYcleGAN(AG-CycleGAN) を用いた分析手法を提案する。そして、AG-CycleGAN の変換過程で注意モデルから得られる注意マップが分類に寄与する領域、変換結果が領域内の形状やパターンの違いを表すと仮定し、AG-CycleGAN が捉えるデータドメイン間の違いを視覚的に分析する。また CNN が抽出した特徴マップに基づいた相互変換の学習を行うために、AG-CycleGAN における生成モデルと注意モデルの Backbone には CNN の Backbone と共通の同じ構造とパラメータを持たせる。しかし、単純に相互変換を学習しただけでは AG-CycleGAN と CNN が同じ違いを認識したことを確認できない。そこで CNN の最終の畳み込み層から得られる特徴マップと誤差逆伝搬法によって得られる勾配から CNN の識別に寄与する領域を獲得する Grad-CAM++[8] と注意マップを比較し、CNN と AG-CycleGAN が捉える領域の類似性についても検証を行う。また、同じデータセットで学習済みの CNN に対して変換前後のデータを用いた分類を行う。提案手法を下記に示す。

1. CNN と共通の Backbone を用いた AG-CycleGAN による相互変換の学習
2. AG-CycleGAN から得た分類に寄与する領域と領域内の形状やパターンの違いの評価
3. Grad-CAM++と注意マップの類似性の評価
4. 変換前後での CNN の分類結果の違いから AG-CycleGAN の変換の有効性を評価

第 4 章

実験内容

ここでは提案手法を検証するために用いたデータセット、対象となる CNN モデル、実験手順について記述する。

4.1 データセット

本実験ではデータセットに胸部 X 線画像データセットと T1 脳構造画像データセットを用いる。以下に各データセットの説明を記述する。

4.1.1 胸部 X 線画像 NIHCC データセット

胸部 X 線画像では NIH Clinical Center が提供する胸部 X 線画像データセット [16] を使用し、以降は NIHCC データセットと表現する。このデータセットでは 112,120 枚の胸部 X 線画像に 14 種類の疾患のラベルが付けられており、症状が検出されなかった画像には検出なし (No finding) のラベルが付けられている。ここでの検出なしの画像はラベルに無い疾患の患者が含まれる可能性があり、健常者とは異なる。図 4.1 は NIHCC データの各ラベルの分布を表すヒストグラムであり、図 4.1(a) は併発なしのデータのみ分布、図 4.1(b) は併発ありのデータを含めた分布である。本研究では GAN の学習に必要なデータ数が多いため、併発ありの条件で各ラベルのデータを抽出している。また検出なしのデータは他のラベルを持たない為、疾患ラベルと検出なしラベルでそれぞれ抽出した場合、データの重複は起こらない。本実験ではそれらのデータのうち、CNN による分類で一定以上の精度が確認でき、GAN の学習に必要なデータ数を確保できる心肥大 (Label: Cardiomegaly) と検出なし

4.1. データセット

(No finding), 気胸 (Label:Pneumothorax) と検出なし (No finding) の 2 パターンの組み合わせを検証に用い, 前処理には $[-1 \sim 1]$ 間へのスケーリングを行った. 学習に用いるにあたり, データセットを分割した結果を表 4.1 に示す. CNN の学習と GAN の学習に用いるテストデータの内容は共通である.

表 4.1: NIHCC データセットの分割内容

	CNN の学習			GAN の学習		合計
	訓練	検証	テスト	訓練	テスト	
心肥大 (Cardiomegaly)	1776	444	556	2220	556	2776
気胸 (Pneumothorax)	3392	849	1061	4241	1061	5302
検出なし (no finding)	38630	9658	12073	48388	12073	60361

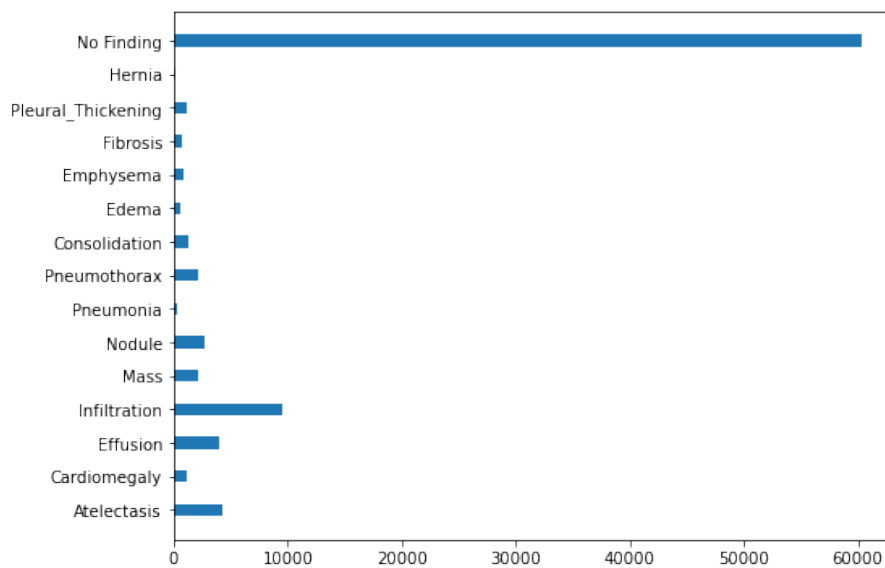
4.1.2 脳構造画像 IXI データセット

MRI で撮像された脳構造画像として IXI データセット [17] を用いる. IXI データセットでは MRI で撮像された健常者の 3 次元脳構造画像が 590 枚あり, T1 脳構造画像に含まれる性別をラベルとして用いて男女間の脳構造の違いを対象に検証する. 前処理では標準化を行った. 学習に用いるにあたり, データセットを分割した結果を表 4.2 に表す. ここでも CNN の学習と GAN の学習に用いるテストデータの内容は共通である.

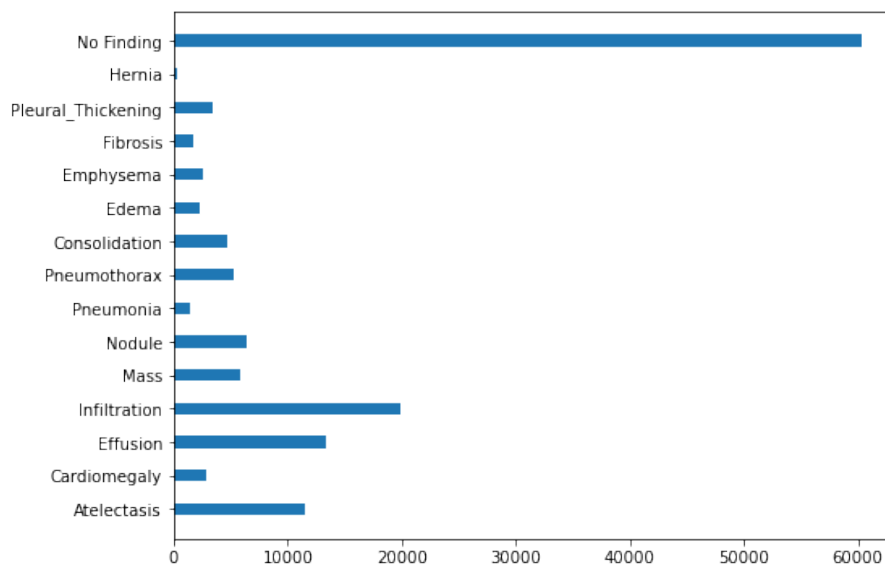
表 4.2: IXI データセットの分割内容

	CNN の学習			GAN の学習		合計
	訓練	検証	テスト	訓練	テスト	
男性	168	42	53	210	53	263
女性	208	53	66	261	66	327

4.1. データセット



(a) 各ラベルのデータ数 (併発なし)



(b) 各ラベルのデータ数 (併発あり)

図 4.1: NIHCC データセット内の各ラベルのデータ分布

No finding: 検出なし, Hernia: ヘルニア, Pleural Thickening: 胸膜肥厚, Fibrosis: 繊維化, Emphysema: 肺気腫, Edema: 浮腫, Consolidation: 浸潤影, Pneumothorax: 気胸, Pneumonia: 肺炎, Nodule: 結節, Mass: 腫瘤, Infiltration: 浸潤, Effusion: 滲出液, Cardiomegaly: 心肥大, Atelectasis: 無気肺

4.2 モデル構造

提案手法の検証を行うにあたり、診断モデル、男女分類モデルとして用いた2値分類を学習するCNNの構造、CNNの分析に用いるGANを構成する生成モデル、注意モデル、識別モデルの構造について記述する。

モデル構造を図示するにあたり Resnet[10] の構造を簡略化するために一般的な層に加えて図4.2(a)の「Bottleneck1」、図4.2(b)の「Bottleneck2」をモデル構造の説明に用いる。また、図4.2のConvolution層を表す「Conv n, a, b 」は「Conv カーネルサイズ, 入力チャンネル数, 出力チャンネル数」を表し、「 $/2, \times 2$ 」は特徴マップが $1/2$ 倍, 2 倍の大きさになることを意味する。以降も同様に表現する。

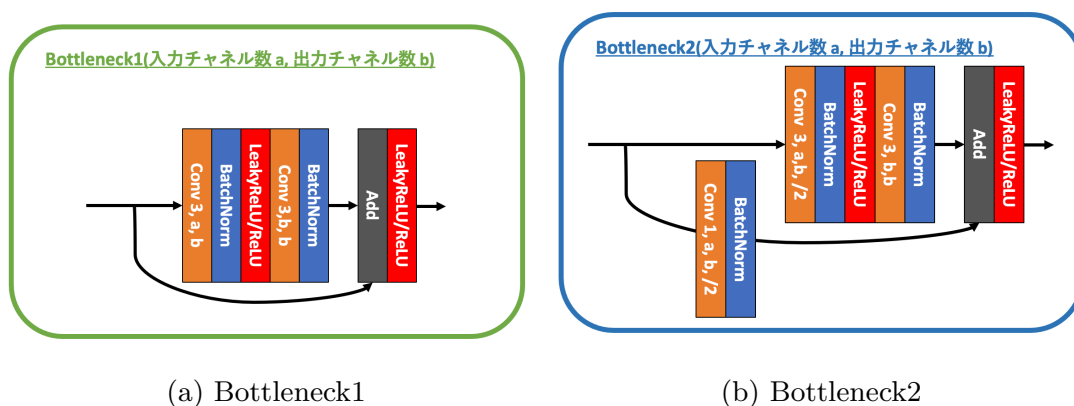


図 4.2: Resnet の図示に用いるモジュール

4.2.1 CNN モデル (診断モデル・男女診断モデル)

本研究では分析対象のCNNとしてHeらが提案したResnet18[10]を用いる。図4.3はCNNのモデル構造を表したものであり、CNNモデルではResnetのBottleneck1,2の活性化関数としてLeakyReLUを使用する。また、本研究では図4.3中の特徴抽出部(バックボーン)部分をGANの生成モデル、注意モデルの特徴抽出部に用いる。

4.2. モデル構造

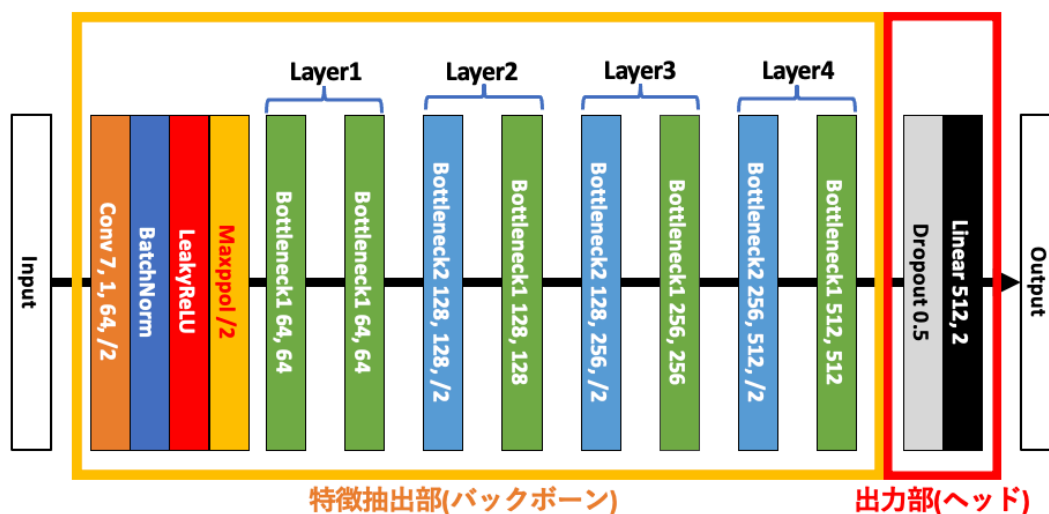


図 4.3: CNN モデルの構造 (Resnet18)

