

令和元年度
学士学位論文

敵対的生成ネットワークによる
MRI 脳画像の性差の分析

**Analysis of Trained Convolutional Neural Network
classifying Gender Differences
from Structural Magnetic Resonance Imaging
using Generative Adversarial Network**

1200336 筒井 康行

指導教員 吉田 真一

高知工科大学 情報学群

要旨

敵対的生成ネットワークによる MRI 脳画像の性差の分析

筒井 康行

人の脳は生まれ持った違いだけでなく生活習慣によって変化しており、脳を見ることでその人の年齢や性別、喫煙の有無といった個人属性を推定できるという考えがある。近年、脳構造画像から性別、年齢などの個人属性を推定する研究が進められている。文献 [1] では畳み込みニューラルネットワーク (CNN:Convolutional Neural Network) によって脳構造画像から男女を 97%の精度で識別できたとされている。CNN では学習後の重みから識別に寄与する領域を特定する手法 (Grad-CAM 等) による識別過程の解明が試みられている。しかし性差の説明には性差に関わる領域だけでなく、領域内での形状やパターンの違いを知る必要があり、領域の特定だけでは性差の説明には十分でない。そこで本研究では敵対的生成ネットワーク (GAN:Generative Adversarial Networks) を応用した Attention-Guided CycleGAN[2] を用いて脳構造画像を男女間で相互変換し、Attention-Guided CycleGAN から得られる注意領域や変換結果から性差の領域だけでなく形状やパターンの分析を行う。その結果として男性から女性への変換では後頭葉を削られるように変換され、女性から男性への変換では後頭葉と小脳の辺りに新たな領域が生成されることを示す。この結果から脳の性差には後頭部辺りの脳の形状が関わる可能性があると考えられる。また、学習過程の注意領域と変換画像を見ることで CNN の学習過程の説明に寄与できることを示す。

キーワード 敵対的生成ネットワーク, 性差, MRI, 機械学習

Abstract

Analysis of Trained Convolutional Neural Network classifying Gender Differences from Structural Magnetic Resonance Imaging using Generative Adversarial Network

Tsutsui Yasuyuki

Shape of brain of human being reflect not only innate differences but also lifestyles such as food or smoking. Ages clearly affect the brain shape and structure and there are studies of sexual dimorphism of brain. Food or smooking may also affect the brain structure. In recent years, research on estimating personal attributes such as gender and age from brain structure images has been advanced. Sasatani et al. have been employed a convolutional neural network(CNN) to obtain 97% accuracy for gender prediction only from brain structure images. Several studies have proposed CNN analysys such as Grad-CAM, which specifies the region which contributes to discrimination from the weights after learning. However, in order to explain gender differences, it is necessary to know not only the regions related to gender differences, but also differences in shapes and patterns within the regions. Therefore, in this research, brain structure images are mutually converted between male and female using Attention-Guided CycleGAN (AG-CycleGAN)[2] applying Generative Adversarial Networks (GAN) and obtained from AG-CycleGAN. From the attention map and the transformation result, not only the gender region but also the shape and pattern are analyzed. As a result, the transformation from male to female is performed so that the occipital lobe is shaved.

The transformation from female to male is performed so that a new region is created around the occipital and the cerebellum. These results suggest that the gender difference of the brain may be related to the shape of the brain around the occiput. It is also found that looking at the attention map and the transformed image of the learning process could contribute to the explanation of the learning process of CNN.

key words GAN, Gender Difference, MRI, Machine Learning

目次

第 1 章	序論	1
第 2 章	関連研究	3
2.1	GAN(Generative Adversarial Networks)	3
2.2	CycleGAN	4
2.3	Attention-Guided CycleGAN	5
2.4	CNN(Convolutional Neural Network)	8
2.4.1	CNN の基本構造	8
2.4.2	畳み込み層 (Convolution layer)	9
2.4.3	プーリング層	12
2.5	PatchGAN	13
2.6	U-Net	14
第 3 章	提案手法	16
第 4 章	実験内容	18
4.1	データセット	18
4.2	学習手順	18
4.3	使用モデル	19
4.3.1	生成モデル	19
4.3.2	識別モデル	19
4.3.3	注意モデル	20
第 5 章	実験結果	22
5.1	全脳	22

目次

5.2	灰白質	26
5.3	白質	30
第 6 章	考察	34
6.1	実験結果に対する考察	34
6.2	注意領域を用いた識別モデルの識別過程の考察	35
第 7 章	結論	39
	謝辞	40
	参考文献	42

目次

2.1	GAN のデータの遷移	4
2.2	CycleGAN のデータの遷移	5
2.3	Attention-Guided CycleGAN のデータの遷移	7
2.4	Attention-Guided CycleGAN のモジュール内でのデータの遷移	7
2.5	3 クラス分類を学習する CNN の基本的な構造	9
2.6	畳み込み層での処理の流れ	10
2.7	入力の特徴マップの各ピクセル値	12
2.8	平均・最大プーリングの例	13
2.9	U-Net 全体のネットワーク構造	15
4.1	生成モデルの構造	20
4.2	識別モデルの構造	20
4.3	注意モデルの構造	21
5.1	全脳の変換結果 男性 → 女性 → 男性	23
5.2	全脳の変換結果 女性 → 男性 → 女性	23
5.3	全脳の変換結果の比較	24
5.4	全脳の元画像と変換画像の平均画像とその差分	24
5.5	全脳における識別モデルの正解率 (全モデル学習時)	25
5.6	全脳における識別モデルの正解率 (注意モデル以外学習時)	25
5.7	灰白質の変換結果 男性 → 女性 → 男性	26
5.8	灰白質の変換結果 女性 → 男性 → 女性	27
5.9	灰白質の変換結果の比較	27
5.10	灰白質の元画像と変換画像の平均画像とその差分	28

図目次

5.11 灰白質における識別モデルの正解率 (全モデル学習時)	28
5.12 灰白質における識別モデルの正解率 (注意モデル以外学習時)	29
5.13 白質の変換結果 男性 → 女性 → 男性	30
5.14 白質の変換結果 女性 → 男性 → 女性	31
5.15 白質の変換結果の比較	31
5.16 白質の元画像と変換画像の平均画像とその差分	32
5.17 白質における識別モデルの正解率 (全モデル学習時)	32
5.18 白質における識別モデルの正解率 (注意モデル以外学習時)	33
6.1 灰白質画像を用いた学習時の元画像 (org) と注意領域 (mask) の遷移	37
6.2 灰白質画像を用いた学習時の元画像 (org) と注意領域 (mask) の遷移	38

第 1 章

序論

人の体には性別や年齢、生活習慣の違いから様々が見られる。性別では体の部位や身長などに性差が見られ、年齢では骨や皮膚のしわなどに違いが見られる。また人の外見だけでなく、体の内部においても同様に違いが見られる。生活習慣では、十分な栄養の摂取、運動、睡眠などを維持して健康的な生活を送っていれば生活習慣病などの病気のリスクが軽減される。逆に栄養、運動、睡眠が不足した生活を続けていると将来的に病気を患うリスクが大きくなる。このように人の体は生活習慣によっても変化している。脳も体の一部であり、同様に性別や年齢、生活習慣による違いが生じると言う考えは自然であり、実際、性別や年齢に関わる研究報告もある [3][4]。そのような考えに沿い、本研究では脳ドックで撮影された脳構造 MRI 画像を対象にそれらの構造の違いと要因の分析に関する研究を行う。このような画像を用いた予測には画像認識が用いられるが、特に近年では医療画像においても畳み込みニューラルネットワーク (CNN:Convolutional Neural Network) と呼ばれる機械学習の手法が取り入れられている。文献 [1] では CNN によって全脳の脳構造画像から男女を 97%の精度で識別できたとされている。このように脳構造画像から年齢や性別、生活習慣などの個人属性が推定できれば、機械学習で脳構造画像からアルツハイマーなどの脳の病気を予測する際に、個人属性と脳の関係に基づいたデータを補正することで不必要な情報を減らし、より高精度な予測を行えるようになると考えられる。医療の進歩によって寿命が長くなり、脳の機能が低下する認知症などの患者が増えている。厚生労働省はアルツハイマー病の患者数が 1999 年の約 3 万人から 2014 年には約 53 万人へと増加していると発表している [5]。これにより、医療現場への負担も大きくなっていると考えられる。そこで機械学習による診断が実用化されれば、医療現場の負担を減らすことができ、高度な医療をより多く

の人に提供することができるようになる。

しかし、個人属性を用いたデータの補正を行うには個人属性と脳の間係を説明できるようになる必要がある。文献 [1] では、識別に寄与する領域を特定する手法として Selvaraju らが提案した Grad-CAM[6] を用いて性差に関わる領域の特定を試みた。この Grad-CAM では、学習後の CNN の重みを利用して識別に寄与する領域の特定を行なっている。しかし、あまり意味のある領域の特定は至っておらず、識別の理由の説明はほとんどできていない。また、性差の説明のためには領域だけでなく領域内での形状やパターンの違いを知る必要があり、領域の特定だけでは十分でない。

そこで本研究では、男女の性差の識別に関わる領域の特定と領域内での形状やパターンの違いを知るためにニューラルネットワークによる分析を行い、認識結果による説明も試みる。そのために用いる手法として敵対的生成ネットワーク (GAN:Generative Adversarial Networks) を応用した Attention-Guided CycleGAN[2] を用いた手法を提案する。ここで Attention-Guided CycleGAN を用いた理由としては、Attention-Guided CycleGAN では注意領域によって変換領域を制限した上で 2 つのデータセットの相互変換を行うことができるため、変換による差から男女の性差に関わる領域の特定や領域内での形状やパターンの分析が行えると考えたからである。よって本研究では、高精度な診断に必要なデータ補正を行うために、Attention-Guided CycleGAN による男女の脳構造画像の相互変換を行い、男女間での脳構造の違いを分析することを目指す。

第 2 章

関連研究

本研究で 3 次元脳構造画像での性差の分析に用いた Attention-Guided CycleGAN の説明のために GAN, CycleGAN, Attention-Guided CycleGAN の順に説明を行う。また, 本研究でモデルを生成するにあたって, モデルの構造に用いた CNN(Convolutional Neural Network), 生成モデルで用いた U-Net, 識別モデルに用いた Patch GAN の説明も行う。

2.1 GAN(Generative Adversarial Networks)

敵対的生成ネットワーク (GAN)[7] は図 2.1 のように生成モデル G と識別モデル D から構成され, G が入力ノイズ z をもとにした元データ x と区別できないようなデータの生成を学習する。Goodfellow らが提案したで GAN の損失関数は Zhu らの研究では Adversarial Loss と呼ばれ, 式 (2.1) のように表すとされている [7][8]。式 (2.1) の第 1 項によって D が元データ x を正しく識別するように学習し, 第 2 項によって D が G によって生成されたデータを元データではないと識別するように学習する。ここでの G と D の関係は通貨の偽造者と偽通貨と本物の通貨を見分ける警察との関係で表されることが多く, 偽造者は警察を騙せる偽通貨を製造できるように努力し, 警察はその偽通貨と本物の通貨を見分けるように努力する。このように偽造者と警察が互いに敵対して競い合うことで偽造者は本物に近い偽通貨の製造方法を学習する。同様に GAN では, G と D が互いに敵対することで G が元データ x に似たデータを生成できるようになる。近年では, GAN を応用することで次に説明する CycleGAN や超解像を行う SRGAN といったモデルも提案されている。

