

修士論文

医用画像読影教育システムのための深層学習による 眼球運動の分類手法

神川 怜史

奈良先端科学技術大学院大学
先端科学技術研究科
情報理工学プログラム

主指導教員：金谷 重彦 教授
計算システムズ生物学研究室（情報科学領域）

令和 2 年 3 月 13 日提出

本論文は奈良先端科学技術大学院大学先端科学技術研究科に
修士(工学) 授与の要件として提出した修士論文である。

神川 怜史

審査委員：

金谷 重彦	教授	(主指導教員, 情報科学領域)
佐藤 嘉伸	教授	(副指導教員, 情報科学領域)
佐藤 哲大	客員准教授	(副指導教員, 情報科学領域)
黄 銘	助教	(副指導教員, 情報科学領域)

医用画像読影教育システムのための深層学習による

眼球運動の分類手法*

神川 怜史

内容梗概

医学の高度な専門化と進歩に伴い、医用画像読影は経験の浅い医療従事候補者にとって限られた訓練期間内で学ぶことは容易ではない。本研究では読影スキルを解明することで、読影中の学習者に自身の読影行動をフィードバックできる教育システムの開発を目指す。まず読影行動を解明するため、Fixation や Saccade などの眼球運動を、教師なしクラスタリングである k-means 法および CNN (Convolutional Neural Network) を用いる 3 つのモデルで分類する手法を提案した。視線行動を記録した公開データを用い、最も高精度な手法を明らかにした。結果として、DeepCNN (DCNN) モデルが従来の視線研究で高精度である gazeNet よりも優れていた。また応用実験では、DCNN モデルを採用し、サンプリング周波数 120Hz の Tobii Pro X3-120 で実測したデータを用いて眼球運動を分類した。被験者は臨床経験のある放射線技師 2 人および放射線技師養成校の学生 2 人とし、30 枚の胸部 X 線画像を読影させた。分類結果では、30 枚全体の画像における Fixation 時間で経験者と学生で有意差があり、先行研究と同様の結果が得られた。今後は、医用画像を読影する際の眼球運動分類アプリケーションに実装し、教育現場で効果があるかを検証する必要がある。

キーワード

眼球運動, Convolutional Neural Network, 機械学習, 医用画像読影

*奈良先端科学技術大学院大学 先端科学技術研究科 修士論文, 令和 2 年 3 月 13 日.

Classification Method of Eye Movement by Deep Learning for Medical Image Interpretation Education*

Satoshi Kamikawa

Abstract

With high specialization and advances in medicine, medical image interpretation is not easy to learn within a limited training period for trainees. The objective of this study is to develop an education system that can provide feedback to trainees by clarifying the reading skills of experts. In our study, we proposed the analysis of fixation and saccade using k-means method and convolutional neural network (CNN) models, and compared several different modeling methods. The most accurate method was clarified using open data which recorded gaze behavior. As a result, The DeepCNN (DCNN) model was superior to gazeNet, which is highly accurate in previous gaze researches. In the application experiment, the DCNN model was adopted, and eye movement was classified using the data measured by Tobii Pro X3-120, with a sampling frequency of 120 Hz. The subjects, two radiological technologists with more than 5 years of clinical experience and two radiography students, read 30 chest X-ray images. In the result, there was a significant difference between radiological technologists and radiography students in fixation duration time in the whole 30 images, and the result was similar to the previous study. In the future, it is necessary to verify whether it is effective in the field of medical image interpretation education.

Keywords:

eye movement, convolutional neural networks, machine learning, medical image interpretation

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, March 13, 2020.

目次

第1章 序論	1
1.1 研究背景	1
1.1.1 Eye tracking	2
1.1.2 眼球運動	3
1.1.3 視線解析	3
1.1.4 胸部 X 線画像	4
1.2 関連研究	5
1.3 研究目的	8
1.4 本論文の構成	8
第2章 理論	9
2.1 眼球運動の研究の歴史	9
2.2 k-means 法	10
2.3 CNN(Convolutional Neural Network)	12
第3章 手法	14
3.1 提案手法	14
3.2 提案モデル	15
3.3 評価指標の算出	21
3.4 応用実験での統計手法	21
第4章 実験	22
4.1 基礎実験	22
4.1.1 データセット	22
4.1.1.1 Lund 2013	22

4.1.1.2 GazeCom	24
4.1.2 教師なしクラスタリング(k-means 法)	24
4.1.3 教師あり学習(CNN)	25
4.2 応用実験	25
4.2.1 データセット	25
4.2.2 視線追跡装置	31
4.2.3 実験環境	31
4.2.4 実験(モデル検証)	33
第5章 結果	34
5.1 基礎実験	34
5.1.1 教師なしクラスタリング(k-means 法)	34
5.1.2 教師あり学習(CNN)	37
5.2 応用実験	42
第6章 考察	49
第7章 結論	51
謝辞	53
参考文献	54
学会発表リスト	59

図目次

図 1	従来法と本研究の仮説	2
図 2	眼球運動	3
図 3	velocity 法の概念図	4
図 4	胸部 X 線画像の一例 (正面像)	5
図 5	gazeNet の訓練の流れ	6
図 6	k-means 法のアルゴリズム概念図	11
図 7	手書き数字データセットの例	13
図 8	グレースケールのパッチ画像	15
図 9	公開データのアノテーション付け	15
図 10	Lund 2013 の視線軌跡一例	23
図 11	胸部 X 線画像を見ている際の視線の軌跡の一例	25
図 12	12 点パッチ画像一例	26
図 13	経験者の ROC、LOC 解析結果	27
図 14	経験者の指摘箇所の詳細結果	28
図 15	学生の ROC、LOC 解析結果	29
図 16	学生の指摘箇所の詳細結果	30
図 17	Tobii X3-120 全体像	31
図 18	実験環境	32
図 19	実験風景	32
図 20	教師なしクラスタリング: Lund 2013 の PCA の結果	34
図 21	教師なしクラスタリング: Lund 2013 の k-means 法結果	35
図 22	教師なしクラスタリング: GazeCom の PCA 結果	35
図 23	教師なしクラスタリング: GazeCom の k-means 法結果	36
図 24	予測結果(左: Fixation、右: Saccade)	42
図 25	病変画像を対象とした Fixation 時間	44
図 26	病変画像を対象とした Saccade 時間	44
図 27	病変画像を対象とした総読影時間	45
図 28	全ての画像を対象とした Fixation 時間	46
図 29	全ての画像を対象とした Saccade 時間	46

図 30	全ての画像を対象とした総読影時間	47
図 31	経験者における病変画像と正常画像での 1 枚あたりの時間	48
図 32	学生における病変画像と正常画像での 1 枚あたりの時間	48
図 33	アプリケーション導入予定の UI	52

表目次

表 1	gazeNet の構造	7
表 2	TL1 の構造	16
表 3	TL1 の出力部分	17
表 4	TL2 の構造	18
表 5	TL2 の出力部分	19
表 6	DCNN の構造	20
表 7	評価指標の 2×2 マトリックス	21
表 8	Lund 2013: train、validation と test のデータの内訳	23
表 9	GazeCom: train、validation と test のデータの内訳	24
表 10	Tobii X3-120 のスペック	31
表 11	各被験者におけるパッチ画像数	33
表 12	教師なしクラスタリング: Lund 2013 の結果	36
表 13	教師なしクラスタリング: Lund 2013 の評価結果	36
表 14	教師なしクラスタリング: GazeCom の結果	37
表 15	教師なしクラスタリング: GazeCom の評価結果	37
表 16	教師あり学習: Lund 2013 の結果	38
表 17	教師あり学習: Lund 2013 の評価結果	38
表 18	教師あり学習: GazeCom の結果	39
表 19	教師あり学習: GazeCom の評価結果	39
表 20	validation: Lund 2013 で学習結果(test:GazeCom)	40
表 21	validation: Lund 2013 の評価結果(test:GazeCom)	40
表 22	validation: GazeCom で学習結果(test:Lund 2013)	40
表 23	validation: GazeCom の評価結果(test:Lund 2013)	41
表 24	各オープンデータでの gazeNet の結果と提案モデルとの比較結果	41
表 25	病変画像を対象とした評価結果	42
表 26	全ての画像を対象とした評価結果	42
表 27	病変画像を対象とした Fixation 時間	43
表 28	病変画像を対象とした Saccade 時間	43
表 29	病変画像を対象とした総読影時間	43

表 30	全ての画像を対象とした Fixation 時間.....	45
表 31	全ての画像を対象とした Saccade 時間.....	45
表 32	全ての画像を対象とした総読影時間.....	45
表 33	経験者における病変画像と正常画像での 1 枚あたりの時間	47
表 34	学生における病変画像と正常画像での 1 枚あたりの時間	47

第1章 序論

1.1 研究背景

現在、医療技術の高度化、チーム医療や医療安全への要求などから、医療従事者の知識や技術が、年々増加し、かつ高度化している。このような時代の進展の中、現役の医療従事者を含め、医療従事者を目指す学生も、必要とされる知識や技術の量が増大している。その中でも、画像診断分野に携わる医療従事者にとって、医用画像読影は必須スキルの1つである。これは、診療放射線技師は「診断価値のある検査画像を取得し、医師の診断に寄与できるかを判断すること」に関係する。医用画像読影スキルは専門性が高く、養成課程の限られた年限で習得することは容易ではない点をふまえ、本研究では、経験の浅い医療従事候補者を効率よく熟練者に近づかせるシステムの実現を目標とする。効率的なスキルの習得において、読影中の学習者に自身の読影行動をフィードバックできる教育システムの開発が効果的であるという観点から、本研究では読影スキルを解明するための手法を提案する。

読影スキルの解明には、注視時間、注視頻度など静的な視線行動に着目した研究が近年、多く行われている [1] [2]。図 1 に近年の研究と本研究での仮説の流れを示す。近年の研究で扱われている、従来から用いられていたアウトカム指標を基にする評価方法として、Receiver operating characteristic (ROC)解析 [3]がある。本研究の開発目標であるプロセスによる読影の評価では、眼球運動に着目することで ROC 解析で測れない能力を測れる可能性がある。これまでの関連研究では、単純な時間集計や画面上の特定エリアへの注視時間、注視回数などの静的な統計が用いられている。本研究では、図 1 に示す知識不足と注意不足の見逃しの違いを検出可能な教育システムの実現を目指す。本研究では動的すなわち眼球運動および視線情報を活用して、リアルタイムにフィードバックできるシステムのアプリケーション化を最終的な目標とする。

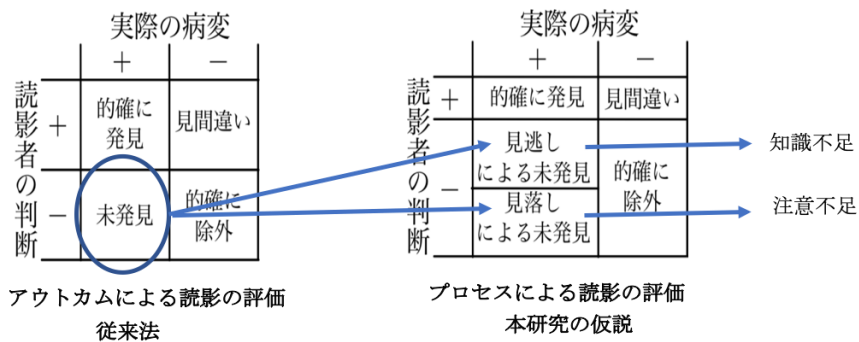


図 1 従来法と本研究の仮説

先行研究として、Laura McLaughlin らは、Radiography students と Experienced radiographers、Experienced reporting radiographers の 3 種の被験者を用いて医用画像読影の差異の統計を明らかにした [1]。評価指標は、病変部位がある画像と全ての画像を読影している際の注視時間、注視数、最初から注視するまでの時間、注視頻度、医師や放射線技師による診断の自信度(信頼度)である。得られた統計処理結果では、Experienced reporting radiographers が他の被験者に比べて有意差が出ており、診断レポートの病理説明が他のグループよりも詳細に書かれていた。その結果、経験の重要性を明らかにした。この研究では、読影頻度をヒートマップに表すことでも各グループの特徴を捉えようと試みた。

また、乳児の視線データを Deep Learning を用いて解析した Kirsten A. Dalrymple らの研究がある [4]。これは、眼球運動の中でも Fixation に着目し、乳児がどの物体をよく見ているのかを特徴抽出し、乳児の年齢を識別するモデルを作成した。

さらに 2018 年に Raimondas Zemblys らは、gazeNet といった眼球運動の識別モデルを提案した [5]。モデルは、RNN(Recurrent neural networks)をベースとしている。現在、RNN をベースとする研究は多くあるが、画像を対象とした CNN(Convolutional neural networks)を用いたもので、眼球運動の識別率の高い研究は、調べた中では行われていない。

1.1.1 Eye tracking

視線追跡は人間の行動分析に用いられており、運転操作者の研究で良く用いられる手法である。

医学の分野では、眼球運動障害の診断や読影中の視点の行動分析に用いられるが、視点を利用した手を使わないデバイスを操作するシステムも開発されている。先行研究の例を挙げると、読影中の経験者と未経験者を比較したものがあり、経験者と未経験者で関心領域をみる回数や時間などに違いがみられた [6]。

1.1.2 眼球運動

視線追跡では人間の眼球運動は以下のように定義されている。図 2 に眼球運動を表した。大きく 3 つに分けられ、Fixation(固視微動)と Saccade(跳躍運動)、Pursuit movement(追従運動)がある [6]。Fixation は、一点を注視中でも無意識に絶えず小さく動く不随意運動である。Saccade とは Fixation と Fixation の間の速度の目の動きである。非常に高速度の眼球運動であり、最高角速度 $300\sim 600^{\circ}/\text{sec}$ に及ぶ。Pursuit movement は、動いている物体を目で追う時に起こる目の動きである。

