

令和 5 年度
修士学位論文

MRI 脳画像からの CNN による年齢推定 モデルの説明性

Explainability of Age Estimation Model Using CNN
from Brain MRI Images

1265111 吉田 翔真

指導教員 吉田 真一

2024 年 2 月 2 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

MRI 脳画像からの CNN による年齢推定モデルの説明性

吉田 翔真

近年, Convolutional Neural Network(CNN) を用いた画像認識による医療診断支援システム Computer Aided Diagnosis(CAD) の研究が盛んである. しかし, CNN を用いた診断ではその診断の説明をすることが困難であり, 現場への導入が進んでいない. 世界的な高齢化社会の進行で, 認知症等の脳の疾患の早期発見が重要である. 脳の年齢の推定はこれらの問題解決の基礎となるものと考えられており, CNN を用いた年齢推定モデルも提案されている. しかしこれまで CNN の分類結果の判断には, 分類に寄与する領域を可視化する Class Activation Map(CAM) が用いられていた. しかし CAM ベースの可視化手法では画像内の判断に寄与したおおまかな位置が分かるだけで, 詳細な画素やその形状の違いやパターンの違いなど, 詳細な情報を取得することが出来ない.

新しい CNN の説明手法として Generative Adversarial Network (GAN) を用いた方法が提案され, 脳画像からの性別, 肺疾患への応用可能性が示されている [1]. そこで本研究では CycleGAN を用い, CycleGAN から得られた相互変換の結果から領域内の形状の違いやパターンの違いを取得することを目的とする. また, 変換前後の分類結果の違いから変換の有用性を検証する.

MRI 脳画像における若年層, 老年層での相互変換を行った結果, 老年層 → 若年層の変換画像は頭頂葉, 前頭葉外側が変化していることが確認できた. 若年層 → 老年層の変換画像では側頭葉や小脳の変化も確認できた. また, 変換後の画像を用いて分類を行った結果, 分類結果も変化していることが確認できた. しかし若年層 → 老年層の変換では変換後の画像が複製されている問題が見られた. 今後は GAN が画像を複製してしまう問題であるモード崩壊について検討する必要があると考える

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, 医療診断支援システム

Abstract

Explainability of Age Estimation Model Using CNN from Brain MRI Images

YOSHIDA, Shoma

In recent years, computer aided diagnosis (CAD) systems for medical diagnosis support using image recognition with convolutional neural networks (CNN) has been actively studied. However, explaining the diagnosis in CNN-based systems is challenging and it is difficult to use in the medical field. The global trend of an aging population emphasizes the importance of early detection of brain disorders such as dementia. Estimating the age of the brain is considered as a foundation to addressing these issues, and age estimation models using CNNs have been proposed.

Previously, class activation maps (CAM) have been used to visualize the regions contributing to CNN classification results. However, CAM-based visualization methods only provide a rough understanding of the areas influencing the decision in the image. They fail to capture detailed information such as pixel differences, shapes, and pattern variations.

A new explanation method for CNNs has been proposed using generative adversarial networks (GANs). This approach has shown applicability in gender prediction and lung disease detection from brain images. In this study, CycleGAN is employed to extract differences in shapes and patterns within regions from the results of mutual transformation. The research aims to verify the usefulness of the transformations based on differences in classification results before and after transformation.

Results of mutual transformations between young and elderly age groups in MRI brain images reveal changes in the parietal lobe and outer frontal lobe in the elderly-to-young transformation. The young-to-elderly transformation shows changes in the temporal lobe and cerebellum. Classification results using the transformed images indicate variations in the outcomes. However, an issue observed in the young-to-elderly transformation is the duplication of images after transformation. Future considerations should address the problem of mode collapse in GANs, where the network duplicates images.

key words generative adversarial network, convolutional neural network, CycleGAN, explainable artificial intelligence, computer-aided diagnosis

目次

第 1 章	はじめに	1
第 2 章	関連研究	3
2.1	CycleGAN を用いた CNN の分類結果の説明性に関する研究	3
2.2	敵対的生成ネットワーク (GAN) モデル	4
2.2.1	敵対的生成ネットワーク (GAN)	4
2.2.2	CycleGAN	4
2.3	Class Activation Mapping(CAM)	5
2.4	Mode Seeking GAN(MSGAN)	6
第 3 章	提案手法	7
第 4 章	実験内容	9
4.1	データセット	9
4.2	機械学習モデル	10
4.2.1	CNN モデル	10
4.2.2	CycleGAN モデル	11
4.2.3	差分画像	11
第 5 章	実験結果	12
5.1	CycleGAN による相互変換	12
5.1.1	灰白質画像	12
5.1.2	白質画像	15
5.1.3	脳領域画像	17
5.2	CNN を用いた変換前後での精度の違い	19

目次

5.2.1	灰白質画像	19
5.2.2	白質画像	19
5.2.3	脳領域画像	20
5.3	CAM ベースの手法との比較	20
5.4	モード崩壊	21
第 6 章	追加実験	23
6.1	追加実験内容	23
6.2	追加実験結果	23
6.2.1	Mode Seeking 正則化項を追加した CycleGAN による相互変換 . . .	23
6.2.2	CNN を用いた変換前後での精度の違い	25
第 7 章	考察	27
7.1	CycleGAN による相互変換	27
7.1.1	灰白質画像	27
7.1.2	白質画像	27
7.1.3	脳領域画像	28
7.2	CNN を用いた変換前後での精度の違い	28
7.2.1	灰白質画像	28
7.2.2	白質画像	28
7.2.3	白質画像	29
7.2.4	脳領域画像	29
7.3	CAM ベースの手法との比較	29
7.4	Mode Seeking 正則化項を追加した CycleGAN による相互変換と年齢推定 の精度の違い	30
第 8 章	おわりに	32

目次	
謝辭	33
参考文献	34
付録 A	36
付録 B	37

目次

2.1	GAN 上でのデータ遷移	4
3.1	CycleGAN による画像変換と差分による説明性	7
3.2	変換した脳画像を用いての年齢分類	8
4.1	CNN モデル	10
4.2	CNN モデル	11
5.1	老年層 → 若年層の変換結果と差分	13
5.2	若年層 → 老年層の変換結果と差分	14
5.3	老年層 → 若年層の変換結果と差分	15
5.4	若年層 → 老年層の変換結果と差分	16
5.5	老年層 → 若年層の変換結果と差分	17
5.6	若年層 → 老年層の変換結果と差分	18
5.7	差分画像と GradCAM の比較	21
5.8	差分画像と GradCAM の比較	22
5.9	差分画像と GradCAM の比較	22
6.1	老年層 → 若年層の変換結果と差分	24
6.2	若年層 → 老年層の変換結果と差分	25
7.1	老年層 → 若年層の変換後と変換前の画像	30
7.2	若年層 → 老年層の変換後と変換前の画像	31

表目次

4.1 IXI データセットの分割内容	9
4.2 ネットワーク構成	10
5.1 変換前後の分類結果の混同行列	19
5.2 変換前後の分類結果の混同行列	19
5.3 変換前後の分類結果の混同行列	20
6.1 変換前後の分類結果の混同行列	26

第 1 章

はじめに

近年、医療現場では人手不足により医師の負担が増える問題がある。そのため Computer Aided Diagnosis(CAD) の研究が盛んに行われており、Convolutional Neural Network (CNN) を用いた画像認識による診断が多く行われている。しかし CNN を用いた医療画像診断では診断過程の説明が困難という課題があり、患者への説明が行えないため現場への導入が進んでいない。高齢化社会の進行で認知症等の脳の疾患の早期発見が重要になってきている。健康な脳の推定年齢と実年齢の差は脳疾患と関係があるとされており Cole ら [2] は年齢の推定を行う CNN のモデルを提案した。Cole らは入力画像として T1 強調画像の前処理をしていない生のデータと前処理された白質画像、灰白質画像を使用している。使用したデータの枚数は 2001 枚で年齢の幅は 18 歳から 90 歳、平均年齢 36.95 で標準偏差が 18.12 である。結果として白質画像で平均絶対誤差が 5.14 歳、灰白質画像で 4.16 歳、白質画像と灰白質画像で 4.34 歳、生のデータで 4.65 歳であった。健常者の年齢は正確に予測することができるかとされており、年齢の予測は脳の加齢過程や認知加齢、神経変性、加齢性脳疾患との関連性を調べることができるとされている。

今までの CNN の診断結果の説明には学習したモデルの重みや勾配を使用して分類に寄与する領域を可視化する Class Activation Map (CAM) ベースの手法が多く用いられてきた。しかし分類に寄与した領域のみでは説明が不十分であり、分類に寄与する領域の詳細な形状の違いやパターンを取得する必要があると考える。

そこで新しい CNN の説明手法として Generative Adversarial Network (GAN) を用いた方法が提案され、脳画像からの性別、肺疾患への応用可能性が示されている [1]。この手法は GAN を用いて脳画像で男女、胸部画像で疾患ありと検出なしの相互変換を行い変換

前後の画像から差分画像を生成する．生成した差分画像から詳細な形状の違いやパターンの取得を可能としている．そこで本研究では CycleGAN を用い，CycleGAN から得られた相互変換の結果から領域内の形状の違いやパターンの違いを取得することを目的とする．CycleGAN はデータ X をデータ Y に変換し，変換したデータ Y がデータ X に戻るようデータドメイン間の関係を学習する教師なしの生成モデルである．データ Y に対しても同様の学習を行う．

本研究では相互変換を行った画像から領域の形状の違いやパターンの違いの獲得を目標とする．変換を行った画像と変換前画像の差分を取り，変換した領域の詳細な違いを獲得する．また，変換の有効性を確認するために変換前後の画像を用いた CNN による識別精度の違いを確認する．

第 2 章

関連研究

本章では関連する研究と、本研究で使用した畳み込みニューラルネットワーク (CNN: Convolutional Neural Network), 敵対的生成ネットワーク (GAN: Generative Adversarial Network) とクラス活性化マップ (CAM: Class Activation Map) について, 記述する.

