

公立はこだて未来大学 2022 年度 システム情報科学実習
グループ報告書

Future University Hakodate 2022 Systems Information Science Practice
Group Report

プロジェクト名

脳をつくるプロジェクト

Project Name

Make Brain Project

グループ名

深層学習モデルの描画研究への応用

Group Name

Application of deep learning in drawing study

プロジェクト番号/Project No.

22-B

プロジェクトリーダー/Project Leader

中村仁 Jin Nakamura

グループリーダー/Group Leader

立山雄晟 Yusei Tateyama

グループメンバ/Group Member

立山雄晟 Yusei Tateyama

前澤榛人 Haruto Maezawa

和泉友人 Tomohito Izumi

池田蓮馬 Renma Ikeda

指導教員

香取勇一 佐々木博昭 加藤譲 ヴラジミール・リアボフ

Advisor

Yuichi Katori Hiroaki Sasaki Yuzuru Kato Volodymyr Riabov

提出日

2023 年 1 月 18 日

Date of Submission

January 18, 2023

概要

描画行為は人間特有の行動であり、その研究は人間特有の創造力の考察に繋がる可能性がある。しかし、既存の描画研究は心理実験とその考察が主であり、数値理論的な手法に欠けている。そのため、描画能力に関して分かっていることはまだ少ない。また近年、深層学習の発達により、深層学習モデルを用いた神経科学の観点から描画研究を行える可能性がある。深層学習モデルから脳へのアプローチの1つとして、ある特定の目的で訓練された深層学習モデルが、脳とよく似た性質を示すといった目的駆動というアプローチが存在する。そこで、本グループでは目的駆動の考え方を取り入れ、十分に訓練された深層学習モデルを用い、描画能力を構成する能力の1つである認識能力に着目して描画研究の心理実験を行った。これによって、深層学習モデルによる心理実験の代用の可能性を示すことを目指す。

前期の活動では、グループメンバーのやりたいことがバラバラであったため、神経科学や脳についての基礎知識を学習したうえで、グループ内で話し合いを行い、グループテーマの決定を行った。前期段階でのグループテーマは、「人間の知覚特性を生かした錯視画像の生成」である。夏季休暇では前期で定めたグループテーマに基づいて文献調査や深層学習モデルの学習・実装を行い、グループテーマを達成するような具体的なアプローチについて考えた。しかし、錯視画像を知覚するような深層学習モデルの実装は可能なものの、人間の知覚特性を応用した画像生成のアプローチは考案出来なかった。そのため、グループテーマの変更を検討し、追加で文献調査を行った。文献調査とグループ内での話し合いの結果、グループテーマを「深層学習モデルの描画研究への応用」に決定した。この目標を実現するために、後期からは描画能力に関する文献調査、深層学習モデルの勉強などを行い、先行研究を参考にし実験を行った。

具体的には、既存の心理実験に対し人間の被験者の代わりに、顔認識を行う深層学習モデルを用いて実験を行った。一つ目の実験では、変形させた顔画像に対する人間の認識精度と平均反応時間を調査した心理実験を、深層学習モデルで行った。二つ目の実験では、顔の特徴的な部分を誇張した顔画像に対する認識精度が、正確な顔画像に対するものより高いことを示した心理実験を深層学習モデルを用いて行い、人間が被験者である場合と同様の結果を示すかを調査した。

一つ目の実験に関しては、すべての条件で傾向が一致することはなかったが、一部の条件においては認識精度の偏りが定性的に類似した。これにより、深層学習モデルは、先行研究で示された人間の知覚特性の一部を再現したと言え、描画実験を深層学習モデルによって一部代用できる可能性が示された。また、二つ目の実験では誇張された顔画像が正確な顔画像よりも精度が低くなり、先行研究とは異なった結果が得られた。今回の実験では、物体認識を行うように事前に訓練されたネットワークをファインチューニングすることによって顔認識のモデルを作成したため、精度が低下したと考えられる。

今回行った二つの実験は、描画能力を構成するうちの一つである認識能力に焦点を当てたものであり、より多くの描画実験に深層学習モデルを代用するには、他の描画能力を構成する要素についても考慮された深層学習モデルの使用が望まれる。また、先行研究と異なる部分に関しては異なる原因に対する考察に基づき、追加で実験を行う必要がある。

キーワード 深層学習, 計算論的神経科学, 描画研究, 心理実験

(※文責: 立山雄晟)

Abstract

The act of drawing is a behavior unique to humans, and its study may lead to a consideration of the creative power unique to humans. However, existing research on drawing is mainly based on psychological experiments with subsequent data analysis and lacks numerical-theoretical methods. Therefore, what we know about drawing ability is still limited. With recent development of deep learning, there is a possibility that drawing research can be conducted from the viewpoint of neuroscience using deep learning models. One of the approaches to the brain from deep learning models is the goal-driven approach, in which deep learning models trained for a specific purpose show properties similar to those of the brain. Therefore, this group decided to adopt the purpose-driven approach and conduct an AI experiment of drawing research using a well-trained deep learning model, focusing on the recognition ability, which is one of the abilities that constitute drawing ability. By doing so, we aim to demonstrate the possibility of substituting psychological experiments with deep learning models.

In the first semester activities, students learned basic knowledge about neuroscience and the brain, and decided on group themes. The group theme in the first semester was the generation of illusion images utilizing human perceptual characteristics. During the summer vacation, the students conducted a literature survey and learned and implemented a deep learning model based on the group theme determined in the first semester. However, although it was possible to implement a deep learning model to perceive illusion images, we could not devise an approach to image generation that applied human perceptual characteristics. Therefore, we considered changing the group theme and conducted an additional literature survey. As a result of the literature review and group discussions, the group theme was decided to be the application of deep learning models to drawing research. In order to realize this goal, from the second semester, the group conducted a literature survey on drawing capabilities, studied deep learning models, and conducted experiments by referring to previous studies.

Specifically, we conducted experiments using a deep learning model that performs face recognition instead of human subjects for existing psychological experiments. In the first AI experiment investigating human recognition accuracy and average reaction time to a deformed face image was conducted with a deep learning model. In the second AI experiment we used a deep learning model to show that recognition accuracy for face images with exaggerated facial features was higher than that for accurate face images and to investigate whether the results were similar to those obtained when human subjects were used.

For the first experiment, the trends did not match in all conditions, but the recognition accuracy bias was qualitatively similar in some conditions. This indicated that the deep learning model reproduced some of the human perceptual characteristics shown in previous studies, and that the drawing experiment could be partially substituted by the deep learning model. In the second experiment, exaggerated face images were less accurate than accurate face images, a result that was different from the previous study. In this experiment, the face recognition model was created by fine tuning a network that had been previously trained to perform object recognition, which may have reduced the accuracy.

The two experiments conducted in this study focused on recognition, one of the components of drawing ability, and in order to substitute deep learning models for more drawing experiments, it would be desirable to use deep learning models that also take into account other components of drawing ability.

Keyword Deep Learning, Computational Neuroscience, Drawing Research, Psychological Experiments

(※文責: 立山雄晟)

目次

第 1 章	はじめに	1
1.1	背景	1
1.2	目的	2
第 2 章	関連研究	3
2.1	深層学習モデルと脳機能に関連する先行研究	3
2.2	物体認識に関連する先行研究	4
2.3	特徴抽出に関連する先行研究	5
第 3 章	方法	7
3.1	提案手法	7
3.1.1	目的駆動	7
3.1.2	ResNet(Residual Network)	7
3.1.3	Saliency Map	8
3.1.4	アフィン変換	9
3.2	実験内容	9
3.2.1	実験 1 変形顔画像実験	9
3.2.2	実験 2 カリカチュア画像の認識精度実験	9
第 4 章	結果・考察	11
4.1	実験 1	11
4.1.1	実験 1 の結果	11
4.1.2	実験 1 に関する考察	11
4.2	実験 2	12
4.2.1	実験 2 の結果	12
4.2.2	実験 2 に関する考察	12
第 5 章	まとめ	14
5.1	プロジェクトの成果	14
5.2	今後の課題	14
第 6 章	外部評価	15
6.1	中間発表	15
6.2	成果発表	16
第 7 章	活動内容	18
7.1	前期活動内容	18
7.1.1	コミュニケーションツールの決定	18
7.1.2	必要な知識の学習	18

7.1.3	グループテーマの決定	18
7.1.4	関連研究の調査	19
7.1.5	前期における活動目標の設定	19
7.1.6	中間発表に向けた準備	19
7.1.7	中間報告書の作成	19
7.1.8	前期の振り返り	20
7.2	夏季休暇活動内容	20
7.2.1	夏季休暇における活動	20
7.2.2	夏季休暇の振り返り	20
7.3	後期活動内容	20
7.3.1	グループテーマの再決定	20
7.3.2	グループにおける役割分担	21
7.3.3	成果発表に向けた準備	21
7.3.4	最終報告書の作成	21
7.3.5	後期の振り返り	22
7.4	プロジェクトにおける各人の主な役割	22
7.4.1	立山	22
7.4.2	前澤	23
7.4.3	和泉	24
7.4.4	池田	25
第 8 章	技術面	26
8.1	開発環境	26
8.1.1	プログラミング言語	26
8.1.2	実行環境	26
8.2	実験方法の詳細	27
8.2.1	顔認識モデル	27
参考文献		29

第 1 章 はじめに

1.1 背景

描画行為は人間に固有の行動である。人間が先史時代から絵を描いていたことが洞窟壁画の調査から確認されている。現在発見されている中で最も古いと考えられている洞窟壁画はインドネシアのスラウェシ島で発見された壁画で、約 45,000 年前に描かれたと推定されている [1]。また、ヨーロッパでも多くの洞窟壁画が発見されている。例えば、フランスに存在する Chauvet 洞窟、Lascaux 洞窟、Niaux 洞窟で洞窟壁画が発見されており、10,000 年から 30,000 年前のものだと推定されている。人間はここから絵画技術を発展させてきた。一方で、他の動物が絵を描くことはない。チンパンジーや象が絵を描く芸を行うことがあるが、いずれも人間による訓練の結果である。