本研究では、静止画像である医用画像の読影を対象とするため、Fixation と Saccade の 2 つの眼球運動が交互に繰り返される。

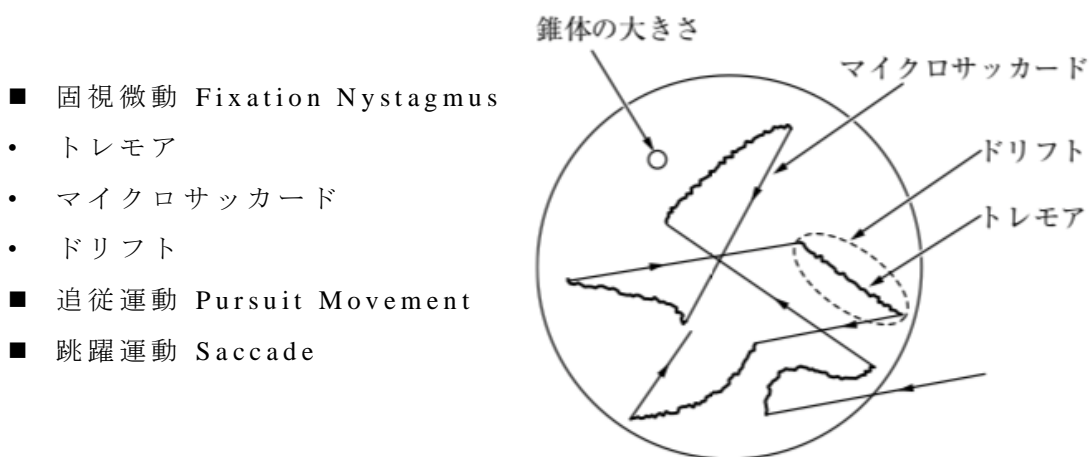


図 2 眼球運動

1.1.3 視線解析

人間の脳の機能が無侵襲に調べるには限界があるため、五感のひとつである視覚を調べることにより心理学や脳神経学の研究がなされている [7]。特に、運転操作中の人の

視線を追跡する研究はよく知られており、ドライバーの無自覚な運転について 視線から情報を得ることができる [6]。その他にも、スポーツ分野の研究や視線操作、商品パッケージの開発など視線解析は様々な分野で応用されている。視線解析ではデバイスに搭載された近赤外線カメラで眼球を撮像し、眼球運動を計測する必要がある。眼球運動研究におけるイベント検出の目標は、カメラなどの計測装置によって提供される元データから、Fixation や Saccade などを堅牢に抽出することである。このイベント検出にはこれまでに、主に 2 つのアルゴリズムが使用されてきた。1 つ目は、高速 Saccade を検出し、残りのデータを Fixation と見なす velocity をベースとしたアルゴリズムである (図 3) [8]。2 つ目は、dispersion をベースとしたアルゴリズムで、元データが特定の決められた最小期間内で、互いに十分に近いデータサンプルを見つけることで、Fixation を検出し、残りを Saccade とみなすものである [8]。

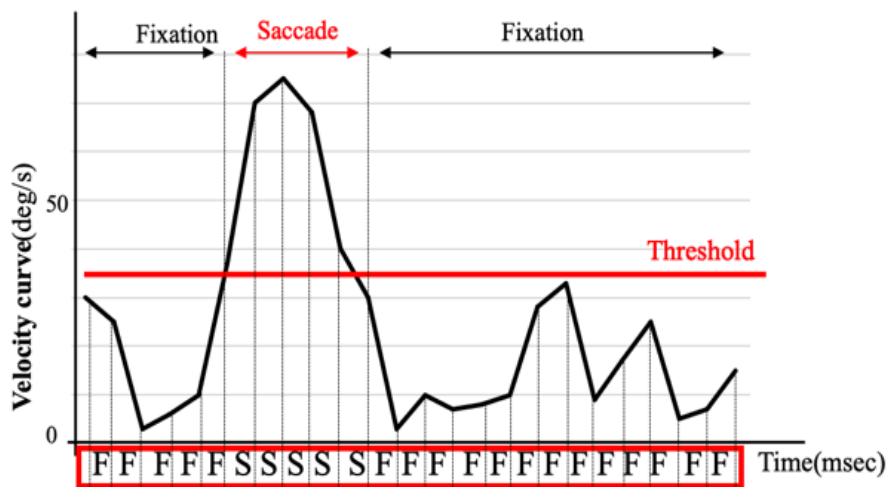


図 3 velocity 法の概念図

1.1.4 胸部 X 線画像

医用画像には、CT や MRI などのデジタル撮影装置を利用して撮影されるデジタル画像と X 線フィルムを用いるアナログ画像がある。アナログ画像は、コンピュータで解析するために A/D 変換を行う必要がある。医療における画像処理技術の本格的な応用は X 線画像で最初に行われた。複雑な濃淡構造を持つ X 線画像への応用は、自動診断やコンピュータ支援診断の研究の先駆けともなった。本研究で扱う胸部 X 線画像の一例 [9] を図 4 に示す。胸部 X 線画像の処理は、1930 年代にフィルムの感性特性などのアナロ

グ処理から始まり、コンピュータによるデジタル処理、そしてコンピュータによる自動診断を意識した研究に変遷した。今日では、間接撮影像や直接撮影法を対象とした処理手法やシステムが数多く開発されるようになった [10]。

胸部 X 線画像には、正面像と側面像がある。本研究では、正面像を用いるので、その読影方法を述べる。中央陰影の読影にあたっては、心臓や大血管の輪郭の異常に注目するとともに、気管や中枢部気管支の透亮像（とうりょうぞう）がきちんと見えているかどうかにも注意する。透亮像とは、周囲に比べて黒く写る所見のことをいう。縦隔などの中央陰影に透過性が低いと、中央陰影に重なる肺野の透過性が低下し、いわゆる低濃度部（白い）肺野の評価が十分でなくなるとともに、縦隔内部の中枢部気道の病変の評価もしにくくなる。また食道の拡張や食道腫瘍を示唆する所見が見られるかどうかなども、重要な確認すべき項目である [11]。



図 4 胸部 X 線画像の一例（正面像）

1.2 関連研究

以下に本研究で比較に用いた手法である gazeNet [5]を紹介する。図 5 に gazeNet の訓練の流れを示す。

gazeNet は、正解ラベル付きの視線追跡データで構成される比較的小さなデータセットを拡張して、最初に生成ニューラルネットワーク gazeGenNet を訓練し、合成ラベル付き視線追跡データを生成する。次に、このデータを生成ニューラルネットワークへの

入力として使用し、イベント検出器を訓練する。その後、実際に計測される眼球運動データを入力とし、視線追跡の専門家と同様の眼球運動イベントを分類する。さらに、結果のイベント検出器は、1つ1つのデータ入力ではなく、すべてのイベントデータを同時に入力することができ、どのイベントが最も可能性が高いかを一度に決定する。

gazeNet のネットワークの構造は、表 1 に示す。このネットワークには、2つの CNN 層があり、その後3つの RNN 層があり、その上に全結合層がある。CNN 層は 2×11 のサイズの 2D フィルターを使用し、実測データから捉えにくい特徴を抽出することを目的としている。一方、RNN 層はイベントシーケンスをモデル化し、Fixation、Saccade、PSO の開始とオフセットを検出する役割を果たす。従来の LSTM 層の代わりに、ゲート付きリカレントユニット (GRU) 層 [12] を使用するの、通常より少ないパラメータで同じパフォーマンスを発揮するためである。

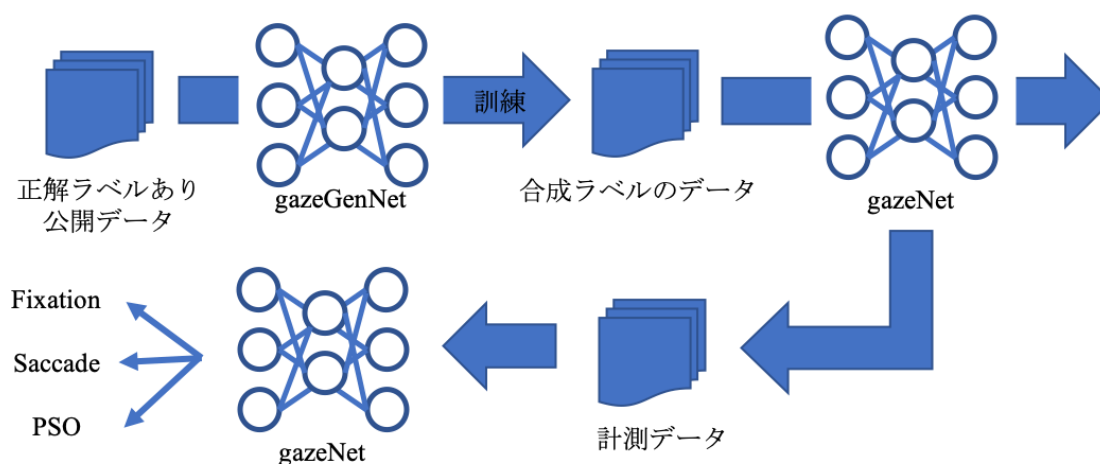


図 5 gazeNet の訓練の流れ

表 1 gazeNet の構造

CNN	Type	“same”
	Layers	2
	Filters	8
	Kernel	2×11
	Stride	[1,1]
	BatchNorm	True
	Dropout	0.25
	Activation	ReLU(clipped at 20)
RNN	Type	GRU
	Layers	3
	Neurons	64
	Bidirectional	True
	BatchNorm	True
	Dropout	0.25
	FC	Layers
	BatchNorm	False
	Bias	False
	N	3
Hyper parameter		
	Optimizer	RMSprop
	Learning rate	0.001
	Batch size	100
	T	100

gazeNet の課題として、

- ・周波数依存がある
- ・評価指標に関する課題 [13]

以上の 2 点が挙げられる。

これらの課題解決方法を以下述べる。周波数依存に関しては、視線データをパッチ画像化させることによってデータに周波数情報が混ざらないようにした。評価指標に関し

て、Friedman の論文では以下のような指摘がある。「gazeNet では、眼球運動を予測したラベルと本来の正解ラベルが一致しない事象を偽陽性または偽陰性のいずれかに分類するとした。しかし実際には、多くが真陰性に分類されている。これが原因で一致しない事象はカッパーが高くなるが、実際に検証すると一致しない事象はカッパーが低くなる」この指摘をふまえ、本研究では gazeNet で使用したサンプルエラー率 (SER) とイベントエラー率 (EER) やコーヘンのカッパーは評価指標として最適ではないと判断し、評価指標には Fixation と Saccade の識別率を用いる。また、公開データを対象として、分類率を gazeNet と比較することで提案するモデルの有用性を検討する。さらに提案モデルを用いて実測データを分類し、従来の医用画像読影のスキル比較と同様の比較を試みる。

1.3 研究目的

本研究では、CNN (convolutional neural network: CNN) を使った 3 つのモデルを用いた分析方法を提案し、モデル比較と工夫の検討によって、読影者の眼球運動をリアルタイムに分類し、フィードバック可能な手法を提案することを目的とする。

本研究では、CNN を用い、視線データの計測中に fixation と saccade を認識できる、迅速かつ正解率の高い分析法を提案する。本研究では、CNN が画像分類に適したモデルである特徴を活用し、医用画像読影中の視線座標の変化を画像として用いる。さらに線分情報 (スカラー、角度、etc) を分類するデータベース MNIST から、転移学習が可能である点を利用したモデル、およびいくつかの工夫を加えた自作の Deep CNN (DCNN) を作成し、精度を比較する。

モデルの比較では、2 つの公開データで従来研究の手法および提案する複数の手法における汎用性を確認し、実際の計測データに適用して Fixation と Saccade の分類することで、既存の医用画像読影評価の研究と比較してその実用性を検証する。

これは、我々の研究の最終目標である医用画像読影教育支援のアプリケーション開発にも繋がる。

1.4 本論文の構成

2 章では本研究で用いる手法の理論を説明し、3 章で提案手法ならびに提案モデルを述べる。4 章の基礎実験でオープンデータを用い提案手法の検証をし、応用実験で我々の実験データを用いる。5 章で各実験の結果、6 章で考察し 7 章で結論を述べる。

第2章 理論

2.1 眼球運動の研究の歴史

以下に眼球運動の研究におけるイベント検出アルゴリズムの歴史を述べる。

眼球運動の研究におけるイベント検出方法は、長い間、2つのアルゴリズムが使用されていた。1つ目は、velocityをベースとしたアルゴリズムである。最もよく知られているのは、Bahill、BrockenbroughとTroost [14]、およびSalvucciとGoldberg [15]のI-VTアルゴリズムであるが、その原理は1960年代のBoyce [16]によるアルゴリズムにまでさかのぼる。2つ目は、dispersionをベースとしたアルゴリズムである [8]。最もよく知られているのは、SalvucciとGoldberg [15]のI-DTアルゴリズムである。dispersionをベースとしたアルゴリズムの中でも、EngbertとKliegl [17]、Nyströmand Holmqvist [18]、およびMould、Foster、AmanoとOakle [19]は、パラメータに閾値を定義し、ノイズレベルが変化する場合に研究者が試行ごとに異なる閾値を設定する必要がないようにした。しかし、これらのアルゴリズムは、限られた範囲のノイズレベルでのみ機能する [20]。もう1つ最近開発されたアルゴリズムとして、Larsson、NystromとStridh [21]およびLarsson、Nystrom、AnderssonとStridh [22]による研究がある。これらの著者は、FixationとSaccadeが混在するイベントを含む、高いサンプリング周波数で記録された元データに対して、Saccade後の振動も自動検出できる新しいアルゴリズムを開発した。さらに、Hessels、Niehorster、KemnerとHooge [20]は、従来の元データだけでなく、幼児研究に特有の非常にノイズの多いデータなどのさまざまなノイズレベルのデータについてもFixationを正常に検出できる、ノイズに頑健な新しいアルゴリズムを発表した。ただし、これらの新しいアルゴリズムは、計算コストが高くなる可能性があり、調整可能なパラメータの数の増加や、エンドユーザーの経験と洞察が必要になるなどの、複雑な設定に依拠している。

一方、近年、機械学習の技術に基づいたイベント検出方法が新しく開発された。これまで、検出器は特定のアルゴリズムに対応するために開発者によって事前に設計されていたが、新しいアルゴリズムでは閾値検出と分類方法はデータ駆動で学習され、イベント検出器によって自動的に設定される。Zembls [23]は、イベント検出のための10種の機械学習アルゴリズムを比較した。それらは、サポートベクターマシン [24]、畳み込みニューラルネットワーク [25]およびランダムフォレスト [26]などである。HoppeとBulling [25]は、単純な1層の畳み込みニューラルネットワーク (CNN) に続けて、最大プーリング層と全結合層を用いることで、単純なdispersionをベースとしたアルゴリ

ズムと **velocity** をベースとしたアルゴリズムおよび PCA ベースとしたアルゴリズムよりも優れていることを示した。Zemblys ら [26]は、既存のイベント検出アルゴリズムから採用された、ランダムフォレストをベースとしたイベント検出器である IRF が、以前の最も性能のよいアルゴリズムよりも優れていることを示した。実際、SMI HiSpeed eye tracking で計測される元データに対して、分類精度は人間の専門家にほとんど達した。Zemblys ら [26]のアルゴリズムによる検出性能は 200Hz までは安定したが、それ以下のサンプリング周波数による検証はなされていない。Zemblys らによる IRF [26]および他の機械学習による方法は、ユーザーが閾値を設定する必要はなく、いったん学習すると非常に高速に計算できる。しかし、イベント検出モデルは、以前のサンプルに割り当てたイベント正解ラベルを認識せず、Post-saccadic oscillations (PSO) などの時間依存性を学習できない。

