

AIST ヒューマンライフテクノロジー研究部門

脳内視覚情報処理における 物体表現の理解を目指して: Deep neural networkの利用と Brain-Machine Interfaceへの応用

林 隆介

システム脳科学研究グループ
産業技術総合研究所

第9回全脳アーキテクチャ勉強会
2015年2月4日

独立行政法人 産業技術総合研究所

AIST ヒューマンライフテクノロジー研究部門

本研究の目的: 脳/神経科学の基礎と応用



- 脳/神経細胞が視覚情報をどのように処理するのか?
 - 計算論的に解明することをめざす
- 神経活動から視覚情報を解読し、制御するBMI技術の開発
 - 機能増強/障害者支援に応用

独立行政法人 産業技術総合研究所

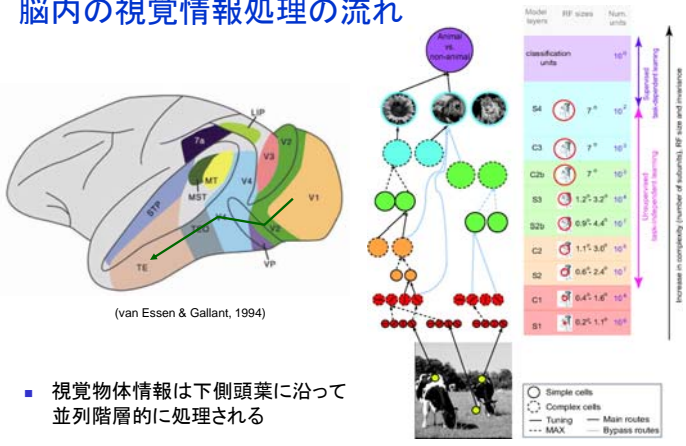
AIST ヒューマンライフテクノロジー研究部門

発表のながれ

- 脳内の視覚情報処理の概要
- Deep neural networkを用いた、
脳内の視覚情報表現の解析
- 今後の展望
 - 視覚概念の理解とメタファー
 - Brain-Machine Interfaceへの応用

独立行政法人 産業技術総合研究所

脳内の視覚情報処理の流れ



Model layers	RF sizes	Num. units
Classification units		10^4
S4	7°	10^4
C1	7°	10^4
C2b	7°	10^4
S3	$1.2^\circ \cdot 3.2^\circ$	10^4
S2b	$0.9^\circ \cdot 4.4^\circ$	10^4
C2	$1.1^\circ \cdot 3.0^\circ$	10^4
S2	$0.8^\circ \cdot 2.4^\circ$	10^4
C1	$0.4^\circ \cdot 1.6^\circ$	10^4
S1	$0.2^\circ \cdot 1.1^\circ$	10^4

- 視覚物体情報は下側頭葉に沿って
並列階層的に処理される

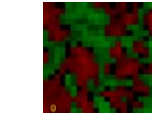
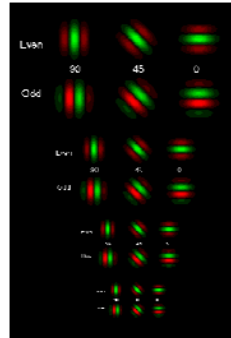
(van Essen & Gallant, 1994)

(Serre, Oliva, Poggio, 2007)

V1 (視覚1次野)の情報処理



(Hubel & Wiesel, 1963)

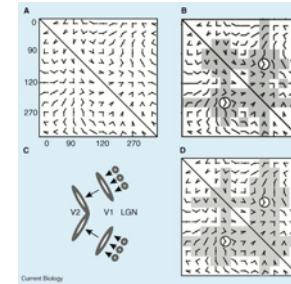


(単純型細胞の受容野)

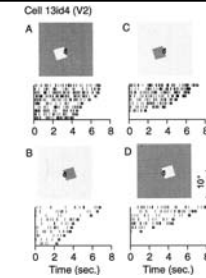
(大阪大、大澤研より)

- それぞれの網膜位置において、さまざまな「エッジの方位」および「空間周波数(縞の間隔)」を同時並列的に検出している。

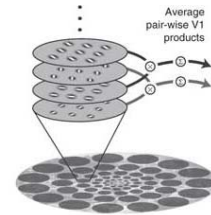
V2の情報処理



(Ito & Komatsu, 2004)



(Zhou et al., 2000)

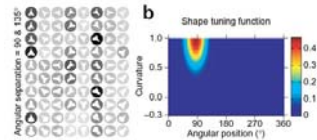


(Portilla & Simoncelli, 2000; Freeman et al., 2013)

- V2になると途端に情報表現が複雑になる
 - 線の曲率表現
 - 図と地の分化、境界の帰属判定
 - テクスチャ情報の表現

V4の刺激選択性

- より複雑な図形特徴に選択的
- 様々な曲線を使い、V4ニューロンの選択性を曲率の観点から調べた研究
- 直行座標での正弦波よりも、極座標や双極線座標での正弦波(Non-Cartesian sinusoidal grating) に選択性を示すという報告
- 実験者が選択した刺激セット内での選択性を反映しているにすぎない

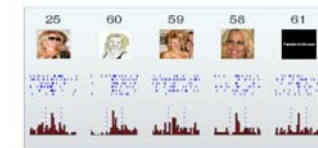
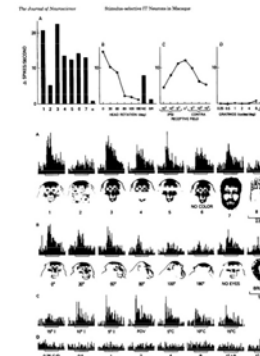