描画行為には様々な能力が関わる。まず、描画を行うためには、物体や風景を視覚を用いて認識しなければならない。そしてその次に、認識した物体を記憶しなければならない。物体を直接見ながら描く場合には短期記憶能力が必要であり、一度見たものについて描く場合には長期記憶能力が必要である。さらに、実際に絵を描くためには手先や筆を思ったように動かすための運動機能が必要である。また、絵の描き方は様々だが、筆を使うにせよ塗料を扱うにせよ、道具使用の能力が必要である。このように、描画行為には、視覚や記憶、運動、道具使用など、様々な能力が関わる。加えて、絵を描くという課題を定め、完成させるための実行機能も不可欠である。

描画行為についての研究は、これまで心理学の分野で行われてきた。その結果、描画行為に関わる要因が数多く明らかになっている。例えば、Kozbelt (2001) は知覚に関するタスクの成績と描画に関するタスクの成績を比較することで、絵の上手な被験者は視覚探索に優れていることを示している [2]。また、Kozbelt (2010) によれば、画家は非画家に比べて物体から有意義な特徴を抽出することに優れている [3]。この結果は、描画能力には注意機構が深く関わることを示している。さらに、Cohen (2005) は画家と非画家間では描画中の眼球運動に差異が見られることを実験によって示している [4]。具体的には、画家は非画家に比べ、描画対象と絵の間で注視を行き来する頻度が高い。

一方で、描画能力全体を俯瞰したような研究はあまり進んでいない。描画行為についてのモデルとして現在最も受け入れられているものは、1989 年に van Sommers によって提案されたモデルである [5]。van Sommers のモデルは二つの階層的なシステムからなる。一つは Marr の視覚計算理論 [6] に基づく視覚に関するシステムであり、もう一つは運動に関するシステムである。また、他にもいくつかのモデルが 1980 年代から 1990 年代に提案されている [7]–[9]。McCrea はこれらのモデルをレビューしたうえで、それぞれを組み合わせた「現代的な」モデルを提案している [10] (図 1.1)。しかし、基本的な構成要素は当時のままであり、近年の知見を反映出来ていない。例えば、視覚の部分では Marr の視覚計算理論が使われているが、視覚に関してはこの 20 年間でより優れたモデルが提案されている。

深層学習技術によって、より精緻な描画行為のモデルを提案出来る可能性がある。我々の知識の限りでは、描画研究の分野でモデルを用いた研究は現在まで行われていない。これは、描画行為が物体認識や運動に関わる複雑な行動であり、描画研究が盛んに行われていた時代ではシミュレーションを行うことが困難であったためと考えられる。その一方で、近年深層学習技術が急速に発展

している。深層学習モデルは数々のタスクで人間と同等、あるいは人間以上のパフォーマンスを実現しており、物体認識や描画もこの例に漏れない [11], [12]。さらに、最近では深層学習技術を用いた神経科学の研究が盛んに行われている [13]–[15]。これについては 2 章で詳しく述べるが、深層学習モデルと視覚野の類似性が数々の研究で示されている [14], [16]。これらのことから、深層学習を描画行為のモデルに適用することで、シミュレーションが可能なモデルを実現できる可能性がある。シミュレーションが可能になれば、モデルのパラメータを制御することで一つ一つの構成要素について詳細に調査することができる等、数多くの利点がある。また、モデルと脳部位の対応関係を調査できる可能性もある。

(※文責: 和泉友人)

1.2 目的

本プロジェクトの目的は、深層学習モデルの描画研究への応用可能性について調査することである。我々は特に、描画行為を構成する能力の中でも視覚を用いた認識の部分に注目した。描画行為は、まず視覚によって物体を認識することから始まる。したがって、深層学習モデルの特性が人間の視覚特性とどの程度類似しているのかは極めて重要な要素である。我々は、このことを調べるために二つの実験を行った。

(※文責: 和泉友人)

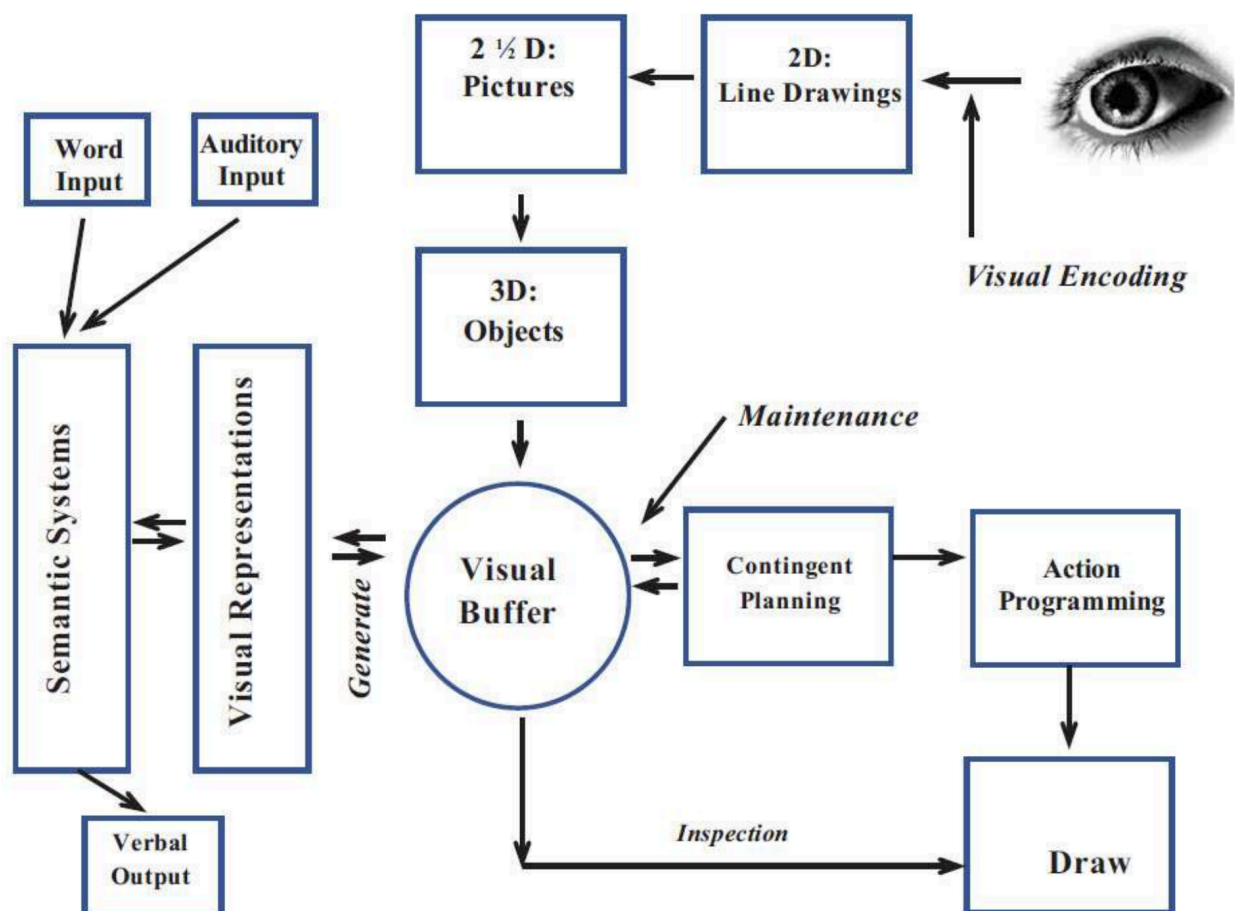


図 1.1: McCrea の提案する描画行為のモデル [10]

第 2 章 関連研究

我々は、深層学習モデルを描画研究に応用できる可能性を調査する。人間の描画行為におけるプロセスには記憶や手続き的知識、物体認識、特徴抽出などの相互作用によるものがある [17]。これらの描画プロセスのうち、物体認識、特徴抽出を中心に、それらに関連する先行研究を調査した。

また、従来人間に対する実験によって行われてきた描画研究を、深層学習モデルによって実現する際に、人間の活動と深層学習モデルの活動の類似性が存在することを示す必要がある。両者の類似性を示すことは、深層学習モデルを描画行為に適用する際に、人間の脳機能と対応づけた研究を行うことが出来る可能性を示すことになる。そのため我々は人間の活動と深層学習モデルの活動を比較した先行研究も同様に調査した。

(※文責: 前澤榛人)

2.1 深層学習モデルと脳機能に関連する先行研究

Yamins ら (2016) によって行われた研究では、CNN と脳の視覚野との関連性が調査された [18]。CNN については 3.1.2 で後述する。既存の CNN は視覚系の神経回路をモチーフに構築されたものであるが、ネットワークが深くなればなるほど、パラメータを多く持つこととなり、視覚野の活動を再現することが困難になっていった。Yamins らは、どのようなパラメータを用いても感覚系の行う活動を解明するのに有効でなければ、それは正しいモデルであるとは言えないと考え、目的駆動型のアプローチによるニューラルネットワークの構築を提案した。目的駆動型の CNN を用いることで、より実際の視覚野の活動を深いところまで再現できるのではないかと考えた。目的駆動の考えとは、特定の目的に基づいてニューラルネットワークを訓練し、最適化することである。目的に応じてネットワークのパラメータを最適化し、最適化されたパラメータと実際の神経活動を比較することでより正確に脳機能との対応付けが出来るのではないかと考えられた。

Yamins らは、物体分類を行うように訓練し、最適化された CNN モデルに、人間の下側頭葉皮質の活動を予測させたり、他のモデルとの比較をする実験などを行った。結果として、物体分類の目的駆動に基づいて訓練された CNN モデルは他のモデルよりもより正確に下側頭葉皮質の反応を予測した。図 2.1 は、様々なモデルにサルの視覚野の V4 領域と IT(下側頭葉皮質) の神経反応を予測させ、その予測の正確性を示したものである。CNN モデルでは、最後から 2 番目の隠れ層が V4 の活動を最も良く予測し、最後の隠れ層が IT の活動を最も良く予測する結果となった。また、下側頭葉皮質や V4 領域だけでなく、V1, V3 領域に対して、CNN の下位層が同様に高精度な予測をすることが確認された。これらの結果により、目的駆動型で訓練された CNN が物体分類において他の CNN よりも優れていることや、人間の視覚野の神経活動により近い性質を持つことが示された。

