■解説（VISION Vol. 29, No. 3, 77–85, 2017）

深層学習を用いた高次視覚機能の理解にむけて

林 隆介

産業技術総合研究所　システム脳科学研究グループ
〒305–8568 茨城県つくば市梅園1–1–1　中央第2
r-hayashi@aist.go.jp

はじめに

深層学習を用いた一般物体認識の成功以来[1]，同手法はさまざまな視覚研究に応用されている[2–4]．本講演では，最近の深層学習研究の動向を紹介しつつ，著者らによる深層ニューラルネットワークの各層と脳の各領域の視覚情報表現を比較した神経科学研究[2, 3, 5]について解説する．その後，将来の展望として，画像の文脈や感性評価など，より高次な視覚機能の理解にむけて，深層学習を利用した「錯視」研究の可能性について議論する．

深層ニューラルネットワークによる一般物体認識

ほんの5, 6年ほど前までは，コンピュータに一般物体認識をさせる（画像に映っている物体を一般的な名称で回答させる）ことは非常に困難な課題であった．それが今では，多段の階層をもつニューラルネットワーク（deep neural network: 深層ニューラルネットワーク）を用意し，大規模な画像データを使って，ニューロン間の結合パラメータをすべて機械的に学習させることで，非常に高い精度で，一般物体認識が実現できるようになった[1]．現在，画像認識の分野で，広く用いられているのは，畳み込み型深層ニューラルネットワーク(Deep Convolutional Neural Network, Deep CNN)である．その元となったのは，福島邦彦先生のネオコグニトロン[6]であり，一次視覚野の単純型細胞と複雑型細胞に関する生理学的知見のアナロジーから，フィルタの畳み込み演算を階層的に繰り返すことで，パターン認識を実現するニューラルネットワークが提案された（本学会誌VISION, **29**(1), 1–5も参照）．その後，ネオコグニトロンの着想を発展させ，畳み込み型の階層型ニューラルネットワークを誤差逆伝播法(Back Propagation[7])により学習し，パターン認識を実現する手法が確立された[8]．現在のdeep CNNは，従来のCNNの基本構造をもとに，階層を増やした構成となっている．過去にも，deep CNNを使った一般物体認識の試みが行われてきたが，学習すべきパラメータ数に比して，画像データが足りなかったり，学習がうまく収束しない（例えば，学習に用いる誤差信号が多階層を伝播する過程で消失してしまう）などの問題が解決できなかったと言われる．

ごく近年になって，一般物体認識が可能になった技術的背景としては，1）インターネット時代を経て大量の学習用画像データが利用可能になったことと，2）計算機能力の向上（特にGPUを用いた並列計算技術の進歩）があげられる．代表的な一般物体認識用の画像データベースとしては，ImageNet[9]が知られる．ImageNetには，Wordnet[10]（英語の概念辞書．単語とその関連語が登録されており，関連語どうしを結んだノードからなるネットワークをたどることで，意味上のつながりが理解できる）内の名詞に対応する画像が登録されており，2017年4月現在，約21,800個の名詞概念と約1,400万枚の画像がWeb上で公開されている(http://www.image-net.org/)．このほか，ImageNetの一

2017年冬季大会．シンポジウム講演．

部（1,000カテゴリの画像）を取り出したILSVRC20XX（ILSVRCは画像認識性能を競うコンテストであり，20XXはコンテストの年度）や，Flickr上で公開された1億枚の画像を集めたYahoo Flickr 100 M[11]，顔画像専用のMegaface[12]やCelebA[13]，さらには，ウェブから収集した画像に，その内容を説明したキャプションを付与したMSCOCO[14]などのデータベースも深層学習による画像認識に広く利用されている．

深層学習の実行は，NVIDIA社のグラフィックカードと，CUDA（GPU向けの並列計算用開発環境）上での動作を基本とする．フレームワークとしては，2～3年前であれば，Caffe (Berkeley Vision and Learning Center)が視覚研究者にとっては，スタンダードな選択肢であったが，現在では，最先端の深層ニューラルネットのコードが，Tensorflow (Google), Torch (Facebook), Chainer (PNI), Keras (MIT), Theanoなど，さまざまなフレームワークで公開されることが多い．上記のフレームワークでは，プログラミング言語に，もっぱらpythonが使われる（Torchの言語はLua）．

