SEMINAR REPORT

自然知覚の脳内情報表現

情報通信研究機構
脳情報通信融合研究センター
研究員　西田 知史　氏
主任研究員　西本 伸志　氏

私たちの日常的な体験、特に視覚体験は、非常に複雑で多様でダイナミックです。私たちは普段から眼球を通して刻一刻と変化する大量の情報を受け取っています。脳の視覚系は、そのような情報を効率的に処理することによって、外界で起きていることを認識し、最終的に柔軟な行動へと結び付けています。

別の言い方をしますと、このような日常で触れるような動画像入力、そのような入力をここでは自然動画と呼びますが、自然動画を処理することが視覚系の存在意義であると言えます。そして、究極的には、自然動画の情報処理がどのような脳機能によって成り立っているのかを定量的に理解したいというのが私の研究の動機であり、長期的な目標です。

研究の大枠

最初に研究の大枠についてご紹介します。

私達は、日常的な視覚認知を支える洗練された神経情報処理のメカニズムを明らかにするために、ヒトの脳機能イメージング実験を用いて検証を行っています。被験者に動画像を見てもらい、動画視聴中のヒト脳活動を知るために、fMRI を用いて連続的に全脳の 3 次元記録を取ります（図 1）。このような実験により、ヒトの知覚体験と脳活動の関係を定量的に結びつける予測モデルを構築します。

両者の関係をモデル化する際には、2 通りの方法が考えられます（図 2）。一つは、知覚体験の表象が脳活動としてエンコード（符号化）される過程のモデル化です。これを私たちはエンコーディングモデルと呼んでおり、脳内の表象を直接的に調べるための枠組みと捉えることもできます。最近では、脳型の人工知能を作る際の一つのお手本、あるいは答え合わせを行うための基盤技術としても注目を集めています。

図１　ヒトと自然知覚を支える脳内表象の定量理解

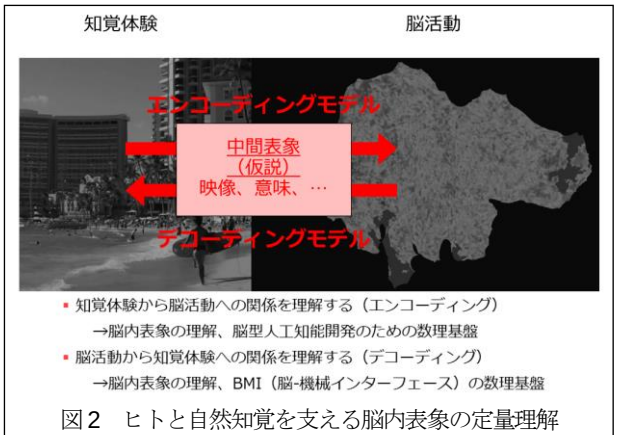

図２　ヒトと自然知覚を支える脳内表象の定量理解

もう一つは、逆方向として、脳活動から知覚体験をデコード（復号化）するモデル化です。これを私たちはデコーディングモデルと呼んでおり、表象理解に対する別のアプローチと捉えることができます。BMI（Brain-Machine Interface）といった次世代の情報伝達技術を実現するための基盤になるとも考えられています。

このようなエンコーディングあるいはデコーディングの問題

を解くことで、脳機能に関するより定量的な知見を得たいというのが私たちの研究の目指すところです。

更に具体的には、この知覚体験と脳活動の関係を上手くモデル化するための中間表象として、どのようなものが有効なのかを考えることが研究の主要な課題となってきています。

図3は、ヒトの脳です。視覚に関連する領野は後頭部に局在していて、目から入った信号は最初に初期視覚野へ到達します。その後、高次視覚野へと信号が移りながら段階的な処理が行われます。前半の初期視覚野は色、動き等の単純で具体的な属性を処理するところであって、後半の高次視覚野は物体認識、印象等の抽象的で高度な情報処理をするところです。

