

ゲーム理論で AI の視点を可視化 画像認識 AI の判断根拠を可視化する新技術

■研究の概要・リード文：

千葉大学大学院情報学研究院の川本一彦教授、計良有志助教、大学院融合理工学府博士前期課程の住安宏介氏（研究当時）の研究チームは、画像認識 AI が画像のどの箇所に注目して判断をしているのかを可視化する、新たな手法を開発しました。従来の手法では、画像ピクセル個々の重要度を考慮していましたが、本手法では画像ピクセル群全体としての重要度を測ることで、より精密な可視化が実現されています。その結果、Vision Transformer などの代表的な画像認識モデルはほとんどの画像において、わずか 4% から 16% 程度の画像領域のみから画像中の物体等を正確に認識することが可能であるとわかりました。この技術により、画像認識 AI の判断根拠をより正確に可視化し、AI 利用の信頼性を高めることが期待されます。

本研究成果は、2024 年 6 月 17 日からシアトル（アメリカ）で開催される、コンピュータビジョン分野の最上位国際会議 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) ^{注1)} で発表されます。

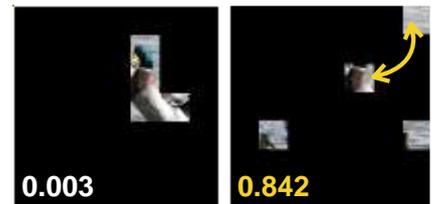
■研究の背景：

画像認識 AI 技術により、画像や動画内にある物体の種類の判別、顔認識、歩行者の検出などの高度な認識が可能になっており、広く応用されています。そのような状況で、AI の認識において画像中のどの箇所が、どの程度重要かを可視化する技術（図 1）は、AI の誤作動等の防止・理解の上で非常に重要です。

従来の可視化手法は、画像中の各ピクセルが認識に与える影響（各ピクセルの貢献度）を計算し、貢献度が上位のピクセルを可視化していました。しかし、貢献度の高いピクセルを集めることがピクセル群全体としての貢献度を最も高めるとは限りません。例えば、海鳥の認識において、海鳥という主情報が最も重要なのはもちろんですが、背景に映り込む海も補助的な情報を持ちます（図 2）。従来法では、海鳥の体やくちばしといった主情報が可視化され、背景等の補助情報はあまり可視化されませんでした。



元画像 「ブルマスティフ」
図 1：画像認識 AI の判断根拠の可視化

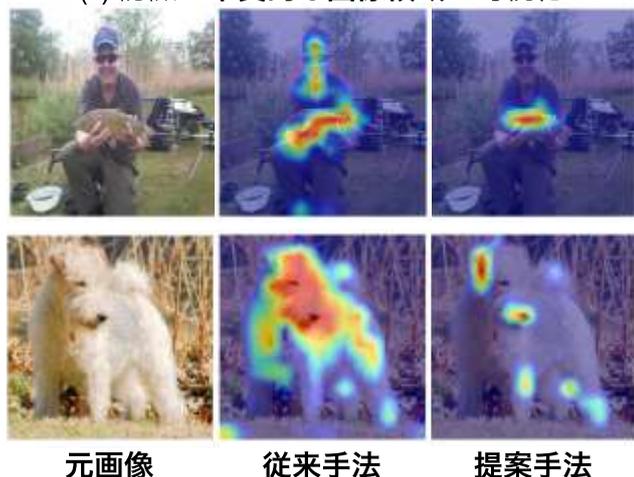


海鳥のみ 海鳥と海
図 2：画像の一部のみを見せた時の画像認識 AI の確信度。海鳥と海の両方を見せることが重要だとわかる

■研究の成果：

本研究では、画像認識 AI 技術の判断根拠を可視化する、新たな技術を開発しました。従来手法が画像内の個々のピクセルの貢献度を計算するのに対し、提案手法では画像ピクセルの集合としての貢献度を計算しています。より具体的には、ゲーム理論^{注2)} で用いられる**シャープレイ値**^{注3)} と**相互作用値**^{注4)} という 2 つの量を画像に対して計算することで、個々のピクセルの貢献度に加えてピクセル間の協調による貢献度を考慮している点が技術的な要点です。ただし、**シャープレイ値**と**相互作用値**という量は、計算にあたり非常に大きな計算コストを必要とします。そこで貪欲法^{注5)} による近似的な最適化を考案し、厳密な計算法を利用した場合と比較して、画像 1 枚の可視化あたり 100 倍程度の高速化に成功しました。従来手法と比較して、提案手法はより精密な可視化が可能です（図 3）。