4.2.2 生成モデル・注意モデルの構造

検証に用いる AG-CycleGAN の生成モデル・注意モデルとしてエンコーダ (特徴抽出部) に Resnet18 の特徴抽出部と共通の構造を持つ U-Net を用いる。図 4.4 は生成モデル・注意モデルの構造を表したものであり、ここでは Bottleneck1,2 の活性化関数として ReLU を用い、出力層では生成モデルは Tanh 関数、注意モデルは Sigmoid 関数を用いた。本研究ではエンコーダ部分のパラメータを対象の CNN モデルと同じにし、デコーダ部分のみを学習させた。

4.2.3 識別モデル

検証に用いる AG-CycleGAN の識別モデルとして、Batch Normalization の代わりに Spectral Normalization を導入した単純な構造の CNN を用いる。図 4.5 は識別モデルの構造を表したものである。

4.3. 対象となる CNN モデル (診断モデル, 男女分類モデル)

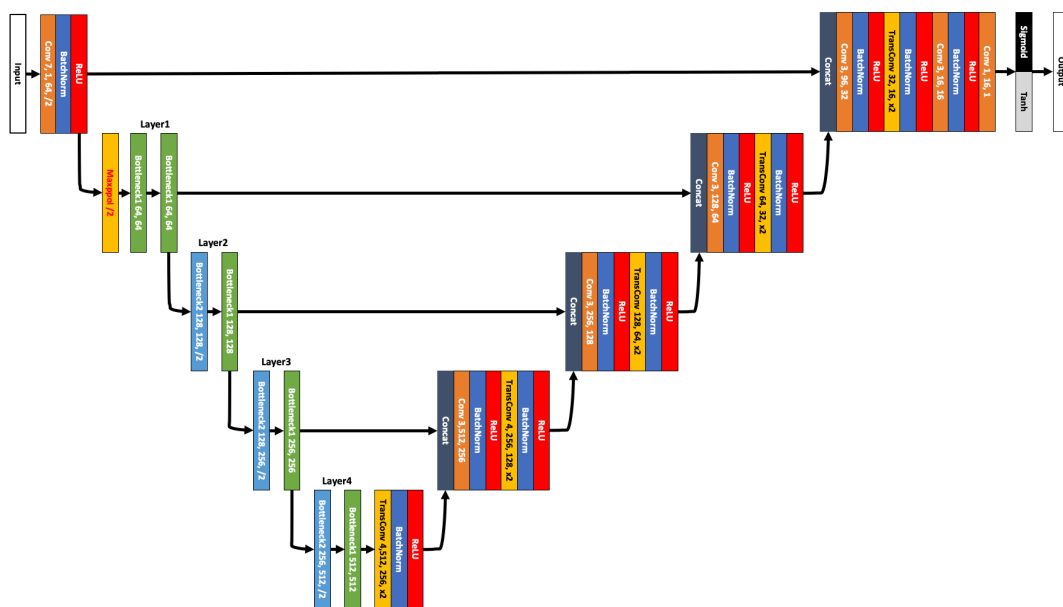


図 4.4: 生成モデル・注意モデルの構造 (U-net)

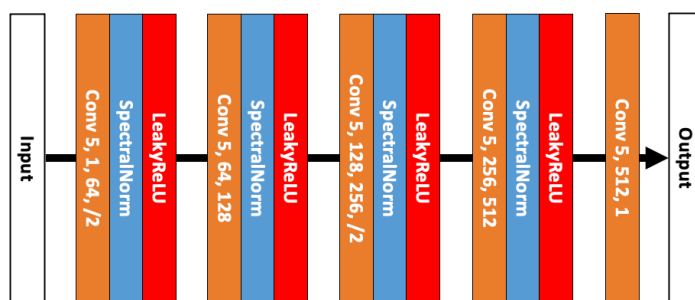


図 4.5: 識別モデルの構造

4.3 対象となる CNN モデル (診断モデル, 男女分類モデル)

本研究では 2 次元胸部 X 線画像における心肥大-検出なし間と気胸-検出なし間, 3 次元脳構造画像 (T1 画像) における男女間で 2 値分類を行う CNN モデルを診断モデル, 男女分類モデルとし, それらを対象に AG-CycleGAN を用いた相互変換による識別過程の分析の検証を行う. そのため, 研究で用いた診断モデル, 男女分類モデルにおける 2 値分類の学習結果を記述する. また, 本研究で用いる診断モデル, 男女分類モデルには入力チャンネルを 1, 出力チャンネルを 2 に設定した Resnet18 を用いた.

4.3. 対象となる CNN モデル (診断モデル, 男女分類モデル)

4.3.1 診断モデル (NIHCC データセット: 心肥大-検出なし)

NIHCC データセット中の心肥大 (Cardiomegaly) と検出なし (No finding) 間で CNN を学習した結果について記述する. まず学習するにあたり, CNN のオプティマイザーとして用いた SGD のハイパーパラメータである学習率, Momentum, Weight Decay の値を決定するために行った LR Range Test と Grid Search の結果を示す. ここで探索条件として LR Range Test では 20epochs, Batch size = 64 で学習率の範囲は $[10^{-7} \sim 10^0]$, Grid Search では Momentum は 0.9, 0.95, 0.99, Weight Decay では $10^{-3}, 10^{-4}, 10^{-5}$ を候補にハイパーパラメータ探索を行った. またクラス間のデータ数の偏りが大きいため, 訓練データと検証データではクラス間のデータ数が均衡になるようにエポックごとに抽出し, 調整した. 表 4.3 は調整した結果のデータ数を示したものである. LR Range Test と Grid Search の結果として付録の図 A.1 で示す結果が得られ, この結果をもとにハイパーパラメータを表 4.4 で示す値に決定した. また CNN の学習には学習率のスケジューラに Cyclical Learning Rate を用いるため, 学習率は Loss が下がり始めた学習率を Base lr, Loss が反発する寸前の学習率を Max lr とした. 次に決定したハイパーパラメータを設定した CNN モデルを学習し, 検証損失が最小のモデルを診断モデルとして扱う CNN モデルとして保存した. 診断モデルの性能評価結果を表 4.5 に示す.

表 4.3: データセット調整結果 (NIHCC: 心肥大-検出なし)

	心肥大	検出なし	調整後の合計
訓練データ	1776	38630	3552
検証データ	444	9658	888
テストデータ	556	12073	12629

表 4.4: ハイパーパラメータ (NIHCC: 心肥大-検出なし)

パラメータ名	epoch	Batch size	Base lr	Max lr	Momentum	Weght Decay
数値	200	64	2×10^{-5}	10^{-3}	0.99	10^{-3}

4.3. 対象となる CNN モデル (診断モデル, 男女分類モデル)

表 4.5: モデルの性能 (NIHCC: 心肥大-検出なし)

	precision	recall	f1-score	support
検出なし	0.98	0.75	0.85	12073
心肥大	0.11	0.66	0.19	556
accuracy			0.75	12629
macro avg	0.55	0.71	0.52	12629
weighted avg	0.94	0.75	0.82	12629

4.3.2 診断モデル (NIHCC データセット: 気胸-検出なし)

同様に, 気胸 (Pneumothorax) と検出なし (No finding) 間の 2 値分類において CNN を学習した結果について記述する. CNN のオプティマイザーとして用いた SGD のハイパーパラメータである学習率, Momentum, Weight Decay の値を LR Range Test と Grid Search を用いて探索を行った. また訓練データと検証データではクラス間のデータ数が均衡になるようにエポックごとに抽出し, 表 4.6 で表すようにデータ数を調整した. 探索結果として付録の図 A.2 に示す結果が得られた. 探索条件は心肥大-検出なしの場合と同様である. さらに, その結果をもとにハイパーパラメータを表 4.7 で示す値に決定した. 次に決定したハイパーパラメータを設定した CNN モデルを学習し, 検証損失が最小のモデルを診断モデルとして扱う CNN モデルとして保存した. 診断モデルの性能評価結果を表 4.8 に示す.

表 4.6: データセット調整結果 (NIHCC: 気胸-検出なし)

	気胸	検出なし	調整後の合計
訓練データ	3392	38630	6784
検証データ	849	9658	1698
テストデータ	1061	12073	13134

4.3. 対象となる CNN モデル (診断モデル, 男女分類モデル)

表 4.7: ハイパーパラメータ (NIHCC: 気胸-検出なし)

パラメータ名	epoch	Batch size	Base lr	Max lr	Momentum	Weght Decay
数値	200	64	2×10^{-5}	4×10^{-4}	0.99	10^{-3}

表 4.8: モデルの性能 (NIHCC: 気胸-検出なし)

	precision	recall	f1-score	support
検出なし	0.96	0.72	0.82	12073
気胸	0.17	0.67	0.27	1061
accuracy			0.71	13134
macro avg	0.57	0.69	0.55	13134
weighted avg	0.90	0.71	0.78	13134

4.3.3 男女分類モデル (IXI データセット: 男性-女性)

同様に, IXI データセットの脳構造画像 (T1 画像) を用いて男性と女性間の 2 値分類において, CNN のオプティマイザーとして用いた SGD のハイパーパラメータである学習率, Momentum, Weight Decay の値を決定するために行った LR Range Test と Grid Search の結果を示す. ここで探索条件は NIHCC データセットでの実験と同様に設定した. 図 A.3 は LR Range Test と Grid Search の結果であり, この結果をもとにハイパーパラメータを表 4.9 で示す値に決定した. また, IXI データセットではデータの偏りが小さいためデータ分布の調整は行わずに実験を行った. 男女分類モデルの性能評価を表 4.10 に示す.

表 4.9: ハイパーパラメータ (IXI: 男性-女性)

パラメータ名	epoch	Batch size	Base lr	Max lr	Momentum	Weght Decay
数値	80	8	4×10^{-5}	4×10^{-4}	0.99	10^{-5}

4.4. 実験手順

表 4.10: モデルの性能 (IXI: 男性-女性)

	precision	recall	f1-score	support
男性	0.91	0.77	0.84	66
女性	0.76	0.91	0.83	53
accuracy			0.83	119
macro avg	0.84	0.84	0.83	119
weighted avg	0.84	0.83	0.83	119

4.4 実験手順

AG-CycleGAN の学習を始めるにあたり、まず診断モデル・男女分類モデルの特徴抽出部のパラメータを生成モデル・注意モデルの特徴抽出部に読み込ませる。その後、CNN の分析のために AG-CycleGAN による相互変換を学習させ、学習後の AG-CycleGAN から出力される注意マップと変換結果から分類に寄与する領域と領域内の形状やパターンの違いからクラス間の違いを捉えた結果が得られたか評価する。男女分類モデルを対象にした実験では、クラス間の違いの獲得に有効な結果が得られたかを確認する。さらに診断モデル・男女分類モデルから得られる GradCAM++ と注意マップを比較し、注目部位の類似性や得られた結果の違いを評価する。最後に変換前後のテストデータと元データのラベルを用いて診断モデル・男女分類モデルに分類させ、AG-CycleGAN の変換の有効性を評価する。

また、AG-CycleGAN の学習を行うにあたり、ハイパーパラメータは表 4.11 で表す値に設定した。表 4.11 中の「n_dis」は、学習の安定化のために Spectral Normalization を用いた識別モデルは学習が遅くなるため、パラメータの更新回数を $G : D = 1 : [n_dis]$ とし、識別モデルを多く学習させるように設定する値である。

4.4. 実験手順

表 4.11: 各データセットごとのハイパーパラメータ (AG-CycleGAN)

データセット名	epoch	Batch size	Optimizer	Learning rate	b1	b2	n_dis
NIHCC	200	1	Adam	2×10^{-4}	0.5	0.999	5
IXI	500	1	Adam	2×10^{-4}	0.5	0.999	1

第 5 章

実験結果

診断モデル，男女分類モデルに対して提案手法の検証を行った結果について記述する．

5.1 AG-CycleGAN による相互変換の学習

AG-CycleGAN による相互変換から得られた注意マップと変換結果を用いて，分類に寄与する領域と領域内の形状やパターンの違いを分析した結果について記述する．また変換結果における各画像と意味は表 5.1 で示す通りである．