2.1 CycleGAN を用いた CNN の分類結果の説明性に関する研究

筒井の研究 [1] では CNN の分類結果の説明のために, GAN を用いて分類に寄与する領域の形状の違いやパターンを獲得する手法を提案した. この研究では 112,120 枚の胸部 X 線画像と 590 枚の MRI 脳画像を使用し, 胸部 X 線画像では心肥大と疾患が検出されない 2 クラス, MRI 脳画像は男性と女性の 2 クラスでの画像変換を行った.

心肥大-疾患なしのタスクでは心肥大の症状に沿った変換が行われ, 可視化手法である Grad-CAM++ と比較して詳細な結果を得ることができた. 変換前後の画像を用いた心肥大-疾患なしの分類でも CNN の分類に有効な変換が行われたことから説明性の高い結果が得られた.

脳 MRI 画像を用いた男性-女性間の変換タスクでは一貫性のある結果が得られず, 有効な変換が行われない結果となった.

2.2. 敵対的生成ネットワーク (GAN) モデル

2.2 敵対的生成ネットワーク (GAN) モデル

2.2.1 敵対的生成ネットワーク (GAN)

GAN[3] は 2014 年に Goodfellow らが提案した新しい生成モデルのフレームワークである。図 2.1 のような構造をしており、生成器 G と識別器 D から構成される。生成器 G は入力されたノイズ z から本物のデータ x に似たデータを生成する。識別器 D は生成器 G が生成した偽のデータと本物のデータ x を本物か偽物かを判断する。これを繰り返すことで生成器は本物に近いデータを生成することができる。

式 2.1 は GAN で用いられている損失関数である。第 1 項は D が元のデータ x を正しく識別しているかを判断するのに用いられ、第 2 項は G によって生成されたデータを D が正しく偽物と判断するために用いられる。

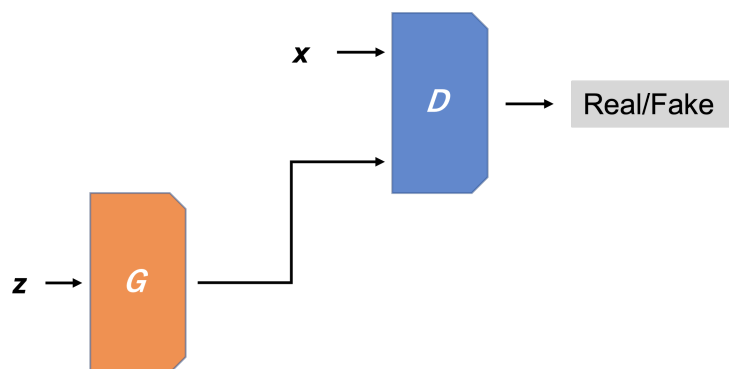


図 2.1: GAN 上でのデータ遷移

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

2.2.2 CycleGAN

CycleGAN[4] は Zhu らが提案した教師なしで学習する GAN モデルである。CycleGAN はデータドメイン X , Y が与えられたときドメイン間の写像 $G: X \rightarrow Y$ と $F: Y \rightarrow X$ を学習する。そして $F(G(x)) \approx x$ および $G(F(y)) \approx y$ となるようにサイクル一貫性損失を追

2.3. CLASS ACTIVATION MAPPING(CAM)

加することで対応していない画像と画像の変換が可能となる．また識別器 D_x , D_y と導入し, D_x は画像 $\{x\}$ と $\{F(y)\}$ を区別し, 同様に D_y は画像 $\{y\}$ と $\{G(x)\}$ を区別する．

CycleGAN の損失関数は式 2.2 で表される． L_{GAN} は GAN で用いられた損失関数であり, 式 2.3 及び式 2.4 で表される． L_{cyc} は変換の一貫性を保持するサイクル一貫性損失を示し, 式 2.5 で表される．サイクル一貫性損失はドメイン X の各画像 x に対して $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ となるように計算される損失である．

$$\begin{aligned} L(G, F, D_x, D_y) = & L_{GAN}(G, D_y, X, Y) \\ & + L_{GAN}(F, D_x, Y, X) + \lambda L_{cyc}(G, F) \end{aligned} \quad (2.2)$$

$$\begin{aligned} L_{GAN}(G, D_y, X, Y) = & \mathbb{E}_{y \sim p_{data}(y)} [\log(D_y(y))] \\ & + \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_y(G(x)))] \end{aligned} \quad (2.3)$$

$$\begin{aligned} L_{GAN}(F, D_x, Y, X) = & \mathbb{E}_{x \sim p_{data}(x)} [\log(D_x(x))] \\ & + \mathbb{E}_{y \sim p_{data}(y)} [\log(1 - D_x(F(y)))] \end{aligned} \quad (2.4)$$

$$\begin{aligned} L_{cyc}(G, F) = & \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] \\ & + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \end{aligned} \quad (2.5)$$

2.3 Class Activation Mapping(CAM)

CNN はなぜそのような結果になったのかが説明できないブラックボックス性の問題がある．この判断の説明のために Bolei[5] らは Class Activation Mapping(CAM) を提案した．CAM は学習済みの CNN のモデルの最後の畳み込み層の特徴マップと最終層の重みや勾配から分類に寄与した領域を顕著性マップとして算出する．しかし CAM はモデルの

2.4. MODE SEEKING GAN(MSGAN)

制約が必要な関係上、物体検出などのタスクでは使用できない問題が挙げられる。そこで Ramprasaath ら [6] は学習時の重みを使用せず、逆伝播時の勾配のみを用いることでモデルの制約を緩和した Grad-GAN を提案した。

2.4 Mode Seeking GAN(MSGAN)

GAN は訓練中にデータの多様性を捉えることができず、変化の乏しいデータのみを生成してしまうモード崩壊と呼ばれる問題がある。モード崩壊が起こる原因としては多様性を出す役割である入力ベクトルノイズを無視してしまうことや、生成器があるデータを生成した際、識別器を騙すことができると生成器はそのデータを複製することで識別器を騙しやすくなる敵対的なネットワークの構造などが挙げられる。そこで Qi ら [7] は損失関数にモード崩壊を防ぐための正則化項を追加することで多様な画像を生成できる Mode Seeking GAN(MSGAN) を提案した。MSGAN の損失関数は式 2.6 で表される。 L_{ori} は GAN で使用される損失関数であり、 λ_{ms} は正則化の強さを調整する項、 L_{ms} は MSGAN で追加される正則化項で式 2.7 で表される。この損失関数は潜在変数間の距離に対する生成画像間の距離の比率を最大にすることを目的にしている。

$$L_{new} = L_{ori} + \lambda_{ms} L_{ms} \quad (2.6)$$

$$L_{ms} = \max_G \left(\frac{d_I(G(c, z_1), G(c, z_2))}{d_z(z_1, z_2)} \right) \quad (2.7)$$

第 3 章

提案手法

本研究では CNN が認識する分類に寄与する領域の形状の違いやパターンの違いの獲得を目指し、筒井の手法を本研究に適用する。まず年齢の相互変換を行うため、これまで広く行われてきた CNN 回帰モデルによる年齢推定を、MRI データを若年層、老年層の 2 クラスに分け、CNN 分類モデルで予測する方法をとる。そして図 3.1 に示すように CycleGAN を用いて若年層、老年層の脳画像の相互変換を学習する。得られた変換結果から差分画像を作成し形状やパターンの違いを捉え、CNN の分類に寄与する領域を特定する。また、図 3.2 に示すように変換の有効性を確認するために相互変換された画像を用いて学習済みの CNN による分類を行い、精度の違いを検証する。