(Pasupathy & Connor, 2002)



(Gallant et al., 1996)

側頭葉のニューロン応答



■ ヒトの側頭葉内側部ニューロン (Quiroga et al., 2005)

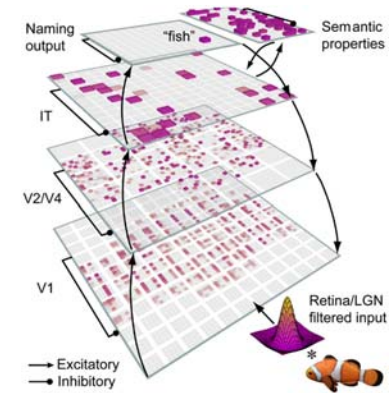
(Desimone et al., 1984)

- 顔や手だけに応答するニューロン
- 特定の物体カテゴリに対し選択的な応答

電気生理学研究の困難

- V2野以降の情報処理の理解が困難な理由
 - 神経活動が符号化する情報が複雑で高次元
 - 神経活動の時空間データに、単一の機能的意味付けができない
- 「意味の無い」脳科学のスヌメ 伊藤浩之 に詳しい
日本神経回路学会誌 巻頭言 vol.21, no.2 (2014), pp.43-49
 - ニューロンの時空間活動データに含まれる統計的/数学的性質を調べあげるといふアプローチもある。
- ヒトと同等の視覚機能を持つ人工ニューラルネットを通して、
脳の情報処理を理解するアプローチが有望と思われる。

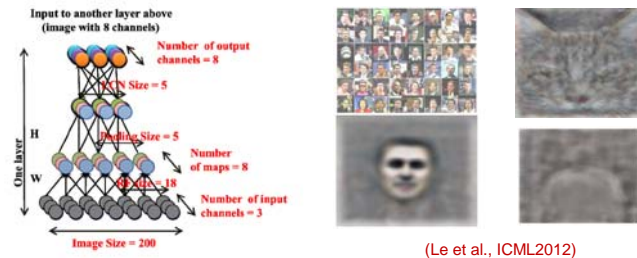
視覚情報の階層的処理のモデル化



- 概念としては昔から存在
- 従来、実際の機能をもつニューラルネットを実装できなかった

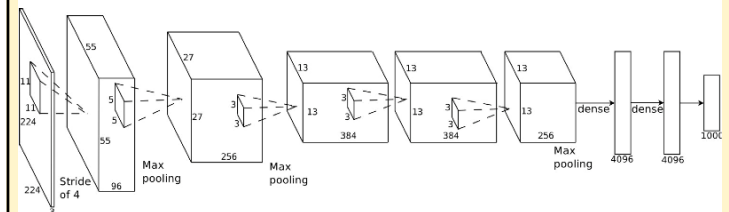
Deep Neural Netを用いた画像特徴量抽出

- Google X Labs
 - 大規模な多層convolution型 autoencoderをつかってYouTube動画を学習
 - ヒトの顔やネコの顔、ヒトの体などを表現するニューロンが形成された
 - 1600コアのクラスターを使った大規模解析 (Le et al., ICML2012)
- GPUクラスター(200万円程度)のシステムでも同等の実装が可能? (Coates et al., 2013)
- Deep convolutional neural netによる一般画像認識の精度向上 (Krizhevsky et al., 2012)



Deep neural networkによる一般画像認識

- 一般物体認識を行うDNNの実装
 - 5層のコンボリューション型演算層 (96~384 filters)
 - 3層の全結合型演算層 (Krizhevsky, Sutskever, Hinton, 2012)
 - ニューロン数65万個
 - 学習データ120万枚、テストデータ15万枚
 - 1000種類の分類課題でエラー率16%の判別精度を達成



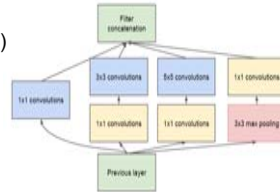
- Cuda-convnetを使い、市販のグラフィックカード1枚+Linux PCで実装可能
- Top5 error=0.200 (Top1 error=0.432)を実現 (ILSVRC2012のベストモデルとほぼ同性能)

最新のDNNによる一般画像認識

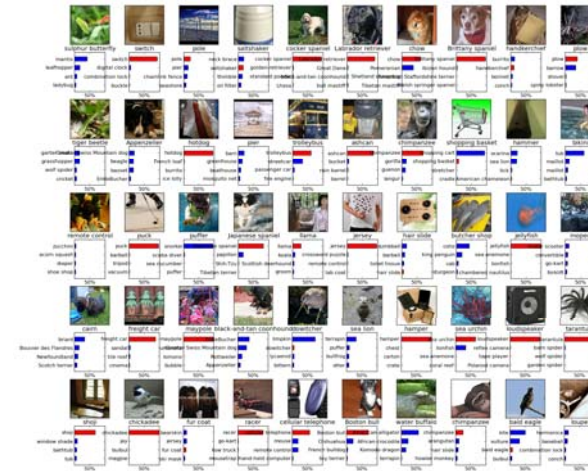


(Szegedy et al., 2014)

- ILSVRC2014の優勝モデル GoogLeNet
 - 22層からなるDNN
 - 一つの層を単純化(パラメータ数が1/12)
 - 中間層からの出力も利用
 - Top5 エラー率 6.67%
- CaffeによるGPU実装も公開 (Wu et al., 2014)
- 精度向上のためのチューニング
- 脳情報処理と比較するモデルとしては、(Krizhevsky et al., 2012)で十分