表 5.1: 変換結果の各画像と意味の対応

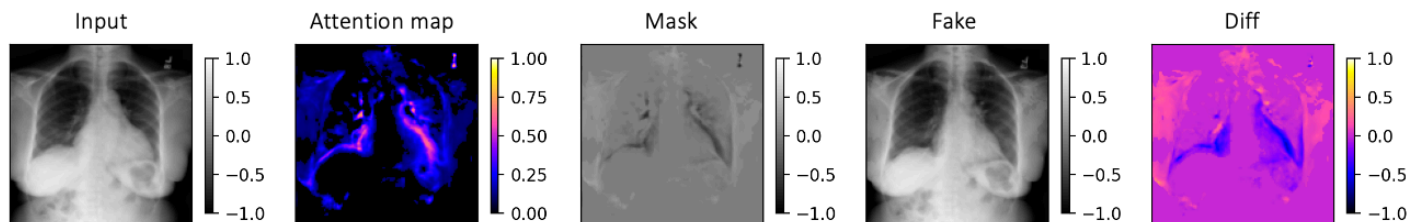
画像名	画像の意味
input	AG-CycleGAN に入力する画像
Attention map	AG-CycleGAN の注意モデルから得られる注意マップ
Mask	注意マップ (Attention map) 内の変換結果
Fake	AG-CycleGAN によるクラス間の変換結果
Diff	Fake から input を引いた差分画像

5.1.1 NIHCC データセット: 心肥大-検出なし

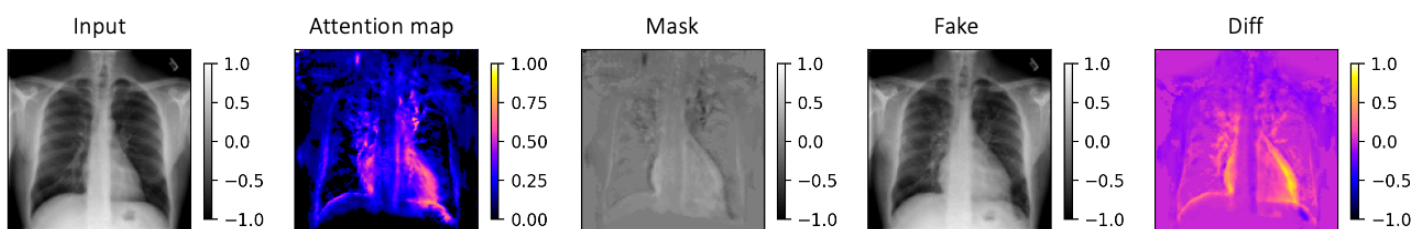
NIHCC データセットにおける心肥大-検出なし間で分類を行った CNN と共通の特徴抽出部を用いて AG-CycleGAN に相互変換を学習させた結果として図 5.1 に示す結果が得られた．結果より，分類に寄与する領域として心臓の側面にあたる領域が得られ，心肥大

5.1. AG-CYCLEGAN による相互変換の学習

→ 検出なしの変換 (図 5.1(a)) では心臓が小さくなるように, 検出なし → 心肥大の変換図 5.1(b)) では心臓が大きくなる様な変換が見られた.



(a) 心肥大 (Cardiomegaly) → 検出なし (No Finding)



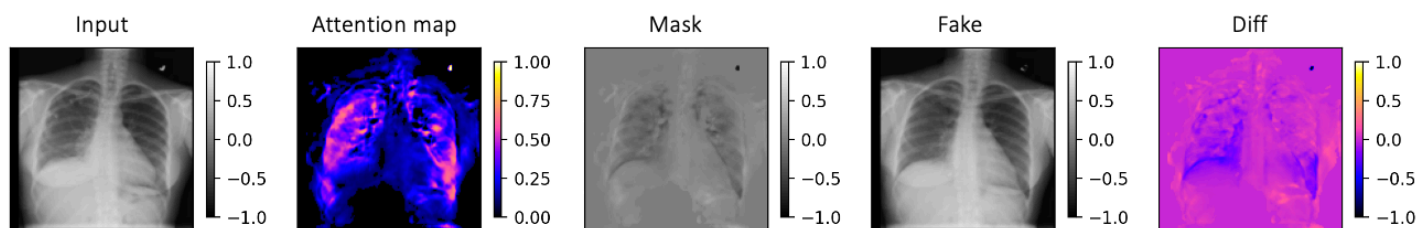
(b) 検出なし (No Finding) → 心肥大 (Cardiomegaly)

図 5.1: 心肥大-検出なし間の相互変換の結果

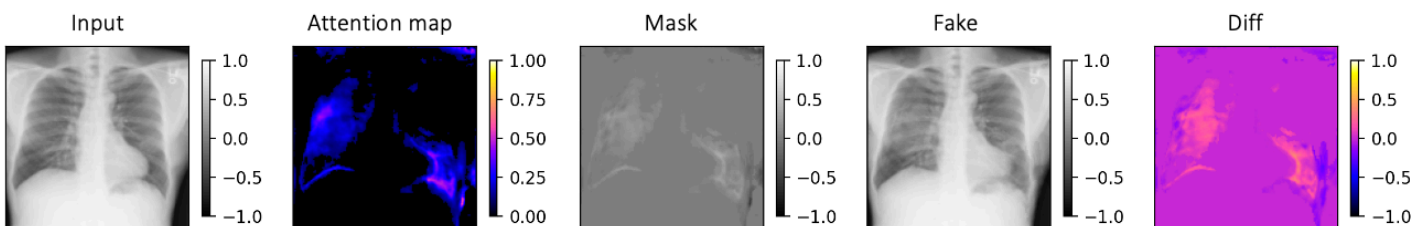
5.1.2 NIHCC データセット: 気胸-検出なし

NIHCC データセットにおける気胸-検出なし間で分類を行った CNN と共通の特徴抽出部を用いて AG-CycleGAN に相互変換を学習させた結果として図 5.2 に示す結果が得られた. 結果より, 分類に寄与する領域として肺領域の周辺部分が得られた. また変換結果として, 気胸 → 検出なしの変換 (図 5.2(a)) では肺領域の周辺を黒くする様な変換が見られ, 検出なし → 心肥大の変換 (図 5.2(b)) では肺領域の周辺を白くする様な変換結果が得られた.

5.1. AG-CYCLEGAN による相互変換の学習



(a) 気胸 (Pneumothorax) → 検出なし (No Finding)



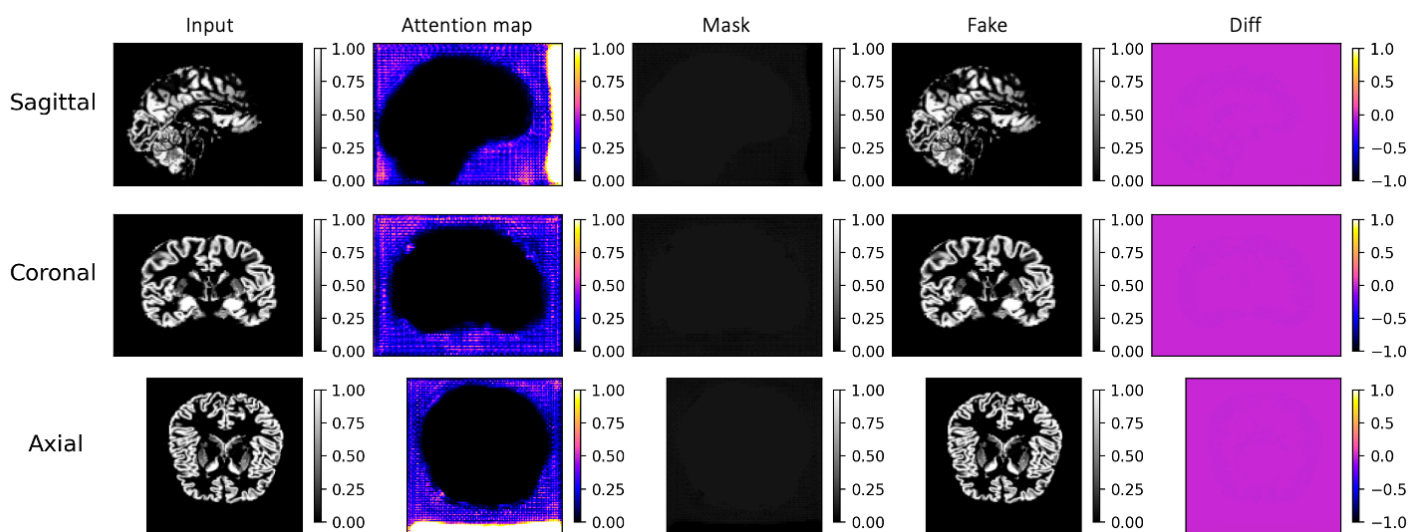
(b) 検出なし (No Finding) → 気胸 (Pneumothorax)

図 5.2: 気胸-検出なし間の相互変換の結果

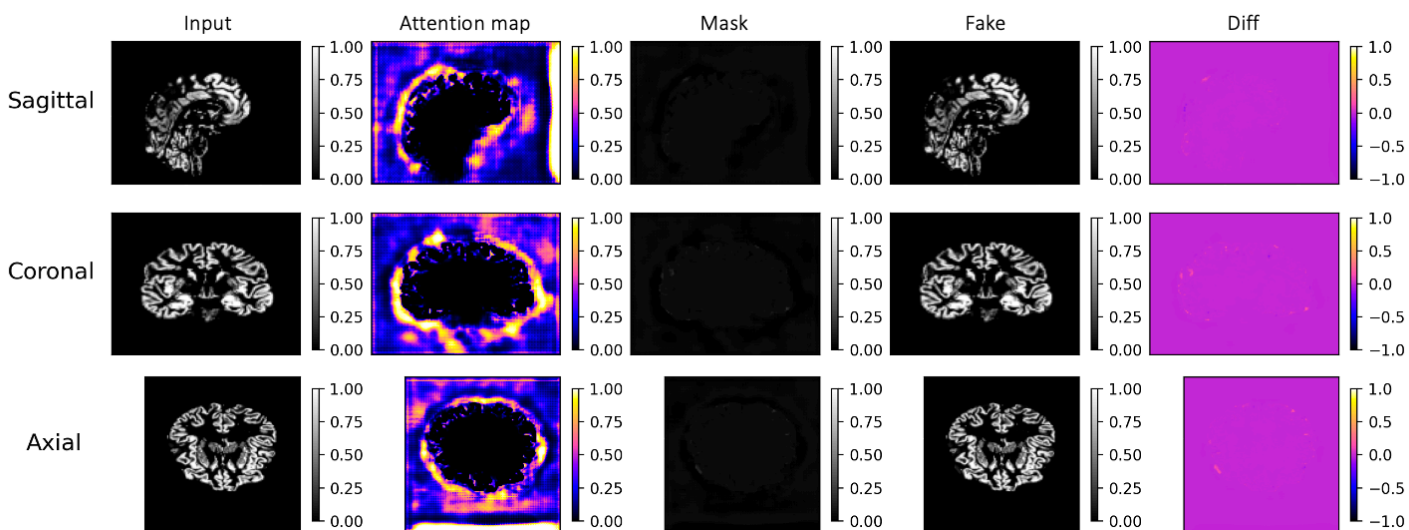
5.1.3 IXI データセット: 男性-女性

IXI データセットにおける男性-女性間で分類を行った CNN と共通の特徴抽出部を用いて AG-CycleGAN に相互変換を学習させた結果として、図 5.3 に示す結果が得られた。結果より、分類に寄与する領域として脳領域の周辺部分が得られた。また変換結果として、男性 → 女性の変換 (図 5.3(a)), 女性 → 男性の変換 (図 5.3(b)) 共に変換が有効に働かなかった。

5.2. GRADCAM++と注意マップの比較



(a) 男性 → 女性



(b) 女性 → 男性

図 5.3: 男性-女性間の相互変換の結果

5.2 GradCAM++と注意マップの比較

診断モデル，男女分類モデルで正しく識別されたデータの GradCAM++と注意マップの比較を行った結果を記述する．

5.2. GRADCAM++と注意マップの比較

5.2.1 NIHCC データセット: 心肥大-検出なし

気胸-検出なし間の分類を学習した診断モデルが正しく分類を行えたデータを対象に Grad-CAM++と注意マップの比較を行った結果, 図 5.4 が得られた. 図 5.4(a) は心肥大, 図 5.4(b) は検出なしのデータを対象に出力したものである.