2.2 CycleGAN

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

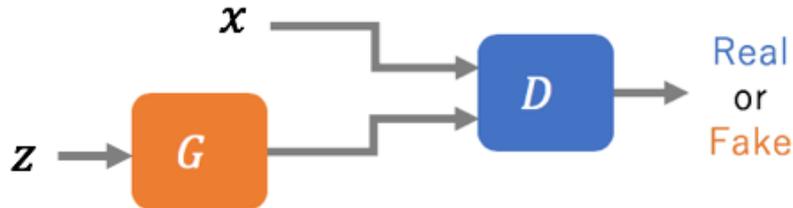


図 2.1: GAN のデータの遷移

2.2 CycleGAN

CycleGAN[8] は GAN を発展させたモデルであり、図 2.2 のように 2 つのデータセット X, Y に対応する生成モデル G_x, G_y と識別モデル D_x, D_y の 4 つのモデルから構成され、 $G_x : Y \rightarrow X$, $G_y : X \rightarrow Y$ という X, Y 間の相互変換を学習する。ここでの相互変換は教師なし学習で行われるため、データセット間の対が必要なくデータセットの制約が緩和されている。そのため、今回のような男性の脳から女性の脳への変換といった解のない変換を行うことも可能だと考えられる。Zhu らの研究では CycleGAN の損失関数は式 (2.2) で表わされ、式 (2.2) 中の最初の 2 項で用いられる L_{GAN} は式 (2.3) で表され、3 項目の L_{cyc} は式 (2.4) で表されるとされている [8]。ここで式 (2.3) は GAN の損失関数である Adversarial Loss であり、この 2 つの Adversarial Loss によってデータセットを相互変換するように学習する。式 (2.4) は変換に一貫性を持たせるための Cycle Consistency Loss であり、式 (2.4) によって 2 度の変換で再構築した際に元画像と再構築画像の差が小さくなるように学習する。図 2.2 は CycleGAN でのデータの遷移の例としてりんごとみかんの 2 次元画像で相互変換を試した時のデータの遷移を表したものであり、変数にハットのついた画像は変換画像を表し、ダブルハットのついた画像は再構成画像を表す。図 2.2 より、Adversarial Loss によって相互変換を学習し、Cycle Consistency Loss によって G_x, G_y による変換の一貫性を持ち画像の再構成ができるように学習されていることを示している。

2.3 Attention-Guided CycleGAN

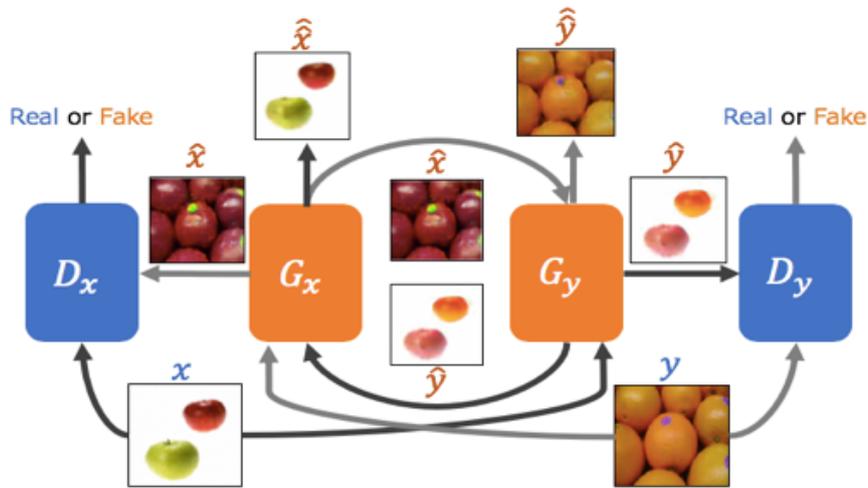


図 2.2: CycleGAN のデータの遷移

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

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

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

2.3 Attention-Guided CycleGAN

Attention-Guided CycleGAN は CycleGAN に注意機構を取り入れたモデルであり、2つのデータセット M, F に対応する生成モデル G_m, G_f と識別モデル D_m, D_f 、注意モデル A_m, A_f の6つのモデルから構成され、CycleGAN と同様に $G_m : F \rightarrow M$, $G_f : M \rightarrow F$ という M, F 間の相互変換を学習する。ここで f を変換した画像 $G_m(f)$ を \hat{m} とし、 f を再構成した画像 $G_f(\hat{m})$ を \hat{f} と表す。また注意モデルは生成モデルの変換領域を制限して識別モデルに誤った識別をさせるように学習することで、識別モデルによる識別に寄与する領域を注意領域として $[0, 1]$ の値で出力する。Mejjiati らの研究 [2] では Attention-Guided CycleGAN では注意モデルが良い注意領域を得られるまで全モデルを学習する間と得られた注意領域を用いて注意モデル以外を学習する間で識別モデルに入力するデータの形式を変えて各モデルの学習が行われている。

2.3 Attention-Guided CycleGAN

まず良い注意領域が得られる間の損失関数は式 (2.5) で表され、式 (2.5) 中の L_{adv}^m は式 (2.6), L_{cyc}^m は式 (2.7) で表されている [2]. $L_{\text{adv}}^f, L_{\text{cyc}}^f$ についても同様である.

$$L(G_m, G_f, D_m, D_f, A_m, A_f) = L_{\text{adv}}^m + L_{\text{adv}}^f + \lambda_{\text{cyc}}(L_{\text{cyc}}^m + L_{\text{cyc}}^f). \quad (2.5)$$

$$L_{\text{adv}}^m(G_m, A_m, D_m) = \mathbb{E}_{m \sim p_M(m)}[\log(D_m(m))] + \mathbb{E}_{m \sim p_F(f)}[\log(D_m(1 - \hat{m}))]. \quad (2.6)$$

$$L_{\text{cyc}}^m(m, \hat{m}) = \|m - \hat{m}\|_1. \quad (2.7)$$

この式 (2.6) の構成は CycleGAN と同様であり, $L_{\text{adv}}^m, L_{\text{adv}}^f$ はデータセット間の相互変換を促す Adversarial Loss を表し, $L_{\text{cyc}}^m, L_{\text{cyc}}^f$ は変換に一貫性を持たせるための Cycle Consistency Loss を表す. また, Attention-Guided CycleGAN の変換は式 (2.8) のように, A_m から得られる注意領域を用いて注意領域を変換し, 注意領域以外は元のデータの値を用いるといった変換を行うとされている [2]. ここで式 (2.8) 中の \odot はアダマール積を表す. A_m は D_m の識別に寄与する領域の割合を $[0, 1]$ の値で注意領域として出力するように学習するため, 式 (2.8) の変換を行うことで D_m の識別に影響する領域のみに制限した変換を行うことができる. \hat{f} についても同様である.

$$\hat{m} = A_m(f) \odot G_m(f) + (1 - A_m(f)) \odot f. \quad (2.8)$$

図 2.3 は Attention-Guided CycleGAN でのデータの遷移の例として, りんごとみかんの 2 次元画像で相互変換を試した時のデータの遷移を表したものである. また図 2.3 中のモジュール M_m, M_f は式 (2.8) での処理を行うモジュールであり, モジュール内でのデータの遷移は図 2.4 で表す. 図 2.3, 図 2.4 より, Attention-Guided CycleGAN でも CycleGAN と同様に相互変換と画像の再構成を学習していることが分かり, 図 2.4 は注意モデル A_m, A_f が変換すべき領域を注意領域として選択した上での変換を行っていることを示している.

次に得られた注意領域を用いて注意モデル以外を学習する間は, 識別モデル G_m に入力する元データ m , 変換データ \hat{m} は式 (2.9) で表すように注意領域の値が閾値 τ を超えた領域内のみデータ $m_{\text{new}}, \hat{m}_{\text{new}}$ を用い, L_{adv}^m は式 (2.10) で表すような Adversarial Loss になるとされている [2]. これにより, 生成モデル, 識別モデルが共に注意領域内に注力した変換や識別を行うことになる.

2.3 Attention-Guided CycleGAN

本研究では、Attention Guided CycleGAN から得られる注意領域や変換画像に着目し、脳構造画像における性差の分析を行う。

$$m_{\text{new}} = \begin{cases} m & \text{if } A_m(m) > \tau \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{m}_{\text{new}} = \begin{cases} \hat{m} & \text{if } A_m(f) > \tau \\ 0 & \text{otherwise} \end{cases}. \quad (2.9)$$

$$L_{\text{adv}}^m(G_m, A_m, D_m) = \mathbb{E}_{m \sim p_M(m)}[\log(D_m(m_{\text{new}}))] + \mathbb{E}_{m \sim p_F(f)}[\log(D_m(1 - \hat{m}_{\text{new}}))]. \quad (2.10)$$

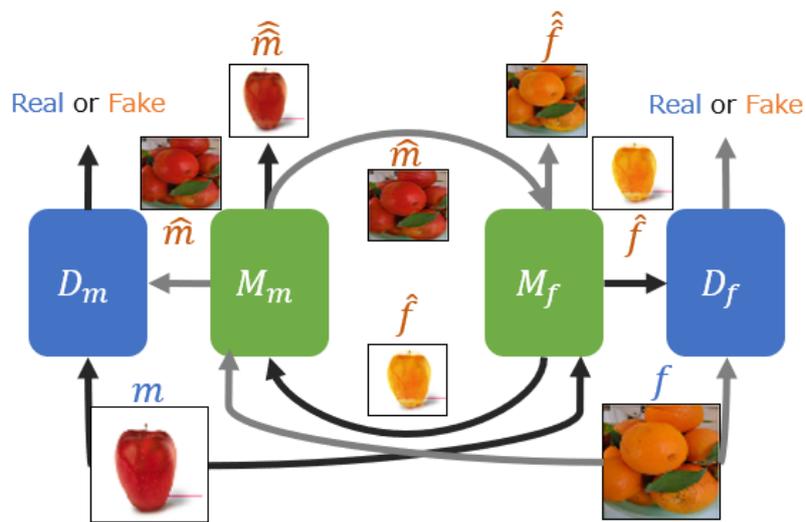


図 2.3: Attention-Guided CycleGAN のデータの遷移

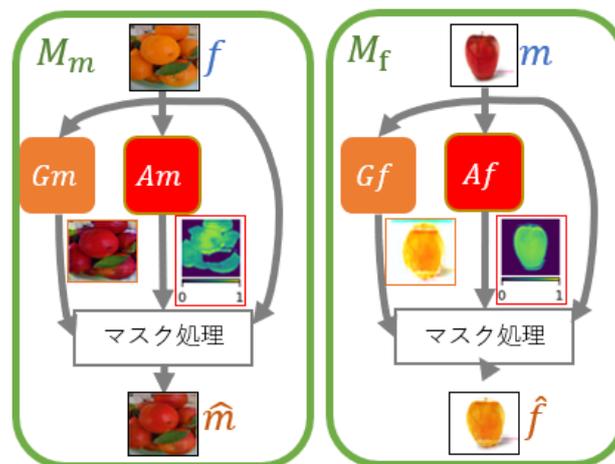


図 2.4: Attention-Guided CycleGAN のモジュール内でのデータの遷移

2.4 CNN(Convolutional Neural Network)

CNN は畳み込みニューラルネットワークと呼ばれ、1次元信号の時系列データや画像などの2次元行列データ、3次元テンソルデータなどを用いた機械学習に用いられる。CNNは1959年頃にHubelとWieselの研究[9]で提案された生物の視覚野の脳神経回路網モデルをもとに1980年頃に福島が考案したNeocognitron[10]というモデルが原型となっている。このNeocognitronは、受容野内の刺激の特徴を抽出する単純型細胞、単純型細胞が抽出した特徴の位置ずれを吸収する複雑型細胞を表す2つの層を交互に重ねることで構成される。Neocognitronでは、異なる受容野で抽出された複数の特徴を用いて複雑な形状を認識できる受容野を形成し、文字などの認識を可能にしている。この単純型細胞と複雑型細胞の役割は現在のCNNでは畳み込み層、プーリング層と呼ばれる層が担っている。また、Neocognitronの学習は教師なし学習で行われ、個々の細胞が全体での認識のために役割を持つ自己組織化という現象によって文字などの提示された刺激パターンを認識できるようになるとされている[10]。その後、1998年頃にLeCunらが手書き文字認識のために、Neocognitronに誤差逆伝搬法を適用して教師あり学習で学習できるようにしたLeNet[11]を提案した。さらに2012年にはImagenetと呼ばれる大規模な画像データセットを用いた画像認識コンペティションでAlexらが提案したAlexNet[12]が従来手法を大きく上回る結果で優勝し、画像認識においてCNNが広く注目されるきっかけとなった。

