2K4-OS-33b-2

深層畳み込みニューラルネットワークが獲得する注意選択モデルに対する心理物理学的タスクの影響

How influence the psychophysical tasks for developing the CNN-based saliency map

我妻伸彦[*1]
Nobuhiko Wagatsuma

日高章理[*1]
Akinori Hidaka

[*1] 東京電機大学理工学部
Tokyo Denki University, School of Science and Engineering

Selective attention is a fundamental function of the brain that allocated its computational resource to the momentarily most important subsets in a visual scene. In this work, we investigated how psychophysical tasks modulated the establishment of a saliency map model based on the Convolutional Neural Network (CNN). Qualitative characteristics of gaze prediction by our CNN-based saliency map models were markedly dependent on the psychophysical tasks. These results suggested the insights into the perceptual characteristics and mechanisms for selective attention and gaze determination.

1. はじめに

外界から投射される膨大な視覚情報は、網膜から大脳視覚皮質へと送られる。そのとき、我々の視覚系は、視野内に投影された全ての情報を均等に処理しているわけではない。その瞬間において、最も注目すべき物体・空間を選択し、それを重点的に処理する。これが選択的注意である[Posner 80]。視覚的注意は、情報を取捨選択し、実環境に適応するために生物が獲得した知覚戦略である[我妻 15]。処理すべき情報を決定するこのような注意選択の大脳皮質メカニズムを理解するための計算論的モデル構築とそのシミュレーション研究が盛んに行われている。注意選択の代表的なモデルとして、Itti らが提案する saliency map (顕著性マップ)が知られている[Itti 98][Itti 00]。これは、主に初期視覚の情報処理機構に基づくモデルであり、ヒトの注意選択特性を良く説明する。近年では、中低次視覚皮質が検出する図領域の統合を起源とする saliency map も提案され、より高い精度のヒト注意選択予測が可能となっている[Russell 14]。

最近、機械学習の手法として、畳み込みニューラルネットワーク(Convolutional Neural Network, CNN)が注目され、画像や音声の自動認識を始めとする様々な分野において、強力な性能を発揮している。CNN に基づくヒトの注意選択領域と視線位置を予測する saliency map モデルが提案されている[Pan 16]。これまでの saliency map モデルのネットワークは、大脳生理学的・心理物理学的知見に基づき、構築される[Itti 98][Russell 14]。一方、CNN ベースの saliency map は、注意選択のためのネットワーク構造を、与えられた学習データに基づき CNN が自ら学習し、獲得する。Pan らが提案した CNN saliency map モデルは、極めて高精度のヒト視線位置予測を実現している。しかし、ヒト被験者への心理物理学的タスクが、CNN に獲得される saliency map モデル構造へ及ぼす影響は明らかになっていない。

本研究では、ヒト被験者に与えられた心理物理学的タスクが、CNN が獲得する saliency map モデルへと及ぼす変調効果を検証した。同一の視覚刺激に対して、異なる心理物理学的タスクを行うヒトの視線追跡データを学習データとする CNN saliency map モデルを構築した。視線予測精度と注意選択特性を核としたモデル間で比較した。具体的には、ヒト被験者が局所的な図方向知覚、もしくは大域的な図地領域知覚を行うときの視線追跡データ[ト部 16]を、それぞれ学習データとして CNN に適用した。異なる特性を持つ視線データから学習された 2 つの CNN saliency model が予測した注意選択領域は、与えられた心理物学的タスクを達成する局所的、または大域的なヒトの注視特性を反映した。また、Itti らが提案した saliency map モデルよりも高い精度でヒトの注視位置を予測した。これらの結果は、心理物理学的タスクがヒトの注意選択と CNN saliency map モデルに与える変調特性を示唆する。

連絡先：我妻伸彦，東京電機大学，埼玉県比企郡鳩山町石坂，Email: nwagatsuma@rd.dendai.ac.jp

2. 提案モデルと学習データ

2.1 CNN が獲得する注意選択の saliency map モデル

本研究で用いたヒトの注意選択特性を獲得する CNN を図 1 に示す。このネットワーク構造は、Pan らにより提案された SalNet [Pan 16]と同様の構造である。畳み込み層における活性化関数は ReLU 関数、プーリングは最大値プーリングを用いた。入力は 320x240 ピクセルの RGB カラー画像であり、畳み込みとプーリングの反復によって 80x60 ピクセルの特徴マップまで縮小された後、Deconvolution 層によって 320x240 ピクセルの出力マップ（saliency map）が得られる。