(a) 認識に本質的な画像領域の可視化



(b) 認識に必要な最小限な画像領域の可視化

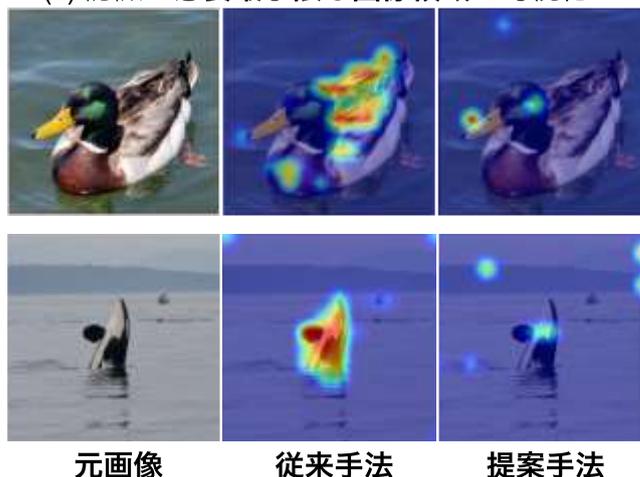


図3：提案法と従来法（Attention Rollout）の比較。(a) 認識に本質的な画像領域の可視化。この部分を隠すと画像認識 AI は対象を認識できないという領域。(b) 認識に必要な最小限な画像領域の可視化。この部分のみで画像認識 AI は対象を認識できるという領域。両方の場合において、従来手法は物体領域を過剰評価しているが、提案手法はより小さな領域で認識可能であることを明らかにしている。

■今後の展望

本研究により、画像認識 AI 技術の判断根拠をより正確に可視化することが可能になりました。今後は、自動運転における歩行者や標識の検出の信頼性を高めるなど、安心・安全な AI を実現するための応用が期待されます。

■用語解説

注 1) IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): AI 分野では研究トレンドの変化が迅速であるため、同分野では論文誌における発表を待たず、プレプリントサーバ（例：arXiv）による速やかな成果の公開、続いて CVPR のようなトップ国際会議への投稿・発表という流れが主流。中でも CVPR はコンピュータービジョン分野における世界最高峰の国際会議である。Google Scholar が提供する h5-index による 2023 年のランキングでは、全学術分野で 4 位（1 位 Nature、2 位 The New England Journal of Medicine、3 位 Science、4 位 CVPR）に位置し、その学術的影響力は非常に大きい。

注 2) ゲーム理論：複数の主体（プレイヤー）が協力的・非協力的に意思決定を行う状況（ゲーム）において、プレイヤーの行動やゲームの展開・帰結を数理的に分析するための理論およびその体系。

注 3) シャーププレイ値：ゲーム理論で提案された、プレイヤーの貢献度の指標の一つ。複数人が協力して得た報酬から、各個人に対しその貢献度に応じて公平に報酬を分配する目的で用いる。あるプレイヤーの貢献度は、ゲームにおいてそのプレイヤーがいた場合といなかった場合の報酬の差から計算される。

注 4) 相互作用値：ゲーム理論で提案された、2 プレイヤーの共同貢献度の指標の一つ。ある 2 プレイヤーの共同貢献度は、ゲームにおいてその 2 プレイヤーが同時に存在した場合と、片方しか存在しない場合との差から計算される。

注 5) 貪欲法：最適化手法の一つ。ある問題に対して報酬を最大化するにあたり、場当たりに最適な選択を進めていく手法。多くの場合、（最適とは限らないが）比較的良好な結果を効率的に得ることができる。

■研究プロジェクトについて

本研究は下記の事業の支援を受けて実施されました。

- ・ 科学研究費助成事業 基盤研究（B）「深層学習における敵対的デザインの開拓」（JP22H03658）
- ・ 科学研究費助成事業 若手研究「深層ゼロ関数学習の計算と理論」（JP22K17962）

■論文情報

タイトル : Identifying Important Group of Pixels using Interactions

著者 : Kosuke Sumiyasu, Kazuhiko Kawamoto, Hiroshi Kera

雑誌名 : Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024

DOI : <https://arxiv.org/abs/2401.03785>

<研究内容について>

千葉大学大学院情報学研究院 助教 計良 宥志

TEL: 043-290-3288 メール : kera@chiba-u.jp

<報道担当>

千葉大学広報室

TEL: 043-290-2018 メール : koho-press@chiba-u.jp