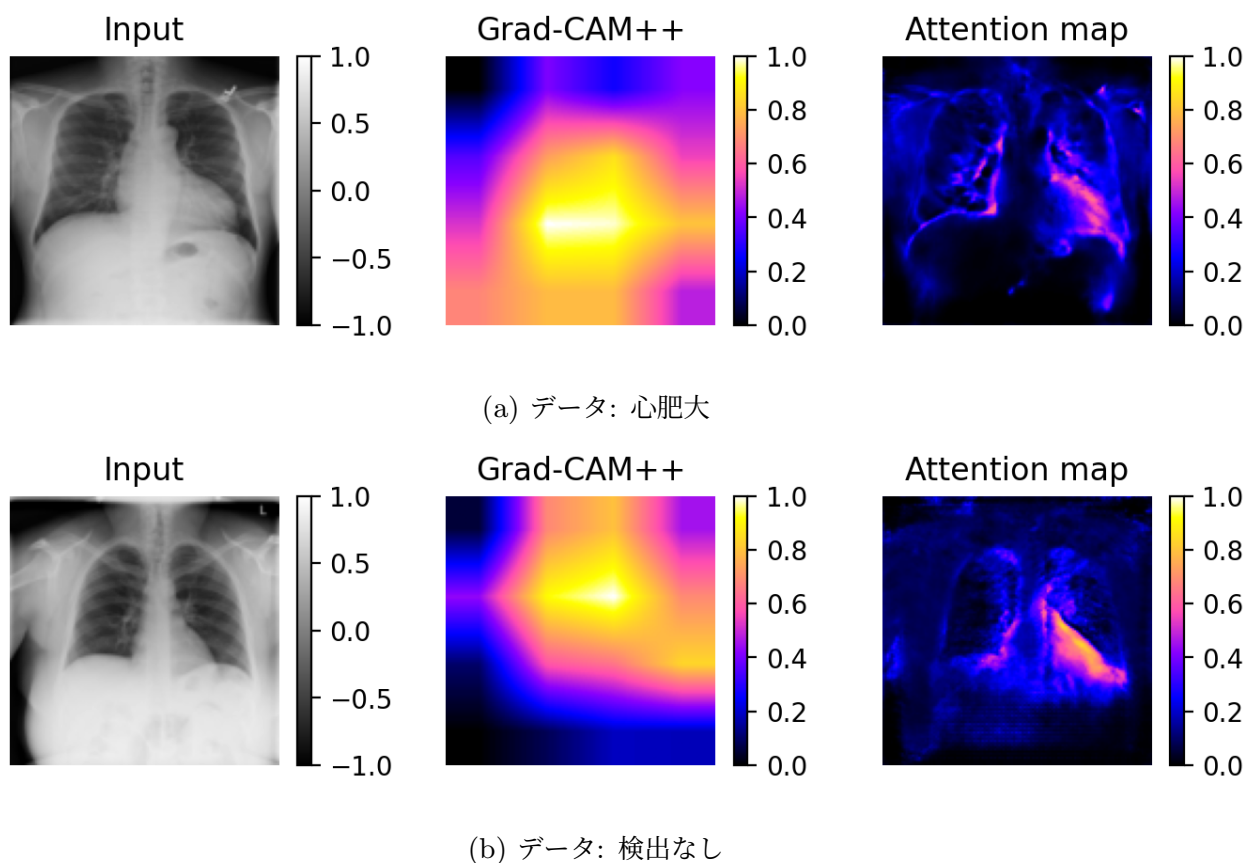
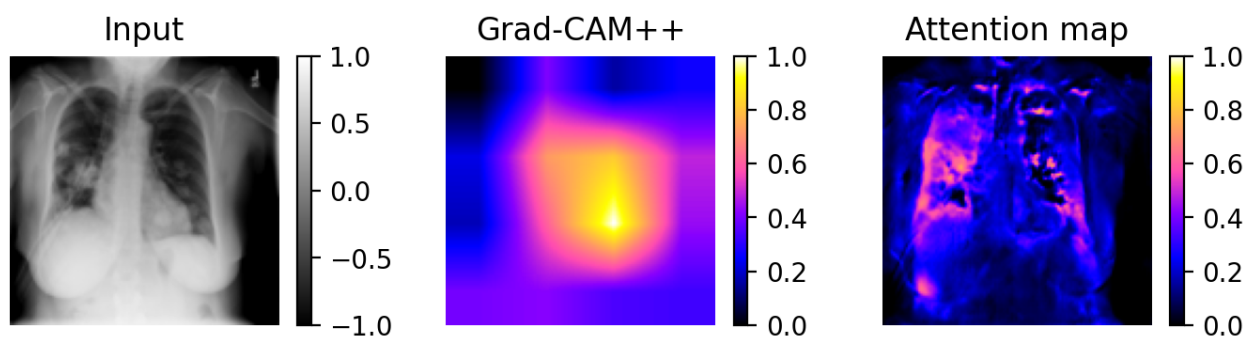


図 5.4: GradCAM++と注意マップの比較 (心肥大-検出なし)

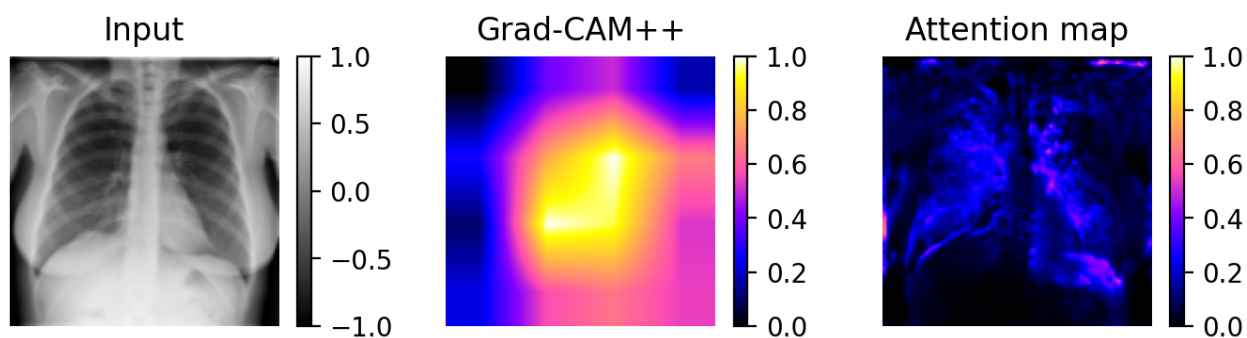
5.2.2 NIHCC データセット: 気胸-検出なし

気胸-検出なし間の分類を学習した診断モデルが正しく分類を行えたデータを対象に Grad-CAM++と注意マップを出力した結果, 図 5.5 が得られた. 図 5.5(a) は心肥大, 図 5.5(b) は検出なしのデータを対象に出力したものである.

5.2. GRADCAM++と注意マップの比較



(a) データ: 気胸



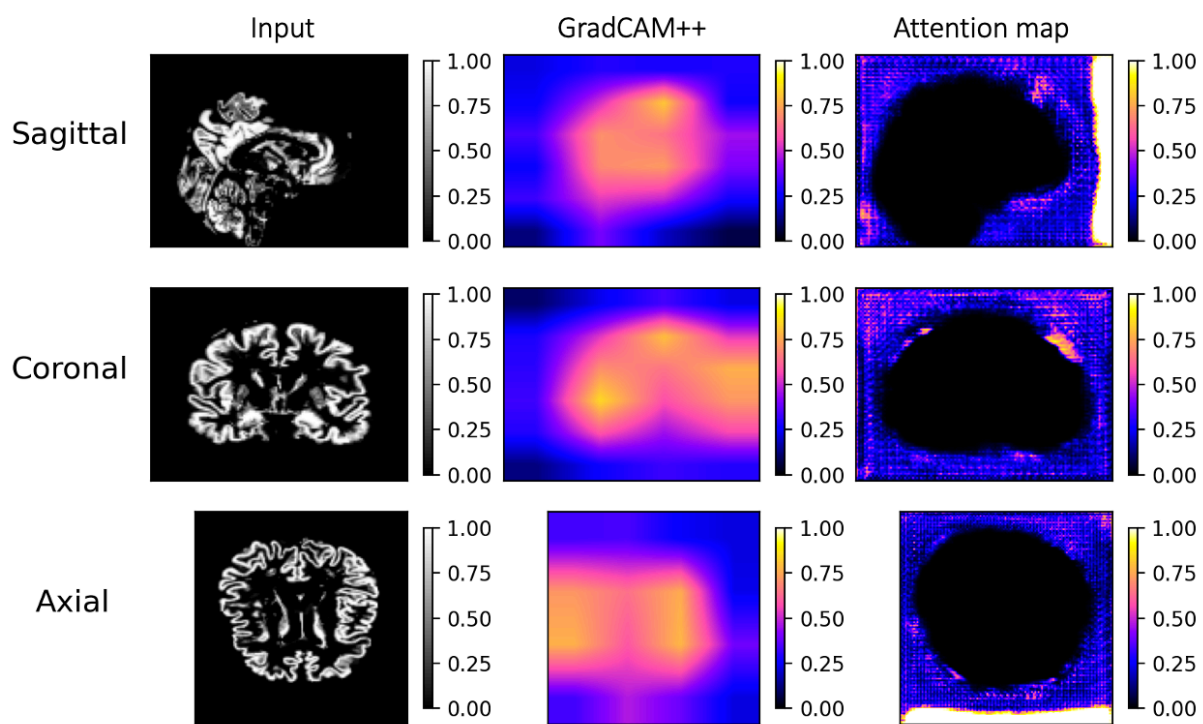
(b) データ: 検出なし

図 5.5: GradCAM++と注意マップの比較 (気胸-検出なし)

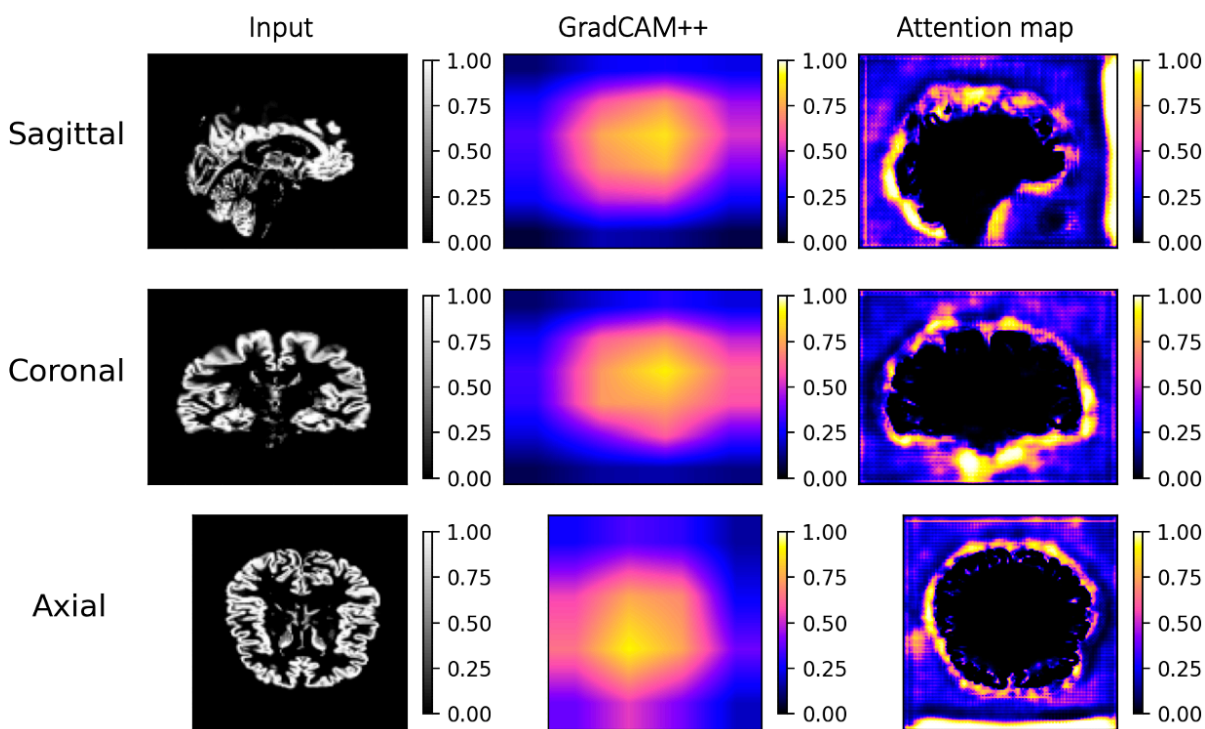
5.2.3 IXI データセット: 男性-女性

男性-女性間の分類を学習した診断モデルが正しく分類を行えたデータを対象に Grad-CAM++と注意マップを出力した結果として、図 5.6 に示す結果が得られた。図 5.6(a) は男性のデータ、図 5.6(b) は女性のデータを対象に出力したものである。

