Chapter VIII Teaching computers how to see

Further reading

(Poggio 1990, Poggio & Anselmi 2016, Serre 2019, Sutton & Barto 2017, Ullman 1984, Ullman 1996)

VIII.1. Recap and definitions

(Abbott et al 1996, Biederman 1987, Brady et al 2008, Deco & Rolls 2004b, Hung et al 2005, Keysers et al 2001, Kirchner & Thorpe 2006, Kriegeskorte et al 2008, Liu et al 2009, Logothetis & Sheinberg 1996, Myerson et al 1981, Nielsen et al 2006, Olshausen et al 1993, Orban 2004, Potter & Levy 1969, Richmond et al 1983, Riesenhuber & Poggio 1999, Rolls 1991, Serre et al 2007a, Serre et al 2007b, Standing 1973, Thorpe et al 1996)

VIII.2. Common themes in modeling the ventral visual stream

(Douglas & Martin 2004, Felleman & Van Essen 1991, Maunsell 1995, Meister 1996, Poggio & Smale 2003, Riesenhuber & Poggio 1999, Serre et al 2007a, Sutherland 1968)

VIII.3. A panoply of models

(Alom et al 2018, Amit & Geman 1999, Amit & Mascaro 2003, Beymer & Poggio 1996, Biederman 1987, Biederman & Kalocsai 1997, Brincat & Connor 2004, Connor et al 2007, Deco & Rolls 2004a, DiCarlo & Cox 2007, Edelman & Duvdevani-Bar 1997, Edelman & Poggio 1991, Eliasmith et al 2012, Foldiak 1991, Freeman & Simoncelli 2011, Hinton 1992, Hinton & Salakhutdinov 2006, Hummel & Biederman 1992, LeCun et al 1998, Marr 1982, Marr & Nishihara 1978, Mel 1997, Oram & Perrett 1994, Poggio & Edelman 1990, Rao & Ballard 1997, Riesenhuber & Poggio 2002, Schwartz 1997, Serre et al 2005a, Serre et al 2007a, Stringer et al 2006, Tanaka 1996, Tank & Hopfield 1987, Tsotsos 1990, Ullman 1996, Valiant 2000, Vogels et al 2001, Wilson 2003, Winston 1975)

VIII.4. A general scheme for object recognition tasks

(Meyers & Kreiman 2011, Poggio & Smale 2003, Singer & Kreiman 2011)

VIII.5. Bottom-up hierarchical models of the ventral visual stream

(Amit & Mascaro 2003, Anselmi et al , Cadieu et al 2007, Callaway 1998, Deco & Rolls 2004b, Fukushima 1980, Hung et al 2005, Krizhevsky et al 2012, Lampl et al 2004, LeCun et al 1998, Mutch & Lowe 2006, Olshausen et al 1993,

Riesenhuber & Poggio 1999, Serre et al 2005a, Serre et al 2007a, Serre et al 2007b, Serre et al 2005b, Simonyan & Zisserman 2014, Sutherland 1968, Szegedy et al 2015, Virga 1989, Wallis & Rolls 1997)

VIII.6. Learning the weights

(Bishop 1995, Goodfellow et al 2016, Hinton 1992, Hinton 2007, Hinton 2010, LeCun et al 2015, Lillicrap et al 2020, Rumelhart et al 1986, Sutton 1988)

VIII.7. Labeled databases

(Alom et al 2018, Barbu et al 2019, Johnson et al 2016, Kay et al 2017, Krizhevsky et al 2012, Lin et al 2015, Russakovsky et al 2014, Soomro et al 2012)

VIII.8. Cross-validation is essential

VIII.9. A cautionary note: lots of parameters!

(Casper S et al 2019, Geirhos et al 2018, Linsley et al 2018, Poggio et al 2019, Recht et al 2019, Zhang et al 2017)

VIII.10. A famous example: digit recognition in a feed-forward network trained by gradient descent

(LeCun et al 1998)

VIII.11. A deep convolutional neural network in action

(Alom et al 2018, He et al 2018, He et al 2015, Krizhevsky et al 2012, Simonyan & Zisserman 2014, van der Maaten & Hinton 2008)

VIII.12. To err is human and algorithmic

(Borji & Itti 2014, Fleuret et al 2011, Geirhos et al 2018, Linsley et al 2017, Rajalingham et al 2018)