2.4.1 CNNの基本構造

次にCNNの基本的な構造について説明を行う。図2.5はサイズが 32×32 でチャンネル数1のグレースケール画像の3クラス分類を行うCNNを表したものである。

CNNは図2.5のように畳み込み層(Conv)とプーリング層(Pooling)を交互に接続した構造を持ち、出力層付近では隣接層間のノードを全て結合した全結合層を1層以上持つとされている[13]。ここで畳み込み層は特徴の抽出、プーリング層は得られた特徴の位置ずれの吸収を行っており、全結合層(Full Connection)は各特徴に対して重みをつけること

2.4 CNN(Convolutional Neural Network)

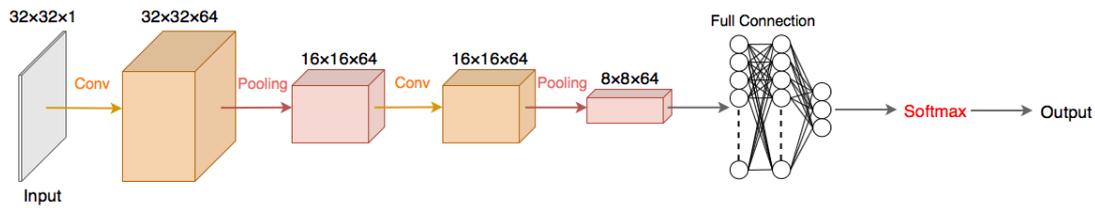


図 2.5: 3 クラス分類を学習する CNN の基本的な構造

で有益な情報を選択した上でのクラスの推定を可能としている。また、CNN の出力を与える最終層では目的のクラス数と同数のノードを配置し、活性化関数であるソフトマックス関数 (Softmax) を通した出力 P_j が最大のノードを推定されたクラスとして出力するとされている [13]。ここでソフトマックス関数の出力は最終層の n 個の各ノードからの入力を $U_j (j \in \{1, \dots, n\})$ と表したとき、式 (2.11) のように計算される。式 (2.11) より、ソフトマックス関数の出力は $0 \leq P_j \leq 1$ であり、 $\sum_{j=1}^n P_j = 1$ であることから、これらの値はそのまま各クラスの確率として用いることができる。CNN 全体としては畳み込み層、プーリング層による特徴抽出を行なった後、各クラスごとに分類に有益な特徴を選択し、その特徴を用いた推定を行う流れでクラス分類を行なっている。

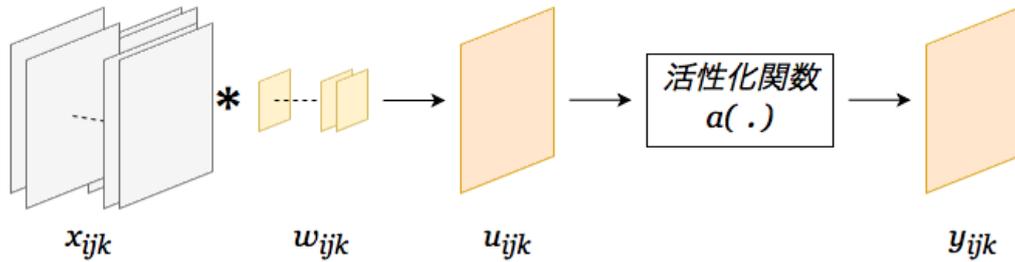
次に畳み込み層、プーリング層について説明を行う。

$$P_j = \frac{e^{U_j}}{\sum_{k=1}^n e^{U_k}}. \quad (2.11)$$

2.4.2 畳み込み層 (Convolution layer)

畳み込み層では入力に対して $L \times L$ のフィルタを用いたフィルタ処理を行い、入力からフィルタが表す特徴を抽出する。これはフィルタ処理を行ってエッジ抽出を行うのと同様の処理である。畳み込み層での処理は、図 2.6 で表すように畳み込みを行った後に活性化関数を通して出力を得ている。ここで畳み込み層へのサイズ $S \times S$ 、チャンネル数 C の入力を $x_{ijk} ((i, j, k) \in \{0, \dots, S-1\} \times \{0, \dots, S-1\} \times \{1, C\})$ と表し、フィルタを $w_{ijk} ((i, j, k) \in \{0, \dots, L-1\} \times \{0, \dots, L-1\} \times \{1, \dots, C\})$ 、各チャンネルごとのバイアス項を $b_k (k \in \{1, \dots, C\})$ と表す。この時、入力に対して畳み込みを行った結果として得られる

2.4 CNN(Convolutional Neural Network)



「深層学習-Deep Learning-」人工知能学会 p.158(2015)を参考に作成

図 2.6: 畳み込み層での処理の流れ

u_{ij} は式 (2.12) のように計算するとされている [13]. ここで P_{ij} は入力中の画素 (i, j) を頂点とする $L \times L$ の領域であり, $P_{ij} = \{(i + i', j + j') | i' = 0, \dots, L - 1, j' = 0, \dots, L - 1\}$ と表せる. この P_{ij} は画素 (i, j) に対して畳み込みを行う際に用いる周囲の画素の集合といえる. 式 (2.12) より, 畳み込み層では入力の全てのチャンネル上のある画素上で畳み込みを行い, 各チャンネルの結果を加算することで u_{ij} を計算していることが分かる.

$$u_{ij} = \sum_{k=1}^C \left[\sum_{(p,q) \in P_{ij}} x_{pqk} w_{p-i, q-j, k} \right] + b_k. \quad (2.12)$$

また入力のサイズが大きい場合, $(0,0), (1,0), (2,0) \dots$ と 1 画素ごとに畳み込みを行うのではなく, $(0,0), (2,0), (4,0) \dots$ と数画素ごとにストライドさせて畳み込みを行うことがある. この畳み込みの間隔をストライド数と呼び, ストライド数は 2 以上の整数であり, ストライド数を s とすると P は $P_{ij} = \{(si + i', sj + j') | i' = 0, \dots, L - 1, j' = 0, \dots, L - 1\}$ と表される. この時, u_{ij} の計算は式 (2.13) のように, $w_{p-i, q-j, k}$ を $w_{p-si, q-sj, k}$ に置き換えることで計算できるとされている [13]. ストライドを用いた場合, s 画素ごとに畳み込みで値を求めるため, u_{ij} のサイズは入力サイズの $\frac{1}{s^2}$ となる. これにより, 以降の処理の入力サイズが小さくなり, ネットワークのサイズを小さくすることができる. しかし, 特徴を取る画素の間隔が開くことが特徴を取りこぼすことに繋がるため, 性能は低下する傾向にあるとされている [13].

2.4 CNN(Convolutional Neural Network)

$$u_{ij} = \sum_{k=1}^C \left[\sum_{(p,q) \in P_{ij}} x_{pqk} w_{p-2i, q-2j, k} \right] + b_k. \quad (2.13)$$

次に、CNN 全体の学習の収束性や学習速度の向上のために活性化関数を用いる。ここでは活性化関に数 ReLU 関数を用いる。畳み込みによって得られた u_{ij} から式 (2.14) で表す ReLU 関数 $a(\cdot)$ を用いて畳み込み層の出力となる y_{ij} を算出する。式 (2.14) から ReLU 関数は負の値を 0 にする関数であることが分かる。この ReLU 関数を用いることで出力が疎になり、勾配が学習の収束を意味する $\mathbf{0}$ になりやすいとされている。

$$y_{ij} = a(x_{ij}) = \max(0, x_{ij}) = \begin{cases} x_{ij} & \text{if } x_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}. \quad (2.14)$$

以上の処理を経て畳み込み層では 1 つのフィルタからサイズ $S \times S$ 、チャンネル数 1 の特徴マップを得ることができ、特徴マップは入力中のどの部分に特徴があったのかを表している。また、畳み込み層では複数個のフィルタを用いて特徴抽出を行うことで 1 つの入力から多様な特徴を抽出できるようになる。その場合、 N 個のフィルタを用いた畳み込み層の最終的な出力の形状は $S \times S \times N$ の特徴マップとなる。

CNN の学習の際には畳み込み層では、 u_{ij} の算出に用いたフィルタを表す w_{ijk} とバイアス項 b_k の値を更新を行っている。また、2 層目以降の畳み込み層は入力に特徴マップを用いるため、異なる特徴を組み合わせたより複雑な特徴を抽出することができる。そのため、畳み込みニューラルネットワークでは入力から見て浅い層で画像中のエッジなどの単純な特徴を抽出し、深い層でそれらの特徴を組み合わせた複雑な形状の特徴を抽出し、最終的に人の顔であれば目や鼻などの特徴が得られるように学習が行われる。

また、特徴マップの各ピクセル値は図 2.7 で表すようにフィルタと同じサイズの入力の一部の領域を需要野として算出されている。後に説明する PatchGAN ではこの仕組みを用いることで、入力時に複数のパッチに分けずに CNN 内でパッチごとに推定できるように実装を行う。

2.4 CNN(Convolutional Neural Network)

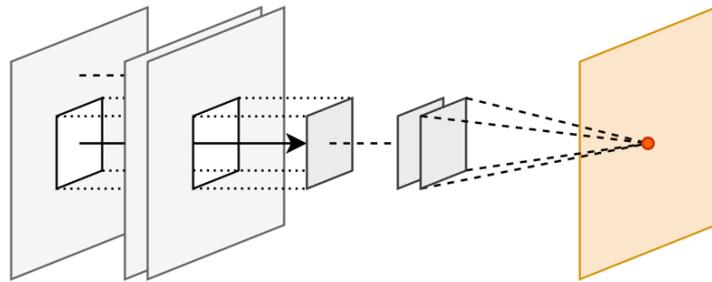


図 2.7: 入力の特徴マップの各ピクセル値

2.4.3 プーリング層

プーリング層の目的は、畳み込みによって得られた $S \times S \times N$ の特徴マップ上の特徴の位置情報を一部捨て、入力ごとの特徴の微小な位置変化に対する応答の不変性を実現することにあるとされている [13]. 例えば人の顔を認識する場合、口といった顔のパーツの位置は人によってある程度異なる。位置情報が全く捨てられない場合、畳み込み層は学習データに従って特徴を抽出するように学習するため、顔のパーツが学習時に用いたデータと異なる位置にある人の顔が人間の顔として認識されなくなる可能性がある。そこでプーリング層では、特徴マップの位置情報を一部捨て大まかに特徴の位置を見ることで、ある程度の位置ズレを許容した上での認識を可能にしている。これにより、学習データに無い特徴のパターンでもある程度許容して認識できるようになる。

ここではプーリングの方法として平均プーリングと最大プーリングについて記述する。プーリング層では特徴マップ上のある画素 (i, j) に対して $L \times L$ の受容野を持ち、ここではその受容野を $P_{ij} = \{(si + i', sj + j') | i' = 0, \dots, L - 1, j' = 0, \dots, L - 1\}$ と表す。また、通常プーリング層では畳み込み層のストライドと同様に数画素ごとに出力を求めるとされている [13]. このストライドの間隔をストライド数 s とすると、畳み込みと同様に得られる出力の大きさは入力の $\frac{1}{s^2}$ 倍になる。ここで特徴マップにおけるチャンネル k のある画素 (i, j) の値を x_{ijk} 、プーリングによって得られる出力値を y_{ij} とする。この時、平均プーリングは式 (2.15) で表すように受容野内の平均値を算出するとされている [13].