ネットワークを学習させるため、自然画像に対するヒトの固視特性画像を ground-truth として用いた（図 2）。各自然画像を入力して得られる出力マップと、その ground-truth（固視特性画像）との間の平均2乗誤差を誤差関数として、誤差逆伝播法により学習を行った。バッチサイズは 2、学習係数は 0.00005、エポック数は 100 とし、最適化手法には Adam [Kingma 14]を用いた。学習データとして適用したデータセットについては、次節で紹介する。

2.2 学習データセット

ヒトの注意選択特性を CNN に獲得させるため、自然画像とそれに対するヒト固視特性画像を学習データとして適用した。本研究では、CAT2000[Borji 2015]、MIT data set [Judd 09]、FIGRIM Fixation Dataset [Bylinskii 15]などの自然画像とそれに対する視線追跡データ、合計約 5700 枚を使用し、ネットワークの学習を実行した。

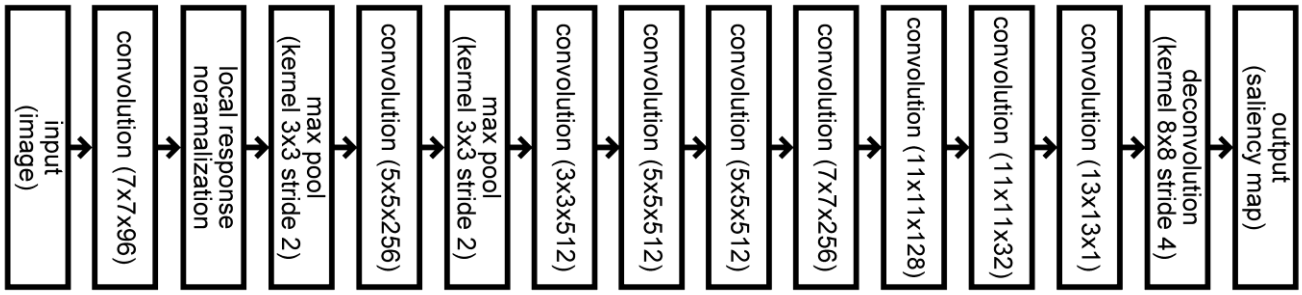

図 1. ヒトの注意選択特性を獲得するための CNN。本研究で用いた CNN の構造は、Pan らにより提案された SalNet と同様である[Pan 16]。活性化関数等の詳細は、本文にて示す。

図 2. 学習データの画像例。MIT data set [Judd 09]より。左図のような自然画像に対して、ヒト被験者の注視位置を計測する。得られた注視位置を中心とするガウシアンを重ねる事で、右図のような固視特性画像が生成される。最も頻繁に被験者が固視した領域は、白で表現されている。自然画像（左）を入力、固視特性画像（右）を ground-truth とする学習データを図 1 の CNN に適用し、ネットワークの学習を行った。

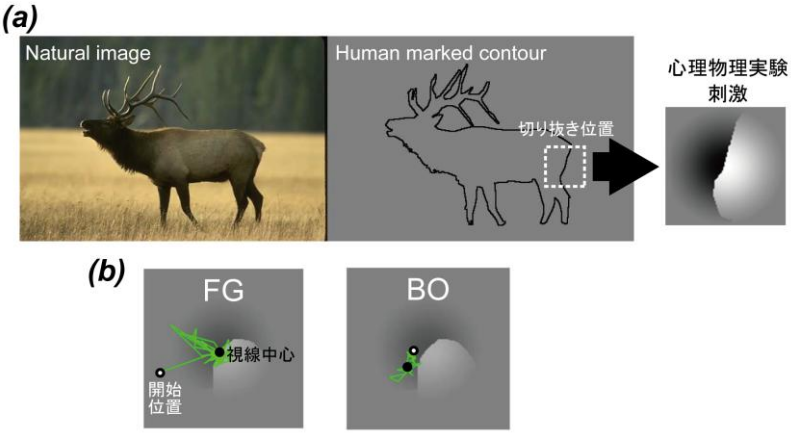

図 3. 卜部らが用いた視覚刺激と心理物理学的タスクに依存する視線移動特性の変化。***(a)***心理物理実験で用いた刺激の作成例[卜部 16]。自然画像からヒトが抽出した輪郭(Human marked contour)上の一点を中心に切り抜いた領域から心理物理実験刺激が生成される。***(b)***実験タスクによる視線移動範囲の変化。図地知覚(FG)では、広範囲に及ぶ視線移動が生じる。一方、図方向知覚(BO)では、局所的な範囲に視線移動が行われる。