VIII.13. Predicting eye movements

(Adeli & Zelinsky 2018, Borji & Itti 2013, Eckstein et al 2000, Fecteau & Munoz 2006, Findlay 1997, Hamker 2005, Itti & Koch 2000, Itti & Koch 2001, Jonides et al 1982, Juan et al 2004, Koch & Ullman 1985, Lee et al 1999, Miconi et al 2016, Mirpour et al 2009, Navalpakkam & Itti 2007, Tsotsos 2011, Walther et al 2002, Walther & Koch 2007, Zelinsky 2008, Zhang et al 2018)

VIII.14. Predicting neuronal firing rates

(Abbasi-Asl et al 2018, Cadieu et al 2007, Carandini & Heeger 2011, Freeman et al 2013, Pospisil et al 2018, Yamins et al 2014)

VIII.15. All models are wrong, some are useful

(Felleman & Van Essen 1991, Markov et al 2014)

VIII.16. Horizontal and top-down signals in visual recognition

(Ahissar & Hochstein 2004, Bullier 2001, Callaway 2004, Chikkerur et al 2009, Compte & Wang 2006, Deco & Rolls 2004b, Deco & Rolls 2005, Douglas & Martin 2004, Felleman & Van Essen 1991, Friston 2003, Frith & Dolan 1997, Gilbert & Li 2013, Hochstein & Ahissar 2002, Hopfield 1982, Itti & Koch 2001, Jones et al 1997, Kar et al 2019, Kersten & Yuille 2003, Kim et al 2019, Kreiman & Serre 2020, Lamme & Roelfsema 2000, Lamme et al 1998, Lee & Mumford 2003, Markov et al 2014, Mely et al 2018, Mumford 1992, Olshausen et al 1993, Panzeri et al 2001, Rao 2004, Rao 2005, Rao et al 2002, Tang et al 2014, Tang et al 2018, Tsotsos 1990, Ullman 1995, Yang et al 2017, Yuille & Kersten 2006)

VIII.17. Predictive coding

(Bastos et al 2012, Choi et al 2018, Hochreiter & Schmidhuber 1997, Lotter et al 2016, Lotter et al 2017, Lotter et al 2020, Nassi et al 2014, Nassi et al 2013, Rao & Ballard 1999, Rao et al 2002)

VIII.18. References

Abbasi-Asl R, Chen Y, Bloniarz A, Oliver M, Willmore B, et al. 2018. The DeepTune framework for modeling and characterizing neurons in visual cortex area V4. *Biorxiv* 1101/465534v1

Abbott LF, Rolls ET, Tovee MJ. 1996. Representational capacity of face coding in monkeys. *Cerebral Cortex* 6: 498-505.

Adeli H, Zelinsky GJ. *CVPR2018*.

Ahissar M, Hochstein S. 2004. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8: 457-64

Alom M, Taha T, Yakopcic C, Westberg S, Sidike P, et al. 2018. The history began from AlexNet: a comprehensive survey on deep learning approaches. *arXiv* 1803.01164

Amit Y, Geman D. 1999. A Computational Model for Visual Selection. *Neural Computational* 11: 24

Amit Y, Mascaro M. 2003. An integrated network for invariant visual detection and recognition. *Vision research* 43: 2073-88

Anselmi F, Leibo J, Rosasco L, Mutch J, Tacchetti A, Poggio T. Unsupervised Learning of Invariant Representations in Visual Cortex (and in Deep Learning Architectures. *PNAS*: 14

Barbu A, Mayo D, Alveiro J, Luo W, Wang C, et al. 2019. *ObjectNet: A large-scale bias-controlled dataset for pushing the limits of object recognition models.* Presented at NeurIPS,

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. 2012. Canonical microcircuits for predictive coding. *Neuron* 76: 695-711

Beymer D, Poggio T. 1996. Image representations for visual learning. *Science* 272: 1905-09

Biederman I. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review* 24: 115-47

Biederman I, Kalocsai P. 1997. Neurocomputational Bases of Object and Face Recognition. *Philosophical Transitions: Biological Sciences* 352: 16

Bishop CM. 1995. *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press.

Borji A, Itti L. 2013. State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35: 185-207

Borji A, Itti L. 2014. *Human vs. computer in scene and object recognition.* Presented at CVPR,

Brady TF, Konkle T, Alvarez GA, Oliva A. 2008. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Science U S A* 105: 14325-9

Brincat SL, Connor CE. 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience* 7: 880-6

Bullier J. 2001. Integrated model of visual processing. *Brain Res Brain Res Rev* 36: 96-107

Cadieu C, Kouh M, Pasupathy A, Connor C, Riesenhuber M, Poggio T. 2007. A model of V4 shape selectivity and invariance. *Journal of Neurophysiology* 98: 1733-50

Callaway EM. 1998. Local circuits in primary visual cortex of the macaque monkey. *Annu Rev Neurosci* 21: 47-74