5.2. GRADCAM++と注意マップの比較



(a) データ: 男性



(b) データ: 女性

図 5.6: GradCAM++と注意マップの比較 (男性-女性)

5.3. 変換前後での CNN の分類結果の違い

5.3 変換前後での CNN の分類結果の違い

各データセットのテストデータに対して AG-CycleGAN による変換を行い、変換前後の CNN の分類を出力した結果を記述する。

5.3.1 NIHCC データセット: 心肥大-検出なし

心肥大-検出なし間の変換前後のテストデータを対象に診断モデルによる分類を行った結果として表 5.2 に表す結果が得られた。

表 5.2: 変換前後の混同行列 (心肥大-検出なし)

		分類結果			
		変換前		変換後	
		検出なし	心肥大	検出なし	心肥大
ラベル	検出なし	9102	2971	4746	7327
	心肥大	188	368	386	170

5.3.2 NIHCC データセット: 気胸-検出なし

心肥大-検出なし間の変換前後のテストデータを対象に診断モデルによる分類を行った結果として表 5.3 に表す結果が得られた。

表 5.3: 変換前後の混同行列 (気胸-検出なし)

		分類結果			
		変換前		変換後	
		検出なし	気胸	検出なし	気胸
ラベル	検出なし	8655	3418	5795	6278
	気胸	352	709	622	439

5.3. 変換前後での CNN の分類結果の違い

5.3.3 IXI データセット: 男性-女性

IXI データセットにおける男性-女性間の変換前後のテストデータを対象に男女分類モデルによる分類を行った結果として表 5.4 に表す結果が得られた。

表 5.4: 変換前後の混同行列 (男性-女性)

		分類結果			
		変換前		変換後	
		女性	男性	女性	男性
ラベル	女性	51	15	50	16
	男性	5	48	5	48

第 6 章

追加実験

IXI データセット 男性-女性間の分類を学習したモデルの分析に対して実験条件の変更を行い、変更した条件で追加実験を行った内容と結果について記述する。

6.1 追加実験内容

IXI データセットでの男女間の分類を行った CNN を対象にした分析結果をもとに、前処理、識別モデルの構造、学習条件等の条件の変更を行った実験内容について記述する。

6.1.1 データセット

データセットでは同様に IXI データセット [17] を用い、前処理では標準化ではなく、 $[-1 \sim 1]$ 間へのスケーリングを行う。また学習に用いるデータの分割は同様に設定する。

6.1.2 識別モデル

同様に Batch Normalization の代わりに Spectral Normalization を導入した CNN を用い、図 6.1 に示すように特徴マップの縮小を 2 層ごとではなく 1 層ごとに各畳み込み層で行うように変更する。

6.1.3 対象となる CNN

同様に、IXI データセットの脳構造画像 (T1 画像) を用いて男性と女性間の 2 値分類において、CNN のオプティマイザーとして用いた SGD のハイパーパラメータである学習率、

6.1. 追加実験内容

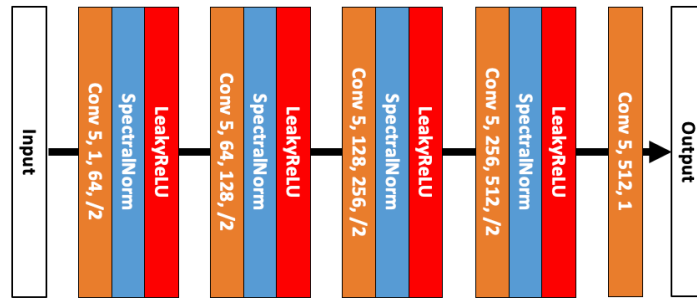


図 6.1: 変更後の識別モデルの構造

Momentum, Weight Decay の値を決定するために行った LR Range Test と Grid Search の結果を示す. ここで探索条件は NIHCC データセットでの実験と同様に設定した. 図 A.4 は LR Range Test と Grid Search の結果であり, この結果をもとにハイパーパラメータを表 6.1 で示す値に決定した. また, IXI データセットではデータの偏りが小さいためデータ分布の調整は行わずに実験を行った. 男女分類モデルの性能評価を表 6.2 に示す.

表 6.1: ハイパーパラメータ (IXI: 男性-女性, 追加実験)

パラメータ名	epoch	Batch size	Base lr	Max lr	Momentum	Weght Decay
数値	80	8	5×10^{-5}	4×10^{-4}	0.99	10^{-3}

表 6.2: モデルの性能 (IXI: 男性-女性, 追加実験)

	precision	recall	f1-score	support
男性	0.85	0.80	0.83	66
女性	0.77	0.83	0.80	53
accuracy			0.82	119
macro avg	0.81	0.82	0.81	119
weighted avg	0.82	0.82	0.82	119

6.2 実験手順

同様に、AG-CycleGAN の学習を始めるにあたり、男女分類モデルの特徴抽出部のパラメータを生成モデル・注意モデルの特徴抽出部に読み込ませる。その後、CNN の分析のために AG-CycleGAN による相互変換を学習させ、学習後の AG-CycleGAN から出力される注意マップと変換結果から分類に寄与する領域と領域内の形状やパターンの違いからクラス間の違いを捉えた結果が得られたか評価する。さらに男女分類モデルから得られる GradCAM++ と注意マップを比較し、注目部位の類似性や得られた結果の違いを評価する。最後に変換前後のテストデータと元データのラベルを用いて男女分類モデルに分類させ、AG-CycleGAN の変換の有効性を評価する。

AG-CycleGAN を用いて分析を行うにあたり、表 6.1 で示すようにハイパーパラメータの設定の変更を行った。表 6.3 中の「n_dis」は、学習の安定化のために生成・注意モデルと識別モデルのパラメータの更新回数の比を $G : D = 1 : [n_dis]$ となる様に設定するハイパーパラメータである。

表 6.3: IXI データセットの追加実験に用いるハイパーパラメータ (AG-CycleGAN)

データセット	epoch	Batch size	Optimizer	Learning rate	b1	b2	n_dis
IXI	500	1	Adam	2×10^{-4}	0.5	0.999	1

6.3 追加実験結果

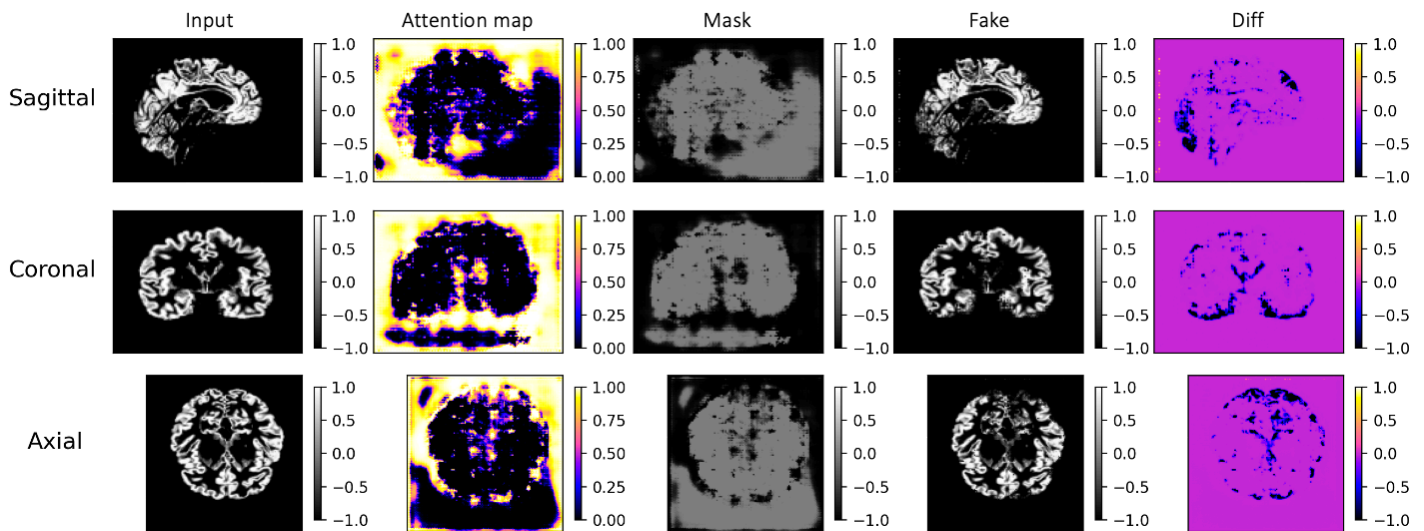
男女分類モデルに対して変更後の実験条件を用いて提案手法の検証を行った結果について記述する。

6.3.1 AG-CycleGAN による相互変換の学習

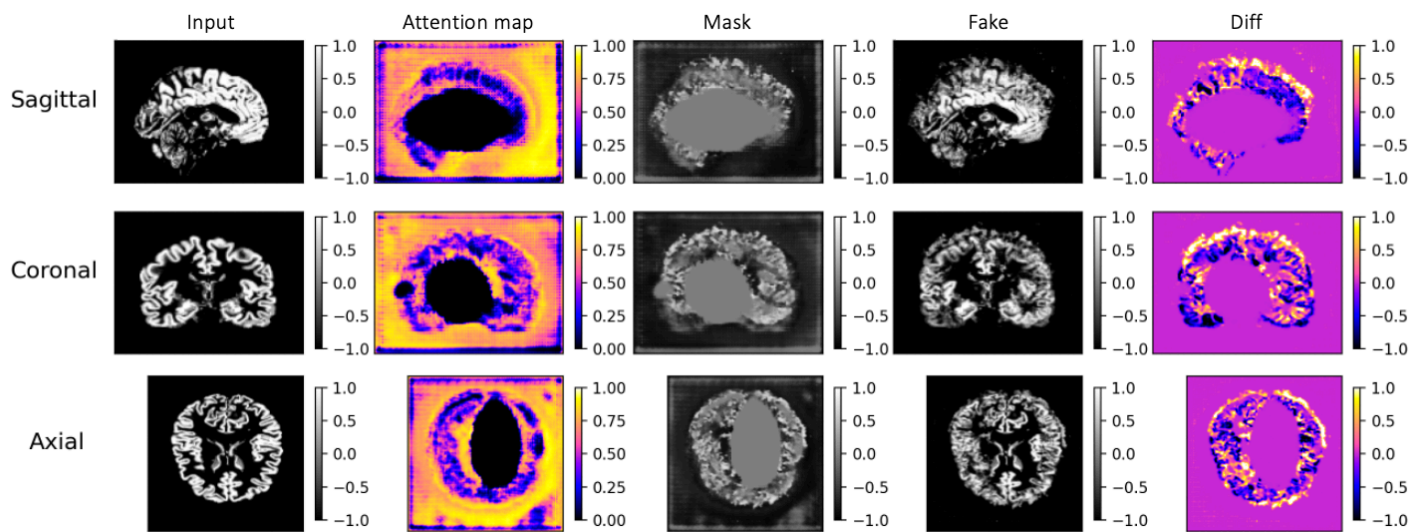
追加実験として IXI データセットにおける男性-女性間で分類を行った CNN と共通の特徴抽出部を用いて AG-CycleGAN に相互変換を学習させた結果として、図 6.2 に示す結果

6.3. 追加実験結果

が得られた・男性 → 女性の変換 (図 6.2(a)) では後頭葉と側頭葉の下部分を中心に脳の外側が削られるような変換が得られた。女性 → 男性の変換 (図 6.2(b)) では、同様に側頭葉の下部分が削られ、前頭葉を大きくするような変換が見られた。



(a) 男性 → 女性



(b) 女性 → 男性

図 6.2: 男性-女性間の相互変換の結果

6.3. 追加実験結果

6.3.2 GradCAM++と注意マップの比較

男性-女性間の分類を学習した診断モデルが正しく分類を行えたデータを対象に GradCAM++と注意マップを出力した結果として、図 6.3 に示す結果が得られた。図 6.3(a) は男性のデータ、図 6.3(b) は女性のデータを対象に出力したものである。