(※文責: 前澤榛人)

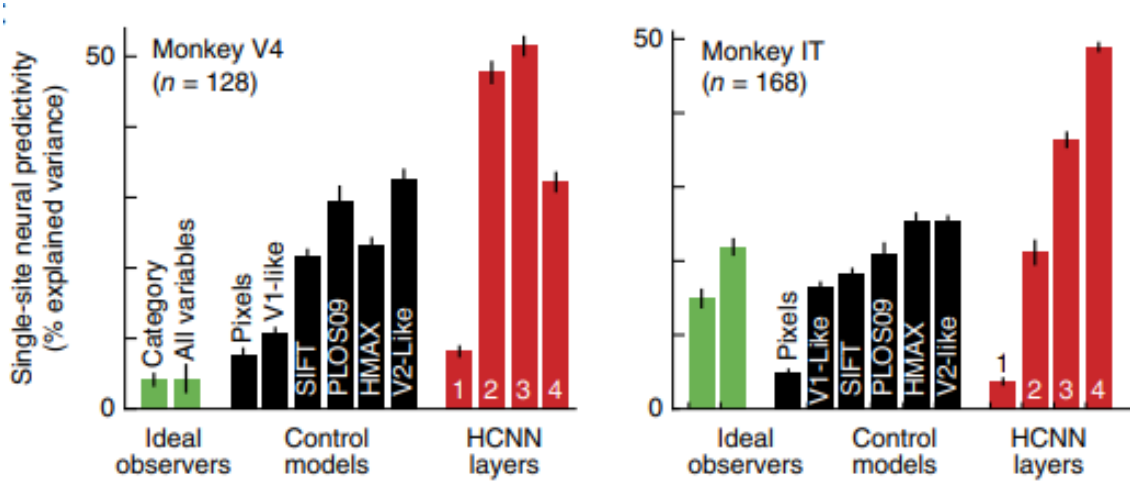


図 2.1: 様々なモデルによる IT と V4 の神経反応の予測率 [18]

2.2 物体認識に関連する先行研究

物体認識について、Hole ら (2002) によって行われた、変形された顔画像を認識する心理実験を行った先行研究がある [19]。この研究では、まず実験用データとして有名人と非有名人の顔画像が用意された。それらの画像に対して倒立やぼかしといった処理や、垂直、水平方向への伸縮や上半分、下半分のみ伸縮など様々な種類の変形を施した画像が作成された (図 2.2)。それらの画像を被験者に提示し、その画像が有名人か非有名人かを判断する実験をおこなった。実験データは非変形状態である正立条件と、画像の変形条件ごとに、平均正答率、平均正答数、回答するまでの平均反応時間、偶然性を示すノンパラメトリック推定の値とそれぞれの標準偏差が記録された。

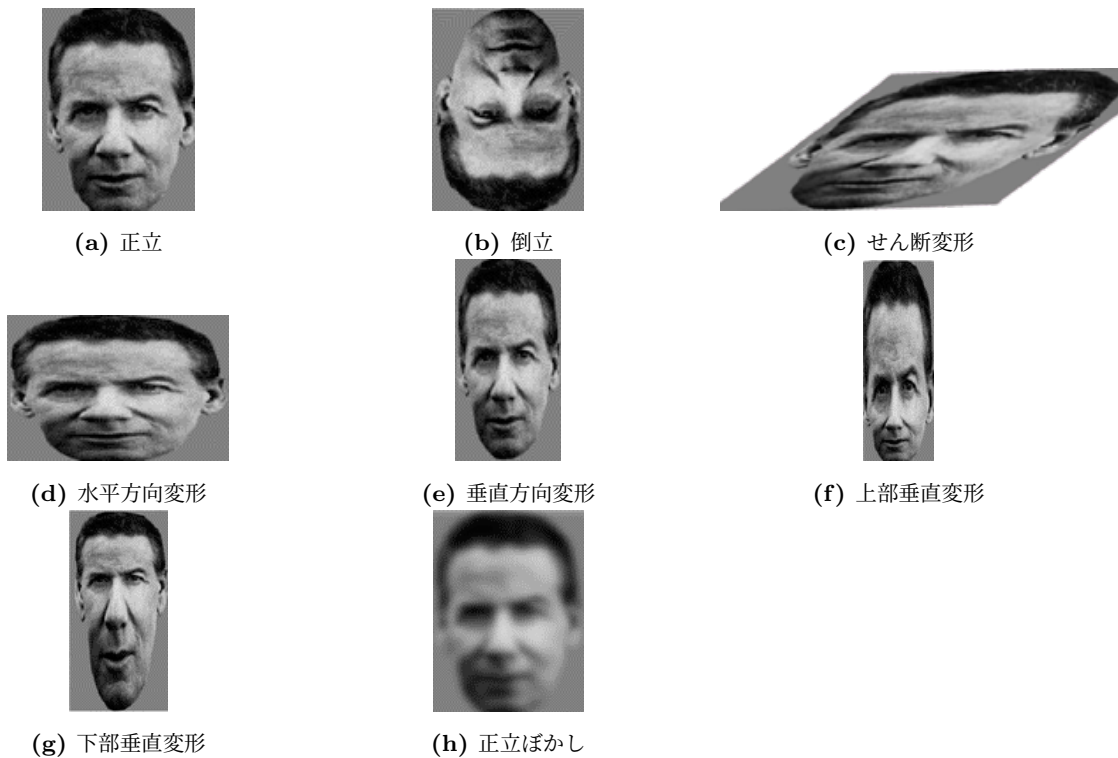


図 2.2: 実験で使用された画像の一例 [19]

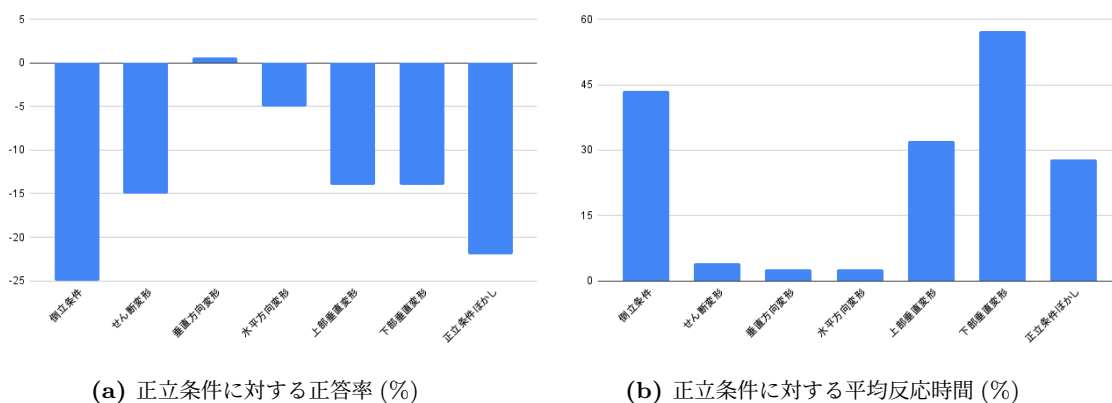


図 2.3: 顔認識実験での正答率 (左) と平均反応時間 (右) の図 [19]

結果を図 2.3 に示す。左は正立条件の正答率を基準として、各変形条件における正答率の差を割合で表したものである。右は正立条件の平均反応時間を基準として、各変形条件における平均反応時間の差を割合で表したものである。正立条件と比較した認識精度の差は、倒立条件で最も低下し、次点で正立条件ぼかし、垂直方向変形ぼかしが低くなった。一方で、垂直方向変形や水平方向変形では認識精度の大幅な低下は見られなかった。また、正立条件と比較した平均反応時間に関しては、下部垂直変形が最も平均反応時間が長く、次点で倒立条件、全体垂直変形ぼかしが長かった。一方でせん断変形、垂直方向変形、水平方向変形条件における平均反応時間は、正立条件と大きな差はなかった。

(※文責: 前澤榛人)

2.3 特徴抽出に関連する先行研究

学習させた刺激	テストされた刺激	
	顔	カリカチュア
正確なカリカチュア		
顔	.601	.641
カリカチュア	.593	.662
正確でないカリカチュア		
顔	.600	.611
カリカチュア	.664	.676

表 2.1: 各刺激に対する MOC 曲線下面積の割合 [20]

特徴抽出について、Mauro ら (1992) によって行われた、カリカチュアを認識する心理実験を行った先行研究がある [20]。カリカチュアとは、人の特徴や性格などを強調するために、意図的に目や鼻などの特定の部位を誇張して描画された絵である。この研究では、まず正常な顔のデータとカリカチュアで描かれた顔のデータをそれぞれ学習させ、顔画像が生成された。その後成人白人男性の平均的な顔のデータに基づいて 2 種類のカリカチュア変換を行った。一つは平均のデータと比べて特徴的な部分を誇張する変換 (正確なカリカチュア)、もう一つは平均のデータと比べて特徴的ではない部分を誇張する変換 (正確でないカリカチュア) である。被験者にはまず 100 枚の画像

Make Brain Project

が提示された。その後、先ほど提示された 100 枚を含む計 300 枚の画像が提示された。被験者は提示される 300 枚の画像のそれぞれに対して、最初の 100 枚で提示された画像と同じ顔が存在していたかを判断した。実験データは、顔画像学習によって生成された顔画像と、正確なカリカチュア、正確でないカリカチュアの 2 種類のカリカチュア画像、カリカチュア画像学習によって生成された顔画像と 2 種類のカリカチュア画像の 8 パターンに対して、被験者の正誤判定によって算出された MOC 曲線の下面積の割合が記録された。

結果は表 2.1 に示す。表は各条件における被験者の実験結果を基に算出された MOC 曲線の下面積の割合を示している。MOC 曲線は、Memory Operating Characteristic の略称であり、正確に記憶されなかった項目数に対して、正確に記憶された項目数を曲線化して表したものである。グラフにおいて、全体の面積に対する MOC 曲線の下部分の面積の割合が指標の一つとなり、この割合が大きいほど正確性が高いことが示される。この先行研究においては、被験者の認識の正確さを表すこととなり、割合が大きいほど、提示された画像の種類 (顔画像やカリカチュア画像) での認識精度をよし悪しが表される。