ニューラルネットの演算技術上の改良点としては，ニューロンの活性化関数としてReLU関数（後述）を導入したことと，学習の際，複数の画像をひとまとめにしたバッチごとに，出力の正規化を行うようになったことが，誤差信号の消失回避に寄与しているといわれている．Deep CNNによる一般物体認識研究の発展はめざましく，ILSVRCの2012年には8層のニューラルネット（通称AlexNet[1]）が84.7%の精度（ただし推定候補top5中に正解が含まれる率）で1,000種類の一般物体認識を実現したのを皮切りに，2013年には19層のニューラルネット(VGG19[15])が92.7%を記録し，2014年には，22層のニューラルネットが，93.3%の精度を実現した(GoogleNET[16])．さらには，96.4%の精度で一般物体認識を行う152層のニューラルネットの実装が行われるに至っている[17]．階層数の増加と認識精度の頭打ちから，最近は一般物体認識以外の課題に注目がシフトしている．

Deep CNNと脳の階層的情報処理の比較

Deep CNNを視覚研究に利用する魅力としては，演算アーキテクチャが脳と似ていることに加え，学習によって獲得されるニューロンの性質と，生体の神経細胞の性質に類似性が認められることがあげられる．たとえば，最下層のニューロンの結合重みをプロットすると，一次視覚野の神経細胞と同様に，さまざまな傾きや空間スケールをもつ二次元ガボールフィルタのような特性が，学習によって獲得されることが確認できる（**図1**A）．さらに，最上位層のニューロンが，もっとも強く応答する画像をプロットすると，複雑で抽象的なレベルで共通特徴をもつ，特定の物体画像に反応していることがわかる（図1B）．そこで，2012年の一般物体認識用deep CNNの登場後，deep CNNの各階層の情報処理を，脳の情報処理と対応付けることで，その情報処理を明らかにしようとする研究が行われることとなった．以下，筆者が過去に行った研究[2,3]に沿って，deep CNNと脳の視覚情報処理を比較する方法について紹介する．

線形回帰による比較

神経科学における視覚研究の目的の一つは，神経細胞の応答が，視覚入力に関するどんな情報を符号化しているのかを解明することである．数理工学的に，符号化モデルを構築する問題としてとらえるならば，画像入力から神経出力への多変量回帰問題として定式化することができる．

モデル回帰の最も基本的な手法は，画像の輝度パターンを入力とした線形回帰であるが，この方法で記述できるのは，一次視覚野の単純型細胞までである．それより高次な視覚野の情報処理をモデル化する場合には，なんらかの非線形性を考慮する必要がある．そこで，画像をdeep CNNのニューロン群の応答が表現する多