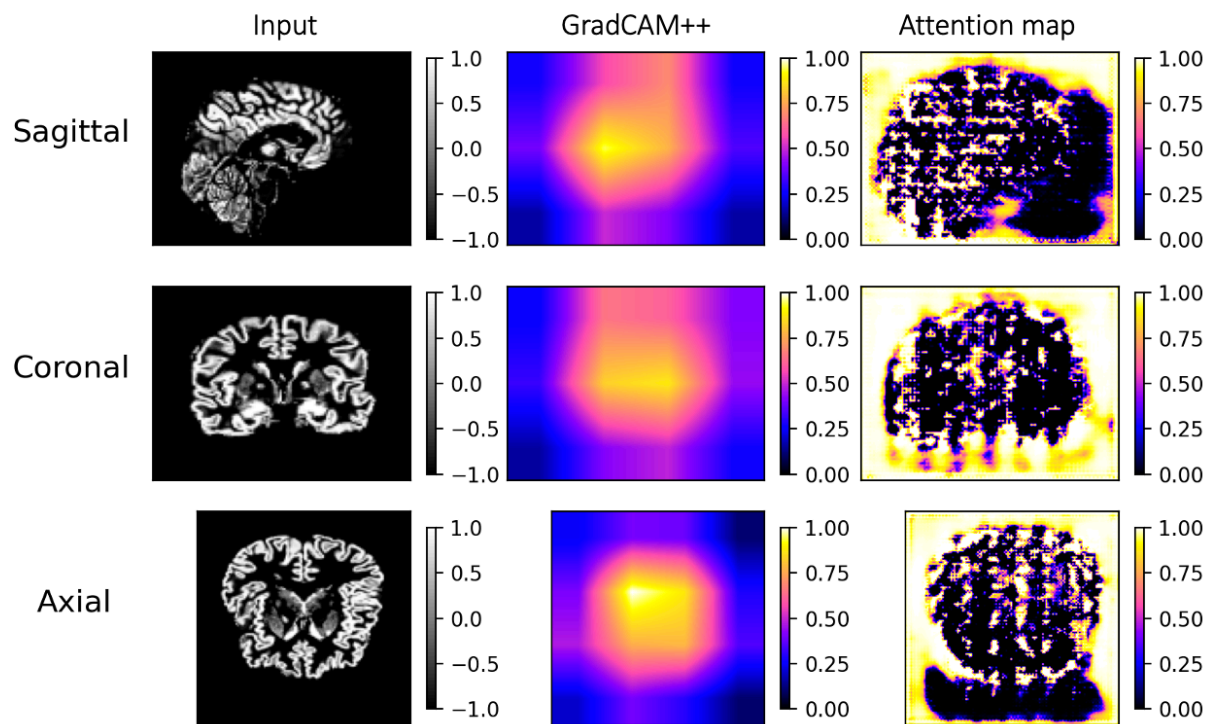
6.3.3 変換前後での CNN の分類結果の違い

IXI データセットにおける男性-女性間のテストデータを対象に変換前後の診断モデルによる分類を行った結果として表 6.4 に表す結果が得られた。

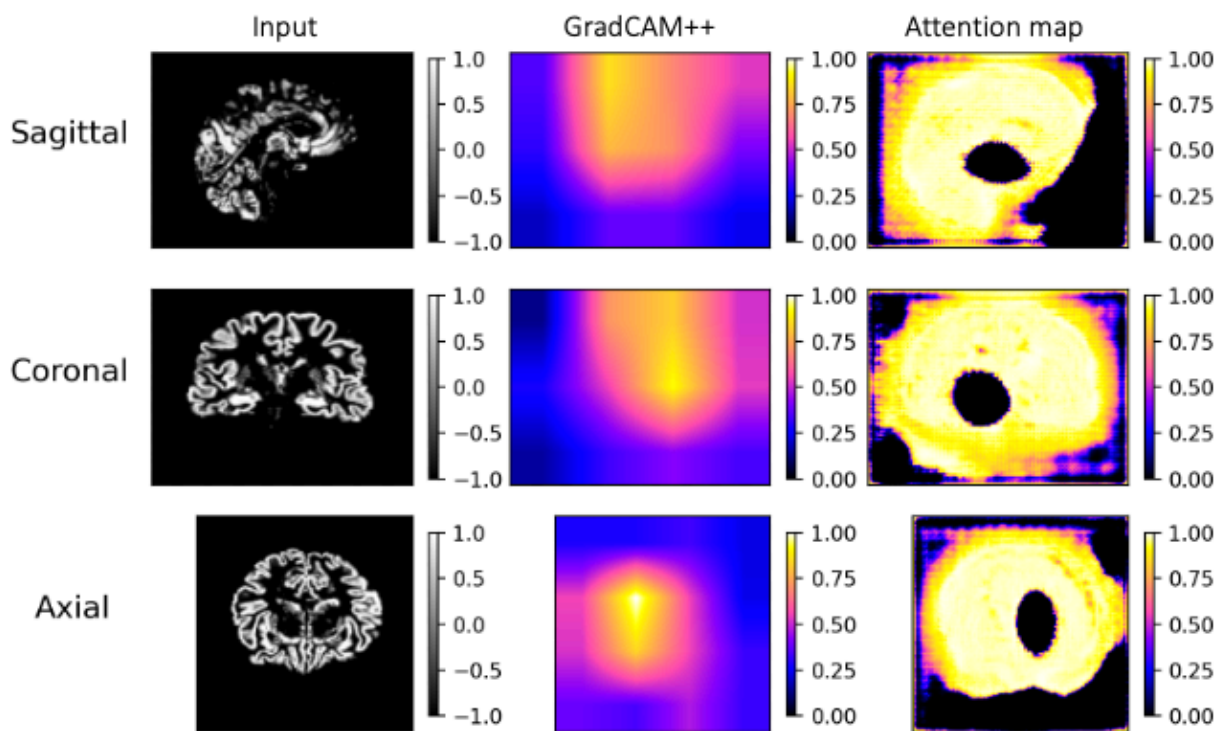
表 6.4: 変換前後の混同行列 (男性-女性, 追加実験)

		分類結果			
		変換前		変換後	
		女性	男性	女性	男性
ラベル	女性	53	13	66	0
	男性	9	44	42	11

6.3. 追加実験結果



(a) データ: 男性



(b) データ: 女性

図 6.3: GradCAM++と注意マップの比較 (男性-女性)

第7章

考察

7.1 AG-CycleGAN による相互変換

各データセットに対して AG-CycleGAN による相互変換を行った結果について考察した内容を記述する.

7.1.1 NIHCC データセット: 心肥大-検出なし

NIHCC データセットにおける心肥大-検出なし間の変換では, 分類に寄与する領域として心臓の側面が得られ, 心臓の大きさを表す部位を選択していると考えられる. 領域内の形状やパターンの違いとしては, 図 5.1(a) の心肥大 → 検出なしの場合では心臓を小さくするような変換, 図 5.1(b) の検出なし → 心肥大の場合では心臓を大きくするような変換が見られ, 心肥大の症状と一致する結果が得られたと考えられる. また, 説明性について考えると注意マップと変換前後の差分から心肥大という症状に対して説明性の高い結果が得られたと考えられる.

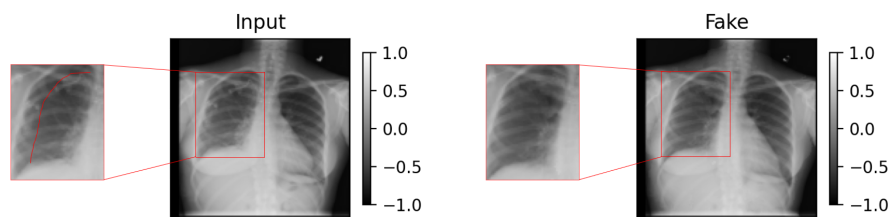
7.1.2 NIHCC データセット: 気胸-検出なし

NIHCC データセットにおける気胸-検出なし間の変換では, 分類に寄与する領域として肺領域の周辺が得られた. 領域内の形状やパターンの違いとしては, 図 5.2(a) の気胸 → 検出なしの場合では肺領域の周辺を暗くするような変換, 図 5.2(b) の検出なし → 気胸の場合では肺の一部を強調するような変換が見られた. X 線画像では空気の多い領域が暗く映る

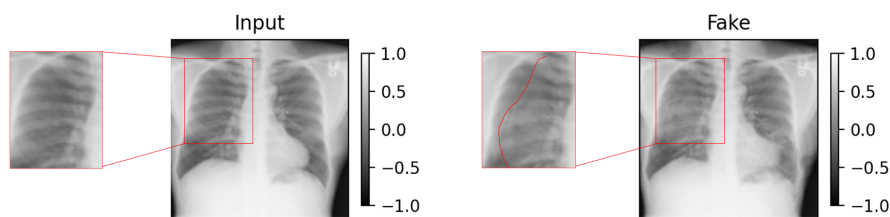
7.1. AG-CYCLEGAN による相互変換

ため、気胸によって漏れ出た空気が溜まる胸腔と空気の入った肺では気胸によって空気が漏れ出た領域の方がわずかに暗く映るが、鮮明に境界を見分けることは難しい。図 7.1 は今回の変換によって得られた胸腔と肺の境界と見られる部位を赤線で図示したものである。図 7.1(a) から気胸 → 検出なしの変換では境界が消えるように、図 7.1(b) では境界ができるように変換された可能性があると考えられる。

また説明性という観点で考えると、単純に領域だけを得られる CAM ベースの手法に比べ、変換前後の差分がクラス間の差を推測する材料として有効に機能していると考えられる。しかし心肥大と比べると説明性が低い結果が得られ、症状の判断の難しさが説明性にも影響する結果となった。今後は注意マップの表現方法として CAM ベースの手法でも用いられる $[0, 1]$ へのスケージングなど、説明のための分類に寄与する領域の獲得方法について検討する必要があると考えられる。クラス間の差の表現についても、変換前後の差分以外にも表現方法について検討する必要がある。



(a) 気胸 → 検出なし



(b) 検出なし → 気胸

図 7.1: 変換前後の気胸による胸腔と肺の境界と見られる領域

7.1. AG-CYCLEGAN による相互変換

7.1.3 IXI データセット: 男性-女性

IXI データセットにおける男性-女性間の変換では、分類に寄与する領域として脳の周辺部分の領域が得られたが、変換による変化が見られず領域内の形状やパターンの違いが得られなかった。その原因の 1 つとして図 7.2 に示す識別モデルの損失が学習途中で減少しなくなっていることが挙げられる。これは生成モデルが生成する変換結果と元のデータの区別ができていない状態であり、構造的に識別モデルの識別能力が足りていない可能性がある。また今回は学習回数の比を識別モデルと生成モデルで 1:1 に設定したがより効果的な比が無いか探索する必要もあると考えられる。

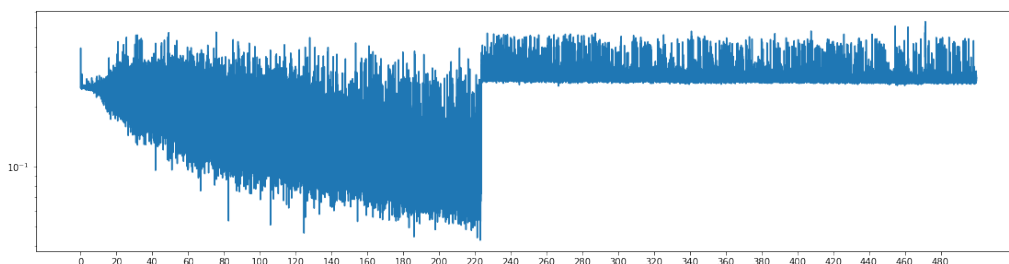


図 7.2: 識別モデルの損失 IXI データセット: 男性-女性

7.1.4 IXI データセット: 男性-女性 追加実験

前処理を標準化から $[-1 \sim 1]$ のスケーリングに、識別モデル内の特徴マップの縮小をより小さくした結果として、注意マップによる領域の選択と変換が学習された。これは標準化では平均 0、標準偏差 1 になるように画素値を変換するが、脳構造画像では画素値が 0 付近の領域が多く、主に $[0 \sim 1]$ の値域に脳領域を表現する画素値が集まっていたが、スケーリングにより $[-1 \sim 1]$ の値域に画素値が分散されるようになったことが変換をより容易にすることにつながった可能性があると考えられる。特徴マップの縮小では、最終畳み込み層に入力される特徴マップの各画素が元画像に対して広い受容野を持つことにつながり、脳全体の形状についての判定が可能になったと考えられる。このことから本手法を用いるにあた