結果として、正確なカリカチュアにおいても正確でないカリカチュアにおいても通常の顔画像よりもカリカチュア画像の方が認識精度が高い傾向にあるという結果が見られた。また、総合して正常な顔のデータを学習して生成された顔画像やカリカチュア画像よりも、カリカチュアで描かれた顔のデータを学習して生成された顔画像やカリカチュア画像の方が認識精度が高い傾向にあることが示された。

(※文責: 前澤榛人)

第 3 章 方法

この章では実験をする際に用いた具体的なアプローチについて紹介する。

3.1 提案手法

3.1.1 目的駆動

目的駆動とは、2.1 でも述べたように Yamins らによって提唱された、ある特定の目的で訓練した深層学習モデルが、脳とよく似た性質を示すといったアプローチである。2.1 の先行研究では、目的駆動によって、人間と深層学習モデルにおける神経活動の類似性が示されていた。後述する二つの実験は、この考え方にに基づき、顔識別モデルとして高い精度を持つ深層学習モデルを使用し、知覚特性の再現を目指した。

(※文責: 立山雄晟)

3.1.2 ResNet(Residual Network)

ResNet とは Residual Network の略称であり、2015 年に Kaiming He によって考案された CNN の一種である [21]。CNN において、より高度で複雑な特徴を抽出するためには、層を深くすることが重要だと考えられている。一方で、単純に層を深くすると、学習がうまくいかなくなる勾配消失問題があったりと、層を深くするにも工夫が必要である。ResNet では残差接続を導入することにより勾配消失問題を解消して多くの層を重ねることに成功した。本プロジェクトの二つの実験では、この ResNet を顔認識をおこなうモデルとして学習させたものを使用する。

- CNN(Convolutional Neural Network)

CNN とは畳み込みニューラルネットワークとも呼ばれ、畳み込み層、pooling 層、全結合層によって構成されている。畳み込み層では、特徴を抽出し特徴マップを作成する。pooling 層では、特徴を残しながら次元削除を行う。全結合層では、特徴が抽出された画像データを一つのノードにまとめる。

- 残差接続 (Residual Connection)

ResNet において、層を深くするために用いられた方法であり、入力 x に対して、畳み込み層に通した出力 $F(x)$ と、入力そのままの値 x を加算したものを次の層への出力とするものである。これにより、学習の逆伝搬時に層を深くしても、大きな誤差が伝搬できるようになり勾配消失問題を解決することが出来る。

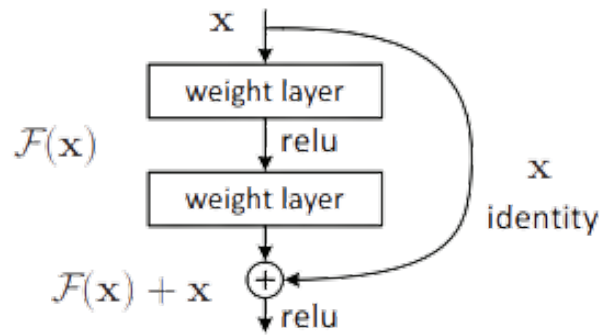


図 3.1: 残差接続のアーキテクチャ [21]

(※文責: 立山雄晟)

3.1.3 Saliency Map

Saliency map とは、顕著性マップとも呼ばれ、特定の画像の特徴的な部分や注目される部分が強調されるようにカラーリングされたものである。Saliency map の生成手法の一つとして Grad-CAM と呼ばれるものがある。Grad-CAM は CNN の最後の畳み込み層に含まれる特徴量を抽出して可視化をするものである。

Grad-CAM の計算方法は以下の手順で行われる [22]

- (1) ニューラルネットワークによる予測結果に対する特徴マップの勾配を計算し、その特徴マップの勾配に対して、Global Average Pooling を計算し、特徴マップの重要度を算出する

$$\alpha_k^c = \overbrace{\frac{1}{Z} \sum_i \sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}} \quad (1)$$

- (2) 特徴マップの重要度を用いて特徴マップの要素を重みづけし,ReLU 関数で計算する

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\underbrace{\sum_k \alpha_k^c A^k}_{\text{linear combination}} \right) \quad (2)$$

(※文責: 立山雄晟)

3.1.4 アフィン変換

実験 1,2 の画像変換の一部に対して、アフィン変換という変形処理を加えた。アフィン変換とは、行列を用いた座標変換であり画像全体の変換だけでなく、1 ピクセルを行列の一つの要素として考えると、画像の部分的な変換が可能になる。本プロジェクトでは主に、画像の下半分を伸ばす下部垂直変形や、上半分を伸ばす上部垂直変形、Grad-CAM によって表現されモデルの重みに合わせた顔画像の変換などに使用した。プログラミング言語は Python で、主に Numpy,OpenCV といったライブラリを使用した。

変形の方法として、OpenCV の関数である remap を用いて、変形させたい部位に対し、変形させたい形になるように出力先の各座標を配列として格納する map を指定する。この map に基づいて、OpenCV の関数である remap を行う。モデル重みに合わせた画像の変換に関しては、Grad-Cam によって与えられたピクセルごとの重みを 0~1 の値で表現した配列を基に、適切に画像が誇張されるように値を調整し、remap 関数を用いて画像の変形を行った。

(※文責: 立山雄晟)

3.2 実験内容

3.2.1 実験 1 変形顔画像実験

Hole ら (2002) によって行われた先行研究では、変形の種類によって顔認識の精度の変化に有意な差が見られた [19]。この結果は、顔認識という人間の知覚において、多くの人に共通した知覚の偏り、特性を示した。我々は深層学習モデルを利用して同様の研究を行い、その結果を先行研究と比較することで、深層学習モデルが人間と類似した知覚の特性を示すかどうかを研究した。

実験では、3.1.2 で説明された ResNet を用いた。初めに、ResNet に約 331 万枚の顔画像を学習データとして、8631 人の顔を識別するように学習を行わせた。顔画像は、VGG-Face2 と呼ばれる顔認識のための顔画像データセットを利用した (8.2.1 で後述)。次に、実験用のテストデータとなる顔画像約 40 万枚を用意し、それぞれの顔画像に対して 2.1 章で示した先行研究で行われた変形の一部をアフィン変換によって施した。最後に、変形させた顔画像とテストデータを ResNet に与え、変形させた顔画像をどれだけ正しく判別できるかを調査する認識精度実験を行わせた。

(※文責: 前澤榛人)

3.2.2 実験 2 カリカチュア画像の認識精度実験

Mauro ら (1992) によって行われた先行研究では、カリカチュアによって顔の特徴が誇張された場合の方が本来の顔よりも認識精度が向上したことが示された [20]。このことは、人間の顔認識と特徴抽出の能力の関係性を補強するものであり人間特有の知覚の特性であるといえる。3.2.1 と同様に、我々は深層学習モデルを利用して同様の研究を行い、その結果を先行研究と比較することで、深層学習モデルが人間と類似した知覚の特性を示すかどうかを研究した。

実験では、前述の実験 1 と同様の ResNet を用いた。まず初めに、実験用のテストデータとなる顔画像に対して、3.1.3 で説明された Grad-CAM に基づいて、特徴的な部分をより誇張するよう

Make Brain Project

に画像を変形させ、カリカチュアを作成した。作成したカリカチュアとテストデータを ResNet に与え、実験 1 と同様にカリカチュアをどれだけ正しく判別できるかを調査する認識精度実験を行った。

(※文責: 前澤榛人)

第 4 章 結果・考察

4.1 実験 1

4.1.1 実験 1 の結果

我々が行った実験の結果と比較用の先行研究の結果を図 4.1 に示す。我々が深層学習モデルを用いて行った実験の結果は図 4.1(a) であり、先行研究で行われた実験の結果は図 4.1(b) である。図はどちらも正立条件の正答率に対する各変形条件の正答率の差を割合で示している。各条件の正答率(正立条件比)は、正立条件:93.626%、倒立条件:0.556% (−93.03%)、せん断条件:20.301% (−73.325%)、垂直方向変形:79.367% (−14.269%)、水平方向変形:68.707% (−24.919%)、上部垂直変形:52.726% (−40.90%)、下部垂直変形:63.615% (−30.011%)、正立条件ばかり:65.0% (−28.626%) となった。

先行研究と類似している点は主に以下の 3 点である。

1. 倒立条件において大幅に認識精度が低下している
2. せん断変形条件において中程度に認識精度が低下している
3. 垂直方向変形条件、水平方向変形条件における認識精度の低下率が小さくなっている

すべての条件で傾向が一致する結果とはならなかったが、一部の条件においては定性的に類似した認識精度の偏りの傾向が見られた。

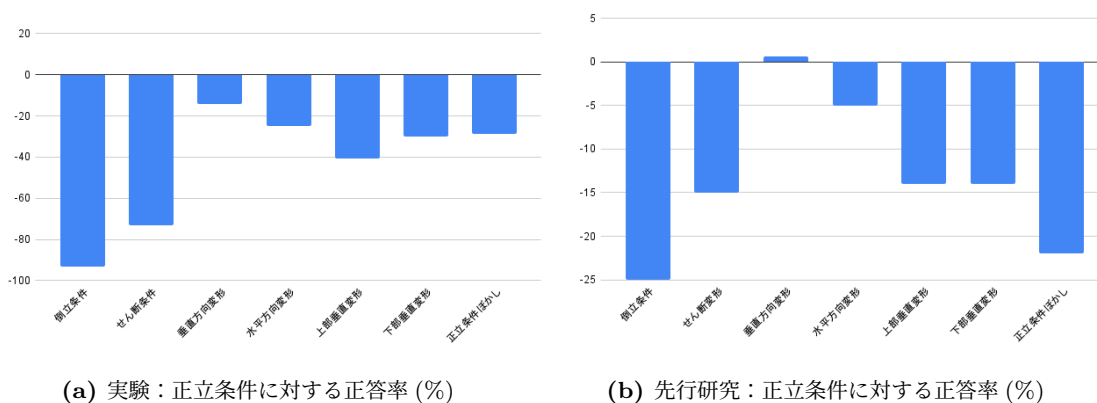


図 4.1: 実験と先行研究における成立条件に対する正答率

(※文責: 前澤榛人)