画像認識機結果 ((Krizhevsky et al., 2012)実装例)



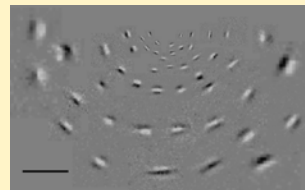
- ラベルづけのノイズや曖昧性を考えると、十分に高精度といえる

Deep Neural Networkと生体の神経回路網

- 第1層のニューロンの特性は、一次視覚野(V1)のニューロンの受容野と類似
- Gabor関数型の受容野特性(方位、空間スケール、位相選択性がある)



Receptive fields of V1 neurons (simple type)
Collected in Ohzawa lab 2004-2009



最上位層のニューロンの選択性

- 最大活動をしめた画像top16
- 高次元画像特徴に選択性をもつ = 下側頭葉のニューロンの特徴と類似

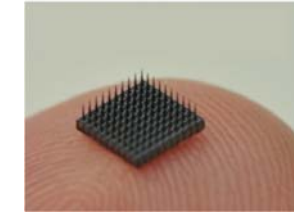


実ニューロンと人工ニューラルネットワークとの対応



- 高次視覚野のニューロンはDNNのどのレイヤーの情報表現に対応するのか？
- DNNとの比較をとおして、視覚情報処理を理解する。

サルを使った神経情報の記録

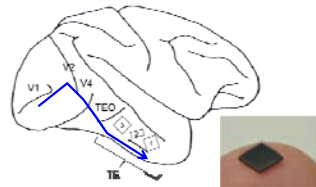


一辺が4mm,電極数 96本のマイクロ電極アレイ

ヒトと同等の視覚機能を有する
マカクザルを利用
脳も大きい
行動課題も学習できる

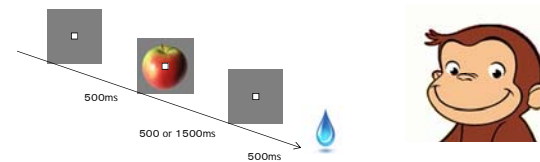
視覚皮質に多数埋め込む
神経活動を大規模に記録する

下側頭葉への電極埋め込み



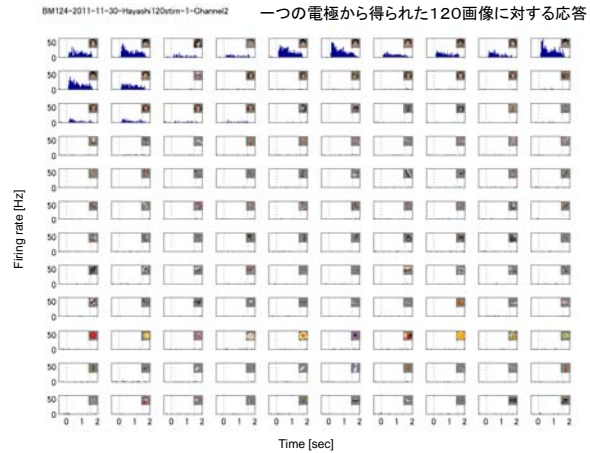
- 視覚物体情報は腹側視覚経路に沿って並列階層的に処理される
- 下側頭葉の前部、中間部、後部にアレイ電極を埋め込む(合計224本)

覚醒行動下での神経活動記録

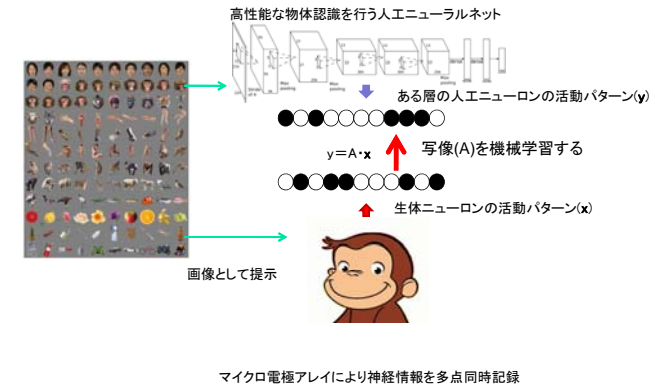


- 顔、身体部位、動物、植物、人工物などのカテゴリを含む120種類の画像を提示

応答例 (顔画像に選択的な応答)



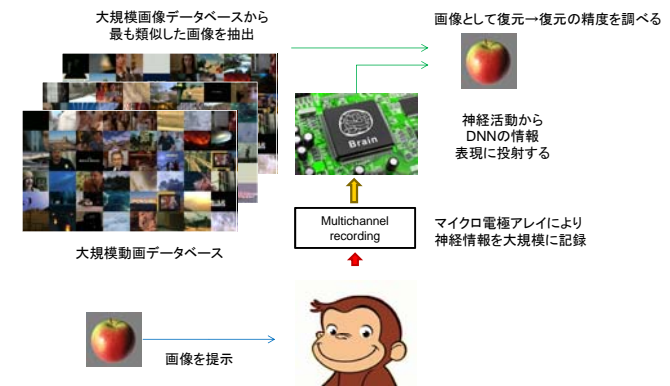
DNN各階層における情報表現との比較



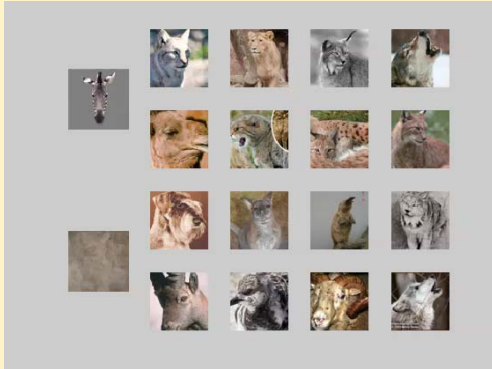
特徴推定の成績(異なるレイヤー間の比較)