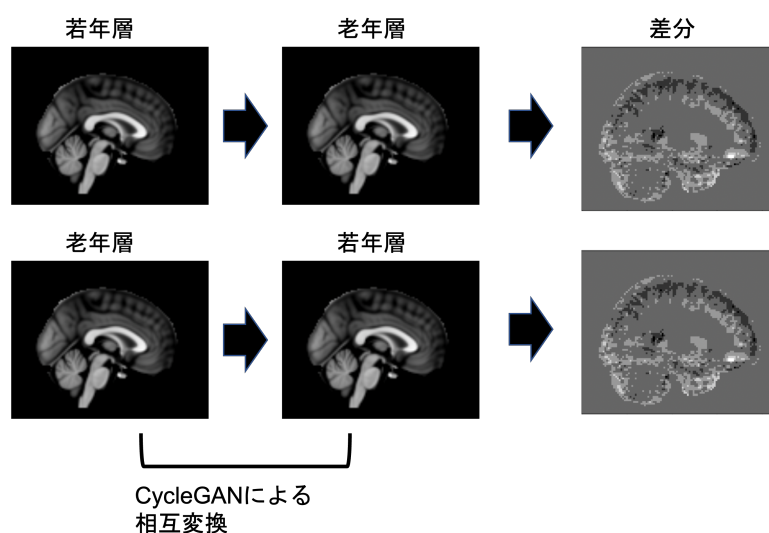


図 3.1: CycleGAN による画像変換と差分による説明性

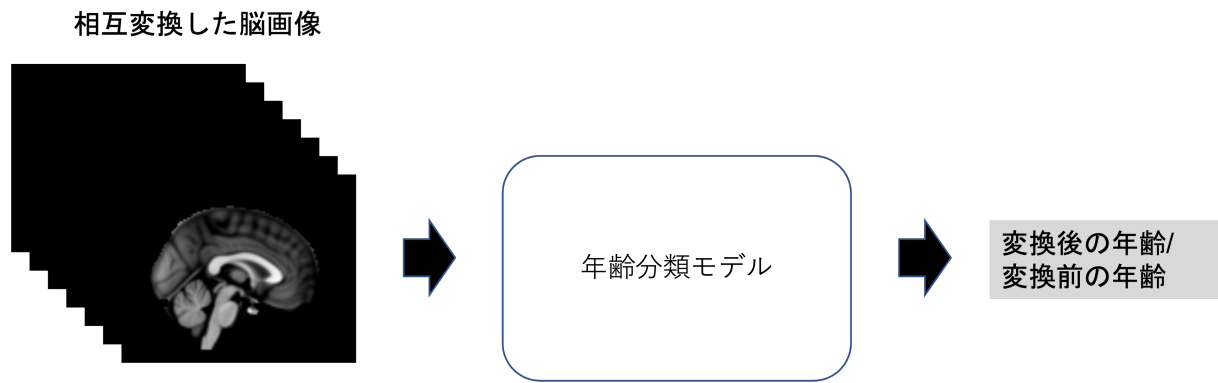


図 3.2: 変換した脳画像を用いての年齢分類

第 4 章

実験内容

本章では本研究で用いたデータセット，年齢分類の CNN モデル，CycleGAN の構成について記述する．

4.1 データセット

本研究では Imperial College London[8] で提供されている IXI データセットを用いる．IXI データセットは MRI で撮影された健常者の画像が 533 枚あり，画像を線形変換で標準脳にレジストレーションし，脳領域，白質，灰白質それぞれを抽出した T1 強調画像を用いる．学習の際に使用するラベルは付属している情報シートに記載されている年齢を使用し，データセットを 45 歳未満と 50 歳以上の 2 組に分割する．歳の間隔を空ける理由としてはデータの境界の違いを効率よく学習させるためである．GAN の学習に使用するデータは 45 歳未満，50 歳以上共に学習用 183 枚，テスト用 50 枚とする．これはデータの枚数の違いによる変換の偏りを抑えるためである．データセットを分割した詳細を表 4.1 に示す．

表 4.1: IXI データセットの分割内容

	CNN の学習		GAN の学習	
	訓練	検証	訓練	テスト
45 歳未満	188	46	183	50
50 歳以上	207	51	183	50

4.2 機械学習モデル

提案手法の検証にも用いた年齢分類を行う CNN の構造, CNN の分析に用いる CycleGAN を構成する生成モデル, 識別モデルの構造について記述する.

4.2.1 CNN モデル

本研究では過去の研究で年齢推定の精度が平均絶対誤差で 6.28 歳であったモデルを使用する. 図 4.1 に CNN のモデルの概要を示す. また表 4.2 にモデルの詳細を示す. 過去の研究では評価指標に平均絶対誤差を用いていたが本実験では若年層, 老年層の分類を行うため 2 クラス分類を行う構造に変更している.

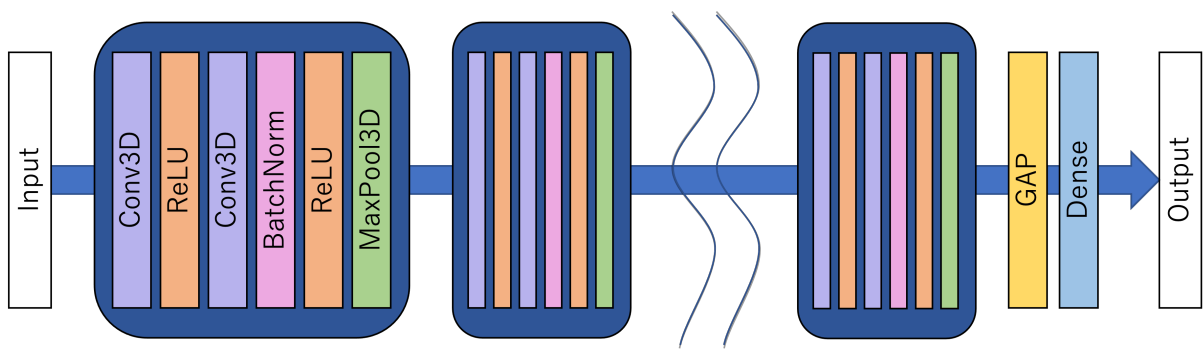


図 4.1: CNN モデル

表 4.2: ネットワーク構成

Block No.	Architecture
1	Conv1(8, 3) → ReLU → Conv2(8, 3) → BN → ReLU → MaxPooling(2)
2	Conv1(16, 3) → ReLU → Conv2(16, 3) → BN → ReLU → MaxPooling(2)
3	Conv1(32, 3) → ReLU → Conv2(32, 3) → BN → ReLU → MaxPooling(2)
4	Conv1(64, 3) → ReLU → Conv2(64, 3) → BN → ReLU → MaxPooling(2)
5	Conv1(128, 3) → ReLU → Conv2(128, 3) → BN → ReLU → MaxPooling(2)
6	Global Average Pooling → Dense(2)

4.2. 機械学習モデル

4.2.2 CycleGAN モデル

本研究では生成モデルとして9層で構成される ResNet を用いた。生成モデルの概要を図 4.2 に示す。

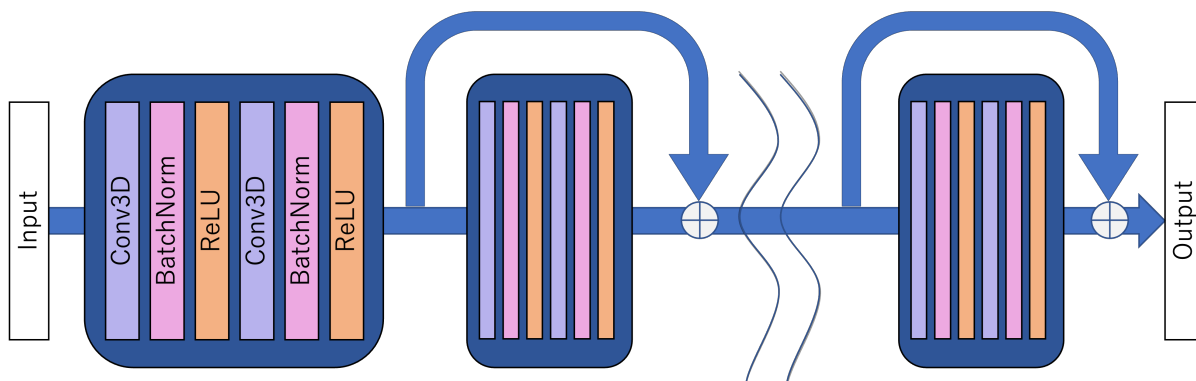


図 4.2: CNN モデル

4.2.3 差分画像

差分画像の算出法を式 4.1 に示す。D は差分画像，I は入力画像，G は変換画像を示す。入力画像から変換画像を除算し差分画像を生成する。そのため，差分画像で負の値を取る領域は増加した領域，正の値を取る領域は減少した領域となる。

$$D = I - G \quad (4.1)$$

第 5 章

実験結果

CycleGAN を用いた相互変換，変換前後の精度の違いについて検証した結果を記述する．

5.1 CycleGAN による相互変換

5.1.1 灰白質画像

CycleGAN による相互変換から得られた変換結果から，分類に寄与する領域と領域内の形状の違いを分析した結果を記述する．

灰白質画像の老年層 → 若年層の変換結果として図 5.1 が得られた．結果より，Saggital では前頭前連合皮質や前頭眼野，補足運動野，一次運動野の領域が増えたことが確認できる．Axial では外側腹側核，内包付近の領域が増え，Coronal では中心前回外側の領域が増えたことが確認できる．

若年層 → 老年層の変換結果として 5.2 が得られた．Saggital では老年層 → 若年層の結果で領域が増えていた前頭前連合皮質や前頭眼野の領域が減少していることが確認できる．Axial では上側頭回や中前頭回，前角，尾状核尾付近の領域が減少している．Coronal では一次運動野や島皮質，前頭弁蓋付近の領域が減少していることが確認できる．すなわち，灰白質の全体的な萎縮が示されていることがわかる．