ヒトは視野内に存在する物体とその位置を認識するために、物体領域(図)と背景領域(地)を分離して知覚する。この知覚を図地分離(Figure-Ground segregation, FG)という。また、輪郭上のある一点から、物体が存在する局所的な方向を図方向(Border-Ownership, BO)と呼ぶ。卜部らは、自然画像のデータセットである Berkeley Segmentation Dataset [Martin 01]を加工した視覚刺激(図 3(a))を用い、図地知覚と図方向知覚におけるヒトの注視特性を計測する心理物理実験を行った[卜部 16]。この心理物理実験は、大脳視覚情報処理メカニズムの理解に重要な知見を示唆しただけでなく、心理物理学的タスクにより、ヒトの注視・視線移動特性が変調されることも示した(図 3(b))。本研究では、前述の自然画像に対する視線追跡データに加えて、心理物理学的タスクの効果を反映した卜部らの実験データ(約 800 枚)を用い、心理物理学的タスクに依存する CNN saliency map モデルを構築した。図方向知覚タスクにより計測された視線データに基づく CNN saliency map を BO モデル、図地知覚タスクに起因する saliency map を FG モデルと呼ぶ。

3. シミュレーション実験と注意選択領域の予測

CNN が獲得する saliency map モデルに対する心理物理学的タスクの影響を検討するため、獲得された FG モデルと BO モデルのそれぞれに自然画像を適用し、saliency map を生成した。ここでは、Toronto dataset に含まれる 120 枚の自然画像をモデルへと適用した[Bruce 09]。得られた saliency map の一部を図 4 に示す。従来の saliency map モデル[Itti 00][Russell 14]と比較して、2 つの saliency map モデルは局所的な空間領域に対して活性化する傾向を示した。また、ヒトの固視特性画像と良く似た画像を saliency map として生成した。これは、CNN が注意選択領域を予測するネットワーク構造を獲得したことを意味する。また、BO モデルよりも、FG モデルはより広い空間領域へと活動を分布させる傾向が確認された。これらの結果は、CNN が獲得する注意選択モデルの構造に、心理物理学的タスクに依存する視線移動特性(図 3(b))が反映されている事を示唆している。

得られた saliency map が予測する注意選択領域の精度を定量的に評価するため、Toronto dataset から提供されるヒト被験者の固視位置情報を指標とする AUC をモデルごとに算出した[Judd 09]。Toronto dataset に対する各モデルの AUC スコアを図 5 に示す。CNN が獲得した2つの saliency map モデルは、従来のモデル[Itti 98][Russell 14]よりも高精度で、ヒトの注意選択特性を再現できることが示された。また、本研究で獲得された BO モデルと FG モデルでは、生成される saliency map の検出特性が異なる(図 4)一方、2 つのモデルから算出された AUC スコアに有意な差はなかった(t-test, P=0.884)。

4. 終わりに

本研究では、ヒト被験者に与えられた心理物理学的タスクが、CNN に獲得される注意選択の saliency map モデルへと及ぼす変調効果を検証した。図地知覚と図方向知覚におけるヒトの注