2.5 PatchGAN

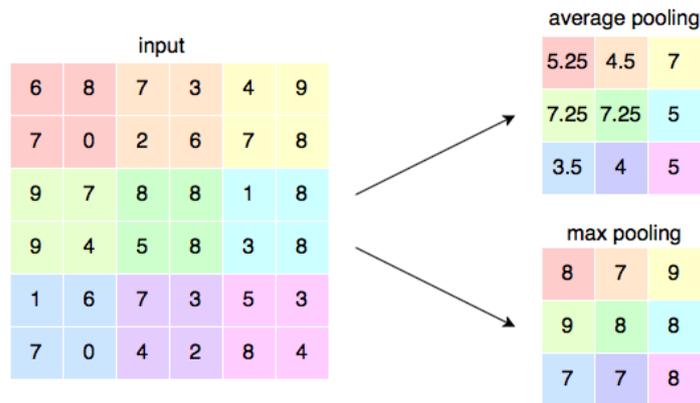


図 2.8: 平均・最大プーリングの例

$$y_{ijk} = \frac{1}{|P_{ij}|} \sum_{(p,q) \in P_{ij}} x_{pqk}. \quad (2.15)$$

また最大プーリングは式 2.16 で表すように受容野内の最大値を算出するとされている [13].

$$y_{ijk} = \max_{(p,q) \in P_{ij}} x_{pqk}. \quad (2.16)$$

図 2.8 は入力 (input) に対して 2×2 の受容野, スライド数 2 の場合の平均プーリング (average pooling), 最大プーリング (max pooling) を行なった場合の出力を表した例である.

2.5 PatchGAN

画像を生成する問題に対して L1, L2 損失を用いた場合にぼやけた画像が生成されるという問題がある. この問題に対して 2 次元画像を対象にした Isola らの研究 [14] では, 画像全体を $N \times N$ のパッチに分けて各パッチごとに GAN の識別モデルが識別する手法である PatchGAN を提案している. PatchGAN を用いずに画像全体を識別する手法では, 画像中の低周波成分を正確に捉えた生成が行われ, 結果としてぼやけた画像が生成されるという問題が発生していた. この問題に対して PatchGAN を用いることで, 識別モデルがパッチという高周波成分を識別するようになり, 結果としてより高品質な結果が得られたとされてい

2.6 U-Net

る [14]. この PatchGAN を実装する際に、実際にパッチごとに識別すると計算にかかる時間やメモリなどのコストが大きくなる。そこで、CNN から得られる特徴マップの各値は入力画像の一部の領域から影響を受けた値になることから、CNN の特徴マップの各値を用いて識別することは PatchGAN と同等であるみなすことができる。よって、PatchGAN の実装では低コスト化のために CNN を用いた手法が用いられる。

2.6 U-Net

U-Net は Ronneberger らの研究 [15] で提案された CNN のネットワーク構造であり、高精度なセグメンテーションを学習する。図 2.9 は 3 段の U-Net のネットワーク構造を表した図である。図 2.9 のように U-Net は特徴マップを縮小する過程と特徴マップをを拡大する過程の 2 つの過程からなり、拡大過程で生成した特徴マップと同じ大きさの縮小過程の特徴マップを結合しながら拡大を行い、最後に畳み込みによって元画像と同じサイズの出力を得るような構造を持っている。ここで図 2.9 で入力画像や特徴マップ上の数字は入力画像のチャンネル数、特徴マップの数を表す。縮小過程では Max Pooling によって 2×2 のフレーム内で最大の値を抽出することで画像サイズを $1/2$ に縮小している。また、拡大過程では 1 つのピクセル値を畳み込みによって 2×2 の大きさに拡大している。U-Net ではこのように異なるスケールの特徴マップを用いて学習することで画像中の様々な大きさの物体に対応したセグメンテーションが可能になっていると考えられる。

2.6 U-Net

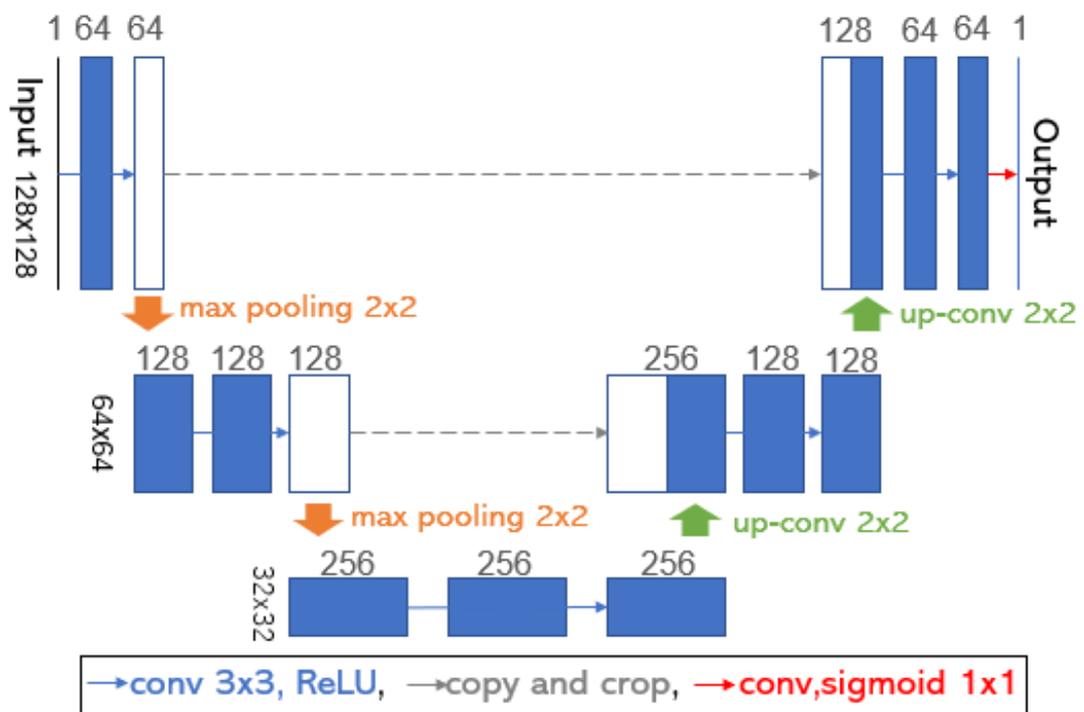


図 2.9: U-Net 全体のネットワーク構造

第 3 章

提案手法

文献 [1] より, CNN を用いることで男女の脳構造画像を高精度で識別できるとされており, CNN の学習後の重みを用いて識別に寄与する領域の特定を行う Grad-CAM を用いた性差の特定も試みられていた. しかし, あまり意味のある領域の特定には至っておらず, 識別の理由の説明はほとんどできていない. また, 性差の説明のためには領域だけでなく領域内での形状やパターンの違いを知る必要があり, 領域の特定だけでは十分でない. そこで性差の説明のために男女の識別に関わる領域を特定し, 領域内での形状やパターンの違いを得ることができる手法が必要となっている.

CNN がどのような形状やパターンを見ているかを知る方法として, CNN のフィルタからどのような特徴を見ているのか調べる方法が考えられる. しかし, CNN では浅い層で抽出した単純な特徴を深い層で組み合わせることでより複雑な特徴を抽出できるように学習するため, CNN 自体が何を見ているかを知るためには複雑な特徴の表現に用いられた複数の単純な特徴を知り, それらからどのような特徴が表現されるかを考える必要がある. また, CNN には各層で多くのフィルタが用いられているため, 複数の層で複数のフィルタに対して同様に考える必要があり, 実際にフィルタから識別過程を説明することは現実的ではない. そこで本研究では, 過程ではなく結果から CNN の識別に関わる形状やパターンを知る方法として敵対的生成ネットワーク (GAN) を用いた手法を提案する.

GAN を用いて性差の分析を行うためには, 男女間での性差を学習できるモデルを用いる必要がある. そこで性差を学習できるモデルとして, GAN を応用したモデルである CycleGAN が挙げられる. CycleGAN では教師なし学習でデータセット間の相互変換を学習するため, データセット間でデータの対が必要なく, 男女の脳構造画像の相互変換といっ

た現実には解のない変換も学習できると考えられる。また、男女の脳構造画像の相互変換を行うことで、変換前後の差分から男女間での形状やパターンの違いを分析できると考えられる。しかし、CycleGAN では画像全体に対して変換を行うため、差分から性差に関わる領域を特定できない。そこで本研究では、CycleGAN に Attention 機構を導入し、変換領域の制限を行なった Attention-Guided CycleGAN を用いた分析を提案する。

本研究では性差の識別に関わる領域の特定と領域内での形状やパターンの違いを分析するために Attention-Guided CycleGAN を用いた分析手法を提案する。ここで Attention-Guided CycleGAN を用いた理由として、Attention-Guided CycleGAN では注意領域によって変換領域を制限した上で2つのデータセットの相互変換を教師なし学習で学習することができる。ここで変換が教師なし学習で学習されるため、CycleGAN と同様に男女の脳構造画像の相互変換といった解のない変換も学習できると考えられる。また Attention-Guided CycleGAN では、注意領域で表される男女の識別に関わる領域を変化させ、一方の性別の脳構造画像をもう一方の性別の脳構造画像に変換することができる。これにより、変換の際の注意領域を見ることで男女の識別に寄与する領域の特定につなげることができると考えられる。また、変換前後の差から男女間での形状やパターンの違いも見ることができると考えられる。よって、性差の分析に Attention-Guided CycleGAN を用いることで男女の性差に関わる領域の特定や領域内での形状やパターンの分析が行えると考え、本研究では実験を行なった。

第 4 章

実験内容

ここでは本研究で提案する Attention-Guided CycleGAN による脳の変換過程の学習モデル構築と、その実験で用いるデータセット、実験手順、使用モデルについて記述する。

4.1 データセット

本研究ではデータセットに男性 207 名, 女性 157 名の健康な脳ドック受診者の T1 画像を使用した。データ構造はチャンネル数 1 の 3 次元データである。

またデータの前処理は以下の手順で行った。

1. 個人差の軽減のために SPM(Statistical Parametric Mapping) を用いて頭の大きさ, 形状を標準脳に合わせる。
2. SPM を用いて全脳, 灰白質, 白質の 3 種類のデータを抽出。
3. データセットを男女に分割。

4.2 学習手順

本研究では, Attention-Guided CycleGAN の学習を以下の手順で行った。

1. データの形状を $256 \times 256 \times 180$ から $128 \times 128 \times 96$ に変形。
2. データの値を平均 0, 標準偏差 1 で標準化
3. エポック数 200, バッチサイズ 2 で全モデルを学習。
4. 識別に有効な注意領域を選択できていると考えられるエポックを選択し, エポック数

4.3 使用モデル

100, バッチサイズ 2 で注意モデル以外を学習.

ここでデータの標準化を行うことで脳構造画像の被験者間の明暗差を小さくしている. また, 識別に有効な注意領域を選択して注意モデル以外を学習する際は Mejjati らの研究 [2] で行われている手法に則り, 注意領域の値が 0.1 以下の場合には注意領域の値を 0 にするように閾値処理を行っている.

4.3 使用モデル

本研究では以下に示す構造の生成モデル, 識別モデル, 注意モデルを用いて Attention-Guided CycleGAN を構成し, 実験に用いた.

