

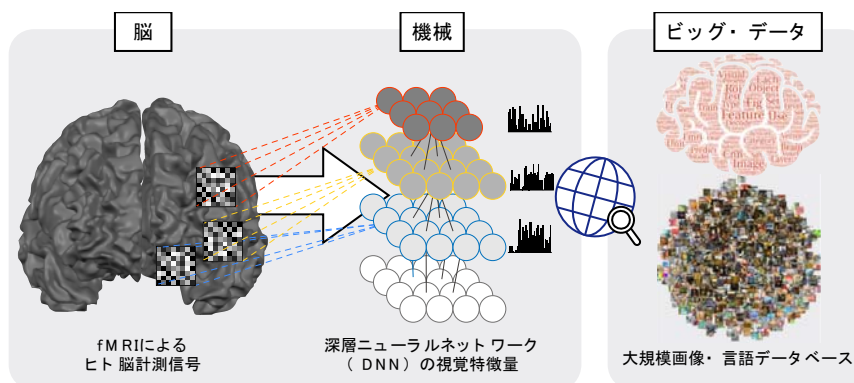


脳から深層ニューラルネットワークへの信号変換による脳内イメージ解読 ～「脳-機械融合知能」の実現に向けて～

神谷之康 株式会社国際電気通信基礎技術研究所（略称 ATR）脳情報総合通信研究所 神経情報学研究室 客員室長（京都大学情報学研究科 教授）と堀川友慈主任研究員は、ヒトの脳活動パターンを深層ニューラルネットワーク（deep neural network model, 以下、DNN）等の人工知能モデルの信号に変換して利用することで、見ている画像に含まれる物体や想像している物体を脳から解読する技術の開発に成功しました。本研究成果は、人工知能の分野で進展が著しい DNN をヒトの脳と対応づけることで、脳からビッグ・データの利用を可能とする先進的技術です。

機能的磁気共鳴画像法（以下、fMRI）等により計測されるヒトの脳活動パターンを機械学習によるパターン認識で解析することで心の状態を解読する技術は「ブレイン・デコーディング」と呼ばれ、当研究グループが世界に先駆けて開発してきました。しかし、脳活動からその時に見ている物体を解読する従来の方法では、あらかじめ脳活動を計測して機械学習モデルをトレーニングした少数の物体カテゴリーしか対象にできませんでした。一方、DNN は、脳の基本素子であるニューロンやシナプスにヒントを得て作られた人工ニューラルネットワークの一種で、近年ではヒトと同レベルの物体認識精度を達成しています。本研究では、画像を見ている時のヒトの脳活動パターンと、同じ画像を入力としたときの DNN の信号パターンに間に相同性を発見し、脳から DNN への信号変換を利用して、任意の物体を脳活動から解読する技術を開発しました。

本研究成果は、人工知能モデルをヒトの脳と対応づけ利用する「脳-機械融合知能」の可能性を切り拓くものです。ブレイン・デコーディングへの応用だけでなく、脳型人工知能の開発にも貢献することが期待されます。本研究は 5 月 22 日 10 時 0 分（英国時間）、英国の科学雑誌 *Nature Communications* 誌に掲載されました。



「脳-機械融合知能」の概念図

ヒトの脳と深層ニューラルネットワークを対応づけることで、脳からビッグ・データの利用を可能とする。

1. 背景

物体認識は、ヒトにおいてもコンピュータにおいても重要な課題であり、ヒトの脳における物体認識メカニズムの研究と、コンピュータによる物体認識精度向上の試みは、相補的に発展してきました。ヒトやサルなどの脳の視覚野の研究では、見ている物体の情報は、脳の階層的な処理を経て、高次視覚野において、物体を構成する要素的な特徴の組み合わせにより分散的に表現されていることが示唆されていました。また、近年ヒトと同レベルの物体認識精度を達成することに成功した DNN においても、入力画像を階層的に処理することで、各階層において段階的な複雑さを有する視覚特徴が獲得されていることが知られており、脳と DNN の情報表現の類似性が広く議論されるようになってきています。