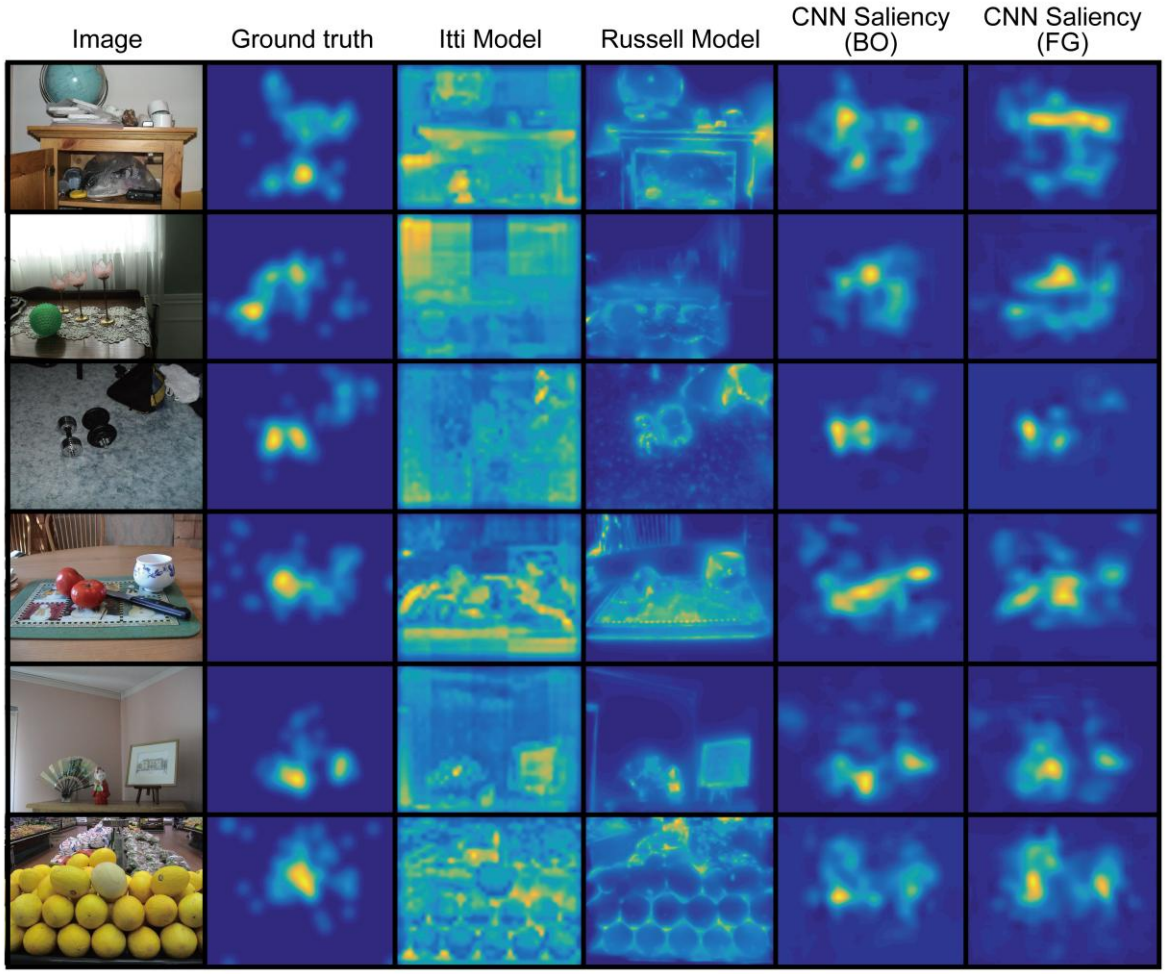

図 4. 入力画像と saliency map モデルが決定する注意選択領域の例。左のコラムから順番に、入力画像、ヒト被験者の固視特性画像(ground-truth)、Itti モデル[Itti 00]、Russell モデル[Russell 14]、図方向知覚(BO)時の視線データに基づく CNN、そして図地知覚(FG)時の視線データに基づく CNN が生成する saliency map を示す。これらの入力画像と固視特性画像は、Toronto dataset [Bruce 09]により提供される。Saliency map において、青い領域が弱い反応、注意が向けられないと予測される領域を意味する。反対に、黄色の領域は、優先的に注意が向けられると予測される領域を示す。従来の saliency map モデルに比べ、CNN が生成する saliency map は局所的な空間領域に対して活性化する傾向を示した。また、ground-truth である固視特性画像と類似する反応傾向が得られた。

視特性を検討した心理物理実験から計測されたヒト被験者の視線データ(図 3, [ト部 16])を学習データとして CNN に適用することで、2 つの saliency map モデルを CNN に獲得させた(図 1)。本研究で獲得された CNN saliency map モデルは、従来のモデル[Itti 98][Russell 14]よりも高精度でヒトの注意選択空間を予測した。また、BO モデルと FG モデルが生成する saliency map は、それぞれ異なる反応特性と注意選択特性を示した。しかし、獲得された2つの CNN が予測する注意選択予測精度に、有意な差はなかった。

ト部らの心理物理実験において、図方向知覚(BO)を行う際、ヒト被験者は局所的な領域を注視する傾向が見られた(図 3(b), BO)。対照的に、図地知覚(FG)では、大域的な範囲へと被験者は視線を移動させる。これらの視線移動・注視位置特性が、BO モデルと FG モデルから生成される saliency map(図 4)に変調をもたらしたと考えられる。

AUC 解析では、視線位置の経時変動などは考慮されない。また、広範囲に活動を分布させるようなモデルに対して、AUC スコアは上昇する傾向がある。Saliency map の評価には、様々な指標が用いられている。AUC スコア以外の異なる指標を用い

