

AI から学ぶ：
エキスパートの知見を導入した
DNN モデルを教師とするインタラクティブ学習に関する研究
TR20011 服部 幸平

指導教授：藤吉 弘亘

1. はじめに

Attention Branch Network (ABN) [1] は、深層学習の判断根拠を人が解釈できるよう、推論時における注視領域をアテンションマップとして可視化する視覚的説明手法である。さらに ABN は、アテンションマップを特徴マップに重みづけをすることで認識性能の向上を実現した。三津原ら [2] は、アテンションマップをエキスパートの知見に基づいて修正し、モデルをファインチューニングすることで視覚的説明性と認識性能をさらに向上させた。

本研究では、視覚的説明を人が解釈するにとどまらず、AI から人が学ぶ Human-in-the-loop を実現することを目的とする。そこで、エキスパートの知見を組み込んだ深層学習モデルを教師役として、学習者がその判断根拠をインタラクティブに学習する方法を提案する。提案手法は、学習者が編集したアテンションマップを利用して、エキスパートの知見を組み込んだ ABN が推論する。この時の推論したスコアが高くなるように、学習者によるアテンションマップの修正と ABN による推論を繰り返す。これにより、ABN に組み込んだエキスパートの判断根拠である注視領域をインタラクティブに学習することが可能となり、高い教育効果を実現する教育アプリとなる。教育アプリを用いて、眼底画像の疾患判定を対象とした被験者による評価実験を行い、提案手法による学習効果について示す。

2. 関連研究

ABN は、推論時の判断根拠となるアテンションマップを特徴マップに乗算することで認識性能の向上を実現している。ABN はアテンションマップを介して視覚的な説明が可能であるが、学習データが少なかったり偏りがあると、注視領域が認識対象以外に発生する場合や認識対象を注視しない場合がある。三津原らは学習済みの ABN の誤ったアテンションマップを人の知見に基づいて修正し、教師データとして再学習することで正確なアテンションマップの獲得を実現した [2]。詳細画像識別タスクにおいて、人の知見を利用してアテンションマップを修正することで、物体のより詳細な領域に着目したアテンションマップの生成を実現し、優れた分類性能を達成した。

3. 提案手法

本研究は、エキスパートの知見に基づいてファインチューニングした ABN を教師とし、学習者がその判断根拠をインタラクティブに学ぶ学習法として、アテンションマップを編集し推論する教育アプリを作成する。

3.1 アテンション機構によるアテンションマップの修正

判断根拠をインタラクティブに学ぶために、ABN のアテンションマップを学習者が繰り返し修正する。その際、出力されるスコアの変動よりどの領域に着目すべきかという知識を学習者が自ら獲得する。学習者が編集したアテンションマップ $M_n \in \{0, 1\}$ を用いて、各クラスに対するスコアを出力するために、Feature Extractor から得られた特徴マップ $g(x)$ に対して重みづけを行う。式 (1) より重みづけされた特徴マップ $g'(x)$ を得る。

$$g'(x) = [g(x) \cdot M_n(x)] + g(x) \quad (1)$$

また、 $M_n \in \{0, 1\}$ は、学習者が注視した領域を 1、注視していない領域を 0 としている。 $g'(x)$ を用いて Perception Branch により、教育アプリで学習者に表示する各クラスにおけるスコアを出力する。学習者はこのスコアの変動から注視領域について学ぶ。

3.2 “AI から” 学ぶ教育アプリ

本アプリでは、学習者が編集したアテンションマップに対し、エキスパートの知見を組み込んだ AI がスコアを提示することで、学習者と AI がインタラクティブに学習することを可能とする。図 1 に提案手法を導入した教育アプリとその操作を示す。提案手法を導入した教育アプリを用いたインタラクティブに学習する手順を以下に示す。

- step.1 データセットから無作為に選択した画像を 1 枚ずつ表示し、学習者がクラス分類を行う。
- step.2 表示した画像に対応するラベルから正答を確認する。
- step.3 学習者がラベルに合わせて、画像の特徴を強調するようにアテンションマップの編集を行う。
- step.4 編集したアテンションマップを用いて、AI が各クラスにおけるスコアを推論する。
- step.5 学習者がラベルに合わせたスコアを確認する。
- step.6 クラス分類をしたスコアの変化より、自身の編集したアテンションマップを時系列的に比較し、スコアが向上する注視領域を学習する。

step.3 から step.6 の処理を繰り返すことで、学習者が AI によりインタラクティブに学習する。また、学習者がラベルに対して正答した場合は step.3 から step.6 を行わない。図 2 に、学習者が教育アプリによりインタラクティブに学習する様子を示す。図 2 より、適切な領域を追加することでスコアが上昇し、不適切な領域を追加するとスコアが下がることがわかる。このように能動的に学習することで、効率良く教育効果を得られると期待できる。