連続的かつ階層的に視覚情報の処理が大脳皮質の視覚系では行われており、入ってきた情報は脳内表象として具体から抽象へ、あるいは客観から主観へと変換されていきます。

この初期視覚野ないし高次視覚野、この階層性の各段階における表象を理解することで、脳のモデル化や脳内情報の解読ができるはずだと考えています。

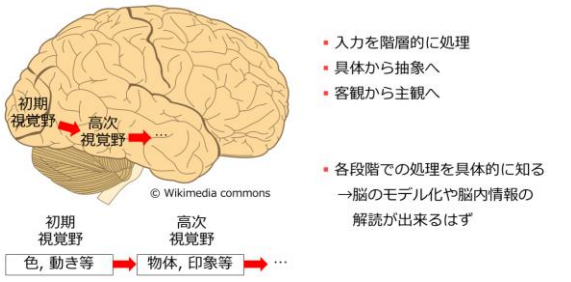

図3　ヒトの大脳皮質視覚野における階層的情報処理

研究内容の紹介

これまで取り組んできた研究について、いくつかのトピックスに分けて紹介します。

(1) 初期視覚野における時空間情報表象と知覚体験のデコーディング

一つ目のトピックは、初期視覚野における時空間情報表象と知覚体験のデコーディングという研究内容です。

研究の枠組みでお話したように、自然動画から脳活動への対応づけをモデル化するにあたり、ここでは中間表象として運動エネルギーモデルを用いました。運動エネルギーモデルとは、平たく言うと、視野内の視覚像の運動成分を取り出してくるフィルタです。脳の初期視覚野あるいは運動情報処理に関わるMT 野と呼ばれている領野における情報処理を模したモデルとなっています。図4がモデルの概要です。

このモデルでは、視野の中の様々な場所の運動情報、つまり時空間周波数情報を取り出してくる6,000個超のフィルタを用意し、これを特徴次元として定義します。入力となる自然動画を、この特徴表現に一旦落とし込んだ上で脳活動を予測するのですが、その間に重み付き遅れ時間フィルタを介します。なぜかと言うと、MRIで取り出せる信号は実際の神経活動より多少遅れて生じることが知られており、その遅れ時間をモデルに取り入れることで、より良い予測モデルの構築が可能だからです。これによって、動画から脳活動を予測するエンコーディングモデルを作り上げています。

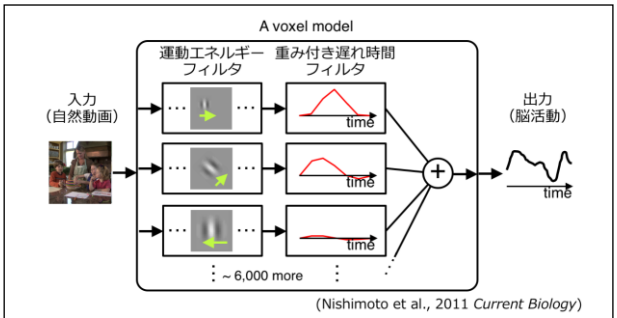

図4　単一ボクセル運動エネルギーモデル

実際、このモデルによって、初期視覚野を初めとする広い視覚領野の脳活動を予測できることが分かりました。さらに、このモデルを使ってデコーディングを行いました。同じ運動エネルギーの中間表象を使って、脳活動から知覚体験を再構成するという試みです。図5の画像が、このデコーディングによって視覚像を再構成した結果になります。左側の画像がヒトの見ている実際の動画像です。右側がそのときのそのヒトの脳活動情報だけから、デコーディングモデルを使って再構成した動画像を表しています。