最新の研究で提案された gazeNet [5]では、入力から出力までを深層学習のアプローチで処理する。これは、生の視線計測データを入力として、Fixation、Saccade、および PSO に分類する。上記の機械学習をベースとしたアルゴリズムとは対照的に、この深層学習によるアプローチは、データからすべての機能と適切な閾値を自動的に学習できる。

深層学習のアプローチによる眼球運動データのイベント検出器の開発には、ボトルネックが伴う課題がある。アルゴリズムの学習データとして使用するには、長時間の眼球運動データに手作業で正解ラベルをつける必要がある。手作業による正解ラベルがついた学習データが少なすぎると、深いネットワークを学習しても適切なイベント検出器が得られない。学習データの準備には、視線追跡の専門家が、膨大な量のデータを注意深く調べる時間が必要である。これには非常にコストがかかり、正解ラベルづけの再現性およびラベル品質を保証できない問題がある。

2.2 k-means法

本研究では、1.3 で示した通り、クラスを Fixation と Saccade の 2 クラスにあらかじめ限定し、その分割方法を最適に探索する。まずクラス分類において代表的な非階層クラスタ分析である k-means 法について以下で述べる。k-means 法は、目的関数 J の最適化問題を解くアルゴリズムである [27]。目的関数 J の式を (2.1) に示す。

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|X_n - \mu_k\|^2 \quad (2.1)$$

各データ点 X_n に対し、対応する2値指示変数 $r_{nk} \in \{0,1\} (k=1, \dots, K)$ を定める。データ点 X_n がクラスター k に割り当てられる場合、 $r_{nk} = 1$ とし $j \neq k$ については $r_{nj} = 0$ とする。

アルゴリズムは、以下の流れである。

1. 各データ $X_n (n=1, \dots, n)$ に対してランダムにクラスターを分割する。
2. 分割したデータを元にし、各クラスターの中心 $\mu_k (k=1, \dots, k)$ を計算する。
3. 各 X_n と各 μ_k との距離を求め、 X_n を最も近い中心のクラスターに分割し直す。
4. 1~3の処理で全ての X_n のクラスターの分割が変化しない場合、収束したと見なす。
5. 4で収束しなかった場合、新しく分割した μ_k を再計算する。

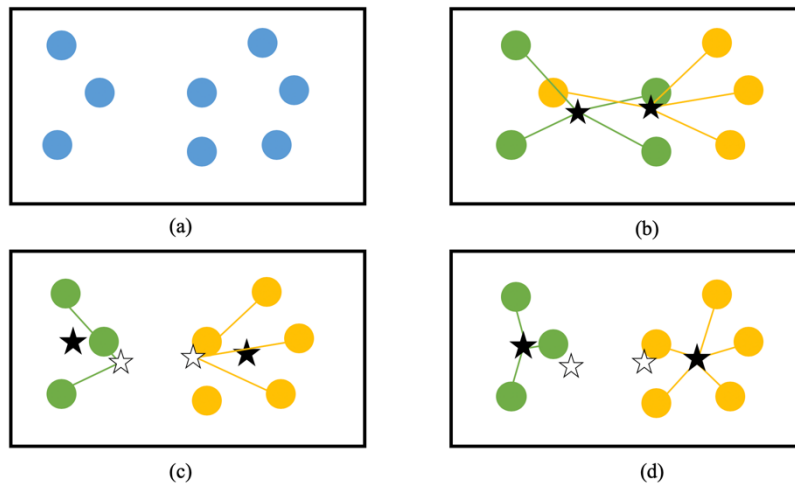


図 6 k-means 法のアルゴリズム概念図

図 6 に k-means 法のアルゴリズム概念図を示す。図の (a)→(b)→(c)→(d) のような順序でクラスターが収束していく。(b)の段階でまず各点に適当にクラスターが割り振られ、その重心が計算される。計算された重心は黒星で表している。(c)ではその重心との距離をもとにし、再度クラスターが割り当てられる。新しく計算された重心は再度黒星で表し、古い重心は白星として表した。この流れを繰り返し(d)のようにクラスター重心が移動しなくなると終了する。

2.3 CNN(Convolutional Neural Network)

CNN とは深層学習の手法の一つで最近発展がめざましい。

近年、計算機の処理能力向上により、深層学習が注目されている。深層学習はデータを学習することによってパターン認識する方法である [28] [29]。その手法の一つに畳み込みニューラルネットワーク(CNN: Convolutional Neural Network)がある。この CNN を利用することにより、人間では判断の難しい細かい特徴を抽出することが可能で、画像・音声認識などで使用されている。医用画像の分野にも応用されており、画像の識別や病変の良悪性判別などの研究が行われている ([30] [31] [32] [33] [34] [35])。ここで CNN を使用した手書き文字認識 (MNIST dataset) の例を図 7 に示す [36]。

代表的な CNN を以下紹介する。1998 年に初めて提案された CNN は、LeNet [37] である。深層学習が注目されるに至った CNN は、2012 年に提案された AlexNet [38] である。AlexNet は、2012 年に開催された大規模画像認識のコンペティション ILSVRC (ImageNet Large Scale Visual Recognition Challenge) で圧倒的な成績で優勝した。他の CNN には、畳み込み層とプーリング層から構成される VGG [39]や、VGG より層が深い GoogLeNet [40]がある。このようなネットワークを用いて、ImageNet などの大きなデータを学習させたモデルは、高い画像分類性能を達成できる。しかし、多層の CNN を用いた画像の学習には多くの時間と計算コストがかかる。その解決策として、転移学習がある。転移学習は、出力層だけを置き換えて他のタスクのために学習し直す方法である。学習すべきタスクが異なるにも関わらず、学習済みパラメータが新たなタスクに対する良い初期値となり学習がスムーズに進行することが知られている [28] [41]。

CNN を用いた病変部位の診断などの研究は数多くあるが、医用画像読影時の視線の動きを分類する研究は、現時点で知る限り同様の研究はない。この研究の成果を応用させることにより、読影教育を効率良く行うことができると考えられる。例えば、読影中の視線をリアルタイムでフィードバックするシステムを開発し、読影中の視点の動きに対し、読影方法の改善を促せると考えられる。それを実現するためにはリアルタイムで視線データを解析し、Fixation と Saccade などの眼球運動に分類することが必要である。

ここで、手書き文字の線が視点の軌跡に類似していることに注目すると、この CNN を用いた MNIST の分類手法が有効に働くのではないかと考え、本研究では MNIST により事前に学習させた重みで転移学習を試みた。

0 1 2 3 4

5 6 7 8 9

図 7 手書き数字データセットの例

第3章 手法

3.1 提案手法

本研究の実験では、サンプリング周波数 120Hz の Tobii pro X3-120 を使用する。Fixation は 100 ms にわたる連続的な視線移動を定義とするため、本研究では 12 点の連続する視線座標位置を使用する。まず前処理として、連続する視線座標から 12 点を 1 点ずつシフトしながら切り出し、11 点をオーバーラップさせる。次に 12 点からなる座標について、重心を中心としたパッチ画像を作成する。MNIST データベースの手書き数字の画像 [36] に合わせるため、パッチサイズは 28×28 画素とする。さらに視点の開始点から終点までは、256 階調のグレースケールで連続的に濃度変化させる(図 8) [42]。以下で述べる公開データの Lund 2013 と GazeCom の詳細の説明は、4.1.1 で述べる。基礎実験で対象とする公開データは Lund2013 が 500Hz で、GazeCom が 250Hz であるが、本研究での実験における 120Hz のデータ取得に合わせ、全て 120Hz 前後にダウンサンプリングする。ダウンサンプリングの方法は、移動平均を用いた。Lund 2013 の場合、前後 4 点のデータを平均し、125 Hz のデータとした。GazeCom は前後 2 点で、同様に 125 Hz とした。公開データから作成するパッチ画像のラベル付与では、12 区間の中間点である 5 または 6 点目が F であるならばその画像は Fixation、5 または 6 点目が S であるならばその画像は Saccade とする。これは、先行研究での正解ラベルは 5 種あり、12 点の場合、FFFFFFSSSS(PSO)S といった例に対処できず、ラベル作成が困難な理由による。5 または 6 点目が F/S でない場合は、Lund 2013 のデータは Post-saccadic oscillations (PSO)、Smooth pursuit、Blinks のラベルを Other とし、GazeCom のデータは Smooth pursuit、Unknown、Noise のラベルを Other として消去した。図 9 にアノテーションの仕方を示した。パッチ画像化することにより、画像の特徴抽出に優れている CNN を用いることができ、RNN(Recurrent Neural Network)ではできない MNIST などの既存のデータベースを用いた転移学習が使用できる。この前処理を行った画像を用いて、教師なしクラスタリング(k-means 法)と CNN を用いた教師あり学習に入力する (2.2 k-means 法、2.3 CNN)。

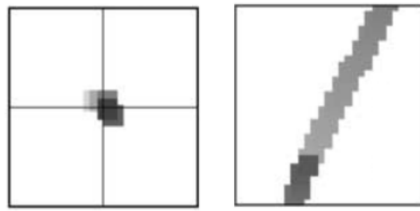


図 8 グレースケールのパッチ画像

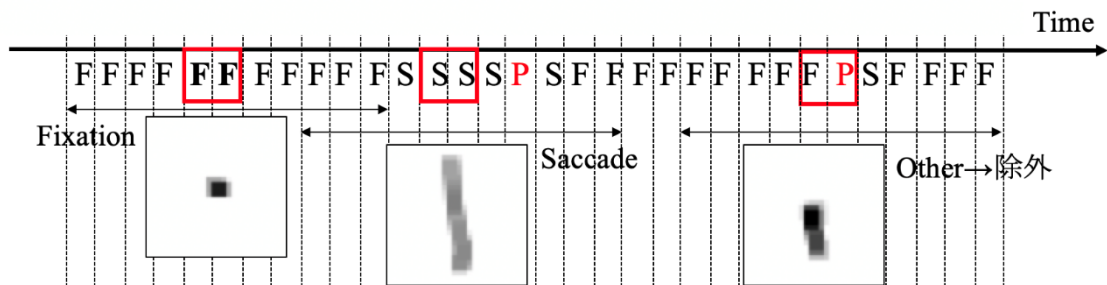


図 9 公開データのアノテーション付け

3.2 提案モデル

提案法のうち 1 つの CNN の構成は、MNIST データベースの手書き数字を分類するため、Tensorflow のチュートリアルに使用されたモデルを採用する (表 2、3 : TL1)。この CNN では、MNIST データベースの手書き数字の認識を行う CNN を転移学習に用いる。特徴抽出部分に MNIST データからの学習済みフィルタを用い、全結合層の部分だけを読影中の視線データから再学習する。TL2 は、TL1 に畳み込み層を 1 層追加したモデルである (表 4、5)。次に、DCNN は表 6 に示す。全てのモデルの Optimizer は、Adaptive moment estimation を用いた。Adam [43] は stochastic gradient descent よりも効率的にパラメータを発見できる方法である [44]。

TL1、2 と DCNN モデルの違いは、

- ① 転移学習を用いたかどうか
- ② 層の数
- ③ Dropout 層を各層に入れるか
- ④ 全結合層の前に Global Average Pooling 層を入れるかどうか

である。

以下、モデルの違いに起因するメリットを挙げる。

- ① 転移学習は実験者側の元データを増やすことなく特徴量を事前に用意できる。
- ② 層を深くすることで、正答率が向上する。
- ③ Dropout 層を入れることで過学習を防ぐ効果がある。
- ④ 学習可能なパラメータ数を減らすことができる [45]。

表 2 TL1 の構造

Layers	Image size [pixel]	Channels	Kernel size	Strides	Zero-padding	Activation function
Input	28×28	1				
Convolution 1	28×28	32	5×5	1	Same	ReLU*
Max-pooling 1	14×14	32	2×2	2	None	
Convolution 2	14×14	64	5×5	1	Same	ReLU
Max-pooling 2	7×7	64	2×2	2	None	
Dropout(rate:0.50)						
Flatten		3136				
Fully connected		64				ReLU
Fully connected		10				Softmax
Output		10				
Parameters	Learning rate : 0.01			Loss function : categorical cross entropy		
	Batch size=32, Epochs=10			Optimizer : Adaptive moment estimation		

ReLU*:Rectified Linear Unit

表 3 TL1 の出力部分

Layers	Image size [pixel]	Channels	Kernel size	Strides	Zero-padding	Activation function
Input		1				
Fully connected		64				ReLU
Fully connected		2				Softmax
Output		2				
Parameters	Learning rate : 0.01 Batch size=32, Epochs=50			Loss function : categorical cross entropy Optimizer : Adaptive moment estimation		

表 4 TL2 の構造

Layers	Image size [pixel]	Channels	Kernel size	Strides	Zero-padding	Activation function
Input	28×28	1				
Convolution 1	26×26	32	3×3	1	None	ReLU*
Max-pooling 1	13×13	32	2×2	2	None	
Convolution 2	11×11	64	3×3	1	None	ReLU
Max-pooling 2	5×5	64	2×2	2	None	
Convolution 3	3×3	64	3×3	1	None	ReLU
Flatten		1024				
Fully connected		64				ReLU
Fully connected		10				Softmax
Output		10				
Parameters	Learning rate : 0.01 Batch size=32, Epochs=10			Loss function : categorical cross entropy Optimizer : Adaptive moment estimation		

表 5 TL2 の出力部分

Layers	Image size [pixel]	Channels	Kernel size	Strides	Zero-padding	Activation function
Input		1				
Fully connected		64				ReLU
Fully connected		2				Softmax
Output		2				
Parameters	Learning rate : 0.01 Batch size=32, Epochs=50			Loss function : categorical cross entropy Optimizer : Adaptive moment estimation		

表 6 DCNN の構造

Layers	Image size [pixel]	Channels	Kernel size	Strides	Zero-padding	Activation function
Input	28×28	1				
Convolution 1	28×28	32	3×3	1	Same	ReLU
Max-pooling 1	14×14	32	2×2	2	Same	
Dropout(rate:0.20)						
Convolution 2	14×14	64	3×3	1	Same	ReLU
Max-pooling 2	7×7	64	2×2	2	Same	
Dropout(rate:0.20)						
Convolution 3	7×7	128	3×3	1	Same	ReLU
Max-pooling 3	3×3	128	2×2	2	Same	
Dropout(rate:0.20)						
Global-average-pooling		128				
Fully connected		128				ReLU
Dropout(rate:0.25)						
Fully connected		2				Softmax
Output		2				
Parameters	Learning rate : 0.01			Loss function : categorical cross entropy		
	Batch size=32, Epochs=50			Optimizer : Adaptive moment estimation		

3.3 評価指標の算出

評価指標は、正解率、Fixation を Fixation と正しく判定した確率(感度)、Saccade を Saccade と正しく判定した確率(特異度)とした。表 7 にもとづいたそれぞれの計算式は (3.1)、(3.2)、(3.3) に示す。