7.2. AG-CYCLEGAN の学習

り、対象となる CNN と識別モデルが同じ広さの受容野を持つようにモデルを調整する必要があると考えられる。

次に変換結果について考える。IXI データセットにおける男性-女性間の変換でも同様に分類に寄与する領域として脳の周辺部分の領域が得られた。そして、変換結果から男性 → 女性への変換では後頭葉と側頭葉の下部分を削り、全体として脳が前方に回転するような変換が行われたと考えられる。女性 → 男性への変換では前頭葉を大きくし側頭葉の下部分を削るような変換が行われ、全体として脳が後方に回転するような変換が行われたと考えられる。また女性 → 男性の変換では、脳の中心部分が変換されていないが、男性 → 女性の変換では変換されているため、実際に脳の中心に違いがないと推測できる結果は得られていないと考えられる。その為、一貫性のある結果が得られるようにモデルや学習条件を調整し、検証を進めていく必要があると考えられる。

7.2 AG-CycleGAN の学習

全体的な相互変換の結果として、心肥大では症状に沿う有効な結果が得られたが、気胸では説明性の低い結果、男女の脳構造画像では CNN の説明に有効な変換結果が得られなかった。これは本研究では GAN のモデルとして用いた AG-CycleGAN の損失関数に用いる Cycle Consistency Loss が変換の一貫性を保持するとともに、変換による画像の変化を抑制する方向に働いた影響が一因として考えられる。今回の場合では、内臓の部位として大きい肺や脳画像全体の変換の際には変換が強く働かなかった。変換によるデータの変化が起きた際に発生する損失には平均絶対誤差とも呼ばれる L1 損失が用いられており、データを大きく変換する際には Cycle Consistency Loss による変換の抑制が強く働く。逆に心肥大のようにデータ全体から見て局所的な変換では、変換の抑制より識別モデルを騙すことが優先され、注意マップによる領域の抽出と変化が有効に働いたと考えられる。そのため、症状が現れる部位の大きさによって変換の容易さが変化しない様に損失の計算を工夫する必要があると考えられるが、その際には変換の一貫性を損なわないように注意する必要がある。ま

7.2. AG-CYCLEGAN の学習

た、クラス間の違いを最大化する対象学習を用いた教師なし相互変換を学習する CUT[18] を用いるなど、データ間の違いを得るために損失関数が変換を抑制しない手法の利用も検討する必要があると考えられる。加えて GAN の学習の安定化について考えると、本研究で用いた AG-CycleGAN では注意モデルが導入されており、注意モデルによる領域の選択に大きな変更が行われた際に図 7.3 の様に損失が大きく変化する箇所が発生する。これにより、GAN の安定化を目指す上で 1 つの指標となるリップシツ連続を満たさない学習となり、安定的な学習が行われていない可能性がある。その為、変換学習と領域選択の学習を独立させるなど、安定化に向けてモデルや損失を工夫する必要があると考えられる。

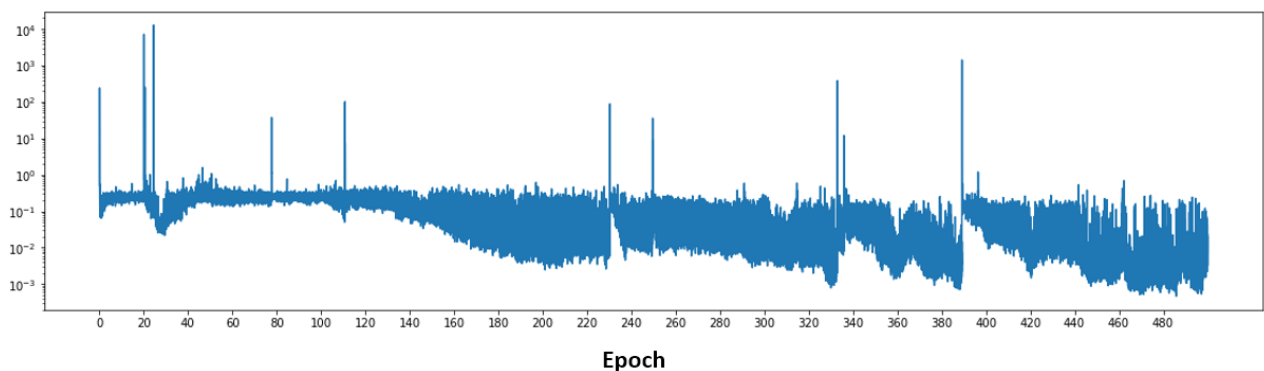


図 7.3: 識別モデルの損失 IXI データセット: 男性-女性 追加実験

また、相互変換を学習させるうえで行う AG-CycleGAN では最適なハイパーパラメータの設定が難しく、今回であれば胸部 X 線画像の場合は識別モデルと生成・注意モデルのパラメータ更新頻度の比が 5:1 の場合で良い生成結果が得られたが、脳画像では 5:1 では生成・注意モデルが十分に学習できず、今回は 1:1 で学習を行った。これらの値の決定には CNN の学習とは異なる視点で損失や生成結果を見て定めることが必要である。今後 GAN を用いた分析手法が一般に用いられるにはこれらの値の決定方法を一般化する手法、動的に調整するような手法について今後検討する必要がある。

加えて、本研究では分析対象として Resnet18 を用いたが、SOTA(State-of-the-Art) モデルと呼ばれるような現時点で最先端な精度を誇るモデルは更新され続けており、実際の CNN の診断に用いるモデルにはより発展したモデルが選ばれると考えられる。その為、よ

7.3. 注意マップと GRAD-CAM++の比較

り複雑なモデルに対しても本手法が有効かを検討することが必要になると考えられる。

7.3 注意マップと Grad-CAM++の比較

NIHCC データセットにおける心肥大-検出なし間では、Grad-CAM++では心臓周辺を注目したという情報のみが得られたが、注意マップでは心臓側面という詳細な部位まで得られた。気胸-検出なし間においても、Grad-CAM++では肺や心臓が集まる領域を注目したという情報のみが得られたが、注意マップでは肺領域とその周辺に限定された領域が得られた。しかし、心肥大に比べると領域の強調が弱く出力された。これは注意マップが変換領域を制限すると同時に変換にも寄与するため、変換の際に値を強く動かさないために強く強調した結果を出力しなかった可能性があると考えられる。脳構造画像における男性-女性の注意マップでは脳以外の画像全体、追加実験では脳の外側を中心に領域が選択されたため、脳の形や向きが性差に影響している可能性があると考えられる。脳の形や向きに性差があると仮定すると、形や向きを整えない場合は特定の部位の性差にあるとしても先に形や向きが優先されて変換され、抽出できない可能性がある。そのため、特定の部位を抽出したデータを用いる方法、標準脳に合わせるだけでなく形や向きを矯正できる方法を検討する必要がある。また、脳の形を矯正する場合はバイオマーカの様な被験者固有の情報が消えないように注意する必要があると考えられる。

また、今回の結果では心肥大や気胸の様な有効な変換が学習された場合は Grad-CAM++と比べて詳細な分類に寄与する領域、領域内形状やパターンの違いで説明性の高い結果を得ることができ、変換を学習できなかった場合もクラス間の違いが一部か全体かを認識できる点で Grad-CAM++と比べて説明性の高い結果が得られたと言えると考えられる。しかし、安定性では AG-CycleGAN から有効な出力が得られるかは GAN の学習に依存しており、計算資源の点では AG-CycleGAN は 6 つのモデルから構成されるため、必要な計算資源が Grad-CAM++と比べて非常に多い。特に学習時よりも多くの GPU メモリが必要になる点で、学習に用いたコンピュータで確実に実行できる Grad-CAM++の方が計算資源の点で

7.4. 変換前後での CNN の分類結果の違い

は優れている。また、一度学習ができればデータ 1 つあたりの結果の出力には実用上大きな時間差は生まれないが、学習の必要がない Grad-CAM++の方が結果の獲得までの時間が短い。これらの点から、省メモリ化と高速化に向けて 2 つの生成モデルと 2 つの注意モデルの合計 4 つのモデルが共通して用いるバックボーンを 1 つに集約するなど、モデル構造に対する工夫が必要である。表 7.1 は GradCAM++と AG-CycleGAN の比較結果を簡単に示したものである。

	GradCAM++	AG-CycleGAN
説明性 (領域)	△ 大まかな領域	○ 詳細な領域
説明性 (形状, パターン)	× 獲得不可能	○ 変換の差より獲得可能
安定性	○	× GAN の学習に依存
計算資源	○ 元モデルと同等	× 元モデルより多い
実行時間 (全体)	○ 1 分未満	× 2~3 日
実行時間 (出力のみ)	○ 18.5ms	△ 23ms

表 7.1: GradCAM++と AG-CycleGAN の比較

7.4 変換前後での CNN の分類結果の違い

NIHCC データセットにおける心肥大-検出なし間、気胸-検出なし間では、変換によって CNN の分類結果の傾向が逆になることから、クラス間を遷移させるような違いが得られていると考えられる。脳構造画像の男性-女性間では追加実験の結果を見ると男性 → 女性の変換では分類結果の傾向が変わるような変換が行われている。しかし、女性 → 男性の変換では反対に全てのデータが女性と識別されるような変換が行われており、変換の学習不足や一貫性を考慮してモデルや学習条件を検討する必要があると考えられる。また、この混同行列についても現段階では変換前後の傾向の違いから判断しており、有効な変換が得られたと考えられる変化量について定量的な基準が必要だと考えられる。

7.4. 変換前後での CNN の分類結果の違い

加えて、今回変換の有効性の検証ために、AG-CycleGAN による変換前後のデータセットを用いて CNN モデルの分類結果の違いをみた際に変換前後でも分類クラスが不変なデータがあった。これらのデータに対して、GAN と CNN で共通の特徴抽出部を用いたことから、CNN が分類を苦手とするデータと GAN が変換を苦手とするデータが類似する可能性があると考えられる。提案手法の分析手法としての妥当性を調べるためにそれらについても今後調査する必要があると考えられる。

第 8 章

結論

本研究では CNN が分類過程の説明可能性の向上を目指して、AG-CycleGAN を用いた分析手法を提案し、説明に必要と考えられる分類に寄与する領域と領域内の形状やパターンの違いの獲得を目指した。胸部 X 線画像と脳構造画像を対象に検証を行った結果、心肥大では症状に沿った変換が行われ、Grad-CAM++と比較して詳細な結果が得られた。混同行列からも CNN の分類に有効な変換が行われ、説明性の高い結果が得られたと考えら得る。しかし、気胸では有効な変換が行われたが説明性が低く、分類に寄与する領域の獲得方法やクラス間の差の表現について検討する必要があると考えられる。さらに脳構造画像を用いた実験では女性 → 男性の変換では有効な変換が行われず、一貫性のある結果が得られなかった。今後は手法の安定的な学習のために、変換の抑制に働く Cycle Consistency Loss への工夫やリプシッツ連続を満たすような構造の検討を進める必要があると考えられる。結果の表現方法についても、混同行列に対して有効な変換が行われたことを判断する定量的な変化量の定め方、元々の違いの判断の難しさが影響しない形での表現方法などを検討していく必要があると考えられる。

謝辞

本研究を進める上でご指導をしてくださった吉田真一教授に御礼申し上げます。研究の方向性に悩んだ際に吉田先生が方向性を示してくださったお陰で本研究に集中して取り組むことができました。また、最初に GAN というテーマに取り組みたいと話をした際に否定するのではなく、GAN の活用に向けて積極的に支援してくださったことで最後まで本研究を進めることができました。吉田先生がつくる「吉研」という空間のおかげで、楽しく充実した研究室生活を送ることができました。大変お世話になりました。

また、副査を引き受けてくださった岩田誠教授と竹内聖悟講師へも御礼申し上げます。学部の卒業研究の頃から、忌憚のない質問や意見、助言で私の理解を深め、研究について考える上で必要な視点を与えてくださったことが本研究の進展をする上で良い影響となりました。頂いた助言の中でやり遂げられなかったことがありますが、研究の進展のために後輩に共有できる様に引き継いでいきたいと思えます。

研究室で自分が行き詰まるたびに相談に乗ってくださった四宮友貴助教へも御礼申し上げます。幅広い知見からご助言くださり、研究に取り組む上で新たな視点が得ることができました。また、優しいながらも忌憚のない意見をぶつけてくださり、研究の本質的な部分について考える機会を与えて頂きました。大変お世話になりました。

ここまで自分を支えてくださった両親にもお礼申し上げます。大学院に進学する際に一度は反対されましたが最後には許して下さい、大学院での勉強に集中できるように様々な面で支えてくださったお陰で大きな不安を抱えることなく、勉学に集中することができました。本当にお世話になりました。

参考文献

- [1] Bo Zhou Daozheng Chen Jun Gao, Qian Jiang. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview. In *Mathematical Biosciences and Engineering*, Vol. 16, pp. 6536–6561, 2019.
- [2] R. Manickavasagam, S. Selvan, and Mary Selvan. Cad system for lung nodule detection using deep learning with cnn. *Medical & Biological Engineering & Computing*, Vol. 60, No. 1, p. 221–228, 2021.
- [3] Dunya Alkurdi, Muhammad Ilyas, and Akhtar Jamil. Cancer detection using deep learning techniques. *Evolutionary Intelligence*, pp. 1–9, 07 2021.
- [4] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, 2016.
- [5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, Vol. 27, pp. 2672–2680. Curran Associates, Inc., 2014.
- [6] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *The Institute of Electrical and Electronics Engineers(IEEE) International Conference on Computer Vision (ICCV)*, Oct 2017.
- [7] Youssef Alami Mejjati, Christian Richardt, James Tompkin, Darren Cosker, and Kwang In Kim. Unsupervised attention-guided image-to-image translation. In *Advances in Neural Information Processing Systems*, Vol. 31, pp. 3693–3703. Curran Associates, Inc., 2018.

