ニュースリリース

令和５年７月３日

報道機関　各位

国立大学法人　電気通信大学

ヒトは複雑な画像特徴量に、より速く、より多く視線を向けることを発見

【ポイント】

※AI 技術によって画像の特徴を複雑さに分けて分析

※それぞれの特徴ごとの注視誘引度（見られやすさ）を空間と時間に分けて評価

※複雑な特徴を持つ場所はより速く、高頻度に見られやすいことを発見

【概要】

電気通信大学情報理工学研究科の赤松和昌研究支援員、西野智博氏（当時学部生）、宮脇陽一教授らは、ヒトは複雑な特徴を持つ場所をより速く、より高頻度に見る傾向にあることを初めて実験的に発見しました。

私たちは多種多様な物体にあふれた生活環境のなかで、時々刻々と目を動かして色々なものに視線を向け、視覚的に情報を取得しています。では、私たちはどのような場所を優先的に見るのでしょうか？どのような場所が見られやすいかを予測する研究は従来から盛んに行われており、視線予測のための様々な計算プログラムが提案されています。しかし、従来の研究では視線を予測することが主目的となっている研究が多く、視線が向けられた場所の特徴自体がどのくらい影響を持っていたのか、そしてどのような種類の特徴が優先して見られやすいのかは分かっていませんでした。

こうした問いに答えるため、私たちの研究グループでは、AI 技術のひとつである深層ニューラルネットワークモデルを用いて、自然なシーン画像のどこに、どの程度の強さで、どのような種類の特徴量が分布しているのかを定量化しました。次に、その画像を観察した際のヒトの視線を計測することで、複雑な画像特徴量は単純な画像特徴量と比べてより速くより高頻度に視線を惹きつけることを発見しました。

本研究の成果は Scientific Reports 誌に 2023 年 5 月 19 日（日本時間）に掲載されました。

【背景】

私たちは多種多様な物体にあふれた生活環境において時々刻々と目を動かして色々なものに視線を向け、視覚的に情報を取得しています。ヒトの目の解像度は一様ではなく視野の中心が最も高いことが知られています。視線を向けるということは、その場所を視野の中心で見るということになります。すなわち視線を向ける場所は、おそらくヒトが視覚的に情報を取得するうえで重要な場所であると考えられます。では、ヒトが視線を向けやすい場所はどのような特徴をもっているのでしょうか？

これまで様々な先行研究においてヒトが視線を向ける場所を予測する計算機プログラムの開発がおこなわれてきました。こうしたプログラムは、単純な明るさのコントラストやエッジが際立っているかという特徴量[1]を用いたものから、深層ニューラルネットワークモデル[2]が獲得した複雑な特徴量を用いたものまで種々存在しています。しかし、これらのプログラムは特徴量とヒトが視線を向けた位置についてのデータを組み合わせて学習することで、視線を向ける場所の予測をしています。こうした手法では、視線予測のプログラムの学習が柔軟

