Understanding images in biological and computer vision.

Andrew J. Schofield[1], Iain D. Gilchrist[2], Marina Bloj[3], Ales Leonardis[4], & Nicola Bellotto[5].

1. School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, a.j.schofield@bham.ac.uk
2. School of Experimental Psychology, University of Bristol, 12A Priory Road, Bristol, UK, BS8 1TU, i.d.gilchrist@bristol.ac.uk
3. School of Optometry and Vision Sciences, University of Bradford, Bradford, BD7 1DP, M.Bloj@bradford.ac.uk
4. School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT, a.leonardis@cs.bham.ac.uk
5. School of Computer Science, University of Lincoln, Brayford Pool, Lincoln, LN6 7TS, nbellotto@lincoln.ac.uk

Introduction

This issue of Interface Focus is a collection of papers arising out of a Royal Society Discussion meeting entitled 'Understanding images in biological and computer vision' held at Carlton Terrace on the 19th and 20th February, 2018. There is a strong tradition of inter-disciplinarity in the study of visual perception and visual cognition. Many of the great natural scientists including Newton [1], Young [2] and Maxwell [3] were intrigued by the relationship between light, surfaces and perceived colour considering both physical and perceptual processes. Brewster invented both the lenticular stereoscope and the binocular camera but also studied the perception of shape-from-shading [4]. More recently Marr's description of visual perception as an information processing problem led to great advances in our understanding of both biological and computer vision [5]: both the computer vision and biological vision communities have a Marr medal. The recent success of deep neural networks in classifying both the images that we see and the fMRI images that reveal the

activity in our brains during the act of seeing are both intriguing. The links between machine vision systems and biology may at sometimes be weak but the similarity of some of the operations is nonetheless striking [6].

This two day meeting brought together researchers from the fields of biological and computer vision, robotics, neuroscience, computer science and psychology to discuss the most recent developments in the field. The meeting was divided into four themes: Vision for action, Visual appearance, Vision for recognition, and Machine learning.

Vision for Action

The meeting opened with a fascinating presentation by Barbara Webb on insect vision for robot navigation considering evidence from ant foraging trajectories to support theories of visual processing (the main topic of the paper presented here [7]) and memory to support robot navigation. The paper describes a trade-off between processing the available image data to make a navigation decision and re-orienting the sensor to gain more information. In the ant such re-orientation involves the whole body and head, whereas for humans an eye movement might suffice. Casimir Ludwig picked up on this issue in his presentation on the timing of visually-guided goal-directed behaviour suggesting that the temporal trigger for action selection comes from the “ongoing task” (i.e. foveal information extraction), rather than a “race-to-threshold” between competing action plans. This result is consistent with the idea that information extraction for action selection proceeds right up to the point of the temporal trigger as shown in an earlier paper [8].

The role of visual memory and sensory input for guiding goal directed behaviour was then reviewed by Mary Hayhoe [9] who argued that the need for sensory input is reduced when reliable information is present in visual memory and that these two sources of information are combined using Bayesian processes. However she suggested that there is also a need to consider the consequences of actions. She highlighted how such estimates might be

achieved in computer vision using convolutional neural networks. The idea that information from different sources must be optimally fused to guide behaviour was the focus of Maurice Fallon's presentation [10] on the dynamic control of quadruped and bipedal robots showing how data fusion can help visual systems deal with adverse conditions such as low lighting or poor imagery.

Visual appearance, shape and illumination

The second session saw a shift in focus towards the ways in which vision is used to judge object appearance and shape. The session was opened by Anya Hurlbert who began by noting that, while without light there would be no colour, our perception of colour is not simply determined by light. Rather, colour perception is the result of a complex interaction between light, surfaces, eyes and brains [11]. Since all brains differ people may perceive the same object to have different colours as was illustrated by the failures of colour constancy seen in the now famous Dress Illusion. The issue of how humans deal with illumination variations when making object selections based on material properties was nicely reviewed by David Brainard [12]. He also presented a model of early retinal processing and showed how this can predict some but not all of the variance in human object selections; the remaining variation being due to post-retinal processing. The desire to assess material properties separately from the illumination profile was also the focus of Graham Finlayson's talk on colour and illumination in computer vision [13]. Arguing that illumination is largely seen as a problem in computer vision, Graham showed how traditional, but simple, algorithms for estimating and removing illumination from an image are biased. However, when such biases are corrected in an exposure invariant way, these relatively simple methods can rival more recent and more complex neural network based approaches.