- リッジ回帰によって推定した特徴表現と実際の画像の特徴表現との間の相関を計算
- 高次層の特徴を利用したとき、推定精度の向上が顕著 ($r=0.57$)
- One way ANOVA $p < 0.01$, 多重比較検定(1.2層、3.4層、5.6.7層、8層)

神経情報を画像として復号化

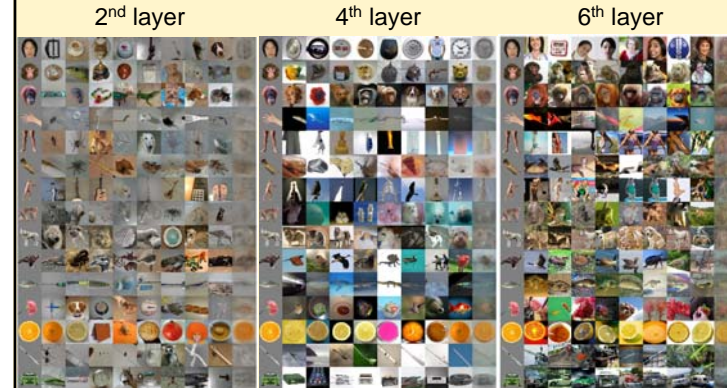


神経活動から高次層への写像を利用した画像復元



- 190chから記録した画像提示後100-1700msの神経データ(190chx17 time bin)をニューラルネット7層の画像表現(4096次元)にリッジ回帰 ($r=0.68$)
- 120万枚の画像データベースから類似画像を検索 (Hayashi & Nishimoto, 2013)
- ターゲット画像と同じカテゴリ画像が類似画像として検索される

他のレイヤーの視覚特徴を用いた画像推定

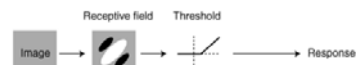


- 1-2層: 背景色が類似した画像が抽出される
- 3-5層: 大まかな輪郭や色が類似した画像が抽出
- 6-8層: 物体の方位に関係なく、カテゴリなど意味的に類似した画像が抽出

V1のモデル

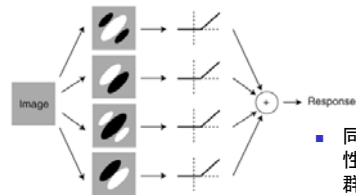
- V1のモデル:
 - さまざまな方位と空間スケール(空間周波数)をもつフィルタバンク
 - 各局所領域にそれぞれフィルタ演算をおこなう=Wavelet表現
 - simple cellは線形出力, complex cellはエネルギー出力で近似できる

A Simple cell



- ガボールフィルタに非線形関数をかけたもの

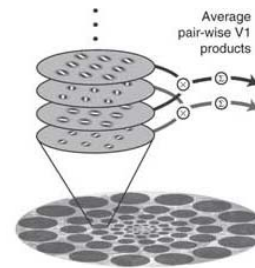
B Complex cell



- 同じ位置、方位、空間スケールに選択性をもつが、位相の異なる単純細胞群の出力和

V2のモデル

- V2のモデル:
 - 近傍の2タイプのV1ニューロンの積を、ある範囲で加算(pooling)して得られる視覚特徴を表現
 - フィルタ間の相関に相当



位置、方位、空間スケールの違うペア積を計算

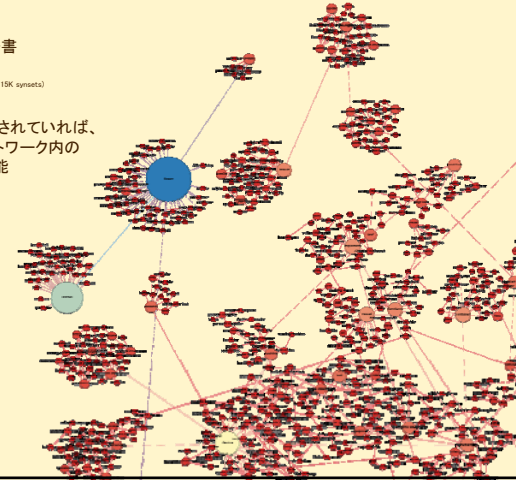
- 周期構造をもつテクスチャの表現に有効
- 電気生理実験やfMRI実験からV2ニューロンの一つの側面を記述するモデルとして受け入れられつつある
- テクスチャの表現に有効

(Portilla & Simoncelli, 2000; Freeman & Simoncelli, 2011; Freeman et al., 2013)



Wordnetに基づく意味類似度評価

- Wordnet: 英語の概念辞書
 - 同義語間のノードにより意味ネットワークを構成 (115K synsets)
- 画像の内容がラベル化されていれば、意味上の類似度をネットワーク内のパス長により定量化可能