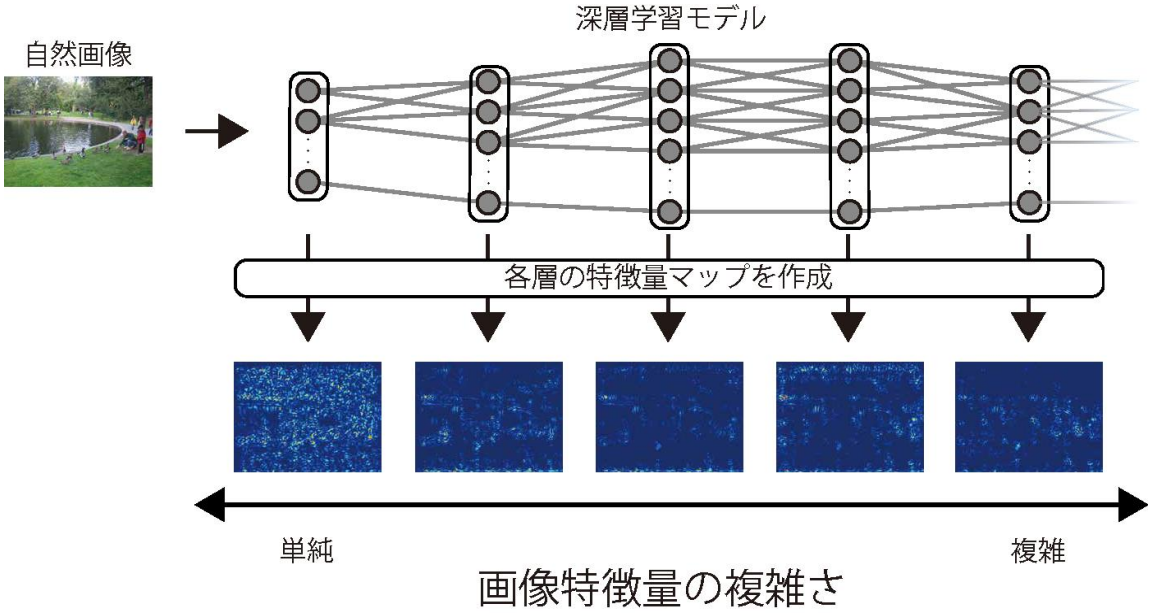

図 1 特徴量マップの作成。自然なシーン画像を深層ニューラルネットワークモデルに入力し、それぞれの層に対応する画像特徴量が画像中のどこにどの程度分布しているのかを表す特徴量マップを作成した。深層ニューラルネットワークモデルでは、層が入力側に近いほど単純な画像特徴量に対応し、入力側から遠くなるほど複雑な画像特徴量に対応する。

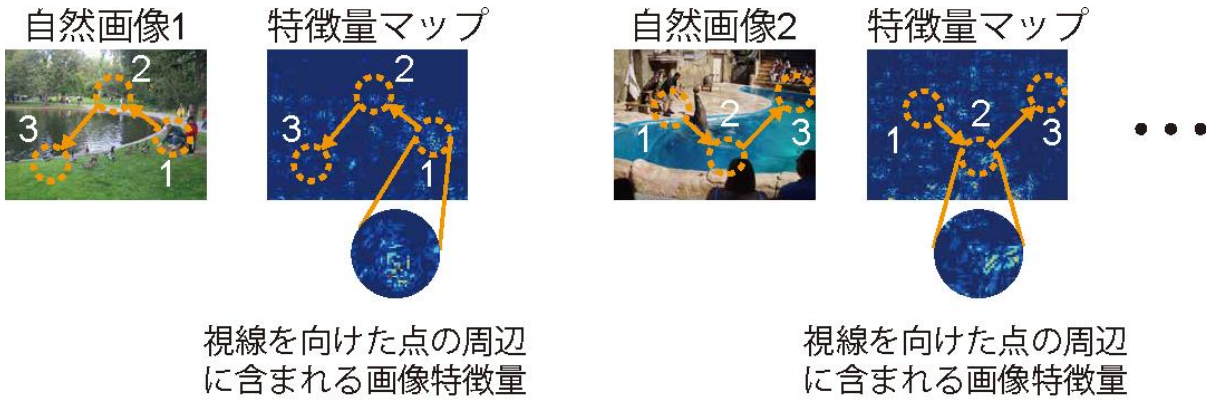

図 2 注視誘引度の計算方法。視線を向けた点の周辺にどのような画像特徴量がどのくらい含まれているのかを計算する。これを画像観察中のすべての視線に対して計算したのち、すべての提示画像に対しての結果を平均することで、どのような種類の画像特徴量がいつ、どの程度よく見られていたかを求めた。

すぎるため、その特徴量に本当に視線が惹きつけられたのか、あるいはプログラムが柔軟に学習してしまったために結果として視線が予測できてしまっているのかの区別がつかない状態でした。このため、どのような特徴量に視線が向けられやすいのかは直接的に検証されていませんでした。

