

国立大学法人北陸先端科学技術大学院大学
国立大学法人東京大学
国立研究開発法人科学技術振興機構 (JST)

深層学習をより簡単に、より正確にする新しいインターフェイスを開発 —AIのバイアスによる影響をワンクリックで軽減—

ポイント

- ユーザがワンクリックで画像に注釈をつけ、人工知能 (AI) が正しい領域に注目して判断できるようにする画期的なヒューマンインザループシステム [用語解説] を開発しました。
- 本システムでは、AI が、学習する際に必要なデータセット [用語解説] で課題となる「共起バイアス」による影響を軽減できます。
- 本システムにより、深層ニューラルネットワーク (Deep Neural Network: DNN) の手法を用いた深層学習をより少ないデータでより正確に実行することが可能となり、DNN のトレーニングの時間とコストの削減につながることを期待されます。

<研究の背景と経緯>

近年、「人工知能 (AI)」という言葉が科学技術の世界でも大きな話題になっています。AI の登場により、これまで人間にしかできなかった作業をコンピュータが実行できるようになりました。現在、様々な研究分野において多種多様な手法を用いて AI を作成することができますが、最も広く用いられている手法の一つとして、深層ニューラルネットワーク (DNN) の手法を用いた深層学習があります。DNN は、ヒトの脳の神経回路の接続と機能を模したモデルで、実社会での運用前にデータセットによって AI をトレーニングします。データセットによるトレーニングを繰り返すことで、AI は画像の特徴を「学習」することができます。例えば、船が写っているデータセット (画像) でトレーニングすると、船が写っている画像を識別できるよう「学習」します。

しかし、深層学習におけるトレーニングでは、データセットが適切に設計されていないなどの理由から、AI が対象を正確に識別することができないなどの問題が生じることがあります。例えば、前述の例では、船の画像は一般的には、水と船が撮影されていますが、DNN におけるトレーニングでは、船ではなく水だけを認識し、画像に船が写っていると判断する

ことがあります。これを「共起バイアス」と呼び、DNN によるトレーニングにおいて、非常に多くの場面で直面する課題です(図1)。

<研究の内容>

北陸先端科学技術大学院大学 (JAIST) (学長・寺野稔、石川県能美市) 創造社会デザイン研究領域の謝浩然講師、Yi He 元研究補助員ならびに中国吉林大学人工知能学院の Xi Yang 准教授、東京大学 (総長・藤井輝夫、東京都文京区) 大学院情報理工学系研究科のチャン チアミン特任講師、五十嵐健夫教授らの研究チームは、今回、この「共起バイアス」を取り除く新たなヒューマンインザループシステムを開発しました (図2)。

これまでも深層学習の共起バイアスによる影響を取り除くため、データセットを再編成することや、AI が正確に特定の領域を認識できるよう、ユーザ (人間) が直接指示するなどの手法がありました。しかし、データセットを再編成することは非常に困難であり、また、ユーザが画像内に直接、注目すべき領域 (ROI) ^[用語解説] を指示するには、画像内にピクセル単位で注釈を付ける作業に人手が必要になり、高いコストがかかります。

そこで、本研究では、ユーザがワンクリックで画像に注釈を付けることができるよりシンプルなアテンション誘導 ^[用語解説] システムを新たに開発しました。これにより、ユーザはワンクリックで画像内の ROI を指定できるようになり、DNN によるトレーニングにかかる時間とコストを大幅に削減することができます。

さらに、従来のアテンション誘導手法では双方向の設計がされておらず、効率が悪かったため、本研究では、ユーザが画像をアノテーションする ^[用語解説] ことにより DNN が ROI を調整する双方向システムも提案しました。本システムでは、ユーザは、マウスのクリックで画像の識別したい部分を左クリックし、必要に応じて無視すべき部分を右クリックします (図3)。例えば、前述した船の画像の例では、ユーザは船を左クリックし、周りの水を右クリックすることになります。これにより、DNN では、船をより正確に識別することができ、データセットの共起バイアスの影響を軽減することができます (図4)。