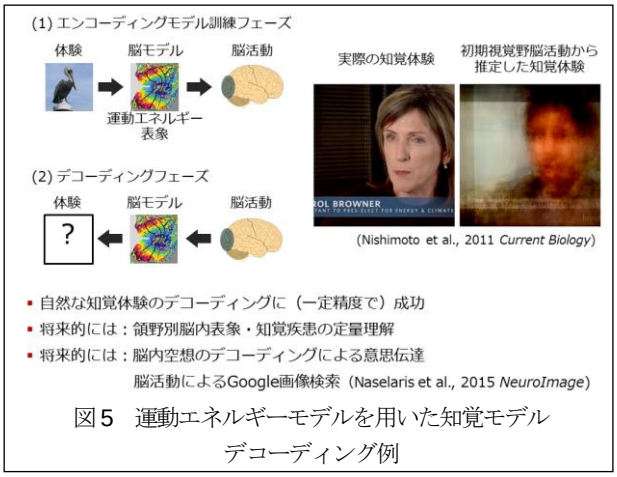

図5　運動エネルギーモデルを用いた知覚モデル
デコーディング例

見てもらえば分かると思いますが、左の画像と右の画像は構成がとてもよく似ています。これは、2011年に西本らが発表した研究成果なのですが、自然な視覚動画像の再構成を世界で初めて脳活動から可能にしたという技法で、その当時は大変注目を浴びました。2011年の時点でも、このレベルで視覚像の再構成が可能でした。

この技術はとても有用で、将来的には、例えば、視覚表象を可視化して知覚疾患の定量理解に結びつけられます。さらに、この研究では実際に視覚映像を知覚しているときの脳活動を使っているのですが、例えば、空想中の脳活動をデコーディングすると、その空想の視覚像が復元できることになります。この関連研究で私たちの共同研究者でもあったトーマス・ナセラリスという人が、「空想中の脳活動を用いたGoogle画像検索」を実現しうる研究成果を2015年に発表しています。

(2) 高次視覚野における意味空間表象と注意による意味空間のワープ

二つ目のトピックは、高次視覚野における意味空間表象と注意による意味空間のワープという研究テーマです。

先ほどの研究は、初期視覚野における情報表現に着目していたのですが、この研究は、もう少し高次の領野における意味処理に着目して、意味の中間表象を用いてモデル化を行ったものです。

この研究に主に取り組んだのは、私たちの共同研究者であるアレックス・フスという研究者です。

例えば、あなたの日常生活における食卓風景を想像してみてください。あなたの目の前にはテーブルがあり、テーブルの上にはグラスや皿や料理などが並んでいるでしょう。また、部屋の中にはテレビや窓があり、窓の外には建物や空が見えるかもしれません。このように、日常世界には様々な物体が存在しています。このような無数に存在する物体が脳内でどのように表現されているのか。これを明らかにしたいというのがここでの私たちの研究のモチベーションです。

グラスや皿のように意味的なまとまりを持った物体のカテゴリーを、私たちは意味カテゴリーと呼びます。脳におけるカテゴリーへの選択性は従来から研究されてきました。図6は、図3と同じように大脳皮質をシート状に開いたものですが、図の左側が後頭部で、初期視覚野と高次視覚野を含んでいます。