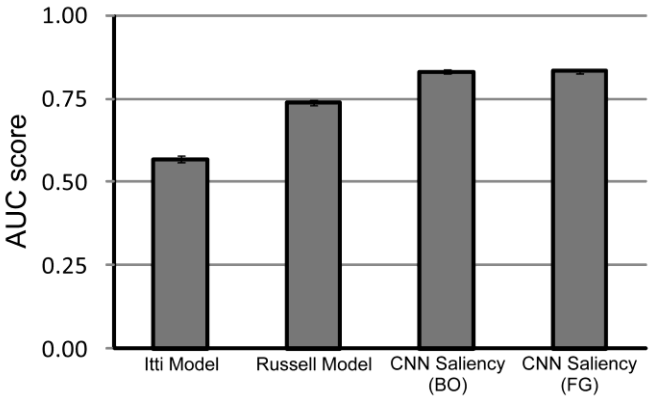

図 5. Toronto dataset [Bruce 09]に対する AUC スコア。自然画像 120 枚の平均を示す。高スコアは、ヒト被験者が示す注意選択特性とモデルの予測がより一致した事を意味する。従来のモデルに比べ、CNN saliency map モデルはヒト注意選択特性をより高精度で予測した。また、2 つの CNN モデル間で AUC スコアに有意差はない。

た CNN モデルの定量的評価が、注意選択モデルを評価するために重要となることが予測される。

学習によって CNN が獲得したネットワーク構造は、生体の神経回路構造と一致する可能性が報告されている[Ranzato 12]。そのため、本研究で獲得されたヒトの注意選択を予測する CNN のネットワーク構造が、選択的注意を仲介する神経回路構造の一端を再現している可能性がある。獲得された CNN の解析による注意選択メカニズムの理解が期待される。我々が構築した CNN は、生体の視覚情報処理系における重要な戦略である注意選択の特性を再現し、その神経回路構造の一端を解明する可能性を示唆した。

謝辞

貴重な心理物理実験データを提供して頂いた筑波大学大学院システム情報工学研究科教授酒井宏教授に感謝する。また、本研究の一部は JSPS 科研費 JP16K16090 の助成を受けたものである。

参考文献

[Borji 15] Borji, A. and Itti, L.: CAT2000: a large scale fixation dataset for boosting saliency research, *arXiv*, 1505.03581 (2015)

[Bruce 09] Bruce, N. D. B. and Tsotsos, J. K.: Saliency, attention, and visual search: an information theoretic approach, *Journal of Vision*, Vol. 9, 5 (2009)

[Bylinskii 15] Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A. and Oliva, A.: Intrinsic and extrinsic effects of image memorability, *Vision Research*, Vol. 116, pp. 165-178 (2015)

[Itti 98] Itti, L., Koch, C. and Niebur, E.: A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans Pattern Anal Mach Intell*, Vol. 20, pp. 1254-1259 (1998)

[Itti 00] Itti, L. and Koch, C.: A saliency-based search mechanism for overt and covert shift of visual attention, *Vision Research*, Vol. 40, pp. 1489-1506 (2000)

[Judd 09] Judd, T., Ehinger, K., Durand, F. and Torralba, A.: Learning to predict where humans look, *IEEE International Conference on Vision*, Vol. 12, pp. 2106-2113 (2009)

[Kingma 14] Kingma, D. and Ba, J.: Adam: a method for stochastic optimization, *arXiv*, 1412.6980 (2014)

[Martin 01] Martin, D., Fowlkes, C. C., Tal, D. and Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *Proceedings of IEEE International Conference on Computer Vision*, Vol. 2, pp. 416-425 (2001)

[Pan 16] Pan, J., McGuiness, K., O'Connor N. E. and Grio-i-Nieto, X.: Shallow and deep convolutional networks for saliency prediction, *IEEE International Conference on Computer Vision and Pattern Recognition* (2016)

[Posner 80] Posner, M.I.: Orienting of attention, *The Quarterly Journal of Experimental Psychology*, Vol. 32, pp. 3-25 (1980)

[Ranzato 12] Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G., Dean, J., Le, Q. V. and Ng, A. Y.: Building high-level features using large scale unsupervised learning, *Proceedings of the 29th International Conference on Machine Learning*, pp. 81-88 (2012)

[Russell 14] Russell, A.F., Mihalas, S., von der Heydt, R., Niebur, E. and Etienne-Cummings, R.: A model of proto-object based saliency, *Vision Research*, Vol. 94, pp. 1-15 (2014).

[卜部 16]卜部みか、酒井宏: 物体領域知覚の皮質メカニズムの究明―図地と図方向知覚の有意性と眼球運動―, Vision(日本視覚学会 2016 年冬期大会概要集), Vol. 28, pp. 59 (2016)

[我妻 15] 我妻伸彦: 視覚皮質の層構造局所回路モデル―大規模シミュレーションが予測した視覚情報処理メカニズム―, 日本神経回路学会誌, Vol. 22, pp. 112-124 (2015)