5.1. CYCLEGAN による相互変換

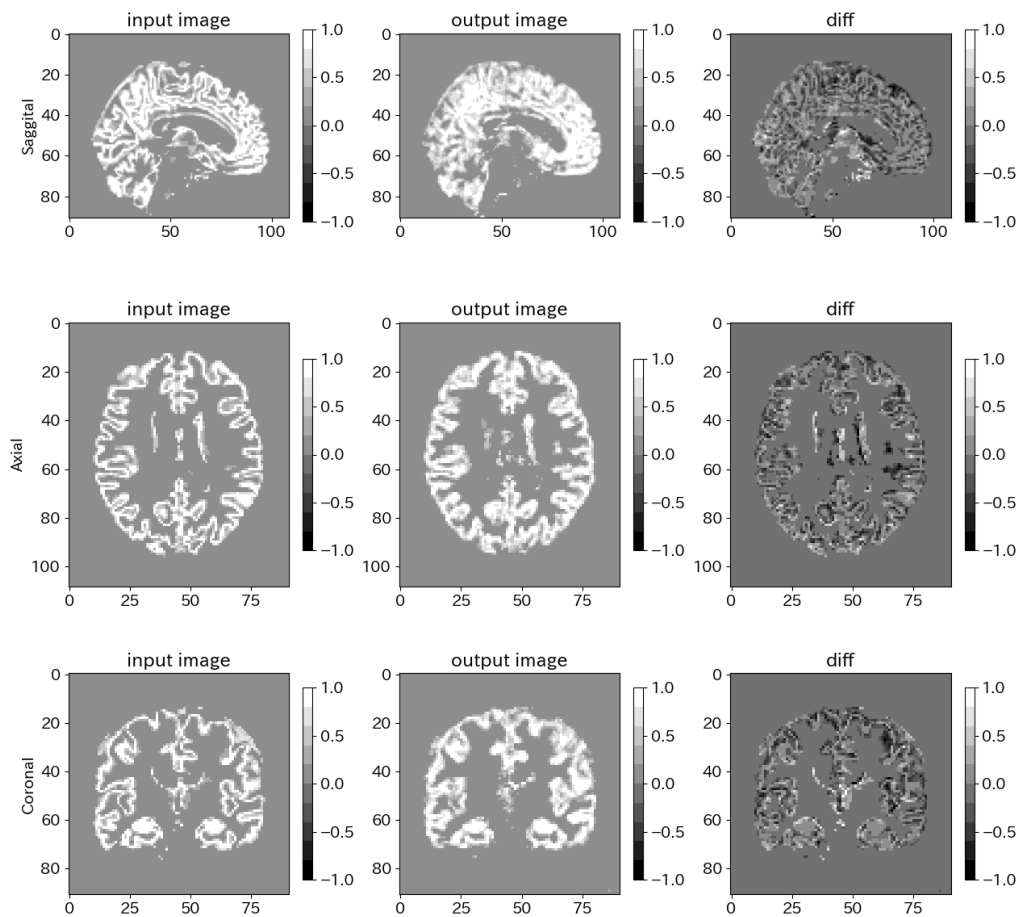


図 5.1: 老年層 → 若年層の変換結果と差分

5.1. CYCLEGAN による相互変換

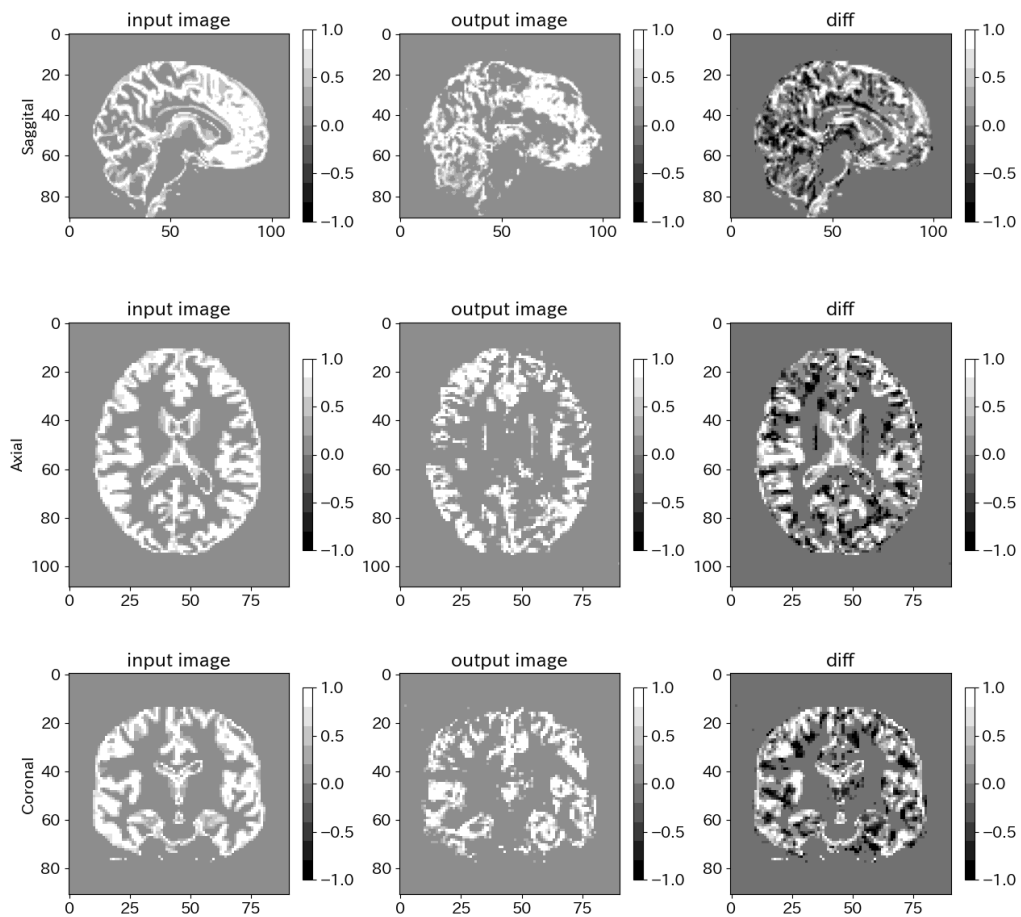


図 5.2: 若年層 → 老年層の変換結果と差分

5.1. CYCLEGAN による相互変換

5.1.2 白質画像

白質画像の老年層 → 若年層の変換結果として図 5.3 が得られた。Sagittal, Axial, Coronal 全ての画像で領域が増加していることが確認できる。しかし output image を見ると脳の構造をしていないことが確認できる。

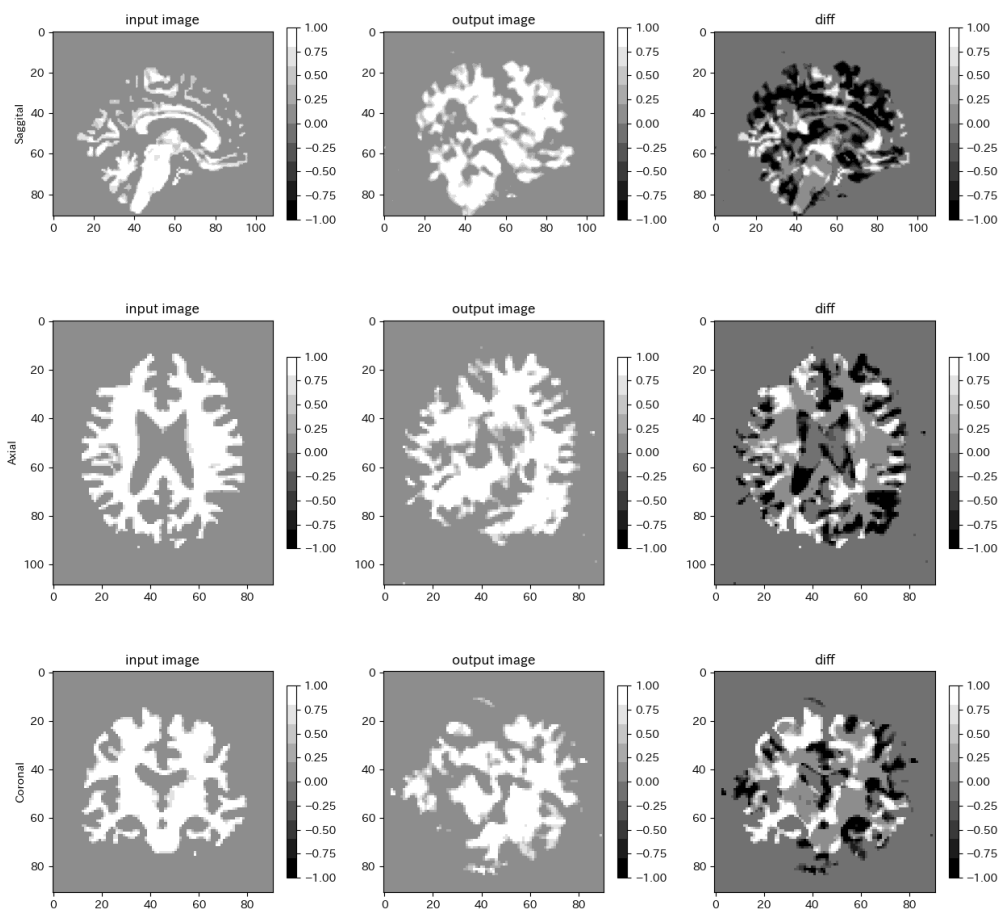


図 5.3: 老年層 → 若年層の変換結果と差分

5.1. CYCLEGAN による相互変換

若年層 → 老年層の変換結果として図 5.4 が得られた。Saggital では脳室付近の領域が少し増加していることが確認できる。Axial, Coronal では大きく増えた領域は確認できないが全体的に少し領域が減少していることが確認できる。