複数指標に基づく推定画像の類似度比較

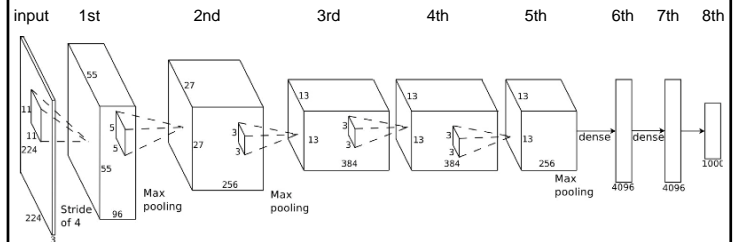
- ニューロン活動から推定した画像と実際の参照画像の類似度を異なる指標で比較
 - 無作為抽出画像と参照画像との相関係数で正規化
- 低次層の特徴により、V1 complex cell様の特徴が復元されるも、高次層の特徴を利用した意味情報ほどの復元精度はない。
- 高次層の視覚特徴を利用すれば、意味的に類似した画像が抽出される
- 5層は、回帰結果の相関は高いが、意味情報の類似した画像は抽出できない

→ 違いは？

Deep Neural Networkの構成

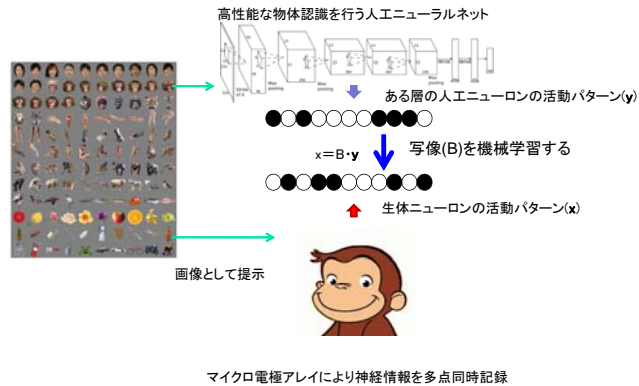
- 第1層から5層まではConvolution型の局所神経結合だけで情報が処理される
 - 物体の大きさや位置のシフト変化に対し不変な情報表現が可能に
- 第6層から8層からは、全ての場所のニューロン同士が神経結合を許容されている
 - 物体の観察方位変化などに依存しない物体認識が可能に

(Krizhevsky, Sutskever, Hinton, 2012)



- 局所的/トポロジカル的ではない大域的な情報統合が物体のカテゴリ情報と、TE野の表現として重要な可能性

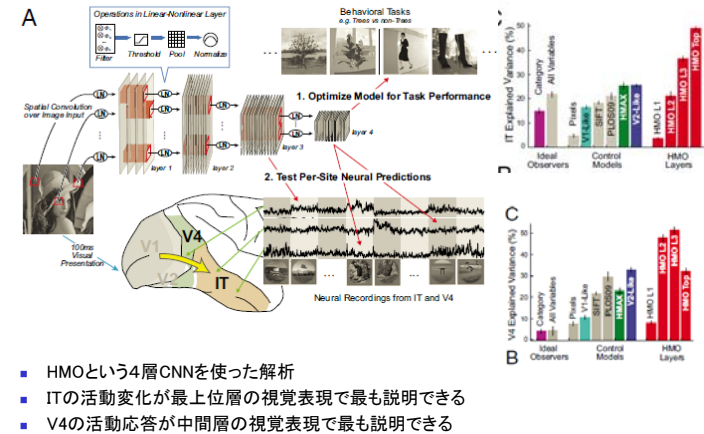
生体ニューロンの応答特性の解析



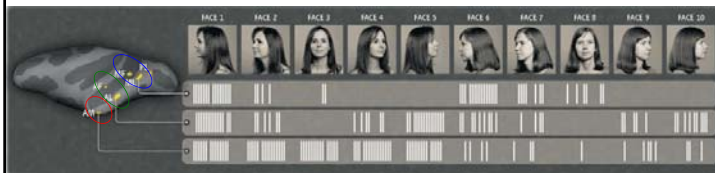
各ニューロンが最大応答を示す画像をデータベースから検索し、刺激選択性を解析した。

類似研究

(Yamins et al., PNAS 2014)



下側頭葉の領域間の機能差



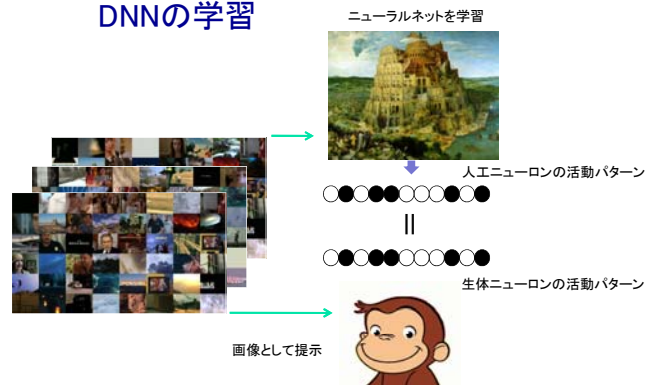
(Freiwald & Tsao, 2010)

- 下側頭葉の各領域では、観察方位の変化に対し異なる表現をしている
- MF & ML: 顔方位に選択的な活動
- AF & AL: 鏡対称な方位に対し同じ活動
- AM: 方位に依存せず、顔のIDに依存した活動
- モデルとの回帰/フィットの際、こうした機能差が失われている可能性。