【手法】

こうした問いに答えるため、本研究では深層ニューラルネットワークモデルを用いて段階的な複雑度をもつ特徴量を画像から抽出し、その画像を見ている時のヒトの注視を調べることにしました。深層ニューラルネットワークモデルは多階層からなっており、各階層では複雑度の異なる画像の特徴量に反応することが知られています。この性質を利用して、まず実験でヒトに見てもらう自然なシーン画像[3]を深層ニューラルネットワークモデルに入力し、そのときの各層の反応を、層を遡るように逆にたどることで、入力した画像中のどこにどの程度の強さで反応の元になった特徴量が分布しているのかを可視化する図（図 1、これを特徴量マップと呼びます）を作成しました。これらの特徴量マップとして、明るさのコントラストやエッジのような単純な画像特徴量に対応する第 1 層から、複雑な画像特徴量に対応する第 5 層までの 5 種類を作成しました。加えて、従来から視線の予測

に使われていた顕著性[4]と呼ばれる画像特徴量も比較に用いました（ここでは顕著性は、明るさのコントラスト、色のコントラストとエッジの鮮明さから計算するものとしました)。

次に、自然なシーン画像 590 枚を実験参加者に提示し、画像を自由に観察している間の視線を視線計測装置[5]で計測しました。実験データは成人 20 名から取得しました。このデータに対して、視線を向けた点の周辺の小領域にどのような画像特徴量がどのくらい含まれているのかを計算しました（図２)。これを画像観察中のすべての視線の移動先に対して求め、さらにすべての画像に対して同じ処理を施したのち、それらの結果を平均しました。これにより、どのような画像特徴量がどの時点でよく見られるのかを求めることができます。これを注視誘引度と定義しました。注視誘引度は、画像特徴量の種類によって変わるので、それぞれの条件ごとに分けて注視誘引度の時間変化を解析しました。

【成果】

注視誘引度の時間変化の解析結果から、深層ニューラルネットワークの対応する層が深くなるにつれて、すなわち画像特徴量が複雑になるにつれて、画像観察中の時間全体的を通して、平均的に高頻度で視線が向けられることがわかりました（図３)。また、画像提示後すぐの時間では、特に視線がよく向けられる傾向も強くなっていくことがわかりました（図３)。そして、本研究で検証した深層ニューラルネットワークの最も深い第５層おいては、この傾向が最も顕著であることがわかりました。従来から視線の予測によく使われる画像特徴量である顕著性と比較しても、第５層に対応する画像特徴量のほうが速く、よく見られることがわかりました（図３)。

以上の結果より、単純な画像特徴量よりも複雑な画像特徴量はより速く、より高頻度に視線が向けられやすいことを初めて実験的に立証することに成功しました。特定の物体の種類ではなく複雑な画像特徴量が注視されやすいことから、個人の興味関心ではなく反射的・無意識的に視線を向けている可能性が示唆されています。