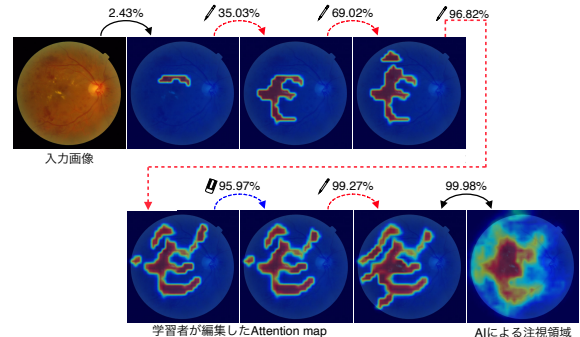


図 2：インタラクティブ学習によるスコアの変化

4. 評価実験

提案手法の効果を検証する。本研究では、眼底画像の疾患である糖尿病網膜症を対象とする。被験者 36 人を教育アプリなし、エキスパートの知見がない教育アプリ、提案手法による教育アプリの 3 つのグループに分け比較実験を行う。実験は、眼底疾患のデータセットから二値分類の出題を行い、正解率から学習効果を評価する。被験者が行う学習前と学習後のテストでは、Messidor のデータセットから正常な眼底画像を 30 枚、疾患のある眼底画像を 30 枚用いる。被験者が行う学習には、Indian Diabetic Retinopathy Image Dataset (IDRiD) のデータセットを使用する。正常な眼底画像を 124 枚、疾患のある眼底画像を 81 枚学習教材として用いて、被験者は学習を行う。

4.1 正解率による評価

3 つのグループの結果をそれぞれ表 1 に示す。ここで、高得点を 80 点以上と設定し、高得点者数を用いて分析を行う。

表 1 より、その他の学習法と比較し、提案手法を用いて学習することで、高得点者数が多いことが確認できる。また、アプリを使用しないグループとエキスパートの知見がない AI モデルを用いたグループでは、得点の標準偏差が

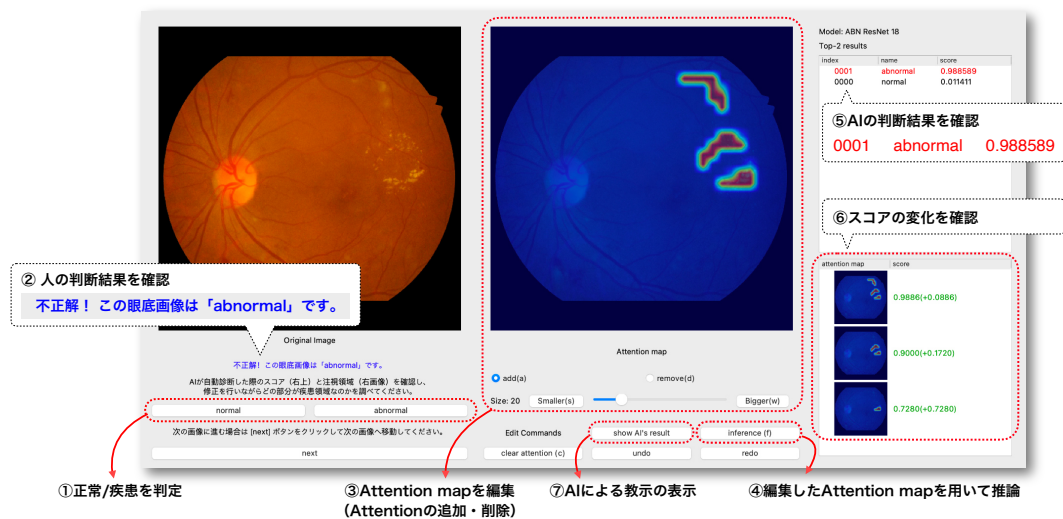


図 1：提案手法による教育アプリ

表 1：評価実験結果

	アプリなし		提案手法			
	学習前	学習後	学習前	学習後	学習前	学習後
80 点未満 [人]	12	9	12	11	12	5
80 点以上 [人]	0	3	0	1	0	7
平均 [%]	55.69	77.78	57.50	69.86	53.47	79.44
標準偏差	11.86	3.65	12.82	9.20	9.88	2.50