表 7 評価指標の 2×2 マトリックス

		Actual label		Total
		Fixation	Saccade	
提案 model	Fixation	a	b	a+b
	Saccade	c	d	c+d
Total		a+c	b+d	a+b+c+d

$$\text{正解率} = \frac{a + d}{a + b + c + d} \quad (3.1)$$

$$\text{感度} = \frac{a}{a + c} \quad (3.2)$$

$$\text{特異度} = \frac{d}{b + d} \quad (3.3)$$

3.4 応用実験での統計手法

Laura McLaughlin らの先行研究 [1]を参考とし、応用実験での評価指標は、5年以上の臨床経験を有する診療放射線技師（以下経験者）と診療放射線技師養成校の学生（以下学生）の病変画像のみ 15 枚に対する注視時間、および全ての画像 30 枚に対する注視時間で有意差検定した。注視数は、実験環境や条件が違うので使用しなかった。統計手法は、先行研究では、3 グループ以上だったため、一元配置分散分析(A one way analysis of variance : ANOVA)を用いていた。しかし、本研究では経験者と学生の 2 グループなので、t 検定を用いる。

第4章 実験

4.1 基礎実験

我々の提案したモデルの汎用性を 2 つの正解ラベル付き(眼球運動の分類ができてい
る)公開データを用いて実験する。また、先行研究である gazeNet [5]の判別結果につい
て、サンプルエラー率 (SER) とイベントエラー率 (EER) やコーヘンのカッパーでは
なく正解率に置き換えた結果や、control の手法として velocity 法と dispersion 法と比
較検証し、考察する。実験で用いる手法は、教師なしクラスタリングの手法である k-
means 法 (2.1 k-means 法) と CNN を用いた教師あり学習 (2.2 CNN) の 2 つである。
教師あり学習では 3 つのモデル (3.2 提案モデル) を用いてそれぞれ比較検討する。

4.1.1 データセット

基礎実験で用いた 2 つの公開データについて以下で説明する。公開データの名称は、
Lund 2013 [21] と GazeCom [46] [47] である。

4.1.1.1 Lund 2013

gazeNet [5]で用いられた公開データ Lund 2013 を、提案モデルの入力のひとつとす
る。このデータは、Lund 大学ヒューマニティーズラボの手動正解ラベル付き視線計測
データである。これは、画像、ビデオ、および移動する点を見た際の、参加者単眼の眼
球運動データで構成されている。すべての参加者の眼球運動は、500 Hz のサンプリング
周波数で SMI Hi-Speed 1250 アイトラッカーで記録された。2 人の眼球運動の専門家
(MN、RA) が、手動でデータを Fixation、Saccade、Post-saccadic oscillations (PSO)、
Smooth pursuit、Blinks、および Undefined events にラベル付与した。この実験で対
象となる静止画像は 20 枚であり、1 枚あたりの静止画像読影時間は 4~10 秒である。
データセットの 20 枚のうち 14 枚は、MN と RA によって個別にラベルを付与された。
本研究における基礎実験では、この 2 名共がラベリングした 14 枚を用いる。実験での
train、validation および test のデータは、表 8 に示す内訳である。内訳の名称は、Lund
2013 内にあるデータフォルダ名による。データの正解ラベルは、本研究では Fixation
と Saccade の 2 値分類問題であるため、Post-saccadic oscillations (PSO)、Smooth
pursuit、Blinks の 3 つはデータから消去した。データの消去方法は 3.1 の通りである。
図 10 は、Lund 2013 のデータセットの中に含まれる視線の軌跡画像の一例である。

Lund 2013 のデータセットの視線軌跡を本研究の手法を用いてパッチ画像にした結果、パッチ画像数は training が 20,990 枚(validation が 3,504 枚)、test が 3,888 枚である。

4.1.1.2 GazeCom

GazeCom データセットは、250 Hz で SR Research EyeLink II で記録された 4.5 時間以上の手動ラベル付き視線計測データである。手動の正解ラベル (Fixation、Saccade、Smooth pursuit、Unknown、Noise) は、2 人の視線追跡の専門家が付与した。

本研究での train、validation と test のデータの内訳は、表 9 に示す。内訳の名称は、GazeCom 内にあるデータフォルダによる。データの正解ラベルは、前節と同様に Fixation と Saccade の 2 値分類問題であるため、Smooth pursuit、Unknown、Noise の 3 つはデータから消去した。データの消去方法は Lund 2013 と同様である。

GazeCom のデータセットの視線軌跡を本研究の手法を用いてパッチ画像にした結果、画像数は training が 11,159 枚(validation が 1,856 枚)、test が 2,174 枚である。

表 9 GazeCom: train、validation と test のデータの内訳

trainSet
AAF_data
AAW_data
APS_data
C 1 K_data
CCB_data
validationSet
CCE_data
testSet
CCF_data

4.1.2 教師なしクラスタリング(k-means法)

3.1 に基づいて前処理した画像に対し、Fixation と Saccade の 2 クラスのクラスタリングを行うために、まず主成分分析(PCA: Principal Component Analysis)を用い、全データを 2 次元まで次元圧縮をした後、k-means 法を用いてクラスタリングを行う。ライ

ブラリーには、Python の scikit-learn を用いた。計算環境は、OS:Ubuntu 16.04.5 LTS、CPU : Intel(R) Xeon(R) CPU E5-1650 v4 @ 3.60GHz、Memory : 32GB である。

4.1.3 教師あり学習(CNN)

前節と同様に、前処理した画像を提案した 3 つのモデル (3.2 提案モデル) に入力し学習させた。全てのモデルで、学習のパラメータとして、Learning rate は 0.01 に設定し、50 epoch 分ネットワークを学習する。バッチサイズは 32 とした。また、Lund2013 と Gazecom とで交差して検証した。検証は、Lund 2013 で学習したものを GazeCom でテストしたものと、GazeCom で学習し Lund 2013 でテストしたもので実施した。計算環境は、OS:Ubuntu 16.04.5 LTS、CPU: Intel(R) Xeon(R) CPU E5-1650 v4 @ 3.60GHz、Memory : 32GB、GPU : NVIDIA GeForce GTX 1080 である。ライブラリー環境は、Tensorflow 1.10.0 (Google Inc.)、Anaconda 1.7.2、Python 3.6.6、Keras 2.2.4 である。

4.2 応用実験

4.2.1 データセット

被験者は、the standard digital image database created by the Japanese Society of Radiological Technology による 30 枚の胸部 X 線画像を読影した [9]。胸部 X 線画像は、病変部位を含む画像を 15 枚、正常画像を 15 枚使用しデータセットとした。図 11 に胸部 X 線画像を読影している際の視線の軌跡の一例であり、図 12 に 12 点パッチ画像の一例を示す。

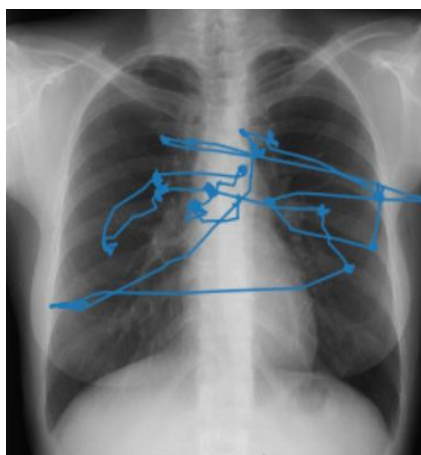


図 11 胸部 X 線画像を見ている際の視線の軌跡の一例

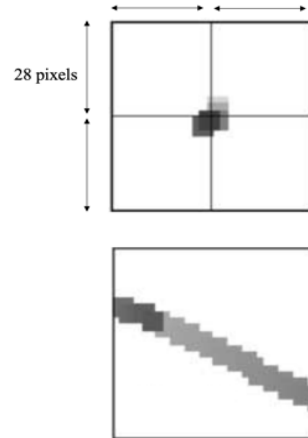


図 12 12点パッチ画像一例

実験の被験者のレベルの一例を従来法である ROC、病変位置の正解を考慮した Location ROC (LOC) 解析で表した物を図 13、15 に示す。また詳細に解析した図を 14、16 に表す。これは、病変部位が正しくマークしたかを詳細に示している。

Classification Table

Prediction	Actual	NP	
	Lesion		
Positive	TP = 11	FP = 8	19
Negative	FN = 4	TN = 7	11
	15	15	N=30

Sensitivity = $11 / 15 = 0.733$

Specificity = $7 / 15 = 0.467$

Positive Predictive Value = $11 / 19 = 0.579$

Negative Predictive Value = $7 / 11 = 0.636$

Accuracy = $(11 + 7) / 30 = 0.600$

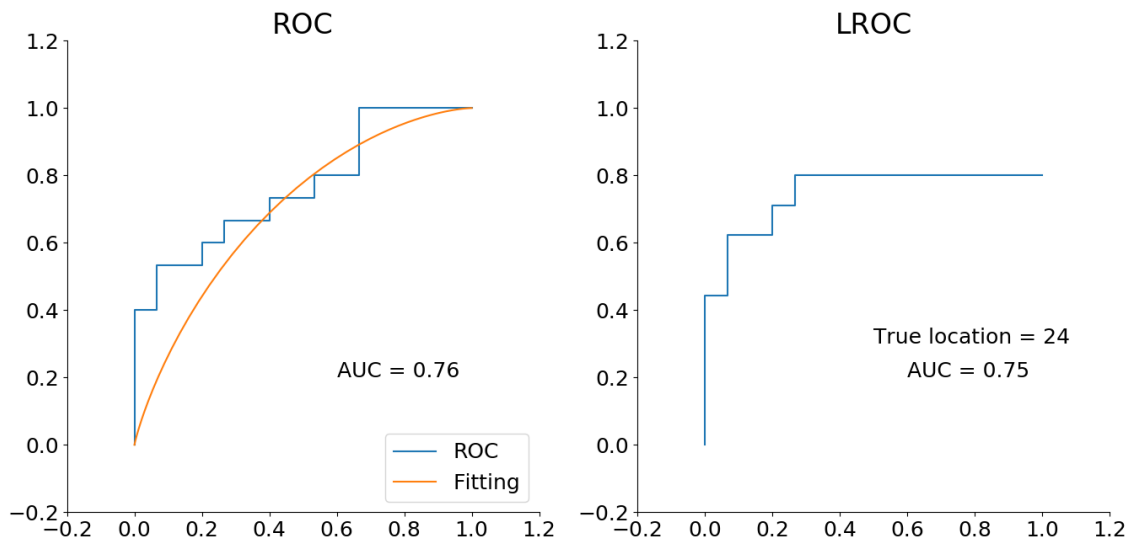


図 13 経験者の ROC、LOC 解析結果

Classification Table

Prediction	Actual Lesion	NP	
Positive	TP = 11	FP = 8	19
Negative	FN = 4	TN = 7	11
	15	15	N=30

Sensitivity = $11 / 15 = 0.733$

Specificity = $7 / 15 = 0.467$

Positive Predictive Value = $11 / 19 = 0.579$

Negative Predictive Value = $7 / 11 = 0.636$

Accuracy = $(11 + 7) / 30 = 0.600$

Interpretation Errors

Prediction	Actual Lesion	NP	Total
Positive	TP = 11 False location = 2	FP = 8 Scanning = 0 Recognition = 4 Decision-Making = 4	19
Negative	FN = 4 Scanning = 1 Recognition = 0 Decision-Making = 3	TN = 7	11
Total	15	15	N=30

図 14 経験者の指摘箇所の詳細結果

Classification Table

Prediction	Actual		NP
	Lesion		
Positive	TP = 9	FP = 5	14
Negative	FN = 6	TN = 10	16
	15	15	N=30

Sensitivity = $9 / 15 = 0.600$
 Specificity = $10 / 15 = 0.667$

Positive Predictive Value = $9 / 14 = 0.643$
 Negative Predictive Value = $10 / 16 = 0.625$

Accuracy = $(9 + 10) / 30 = 0.633$

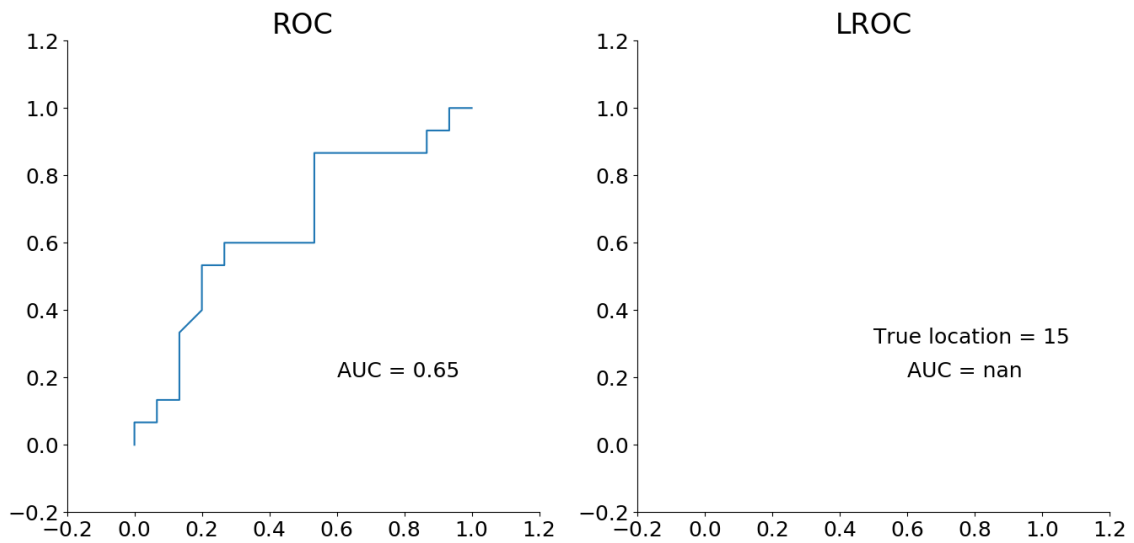


図 15 学生の ROC、LOC 解析結果

Classification Table

Prediction	Actual Lesion	NP	
Positive	TP = 9	FP = 5	14
Negative	FN = 6	TN = 10	16
	15	15	N=30

Sensitivity = $9 / 15 = 0.600$

Specificity = $10 / 15 = 0.667$

Positive Predictive Value = $9 / 14 = 0.643$

Negative Predictive Value = $10 / 16 = 0.625$

Accuracy = $(9 + 10) / 30 = 0.633$

Interpretation Errors

Prediction	Actual Lesion	NP	Total
Positive	TP = 9 False location = 9	FP = 5 Scanning = 0 Recognition = 4 Decision-Making = 1	14
Negative	FN = 6 Scanning = 6 Recognition = 0 Decision-Making = 0	TN = 10	16
Total	15	15	N=30