参考文献

- [8] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847, 2018.
- [9] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *The Institute of Electrical and Electronics Engineers(IEEE) International Conference on Computer Vision (ICCV)*, Oct 2017.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [11] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 of *LNCS*, pp. 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).
- [12] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018.
- [13] Martín Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.
- [14] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, Vol. 13, No. 10, pp. 281–305, 2012.

参考文献

- [15] Leslie N. Smith. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472, 2017.
- [16] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3462–3471, 2017.
- [17] Imperial College London Biomedical Image Analysis Group. Ixi dataset – brain development, 2022. <https://brain-development.org/ixi-dataset/>.
- [18] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*, 2020.

付録 A

ハイパーパラメータ探索の結果

以下に LR Range Test と Grid Search の結果を示す。ここで各結果において、グラフの上のタイトルは”SGD_[Weight decay の値]”を示し、各グラフ中の各線は Momentum の値ごとの Loss の変化を表す。

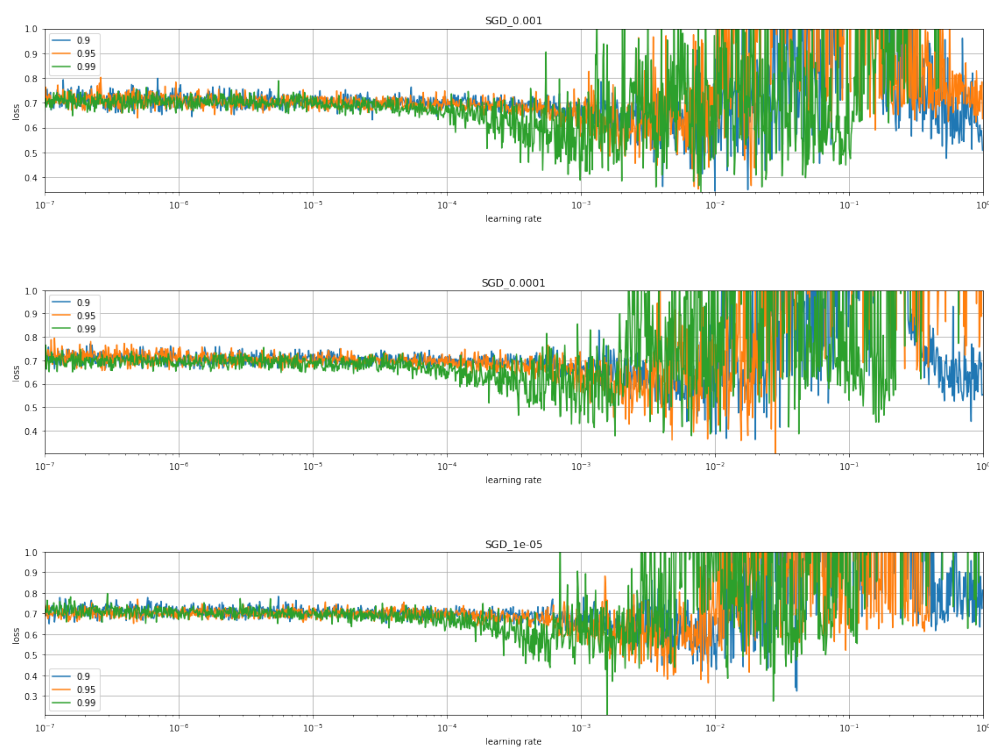


図 A.1: LR Range Test + Grid Search の結果 (NIHCC: 心肥大-検出なし)

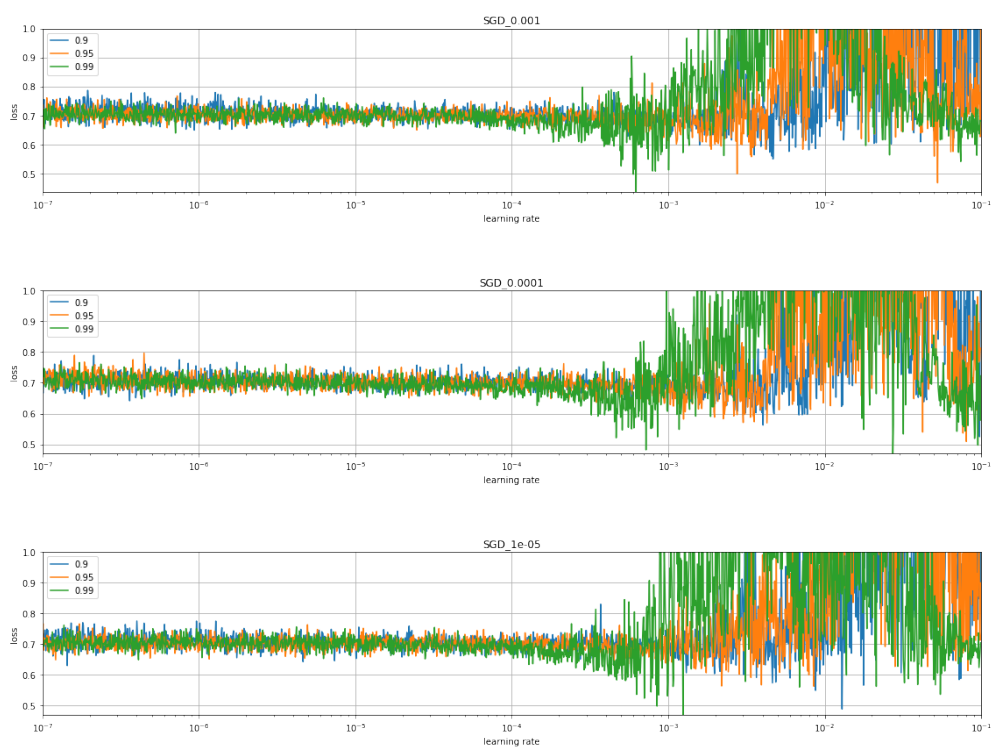


図 A.2: LR Range Test + Grid Search の結果 (NIHCC: 胸部肥厚-検出なし)

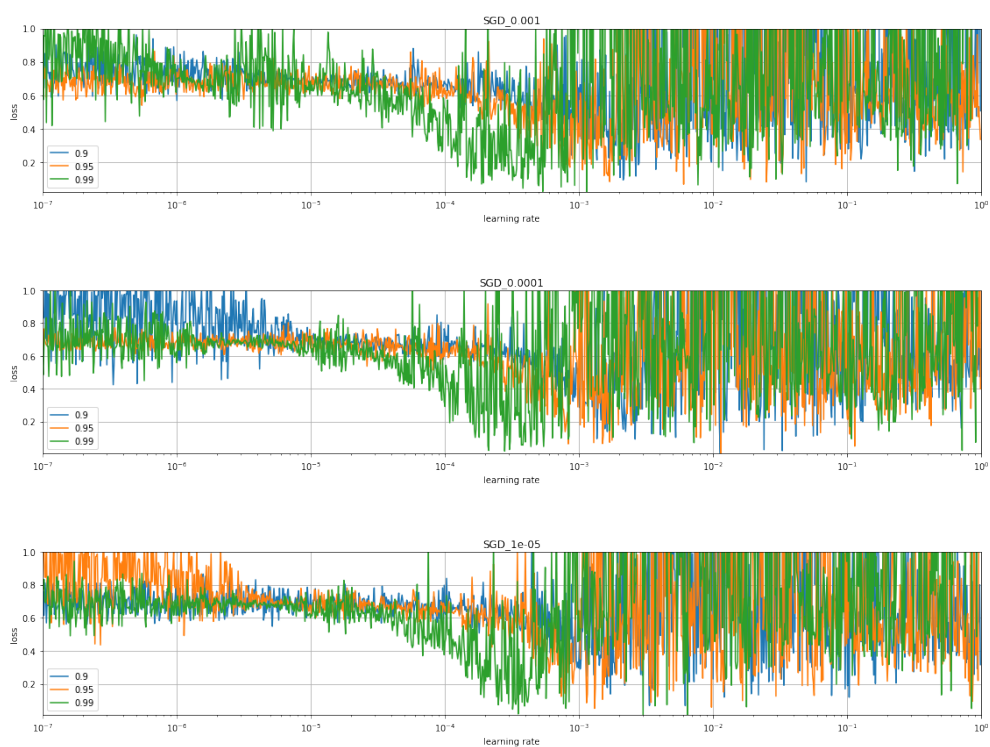


図 A.3: LR Range Test + Grid Search の結果 (IXI: 男性-女性)

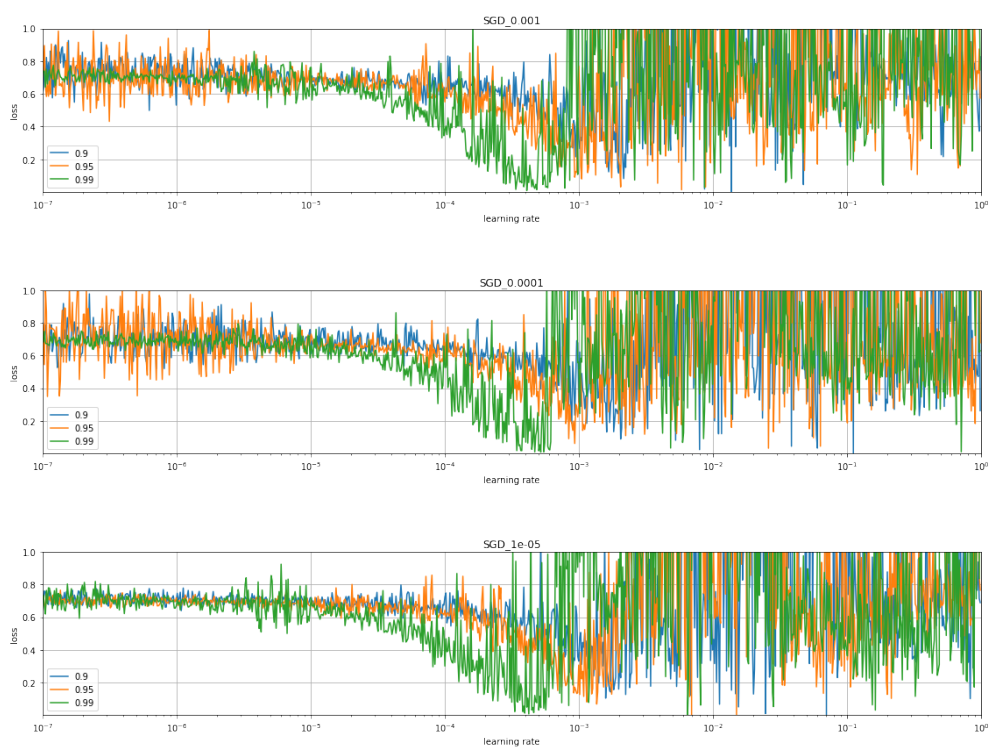


図 A.4: LR Range Test + Grid Search の結果 (IXI: 男性-女性) 追加実験