4.1.2 実験 1 に関する考察

前述の 4.1.1 より、深層学習モデルは、先行研究で示された人間の知覚特性の一部を定性的に再現したと言える。更にこのことによって、同様に顔認識を主題にした心理実験を必要とする研究において、人間の被験者の代わりに人間と類似した知覚特性を一部持つ深層学習モデルによって代用が出来る可能性が示される。深層学習モデルによる心理実験の代用は、人間の被験者で心理実験を

行った際に発生する認知バイアスやスキーマ、プライミングといった様々なパラメータの差異やそれによる結果のズレが発生するといった問題点やリスクを大幅に低減することが出来るといった利点が存在する。このことは、上記のような心理実験の問題点を解消し、より正確性の高い実験結果を得られる可能性に繋がる。

また、描画能力と実験結果の関係性について、我々の行った実験において上部垂直変形条件と下部垂直変形条件の間に正答率の差が見られた。一方、既存の描画研究において人が顔の絵を描画する際に目の位置を高く描いてしまうという偏りがみられることが示されている [23]。この原因に関しては未だ説明されていないが、描画プロセスの一つである物体認識が原因である可能性が示唆されている。

この可能性と我々の研究を関連付けると、下部垂直変形の方が人間が認識しやすい要素を含んでいるのではないかと考える。下部垂直変形は、顔の大きさに対して目の高さが高くなる。我々の実験において、下部垂直変形条件で認識精度が向上したことが、顔の大きさに対して目の高さが高くなったことに起因すると仮定するならば、人が顔を描く際に目の高さを高く描いてしまう偏りがあることは、無意識的に認識しやすい顔を描くようになっているという可能性が考えられる。

しかし、先行研究においては上部垂直変形と下部垂直変形において精度の差が見られなかったこと、上部垂直変形と下部垂直変形の平均反応時間を比較した際に下部垂直変形の方が平均反応時間が大きいことや、先行研究と我々の実験の両方において、上部垂直変形と下部垂直変形を比較した際には下部垂直変形の方が精度は高いが、どちらも正立条件に対しては精度の低下がみられることなど、上記の可能性を支持するには十分ではない要素も含まれており、更なる研究を進めていく必要がある。

(※文責: 前澤榛人)

4.2 実験 2

4.2.1 実験 2 の結果

変形を行ってない画像の認識精度は 93% に対し、Grad-CAM によって変形を行った画像による認識精度は 85% と、変形を行っていない画像に対し 8% 低下した。

(※文責: 和泉友人)

4.2.2 実験 2 に関する考察

4.2.1 で示されたように、変形後の画像に対する認識精度はオリジナルの画像に対する認識精度を下回った。この原因として、脳とニューラルネットワークにおける顔の表現のされ方の違いが考えられる。神経科学や心理学の分野において、ノルム基準顔符号化 (norm-based face encoding) という神経符号化の仮説が提案されている [24], [25]。この仮説によれば、特定の個人の顔は、プロトタイプ (Prototype) と呼ばれる規範的な顔からどの程度逸脱しているかによって表現されている。Mauro と Kubovy の論文は、カリカチュア的な変形によって認識速度が向上することは、この符号化の様式から説明できるとしている [20]。すなわち、変形によって顔の特徴が誇張されることでプロトタイプからの逸脱度合いが大きくなり、その結果認識されやすくなるということである。一方で、この仮説はより広範な物体認識の領域では提案されていない。今回の実験では、物体

Make Brain Project

認識を行うように事前に訓練されたネットワークをファインチューニングすることによって顔認識のモデルを作成した。そのために、我々のモデルはノルム基準の符号化を獲得しなかった可能性がある。その結果、特徴の誇張が精度向上の助けとならず、単に精度を低下させたと考えられる。したがって、今後の実験でカリカチュア効果の再現を試みる場合には、学習則や訓練方法を改良することによって、ノルム基準の符号化をモデルに獲得させる必要があると考えられる。

(※文責: 和泉友人)

第5章 まとめ

5.1 プロジェクトの成果

我々は深層学習モデルを用いて描画研究の心理実験を再現することを目的として2つの実験を行った。実験1では先行研究の結果と深層学習モデルを使った実験の結果の一部に偏りの傾向が見られたため、深層学習モデルが人間の知覚特性の一部を定性的に再現したと言える。また、描画能力に関して人が顔の絵を描画する際に目を実際の位置よりも上に描くことから下部垂直変形の認識精度が高くなると予想していた。だが、結果はそれを支持するには不十分であった。実験2では変形後の画像に対しての認識精度が元画像よりも下回った。原因は顔の符号化様式が脳と今回利用したネットワークで異なるからである。人間は規範的な顔からどの程度逸脱しているかによって顔の特徴を検出しているが、今回の実験で用いたネットワークはそのような方法で人の顔を判断していない。今回行った実験で、描画研究における深層学習モデルの代用可能性を示すことができた。加えて、予想と異なる結果から原因の考察を行った。このことから追加実験を行う必要がある。

(※文責: 池田蓮馬)

5.2 今後の課題

本プロジェクトにおける今後の課題は4つある。

- 1つ目は、代用可能性を示せた深層学習モデルを用いて、パラメータの差異などによって起こりうる、人間の被験者では行うことが困難な心理実験を行うことである。
- 2つ目は、本プロジェクトで行った実験や、一つ目で述べたような実験の深層学習モデルの重みなどから、神経科学的な観点より描画研究のメカニズムが明らかにすることである。
- 3つ目は、実験1や実験2の一部は心理実験とは違った結果が出ており、このようなことが起きた原因の考察に基づき追加の実験を行う必要がある。
- 4つ目は、今回調査した認識能力は、描画能力を構成するうちの1つであり、描画能力を深層学習モデルで完全に代用するためには、長期記憶や、実際に描くというプロセスなど、他の描画能力を構成する要素についても追加の実験などを行う必要がある。

(※文責: 立山雄晟)

第 6 章 外部評価

6.1 中間発表

中間発表では、Google フォームでのアンケートを用いて、評価を記入してもらった。評価は発表技術と発表内容についての 2 項目で行う。それぞれに 10 段階での評価とその理由や感想を自由記述にて書いてもらった。結果的に 46 人の方々からフィードバックを得ることができた。まず、発表技術についての評価を示す。46 人の平均評価点は 7.8 点であった。発表内容の平均評価点は 8.2 点であった。(どちらも小数点以下第二位で四捨五入) また、評価理由の記述を一部以下に示す。

発表技術

- モニターを見ながら発表するのであれば、もっと大きな声で発表すべき (全員ではない) 発表練習不足なのか聴衆を見ていない発表者がいた。
- スライドは工夫されていましたが説明が不足しているのでは、と思われる箇所がありました。
- グループ B の発表がとてもわかり易かった。質問も丁寧にさせていただき専門外でしたが理解することができました。
- 話す速度、スライドの内容も良かった。スライドについて、一瞬流れるのが早く感じた。補足スライドを用意していて良いと感じた。
- 例えや図などを活用していて、機械学習について知識の少ない人でもわかりやすいような発表をしていて良かった。多くの先行研究と関連づけて発表を進めており説得力があった。

発表内容

- 脳を作るというのは、プログラムで人間の脳みそを作るといよりは、人間らしい振る舞いをするプログラムを作るということ? プロジェクトの名前とサービスが思っていたものと違ってた。それについての説明を最初にして欲しかったです。
- 専門用語に関する説明を図や絵などを用いていたので具体的でわかりやすかった。質問に対する返答も質問の的を得た説明を行っていたため対応がよかった。
- もう少しわかりやすく目的や方法を示したほうが良いと思う。また、内容が難しく伝わりにくいので簡潔にわかりやすくしたほうが良いと思う。(どちらのグループも)
- グループ B の発表で錯視についての資料が多く人間の脳が勘違いするような絵というのが伝わらなかった。
- 難しい単語が多く、何も知らない人は理解しにくいと思いました。

総評としては得られた評価は 7 点以上が全体を占めており、1~3 点などの低い評価は少なかった。発表技術については、スライドが見やすく理解しやすいもので、補足スライドについてもいい評価を得られた。しかし、一部聴講者から、発表者の声が聞き取りづらいとの意見も得られた。このこ

とから、全体の発表練習不足であったと考えられる。発表内容については、目的や内容がわかりづらいといった意見が多く見受けられた。前期活動は、基礎知識の修得をメインに作業していたため、具体的な最終目標を定めることが遅くなったことで、成果物の制作に着手できなかったことが原因として考えられる。事前知識のない聴講者にとって難しい言葉などの説明をするべきであった。また、プロジェクトのテーマが認識と違っていたことで混乱したという意見が見受けられた。最初にプロジェクトテーマについて説明することでグループごとの活動を理解しやすいものにするべきであった。資料作成において、グループ内のみで完結していた。別のグループと資料の添削を行うことで多様な意見を得ることができたと考える。また、ポスターは一目見て内容を理解することが難しいものになっており、聴講者の興味を引きにくいものであった。中間発表は理解の難しさが反省点にあげられる。反省点を踏まえて後期からの活動に生かしていきたい。

(※文責: 池田蓮馬)

6.2 成果発表

成果発表では、Google フォームでのアンケートを用いて、評価を記入してもらった。評価は発表技術と発表内容についての2項目で行う。それぞれに10段階での評価とその理由や感想を自由記述にて書いてもらった。結果的に50人の方々からフィードバックを得ることができた。発表技術の平均評価点は7.7点、発表内容の平均評価点は8.3点であった。(どちらも小数点以下第二位で四捨五入) また、評価理由の記述を一部以下に示す。

発表技術

- 発表箇所がどこなのか指で指示していたのでわかりやすかった。Bグループの方で声が少し小さかったため、少し分かりにくかった。
- スライドが見やすい、スライドの色使いが見やすい、文字が少なく図が多かったためわかりやすかった。
- 観客に視線を配りつつ、スライドの該当箇所を示しながら発表を行っていたため非常に良いと感じた。問題点としてはPCを持ちながらでは不恰好である点と時間配分に問題があると感じた。
- ロゴがかっこいい。
- 発表者によってわかりやすい発表だったり、逆に分かりにくく聞き取りづらい発表であったりと、差が激しいものになっていました。発表を行う際は、画面ではなく聴衆に顔を向けた方がいいと思います。また、少々辿々しい印象を受けました。
- 難しい内容ではあったが、概要図や動画で分かりやすくする工夫があった。少し聞こえにくかった。