図 16 学生の指摘箇所の詳細結果

4.2.2 視線追跡装置

本研究では赤外線追跡装置を用いて読影時の視線を追跡し、視点データを取得した。装置は Tobii pro X-120 [48]を用いた。全体図は図 17 に示し、スペックは表 10 に示す。



図 17 Tobii X3-120 全体像

表 10 Tobii X3-120 のスペック

サンプリングレート	120 Hz
瞳孔検出能力 (Trackability)	97 %
ロバスト性	92 %
正確度(Accuracy)	0.4°
精密度(Precision)	0.24°
許容頭部振れ幅	左右 50cm×上下 40cm(19.7×15.7 インチ) 瞳孔検出距離 80cm の場合、50cm~90cm
遅延時間	トータルシステム遅延：11ms 以下
リカバリー時間	瞬き：瞬時 トラッキング・ロスの後：100ms 以下
推奨スクリーンサイズ	25 インチまで(16:9)

4.2.3 実験環境

実験環境を図 18 に示す。実験室内は、医療用高精細モニターを背にして蛍光灯のみを使用する。モニターの画面は垂直にし、人物の目との距離を約 700 mm とする。Tobii pro X3-120 とモニターをおいている台との角度は 67° とする。実際の実験風景は、図 19 に示す。



室内光はPCを背にし、
蛍光灯のみ使用

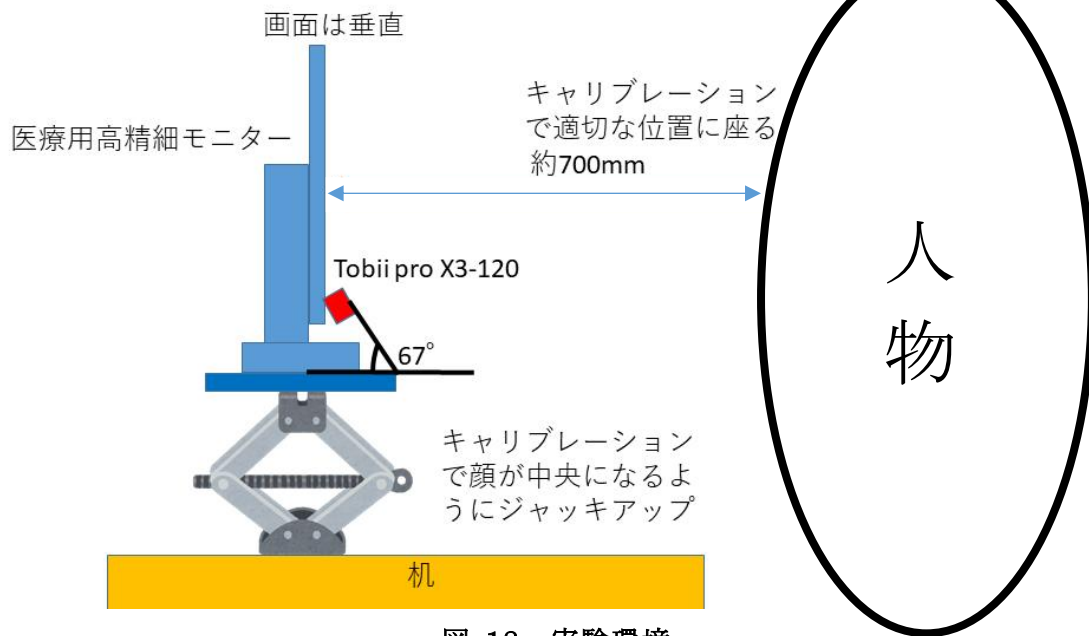


図 18 実験環境

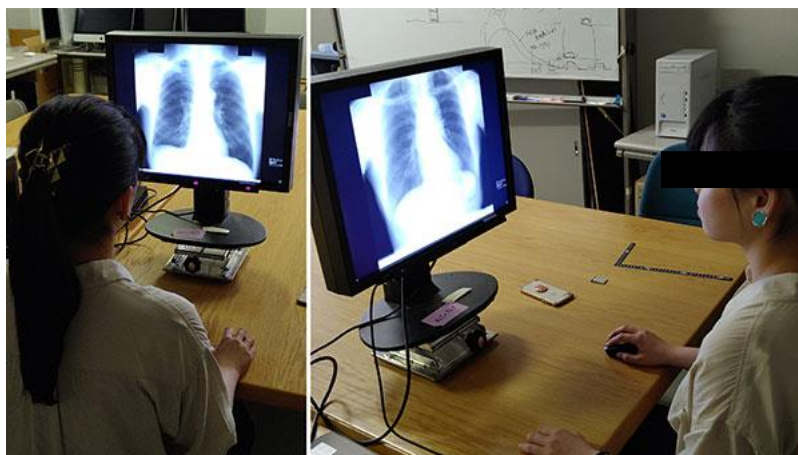


図 19 実験風景

4.2.4 実験(モデル検証)

被験者は、5年以上の臨床経験を有する診療放射線技師2名、診療放射線技師養成校の学生2名の計4名である。1枚あたりの検索時間は最大30秒に制限した。本実験の眼球運動の正解ラベルは、4.1で学習させたDCNNモデルを用い、FixationとSaccadeを分類した。これは、gazeNetがLund 2013を学習させてモデルを作成していたため、本実験でもLund 2013で学習させた。本研究で扱う実験データは、群馬県立県民健康科学大学の倫理審査委員会で承認され、参加者に同意を得て研究を実施した。なお実験データは匿名化して扱っている。実験で得られた視線軌跡データから前処理を行なったデータのパッチ画像数を以下の表11に示す。

表 11 各被験者におけるパッチ画像数

被験者	画像枚数[枚]
経験者 1	43,830
経験者 2	57,568
学生 1	27,508
学生 2	33,163

第5章 結果

5.1 基礎実験

5.1.1 教師なしクラスタリング(k-means法)

図 20、22 に PCA の結果を示す。横軸が第一主成分軸で、縦軸が第二主成分軸である。図の特徴として、原点より左が Fixation、右が Saccade となった。PCA の前処理結果に、k-means 法で分類した結果を図 21、23 に示す。青点は Saccade、赤点は Fixation を表す。黒の×マークがそれぞれのクラスターのセントロイドである。図 21、23 のセントロイドに着目すると、2 つのセントロイド間の距離は重なっておらず、Fixation と Saccade の 2 つにおおむね分類できている。公開データで専門家がラベル付けした結果 (4.1.1 データセット) を正解ラベルとし、Lund 2013 と GazeCom のデータを分類した結果のコンフュージョンマトリックスを表 12、表 14 に示す。表から、コンフュージョンマトリックス上においても Fixation と Saccade が判別できている。表 12、14 から感度、特異度、正解率を算出した結果をそれぞれ表 13、15 に示す。Lund 2013 の正解率は 88.4 %であり、GazeCom の正解率は 89.7 %であった。感度は、それぞれ 88.6 %と 92.3 %であり、特異度は、それぞれ 86.6 %と 75.3 %であった。

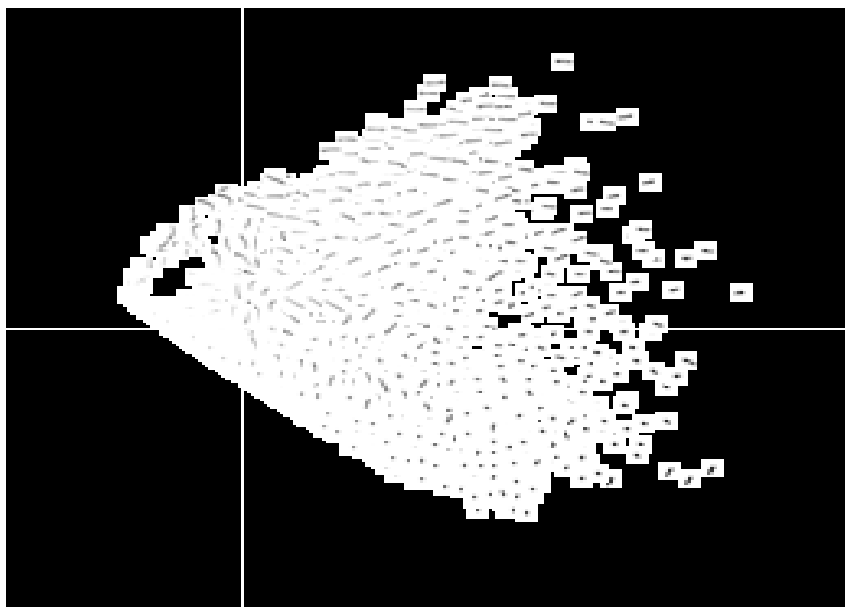


図 20 教師なしクラスタリング: Lund 2013 の PCA の結果

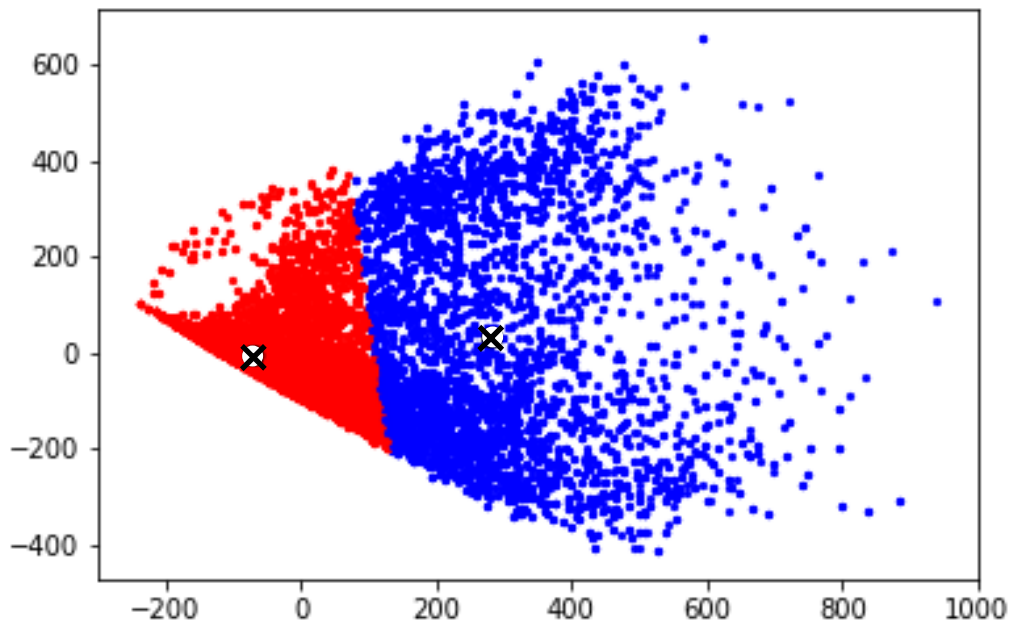


図 21 教師なしクラスタリング: Lund 2013 の k-means 法結果

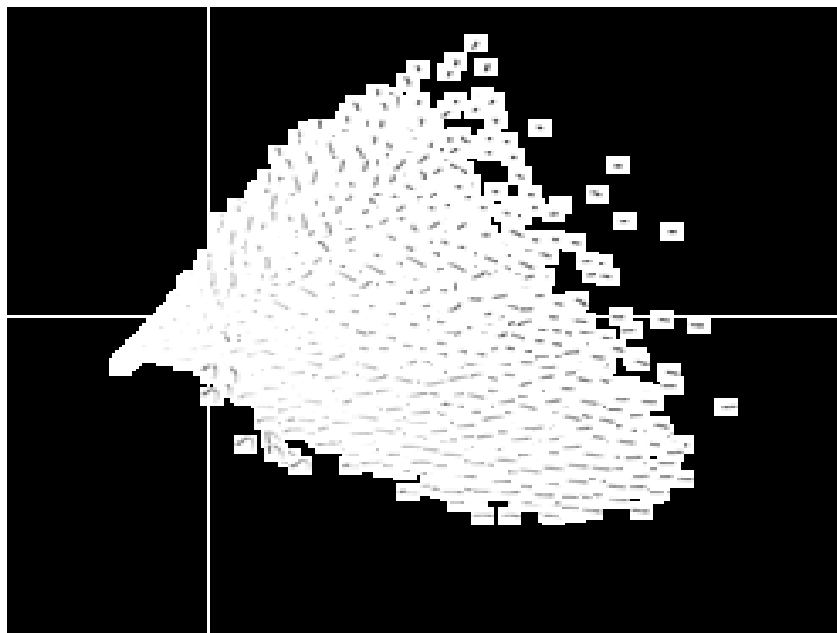


図 22 教師なしクラスタリング: GazeCom の PCA 結果

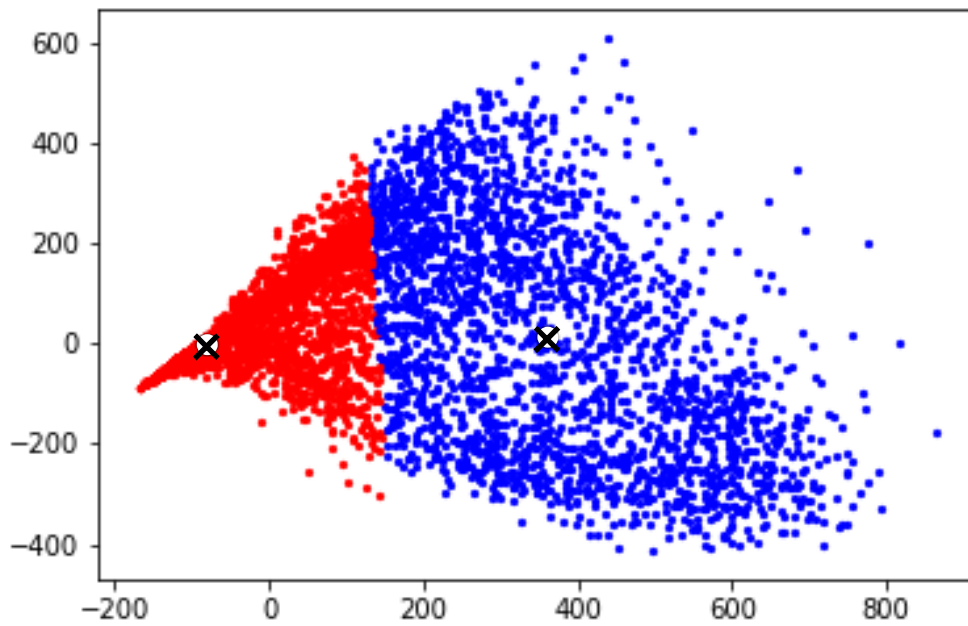


図 23 教師なしクラスタリング: GazeCom の k-means 法結果

表 12 教師なしクラスタリング: Lund 2013 の結果

		Actual label		Total
		Fixation	Saccade	
Lund 2013	Fixation	19,643	380	20,023
	Saccade	2,518	2,458	4,976
Total		22,161	2,838	24,999