発表のながれ

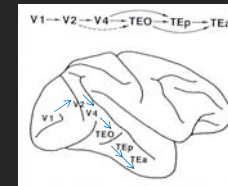
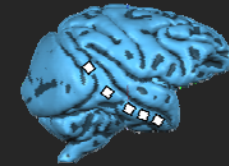
- 脳内の視覚情報処理の概要
- Deep neural networkを用いた、脳内の視覚情報表現の解析
- 今後の展望
 - 視覚概念の理解とメタファー
 - Brain-Machine Interfaceへの応用

生体ニューロンの応答を教師信号とした DNNの学習



- 高次視覚野のニューロンのモデル化が可能になる。
- 学習したDNNを解析することで、脳内の情報処理を理解する。

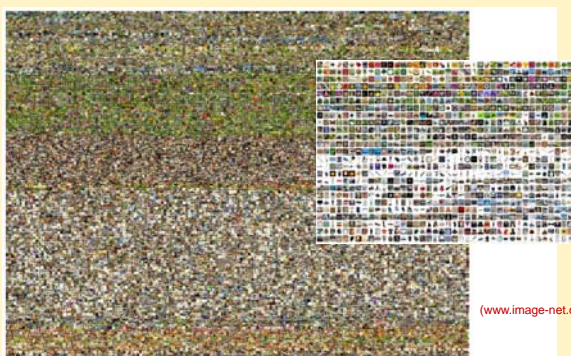
大規模神経活動に向けた電極数の拡張



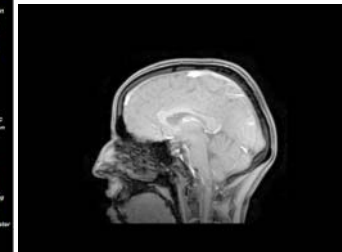
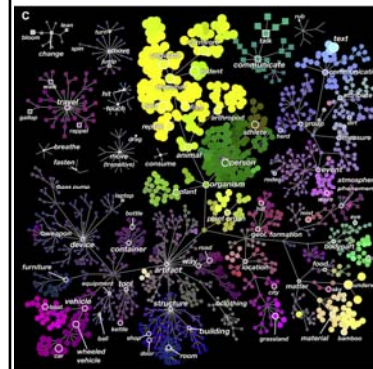
- 任意の画像復元にはより多くの神経情報が必要
- マイクロ電極アレイを5個: 合計480本の電極を埋め込む
- 下側頭葉全体 (V4野、TEO野、TE野後部、中央部、前部) から神経活動を大規模に記録する

より多様な映像を提示

- DVDからリッピングした動画画像 2TB
- Caltech256などの約40000枚の画像
- YouTubeからダウンロードした動画画像 500MB
- IMAGENETの画像データベースを利用 128,1167枚の画像



言語表現と脳情報表現の比較



(Huth et al., 2012)

- 脳活動のネットワーク解析の結果が語彙ネットワーク(word net)と対応
 - 類似概念どうしは、脳活動パターンとしても類似している
 - 神経表現の類似性は、どのようにして決まるのか？

カテゴリ間の関係

1. 視覚的類似性
 2. 自然界における物体の共起確率に基づくカテゴリ同士の連合学習
- 大規模な画像やテキストデータのなかに反映されている



自然言語解析手法の利用



犬 ボール 芝生

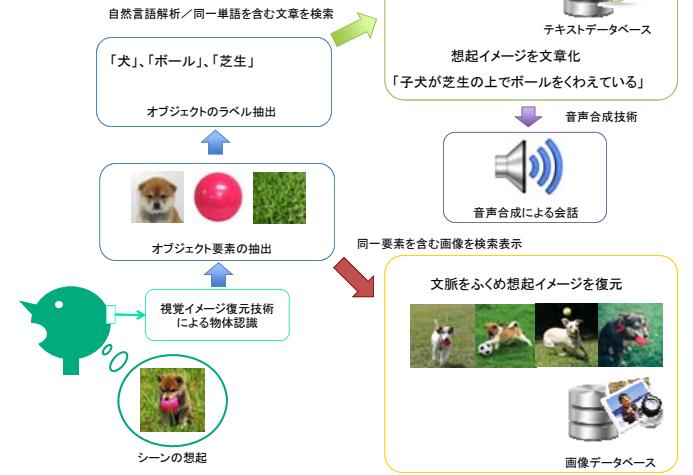


Google books Ngramsなどでの単語の共起関係と神経活動の相関

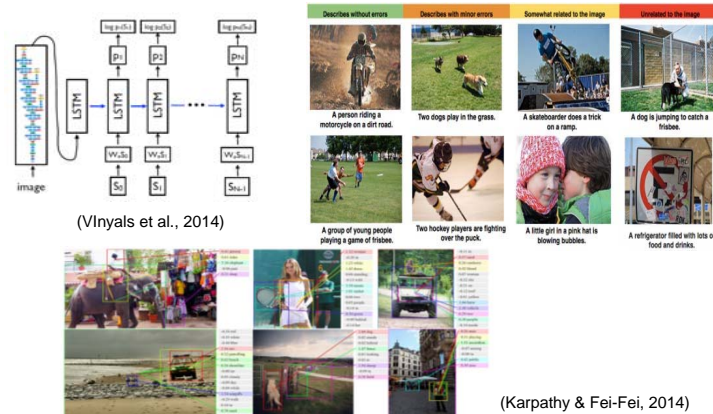
「犬がボールをくわえている」「芝生を走る犬」
「サッカーボールが芝のフィールドを転がる」

