

サル高次視覚野における物体像の表現 ～畳み込みニューラルネットは霊長類の視覚モデルになるか～

谷藤学[†]

福島邦彦のネオコグニトロンに起源をもつ Deep Convolutional Neural Network (DCNN)は、物体像のカテゴリ弁別に優れたネットワークとして着目されている (Krizhevsky, et al., 2012)。このような DCNN は、畳み込み層と局所プーリング層、規格化層から成るセットを一つの階層として、それを複数積み重ねた構造をしている。一方、霊長類の脳には、初期視覚野に畳み込み層や局所プーリング層に相当する単純型細胞と複雑型細胞があり、初期視覚野から複数の領野を経由して認識に必要な物体像の表現が高次視覚野に作られる。各領野が DCNN の各階層に相当すると考えれば、DCNN は脳における物体像処理の近似的なモデルになっているとも考えられる。実際、DCNN の物体表現層の応答特性とサル高次視覚野の物体像に対する応答特性との間には高い相関があるという報告もある (Yamins, et al., 2014)。しかし、このような見かけの類似性から DCNN は脳のよいモデルになっていると考えてよいのだろうか？いくつかの観点からこの問題について考えてみたい。

まず、霊長類の研究においては、共通の反応特性を持つ細胞が空間的にどのような構造を作っているかに注意が注がれる。例えば、初期視覚野の細胞は視野の局所にある単純な特徴（輝度勾配の向き）に反応し、それらの細胞の規則的な配列によって初期視覚野には視野のマップが作られている。一方、高次視覚野の細胞は、中程度に複雑な図形特徴に反応し、視野上の位置にはあまり依存しない。高次視覚野の中でのこれらの細胞の並び方もでたらめではないと考えられている。少なくとも行動学に意味のある物体カテゴリについて、同じカテゴリに反応する細胞は集まってクラスターを作っている。DCNN においても低次の階層には視野のマップがある。高次の物体表現層には特定の内部構造があるのだろうか？

次に、多くの研究者が高次視覚野の細胞がどのような特徴に反応しているのかという問題に取り組んできた。これらの定性的あるいは定量的な研究から、高次視覚野の細胞は物体像に含まれる中程度に複雑な図形特徴を符号化していることが示唆されている。それらの図形特徴の一つ一つは物体を特定するために十分な情報を含んでいないので、物体像は複数の図形特徴の組合せによって表現されていることになる。私たちの周りにある物体像は、照明の違い、他の物体の陰になっている、視点が違うなど様々な要因のために、同じ物体像であっても異なる画像（「見え」）として目に映る。しかし、画像として異なるにも拘わらず、私たちは正しく個々の物体を同定し、また、カテゴリを判別することができる（不変的な物体認識）。いくつかの研究から、高次視覚野における図形特徴の組合せによる物体表現は不変的な物体認識を可能にしていることが示唆されている。DCNN の物体表現層においてもノードの組合せによって不変的な物体認識が可能になっていると思われるが、それを可能にしている各ノードはどのような特徴を符号化しているのだろうか？

基本的には、現在の DCNN は自然画像の持つ統計的な性質に基づくスタティックなパターン分類である。これに対して、ヒトの行動をみると、物体認識は目的に応じてダイナミックに変化するプロセスであるように思われる。たとえば、ヒトの物体認識には空間的な注意が重要な役割を果たしている。私たちは、注意をあちこちに向けてくことで様々な物体が混在する自然画像の中で目標となる物体を探索する。また、注意を向けている場所にある物体像は、注意を向けていない場所の物体像と比較して容易に検出できることも心理学実験により示されている (Posner, 1980)。このようなヒトの行動を考えると、DCNN とは違って、霊長類の視覚情報処理はダイナミックで、その最終段に当たる高次視覚野における物体像の表現もまた目的に応じて状態依存的に変化することが示唆される。

最後に、DCNN の評価に使われたベンチマークテストでは、人の不変的な物体認識の研究で問題とされてきたことが必ずしも反映されていないことを挙げておきたい。例えば、顔の認識については、人が優れた能力を発揮することが知られているが、1000 個あるベンチマークテストのカテゴリには、「顔」というカテゴリはないらしい。ましてや、異なる向きの顔や表情の異なる顔を DCNN がうまく識別できるかは不明である。DCNN をより人間の視知覚に合わせたものに調整できるだろうか？

本講演では、霊長類、特にサルを対象とした物体視の研究を概説することで、これらの問題を考える礎としたい。物体認識研究の究極の目標は、不変的な物体像認識機構を明らかにすることにある。サルを対象とする研究は人ではできない神経細胞レベルの電気活動の記録を可能にする。また、人を対象とする研究はサルにできない行動レベルでの解析（心理実験）を容易にする。両者は相補的な関係にある。もし、機械（DCNN）が霊長類の視覚系のよいモデルになるとすれば、機械は、人・サルでは難しいアプローチを可能にするだろう。DCNN の登場によって、私たちは、新たな相補的な関係を作ることができるようになった。ただし、そのためには、DCNN を人の不変的な物体認識の研究で使われてきた課題に合わせて再調整して、不変的な物体像認識機構につながる共通の基盤を作らなければならない。DCNN のメリットは、サルや人ではできないネットワークの構造の操作が容易なことである。DCNN が霊長類のモデルになるかという問いかけより、むしろ、機械・人・サルの 3 者を相補的に連関させることで物体認識研究の本質に迫ることに期待がある。

[†] 理化学研究所 脳科学総合研究センター（理研 BSI）
Riken Brain Science Institute