表 13 教師なしクラスタリング: Lund 2013 の評価結果

	k-means 法[%]
Sensitivity(Fixation 識別率)	88.6
Specificity(Saccade 識別率)	86.6
Accuracy	88.4

表 14 教師なしクラスタリング: GazeCom の結果

		Actual label		Total
		Fixation	Saccade	
GazeCom	Fixation	10,113	499	10,612
	Saccade	839	1,523	2,362
Total		10,952	2,022	12,974

表 15 教師なしクラスタリング: GazeCom の評価結果

k-means 法[%]	
Sensitivity(Fixation 識別率)	92.3
Specificity(Saccade 識別率)	75.3
Accuracy	89.7

5.1.2 教師あり学習(CNN)

表 16、18 に TL1、2、DCNN model それぞれで公開データの test set を分類した結果のコンフュージョンマトリックスを示す。表 16、18 から感度、特異度、正解率を算出した結果をそれぞれ表 17、19 に示す。

Lund 2013 のデータセットについて、MNIST データベースを用いた転移学習の正答率は TL1 で 95.4 %、TL2 で 95.7 % であり、DCNN model では正答率は 96.1 % であった。感度は、それぞれ 97.0 %、97.3 % と 97.7 % であり、特異度は、それぞれ 85.6 %、85.1 % と 86.4 % であった。3 つのモデルを比較すると、DCNN model がどの評価指標でも高かった。

次に GazeCom のデータセットでは、TL1 で 97.2 %、TL2 で 97.2 % であり、DCNN model では正答率は 97.6[%] であった。感度は、それぞれ 97.9 %、97.9 % と 98.2 % であり、特異度は、それぞれ 91.9 %、92.2 % と 93.8 % であった。3 つのモデルを比較すると、こちらも Lund 2013 と同様で DCNN model がどの評価指標でも高くなった。

表 16 教師あり学習: Lund 2013 の結果

		Actual label		Total
		Fixation	Saccade	
TL 1 model	Fixation	3,258	76	3,334
	Saccade	101	453	554
TL 2 model	Fixation	3,269	79	3,348
	Saccade	90	450	540
DCNN model	Fixation	3,281	72	3,353
	Saccade	78	457	535
Total		3,359	529	3,888

表 17 教師あり学習: Lund 2013 の評価結果

	TL 1 model[%]	TL 2 model[%]	DCNN model[%]
Sensitivity(Fixation 識別率)	97.0	97.3	97.7
Specificity(Saccade 識別率)	85.6	85.1	86.4
Accuracy	95.4	95.7	96.1

表 18 教師あり学習: GazeCom の結果

		Actual label		Total
		Fixation	Saccade	
TL 1 model	Fixation	1,863	22	1,885
	Saccade	39	250	289
TL 2 model	Fixation	1,862	21	1,883
	Saccade	40	251	291
DCNN model	Fixation	1,867	17	1,884
	Saccade	35	255	290
Total		1,902	272	2,174

表 19 教師あり学習: GazeCom の評価結果

	TL 1 model[%]	TL 2 model[%]	DCNN model[%]
Sensitivity(Fixation 識別率)	97.9	97.9	98.2
Specificity(Saccade 識別率)	91.9	92.2	93.8
Accuracy	97.2	97.2	97.6

2つの公開データを交差した検証の結果を表 20、21、22、23 に示す。検証は、Lund 2013 で学習し GazeCom でテストした結果および、GazeCom で学習し Lund 2013 でテストした結果の 2 つとした。正解率は、全て約 90 %を超えた。しかし、Lund 2013 で学習し GazeCom でテストした結果では、特異度が 75%を超えなかった。これは、gazeNet の 75.7 %に劣る。一方、感度は 97 %を超えた。これは、gazeNet の 84.0 %よりも高くなった。gazeNet では、GazeCom で学習し Lund 2013 でテストした結果がないため比較できないが、この場合においても感度と特異度とも 85~90 %と比較的高くなった。

表 20 validation: Lund 2013 で学習結果(test:GazeCom)

		Actual label		Total
		Fixation	Saccade	
TL 1 model	Fixation	1,857	73	1,930
	Saccade	45	199	244
TL 2 model	Fixation	1,861	84	1,945
	Saccade	41	188	229
DCNN model	Fixation	1,852	69	1,921
	Saccade	50	203	253
Total		1,902	272	2,174

表 21 validation: Lund 2013 の評価結果(test:GazeCom)

	TL 1 model[%]	TL 2 model[%]	DCNN model[%]
Sensitivity(Fixation 識別率)	97.6	97.8	97.4
Specificity(Saccade 識別率)	73.2	69.1	74.6
Accuracy	94.6	94.3	94.5

表 22 validation: GazeCom で学習結果(test:Lund 2013)

		Actual label		Total
		Fixation	Saccade	
TL 1 model	Fixation	3,042	78	3,120
	Saccade	317	451	768
TL 2 model	Fixation	3,019	71	3,090
	Saccade	340	458	798
DCNN model	Fixation	3,027	62	3,089
	Saccade	332	467	799
Total		3,359	529	3,888

表 23 validation: GazeCom の評価結果(test:Lund 2013)

	TL 1 model[%]	TL 2 model[%]	DCNN model[%]
Sensitivity(Fixation 識別率)	90.6	89.9	90.1
Specificity(Saccade 識別率)	85.3	86.6	88.3
Accuracy	89.8	89.4	89.9

次に、各公開データに対する gazeNet の結果と、本研究による提案モデルとで、感度と特異度の比較を行う。gazeNet の論文で使用している評価指標であるサンプルエラー率 (SER) とイベントエラー率 (EER) やコーヘンのカッパーを指摘した論文 [13] をもとに、本研究では感度と特異度で他のアルゴリズムと比較した結果の一覧を表 24 に示す。

表 24 各オープンデータでの gazeNet の結果と提案モデルとの比較結果

Dataset	Algorithm	Fixation 識別率[%]	Saccade 識別率[%]
Lund 2013	TL 1 model	97.0	85.6
	TL 2 model	97.3	85.1
	DCNN model	<u>97.7</u>	<u>86.4</u>
	gazeNet	98.6	83.0
	Velocity	90.4	99.2
	Dispersion	90.5	15.5
GazeCom	TL 1 model	<u>97.9</u>	91.9
	TL 2 model	<u>97.9</u>	<u>92.2</u>
	DCNN model	98.2	93.8
	gazeNet	84.0	75.7
	Velocity	85.8	14.0
	Dispersion	88.3	21.1

表 24 より、我々の提案手法と gazeNet を比較したところ、3 つのモデル全てにおいて感度は 95 %、特異度は 85 % を上回り、公開データに対して汎用性があった。提案モ

デルの中では、DCNN model の精度が最も高いため、以降で述べる応用実験で用いるのは、DCNN model とした。

5.2 応用実験

提案する DCNN モデルで、我々が実測した視線データの分類結果を以下に示す。表 25 には病変画像のみ(15 枚)の結果を示す。表 26 は画像全て(30 枚)の結果である。図 24 には分類した画像を示す。表 25、26 から病変画像のみと画像全てともに経験者は学生よりも Fixation 時間および総読影時間が長かった。

表 25 病変画像を対象とした評価結果

被験者	Fixation の枚数	Fixation 時間 [sec]	Saccade の枚数	Saccade 時間 [sec]	総読影時間 [sec]
経験者 1	15,029	125.24	2,297	19.14	144.38
経験者 2	20,974	174.78	6,794	56.62	231.40
学生 1	6,384	53.20	8,429	70.24	123.44
学生 2	8,544	71.20	4,955	41.29	112.49

表 26 全ての画像を対象とした評価結果

被験者	Fixation の枚数	Fixation 時間 [sec]	Saccade の枚数	Saccade 時間 [sec]	総読影時間 [sec]
経験者 1	37,254	310.45	6,576	54.80	365.25
経験者 2	43,677	363.98	13,891	115.76	479.73
学生 1	14,175	118.13	18,988	158.23	276.36
学生 2	17,474	229.23	10,034	83.62	229.23

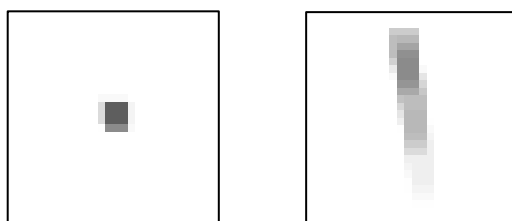


図 24 予測結果(左 : Fixation、右 : Saccade)

経験者と学生の読影行動には差があるという仮説を立て、t 検定した結果を表 27～32 に示す。1 枚あたりの時間は $1/120[\text{Hz}]=0.0083[\text{s}]$ であるので、枚数 $\times 0.0083$ としてそれぞれの時間を算出した。また t 検定の結果をグラフ化したものを図 25～30 に示す。p 値は病変画像 15 枚の対象では Fixation 時間で 0.0790(>0.05)、Saccade 時間で 0.529(>0.05)、総読影時間で 0.252(>0.05)となり、全てにおいて有意水準 5 %で統計的に有意な差が見られなかった。しかし、Fixation 時間に関しては片側検定を行うと 0.040(<0.05)となり有意水準 5 %で統計的に有意であった。全ての画像 30 枚を対象にした検定では、Fixation 時間での p 値が 0.0208(<0.05)となり有意水準 5 %で統計的に有意な差がみとめられた。Saccade 時間の p 値は 0.536(>0.05)、総読影時間は 0.111(>0.05)であり有意水準 5 %では統計的に有意な差が見られなかった。