Callaway EM. 2004. Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Netw* 17: 625-32

Carandini M, Heeger DJ. 2011. Normalization as a canonical neural computation. *Nature Reviews Neuroscience* 13: 51-62

Casper S, Boix X, D'Amario V, Guo L, Schrimpf M, et al. 2019. Removable and/or repeatable units emerge in overparametrized neural networks. *arXiv* 1912.047832

Chikkerur S, Serre T, Poggio T. 2009. A Bayesian inference theory of attention: neuroscience and algorithms, MIT, Cambridge

Choi H, Pasupathy A, Shea-Brown E. 2018. Predictive Coding in Area V4: Dynamic Shape Discrimination under Partial Occlusion. *Neural Comput* 30: 1209-57

Compte A, Wang XJ. 2006. Tuning curve shift by attention modulation in cortical neurons: a computational study of its mechanisms. *Cerebral Cortex* 16: 761-78

Connor CE, Brincat SL, Pasupathy A. 2007. Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology* 17: 140-7

Deco G, Rolls ET. 2004a. *Computational Neuroscience of Vision*. Oxford Oxford University Press.

Deco G, Rolls ET. 2004b. A neurodynamical cortical model of visual attention and invariant object recognition. *Vision research* 44: 621-42

Deco G, Rolls ET. 2005. Attention, short-term memory, and action selection: a unifying theory. *Prog Neurobiol* 76: 236-56

DiCarlo JJ, Cox DD. 2007. Untangling invariant object recognition. *Trends Cogn Sci* 11: 333-41

Douglas RJ, Martin KA. 2004. Neuronal circuits of the neocortex. *Annu Rev Neurosci* 27: 419-51

Eckstein MP, Thomas JP, Palmer J, Shimozaki SS. 2000. A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & psychophysics* 62: 425-51

Edelman S, Duvdevani-Bar S. 1997. A model of visual recognition and categorization. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 352: 1191-202

Edelman S, Poggio T. 1991. Models of object recognition. *Current Opinion in Neurobiology* 1: 270-3

Eliasmith C, Stewart T, Choo X, Bekolay T, DeWolf T, et al. 2012. A Large-Scale Model of the Functioning Brain. *Science* 338: 1-38

Fecteau JH, Munoz DP. 2006. Salience, relevance, and firing: a priority map for target selection. *Trends Cogn Sci* 10: 382-90

Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1: 1-47

Findlay JM. 1997. Saccade target selection during visual search. *Vision research* 37: 617-31

Fleuret F, Li T, Dubout C, Wampler EK, Yantis S, Geman D. 2011. Comparing machines and humans on a visual categorization test. *Proceedings of the National Academy of Sciences of the United States of America* 108: 17621-5

Foldiak P. 1991. Learning Invariance from Transformation Sequences. *Neural Computation* 3: 194-200

Freeman J, Simoncelli EP. 2011. Metamers of the ventral stream. *Nature Neuroscience* 14: 1195-201

Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA. 2013. A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience* 16: 974-81

Friston K. 2003. Learning and inference in the brain. *Neural Netw* 16: 1325-52

Frith C, Dolan RJ. 1997. Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society of London* 352: 1221-30

Fukushima K. 1980. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36: 193-202

Geirhos R, Temme C, Rauber J, Schütt H, Bethge M, Wichmann F. 2018. Generalisation in humans and deep neural networks. *arXiv* 1808.08750

Gilbert CD, Li W. 2013. Top-down influences on visual processing. *Nat Rev Neurosci* 14: 350-63

Goodfellow I, Bengio Y, Courville A. 2016. *Deep learning*. Cambridge, Massachusetts: The MIT Press. xxii, 775 pages pp.

Hamker FH. 2005. The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex* 15: 431-47

He K, Gkioxari G, Dollar P, Girshick R. 2018. *Mask R-CNN.* Presented at IEEE Trans. Pattern Anal. Mach Intell.,

He K, Zhang X, Ren S, Sun J. 2015. Deep residual learning for image recognition. *arXiv* 1512.03385

Hinton G. 1992. How neural networks learn from experience. *Scientific American* 267: 145-51

Hinton GE. 2007. Learning multiple layers of representation. *Trends Cogn Sci* 11: 428-34

Hinton GE. 2010. Learning to represent visual input. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365: 177-84

Hinton GE, Salakhutdinov RR. 2006. Reducing the dimensionality of data with neural networks. *Science* 313: 504-7

Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Computation* 9: 1735-80

Hochstein S, Ahissar M. 2002. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36: 791-804