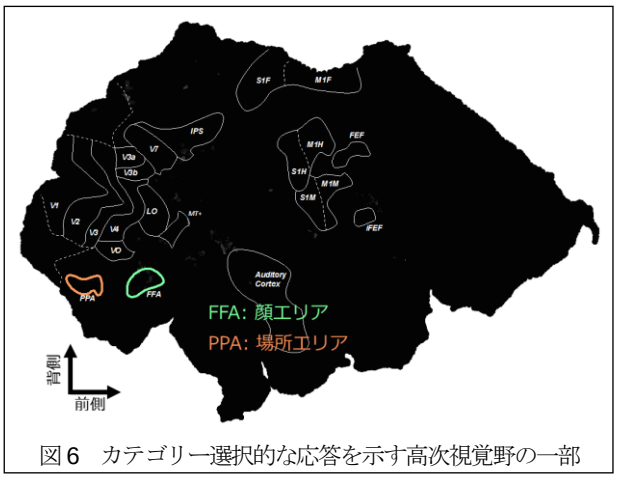

図6　カテゴリー選択的な応答を示す高次視覚野の一部

高次視覚野にはカテゴリー選択性を示す領野があって、例えば、紡錘状回顔領域（FFA：Fusiform Face Area）は、顔に関連した視覚情報に選択性を示す顔エリアと呼ばれる領野です。海馬傍回場所領域（PPA：Parahippocampal Place Area）というのは、場所に関連した視覚情報に選択性を示す場所エリアと呼ばれる領野です。顔エリア、場所エリアのような単純なカテゴリー選択性を示す領野は、今までから報告されていましたが、しかし、顔カテゴリー、場所カテゴリーというのは、自然界に存在する無数のカテゴリーの内の極々一部に過ぎません。そのようなカテゴリーだけでなく、自然界で無数に存在するカテゴリーが脳内でどのように表現されているか、どのように情報として貯えられているかを明らかにすることがこの研究の目的です。

私たちの研究では、脳内意味空間というものを提唱しました（図7）。例えば、犬という意味カテゴリーがあって、猫という意味カテゴリーがあって、この2つは似ています。そこに海というカテゴリーがあると、犬猫とは大きく離れているのが直感的に理解できまし、脳内でもきっとそのように表現されていると思います。それでは、車のカテゴリーはこれらのどこに位置するでしょうか。

私たちは、このような多様なカテゴリーが脳内でどのような距離関係をもって表現されているのかを表わす空間を可視化し、それを脳内意味空間と呼びました。

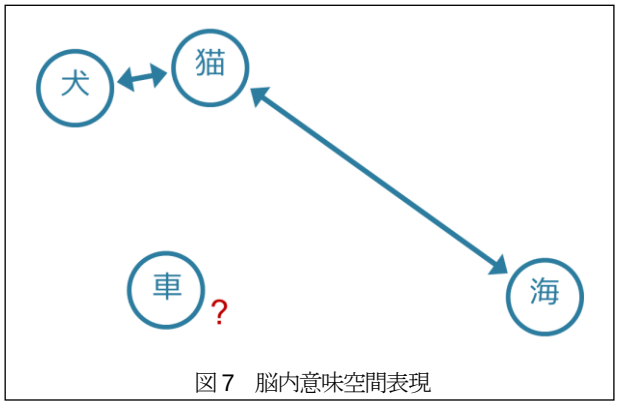

図7　脳内意味空間表現

具体的には、図8のようなカテゴリーモデルを使って脳活動のモデル化を行いました。図4とよく似た枠組みで、動画像と脳活動の対応づけを調べるために、ここでは中間表象としてカテゴリーラベルを使います。カテゴリーラベルというのは、例えば、図8に示したシーンで、Childが出ている／出ていない、Phoneが出ている／出ていないのように、1,700ほどのカテゴリーに対して、出ている／出ていないを1, 0で示す特徴表現となっています。

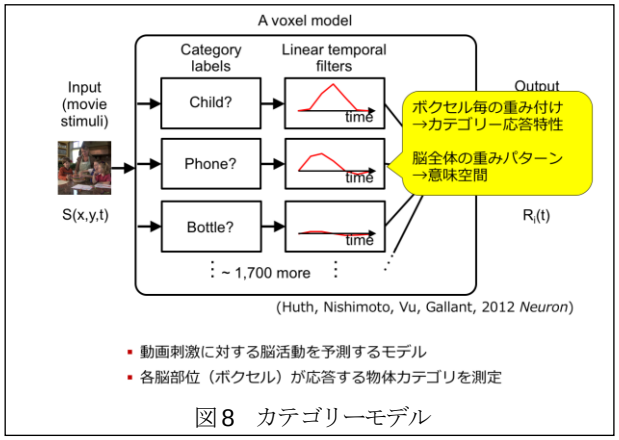

図8　カテゴリーモデル

その特徴表現を中間表象として、脳活動の時系列を予測するモデルを立てたところ、視覚野の広い領域、特に高次視覚野の脳活動を予測することができました。

このモデルにおける重み付け、つまりfMRIの計測単位であるボクセルごとのカテゴリーラベルに対する重み付けは、それぞれのカテゴリーに対するそれぞれのボクセルの選択性を示していると言えます。