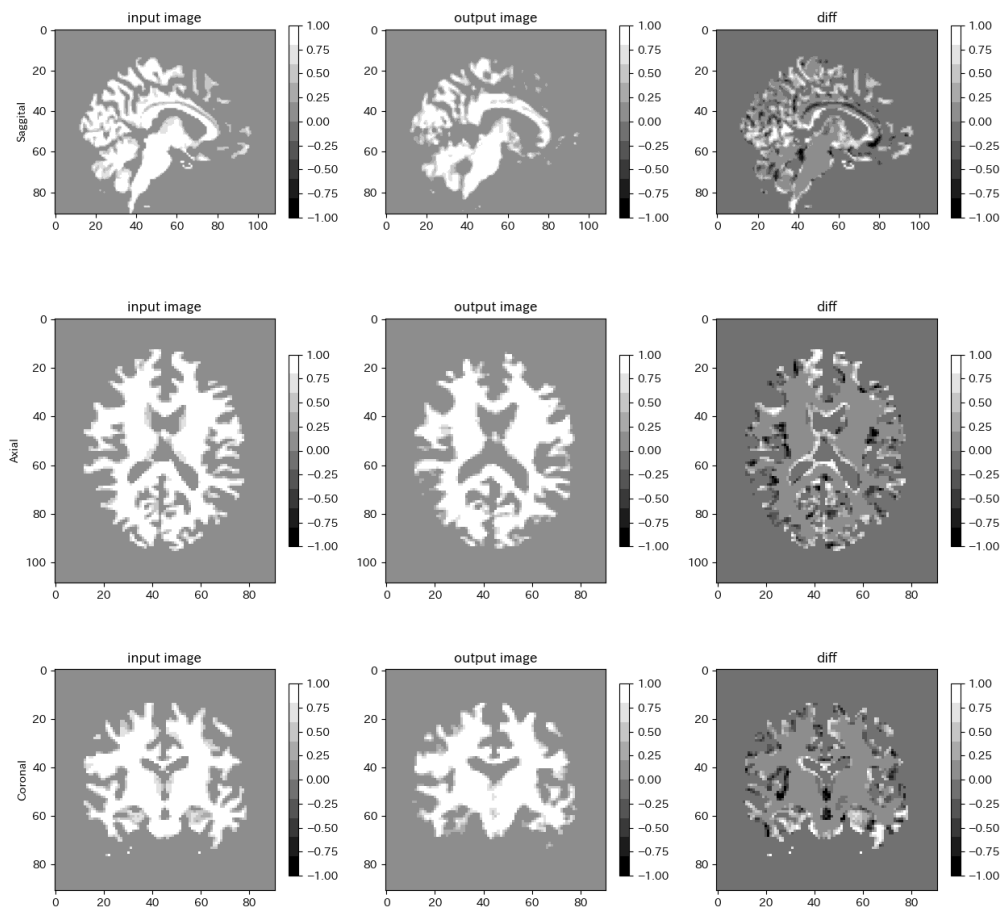


図 5.4: 若年層 → 老年層の変換結果と差分

5.1. CYCLEGAN による相互変換

5.1.3 脳領域画像

脳領域を抽出した画像の老年層 → 若年層の変換結果として図 5.5 が得られた。白質、灰白質と比べ大きな領域の変化は見られないが Saggital, Axial, Coronal で脳室付近の領域が増加していることが確認できる。

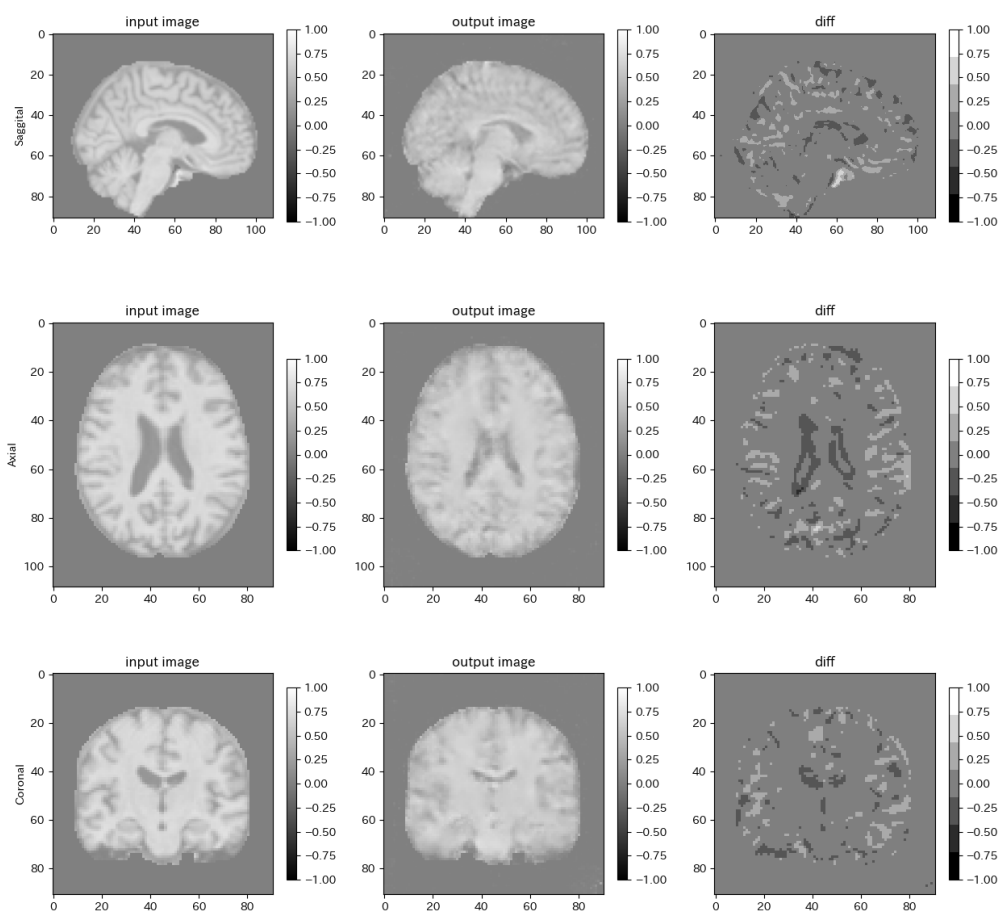


図 5.5: 老年層 → 若年層の変換結果と差分

5.1. CYCLEGAN による相互変換

若年層 → 老年層の変換結果として図 5.6 が得られた。大きな変化は見られないが老年層
→ 若年層の変換結果で減少していた領域が増加していることが確認できる。

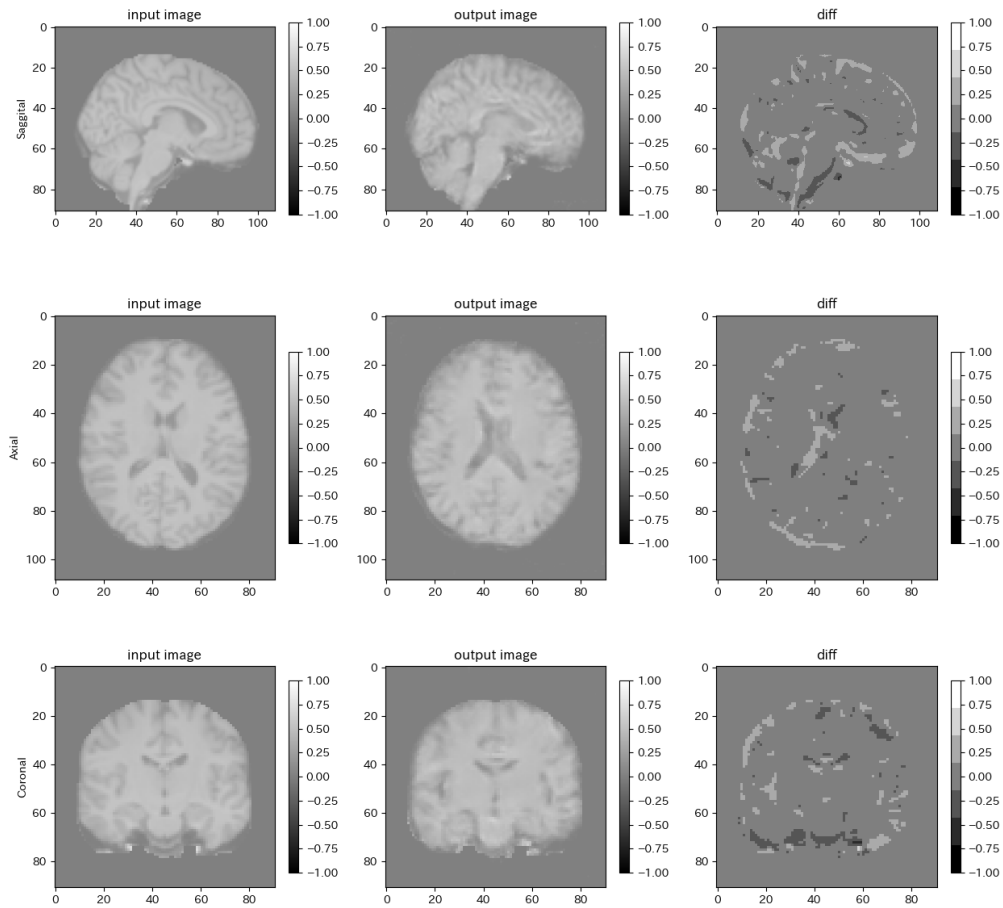


図 5.6: 若年層 → 老年層の変換結果と差分

5.2. CNN を用いた変換前後での精度の違い

5.2 CNN を用いた変換前後での精度の違い

5.2.1 灰白質画像

変換前後の画像を用いて年齢分類を行った結果を表 5.1 に示す。変換前では老年層で 13 人、若年層で 4 人誤分類しているが変換後では老年層で 2 人、若年層で 17 人誤分類している結果となった。

表 5.1: 変換前後の分類結果の混同行列

		分類結果			
		変換前		変換後	
		老年層	若年層	若年 → 老年	老年 → 若年
ラベル	老年層	170	4	181	17
	若年層	13	179	2	166

5.2.2 白質画像

変換前後の画像を用いて年齢分類を行った結果を表 5.1 に示す。変換前では老年層で 13 人、若年層で 4 人誤分類しているが変換後では老年層で 2 人、若年層で 17 人誤分類している結果となった。

表 5.2: 変換前後の分類結果の混同行列

		分類結果			
		変換前		変換後	
		老年層	若年層	若年 → 老年	老年 → 若年
ラベル	老年層	182	1	183	25
	若年層	1	182	0	158

5.3. CAM ベースの手法との比較

5.2.3 脳領域画像

変換前後の画像を用いて年齢分類を行った結果を表 5.1 に示す。変換前では老年層で 13 人、若年層で 4 人誤分類しているが変換後では老年層で 2 人、若年層で 17 人誤分類している結果となった。