したがって、全ての画像を対象とした Fixation 時間に限って経験者と学生との有意な差がみとめられた。

表 27 病変画像を対象とした Fixation 時間

グループ	標本数	平均[sec]	標準偏差
経験者	2	150.01	24.77
学生	2	62.20	9

表 28 病変画像を対象とした Saccade 時間

グループ	標本数	平均[sec]	標準偏差
経験者	2	37.88	18.74
学生	2	55.77	14.48

表 29 病変画像を対象とした総読影時間

グループ	標本数	平均[sec]	標準偏差
経験者	2	187.89	43.51
学生	2	117.97	5.48

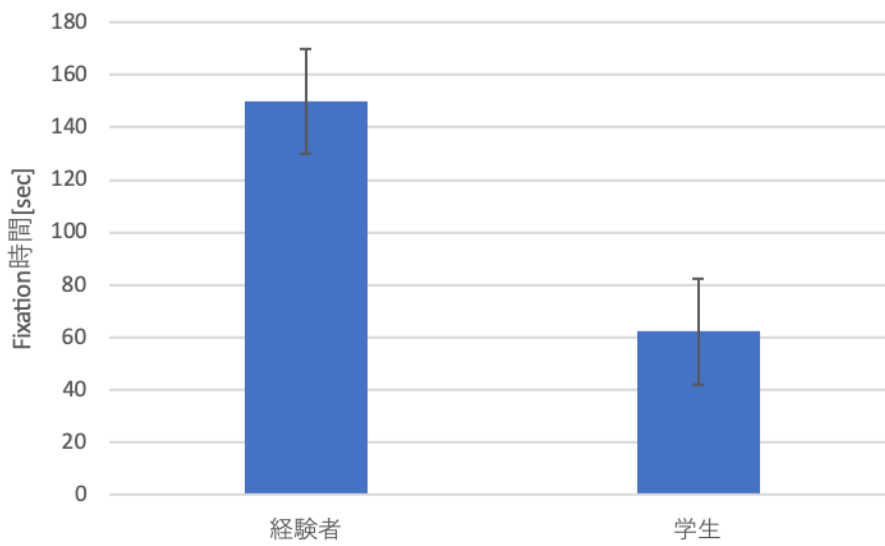


図 25 病変画像を対象とした Fixation 時間

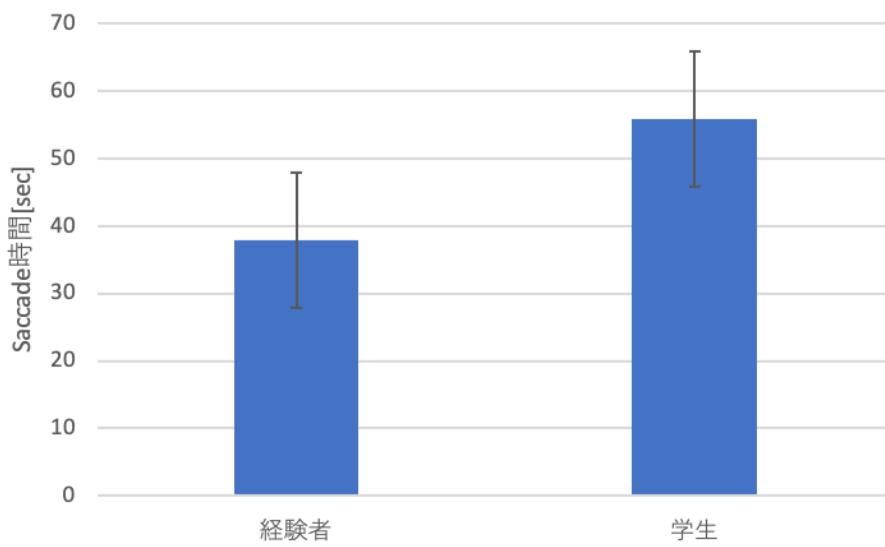


図 26 病変画像を対象とした Saccade 時間

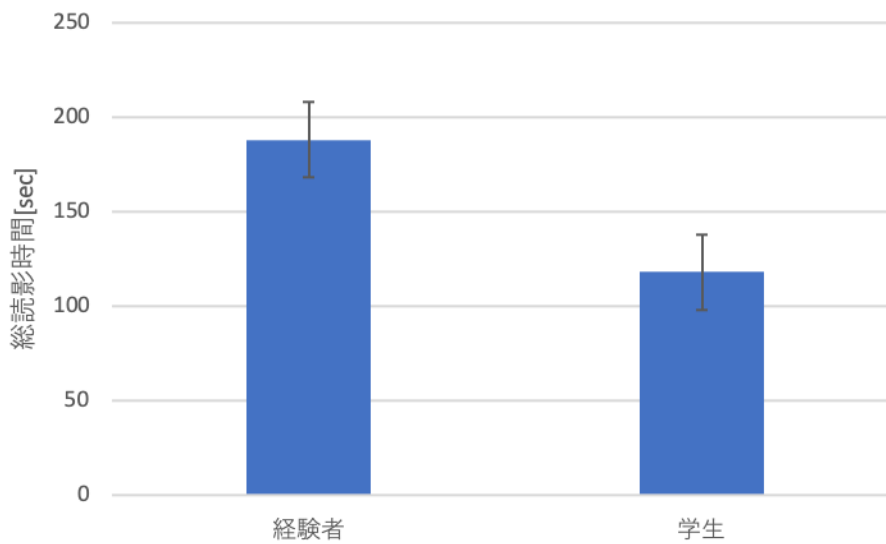


図 27 病変画像を対象とした総読影時間

表 30 全ての画像を対象とした Fixation 時間

グループ	標本数	平均[sec]	標準偏差
経験者	2	337.21	26.76
学生	2	131.87	13.75

表 31 全ての画像を対象とした Saccade 時間

グループ	標本数	平均[sec]	標準偏差
経験者	2	85.28	30.48
学生	2	120.93	37.31

表 32 全ての画像を対象とした総読影時間

グループ	標本数	平均[sec]	標準偏差
経験者	2	422.49	57.24
学生	2	252.80	23.57

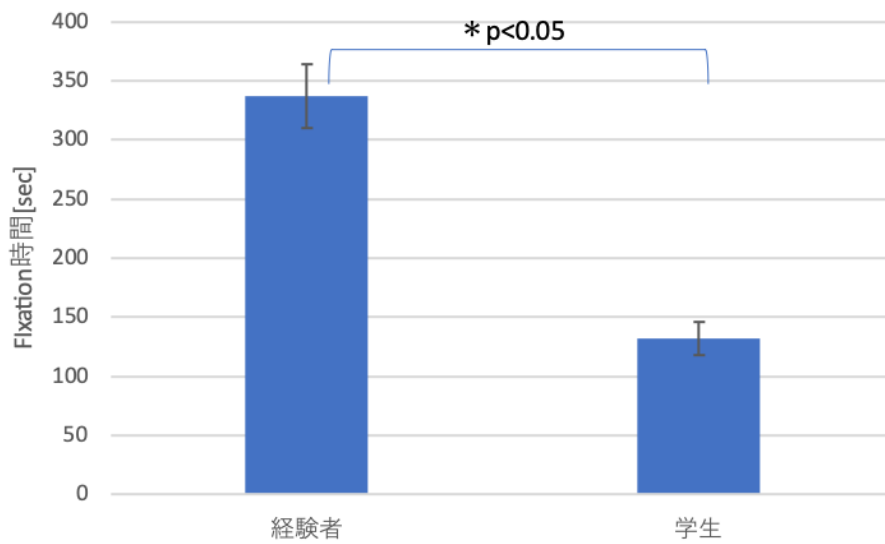


図 28 全ての画像を対象とした Fixation 時間

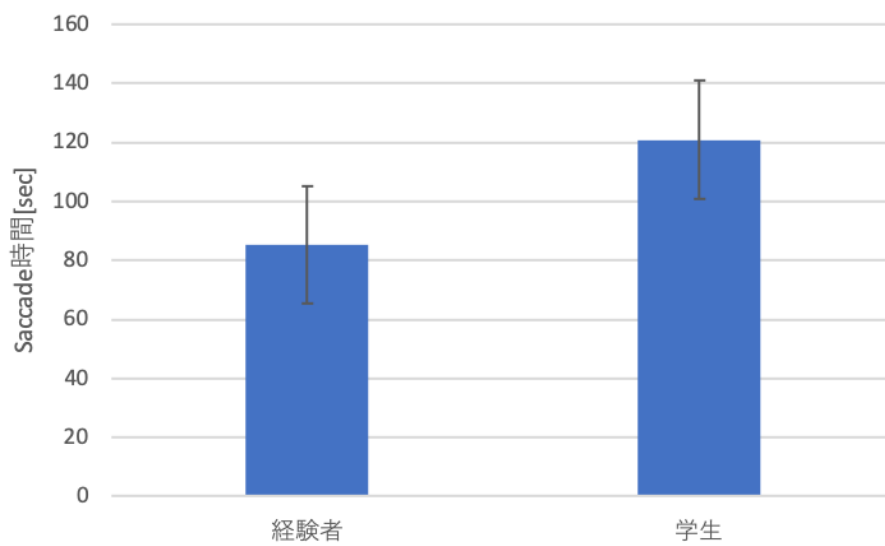


図 29 全ての画像を対象とした Saccade 時間

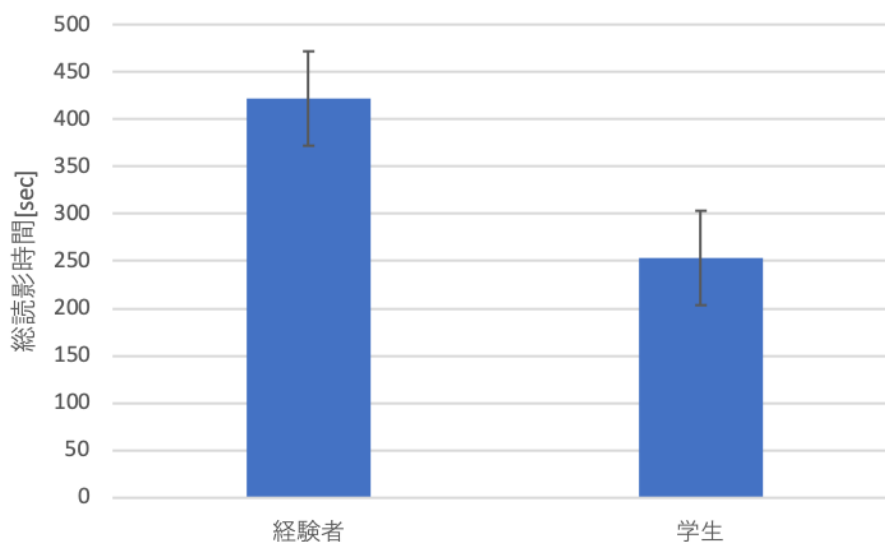


図 30 全ての画像を対象とした総読影時間

ここで、経験者と学生が各々で病変画像を読影している時間と正常画像を読影している時間において分散の違いがあるのか F 検定を行った。その結果を表 33、34 と図 31、32 に示す。

表 33 経験者における病変画像と正常画像での 1 枚あたりの時間

グループ	標本数	平均[sec]	標準偏差	分散
病変画像	15	25.05	5.80	33.66
正常画像	15	31.28	1.83	3.35

表 34 学生における病変画像と正常画像での 1 枚あたりの時間

グループ	標本数	平均[sec]	標準偏差	分散
病変画像	15	15.73	0.37	0.13
正常画像	15	8.99	2.41	5.82

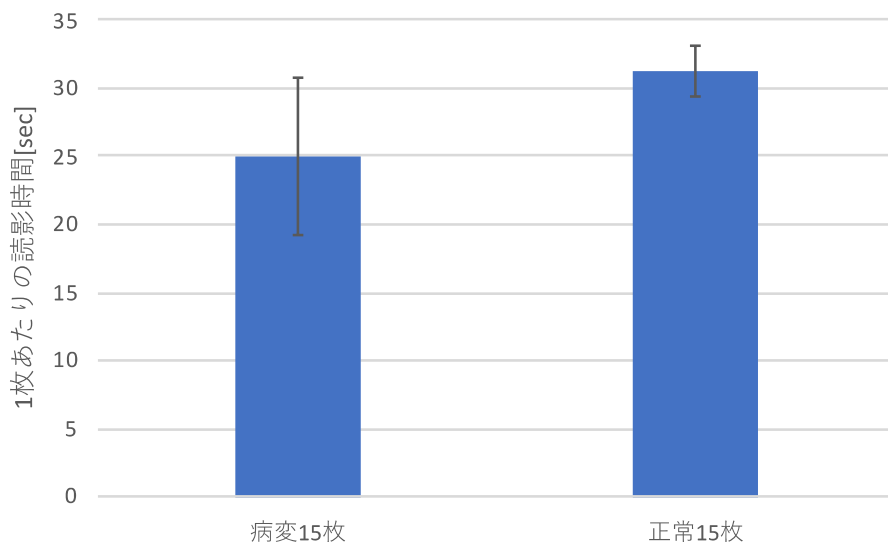


図 31 経験者における病変画像と正常画像での 1 枚あたりの時間

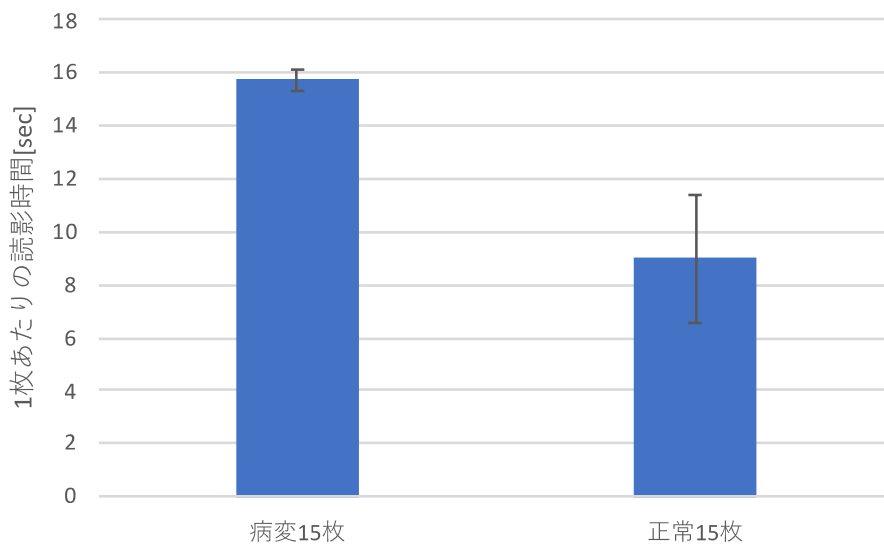


図 32 学生における病変画像と正常画像での 1 枚あたりの時間

図 31、32 より、F 値がそれぞれ 0.195(>0.05)、0.187(>0.05)となり、分散の違いにおいて有意水準 5 %で統計的に有意な差が見られなかった。したがって、被験者それぞれで病変画像と正常画像ともに 1 枚を読影している時間に差があるのかどうかは言及できないが、経験者の総時間は学生に比べて長いことは分かった。

第6章 考察

教師なしクラスタリングである k-means 法と CNN を用いた教師あり学習では、正解率、Fixation を Fixation と正しく判定した確率(感度)、Saccade を Saccade と正しく判定した確率(特異度)の全てにおいて、CNN を用いた教師あり学習が高い結果となった。PCA の図では、セントロイドに着目すると重なっていないが、Fixation と Saccade の境界を明らかにするのは困難である。画像の特徴抽出に優れた CNN を用いたことで、高い精度が得られたと考えられる。MNIST データベースを用いて転移学習させるモデルと DCNN の比較では、正答率において DCNN model の方が優れていた。MNIST データベースを用いて転移学習させるモデルに対し、3.2 提案モデルで記述したように、層の内容を工夫した DCNN model では、正答率、感度および特異度の高さから、視線を分類するための高い能力を有するということが明らかになった。

公開データである Lund 2013 のデータセットに対しては、DCNN model よりも gazeNet は感度が高く、従来手法である velocity 法は特異度が高い結果となった。しかし、Lund 2013 と GazeCom のデータを用いた gazeNet の結果と提案モデルとの比較をすると、GazeCom は 250 Hz のデータセットであるため、Lund2013 で学習した gazeNet(500 Hz)では周波数が合わず、モデルの汎用性がないことがわかった。我々の提案手法はパッチ画像を用いており、周波数に依存することはなかった。

GazeCom のデータセットに対しては、DCNN model が感度、特異度ともに他のモデルよりも高く、最適であった。加えて、Lund 2013 のデータセットについて、感度と特異度どちらともが 2 番目に高くなった DCNN model は、汎用性の高さが示唆できる。したがって、公開データの種類を問わずモデルを適用することができる。これは、各公開データを交差させた検証からも確認された。

応用実験では画像全てを対象とする Fixation 時間で、グループ間で有意差があった。これは、経験者は学生よりも注視時間が多く丁寧に医用画像を見ていることに起因しており、従来の ROC 解析の正当性を裏付けると考えられる (図 13、15)。

先行研究 [1]による検証結果と比較すると、応用実験において画像全てを対象とする Fixation 時間以外でグループ間での有意差はなかった。先行研究では、画像全てと病変画像の Fixation 時間でグループ間の有意差がみとめられたが、同じではなかった。そのため、今後先行研究に基づき各グループ 20 人ずつ被験者数を増やし検証する必要がある。

応用実験では、先行研究と同様の条件で比較するために画像を加増していない。その

ため、Fixation と Saccade の画像枚数に差が生じている。CNN では画像の偏りすなわち不均衡データによって精度が落ちることが知られており [49]、そのために本研究で提案した DCNN モデルの感度と特異度の差異を生じたことが考えられる。

第7章 結論

本研究は、医用画像を読影する際の眼球運動識別アプリケーションを作成するための基礎研究である。アプリケーションに導入予定の UI の一部を、図 33 に示す。現状の医用画像読影の評価の研究では、何回病変部位を見たか、何秒間見たかなどの静的な統計集計が行われてきた。本研究では眼球運動に着目することで **Fixation** や **Saccade** などを用いた動的な解析を行い、既存の静的な評価の研究と同様の結果を得ることができるかを検証した。図 33 を見ても分かる通り、**Fixation** と **Saccade** を判別することができれば、UI 上で **Fixation** と **Saccade** のそれぞれの動きがどのタイミングで起こったのかが分かる。本研究の動的な解析は、医用画像読影の研究以外に眼球運動や視線の研究で盛んに行われている静止画や動画を見る、あるいは車の運転中の眼球運動を対象とした研究を参考にして、機械学習や CNN を用いて分類する手法を用いた。まず初めに、提案する識別器の性能を、先行研究 gazeNet と同様に、公開データである Lund 2013 と GazeCom を用いて比較した。結果として、先行研究よりも高い正解率を実現した。また、教師なしクラスタリングである k-means 法と CNN を用いた 2 つの転移学習モデルと DCNN モデルを比較した結果、DCNN モデルが最も良い精度を示した。この DCNN モデルを用いて、Lund 2013 のデータセットで重みを学習し、実際に視線行動を測定した実験で得たデータを **Fixation** と **Saccade** に分類した。病変画像に限っては先行研究と異なり、**Fixation** 時間では経験者と学生の間には有意差がなかった。しかし画像全てに対しては、先行研究と同じく **Fixation** 時間で経験者と学生の間には有意差があった。本研究では、提案する識別器の精度を測ることを目的とし、提案した中で高精度を出した DCNN モデルを採用し試行した。本研究の課題は 2 つある。1 つ目は、経験者 2 人と学生 2 人の計 4 人のみしか計測実験を行っていないため、被験者を増やした実験を今後実施する必要がある。2 つ目は、病変部位を特定する行為そのものの、確からしさを検証することである。今後の展望として、医用画像を読影する際の眼球運動識別アプリケーションを作成し、教育効果があるかを実証予定である。