発表内容

- 深層学習について上手くまとめられていたり、先行研究などを用いて結果などをわかりやすく示せており良いと感じた。
- Bグループはただ追実験したのではなく、結論として今後活かせるようなことをきちんと挙げていてよかった。
- 発表中に「心理学は理論的ではない」という発言がありましたが、そのように言ってし

まってよいのでしょうか？

- グループ B プロジェクト学習で画像処理を用いた機械学習の観点で考えると、レベルの高い内容であったと思う。しかし、学習させた人工知能を人間の脳と同様に扱うことが可能になるという考えは、結果が学習モデルに依存するため、難しいのではないかと思う。全体他のプロジェクトと比較してもかなりレベルの高い内容でした。この内容を上げるためにかなり大変苦労して活動してきたと思います。
- 心理実験データと対応するのは面白い。どのようなモデルパラメタの範囲で類似度が高まるのかを明示できるとよりよいと思った
- プロジェクトの目標が脳を作るだったので、そのゴールが設定されていたらわかりやすかったと思いました。人間の研究と比べて差がほぼなかったらゴール？ それとも他のモデルと比較して良かったらゴール？もしかしたら別なゴールがあるかも知れません。それを提示してこれぐらい出来ましたと言ってくれればすごい！となると思います。

評価点数に関しては、8点や9点の評価が多く見受けられた。それに加えて、1~4点といった極めて低い評価はなかった。各評価の平均点を中間発表と比較すると、発表技術は0.1点下がり、発表内容は0.1点上がっており、中間発表との大きな変化は見受けられなかった。発表方法は中間発表と大きく変化しており、前期で行ったスライドを用いて発表する形式に加えて、グループ用のポスターと興味を引くためのデザインを重視したポスターを制作した。前期との変化によっての発表中の大きなトラブルは無かったが、準備の作業が増えたことに加え、成果物の作成に時間を取られたため、前期よりも発表準備の時間を確保することができず、全体での発表練習が減ってしまった。そのため分かりやすい発表であったり、聞きづらい発表であったりと聴講者によって意見の違うコメントが見受けられたと思う。中間発表の反省点であった、準備のための時間を確保することが達成できなかったのは心残りである。発表内容については、スライドの最初にプロジェクトのテーマ紹介を行ったことで、各グループで大きく異なったテーマによって聴講者が混乱を阻止したと考えられる。しかし、発表内容が増えたことによってグループごとの発表時間が減り、短い時間で研究発表をまとめなければならなかった。聴講者のコメントでは、発表内容が理解できないなどの意見は見受けられない。内容の理解しやすさは中間発表での反省点の一つとして挙げられていたことであったので、その点については、前期の反省を十分に活かし改善することができたと思う。その一方で結果に関する意見がたびたび見受けられる。結果を支持する先行研究の説明を省略したことが原因であると考えられる。発表の反省としてグループごとに発表を分けるべきであると意見があったので、考慮して検討すべきと思われる。しかし、良い意見も多くあったことから、発表を通して活動内容を理解していただけたと考える。

(※文責: 池田蓮馬)

第 7 章 活動内容

7.1 前期活動内容

前期では以下の活動を行った

7.1.1 コミュニケーションツールの決定

連絡手段としては、プロジェクト全体としても使っていた Slack と、音声会議用に Discord を使用した。また、クラウド型ワークスペースである Notion を用いて、活動毎に議事録作成を行った。

(※文責: 池田蓮馬)

7.1.2 必要な知識の学習

前期では、主に活動に必要な知識の学習を行った。本グループは人数の関係上、神経回路を用いたシミュレーションを希望するメンバーと、深層学習モデルを用いた画像生成を希望するメンバーによって構成されたグループであった。そのため、すぐにグループテーマの決定をするのではなく、脳や神経科学について知識を付けたうえで、メンバーごとで案を出し合い、話し合いをすることでテーマを決定する方針を執った。学習方法として、「カールソン神経科学テキスト 脳と行動」と「計算論的精神医学」の二つのテキストを主に使用することとした。カールソン神経科学テキストでは脳や神経科学について我々が習得すべき知識に関して描かれている章を 7 つほど選び学習をおこなった。学習方法としては、プロジェクトの時間を使い全員で読書をした後、章末にある復習問題を分担して解き、自分が担当した問題を他のグループメンバーに説明するという方法をとった。計算論的神経科学では、一人がプロジェクト時間外に事前学習として内容を理解たうえでスライドを作成し、プロジェクト内で発表をするといった方法を取り、効率的に知識の習得を進めていった。また、テーマ決めで多様なアイデアを出すために、各自でテーマ決めのための本を決め、その本の内容で気になった部分などをまとめてグループで発表を行った。

(※文責: 池田蓮馬)

7.1.3 グループテーマの決定

脳や、神経科学についての学習を行った後、それぞれプロジェクトでやりたいことについて意見を出し合い、グループテーマを決定した。意見を基に関連研究の調査を行っていく中で、人間におこる知覚のエラーによって引き起こされる、視覚に関する錯覚である錯視に注目した。さらに話し合いや錯視に関連する文献調査を行い、グループテーマは「人間の知覚特性を生かした錯視画像の生成」とし、人工知能による人間の視覚特性を利用したアート生成の実現可能性を示すことを目標に置いた。また、グループテーマが絵を生成するという数値的な成果が見えづらいものであったため、「人間の知覚特性である錯視を知覚再現できるモデルを使用する」「人間らしさを評価できるような指標を検討し、それにより絵を評価する」といった成果における指標を定めた。

(※文責: 池田蓮馬)

7.1.4 関連研究の調査

グループテーマに基づき関連研究を調査した結果、錯視画像を脳機能の理論の予測符号化に基づいて作られた深層学習モデルである「PredNet」が動きを予測したという研究を見つけた。我々は、この PredNet と、同様に錯視に関する先行研究が存在した、次元圧縮を行う深層学習モデル AutoEncoder を用いて、錯視画像の生成を目指した。

(※文責: 池田蓮馬)

7.1.5 前期における活動目標の設定

グループテーマの決定に伴い、前期における活動目標を以下の三つに設定した。

- テキストを読み、神経科学についての基礎知識を付ける
- 錯視に関する論文を読み、錯視の作成に必要な知識を獲得する
- PredNet に関する論文を読み、PredNet 実装に必要な知識を獲得し実装を行う

(※文責: 立山雄晟)

7.1.6 中間発表に向けた準備

発表の2週間前からは中間発表に向けたポスターやスライド資料、プレゼンテーションの準備を行った。ポスターに関しては、プロジェクト単位で作成するもののため、グループ A とデザインや章立てに関して話し合いを行ったうえで作成し、教員からのフィードバックを頂き修正といった工程をとった。スライド資料では、章ごとに担当を決め作業したうえで全員で内容を見直した。プレゼンテーションの準備では、当日を想定してモニターを使用しての発表練習をした。また、中間発表の前日には他のグループと合同で発表練習を行った。発表内容は、前期では知識の習得と、グループテーマの決定が主な活動だったため、背景や目的、深層学習モデルの具体的な説明を主にを行った。また、質問に対して答えられなかったりといったことを防ぐために、質問想定とその回答を用意した。

(※文責: 立山雄晟)

7.1.7 中間報告書の作成

章立てに関しては昨年の報告書を参考にし、教員に相談をしたうえで決定した。中間報告書においても、中間発表でのスライド作成と同様に章ごとに役割分担を行い最後に全員で内容の見直しを行うという方法で取り組んだ。Tex 形式での記載が推奨されていたため、Overleaf という Tex を共同で編集できるツールを使用した。内容については、知識の習得、文献調査と前期で行えた活動が少なかったため、前期段階で記載できる内容が少なかった。

(※文責: 立山雄晟)

7.1.8 前期の振り返り

前期は知識の習得とグループテーマの決定が主であり、中間発表、中間報告書では成果物を発表することはできなかった。特に、全員やりたいことが微妙に違いグループテーマの決定にかなりの時間を使ってしまった。また、中間発表、中間報告書ともにスケジュールの見通しが甘く、直前になって慌てることとなってしまった。一方で、知識獲得に十分に時間をかけたため、グループメンバーの共通知識を高めることが出来た。これにより、目的の制定において多様な案が提案され、メンバーが納得するグループテーマが立てられた。後期で結果的にグループテーマは変わってしまったが、後期で再度グループテーマを決定する際には、前期での経験を踏まえて比較的スムーズに話し合いを進めることが出来た。

(※文責: 立山雄晟)

7.2 夏季休暇活動内容

7.2.1 夏季休暇における活動

夏季休暇では、前期の進捗具合から後期だけでは十分な成果を得ることが難しいと考えたため活動を行った。前期終了段階で、グループテーマや必要な習得技術については大方見通しがついていたが、錯視画像の生成をどのように行うかといった、具体的な手法については決定していなかった。そのため、錯視に関する文献調査を行い具体的な手法を決定するメンバーと、深層学習用フレームワークである PyTorch を用いた深層学習モデルの学習と、PredNet の実装を行うメンバーに分かれ作業を行った。また、お互いの作業状況を把握するために、週に一回程度オンラインで進捗報告を行うようにした。

(※文責: 立山雄晟)

7.2.2 夏季休暇の振り返り

夏季休暇ではメンバーの帰省や、インターンシップへの活動などにより想定よりも活動に使える時間が少なかった。そのため、具体的な手法の決定や、PredNet の実装に関しては後期へと持ち越すこととなった。

(※文責: 立山雄晟)

7.3 後期活動内容

後期では以下の活動を行った

7.3.1 グループテーマの再決定

当初は夏季休暇中に具体的なアプローチを決定し、後期から取り組むという活動計画を立てていた。しかし、錯視画像を知覚するような深層学習モデルの実装は可能なものの、人間の知覚特性を