表 5.3: 変換前後の分類結果の混同行列

		分類結果			
		変換前		変換後	
		老年層	若年層	若年 → 老年	老年 → 若年
ラベル	老年層	179	4	182	2
	若年層	4	179	1	181

5.3 CAM ベースの手法との比較

CAM を用いて老年層、若年層の分類を可視化した結果と灰白質画像の差分画像を用いた結果の比較を図 5.7 に示す。GradCAM を用いて可視化した画像では老年層では前頭葉内側から頭頂葉内側の関与が示されている。若年層では前頭葉、頭頂葉内側と老年層よりも広い範囲の関与が示されている。相互変換を行った差分画像からは老年層 → 若年層の平均画像では補足運動野、一次運動野付近の領域が少し増加したのが確認できる。若年層 → 老年層の平均画像からは前頭前連合皮質や前頭眼野の領域が減少していることが確認できる。GradCAM では分類に寄与した領域が示されているが領域内の形状の違いなど分類に寄与した詳細な領域の取得はできていない。対して差分画像からは変化した領域の詳細が取得できていることが確認できる。

5.4. モード崩壊

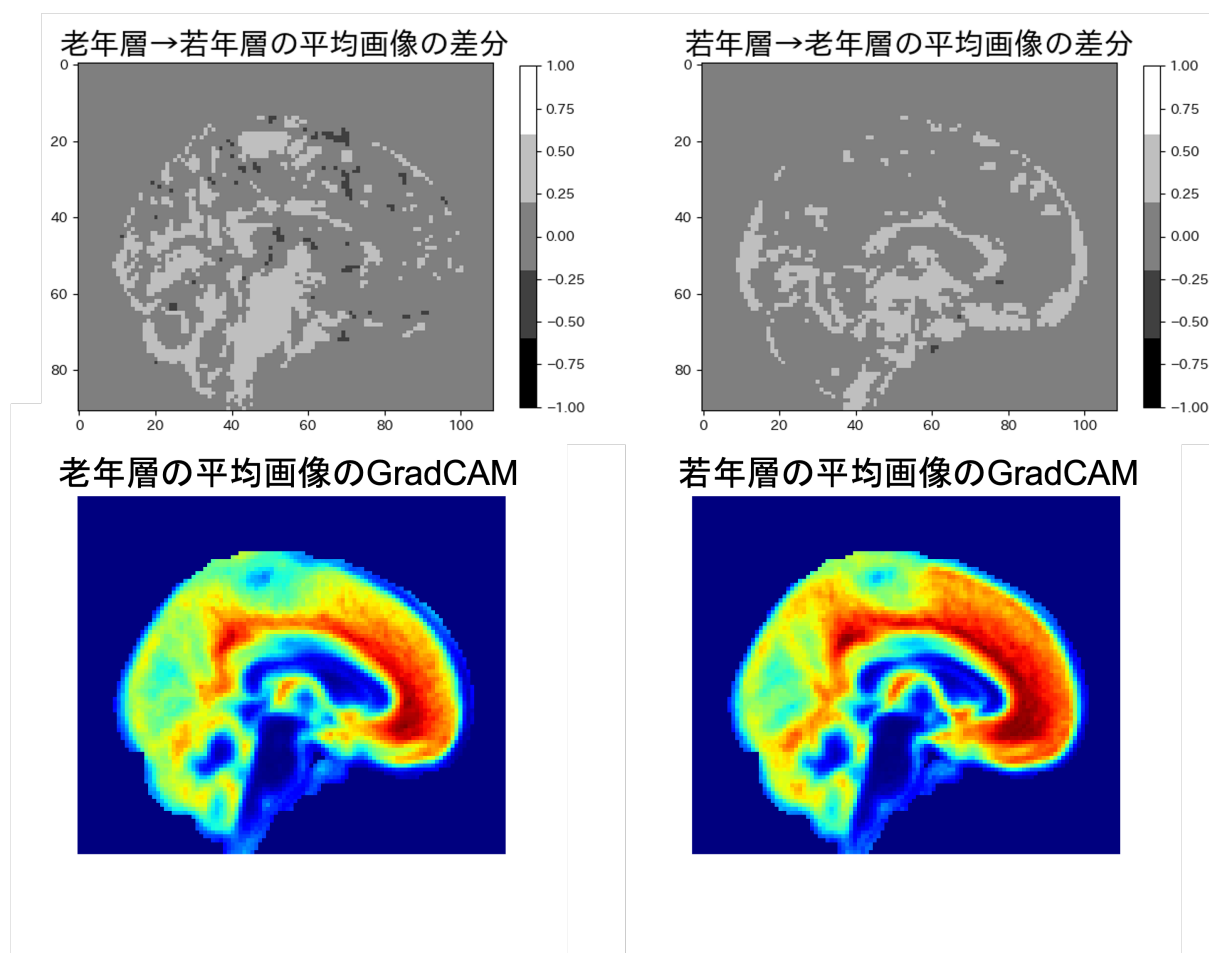


図 5.7: 差分画像と GradCAM の比較

5.4 モード崩壊

相互変換を行った画像を用いて年齢の分類を行った結果、脳領域画像、白質、灰白質全ての画像で年齢推定の結果が変化していることが確認できた。しかし差分画像を見ると変化量が少ない変換と変化量が多い変換があることが確認できる。そこで若年層 → 老年層の画像 A と B 、老年層 → 若年層の画像 C と D のそれぞれ差分画像を図 5.9 と図 5.8 に示す。老年層 → 若年層の変換では入力画像が異なると変換された画像も形状が異なっているのが差分画像からわかる。しかし若年層 → 老年層の変換では変換された画像に多様性がなく、似た画像が生成されている。これはモード崩壊が起り、画像を複製しているからと考えられる。そのため灰白質画像の若年層 → 老年層の変換結果や白質画像の老年層 → 若年層の変

5.4. モード崩壊

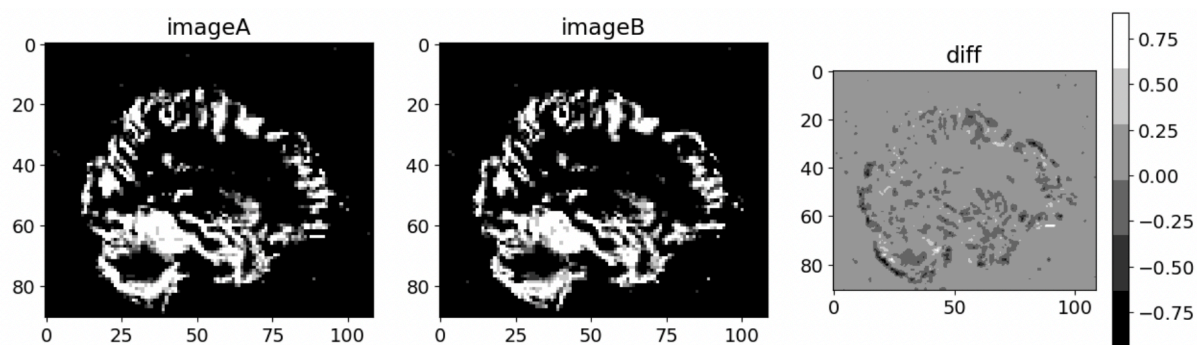


図 5.8: 差分画像と GradCAM の比較

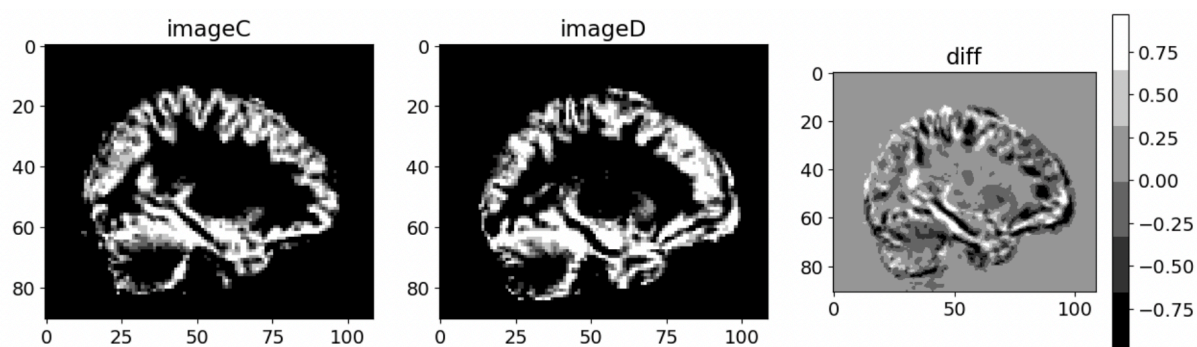


図 5.9: 差分画像と GradCAM の比較

換結果の差分ような領域全体が変化してる画像が生成されていると考える。

第 6 章

追加実験

CycleGAN を用いて変換した画像がモード崩壊をしていることから,MSGAN の Mode Seeking 正則化項を CycleGAN に適用し追加実験を行なった内容と結果について記述する.

6.1 追加実験内容

IXI データセットの灰白質画像を用いて CycleGAN による相互変換を行う際, Mode Seeking 正則化項を CycleGAN に適用し画像の変換を行う. また変換後の画像を用いて年齢分類を行い, 変換の有効性を確認する.

6.2 追加実験結果

6.2.1 Mode Seeking 正則化項を追加した CycleGAN による相互変換

Mode Seeking 正則化項を追加した CycleGAN による変換結果をと差分を図 6.1 及び図 6.2 に示す. 老年層 → 若年層の変換では Saggital, Axial, Coronal 全てで脳領域全体が増加していることが確認できる. 若年層 → 老年層の変換では Saggital, Axial, Coronal 全てで大きな領域の変換は見られない.