大きい。これは、学習やその後の解答が個人の能力に依存しているためと考えられる。一方、提案手法の教育アプリのグループは標準偏差が小さいことから、多くの学習者にとって教育効果が高い。

4.2 アテンションマップによる評価

被験者が編集したアテンションマップを比較し、適切な領域に注目しているかの学習効果を検証する。図 3 に各グループの被験者が編集したアテンションマップの例を示す。

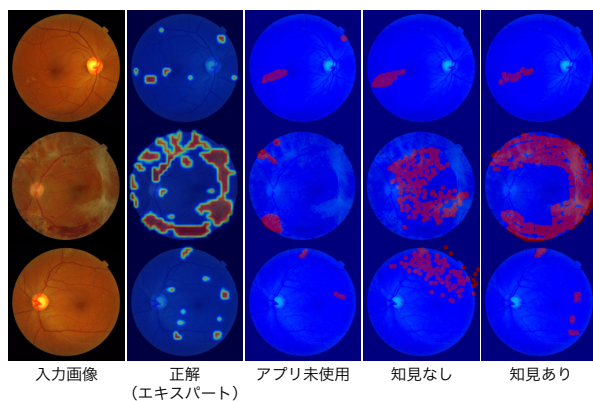


図 3：アテンションマップの比較図

提案手法により学習を行った被験者はエキスパートのアテンションマップと類似したアテンションマップを編集している。一方で、そのほかの手法では、異なる領域にアテンションを追加する場合や、一部エキスパートのアテンションと同じ箇所を編集しているものの、アテンションを追加していない領域などが見受けられる。

また、アテンションマップを用いた定量的評価として、30 枚の疾患画像に対するエキスパートと被験者のアテンションマップの疾患領域に対するクラス IoU を表 2 に示す。提案手法は学習後のクラス IoU が最も大きく、最もエキスパートの知見に近いアテンションマップを編集していることがわかる。以上より、エキスパートの知見を導入した AI モデルによる学習により被験者がインタラクティブにどこに注視すべき領域であるかを学習したことが確認できる。また、AI が示すアテンションマップにより、疾患判定の判断根拠を学習者に示すことで、高い学習効果が得られることを確認した。

表 2：エキスパートと被験者の IoU

	教育アプリなし		提案手法			
	学習前	学習後	学習前	学習後	学習前	学習後
平均	0.0958	0.1079	0.0895	0.0835	0.1107	0.1314

4.3 被験者からのフィードバック

教育アプリに対する定性的評価として、知見を導入した教育アプリを使用した被験者からの実験終了後のフィードバックを以下に示す。

- 能動的な学習により、疾患がどういったものかわかるようになった

一方、エキスパートの知見がない AI モデルを用いて学習した被験者からのフィードバックを以下に示す。

- 細かい領域に着目することができず、疾患の特徴に対する思い込みを修正するのが難しかった

以上より、教育アプリに知見を導入した AI モデルを併せて活用することで、判断根拠となる疾患領域を明確にし、提案手法の学習効果を促進することを示した。

5. おわりに

本研究では、エキスパートの知見を導入した深層学習モデルを教師として、学習者である人がその判断根拠を“AI から”学ぶ手法を提案した。提案手法では、ABN のアテンションマップ及びアテンション機構を活用し、ネットワークへエキスパートの知見を導入した。知見を導入した ABN を用いて、学習者が編集したアテンションマップに対しアテンション機構の活用による分類結果を確認することで、AI の判断根拠のもとに学習を行うことを可能とした。また、この学習法を容易に行うために、教育アプリケーションを作成し、被験者実験を行った。評価実験により、提案手法を用いることで、より効率的に学習を行うことができ、エキスパートの知見と同様の疾患領域に着目し判定することが可能となることを示した。本手法は、医療画像の疾患判定だけではなく、外観検査をはじめとする高い専門性を必要とするタスクへの活用が期待できる。

参考文献

- [1] H. Fukui, *et al.*, “Attention Branch Network : Learning of Attention Mechanism for Visual Explanation”, CVPR, 2019.
- [2] M. Mitsuhashi, *et al.*, “Embedding Human Knowledge into Deep Neural Network via Attention Map”, International Conference on Computer Vision Theory and Applications, pp. 626-636, 2021.

研究業績

- [1] 服部幸平, 平川翼, 山下隆義, 藤吉弘巨, “AI から学ぶ：エキスパートの知見を導入した DNN モデルを教師とする学習者のためのインタラクティブ学習法の提案”, 電子情報通信学会パターン認識・メディア理解研究会 (PRMU), 2022.