応用した画像生成のアプローチは考案出来なかった。そのため、グループテーマの変更を検討するために、追加でさらに文献調査を行った。その過程で、絵画の練習をした経験のない人が人間の顔の絵を描く際、目を実際の位置よりも上に描いたり、顔を丸く描いたりするなどの誤りを共通して作るという論文を発見した [26]–[28]。我々はこの現象をグループテーマと関連付けられないかと考え、描画行為における関連研究の調査を進めた。その結果、描画研究では心理実験とその結果に対する考察がほとんどで、計算論的な方法が取られていないことが分かった。そのため、ほとんどの研究はある現象と複数の心理学的な構成概念との関連性を曖昧に述べるだけに留まっており、現象に対する明確な原因やその現象を左右するパラメータ、関連する脳領域などは分かっていなかった。加えて、グループメンバー共通のやりたいこととして、人間特有だと思われるものに対し、人工知能を使ってアプローチするというものがあった。そのため、グループテーマを「深層学習モデルの描画研究への応用」に変更し、特に描画能力を構成する能力の1つである認識能力に着目することとした。

(※文責: 和泉友人)

7.3.2 グループにおける役割分担

テーマが決定した後は、深層学習モデルで代用する心理実験を決めるため、描画能力に関連する文献調査を行い、論文をまとめて他のグループメンバーに説明する人、心理実験に用いる深層学習モデルの実装・訓練を行う人、画像変形や Grad-CAM のプログラムを作成する人というように、グループ内で役割分担を行い、並行して作業を行った。深層学習モデルの実装・訓練に少し時間がかかったが、その後の実験に関してはスムーズに進めることが出来た。

(※文責: 和泉友人)

7.3.3 成果発表に向けた準備

前期の活動で発表に向けての準備の時間を十分にとることが出来なかった反省を生かし、11月からは成果発表に向けたスライド作成や、プレゼンテーションの準備も並行して進めた。また、11月中旬ごろからは、全体ポスターなどプロジェクト単位で成果発表に向けてやるべきことを、グループ A と協力して行った。発表の二週間前からは、プロジェクト毎に、担当教員の前で発表練習を A グループと合同で行い、発表内容についてのフィードバックを頂いた。また、最終報告書に関しても年末には第 1 稿を先生方に校正をお願いするなど、早めに取り組むことを心掛けた。

(※文責: 和泉友人)

7.3.4 最終報告書の作成

前期とは異なるグループテーマとなったため、前期での中間報告書に内容を付け足すのではなく、新たに章立てから決定することとした。前年のグループ報告書等を参考にしながら、章立てと大方の記入内容を決定したうえで、先生にフィードバックを頂いた後、本格的に作業に取り組んだ。ツールは前期と同様 Overleaf を使用した。

(※文責: 立山雄晟)

7.3.5 後期の振り返り

夏季休暇の活動では計画通りに進まず、具体的な手法も決定していない状態で迎えた後期の活動であったが、無事成果発表までに実験とその考察を行うことが出来た。うまくいった要因としては、前期でのグループテーマに固執せずメンバーのやりたいことや文献調査を基に柔軟にグループテーマの変更を行ったことが挙げられる。また、テーマ決定後には、やるべきことを明確にしたうえで適切な役割分担やスケジュール管理が出来ていた。成果発表においても前述したとおり前期の反省を生かし、かなり早めから取り組んだ。しかしながら、A グループを加えた全体通しての発表練習を数回しか行えず、結果的に成果発表において全体での時間管理がうまくいかなかったことから、グループ間でのコミュニケーションを十分にすべきであったという反省点もある。

(※文責: 立山雄晟)

7.4 プロジェクトにおける各人の主な役割

7.4.1 立山

グループリーダー

グループリーダーとして、プロジェクトリーダー、A グループのグループリーダーとコミュニケーションをとり、決定事項をグループメンバーに伝えるといった活動を行った。また、プロジェクト開始時には本日の活動予定、終了時には本日の活動内容の共有を A グループと教員に対して行った。さらに、中間・期末発表や報告書など、期限が決まっており計画を立てる必要があるものに関しては、少し余裕を持たせて終わるようにスケジュールを組むことを心掛けた。日々の活動を議事録として記録した。

テキストの要約

前期での必要な知識の習得の際、一部のテキストに関しては全体で学習を行うのではなく、一人が事前に内容を理解しスライドにまとめたうえで発表を行うという学習方法をとった。私は、計算論的精神医学における第四章の内容である、生物物理学モデルについて要約を行いグループメンバーに説明を行った。

PredNet の調査・実装

PredNet のアーキテクチャを説明している論文 [29] を読み、スライドにまとめグループ内で発表を行った。また、蛇の回転錯視と呼ばれる錯視画像の動きを PredNet が知覚したという先行研究 [30] から、実際に追実験を行い、先行研究と同様の結果が得られるか調査することとした。プログラムは、先行研究の論文執筆者が公開していたものを使い、プログラミング言語は Python、深層学習に用いるフレームワークは PyTorch を使用した。結果は、先行研究と同様に蛇の回転錯視に対し動きを予測し、蛇の回転錯視に似た人間が動きを認識しない絵に対しては動きを予測しないという結果が現れた。

画像変形プログラムの作成

主に 4.1.2 で述べたアフィン変換という手法を用いた。実験 1 においては、倒立変形、せん断変形、垂直方向変形、水平方向変形、下部垂直変形、上部垂直変形、成立ぼかしの 7 種類の先行研究と同様の顔変形を行うプログラムを作成した。実験 2 においては重みに基づいて画像の変形を行うためのプログラムを作成した。使用言語は Python で、主に Numpy や OpenCV といったライブラリを使用した。

最終報告書の章立て

前期とはグループテーマが異なったものとなったため、前期の報告書の内容を補足していくのではなく、報告書を章立てから行うこととした。前年の最終報告書を参考に本グループでの活動内容に合うように章立てを行い、グループメンバーから意見を求めた。グループ内で共有し修正を行った後は、記入する大まかな内容も踏まえて、教員からフィードバックを頂いた。

(※文責: 立山雄晟)

7.4.2 前澤

論文調査

実験の基となる研究や基礎知識に関する論文を中心に読んだり、まとめて発表などをした。脳科学や神経科学、心理学、深層学習など様々な分野の論文を読み、プロジェクトの活動において必要になる知識や指標となる研究を調査した。論文で示された実験のデータを集約したり、複数の論文の内容をまとめて報告書に落とし込むなど、調査して得た内容を実際の活動と絡めて提示することで、プロジェクトの研究内容の成果をより補強した。報告書執筆の際にも、実験結果に対する考察の足掛かりとして、論文を読み返したり、新たな文献を調査するなどを行い、プロジェクト全体を通して既存の関連研究への調査を続けた。

実験データの整理

実験 1 で行われた深層学習モデルによる顔認識精度実験の結果データをまとめて図表化した。前述の論文調査の際に行った先行研究での実験データの図表化と合わせて、両実験の比較、検討が行えるようにした。図表化に際して、先行研究との比較を前提とした実験データの抽出や整理を行い、プロジェクトの活動として最終的な結論を見据えて調整を行った。

プロジェクトロゴ制作

本プロジェクトのロゴマークを作成した。プロジェクト名である「脳をつくる」をテーマに、ロゴ一つでなるべくプロジェクトの要旨が伝わるようなデザインを考えた。コンピュータによる脳機能の再現を目指すことから、コンピュータと脳を組み合わせたデザインを考え、ロゴの制作を行った。ロゴ制作、また後述のデザインポスター制作においては、Clip Studio Paint PRO を用いて制作を進めた。制作過程では、いくつかのデザイン案をメンバーに提示してその中から最終的なデザインの決定稿を選んでもらうなどし、単独の作業ではなくプロジェクトの活動として進行した。

デザインポスター制作

成果発表会におけるポスターの一つを制作した。メインのプロジェクトの内容を説明するポスターとは別に、デザイン面を重視することを目的にポスターを1枚作成することになった。デザインポスターでは、ロゴマーク同様、簡潔にプロジェクトの雰囲気が伝わるようなデザインを考えつつ、ポスターであることを意識してより目に留まりやすいような目立つデザインを意識して作成した。デザインポスター制作においては、プロジェクトロゴ制作の際も利用した Clip Studio Paint に合わせて、MS Power Point も利用して制作を進めた。前述のプロジェクトロゴと同様に、メンバーにポスターデザイン案を複数提示して意見を募った。最終的にはデザイン案として提示した3種類のポスターが成果発表会に利用されることとなった。

(※文責: 前澤榛人)

7.4.3 和泉

事前学習の際の書籍選定

プロジェクト活動の初期に行った必要な知識の学習 (§7.1.2) の際、学習に使用する書籍の選定を行った。選定は難易度、扱う範囲、プロジェクトとのマッチ具合などを基準として行った。「カールソン神経科学テキスト 脳と行動」は神経科学に関する内容を細胞生物学から高次脳機能まで広く扱っており、一章あたりの分量も輪講を行うにあたって適切だったため採用した。「計算論的精神医学」は計算論的神経科学に関するトピックを広く扱っており、テーマの方向性を決定するにあたって参考になるため採用した。

文献調査

具体的なグループテーマの策定や実験の計画、結果の考察のために文献調査を行った。文献調査の際には情報ライブラリーや Google Scholar を利用した。さらに、読んだ論文は Paperpile という文献管理ソフトウェアで管理し、いつでも読み直せるようにしておいた。また、調査した文献の一部をスライドに整理し、グループメンバーに向けて発表した。

実験計画

文献調査から得た知識を基に実験 1 (§4.1)、実験 2 (§4.2) の内容を考案した。

プログラムの実装

実験 1 (§4.1)、実験 2 (§4.2) で用いた顔認識モデルの実装を、Python を用いて行った。実装の詳細については §8 に詳しく記す。また、実装した顔認識モデルと立山の実装した顔変形プログラムを組み合わせることで顔認識モデルを評価するプログラムの実装を行った。

作業の分担

考案した実験を実施するために作業を分担する作業を行った。実験で用いる一部のプログラムの実装や結果の整理など、実験のプロセスをいくつかに分け、負担が均等になるようメンバーに割り振った。

発表準備

成果発表会のためのスライドを他メンバーと分担して行った。スライドを制作する際は、特にストーリーが伝わることを意識した。また、発表練習や想定質問の作成も行った。

(※文責: 和泉友人)