6.2. 追加実験結果

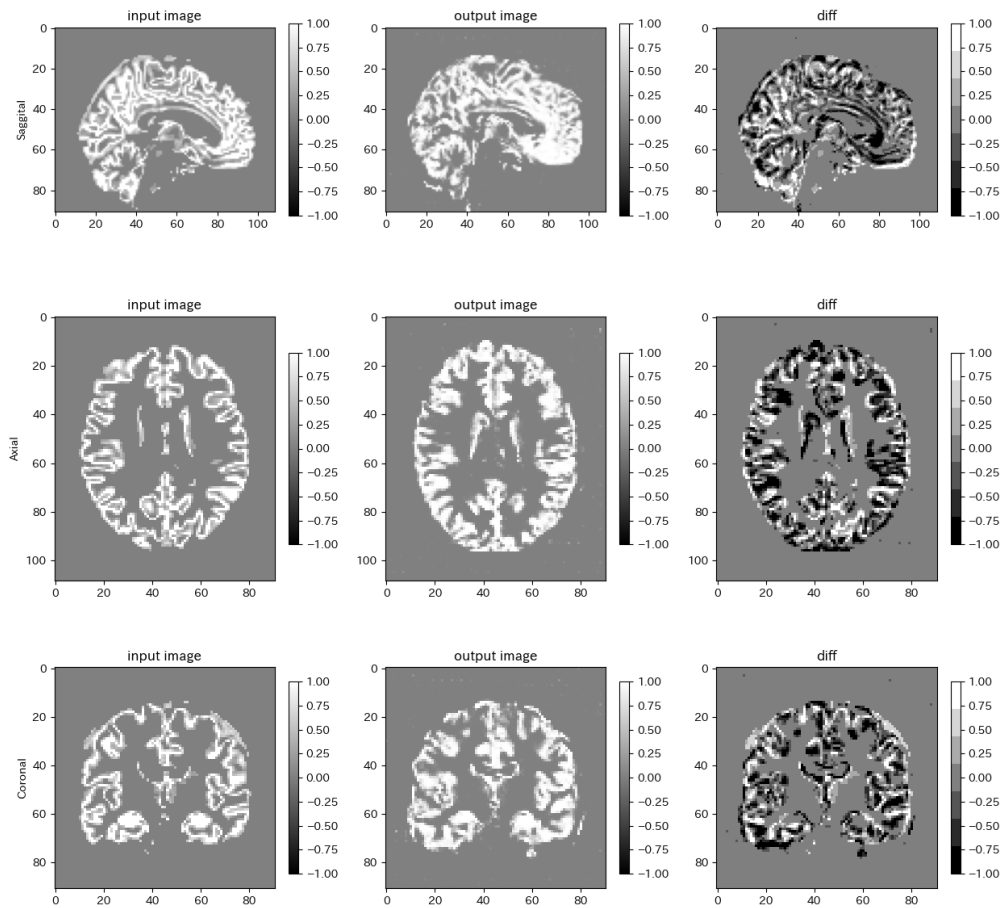


図 6.1: 老年層 → 若年層の変換結果と差分

6.2. 追加実験結果

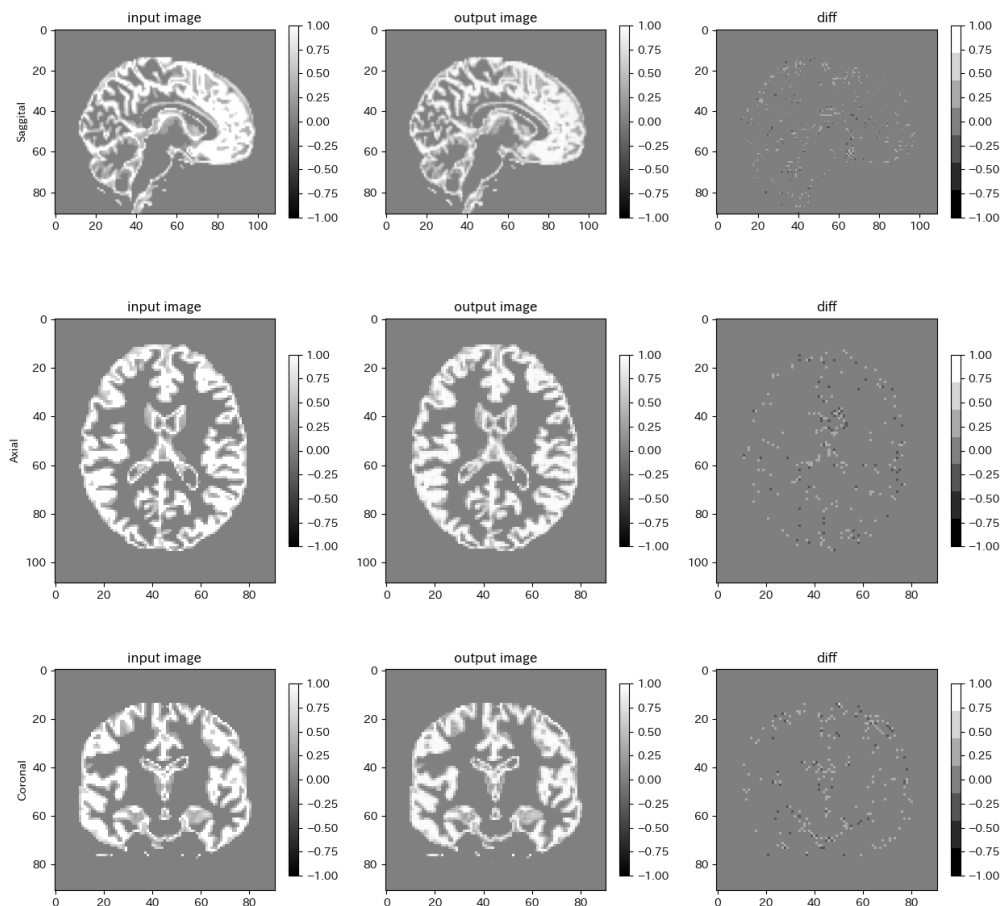


図 6.2: 若年層 → 老年層の変換結果と差分

6.2.2 CNN を用いた変換前後での精度の違い

Mode Seeking 正則化項を追加した CycleGAN による変換画像を用いて CNN による年齢分類を行なった結果を表 6.1 に示す。老年層 → 若年層では 33 人を誤分類しており、若年層 → 老年層では 182 人誤分類していることが確認できる。

6.2. 追加実験結果

表 6.1: 変換前後の分類結果の混同行列

		分類結果			
		変換前		変換後	
		老年層	若年層	若年 → 老年	老年 → 若年
ラベル	老年層	170	4	1	33
	若年層	13	179	182	150

第7章

考察

7.1 CycleGAN による相互変換

7.1.1 灰白質画像

老年層、若年層の脳画像を相互変換した結果として老年層 → 若年層では差分から前頭前連合皮質や前頭眼野、補足運動野、一次運動野の領域が増えたことが確認できた。また、全体的に脳領域が増えていることが確認できた。このことから CycleGAN は年齢を重ねるごとに前頭葉、頭頂葉付近の灰白質が減少することを学習していると考えられる。また、この部位は Good らの研究 [9] で報告されている加齢によって灰白質の容積が低下する部位と一致している。

若年層 → 老年層の変換では前頭葉の領域が減少し、後頭葉、海馬、小脳付近の領域が増加していることが確認できる。これらの部位も文献 [9] で報告されている加齢によって減少しない領域と一致している。

7.1.2 白質画像

老年層 → 若年層の変換では差分画像から脳領域全体の増加が増加していることが確認できる。しかし変換後の画像を見ると脳の構造をしていない画像が生成されており、老年層と若年層の白質の容積の違いを学習していない可能性が高いと考えられる。

若年層 → 老年層の変換では脳全体の領域が増加していることが確認できる。これらの結果は加齢によって灰白質が減り白質が増えていることを CycleGAN で学習できているとい

7.2. CNN を用いた変換前後での精度の違い

えると考える。

7.1.3 脳領域画像

脳領域画像では老年層 → 若年層、若年層 → 老年層の変換共に大きな領域の増減は見られなかった。細かな変化だと老年層 → 若年層では脳室付近と尾状核付近の領域が増加しており、若年層 → 老年層では尾状核付近と脳の輪郭の領域の減少が確認できる。

7.2 CNN を用いた変換前後での精度の違い

7.2.1 灰白質画像

灰白質画像での変換前後の年齢分類の精度の違いとして若年層 → 老年層では誤分類が2人と高精度で分類していることが確認できた。一方老年層 → 若年層では17人を誤分類すると精度の低下が見られた。この結果は若年層 → 老年層では画像を複製してしまうモード崩壊が確認されており、複製された画像が老年層の特徴を持っていたため、高精度での推定になったと考えられる。老年層 → 若年層ではモード崩壊は確認されておらず、若年層の脳に変換できていない画像が含まれていたため、誤分類が多くなったと考える。

7.2.2 白質画像

白質画像での年齢分類の精度の違いとしては若年層 → 老年層の変換で誤分類が無いという結果になった。変換された画像を見てもモード崩壊が起こっておらず、変換が有効であったと言える。全ての変換が有効であった理由として、加齢によって白質が増えることを CycleGAN が学習し、白質が増えたことにより老年層の脳の構造に近くなったからであると考える。一方老年層 → 若年層では25人を誤分類している。図 5.3 から確認できるように変換後の画像が脳の構造を保てておらず、推定の精度に影響が出たと考える。