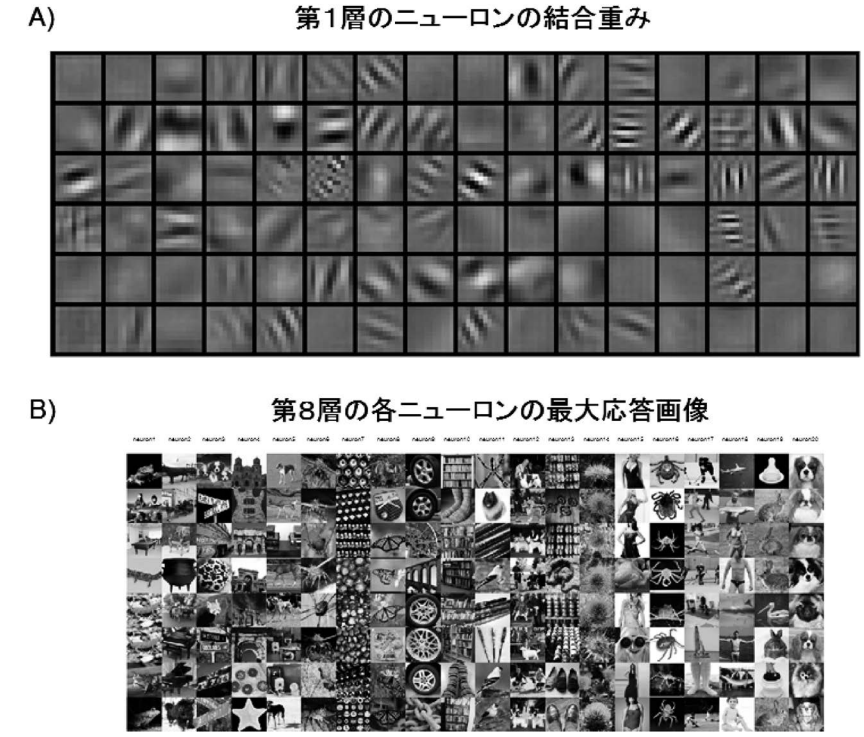


図1　AlexNet[1] のニューロン特性．A）第1層のニューロンの結合重み　B）第8層のニューロンの最大応答画像．

次元特徴空間に投射したうえで，線形回帰を行う手法がとられる．神経応答をy，視覚入力の非線形変換表現をX，回帰係数ベクトルをβ, εを誤差とした場合，線形回帰を行うモデルは次式であらわされる．

$$y=X\beta+\varepsilon$$

訓練データとの誤差だけを最小化するようにモデル回帰すると，なるべく多くの回帰係数を使って，個々の訓練データと合致するように学習が進むあまり，未学習のデータに対しては，むしろ誤差が大きくなってしまう．こうした過学習を避けるために行われるのが，学習の自由度を制約する正則化手法であり，回帰係数が大きくならないような罰則項を学習の目的関数に加える．実際のモデル化の際によく用いられるL2正則化つき回帰（リッジ回帰）とは，「誤差の二乗和」と「回帰係数の二乗和」の荷重和（ハイパーパラメータλ）からなる以下の目的関数を最小化するβを求めることである．

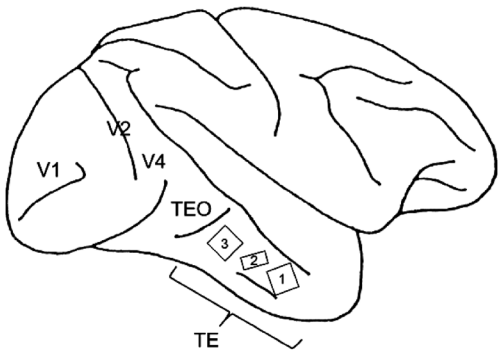
$$\min_{\beta}\|y-X\beta\|_2^2+\lambda\|\beta\|_2^2$$

モデルの妥当性は，交差検証法(Cross Validation)を用いた，未学習の入力に対する予測精度によって評価することとなる．すなわち，実験データをあらかじめ訓練データとテストデータにランダムに2分してから，訓練データだけを使って，モデル回帰係数を推定したのち，同じ係数がテストデータの神経応答をどれだけ正しく予測できるか評価する．同じデータ内で，訓練データとテストデータの分割を繰り返し行うことで，平均予測誤差を求め，ハイパーパラメータλは，この予測精度が最も高くなる値を選択する．

視覚システムを符号化モデルの形で記述することは，その逆問題である復号化を行うことと相補的な関係にある[18, 19]．復号化の場合，脳神経細胞の応答から，入力画像（あるいはそのdeep CNN表現）を推定する線形回帰モデルの構築を行うこととなる．

サル下側頭葉の神経細胞応答記録とdeep CNNの実装

下側頭葉は，物体認識に密接に関わることが示唆されているが，その情報処理をモデル化する場合，deep CNNのどの階層の特徴量表現を使ってモデル回帰すればよいかが問題となる．また，モデル回帰に有効な特徴量表現がわかったのち，神経活動データからどこまで入力画像が復号化できるか検証することは，工学応用の観点から興味深いテーマである．筆者らは，主に復号化を目的として，deep CNNを利用した下側頭葉の神経情報の解析を行った．以下では，さまざまな画像に対する多数の神経細胞群の電気的活動を時系列のベクトルデータとして記録したのち，提示画像に対するdeep CNN各層の特徴量表現が，神経活動データの線形回帰によってどの程度予測できるか検討した．そして，予測した特徴量表現と類似した画像をデータベースから検索することで，見ている画像を復元した事例を紹介する[2, 3]．