4.3.1 生成モデル

生成モデルでは図 4.1 に表すようなモデルを用いた. 図 4.1 中の u はアップサンプリングの倍率を表し, f は特徴マップの数, k はカーネルサイズ, s はストライド数, d はドロップアウトの閾値, α は Leaky ReLU の負の部分の傾きを表す. また生成モデルの構造にはセグメンテーションなどに用いられる U-Net を用いており, それによって様々なスケールの特徴を反映した変換が可能になると考えられる. 本研究では, 生成モデルは入力された 3 次元脳画像をもう一方の性別の脳になるように変換し, 2 つの生成モデルを用いた男女間の相互変換を学習させる.

4.3.2 識別モデル

識別モデルでは図 4.2 に表すようなモデルを用いた. 図 4.2 中の f は特徴マップの数, k はカーネルサイズ, s はストライド数を表す. また, 識別モデルでは PatchGAN の実装のために最後の出力層は畳み込み層となっている. 本研究では, 識別モデルは各性別ごとに変換されていない 3 次元脳画像と変換された 3 次元脳画像を識別するように学習させる.

4.3 使用モデル

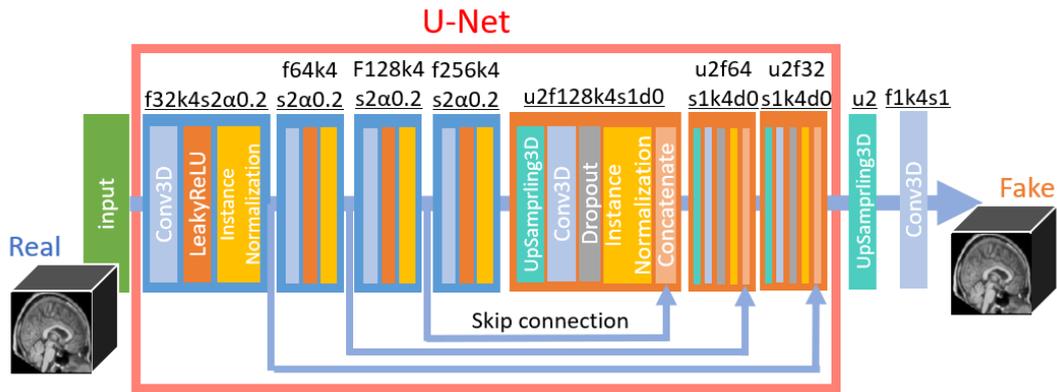


図 4.1: 生成モデルの構造

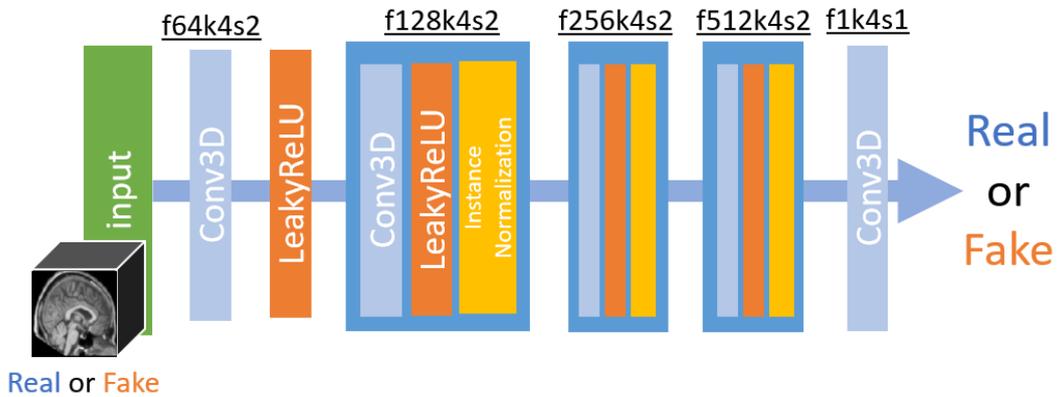


図 4.2: 識別モデルの構造

4.3.3 注意モデル

注意モデルでは図 4.3 に表す構造のモデルを用いた。図 4.3 中の u はアップサンプリングの倍率を表し、 f は特徴マップの数、 k はカーネルサイズ、 s はストライド数、 m は Batch Normalization における慣性項の係数を表す。また、3次元畳み込み層 (Conv3D) の初期化には平均 1、標準偏差 0.02 の切断正規分布から生成される乱数を用いた、本研究では、この注意モデルが出力する $[0, 1]$ の値を持つ 3次元の注意領域を用いて生成モデルの変換領域を制限するように学習させる。また、この注意領域を見ることで変換される領域を特定することができ、その領域内の 3次元脳画像の形状やパターンの変化を見て性差の分析を行う。

4.3 使用モデル

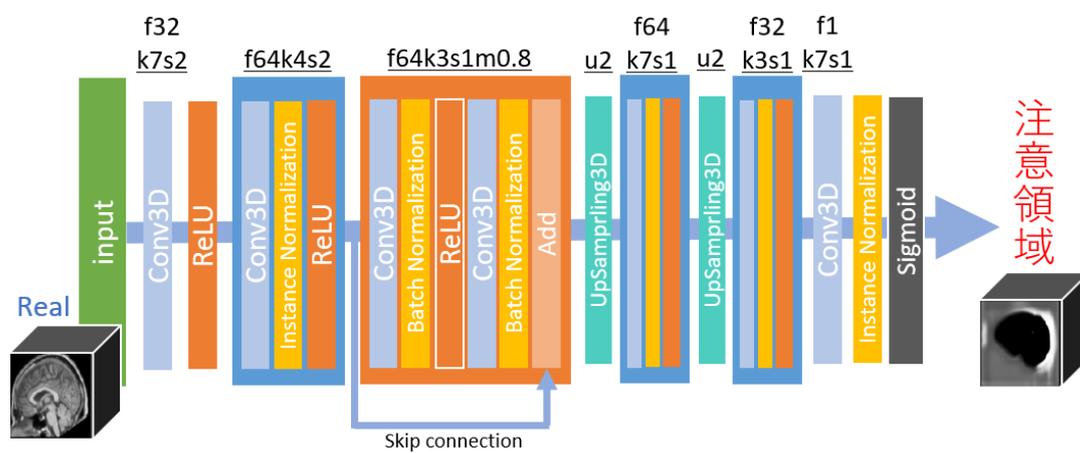


図 4.3: 注意モデルの構造

第 5 章

実験結果

以下に実験の結果を全脳, 灰白質, 白質に分けて示す. また結果の図における Coronal(冠状断面) は脳を前後に分割した時の断面を表し, Axial は脳を上下に分割したときの断面を表し, Sagittal(矢状断面) は脳を左右に分割したときの断面を表す.

5.1 全脳

Attention-Guided CycleGAN を用いた全脳の男女相互変換を行った結果, 男性 → 女性 → 男性 という変換では図 5.1 のような結果が得られた. また, 女性 → 男性 → 女性 という変換では図 5.2 のような結果が得られた. また, 比較のために図 5.1, 5.2 を並べた結果として, 図 5.3 のように 男性 → 女性 → 男性 の変換で後頭葉を削るような結果が得られた. さらに, データセット全体での変換傾向を見るために元画像, 変換画像の平均画像と平均画像から元画像を引いた差分画像を作成し値域を-0.1 から 0.1 として描画した結果として, 図 5.4 が得られ, こちらも同様に 男性 → 女性 → 男性 の変換で後頭葉を削るような結果が見られた. また 女性 → 男性 → 女性 の変換では頭頂葉から後頭葉にかけて削るような変換が見られた.

また図 5.5 は最初に全てのモデルを 200 エポック学習させた際の識別モデルの正解率を表しており, 2つの識別モデルの正解率 (Accuracy) の平均をとって算出している. 今回はこの図 5.5 より全脳では, 識別に寄与する領域を選択できているエポックとして 150 エポック目の注意領域を選択して, 識別モデルと生成モデルを追加で 100 エポック学習した. 図 5.6 はその間の正解率の遷移を表している.

5.1 全脳

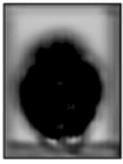
男性→女性→男性					
	元画像	注意領域	変換画像	注意領域	再構成画像
Coronal (冠状断面)					
Axial (体軸断面)					
Sagittal (矢状断面)					

図 5.1: 全脳の変換結果 男性 → 女性 → 男性

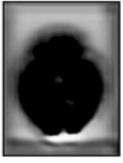
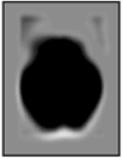
女性→男性→女性					
	元画像	注意領域	変換画像	注意領域	再構成画像
Coronal (冠状断面)					
Axial (体軸断面)					
Sagittal (矢状断面)					

図 5.2: 全脳の変換結果 女性 → 男性 → 女性

5.1 全脳

	男性→女性→男性			女性→男性→女性		
	元画像	変換画像	再構成画像	元画像	変換画像	再構成画像
Coronal (冠状断面)						
Axial (体軸断面)						
Sagittal (矢状断面)						

図 5.3: 全脳の変換結果の比較

	男性→女性→男性			女性→男性→女性		
	平均画像	変換画像	差分画像	平均画像	変換画像	差分画像
Coronal (冠状断面)						
Axial (体軸断面)						
Sagittal (矢状断面)						

図 5.4: 全脳の変換結果の平均画像とその差分

5.1 全脳

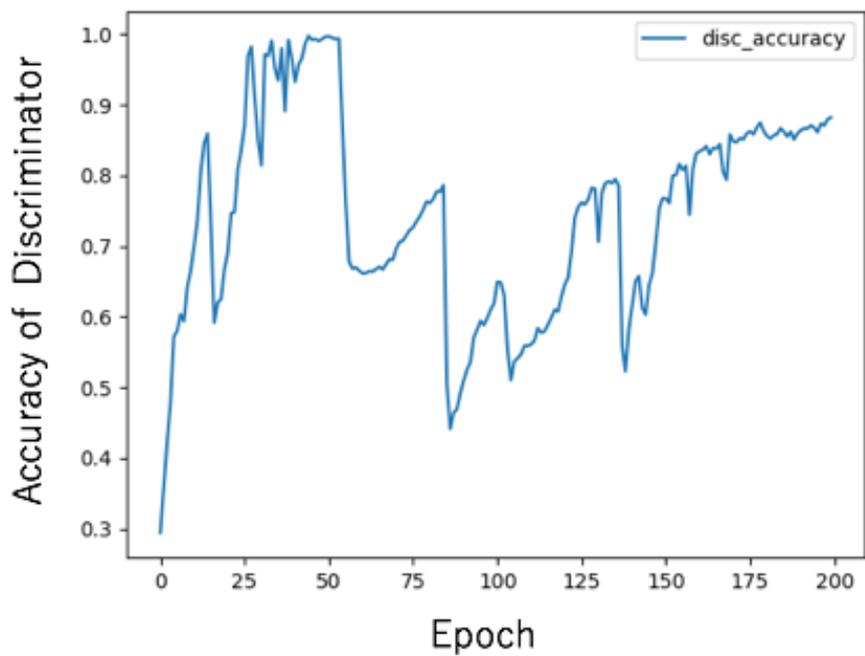


図 5.5: 全脳における識別モデルの正解率 (全モデル学習時)