さらに今回、アノテーションが必要な画像の数を減らすため、ガウス混合モデル (GMM) ^[用語解説] を使用した新たなアクティブラーニング方法 ^[用語解説] についても考案しました。この方法は、これまでのどの方法より精度が高いことがユーザアンケートにより示され、その結果として、人間が画像内の ROI をクリックして注釈を付けるシステム (クリックベースのシステム) ではアテンション誘導に必要な時間が 27%削減されたこと、学習の正確性を大幅に向上させることが示されました (図5)。

本研究成果は、2023年3月27日から31日までオーストラリアのシドニーで開催の、人工知能とインタラクション技術に関する国際会議 ACM IUI 2023 (28th Annual Conference on Intelligent User Interface) で発表されました。

<今後の展開>

今後は、本研究成果を用いて実社会のアプリケーションにおける DNN の転移性と解釈性を大幅に向上させることができます。システムがより正確で明確な判断を下すことができれば、ユーザが AI に対して持つ信頼度が高まり、これらのシステムをより簡単に社会応用へと展開できるようになります。

今後の課題としては、DNN の信頼性を高めることに焦点を当てており、AI 技術の適用と開発に大きな影響を与えうる社会実装を目指します。本研究成果が AI 業界に大きな影響を与え、近い将来に AI 技術のさらなる進展を可能にすることが期待されます。

<論文情報>

論文題目：Efficient Human-in-the-loop System for Guiding DNNs Attention

発表先：28th Annual Conference on Intelligent User Interfaces (ACM IUI 2023)

著者：Yi He, Xi Yang, Chia-Ming Chang, Haoran Xie, Takeo Igarashi

論文掲載日：2023年3月27日

DOI：https://doi.org/10.1145/3581641.3584074

解説動画：https://youtu.be/2MD-z6vXKJ4

解説ホームページ：https://yang-group.github.io/#/ProjectPageIUI2023

ソースコード：https://github.com/ultratykis/Guiding-DNNs-Attention

<用語解説>

ヒューマンインザループシステム (HITL: Human-in-the-Loop) :

工学やコンピュータ科学などさまざまな分野で広く用いられている用語で、「人間がループ (システム) の中に組み込まれている」という言葉どおり、ループ (システム) の中に人間との相互作用 (interaction) が含まれることを指します。例えば、機械学習分野において人間の介入によりデータ準備や学習効果の向上へと繋がります。

データセット (Dataset) :

何らかの目的や対象について収集され、一定の形式に整えられたデータの集合。機械学習などコンピュータによる自動処理を行うために用意された大量の標本データのことなどを指します。

ROI (Region of Interest) :

画像処理でよく使われる用語で、画像中における興味のある領域を指します。本研究では、DNN が注目すべき領域のことを指しています。

アテンション誘導 (Attention Guidance) :

データセットの共起バイアスにより、DNN が誤った領域に注目して判断することがあります。しかし、人間が行うアノテーションを利用し、DNN が注目すべき領域を調整することで、正しい領域に注目して判断できるようになります。

アノテーション (Annotation) :

データセットに含まれる各データに対して、人間が正解となるラベルや注釈などの情報を付与することを指します。本研究では、マウスによるシンプルなクリックで、画像領域に対して ROI の中心に注釈をつけることができます。

ガウス混合モデル (GMM) :

Gaussian Mixture Model の略であり、クラスタリングや異常検知によく使われるアルゴリズムの一つ。複数のガウス分布の重み付き平均を利用して、複雑なデータ分布をモデル化する確率モデルです。

アクティブラーニング (Active Learning) :

機械学習の予測精度を向上させるために、有用なデータを少量サンプリングすることでアノテーションコスト (人間がする作業や介入する時間のことなど) を低減する手法を指します。