7.3. CAM ベースの手法との比較

7.2.3 白質画像

白質画像での年齢分類の精度の違いとしては若年層 → 老年層の変換で誤分類が無いという結果になった。変換された画像を見てもモード崩壊が起こっておらず、変換が有効であったと言える。全ての変換が有効であった理由として、加齢によって白質が増えることを CycleGAN が学習し、白質が増えたことにより老年層の脳の構造に近くなったからである。一方老年層 → 若年層では 25 人を誤分類している。図 5.3 から確認できるように変換後の画像が脳の構造を保てておらず、推定の精度に影響が出たと考える。

7.2.4 脳領域画像

脳領域画像での年齢分類の精度の違いとして若年層 → 老年層で 1 人、老年層 → 若年層で 2 人のみの誤分類と高精度での分類が行えた。図 5.5 と図 5.6 では大きな領域の変化は見られなかったが、分類が高精度で行えていることから脳室付近や尾状核付近の領域、脳の輪郭が年齢推定に影響していると考えられる。

7.3 CAM ベースの手法との比較

GradCAM を用いて老年層、若年層の分類に寄与した領域の可視化と、差分画像を用いた領域の可視化の結果から GradCAM と差分画像では可視化される領域が異なっていることが確認できる。GradCAM は分類に寄与した領域を学習済みモデルの勾配から算出しているが CycleGAN は脳の構造自体を変換しているため、このような違いが現れたと考える。しかし、前頭葉や頭頂葉付近など GradCAM と差分画像で共通している領域も確認できる。この共通する領域は年齢の分類に寄与した領域として特に重要な部位であると考えられる。

7.4 Mode Seeking 正則化項を追加した CycleGAN による相互変換と年齢推定の精度の違い

モード崩壊を抑制するため Mode Seeking 正則化項を追加した CycleGAN で老年層、若年層の変換を行なった結果として老年層 → 若年層では脳領域の全体が増加し、若年層 → 老年層では脳領域全体であまり変換が見られない結果となった。年齢の分類結果では老年層 → 若年層では 33 人、若年層 → 老年層では 182 人誤分類する結果となった。老年層 → 若年層では 7.1 から確認できるように小脳が肥大化したり、前頭葉が尖るなど脳の構造をしていないのが確認できる。これが精度に影響を与えていると考える。

若年層 → 老年層では 7.2 から確認できるように元画像と変換画像で変化している領域が確認できない。これがほぼ全ての画像を誤分類した要因だと考える。

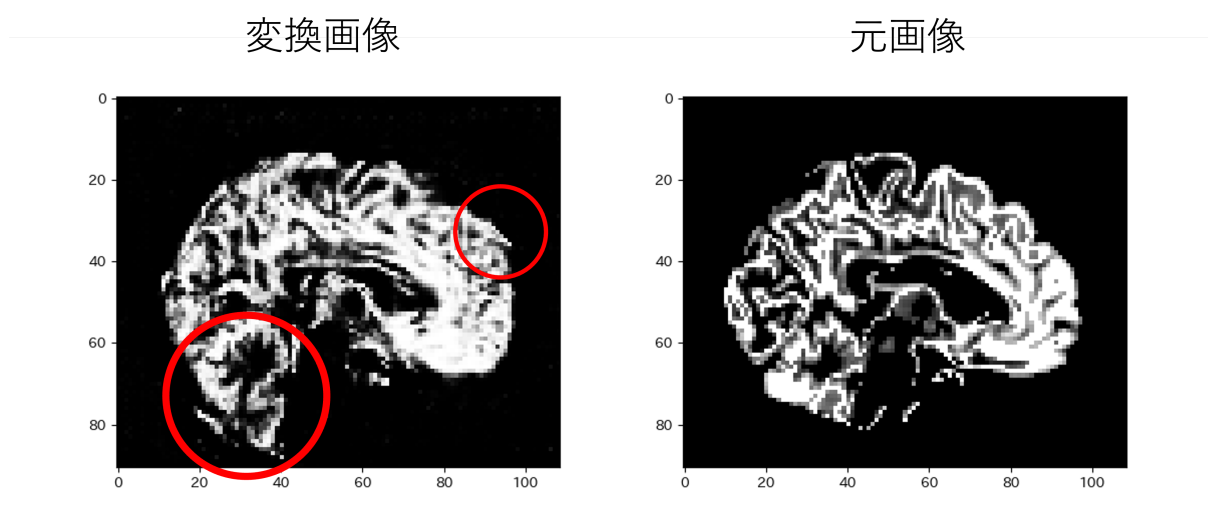


図 7.1: 老年層 → 若年層の変換後と変換前の画像

7.4. MODE SEEKING 正則化項を追加した CYCLEGAN による相互変換と年齢推定の精度の違い

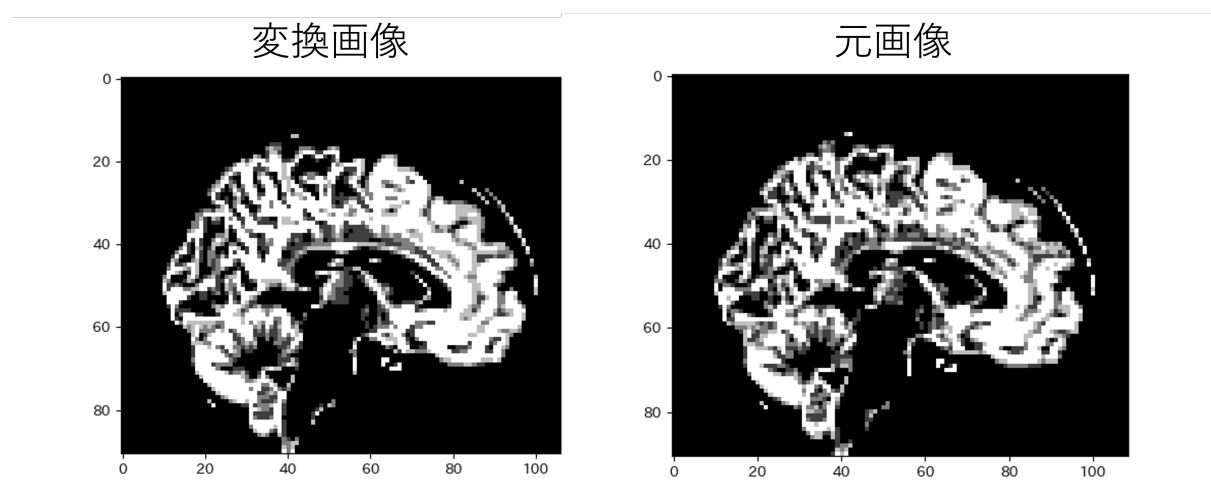


図 7.2: 若年層 → 老年層の変換後と変換前の画像

第 8 章

おわりに

本研究では GAN を用いて画像内の領域の違いやパターンの違いを取得する手法を本研究に適用し、年齢分類を行う CNN のモデルの説明性の向上を目指した。MRI 脳画像を対象に検証した結果、年齢の変化における脳領域の変化を取得することができ、Grad-CAM と比較しても詳細な結果が得られた。変換前後の画像を用いた若年層-老年層の分類でも CNN の分類に有効な変換が行われたことが確認でき、説明性の向上につながったと考えられる。しかし、使用する画像の違いや若年層 → 老年層と老年層 → 若年層の変換過程の違いで説明性の低い画像が生成されることが確認できた。また、変換の一部では生成器が似た画像を生成するモード崩壊が起こったことが確認できた。今後は画像の違いや変換の過程の違いによる説明性の低下の問題や、モード崩壊の問題を解決するための構造の追加を検討する必要があると考える。また生成した画像が正しい脳の構造をしているかの定量的な評価を行う指標も必要であると考えられる。

謝辞

本研究を行うにあたり、日頃からご指導いただいた吉田真一教授に御礼申し上げます。4年間の研究室生活を無事に過ごせたのは吉田先生の優しさがあってこそだと感じています。研究が進まない時も一緒に解決案を探してくださったり、就職活動も親身になって相談に乗ってくださり、充実した研究室生活を過ごすことができました。

本研究で副査を引き受けてくださった中原潔教授・門田宏准教授にも心から感謝致します。自分では気づかない問題点や改善案などを提案してくださり、研究に対する知見が深まる機会が多くありました。

知能情報学研究室の方々にもお世話になりました。修士になってからは教えることが多くなりましたが、教えることで自分の知識の再確認にもなり、また逆に自分が学ぶ機会も多くなりました。研究活動のみでなく、普段の生活でも賑やかで楽しい生活をすることができました。

大学、大学院での生活を無事に終えられたのは両親を始め様々な人に支えられたお陰だと痛感しています。最後になりますが心から感謝致します。

参考文献

- [1] Shinichi Yoshida Yasuyuki Tsutsui. Analysis of trained convolutional neural network using generative adversarial network. In *IWACIII*, 2021.
- [2] James H. Cole, Rudra P.K. Poudel, Dimosthenis Tsagkrasoulis, Matthan W.A. Caan, Claire Steves, Tim D. Spector, and Giovanni Montana. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, Vol. 163, pp. 115–124, 2017.
- [3] 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.
- [4] 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.
- [5] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization, 2015.
- [6] 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. *International Journal of Computer Vision*, Vol. 128, No. 2, p. 336–359, October 2019.
- [7] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis, 2019.
- [8] Imperial College London Biomedical Image Analysis Group. Ixi dataset – brain

参考文献

development, 2022. <https://brain-development.org/ixi-dataset/>.

- [9] Catriona D. Good, Ingrid S. Johnsrude, John Ashburner, Richard N.A. Henson, Karl J. Friston, and Richard S.J. Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, Vol. 14, No. 1, pp. 21–36, 2001.

付録 A

付録 B