- 画像だけでなく自然言語解析手法と絡めた神経データ分析により、脳内における物体概念の表象様式の一端が解明できると期待される。

画像文脈判断と音声化

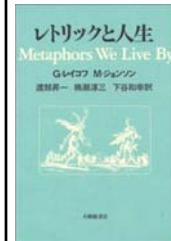


画像認識と言語処理の統合



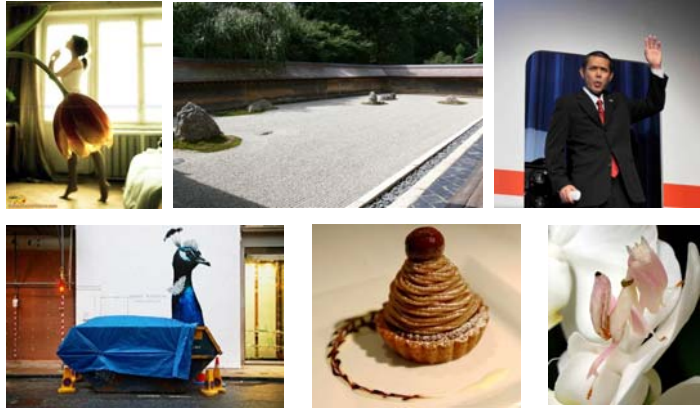
- 画像認識のCNN出力と文章生成のRNNを統合
- 画像に対する自動キャプション生成技術が急速に進展

メタファーによる概念体系の構造化説



- 修辞技法の一つ。物事のある側面を、より具体的なイメージを換気する言葉で置き換え簡潔に表現する機能をもつ (Wikipediaより)。
 - コミュニケーションも、思考や行動の場合と同じ概念体系に基づいて行われているのであるから、言語は、この概念体系の実体を知るための重要な情報源である。
 - メタファーの本質は、ある事柄を他の事柄を通して理解し、経験することである。
 - 大部分の概念は、(知覚-運動経験により獲得されるより直接的でわかりやすい)他の概念をとおして部分的に理解される(構造をあたえられる)。
 - 例) 時間は空間概念によって理解される。
 - ささまざまな経験は、構造を与えられた統一体へとまとめられる。
 - 経験の中の「重要な」部分を識別することによって、我々は経験をカテゴリ-に分類し、理解し、記憶することができる。
- 言語的にメタファーによって構造化された概念を、DNNにより実装可能では？
 - DNN: 経験は、類似性に基づき、重ね合わされて、一貫した概念体系に。

視覚的メタファー(見立て、double meaning、ものまね)



DNNにみられる視覚的メタファーの萌芽

- 女性とフラミンゴ
- プレッツェルとヘビ
- 視覚的メタファーの基盤ともよべる情報表現が既存のDNNで実装できている。
- 言語をふくめた概念体系の理解が期待される。



発表のながれ

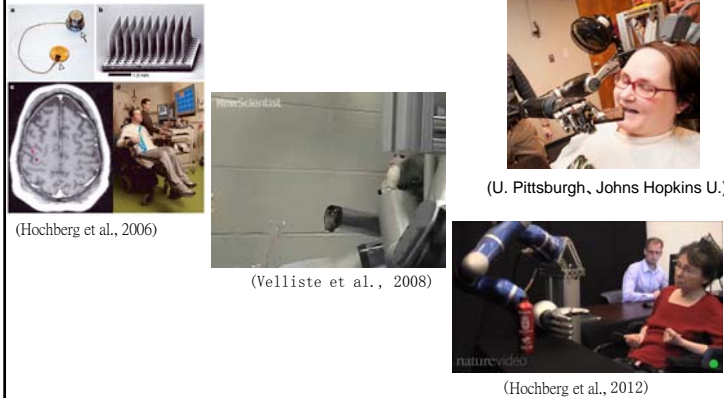
- 脳内の視覚情報処理の概要
- Deep neural networkを用いた、脳内の視覚情報表現の解析
- 今後の展望
 - 視覚概念の理解とメタファー
 - Brain-Machine Interfaceへの応用

ALSなどの身体麻痺患者が直面する困難



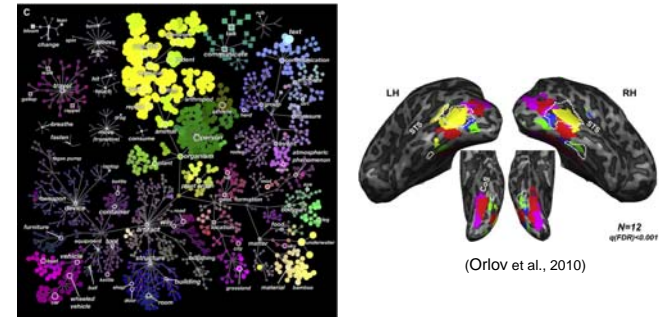
- コミュニケーション手段が失われる
- 摂食行動など日常の運動ができない
- 残存する脳機能活動を利用し、思念するだけで補助器具が操作できるBMI*技術が注目される。
- * Brain-Machine-Interfaceの略

侵襲型BMI技術の進展



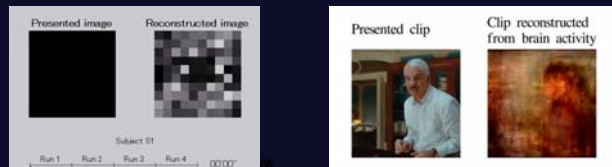
- 直接脳内に電極を埋め込み、記録した神経活動を利用するBMI技術が進歩
- 特にアメリカでヒトの臨床試験がすすむ: 運動信号の読み出しで大きな進展