<参考資料>

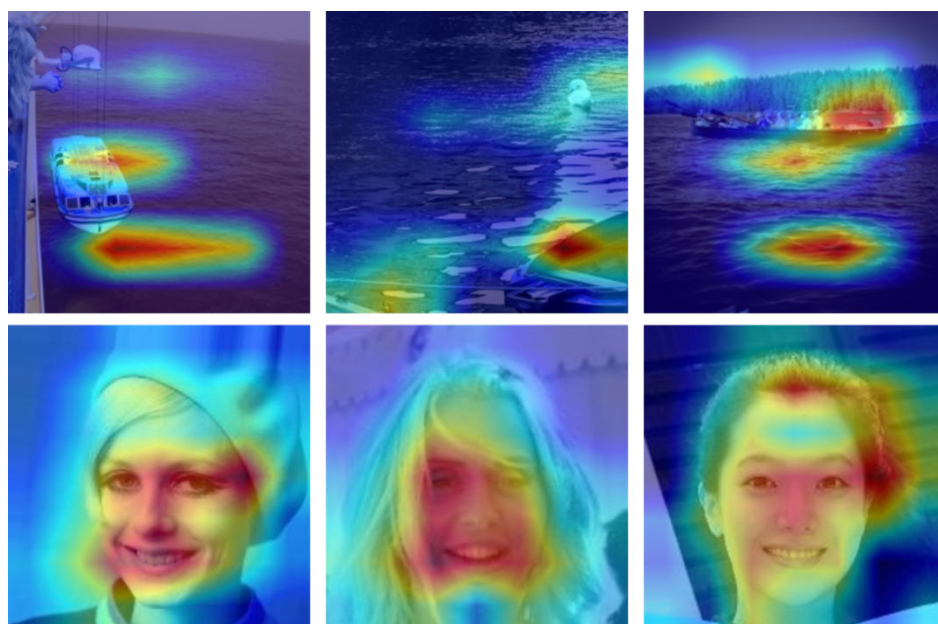


図1. AIトレーニングデータセットの共起バイアス

上段：DNNは船自体ではなく、水の波、海岸の輪郭、または船と水の境界に注目することで、「船」を分類します。

下段：DNNは「口紅」属性を分類する際に口の領域だけでなく、目や眉毛にも注目します。
(画像出典：COCO および CelebA データセット)