サルの下側頭葉（TE野）の電極埋め込み位置

図2　アレイ型微小電極の埋め込み位置．

筆者は，サルの下側頭葉（TE野）の前部，中央部，後部の3ヵ所にアレイ型微小電極を埋め込み（**図2**参照），総数にして224本の電極針から神経細胞群の電気的活動を同時記録した．実験中，サルはコンピュータのモニタ中央を注視しており，ランダムに繰り返し提示される120種類の物体画像観察中に生じる神経細胞群の電気信号が計測された．

比較対象としたdeep CNNは，2012–2013年当時，一般物体認識課題で最高性能だったAlexNet[1]である（**図3**）．当時はcuda-convnetをフレームワークとして，ライブラリを修正しながらモデルの学習を行ったが，今ではCaffeなどを利用して，公開されている学習済みパラメータを読み込めば，簡単に実装可能である．AlexNetは，5つの畳み込み層と3つの全結合層からなるニューラルネットで，1層と2層，5層の畳み込み層の後には，それぞれMax pooling層があり，一定範囲内にあるニューロン群の出力を集約し，その最大値だけを後層に出力する（これは，複雑型細胞にみられる局所の空間統合処理に相当し，画像の微小な位置変化に対する恒常性を持たせる効果がある．ただし，最近ではpooling層を挿入しない，畳み込み層だけからなるニューラルネットでも実現できるといわれている[20]）．そして，最終層では，Softmax関数による多項ロジスティック回帰で

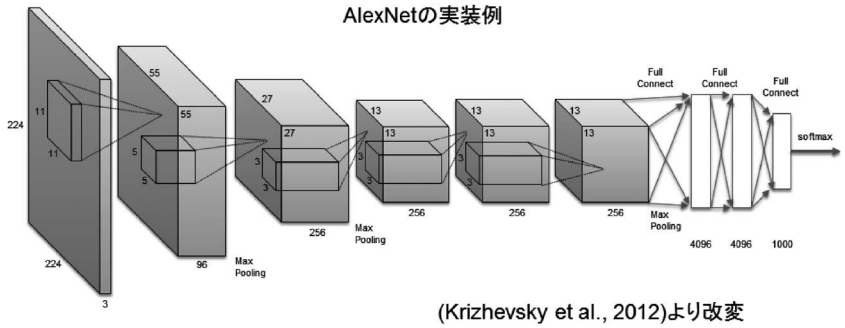


図3　AlexNetの実装例．

図4　Deep CNNの第8層の特徴量推定を利用した入力画像の復元.

クラス分類を行う．学習にはILSVRC2012の画像データセット（1,000種，120万枚の画像）を用い，1,000枚の画像を1バッチとして，確率的勾配降下法(Stocastic Gradient Descent, SGD)によりパラメータを更新した．それぞれのニューロンの活性化関数はReLU (Rectified Linear, max (0, x))関数である．また，パラメータの過学習を避けるためdrop out法を用いた（毎回ランダムに半分のニューロンの出力をゼロにしながら学習を行う．学習終了後はすべての結合重みを半分の値にする．drop out法は，過学習の低減に有効だが，最近ではdropout法を用いない実装例も多い）．

入力画像に対するdeep CNN各層のニューロン群の活動パターンをPCAで次元削減（累積寄与率95%以上となる主成分数）した表現を，各層における画像の特徴量表現とした．そして，実験に用いた120枚のうち，119枚の画像の特徴量表現と記録した神経活動データ（画像提示後200–400 ms間のスパイク発火頻度）を訓練データとして回帰係数を決定したのち，未学習の画像に対する神経活動データ（＝テストデータ）が，入力画像の特徴量表現の真値をどの程度正しく推定できるか，両者の相関係数を計算した．これを120枚の各画像に対する神経活動データについて繰り返す，Leave one outによる交差検証を行い，予測精度を評価した．神経活動データからdeep CNN各層の特徴量表現がどれだけの精度で予測できるか比較すると，上位層へ行くほど予測精度が漸次向上し，第8層の特徴量表現との対応が最も高い（相関係数＝0.57±SE0.01）ことを明らかにした．