Steven Zucker's talk shifted the emphasis away from colour constancy towards the stable perception of shape based on shading cues under changes in illumination, rendering

methods, and small changes in scene viewpoint [14]. He notes that, while much of the image changes under such manipulations, some features remain relatively constant and these tend to be contours that are used in line drawings. These invariant, critical contours may underlie the perception of shape-from-shading across a wide range of image variations.

Vision for Recognition.

Kalanit Grill-Spector started the session on vision for recognition with a talk about face recognition. In her paper here [15] she considers the neural substrates in human vision where she has characterised perceptual field structures for face processing using fMRI methods. She compares these structures to deep convolutional neural networks (DCNN) trained for face recognition and highlights a number of structural similarities and important processing differences, speculating that altering the structure of future DCNN models may improve their performance and increase our understanding of human vision. The idea of using DCNNs to understand human physiology was explored further by Rual Vicente who looked for frequency-resolved correlates of object recognition using DCNNs as an analysis tool, showing how gamma band oscillations in EEG data correlate with object recognition processes in human vision [16]. Continuing the neural network theme, Jitendra Malik presented a review of object recognition methods in machine vision including DCNNs and outlined the ways in which such systems still fall short of biological vision. He argued that further advances in computer vision may/will come from adopting developmental approaches and performing learning in active and embodied settings [17]. Much of the success of DCNNs has been at the whole object recognition level. In his talk, Shimon Ullman [18] sought to expand recognition both downward to uncover the minimal images required by humans to recognise objects and upwards to consider also the minimal images to recognise social interactions. In both cases the size of these minimal images was surprisingly small in comparison to whole object or whole scene representations.

Future Direction: Machine learning.

The final session of the meeting dealt with future directions for the field with an emphasis on machine learning. Thomas Serre opened the session by outlining a number of instances where DCNNs are incapable of learning correct classification that present humans and other less complex animals with no difficulty [19]. Considering a taxonomy of visual tasks, those involving same-different relations emerged as particularly problematic for DCNN models, suggesting that the addition of feedback mechanisms, attention and working memory may be needed to solve such problems. Andrew Fitzgibbon eschewed neural network methods altogether showing how 3D shape can be recovered from 2D silhouettes using more traditional ellipse fitting techniques.

The final two talks of the meeting turned their attention to spiking neural networks. Most neural network models ignore the fact that biological neurons communicate via discrete 'digital' spikes and thus lose critical information contained in relative spike timings. Simon Stringer noted that relative spike timing across a number of neurons – poly-synchrony – can indicate not just that two neurons are active at the same time but that the activity of one is causal to the activation of the other, this causal relationship being observed by a third neuron [21]. Further, such networks can solve the binding problem indicating which parts belong to which objects. Novel spiking neural networks such as described by Stringer [21] might benefit from novel computing architectures centred on the concept of spiking neurons such as the SpiNNaker system described by Steve Furber [22] who then outlined the benefits of event based processing and explored synaptic plasticity as a route to unsupervised on-line learning in such systems.

Conclusion

There can be little doubt that computer vision has ‘come of age’ with performance on a number of machine perception tasks, now surpassing that of human vision. These advances, enabled in part by DCNN technologies, have been paralleled by a much deeper understanding of neural processing. There remains a symbiosis between computer and human vision with DCNN tools being used to understand biological processing, revealing the similarities and dissimilarities between the two. As became clear at the Royal Society meeting in February 2018, neural networks are not always the only or even the best solution in many cases. There remain problems for which these methods are not well suited and where biological vision has the edge. Many of the papers in this special issue point the way to new advances that might circumvent some of these challenges through collaboration between the disciplines.