そして、脳全体のボクセルの重みパターンが意味空間に対応します。図9が意味空間の可視化を行った結果です。ここでは、先ほど求めた全皮質のボクセルの重みパターンを、主成分分析と呼ばれる可視化方法を使って3次元空間で表しています。すべてのカテゴリーが脳の中でどのような位置関係、構造を持っ

て表現されているかを表す図になっています。

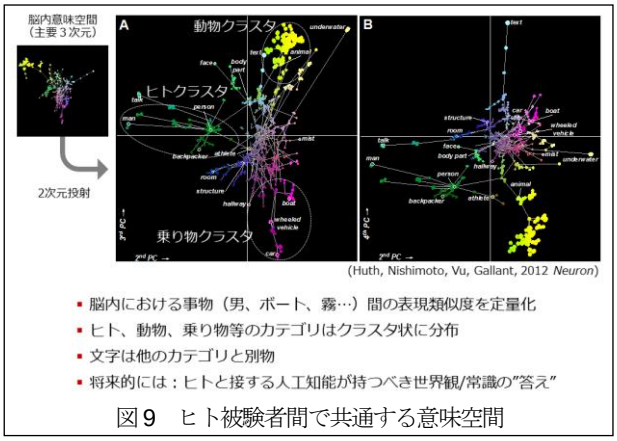

図９　ヒト被験者間で共通する意味空間

図Ａは、その内の２次元の平面を取り出しています。一つ一つの点がその一つ一つのカテゴリーを表しています。これらの点の位置は、脳活動モデルから推定した脳内意味空間におけるカテゴリーの位置と言えます。例えば、図Ａの上部には動物を表すカテゴリーが集まっています。中央左側にはヒトに関連するカテゴリー、下部には乗り物に関連するカテゴリーが集まっています。このような集合のことをクラスタと呼びます。このように、ヒトの脳内ではカテゴリーがクラスタを形成していることが分かります。

この結果は非常に興味深いことで、ヒトというのは生物学では動物の一種、動物が進化したものと考えられていますが、ヒトの意味空間の中では動物とヒトの距離はヒトと乗り物の距離と同じほど離れています。すなわち、ヒトは自分たちが動物とは違う特別な生き物であると捉えているのかもしれないということが、この分析から分かりました。

さらに、この３次元空間から別の２次元の平面を取り出しました。図Ｂでも同じようにクラスタができていますが、見て分かるように、テキストというカテゴリーが大きく外れた位置に表現されています。テキストとは文字であって、文字は私たち人間が人類の歴史、あるいは地球の歴史上で極めて最近になってから獲得したものです。故に、他の自然なカテゴリーに比べると、脳内ではとても離れたものとして表現されているのかもしれません。この意味空間の可視化によって、このようなことが明らかになります。

また、この脳内意味空間は、ヒトと接する人工知能が持つべき世界観、常識の模範を提供するものになるかもしれません。例えば、人工知能が生物学のことをよく理解していて、動物とヒトを同じように扱ったとしましょう。そうすると、その人工知能は「何て人間味のない人工知能なのだ。」というように人間に反発されるかもしれません。しかし、逆に人工知能がヒトのことを動物と違う特別なものとして扱ったのであれば、人間は「あ、こいつは人間らしい、いいやつだ。」と思うかもしれません。そういった、人工知能が持つべき常識といったものを、この脳の空間から何かしら獲得できるかもしれません。

また、カテゴリーの情報がヒトの大脳皮質でどのように表現されているかを表すマップも構成することができました（図10）。図の左端が視覚野です。その右側に高次視覚野があり、図中の色は、図９の意味空間における色と対応しています。例えば、緑色の部分はヒトに関連する部分で、この領野でヒトに関連する情報がカテゴリー表現されています。