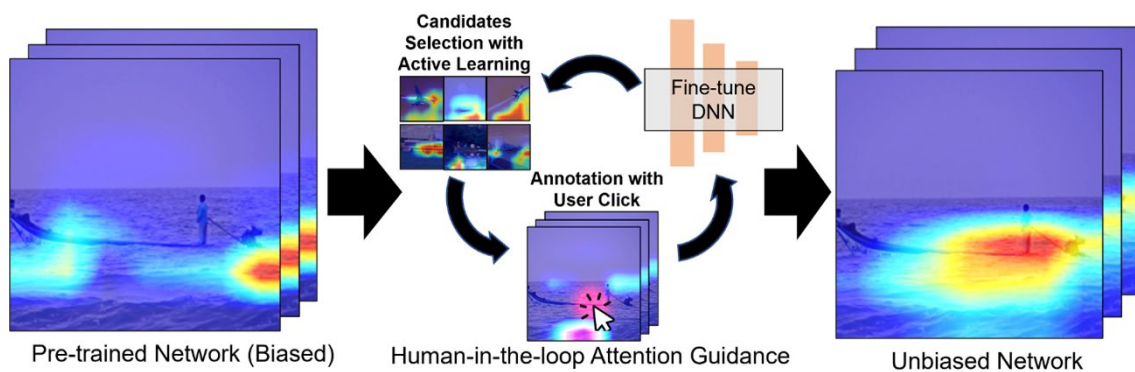


図2. 提案のクリックベースのAIトレーニングシステム

単一クリックの注目誘導を用いたユーザインターフェイスと新たなアクティブラーニング手法を利用することで、DNNをより正確かつ効率的にトレーニングできます。

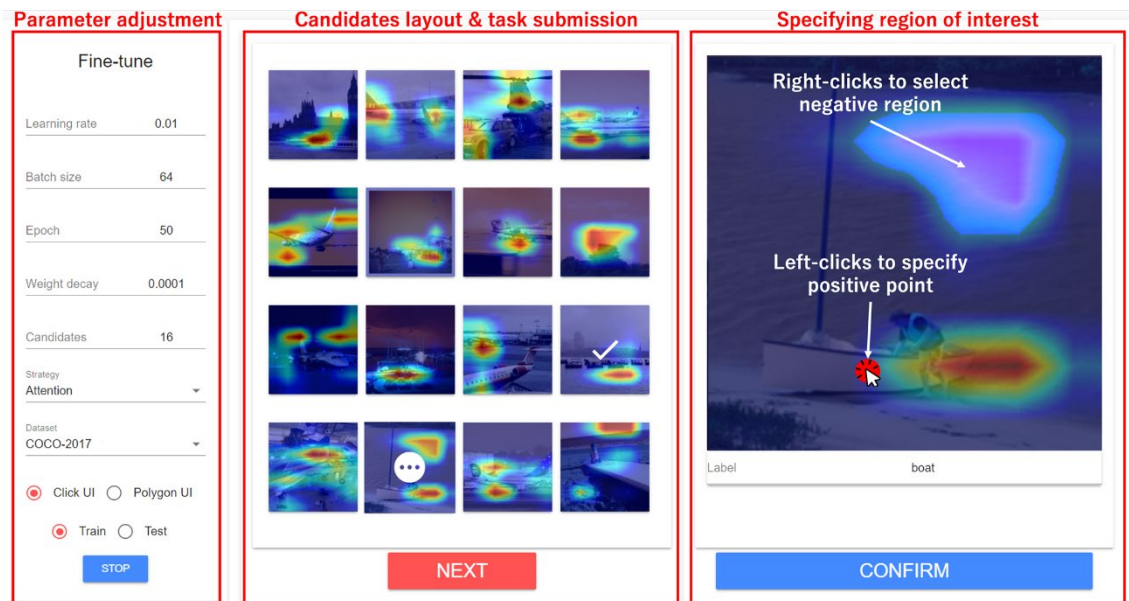


図3. AIトレーニングシステムの提案インターフェイス

図左：ユーザはパラメータを修正し、アクティブラーニングに使用する手法を選択し、データセットを選択できます。

図中央、右：ユーザは右側のアテンションマップの精度を上げるために左クリックして正の注目領域（画像識別したい部分）の中心を指定し、右クリックして負の領域（無視すべき領域）を指定できます。



図4. 提案のトレーニングシステムにより良い結果が得られました

CelebA データセットの「口紅」に関するデータセットでネットワークが示したアテンションマップを比較。

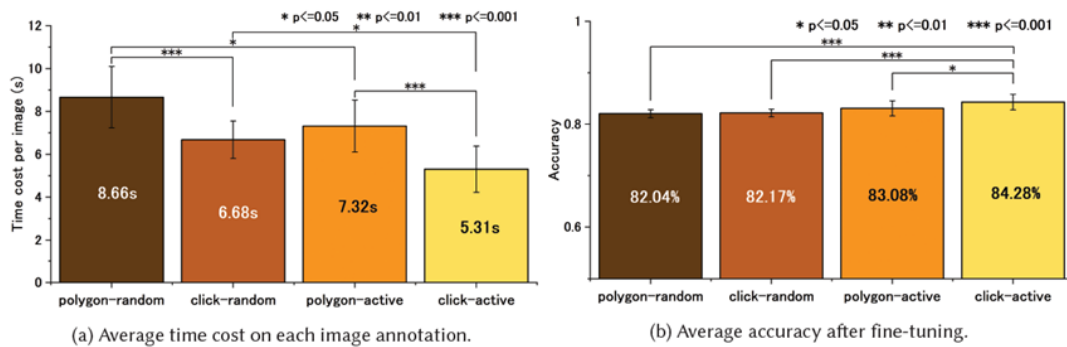


図5. クリックベースのAIトレーニングシステムによる平均利用時間と正確性
 polygon-random: アクティブラーニング未使用の従来法（目標対象の輪郭をなぞるインターフェイス）

click-random: アクティブラーニング未使用のクリックベースの方法

polygon-active: 提案のアクティブラーニングを利用した従来法

click-active: 提案手法

(a) : クリックベースの方法は、従来法と比較して注釈に必要な時間を大幅に短縮しました。

(b) : 提案のアクティブラーニングを利用した手法(polygon-active, click-active)は、アテンションを集めるタスクの正確性を大幅に向上させます。

<助成情報>

本研究は、以下の事業・研究領域・研究課題の支援を受けて行われました。

科学技術振興機構（JST） 戦略的創造研究推進事業 チーム型研究（CREST）

研究領域：「人間と情報環境の共生インタラクション基盤技術の創出と展開」

（研究総括：間瀬 健二（名古屋大学 数理・データ科学教育研究センター 特任教授））

研究課題：「データ駆動型知的情報システムの理解・制御のためのインタラクション」

（JPMJCR17A1）

研究代表者：五十嵐 健夫（東京大学 大学院情報理工学系研究科 教授）