[1] Newton, I. (1704) Opticks. London, England: S. Smith and B. Walford.

[2] Young, T (1804). Bakerian Lecture: Experiments and calculations relative to physical optics. *Phil.Trans. Roy. Soc.* ***94****: 1–16.* https://doi.org/*10.1098/rstl.1804.0001*.

[3] Longair, M.S. (2008). Maxwell and the science of colour, *Phil. Trans. Roy. Soc. A*, 366, 1685-1696. https://doi.org/10.1098/rsta.2007.2178

[4] Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios, and intaglios into cameos, with an account of other analogous phenomena. Edinburgh Journal of Science, 4, 99-108.

[5] Marr, D. (1982) Vision, San Francisco, CA: Freeman.

[6] Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, Solomon EA, et al. (2014) Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLoS Comput Biol 10(12): e1003963. https://doi.org/10.1371/journal.pcbi.1003963

[7] Stone, T., Mangan, M., Wystrach, A., Webb, B. (2018) Rotation invariant visual processing for spatial memory in insects. Interface focus.

[8] Ludwig, C.J.H., Davies, R.L., Eckstein, M.P. (2014) Foveal analysis and peripheral selection during active visual sampling, Proc. Nat. Acad. Sci., 111, E291-E299; https://doi.org/10.1073/pnas.1313553111

[9] Hayhoe, M.M., and Matthis, J.S. (2018) Vision in the context of natural behaviour, Interface focus.

[10] Fallon M. (2018) Accurate and Robust Localization for Walking Robots Fusing Kinematics, Inertial, Vision and LIDAR, Interface Focus.

[11] Pearce B, Crichton S, Mackiewicz M, Finlayson GD, Hurlbert A (2014) Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. PLoS ONE 9(2): e87989. https://doi.org/10.1371/journal.pone.0087989

[12] Brainard, D., Cottaris, N., and Radonjic, A. (2018) The perception of color and material in naturalistic tasks, Interface Focus.

[13] Finlayson, G.D. (2018) Colour and illumination in computer vision. Interface Focus.

[14] Zucker, S. (2018) Alexandr, E., Cholewiak, S, Holtmann-Rise, d., Kunsberg, B., Fleming, R., Zucker, S.W. (2018) Color, contours, shading and shape: flow interactions reveal anchor neighborhoods, Interface Focus.

[15] Grill-Spector, K., Weiner, K., Gomez, J., Stigliani, A., Natu, V. (2018) The functional neuroanatomy of face perception: From brain measurements to deep neural networks. Interface Focus.

[16] Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J.P., Baciu, M., Kahane, P., Rheims, S., Vidal, J.R., Aru, J. Activations of deep convolutional neural network are aligned with gamma band activity of human visual cortex. bioRxiv 133694.

[17] Malik, J., Arbelaez, P., Carreira, J., Fragkiadaki, K., Girshick, R.B., Gkioxari, G., Gupta, S., Hariharan, B., Kar, A., Tulsiani, S. (2016) The three R's of computer vision: Recognition, reconstruction and reorganization. Pattern Recognition Letters 72: 4-14

[18] Ben-Yosef, G., & Ullman, S. (2018) Image interpretation above and below the object level. Interface focus.

[19] Kim, J., Ricci, M., & Serre, T. (2018) Learning same-different relations strains feedforward neural networks. Interface focus.

[20] Cashman, T.J., & Fitzgibbon, A.W. (2012) What Shape Are Dolphins? Building 3D Morphable Models from 2D Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1), 232-244, https://doi.org/10.1109/TPAMI.2012.68

[21] Isbister, J., Eguchi, A., Ahmad, N., Galeazzi, J., Buckley, M., Stringer, S. (2018) A new approach to solving the feature binding problem in primate vision. Interface F ocus.

[22] Hopkins, M., Pineda-Garcia, G., Bogdan, P.A., & Furber, S.B. (2018) Spiking Neural Networks for Computer Vision. Interface Focus.