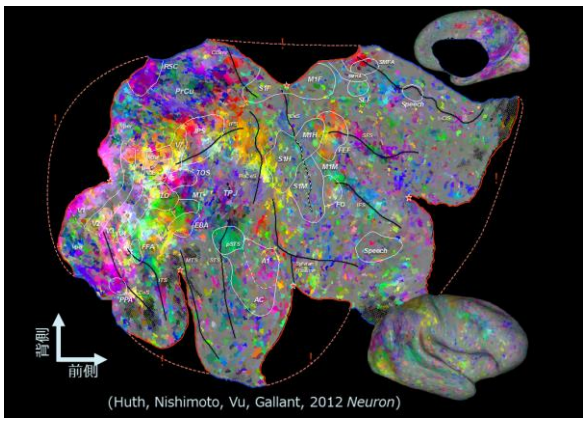

図10　意味空間情報表現の皮質分布

見てのとおり、とても広い範囲にわたってカテゴリー表現が分布していることがわかります。かつ、色が連続的に変化していることもわかります。すなわち、人間の大脳皮質では、無数に存在するカテゴリーの情報は、ある程度連続性を持って広い領域にわたり表現されていることがわかりました。

意味空間の可視化は、他の科学研究にも応用できます。視覚的注意というのをご存知でしょうか。視覚的注意は、視野内の特定の情報を選択的に抽出して処理するための、ヒトが有する認知機能の一つです。視覚的注意があるカテゴリーに向けられたとき、そのカテゴリーの表現がどのように変化するかは、今まで諸説ありました（図 11）。例えば、動画の中から「ヒト」を探索中の脳活動の意味空間を可視化すると、「ヒト」の表象が大きく広がることがわかりました(b)。すなわち、「ヒト」の方に選択性がシフトします(e)。さらに、「乗り物」を探索中の脳活動からは、「乗り物」に関連したカテゴリーの表現が大きくなることが分かりました(c)(f)。

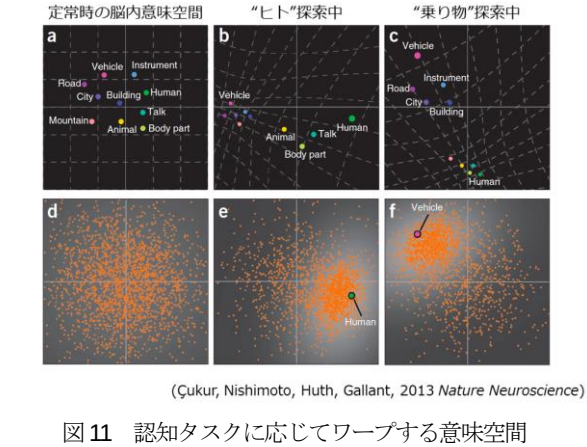

図11　認知タスクに応じてワープする意味空間

つまり、意味空間が視覚的注意によってワープすることが明らかになり、それまで諸説粉々あった視覚的注意の仮説に、いわば一つの答えを与えることができたわけです。意味空間の可視化技術は、このような科学研究のツールとしても使えることを強調しておきたいと思います。

脳と人工知能

ここからは、脳活動モデルと人工知能をどのように組み合わせていくかという話に移ります。

私たちの脳は、世界の情報を解釈して、それに働きかけるプロセスを行っています。人工知能も同様のプロセスを実行しています。これらのプロセスは、それぞれ別個に研究されていますが、お互いの情報、研究成果を共有することで、相互に理解を高めていけるのではないかと考えています。

例えば、図 12 のような自然なシーンを見たときの脳の活動にモデルを適用して、脳内における知覚内容の情報表現とその関係性を得られます。同じシーンを見ているときの人工知能の内部情報表現を、脳内の情報表現に近づけていくことで、人工知能をより人間らしく思考するものに変えていくことが可能です。

一方で人工知能は、回路を組み替えることで様々なルールに則った知覚情報処理をテストすることができます。そのテストの過程で、より自然な知覚処理を行う人工知能が持つ内部表現を用いて、人間の知覚情報表現をモデル化することで、人間の脳に対する理解を深めるという試みも可能です。私たちは、このような相互参照により、人工知能の進化と脳の理解を同時に進めていくことができるのではないかと考えています。