このように，TE野の神経細胞群による視覚情報表現がdeep CNNの上位層の情報表現とよく対応するのであれば，神経活動データから入力画像の特徴量表現を推定し，類似の特徴量表現を持つ画像をデータベースから検索することで，元の提示画像を可視化することが可能となる[2, 3]．**図4**では，神経活動データから提示画像の第8層における特徴量表現を推定したのち，ILSVRC2012画像データベースから，類似特徴量表現を持つ画像を検索した結果を示す．deep CNN上位層の特徴量表現への回帰という極めてシンプルな手法だけで，脳の電気信号から，今見ている画像の，とりわけ物体カテゴリの内容を極めて正確に推定できている．以上の結果はTE野が脳において物体認識処理の終端であるという知見と整合する．

画像間（非）類似度行列による比較

神経応答とdeep CNNの特徴量表現との間で線形回帰を行う前述の手法では，両者の比較の際に，全結合層を1つdeep CNNに加えたことに相当する．このため，微妙な情報処理の階層差が比較検証できない可能性がある．筆者は，120種類の物体画像の他に，男女18人の顔がさ

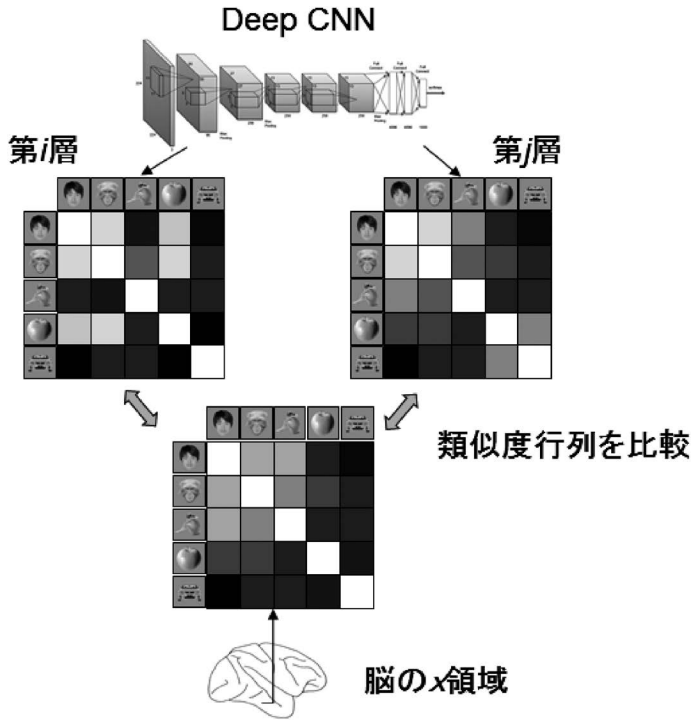


図5　画像間の類似度行列を利用したシステム間の情報表現の比較.

まざまな方位（左90度から右90度まで30度刻み）を向いた画像群に対する応答も記録した．顔画像応答の解析をとおして，下側頭葉の3つの記録部位で，顔の方位情報と個人識別情報の表現様式が異なることを明らかにした[21]．しかしながら，線形回帰を介して，deep CNNと比較した場合，記録部位の違いによる階層差を認めることができなかった．

これに対し，Kriegeskorteら[5]は，異なるシステムの情報処理を比較する手法として，入力画像どうしの符号化表現の類似度関係を指標に用いることを提唱している（**図5**）．すなわち，ある脳領域の神経活動とdeep CNNのニューロン応答を直接比較するのではなく，まず，個々のシステム内で，入力画像群の表現形式が，互いにどれだけ類似（あるいは非類似）しているか，画像間の（非）類似度行列を計算したのち，システム間で画像間類似度行列の類似度を評価する手法である．

そこで，筆者は，Alexnetよりも物体認識精度が高く，階層数も多い，VGG19[15]を比較モデルとしてとりあげ，その各階層における画像間類似度行列と，3つの脳領域別の画像間類似度行列を比較する解析を行った．その結果，前部下側頭葉の情報表現は，deep CNN上位層と高い相関があるのに対し，中央部と後部下側頭葉の情報表現はdeep CNNの低次から中位層と相関が高いことが明らかになった．さらに，神経活動の時間別解析をしていくと，後部と中央部における，情報表現の時間特性の違いが確認された．このように，画像間類似度行列を介したdeep CNNとの比較から，わずか2 cmほどの小さな脳領域の中にも情報処理の階層性とダイナミックスの違いが明らかとなった．

深層ニューラルネットワークを用いた視覚研究の展望

Deep CNNと脳活動を単に階層間で比較する研究は，その後，多くの研究者が利用しており[4, 5]，すでに解析手法そのものの新規性は失われつつある．また，多くの研究結果は，脳における情報処理の階層性の追認に留まってお

図6　使用したモーフィング画像例.