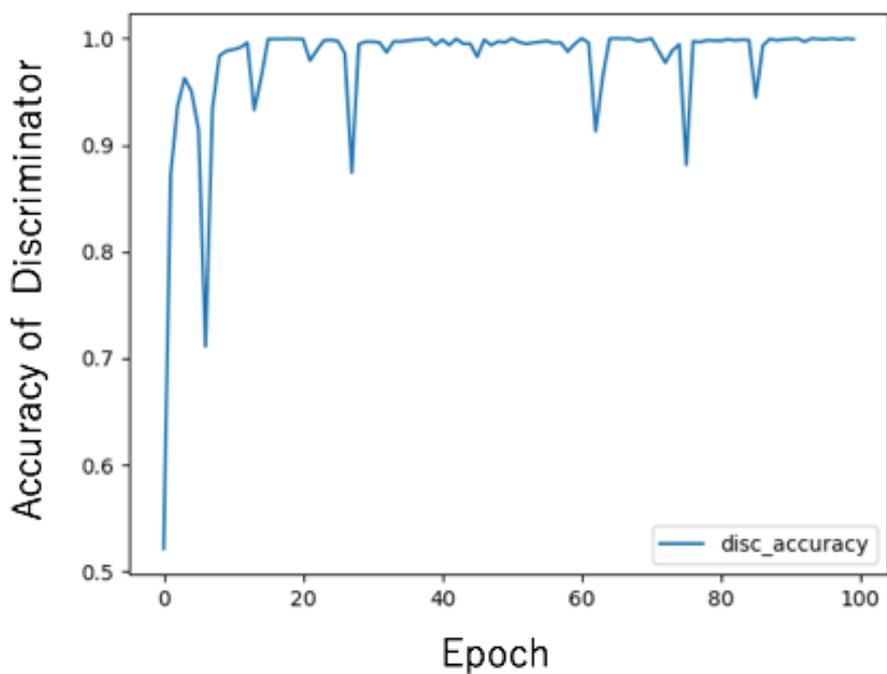


図 5.6: 全脳における識別モデルの正解率 (注意モデル以外学習時)

5.2 灰白質

同様に Attention-Guided CycleGAN を用いた灰白質の男女相互変換を行った結果、男性 → 女性 → 男性 という変換では図 5.7 のような結果が得られ、女性 → 男性 → 女性 という変換では図 5.8 のような結果が得られた。また、比較のために図 5.7, 5.8 を並べた結果として、図 5.9 が得られ、男性 → 女性 → 男性 の変換では小脳を削るような変換が見られた。さらに、変換画像の平均画像と平均画像から元画像を引いた差分画像を作成し値域を-0.1 から 0.1 として描画した結果として図 5.10 が得られ、平均画像でも図 5.9 と同様に小脳の下部を削るような変換が見られた。しかし、女性 → 男性 → 女性 の変換では変化が見られなかった。

また図 5.11 は最初に全てのモデルを 200 エポック学習させた際の識別モデルの正解率を表している。今回はこの図 5.11 より灰白質では、識別に寄与する領域を選択できているエポックとして 70 エポック目の注意領域を選択して識別モデルと生成モデルを追加で 100 エポック学習した。図 5.12 はその間の正解率の遷移を表している。

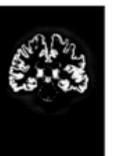
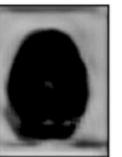
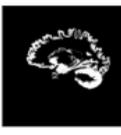
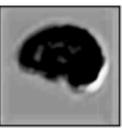
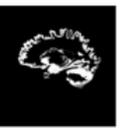
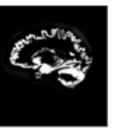
	男性→女性→男性				
	元画像	注意領域	変換画像	注意領域	再構成画像
Coronal (冠状断面)					
Axial (体軸断面)					
Sagittal (矢状断面)					

図 5.7: 灰白質の変換結果 男性 → 女性 → 男性

5.2 灰白質

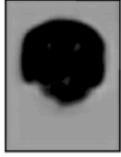
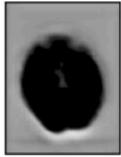
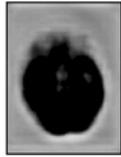
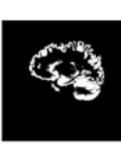
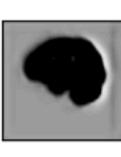
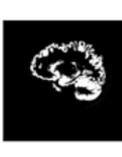
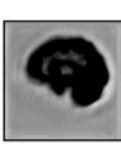
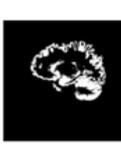
		女性→男性→女性				
		元画像	注意領域	変換画像	注意領域	再構成画像
Coronal (冠状断面)						
Axial (体軸断面)						
Sagittal (矢状断面)						

図 5.8: 灰白質の変換結果 女性 → 男性 → 女性

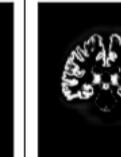
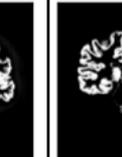
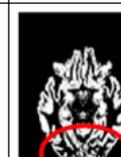
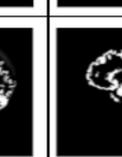
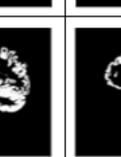
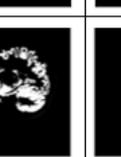
		男性→女性→男性			女性→男性→女性		
		元画像	変換画像	再構成画像	元画像	変換画像	再構成画像
Coronal (冠状断面)							
Axial (体軸断面)							
Sagittal (矢状断面)							

図 5.9: 灰白質の変換結果の比較

5.2 灰白質

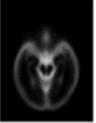
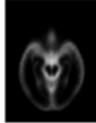
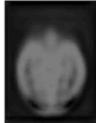
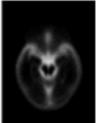
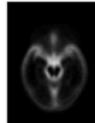
	Male→Female→Male			Female→Male→Female		
	元画像	変換画像	差分画像	元画像	変換画像	差分画像
Coronal (冠状断面)						
Axial (体軸断面)						
Sagittal (矢状断面)						

図 5.10: 灰白質の元画像と変換画像の平均画像とその差分

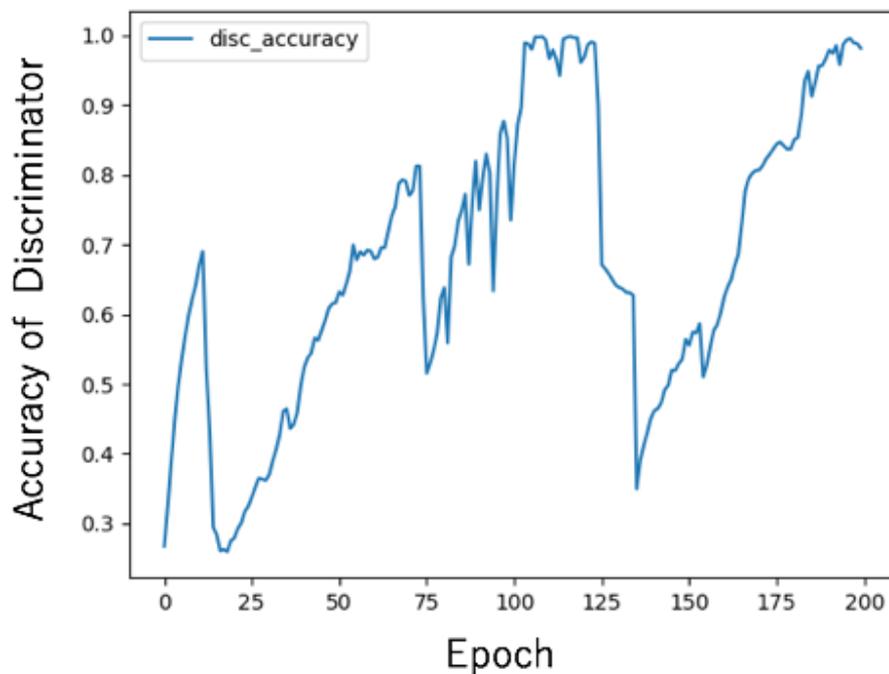


図 5.11: 灰白質における識別モデルの正解率 (全モデル学習時)

5.2 灰白質

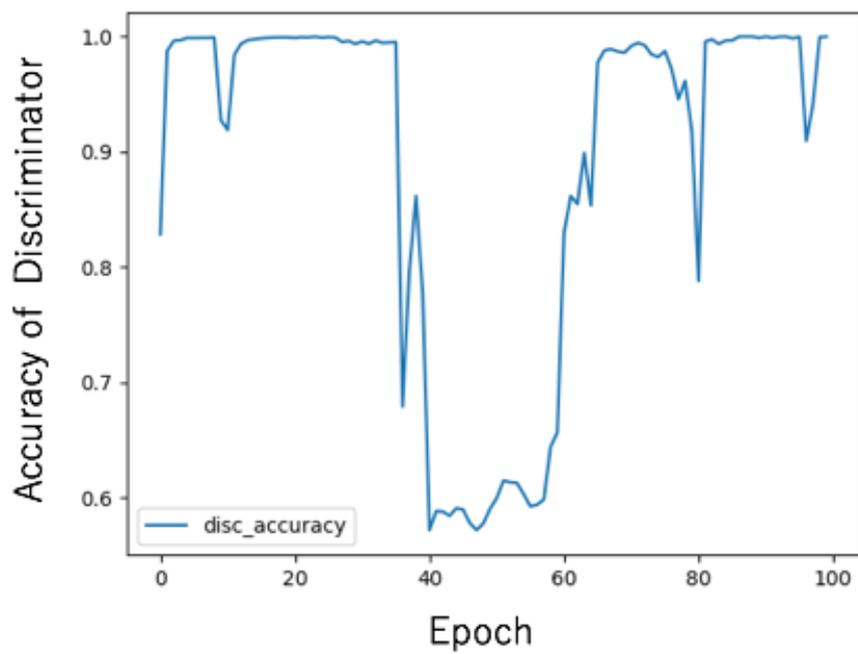


図 5.12: 灰白質における識別モデルの正解率 (注意モデル以外学習時)

5.3 白質

同様に Attention-Guided CycleGAN を用いた白質の男女相互変換を行った結果、男性 → 女性 → 男性 という変換では図 5.13 のような結果が得られ、女性 → 男性 → 女性 という変換では図 5.14 のような結果が得られた。また、比較のために図 5.13, 5.14 を並べた結果として図 5.15 のような結果が得られ、男性 → 女性 → 男性 の変換では後頭葉を削るような変換が見られた。さらに、変換画像の平均画像と平均画像から元画像を引いた差分画像を作成し値域を-0.1 から 0.1 として描画した結果として図 5.16 が得られ、男性 → 女性 → 男性 の変換では後頭葉を削るような変換が見られた。また、女性 → 男性 → 女性 の変換では後頭葉に新たな領域を生成するような変換が見られた。

また図 5.17 は最初に全てのモデルを 200 エポック学習させた際の識別モデルの正解率を表している。今回はこの図 5.17 より灰白質では、識別に寄与する領域を選択できているエポックとして 130 エポック目の注意領域を選択して識別モデルと生成モデルを追加で 100 エポック学習した。図 5.18 はその間の正解率の遷移を表している。

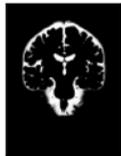
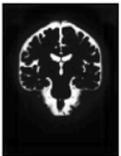
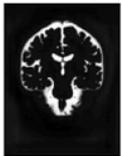
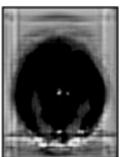
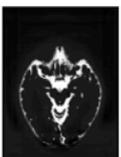
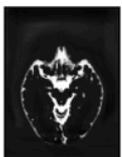
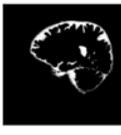
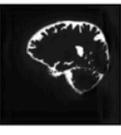
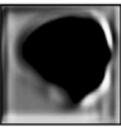
	男性→女性→男性				
	元画像	注意領域	変換画像	注意領域	再構成画像
Coronal (冠状断面)					
Axial (体軸断面)					
Sagittal (矢状断面)					

図 5.13: 白質の変換結果 男性 → 女性 → 男性

5.3 白質

女性→男性→女性					
	元画像	注意領域	変換画像	注意領域	再構成画像
Coronal (冠状断面)					
Axial (体軸断面)					
Sagittal (矢状断面)					