ヒトの脳における物体認識の仕組みは、脳の活動を計測した信号を機械学習によるパターン認識で解析することで心の状態を解読することを可能にする「ブレイン・デコーディング」技術によっても調べられてきました。これまでの研究では、fMRI 等により計測されるヒトの脳活動パターンから、どのような物体を見たり想像したりしているかを予測できることが示されていました。しかし、従来の方法では、あらかじめ脳活動を計測して機械学習モデルをトレーニングした少数の物体カテゴリーの情報しか予測の対象とできず、任意の物体カテゴリーの情報を予測することはできませんでした。モデルの学習に用いていない情報への予測を扱う問題は「ゼロショット・ラーニング (zero-shot learning)」として定式化されており、機械学習の分野においても重要な課題とされています。

そこで本研究では、脳と DNN の情報表現の類似性を利用することで、モデルの学習に用いていない物体カテゴリーを含む任意の物体カテゴリー情報を脳活動から予測する技術（一般物体デコーディング）の開発を行いました（図 1）。

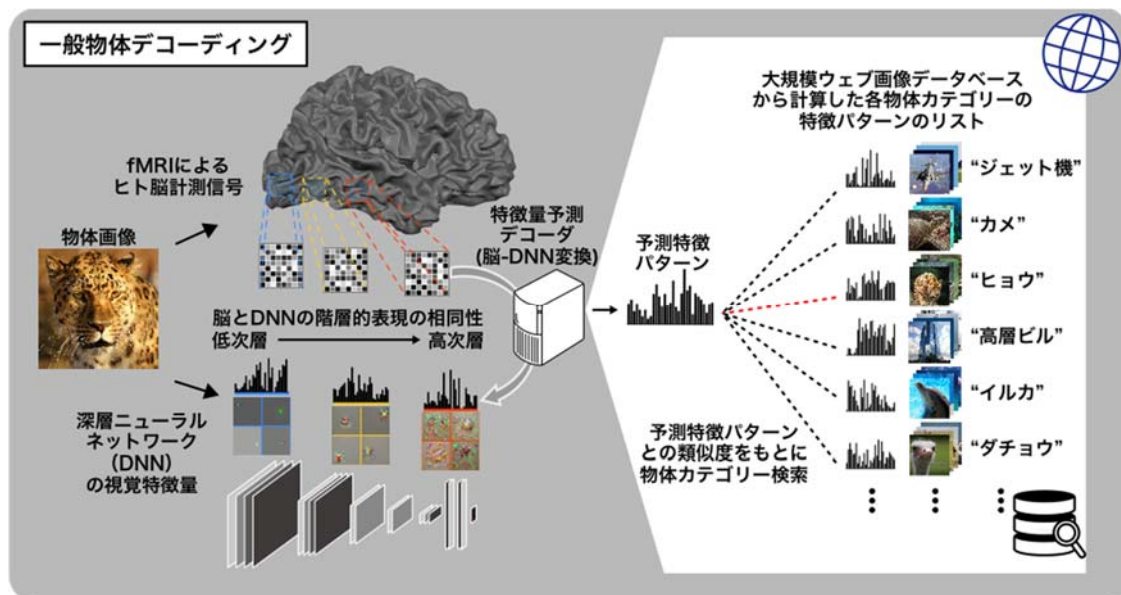


図 1 : 「一般物体デコーディング」の概要図。物体画像を深層ニューラルネットワーク (DNN) による階層的視覚特徴パターンとして表現し、見ている物体画像の視覚特徴パターンを脳活動から予測するよう学習したモデル (デコーダ) を用いて、脳活動から特徴パターンへの予測 (変換) を行う。予測特徴パターンと大規模画像データベースの画像群から計算した各物体カテゴリーの特徴パターンを照合することで、任意の物体カテゴリーの予測を行う。