り，個別の情報処理理解の深化には至っていないのが現状である．そのため，deep CNNを用いた視覚研究の新たな方向性が模索されている．Yamins & DiCarlo[22]は，今後の研究展開として1）一般物体認識以外の，より高次な視覚機能の理解，2）脳の機能的構成に基づくアーキテクチャの改善，3）あたらしい学習則の利用などが課題となると指摘している．筆者は，1）に関しては，物体のカテゴリ情報よりもさらに高度な，意味論的表現（セマンティクス）や画像の文脈理解などに注目している．2）については，脳に見られる視覚機能コラムなどの構造化された機能をいかに実現するか，そして，マルチタスクに適した情報処理の分岐をどのように実装するか興味がある．3）については，現在Generative Adversarial Neural Networkと呼ばれる（半）教師なし学習法が目覚ましい発展を遂げており[23,24]，今後教師なし学習によって獲得されたdeep CNNの情報表現利用が視覚研究の主流になると予測している（詳細は，本解説では触れない）．第4の展望として，筆者は，モデルの妥当性の検証法についても，従来の神経活動データの予測精度にくわえ，機能的な予測性がより重要になると考えている．たとえばdeep CNNが，学習データに含まれない視覚入力に対し，われわれと同様の「錯視」現象を再現するか検証するといったアプローチである．

セマンティクスに関連した深層ニューラルネットワーク研究としては，入力画像に対し，その内容を説明するキャプションを自動生成する研究が知られる[25]．キャプションの自動生成には，画像認識用のdeep CNNと文章生成用の再帰ニューラルネットをつなぎあわせ，MSCOCOのように画像とキャプションがセットになったデータベースを使って学習することで実現する．視覚研究のツールとしては，文章生成部分を省略し，画像に含まれるさまざまな視覚概念を単語として出力するニューラルネットワークなども研究されている[26]．筆者は，視覚概念を単語として出力するニューラルネットワークを使って，「不気味の谷」現象（ロボットやCGアバターなどの外観が人間に近づくと，ある時点で強い嫌悪感が生じる現象[27]）を説明する研究に取り組んでいる．顔画像とその他のオブジェクト画像とのモーフィング画像に対し（**図6**），視覚概念を単語出力する深層ニューラルネットが，どのような応答を示すのかを解析した結果，モーフィング度が中間レベルになると名詞概念の混乱が生じ，それにあわせて形容詞表現のなかでも嫌悪感に関連した単語の出力が上昇することが確認された．この結果は，通常の視覚体験でも生じる概念間の対立や不整合が嫌悪感と関わっており，同メカニズムが「不気味の谷」の認知基盤となりうることを示唆している．

おわりに

深層ニューラルネットの技術発展により，脳と等価な情報処理システムを人工的に構築することで，構成論的に視覚情報処理を理解する試みが，有効となりつつある．GPGPU用のオー

プンソースを使って，1台のPCで手軽に深層ニューラルネットが実装できるようになったことも，基礎視覚研究者の側から，構成論的に視覚システム研究を目指す追い風となっている．構成論的アプローチによって期待されるのは，未学習な視覚入力に対して高い予測精度を実現する，ニューラルネット・アーキテクチャの構築である．今後は，深層ニューラルネットを脳から記録した神経活動と照合するだけでなく，錯視の再現を通した視覚システムのモデル同定など，視覚心理学研究を含めた領域横断的な研究がますます重要となるであろう．