図 5.14: 白質の変換結果 女性 → 男性 → 女性

	男性→女性→男性			女性→男性→女性		
	元画像	変換画像	再構成画像	元画像	変換画像	再構成画像
Coronal (冠状断面)						
Axial (体軸断面)						
Sagittal (矢状断面)						

図 5.15: 白質の変換結果の比較

5.3 白質

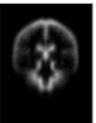
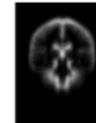
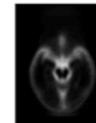
	Male→Female→Male			Female→Male→Female		
	元画像	変換画像	差分画像	元画像	変換画像	差分画像
Coronal (冠状断面)						
Axial (体軸断面)						
Sagittal (矢状断面)						

図 5.16: 白質の元画像と変換画像の平均画像とその差分

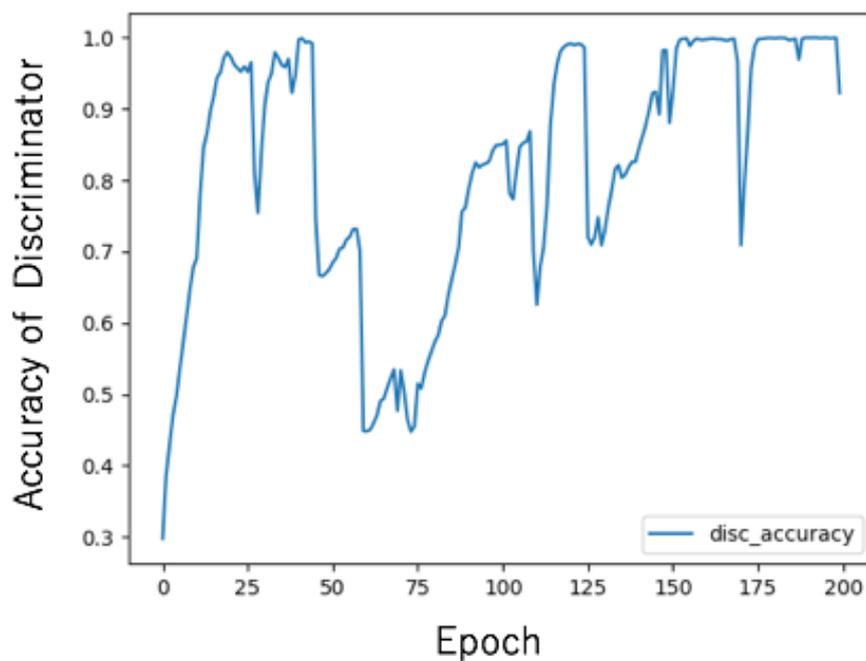


図 5.17: 白質における識別モデルの正解率 (全モデル学習時)

5.3 白質

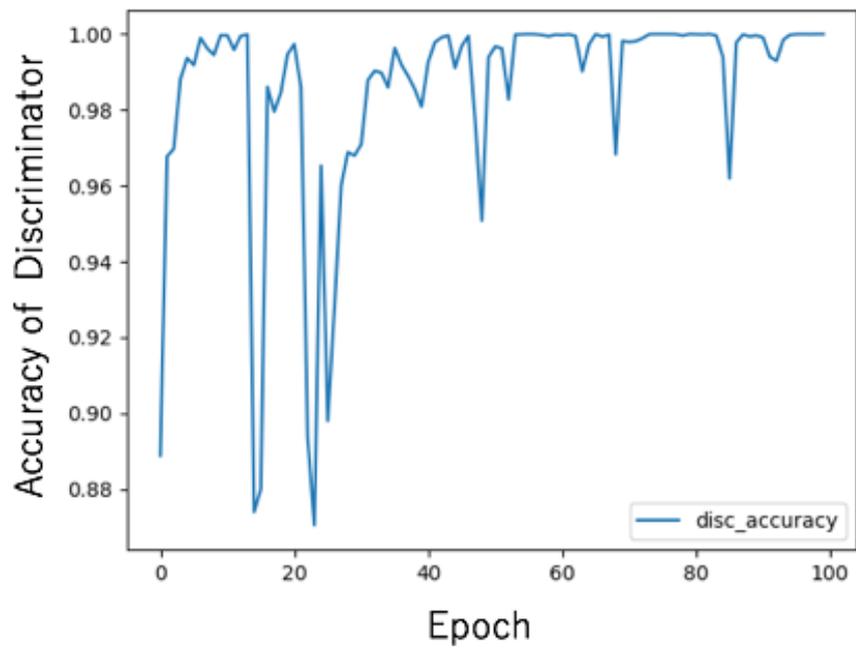


図 5.18: 白質における識別モデルの正解率 (注意モデル以外学習時)

第 6 章

考察

6.1 実験結果に対する考察

Attention-Guided CycleGAN での全脳、灰白質、白質画像の相互変換の結果より、男性から女性への変換の際に後頭葉が削られるように変換された。また女性から男性への変換の際には、全脳では頭頂葉から後頭葉にかけて削るように、白質では後頭葉に新たな領域を生成するように変換がされた。この結果から性差には脳の後頭部の形状が関わっている可能性が見出せた。しかし、変換結果が鮮明でなく性差を特定する上での形状やパターンの違いを見ることができなかった。そのため、今後はモデルのネットワーク構成としてモデルの構造を ResNet や FPN (Feature Pyramid Network) などのネットワークに変更したり、ハイパーパラメータの調整を行う必要があると考える。また、今回用いた手法についても信頼性の検証を行っていないため、別データを用いた検証を行い、データ間の差の分析に Attention-Guided CycleGAN が適しているかを調べる必要があると考える。さらに現在の識別機は単純な CNN で構成されており、識別の際にスケールを考慮していない可能性があると考えられる。そのため、灰白質の変換でも同じスケールで生成された領域なのかを考慮する必要があると考える。今後の課題として、生成モデルで取り入れた U-Net のような異なるスケールの特徴を考慮できるネットワークを識別モデルに取り入れた場合の変換の違いを調べる必要があると考える。

さらに、本実験では全モデルのオプティマイザに Adam を使用し、学習率などのハイパーパラメータも同じにして学習をした。しかし、個々のモデルによって課題が異なるため、各モデルに適した学習率が存在し、モデル間の相互関係も考慮した学習率の設定が必要

6.2 注意領域を用いた識別モデルの識別過程の考察

だと考える。よって、今後の課題として Attention-Guided CycleGAN 全体として最適な 옵ティマイザや学習率を探索する必要があると考える。また、Isola らの研究 [14] では、PatchGAN のパッチのサイズによっても輪郭画像から画像を生成する pix2pix の生成結果の鮮明さなどが異なっている。このことから、本研究に関しても異なるパッチサイズでの学習を行い、最適なパッチのサイズを求める必要があると考えられる。

次に注意領域の選択の際に、本研究では識別モデルの正解率の動きや注意領域の変化を 10 エポックごとに見ながら選択した。しかし、選択の基準が明確でないと分析する人によって結果に差が生じる可能性がある。そのため、データの分析手法として Attention-Guided CycleGAN を用いるためには選択の基準を設定する必要があると考えられる。また、結果の評価についても本研究では見た目の変化だけで判断する定性的評価しかできていない。そのため、今後の課題として今回の手法に対してどのように定量的評価を行うかを考える必要があると考えられる。

6.2 注意領域を用いた識別モデルの識別過程の考察

本実験では Attention-Guided CycleGAN を用いた脳の部位を表す 3 次元画像の相互変換を行うことで性差の分析を行った。その学習過程で得られる元画像と注意領域は灰白質における男性から女性への変換では図 6.1, 6.2 で表すように遷移した。注意領域を見ると、多くの場合で脳の部分は注意領域の値が 0 になっており、生成モデルによる変換の影響を受けない領域となっている。しかし、図 5.11 で表す識別モデルの正解率では 150 エポック目あたりで正解率が大きく低下しており、図 6.2 では 150 エポック目あたりで後頭部の注意領域の値が大きくなっている。注意領域の値が大きくなっていることから後頭部で変換された値の影響を強く受けるため、後頭部の変換が識別に大きな影響を与えたと考えられる。この結果から注意領域を見ることで識別モデルがどこに着目していたかを推定できるのではないかと考えられる。

また注意領域の傾向として背景部分の注意領域が徐々に黒くなっていることが挙げられ

6.2 注意領域を用いた識別モデルの識別過程の考察

る。これは背景が変換されることで識別モデルが元画像の背景と変換画像の背景を識別できるようになることから、エポックが進むにつれて背景を無駄に変換しないように注意領域を生成していると考えられる。しかし、最終的な注意領域を見ても背景が黒になっておらず、背景を変換しないような注意領域の生成ができていない。よって、より変換すべき領域を限定できるように、良い注意領域が得られた後に注意モデル以外を学習する際に行う閾値処理の閾値をより大きな値にする必要があると考えられる。しかし、閾値を大きくし過ぎると変換すべき領域を変換できなくなる可能性がある。そのため、閾値についても最適な値を探索する必要があると考えられる。