2. 研究手法・成果

本研究では、任意の物体カテゴリーの予測を可能にするため、コンピュータ・ビジョンの分野で用いられている“物体概念を視覚特徴のパターンによって表現する”という考え方を採用し、画像を DNN で処理することで得られる視覚特徴量のパターンによって、物体画像を表現しました。そして、ブレイン・デコーディングを応用して、被験者が物体画像を見ているときに fMRI で計測した脳活動パターンから、DNN の信号の予測（脳から DNN への変換）を行う機械学習モデル（DNN 特徴デコーダ）を学習させました。学習済みの DNN 特徴デコーダによって脳活動から予測された特徴パターンと大規模画像データベースから得られる物体カテゴリーと視覚特徴パターンのリストを、パターン間の類似度をもとにマッチングすることで、画像を見たり想像したりしているときの脳活動から、任意の物体カテゴリーの解読（検索）が可能になります。

このアプローチの有効性を検証するため、われわれは第一に、脳活動から DNN の特徴量を予測できるかどうかを確認しました。脳の視覚野の異なる部位の活動を使ってデコーダの学習を行ったところ、低次視覚野の活動を使って学習したデコーダは、DNN の低次層の特徴量をよく予測でき、逆に高次視覚野の活動を使って学習したデコーダは、高次層の特徴量をよく予測できることがわかりました（図 2）。この結果から、脳と DNN の間の相同性（ホモロジー）を確認することができました。

さらに、デコーダを、物体を想像中の脳活動に対して適用すると、DNN の高次層の特徴量を予測した成績のピークのタイミングが、低次層の予測成績のピークのタイミングに先行するという結果が得られました（図 3）。この解析では、物体画像を想像しているときに、想像中の物体の視覚特徴が実際に画像を見ているときと共通の脳活動パターンによって表現されていることが明らかにされるとともに、脳が