Information

JPCLN001.png

Lesion: (x, y) = (957, 405)

False Location

Click: (x, y) = (484, 462) (Dist: 476.8)

VAS: 0.960

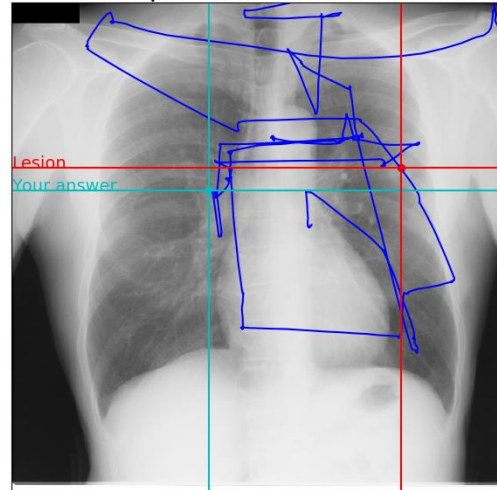
Total time [s] 10.52

Fixation duration [s] 6.34

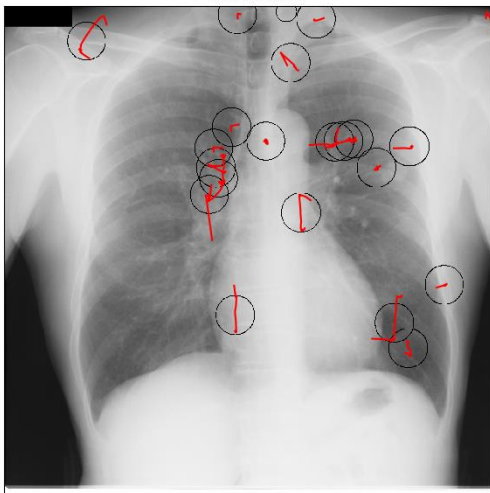
Fixation counts 21

Time of first fixate [s] 6.66

Clickpoint and Gaze line



Fixation



Saccade

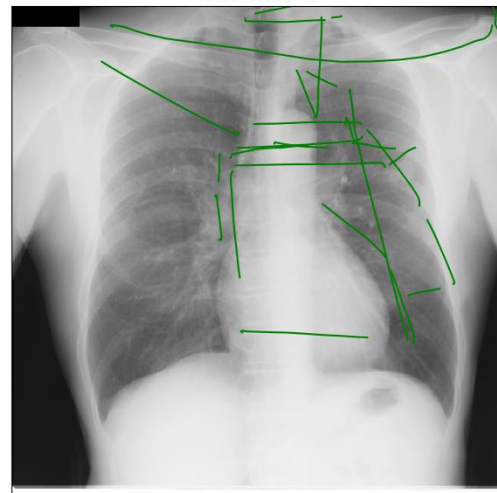


図 33 アプリケーション導入予定の UI

謝辞

本研究は多くの方々からの御指導、御協力をいただき、進めることができました。本研究を行うにあたり、丁寧な御指導と研究環境を与えてくださった奈良先端科学技術大学院大学先端科学技術研究科計算システムズ生物学研究室 金谷 重彦 教授に深く感謝致します。他大学の学部生であった私に快く研究室にて勉強する機会を与えてくださったことが、私のこれまでの人生の中で大きな転機となりました。また、研究室ゼミでは的確な御意見を頂きました先端科学技術研究科 Md. Altaf-Ul-Amin 准教授、先端科学技術研究科 小野 直亮 准教授、先端科学技術研究科 黄 銘 助教に深く感謝致します。

研究発表の際、貴重な御意見及び助言を賜りました奈良先端科学技術大学院大学先端科学技術研究科 佐藤 嘉伸 教授に厚く御礼申し上げます。

学部生の頃に分野外であった私を受け入れてくださり、様々な面から御助言、熱心な御指導いただきました先端科学技術研究科 佐藤 哲大 客員准教授に深く御礼申し上げます。大学院入学前から、長い間、大変お世話になりました。

並びに、応用実験で使用した視線データの提供や、医療という観点から貴重な御助言を頂きました、群馬県立県民健康科学大学診療放射線学科 寺下 貴美 講師に深く感謝致します。

計算システムズ生物学研究室の皆様には、大学院生活全般にわたり様々な面で支えていただきました。皆様のおかげで充実した大学院生活を過ごすことができました。誠にありがとうございました。

高校の友人でありメルボルン大学博士課程 石原 輝人 君には研究者としての心得やたわいも無い話をする事で大学院生活を有意義に過ごすことができました。感謝しております。

最後に、大学院修了まで多大なる支援を頂いた家族に深く感謝し、ここに謝辞とさせていただきます。

参考文献

- [1] McLaughlin L, Bond R, Hughes C, McConnell J, McFadden S, "Computing eye gaze metrics for the automatic assessment of radiographer performance during X-ray image interpretation," *Int J Med Inform.* 105, pp. 11–21, 2017.
- [2] Mall S, Brennan PC, Mello-Thoms C, "Modeling visual search behavior of breast radiologists using a deep convolution neural network," *J Med Imaging (Bellingham).* 5(3), pp. 035502., 2018.
- [3] Ogawa W, *Image evaluation for radiography [in Japanese]*, Iryokagakusha (Tokyo), 2007.
- [4] Dalrymple KA, Jiang M, Zhao Q, Elison JT, "Machine learning accurately classifies age of toddlers based on eye tracking," *Scientific Reports*;9(1):pp.6255. , 2019.
- [5] Zemblys R, Niehorster DC, Holmqvist K, "gazeNet: End-to-to-end eye-movement event detection with deep neural networks," *Behav Res Methods.* 51(2), pp.840-864, 2019.
- [6] 内川 恵二, *視覚心理入門-基礎から応用まで-*, 株式会社オーム社, 2009.
- [7] トビー・テクノロジー, "アイトラッキング活用領域," [オンライン]. Available: <https://www.tobiipro.com/ja/fields-of-use/>. [アクセス日: 28 1 2020].
- [8] Holmqvist K, Nystrom M, Andersson R, Dewhurst R, Jarodzka H, *EYE TRACKING A COMPREHENSIVE GUIDE TO METHODS AND MEASURES*, Oxford University Press, 2015.
- [9] Shiraishi J, Katsuragawa S, Ikezoe J, et al, Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules., *Am J Roentgenol.* 174(1),pp.71-74, 2000.
- [10] 医用画像工学ハンドブック 編集委員会, "医用画像工学ハンドブック," 日本医用画像工学会, 2012.
- [11] 櫛橋 民生, *ポータブルを含む各種写真の正しい理解 胸部 X線写真の読影*, 中外医学社, 2013.

- [12] Cho K, Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y, “Learning phrase representations using RNN encoder-decoder for statistical machine translation. ” CoRR, arXiv:1406.1078, 2014.
- [13] Friedman L, "TWO ERRORS AND TWO PROBLEMS IN A EWCENT PAPER: gazeNet: END-TO-END EYE-MOVEMENT EVENT DETECTION WITH DEEP NEURAL NETWORKS," arXiv:1908.11436v2, 2019.
- [14] Bahill AT, Brockenbrough A, & Troost, B. T, "Variability and development of a normative data base for saccadic eye movements," *Investigative Ophthalmology & Visual Science*, 21(1), pp.116–125., 1981.
- [15] Salvucci DD, Goldberg JH, "Identifying fixations and saccades in eye-tracking protocols," In *Proceedings of the 2000 symposium on eye tracking research & applications, ETRA '00* (pp. 71–78)., 2000.
- [16] Boyce PR, “Monocular fixation in human eye movement,” *Proceedings of the Royal Society of London B: Biological Sciences*, 167(1008), pp.293-315., 1967.
- [17] Engbert R, Kliegl R, “Microsaccades uncover the orientation of covert attention,” *Vision Research*, 43(9), pp.1035-1045., 2003.
- [18] Nystro ¯m M, Holmqvist K, “An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data,” *Behavior Research Methods*, 42(1), pp.188-204., 2010.
- [19] Mould MS, Foster DH, Amano K, Oakley JP, “A simple nonparametric method for classifying eye fixations,” *Vision Research*, 57, pp.18-25., 2012.
- [20] Hessels RS, Niehorster DC, Kemner C, Hooge ITC, “Noise-robust fixation detection in eye movement data: Identification by two-means clustering (i2mc),” *Behavior Research Methods*, 49(5), pp.1802-1823., 2017.
- [21] Larsson L, Nystro ¯m M, Stridh M, “Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit,” *IEEE Transactions on Biomedical Engineering*, 60(9), pp.2484- 2493., 2013.
- [22] Larsson L, Nystro ¯m M, Andersson R, Stridh M “Detection of fixations and smooth pursuit movements in high- speed eye-tracking data,” *Biomedical*

- Signal Processing and Control, 18, pp.145-152., 2015.
- [23] Zemblys R, “Eye-movement event detection meets machine learning,” In Biomedical engineering (pp. 98-101)., 2016.
- [24] Anantrasirichai N, Gilchrist ID, Bull DR, “Fixation identification for low-sample-rate mobile eye trackers,” In 2016 IEEE international conference on image processing (ICIP) (pp. 3126-3130). IEEE., 2016.
- [25] Hoppe S, Bulling A, “End-to-end eye movement detection using convolutional neural networks,” ArXiv e-prints., 2016.
- [26] Zemblys R, Niehorster DC, Komogortsev O, Holmqvist K, “Using machine learning to detect events in eye-tracking data,” Behavior Research Methods, 50(1), pp.160-181., 2018.
- [27] C・M・ビショップ, パターン認識と機械学習 下 ベイズ理論による統計的予測, 丸善.
- [28] 斉藤 康毅, ゼロから作る Deep Learning -Python で学ぶディープラーニングの理論と実装, 株式会社オライリー・ジャパン. 株式会社オーム社, 2017.
- [29] 松尾 豊, 松原 仁 (人工知能学会), 深層学習-Deep Learning-, 株式会社近代科学社, 2016.
- [30] 堂下 晃広, 藤田 広志, 津坂 昌利, “ニューラルネットワークによる医用 X 線画像処理法の基礎的研究: ボケ画像の回復.,” 医用画像情報学会雑誌. Vol. 10. No.3. pp.199- 128., 1993.
- [31] 真下 智行, 蔡 篤儀, 山田 功, 小島 克之, “白黒濃淡画像におけるニューラルネットワークを用いた動的領域分割とその医用画像への適用,” 医用画像情報学会雑誌. Vol. 16. No.3. pp.179- 185., 1993.
- [32] 李 鎔範, 蔡 篤儀, “微小石灰化像良悪性鑑別のための人工ニューラルネットワーク法とファジィ推論法の ROC 比較評価,” 医用画像情報学会雑誌. Vol.21. No.1. pp.122- 130., 2004.
- [33] 李 文光, 張 学軍, 兼松 雅之, 原 武史, 周 向荣, 藤田 広志, 加藤 博基, 近藤 浩史, 星 博昭, “腹部 MR 画像における肝硬変の自動識別法の開発,” 医用画像情報学会. Vol.21. No.2. pp.194- 200., 2004.

- [34] 國枝 琢也,内山 良一,原 武史,藤田 広志,加藤 博基,浅野 隆彦,兼松 雅之,星 博昭,岩間 亨,紀ノ定 保臣,横山 和俊,篠田 淳, “脳 MR 画像におけるラクナ梗塞と血管周囲腔拡大の鑑別法,” 医用画像情報学会. Vol.26. No.3. pp.59- 63, 2009.
- [35] 庄野 逸,鈴木 聡志,木戸 尚治, “ディープラーニングの医用画像への応用,” 医用画像情報学会雑誌 vol.33 No.4. pp.48- 53, 2017.
- [36] LeCun Y, Cortes C, Burges CJC, “Classification of handwritten character for Mixed National Institute of Standards and Technology (MNIST) database,” [オンライン]. Available: <http://yann.lecun.com/exdb/mnist/>. [アクセス日: 28 1 2020].
- [37] LeCun Y, Bottou L, Bengio Y, Haffner P, “GradientBased Learning Applied to Document Recognition,” Proceedings of the IEEE 86, pp.2278-2324., 1998.
- [38] Krizhevsky A, Sutskever I, Hinton GE, “ImageNet Classification with Deep Convolutional Neural Networks,” In F.Pereira, C.J.C.Burges, L.Bottou, & K.Q.Weinberger, eds. Advances in Neural Information Processing System 25. Curran Associates, Inc., pp.1097-1105., 2012.
- [39] imonyan K, Zisserman A, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv:1409.1556, 2015.
- [40] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A, “Going Deeper with Convolutions,” In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ., 2015.
- [41] 瀧 雅人, これならわかる深層学習 入門, 講談社, 2017.
- [42] Terashita T, Sato T, Tsutsumi S, Sato M, Ogura T, Doi K, Eye movement analysis using deep learning for medical image interpretation, ECR 2019 EPOS. C-0155, 2019.
- [43] McClure N, Tensorflow Machine Learning Cookbook, Packt Publishing(UK), 2017.
- [44] Kingma DP, Ba JL, “Adam: A Method for stochastic optimization.,” arXiv: 1412.6980, 2014.
- [45] Yamashita R, Nishio M, Do RKG, Togashi K, “Convolutional neural networks:

- an overview and application in radiology,” *Insights Imaging*. 9(4), pp. 611-629., 2018.
- [46] Startsev, M., Agtzidis, I., & Dorr, M, “Smooth pursuit,” [オンライン]. Available: <http://michaeldorr.de/smoothpursuit/>. [アクセス日: 28 1 2020].
- [47] Startsev, M., & Agtzidis, I, “ Manual & automatic detection of smooth pursuit in dynamic natural scenes,” In *Proceedings of the European conference of eye movements.*, 2017.
- [48] トビー・テクノロジー, “Tobii Pro X3-120,” [オンライン]. Available: <https://www.tobiipro.com/product-listing/tobii-pro-x3-120/>. [アクセス日: 28 1 2020].
- [49] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP, “SMOTE: Synthetic minority over-sampling technique,” in *JAIR*, 2002.

学会発表リスト

- [1] 神川 怜史, 佐藤 哲大, 寺下 貴美, 江口 遼平, 金谷 重彦 : “教師なし学習を用いた医用画像読影時の視線行動の分類”, 電子情報通信学会 医用画像研究会, 2019 年 1 月
- [2] Satoshi Kamikawa, Takayoshi Terashita, Shigehiko Kanaya, Tetsuo Sato, “Comparing convolutional neural networks for eye movement analysis in medical image interpretation”, 生体医工学シンポジウム, 2019 年 9 月
- [3] Satoshi Kamikawa, Takayoshi Terashita, Shigehiko Kanaya, Tetsuo Sato, “Comparing convolutional neural networks for medical image interpretation based on eye movement analysis”, ECR2020 (European Congress of Radiology), 2020 年 7 月