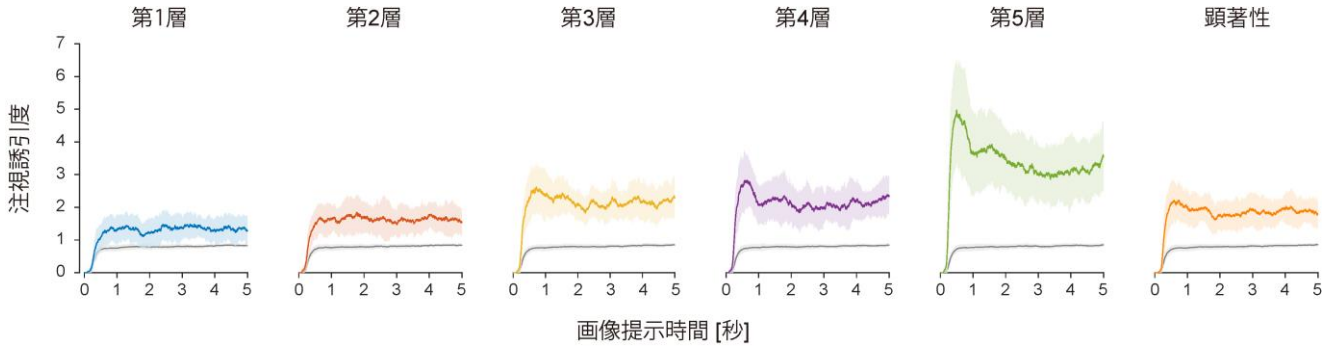

図３　各画像特徴量に対する注視誘引度の時間変化（横軸の 0 は画像提示開始時刻を表す)。深層ニューラルネットワークの第５層に対応する画像特徴量の注視誘引度は、画像提示直後に大きなピークを示し、かつ全時刻にわたって平均的に高い。すなわち、深層ニューラルネットワークの第５層に対応する画像特徴量は、より速く、より高頻度に見られる、ということを示している。

【今後の期待】

複雑な画像特徴量は単純な画像特徴量と比べて、速く、高頻度に注視されやすいことがわかりました．応用的な観点からは、複雑な画像特徴量がより多く含まれるように人工的に操作した画像を生成することによって、所定の場所へと注視を誘導できる可能性があります。例えば、道路標識のデザインであったり、使いやすい見た目のユーザーインターフェースであったり、効果的な広告の作成に役立てることができるかもしれません。

また基礎的な観点からは、複雑な画像特徴量を注視しやすいという現象がヒトの脳内のどこでどのように生じているのかを調べることが重要です。通常、複雑な特徴量は多段の処理を経て形作られるものなので、認識されるまでの時間がかかるはずですが、なぜ素早く視線が誘引されるのかはまだ分かっていません。こうした問題にさらに挑戦することによって、ヒトの視覚情報処理と視線移動の仕組みがより詳しく解明されると期待されます。

論文結果紹介動画 URL:　https://www.researchsquare.com/article/rs-3121604/v1
https://www.youtube.com/watch?v=yfyXGewId5Y
https://vimeo.com/835593892

（論文情報）

Kazuaki Akamatsu, Tomohiro Nishino, Yoichi Miyawaki, "Spatiotemporal bias of the human gaze toward hierarchical visual features during natural scene viewing," Scientific Reports, 13, 8104 (2023).

（外部資金情報）

本研究は、JST さきがけ（JPMJPR1778)、科研費基盤研究 A（20H00600)、科研費国際共同研究加速基金(国際共同研究強化(A)）（18KK0311)、科研費基盤研究 B（17H01755）および矢崎科学技術振興記念財団の支援を受けて実施されました。

（用語説明）

[1] 特徴量　ここでは画像中の特徴量である画像特徴量のことを指す。画像特徴量とは、明るさ、色、それらの隣接領域間での差異、線分要素の傾きなど、画像を特徴づける量のことを指す。

[2] 深層ニューラルネットワークモデル　多数の層を持つニューラルネットワークモデル。本研究では、物体の写っている画像を物体の種類ごとに分類できるように訓練された深層ニューラルネットワークモデルを利用した。

[3] 自然なシーン画像　屋外や屋内の日常的なシーンを撮影した画像。本研究では ADE20K（Zhou et al., 2017）および PASCAL-Context（Mottaghi et al., 2014）と呼ばれる大規模なシーン画像データベースから多数の物体が写っている画像を選定した。

[4] 顕著性　英語では saliency と呼ばれる画像特徴量のひとつ。定義はいろいろあるが、代表的な定義は Itti et al. (1998)の論文中で導入された、明るさ、色、線分方位の局所的なコントラストに基づくものである。本研究で用いている顕著性はこの定義に従っている。なお視線は顕著性の高いところに惹きつけられるという通説が広く受け入れられている。

[5] 視線計測装置　視線を計測する装置のことであり、本研究では赤外線を眼に放射して、その反射光を計測し、その結果から視線位置を計算する光学式システムを用いている。

【連絡先】

＜研究内容に関すること＞
電気通信大学　大学院情報理工学研究科
【職名】教授
【氏名】宮脇　陽一
Tel:　042-443-5982　　E-Mail: yoichi.miyawaki@uec.ac.jp

＜報道に関すること＞
電気通信大学　総務企画課　広報係
Tel:　042-443-5019　Fax:　042-443-5887
E-Mail kouhou-k@office.uec.ac.jp