脳内の高次概念表現を利用したBMI



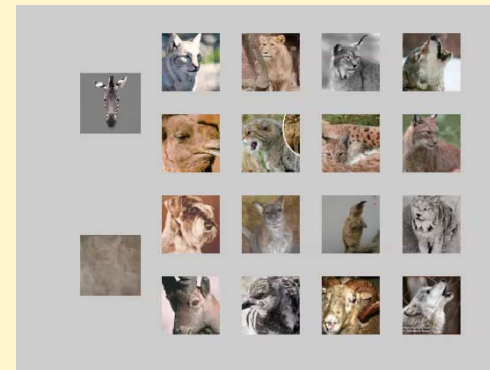
- (Huth et al., 2012)
- 運動野: 運動信号の読み出ししかできない
 - 視覚野: 高次概念を表象している。身体表現も含む
- ↓
- 視覚情報の解釈と制御がBMI技術として有望

Image reconstruction from neural/brain activity



- fMRI信号をつかった復元 (Miyawaki et al., 2008)
- fMRI信号をつかった復元 (Nishimoto et al., 2011)
- 複雑な物体像を高精細かつ実時間で復元へ
- 小型計測機器による実装
- LGNニューロンからの復元 (Stanley et al., 1999)

神経活動から高次層への写像を利用した画像復元



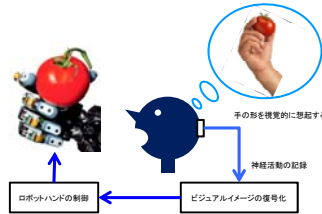
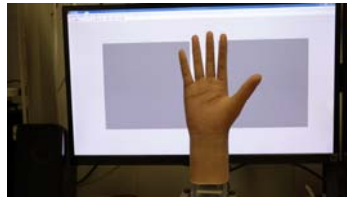
- DNNを利用することにより、見ている画像の物体情報が精度高く推定できた。
- 逆変換の学習により、データベースからの検索を利用せずに見ている画像を復元することも可能 (Hayashi & Nishimoto, 2013)

ロボットハンドの視覚脳活動制御

- 提示画像が「グー・チョキ・パー」のいずれか判別する
- 10Hzの時系列データで80-85%を正しく判別
- ロボットハンドの制御に利用できることを実証

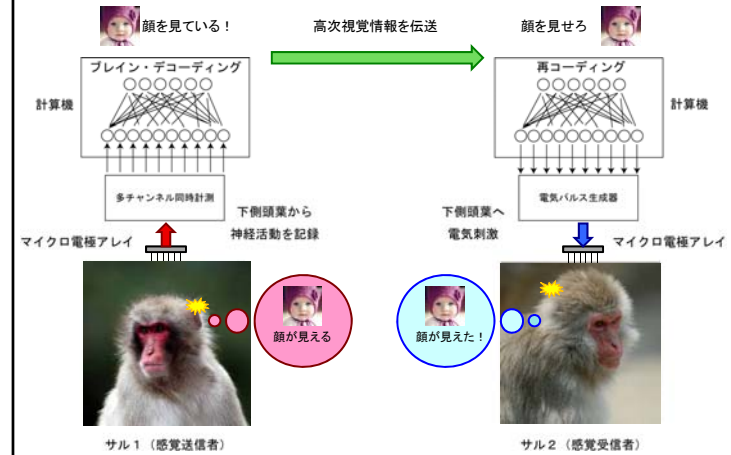


- 視覚フィードバックによるロボットハンド制御へ
- 本物の手のような見た目と動きを再現する必要



(Hayashi & Saga, 2014)

BMIを介した知覚共有技術の実現へ



総括と展望

- 下側頭葉ニューロンの情報表現が、deep convolution neural networkの高次層のニューロンと良く対応する
 - BMIに有用な視覚情報の復号化技術が実現できる。
- 多次元空間で表現される高次ニューロンの情報処理は、deep neural networkを介して理解するしかないように思われる
 - 一方で、神経活動の変動をより良く説明するモデルを実装するだけでは、理解はすまない。
 - deep neural network内の情報処理を解析する数理的なツールが必要
 - 特定の目的をもって実験者が選んだ画像刺激を使って検証
- 脳内の視覚物体情報表現は、われわれの物理世界に対する概念体系を表現
 - 映像だけでなくテキストを含めたビッグデータを利用するアプローチが有効と予想する
 - 脳というハードウェアがもつ制約条件下での情報表現の構造の理解が重要
- Deep neural networkの枠組みと神経活動記録技術を介して、
脳とコンピュータの情報処理がシームレスに統合されていく

Acknowledgements

Surgery Assist
Aya Takemura
Yasuko Sugase-Miyamoto
Yumi Murata

Collaborators
Shinji Nishimoto (NICT)
Haruo Hosoya (ATR)

Eye tracking system, MRI scanning
Keiji Matsuda
X-ray CT scanning
Takafumi Minamimoto (NIRS)
Animal care, Surgery support
Toshiharu Takasu
Ai Muramats

Funding
JST PRESTO
CANON Foundation
MEXT Grant-in-Aid for Scientific Research on
Innovative areas "The Science of Mental
Time"
JSPS Grain-in-Aid for Scientific Research C
Tateishi Foundation

Special Thanks

Manabu Tanifuji (RIKEN Brain Sci. Inst.)
Takayuki Satho (RIKEN Brain Sci. Inst.)
Naohisa Miyakawa (NCNP)

Kazutaka Takahashi (Chicago Univ.)
Keisuke Kawasaki (Niigata Univ.)
Yuuji Ichisugi (AIST)