文　　献

1) A. Krizhevsky, I. Sutskever and G. E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 1106–1114, 2012.
2) R. Hayashi and S. Nishimoto: Decoding visual information in monkey IT cortex using deep neural network. *Proceedings of Life Engineering Symposium 2013 (LE2013)*: 511–514, 2013.
3) R. Hayashi and S. Nishimoto: Image reconstruction from neural activity via higher-order visual features derived from deep convolutional neural networks, *Neuroscience 2013 (The 43rd Annual Meeting of the Society for Neuroscience)*, San Diego, USA, Nov. 12, 2013.
4) D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert and J. J. DiCarlo: Performance-optimized hierarchical models predict neural responses in higher visual cortex, *Proceedings of National Academic Science U.S.A.*, **111**, 8619–8624, 2014.
5) N. Kriegeskorte, M. Mur and P. Bandettini: Representational similarity analysis-connecting the branches of systems neuroscience, *Frontiers in Systems Neuroscience*, **2**, 1–28, 2008.
6) K. Fukushima: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biological Cybernetics*, **36**, 193–202, 1980.
7) D. E. Rumelhart, G. E. Hinton and R. J. Williams: Learning representation by back-propagating errors, *Nature*, **323**, 533–536, 1986.
8) Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Handwritten digit recognition with a back-propagation network, *Advances in Neural Information Processing Systems*, 1990.
9) J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei: ImageNet: A Large-Scale Hierarchical Image Database. *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2009.
10) G. A. Miller: Wordnet: A Lexical Database for English, *Communications of the ACM*, **38**, 39–41, 1995.
11) B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth and L. Li: YFCC100M: The New Data in Multimedia Research, *Comunications of the ACM*, **59**, 64–73, 2016.
12) I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller and E. Brossard: The MegaFace Benchmark: 1 Million Faces for Recognition at Scale, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
13) S. Yang, P. Luo, C. C. Loy, and X. Tang: From Facial Parts Responses to Face Detection: A Deep Learning Approach, *IEEE International Conference on Computer Vision (ICCV)*, 2015.
14) T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick and P. Dollar: Microsoft COCO: Common objects in context, *arXiv*, 1405.0312v3, 2015.
15) K. Simonyan and A. Zisserman: Very deep convolutional networks for large-scale image recognition, *arXiv*, 1409.1556v6, 2014.
16) C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. Reed,

D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich: Going deeper with convolutions, CoRR, *arXiv*, 1409.4842, 2014.

17) K. He, X. Zhang, S. Ren and J. Sun: Deep residual learning for image recognition, *arXiv*, 1512.03385.
18) Y. Miyawaki, H. Uchida, O. Yamashita, M. Sato, Y. Morito, H. C. Tanabe, N. Sadato and Y. Kamitani: Visual image reconstruction from human brain activity using a combination of multi-scale local image decoders. *Neuron*, **60**, 915–929, 2008.
19) S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu and J. L. Gallant: Reconstructing visual experiences from brain activity evoked by natural movies, *Current Biology*, **21**, 1641–1646, 2011.
20) J. T. Springenberg, A. Dosovitskiy, T. Brox and M. Riedmiller: Striving for Simplicity: The all Convolutional Net, *arXiv*, 1412.6806, 2015.
21) R. Hayashi: Hierarchical processing of face across the surface of macaque inferior temporal cortex: multi-electrode array recording study, *Neuroscience 2012 (the 42nd Annual Meeting of the Society for Neuroscience)*, New Orleans, USA, Oct 2012.
22) D. L. K. Yamins and J. J. DiCarlo: Using goal-diriven deep learning models to understand sensory cortex, *Nature Neuroscience*, **19**, 356–365, 2016.
23) I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio: Generative adversarial nets, *arXiv*, 1406.2661, 2014.
24) A. Radford, L. Metz and S. Chintala: Unsupervised representation learning with deep convolutional generative adversarial networks, *arXiv*, 1511.06434, 2015.
25) H. Fang, S. Gupta, F. Iandola, R. Srivastava, L. Deng, P. Dollár, J. Gao, X. He, M. Mitchell, J. C. Platt, C. L. Zitnick and G. Zweig: From captions to visual concepts and back, *arXiv*, 1411.4952, 2015.
26) I. Misra, C. L. Zitnick, M. Mitchell and R. Girshick: Seeing through the human reporting bias: Visual classifiers from noisy human-centric labels, *arXiv*, 1512.06974, 2016.
27) M. Mori: The uncanny valley, *IEE Robotics and Automation*, **19**, 98–100, 2012.