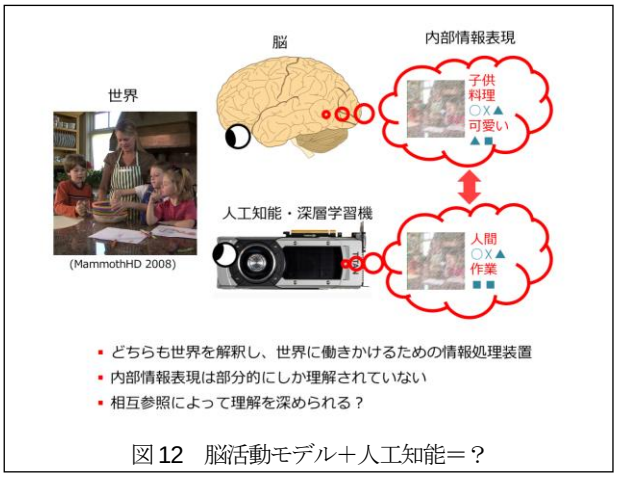

図 12　脳活動モデル＋人工知能＝？

最近、deep learning、深層学習というものが注目を浴びています。先日、Google が開発した、深層学習を取り入れた人工知能の AlphaGo が、碁の世界チャンピオンに圧勝するという快挙もありました。深層学習というのは、ニューロンの層を何重にもつなげて、例えば、画像の識別といったタスクに特化して、とても優秀な成績を収めることが知られているモデルです。

最近の研究では、深層学習の階層的な内部表現が、脳の視覚システムにおける階層的な内部表現と似ていることが報告されました。図 13 は、深層学習モデルの各階層のニューロンが、視覚経路における各階層の脳活動とよく似た表現を持つことを表しています。そして、それらは低い階層から高い階層にいたるまで、連続的な類似関係を持つことが分かりました。

深層学習の表現が脳の表現と似ているというのは、私たちにとって驚きの結果でした。深層学習は確かに脳を模したアーキテクチャを持っていますが、画像認識に特化したモデルです。しかし、人間に匹敵するパフォーマンスを示すようになるまで、動画像の認識タスクに対して最適化を行うと、そのモデルの内部表現が自動的に脳の内部表現に似てくることが分かり、それは私たちにとって興味深い発見でした。

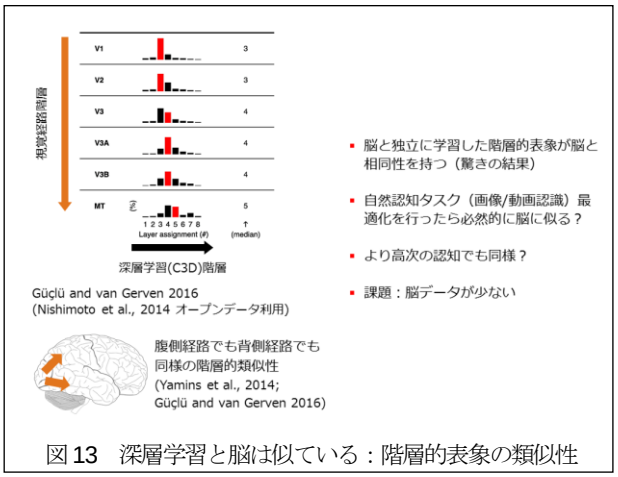

図 13　深層学習と脳は似ている：階層的表象の類似性

深層学習と脳を組み合わせた試みの一つとして、現在私たちが取り組んでいる研究を紹介します。図 14 に示しましたモデルは深層学習の一つで、画像や動画に対するアノテーション文を自動生成するというものです。例えば、図 14 左側の画像に対しては、”boy is playing baseball in the field.”のような文章が自動生成されます。