7.4.4 池田

文献調査

プロジェクトの具体的なグループテーマの制定のために描画研究の先行研究の調査を行った。その結果、描画研究では心理実験がほとんどであることが分かった。文献調査においては情報ライブラリーと Google Scholar を主に利用した。読んだ文献は、スライドにまとめ、グループ内で発表を行うことで、進捗状況の確認と情報共有を図った。

発表準備

成果発表のためのスライド制作を行った。スライドは理解のしやすさを第一に考えて作成した。自分の担当外のスライドやポスターについては意見や質問などで積極的に取り組み、ポスターの精査を行った。スライドの完成後は、発表練習を行った。

深層学習の習得

今回、実験で用いた深層学習モデルの PredNet を実装するために、深層学習の勉強を行った。深層学習の本を用いて、深層学習のメカニズムの学習を行い、その結果をスライドにまとめて、グループ内で発表を行った。その上、CNN の実装を行うことで学習を行った。

(※文責: 池田蓮馬)

第 8 章 技術面

8.1 開発環境

8.1.1 プログラミング言語

本グループは Python によって実験等に用いるプログラムを実装した。また、pyenv (<https://github.com/pyenv/pyenv>) というバージョン管理マネージャを使用することでグループ全体の環境をバージョン 3.8.12 に統一し、複数の PC 上で円滑にプログラムを実行出来るようにした。さらに、プログラムを実装する際、PyTorch や OpenCV、Numpy、Matplotlib などのライブラリを使用した。PyTorch は深層学習フレームワークの一つであり、ニューラルネットワークの実装や訓練を簡単に行うことが出来る。OpenCV は画像処理を行うライブラリであり、顔画像の変形を行う際に用いた。Numpy はベクトル・行列の演算に特化したライブラリであり、通常 Python で実装すると低速になるベクトル・行列の演算を高速に行える。本グループでは特に、画像を扱う用途で利用した。Matplotlib はグラフの描画を行うためのライブラリであり、顔認識モデルの学習の推移などを可視化するために利用した。

(※文責: 和泉友人)

8.1.2 実行環境

本グループでは実装・実験を行うために 3 台の PC を使用した。それぞれの仕様は以下の通りである。

- PC 1
 - CPU: AMD Ryzen 9 5950X
 - GPU: NVIDIA RTX 3070
 - RAM: DDR4 64GB
- PC 2
 - CPU: Intel Xeon Gold 6126
 - GPU: NVIDIA Tesla V100
 - RAM: DDR4 128GB
- PC 3
 - CPU: AMD Ryzen 7 3700X
 - GPU: NVIDIA RTX 3070
 - RAM: DDR4 64GB

OS には Ubuntu 20.04 LTS を全ての PC で共通して利用した。また、CUDA のバージョンは 11.1 に統一した。

(※文責: 和泉友人)

8.2 実験方法の詳細

8.2.1 顔認識モデル

我々は ResNet-50 というニューラルネットワークを訓練することで、8631 人の顔画像を識別する顔認識モデルを作成した。顔画像には VGGFace2 というデータセットを用いた。

ResNet-50 は ResNet のバリエーションの一つであり、50 層からなる畳み込みニューラルネットワークである。224 × 224 × 3 の入力に対応しており、典型的には 224 × 224 の RGB 画像が入力として用いられる。今回の実験では、ResNet-50 をファインチューニングすることで顔画像を識別するように訓練した。ファインチューニングは転移学習手法の一つで、事前に別のデータで訓練した学習済みモデルの出力層だけを置き換え、新たなデータでモデルを訓練する手法である。今回の実験では ImageNet という物体認識データセットで訓練した ResNet-50 を学習済みモデルとして用いた。

VGGFace2 はオックスフォード大学の Visual Geometry Group が提供する大規模な顔画像のデータセットである。VGGFace2 には 9131 人分、約 331 万枚の顔画像が含まれている。1 人あたりの顔画像数はおよそ 362 枚である。VGGFace2 は訓練セットとテストセットに分けられており、訓練セットには 8631 人分、テストセットには 500 人分のデータが含まれている。今回の実験ではこのうち訓練セットだけを使用した。

(※文責: 和泉友人)

参考文献

- [1] A. Brumm, A. A. Oktaviana, B. Burhan, *et al.*, “Oldest cave art found in sulawesi,” en, *Sci Adv*, vol. 7, no. 3, Jan. 2021.
- [2] A. Kozbelt, “Artists as experts in visual cognition,” *Vis. cogn.*, vol. 8, no. 6, pp. 705–723, Dec. 2001.
- [3] A. Kozbelt, A. Seidel, A. ElBassiouny, Y. Mark, and D. R. Owen, “Visual selection contributes to artists’ advantages in realistic drawing,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 4, no. 2, pp. 93–102, May 2010.
- [4] D. J. Cohen, “Look little, look often: The influence of gaze frequency on drawing accuracy,” en, *Percept. Psychophys.*, vol. 67, no. 6, pp. 997–1009, Aug. 2005.
- [5] P. van Sommers, “A system for drawing and drawing-related neuropsychology,” *Cogn. Neuropsychol.*, vol. 6, no. 2, pp. 117–164, Mar. 1989.
- [6] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY, USA: Henry Holt and Co., Inc., 1982, ISBN: 0716715678.
- [7] S. Roncato, G. Sartori, J. Masterson, and R. Rumiati, “Constructional apraxia: An information processing analysis,” *Cogn. Neuropsychol.*, vol. 4, no. 2, pp. 113–129, May 1987.
- [8] F. La Femina, V. P. Senese, D. Grossi, and P. Venuti, “A battery for the assessment of visuo-spatial abilities involved in drawing tasks,” en, *Clin. Neuropsychol.*, vol. 23, no. 4, pp. 691–714, May 2009.
- [9] F. Guérin, B. Ska, and S. Belleville, “Cognitive processing of drawing abilities,” en, *Brain Cogn.*, vol. 40, no. 3, pp. 464–478, Aug. 1999.
- [10] S. McCrea, “A neuropsychological model of free-drawing from memory in constructional apraxia: A theoretical review,” en, *Am. J. Psychiatry Neurosci.*, vol. 2, no. 5, p. 60, 2014.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Dec. 2015. arXiv: 1512.03385 [cs.CV].
- [12] P. Xu, T. M. Hospedales, Q. Yin, Y.-Z. Song, T. Xiang, and L. Wang, “Deep learning for Free-Hand sketch: A survey,” Jan. 2020. arXiv: 2001.02600 [cs.CV].
- [13] B. A. Richards, T. P. Lillicrap, P. Beaudoin, *et al.*, “A deep learning framework for neuroscience,” en, *Nat. Neurosci.*, vol. 22, no. 11, pp. 1761–1770, Nov. 2019.
- [14] N. Kriegeskorte, “Deep neural networks: A new framework for modeling biological vision and brain information processing,” *Annu. Rev. Vis. Sci.*, vol. 1, no. 1, pp. 417–446, Nov. 2015.
- [15] M. Schrimpf, J. Kubilius, M. J. Lee, N. A. Ratan Murty, R. Ajemian, and J. J. DiCarlo, “Integrative benchmarking to advance neurally mechanistic models of human intelligence,” en, *Neuron*, vol. 108, no. 3, pp. 413–423, Nov. 2020.

- [16] G. W. Lindsay, “Convolutional neural networks as a model of the visual system: Past, present, and future,” en, *J. Cogn. Neurosci.*, vol. 33, no. 10, pp. 2017–2031, Sep. 2021.
- [17] A. Kozbelt and W. P. Seeley, “Integrating art historical, psychological, and neuroscientific explanations of artists’ advantages in drawing and perception,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 1, no. 2, pp. 80–90, May 2007.
- [18] D. L. Yamins and J. J. DiCarlo, “Using goal-driven deep learning models to understand sensory cortex,” *Nature neuroscience*, vol. 19, no. 3, pp. 356–365, 2016.
- [19] G. J. Hole, P. A. George, K. Eaves, and A. Rasek, “Effects of geometric distortions on face-recognition performance,” *Perception*, vol. 31, no. 10, pp. 1221–1240, 2002.
- [20] R. Mauro and M. Kubovy, “Caricature and face recognition,” *Memory & Cognition*, vol. 20, no. 4, pp. 433–440, 1992.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [22] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [23] J. Ostrofsky, D. J. Cohen, and A. Kozbelt, “Objective versus subjective measures of face-drawing accuracy and their relations with perceptual constancies,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 8, no. 4, p. 486, 2014.
- [24] T. Valentine and V. Bruce, “The effects of distinctiveness in recognising and classifying faces,” en, *Perception*, vol. 15, no. 5, pp. 525–535, 1986.
- [25] D. A. Leopold, I. V. Bondar, and M. A. Giese, “Norm-based face encoding by single neurons in the monkey inferotemporal cortex,” en, *Nature*, vol. 442, no. 7102, pp. 572–575, Aug. 2006.
- [26] K. Smith, V. Kempe, and L. Wood, “Eye placement bias is remarkably robust,” en, *Iperception*, vol. 12, no. 3, p. 20416695211017564, May 2021.
- [27] J. Ostrofsky, A. Kozbelt, M. Tumminia, and M. Cipriano, “Why do non-artists draw the eyes too far up the head? how vertical eye-drawing errors relate to schematic knowledge, pseudoneglect, and context-based perceptual biases,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 10, no. 3, pp. 332–343, Aug. 2016.
- [28] N. R. Harrison, J. Jones, and S. J. Davies, “Systematic distortions in vertical placement of features in drawings of faces and houses,” en, *Iperception*, vol. 8, no. 1, p. 2041669517691055, Jan. 2017.
- [29] W. Lotter, G. Kreiman, and D. Cox, “Deep predictive coding networks for video prediction and unsupervised learning,” *arXiv preprint arXiv:1605.08104*, 2016.
- [30] E. Watanabe, A. Kitaoka, K. Sakamoto, M. Yasugi, and K. Tanaka, “Illusory motion reproduced by deep neural networks trained for prediction,” *Front. Psychol.*, vol. 9, pp. 345–356, Mar. 2018.