6.2 注意領域を用いた識別モデルの識別過程の考察

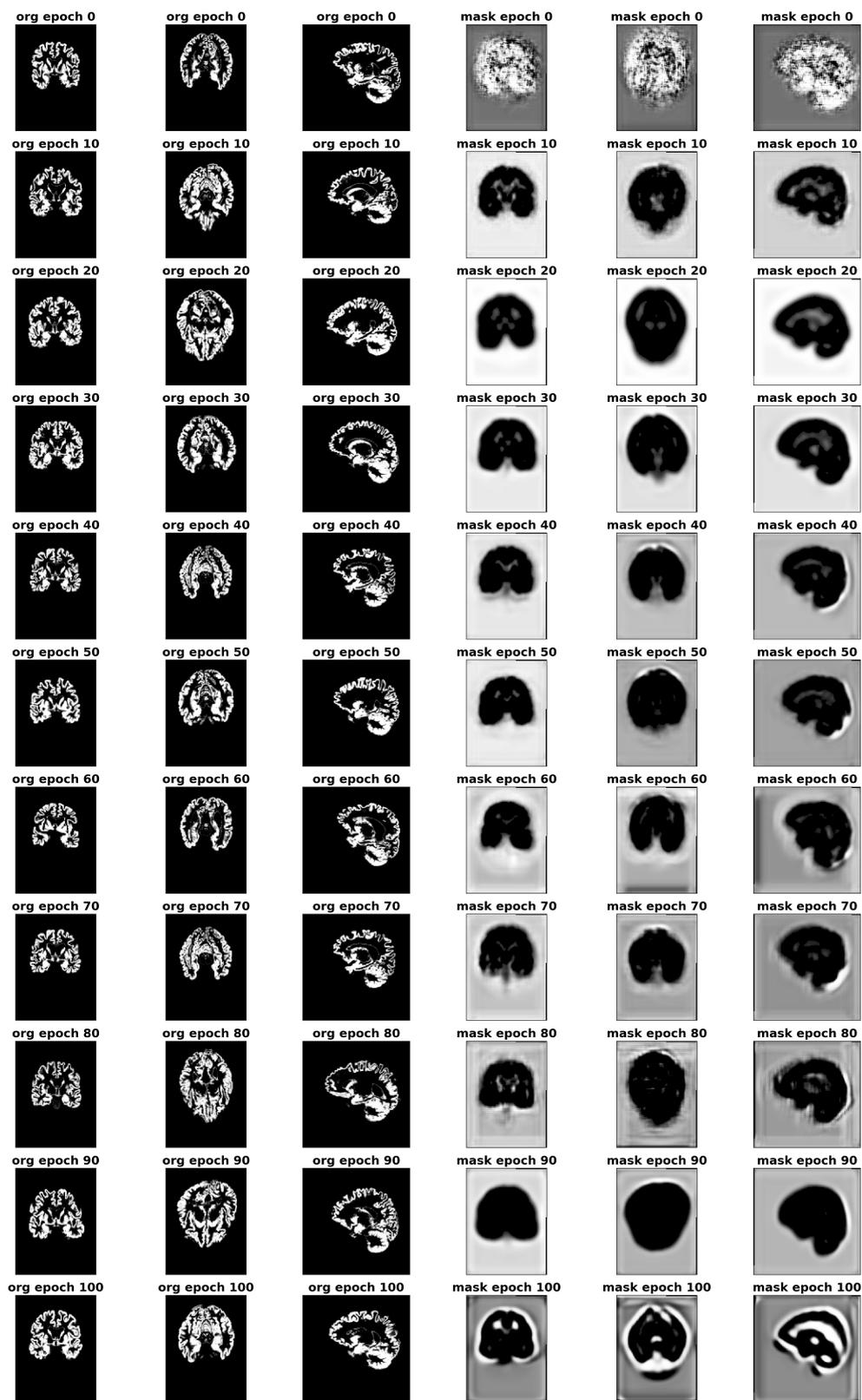


図 6.1: 灰白質画像を用いた学習時の元画像 (org) と注意領域 (mask) の遷移

6.2 注意領域を用いた識別モデルの識別過程の考察

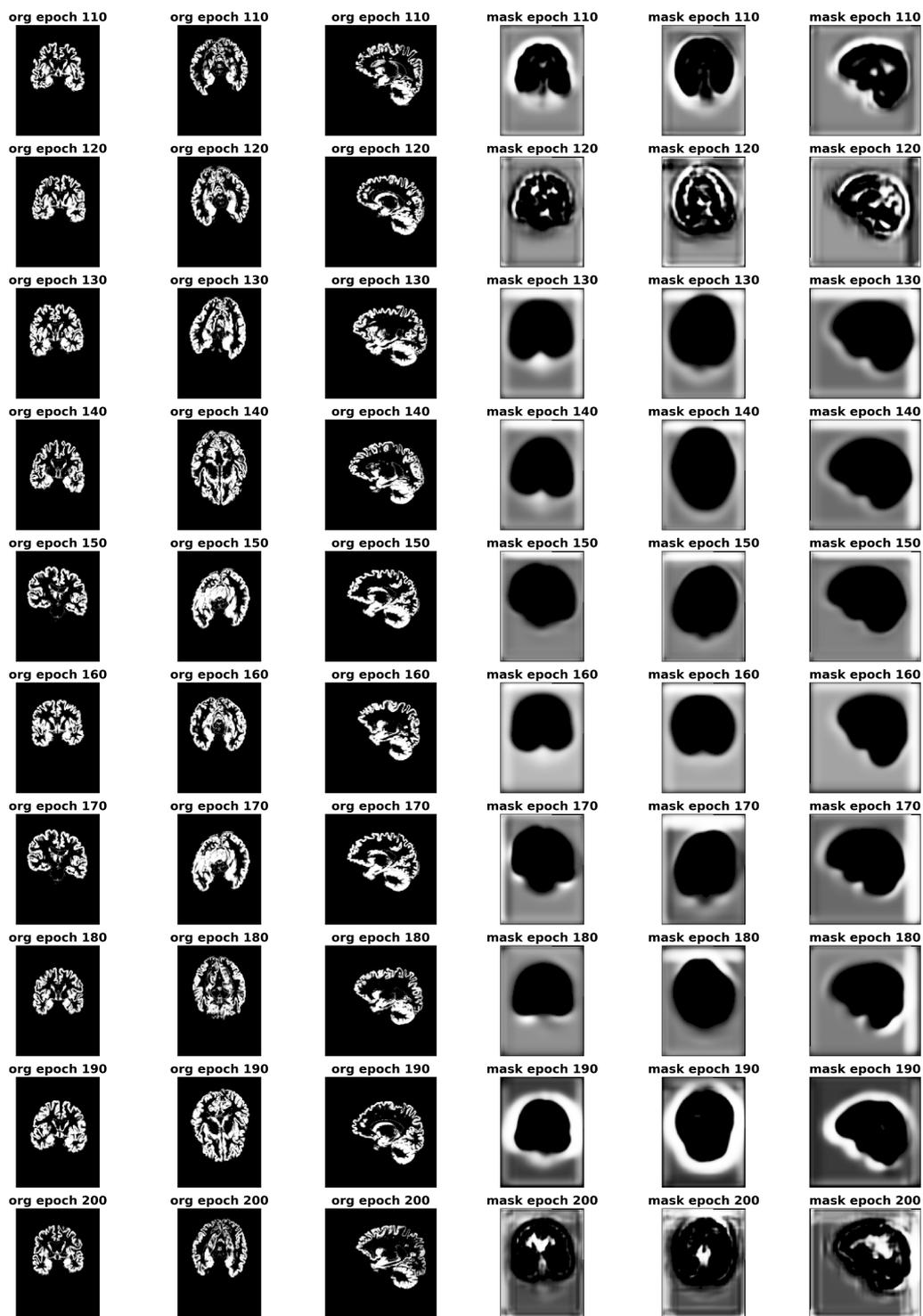


図 6.2: 灰白質画像を用いた学習時の元画像 (org) と注意領域 (mask) の遷移

第7章

結論

敵対的生成ネットワークによる学習ベースの相互変換を通じて脳画像の性差の分析を試みた。その結果、CNNによる男女の脳構造画像の識別には脳の後頭部の形状が関係する可能性が見い出せた。また、学習時の注意領域や識別モデルの正解率から、識別モデルが男女の識別の際に着目した部分を推定できると考えた。しかし変換結果や手法の信頼性に難があるため、今後は各モデルの改良による精度の向上、より多くのデータを用いた検証、別データによる結果の信頼性の検証に取り組む必要がある。

謝辞

2年間の研究室活動にあたり、就職活動や卒業研究にあたって多忙ながらも真摯に指導してくださった高知工科大学情報学群の吉田真一准教授に感謝いたします。また、機械学習のプログラミングに疎かった自分の質問に、最初から最後まで丁寧に答えてくださった高知工科大学情報学群の四宮友貴助教にも感謝いたします。正直ノリで始めた研究ですが、研究を進める上で2人の先生方が支えてくださったおかげで最後まで研究を続けることができました。吉田先生に関しては3年生の1月まで就職を考えてた自分の就活相談に乗ってくださり、ありがとうございました。結局は就職せずに進学をする道を選びましたが、来年度は2回目の就活があるので相談に乗ってくださった成果を発揮したいと思います。また本研究で扱った脳ドック受診者のデータを提供してくださった高知検診クリニックの朴啓彰先生にも感謝いたします。朴先生のデータのおかげで本研究を進めることができました。

また、お忙しい中にも関わらず本研究の副査を引き受けてくださった岩田誠教授、竹内聖悟講師にも感謝いたします。岩田誠教授には自分の研究を話す上でどのようなところを前面に出して話した方が良いといったアドバイスや今後の研究方針について提案していただき、大変参考になりました。竹内聖悟講師には研究内容を再認識させるような質問を投げかけていただき、再度自分の研究について考える機会を得ることができました。今後またお世話になる機会があれば、その時はよろしくお願いします。

研究室の方々に関しては、まずは卒業された方も含めITNewsや輪講、行事などの様々な場面で助けてくださった先輩方に感謝いたします。先輩方の印象としては凄くスイッチの切り替えが速くて、さっきまで研究に打ち込んでいると思ったら今度は楽しそうに会話を始めたりとオンとオフの差が大きいと感じていました。自分はその辺の切り替えが不得意で、集中した後は神経質になりがちなのでこれからも見習っていきたいと思います。また、研究室に来なくても家で研究したりと見えないところで努力していて、自然と自分も頑張らないという気分させてくれる方々でした。

謝辞

同級生はかなり個性的なメンバーが集まったと感じていて、自由奔放なよしけんらしい雰囲気の中で研究室での生活を送れたと思います。東は研究室とそれ以外のことに関しての両立が上手だなあと感心しています。その感じで来年以降も頑張ってください。高橋くんに関してはページ数の都合により割愛させていただきます。高橋さんは周りとは方向性の違う研究テーマに取り組む中でコツコツと進捗を出していて、よしけんには数少ないしっかりした人だと思っています。春から社会人ですが、持ち前のコツコツさで GANGAN 出世しちゃってください。元木は前処理やネットワーク系の問題の際に周りのために動いてくれて有り難かったです。高橋さんと同じく春から社会人ですが、新生活を楽しんでください。新田はこれまた数少ない常識人でみんなを叱ってくれる存在でした。ちょいちょい愚痴も聞いてくれてとても助かりました。今後もよろしくお願いします。野田には何度か知識的な面で頼ることがあって、その時は普段の適当さとは真逆に丁寧に教えてくれてありがとうございました。今後も頼ることがあると思うのでその時はよろしくお願いします。

後輩については普段はしっかりしてる人が多くて、オープンキャンパスやパソコンのセッティングなどの研究室での仕事もしっかりこなしてくれてすごく頼りになると思っています。ただ飲み会の場では終始ドン引くくらい飲むので、健康に気をつけて酒に飲まれないように楽しんでください。来年度は卒業研究、就活と大変な一年になると思いますが、頑張れば結果がついてくると思います。楽しいよしけんライフを過ごしてください。

またこの研究を進めるためには、親のサポートや恵まれた設備など欠かせないものがありました。そのような恵まれた環境にも感謝しつつ、修士課程での2年間も励んでいきたいと思っています。

参考文献

- [1] Shinichi Yoshida, Yuki Shinomiya, Takenobu Sasatani, and Kaechang Park. Prediction of human characteristics from brain structural MRI using CNN. In *International Symposium on Frontier Technology (ISFT)*, Aug 2019.
- [2] 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.
- [3] N.JonShah Gereon R.Fink Martina Piefke Martin Schulte-Rüther, Hans J.Markowitsch. Gender differences in brain networks supporting empathy. In *NeuroImage*, Vol. 42, pp. 393–403, 2008.
- [4] View ORCID ProfileAleksi Tiulpin Florian Dubost Marleen de Bruijne Meike W. Vernooij Hieab H. H. Adams M. Arfan Ikram Wiro J. Niessen Johnny Wang, View ORCID ProfileMaria J. Knol and View ORCID ProfileGennady V. Roshchupkin. Gray matter age prediction as a biomarker for risk of dementia. In *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, Oct 2019.
- [5] 厚生労働省. 平成 28 年版厚生労働白書 – 人口高齢化を乗り越える社会モデルを考える –, 2016. <https://www.mhlw.go.jp/wp/hakusyo/kousei/16/index.html>.
- [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. In *The Institute of Electrical and Electronics Engineers (IEEE) International Conference on Computer Vision (ICCV)*, Oct 2017.

参考文献

- [7] 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.
- [8] 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.
- [9] David H. Hubel and Torsten N. Wiesel. Receptive fields of single neurons in the cat’s striate cortex. *Journal of Physiology*, Vol. 148, pp. 574–591, 1959.
- [10] Kuniyuki Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, Vol. 36, pp. 193–202, 1980.
- [11] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the Institute of Electrical and Electronics Engineers(IEEE)*, pp. 2278–2324, 1998.
- [12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1097–1105. Curran Associates, Inc., 2012.
- [13] 人工知能学会. 深層学習-Deep Learning-. 近代科学社, 2015.
- [14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *The Institute of Electrical and Electronics Engineers(IEEE) Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [15] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-*

参考文献

Assisted Intervention (MICCAI), Vol. 9351 of *LNCS*, pp. 234–241. Springer, 2015.
(available on arXiv:1505.04597 [cs.CV]).