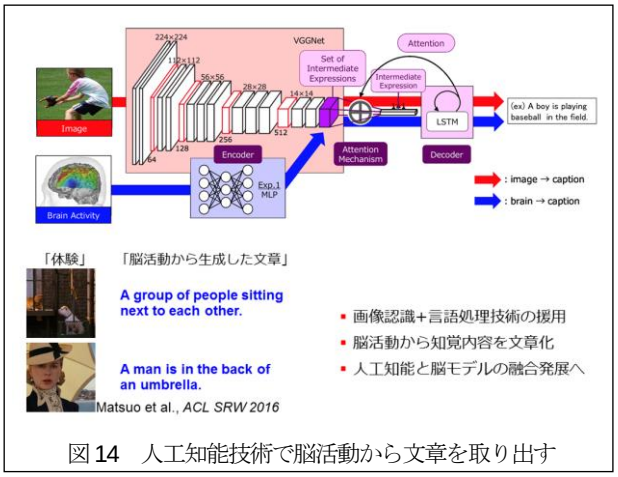

図 14　人工知能技術で脳活動から文章を取り出す

ここに脳を組み込み、脳活動から文章を生成するモデルの構築を試みました。「体験」は、実際に被験者が目にした画像です。「脳活動から生成した文章」は、このモデルを介して脳活動から推定したアノテーション文です。上の画像では、”A group of people sitting next to each other.”となっています。実際には動物であって people ではないですが、うまくシーンを記述したアノテーション文が生成されています。下の画像では、”A man is in the back of an umbrella.”と生成されて、woman のところが man となっていますが、これもシーンをうまく記述する文章となっています。脳活動と深層学習モデルを組み合わせることで、このような応用も可能となります。最終的には、人工知能と脳モデルの融合的発展へと繋げたいと考えています。

まとめ

私たちの研究における問題は、自然な知覚体験と脳画像の対応づけを獲得することです。それには、エンコーディングとデコーディングの２つのアプローチがあります。この問題の本質は、知覚体験と脳活動の間にある中間表象を定量的にモデル化することにあります。本日お見せした研究成果は、モデル化を用いることによって、脳内情報表現とその表現の皮質上マップの定量化、さらには知覚内容の可視化を行った、これまでの一連の取り組みです（図 15）。今後は、人工知能研究との融合も視野に入れながら、基礎研究から応用研究まで幅広く取り組んで行きたいと考えております。

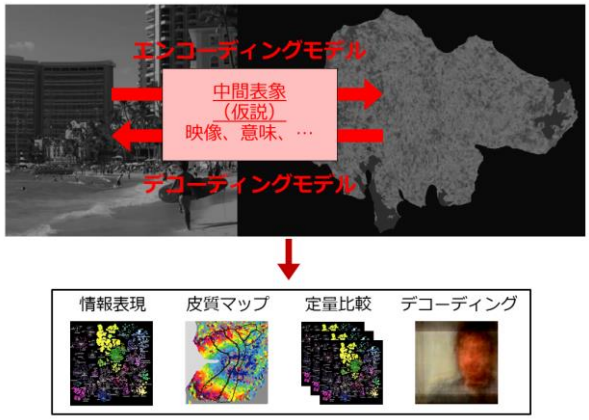

図 15　自然な知覚体験と脳画像の対応づけ

本日のお話は、主として CiNet の西本ラボで行った研究内容です。CiNet は、ヒトを対象としたシステム神経科学では国内でも有数の研究機関です。MRI 関連の研究者を多数抱えていて、最近は神経科学と人工知能の融合にも注力しています。また本日お話した研究成果の多くは、カリフォルニア大学バークレー校のジャック・ギャラント研究室との共同研究でもあります。

本講演録は、平成 28 年 7 月 15 日に開催されたＳＣＡＴ主催「第 98 回テレコム技術情報セミナー」のテーマ、「人工知能と脳神経科学との接点について」の講演内容です。

＊掲載の記事・写真・イラストなど、すべてのコンテンツの無断複写・転載・公衆送信等を禁じます。