脳の低次/高次の部位から DNN の低次/高次層の予測成績が高い

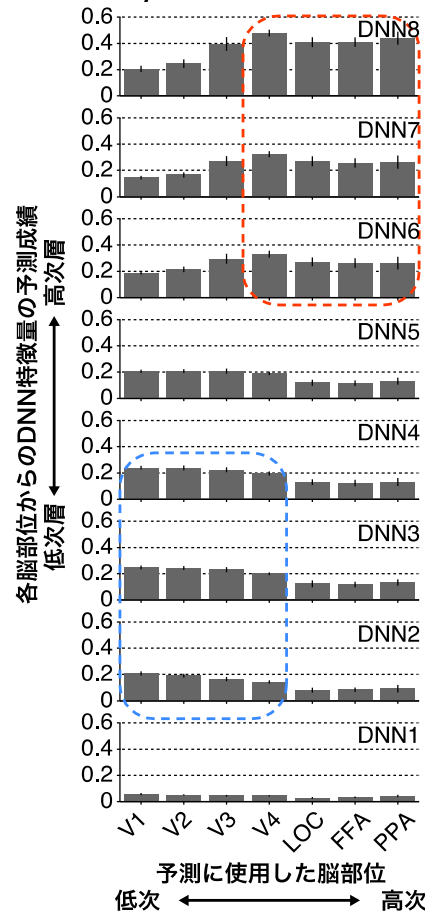


図 2：視覚野の各脳部位から DNN の各階層の視覚特徴量を予測した結果。低次脳部位からは低次視覚特徴、高次脳部位からは高次視覚特徴をよく予測できている。

高次層から低次層に漸進的に予測成績のピークがあらわれる

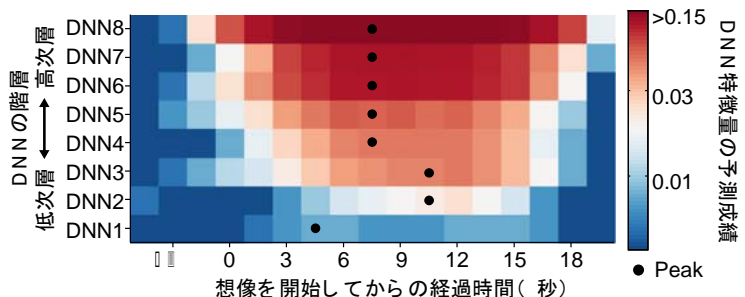


図 3：物体を想像中の脳活動から視覚特徴量を予測した結果。DNN の高次層の予測成績のピークが、低次層の予測成績のピークに先行している。

高次から低次の順にトップダウン的に階層的特徴表現を活用している様子を可視化することに成功しました。

これらの結果をもとに、提案するアプローチを用いて、ランダムに選択した 1000 個の物体カテゴリから、知覚・想像している物体カテゴリを正しく予測（検索）できるか検証した結果、正解の物体カテゴリが検索ランキングの上位にランクインするという結果が得られました（図 2 右下）。また、予測が外れた場合であっても、正解のカテゴリと意味的によく似たカテゴリを予測できていることがわかりました（例：“duck”が正解の場合に他の鳥類のカテゴリが予測として与えられる）。

本研究により、モデルの学習に用いていない任意の物体に関しても、ヒトが物体を知覚・想像しているときの脳活動から物体のカテゴリ情報を予測することが可能になりました。さらに、脳の視覚野と DNN の階層的相同性が示されるとともに、物体を想像しているときには、階層的情報表現が高次から低次に向けて段階的に活用されていることが明らかになりました。

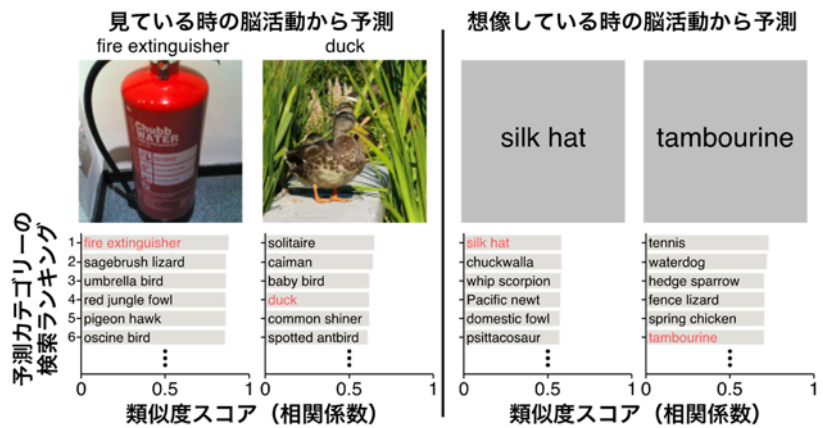
3. 波及効果、今後の予定

本研究では、ブレイン・デコーディング、DNN、大規模画像データベースを組み合わせることで、脳活動パターンから、知覚・想起している任意の物体を解釈する方法を開発しました。人工ニューラルネットワークは脳の構造にヒントを得て作られた数理モデルですが、近年では脳のモデルとしてよりも汎用的な機械学習手法として利用されてきました。今回の成果によって、人工ニューラルネットワークが再び実際の脳と対応づけられることとなり、ブレイン・デコーディングへの応用だけでなく、脳型人工知能の開発にも貢献することが期待されます。今後は、実際の脳と機械を融合する新たな知能システム（「脳-機械融合知能」）の実現を目指し、神経科学と人工知能を融合する研究を進める予定です。

4. 研究プロジェクトについて

本研究は以下の支援を受けて行われました。

- ・ 国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO) 「次世代人工知能・ロボット中核技術開発」プロジェクト
- ・ 科学研究費補助金 JP26119536, JP26870935, JP15H05920, JP15H05710
- ・ 総合科学技術・イノベーション会議が主導する革新的研究開発推進プログラム (ImPACT)



正解のカテゴリが検索上位にランクイン

図 4：知覚・想像している物体カテゴリの予測結果。正解カテゴリが検索上位にランクインしており、予測が外れた場合でも正解カテゴリに似たカテゴリの予測ができています。

<論文タイトルと著者>

タイトル : Generic decoding of seen and imagined objects using hierarchical visual features

著者 : Tomoyasu Horikawa, Kamitani Yukiyasu

掲載誌 : *Nature Communications*