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79: 2554-58

Hummel JE, Biederman I. 1992. Dynamic binding in a neural network for shape recognition. *Psychol Rev* 99: 480-517

Hung CP, Kreiman G, Poggio T, DiCarlo JJ. 2005. Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. *Science* 310: 863-66

Itti L, Koch C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research* 40: 1489-506

Itti L, Koch C. 2001. Computational modelling of visual attention. *Nat Rev Neurosci* 2: 194-203

Johnson J, Hariharan B, van der Maaten L, Fei-Fei L, Zitnick C, Girshick R. 2016. CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning. *arXiv* 1612.06890

Jones MJ, Sinha P, Vetter T, Poggio T. 1997. Top-down learning of low-level vision tasks. *Current Biology* 7: 991-4

Jonides J, Irwin DE, Yantis S. 1982. Integrating visual information from successive fixations. *Science* 215: 192-4

Juan CH, Shorter-Jacobi SM, Schall JD. 2004. Dissociation of spatial attention and saccade preparation. *Proceeding of the National Acedemy of Sciences of the United States of America* 101: 15541

Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ. 2019. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience* 22: 974-83

Kay W, Carreira J, Simonyan K, Zhang B, Hiller C, et al. 2017. The kinetics human action video dataset. *arXiv*: arXiv:1705.06950v1

Kersten D, Yuille A. 2003. Bayesian models of object perception. *Current Opinion in Neurobiology* 13: 150-8

Keysers C, Xiao DK, Foldiak P, Perret DI. 2001. The speed of sight. *Journal of Cognitive Neuroscience* 13: 90-101

Kim J, Linsley D, Thakkar K, Serre T. 2019. Disentangling neural mechanisms for perceptual grouping. *arXiv* 1906.01558

Kirchner H, Thorpe SJ. 2006. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision research* 46: 1762-76

Koch C, Ullman S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol* 4: 219-27

Kreiman G, Serre T. 2020. Beyond the feedforward sweep: feedback computations in the visual cortex. *This Year in Cognitive Neuroscience*

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, et al. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60: 1126-41

Krizhevsky A, Sutskever I, Hinton G. 2012. *ImageNet Classification with Deep Convolutional Neural Networks.* Presented at NIPS, Montreal

Lamme VA, Roelfsema PR. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 23: 571-9

Lamme VAF, Super H, Spekreijse H. 1998. Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinon in Neurobiology* 8: 529-35

Lampl I, Ferster D, Poggio T, Riesenhuber M. 2004. Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J Neurophysiol* 92: 2704-13

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521: 436-44

LeCun Y, Bottou L, Bengio Y, Haffner P. 1998. Gradient-based learning applied to document recognition. *Proc of the IEEE* 86: 2278-324

Lee D, Itti L, Koch C, Braun J. 1999. Attention activates winner-take-all competition among visual filters. *Nature Neuroscience* 2: 375-81

Lee TS, Mumford D. 2003. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis* 20: 1434-48

Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G. 2020. Backpropagation and the brain. *Nat Rev Neurosci*

Lin T, Maire M, Belongie S, Bourdev L, Girshick R, et al. 2015. *Microsoft COCO: Common Objects in Context.* Presented at arxiv:1405.0312v3,

Linsley D, Eberhardt S, Sharma T, Gupta P, Serre T. *IEEE ICCV Workshop on the Mutual Benefit of Cognitive and Computer Vision2017*.

Linsley D, Kim J, Berson D, Serre T. 2018. Robust neural circuit reconstruction from serial electron microscopy with convolutional recurrent networks. *arXiv* 1811.11356

Liu H, Agam Y, Madsen JR, Kreiman G. 2009. Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62: 281-90

Logothetis NK, Sheinberg DL. 1996. Visual object recognition. *Annual Review of Neuroscience* 19: 577-621

Lotter W, Kreiman G, Cox D. *International Conference on Learning Represenations (ICLR), Puerto Rico, 2016*.

Lotter W, Kreiman G, Cox D. *International Conference on Learning Representations (ICLR), Toulon, France, 2017,* arXiv:1605.08104.

Lotter W, Kreiman G, Cox D. 2020. A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nature Machine Learning* 2: 210-19

Markov NT, Ercsey-Ravasz MM, Ribeiro Gomes AR, Lamy C, Magrou L, et al. 2014. A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex. *Cerebral Cortex* 24: 17-36

Marr D. 1982. *Vision*. San Francisco: Freeman publishers.

Marr D, Nishihara HK. 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc Lond B Biol Sci* 200: 269-94

Maunsell JHR. 1995. The brain's visual world: representation of visual targets in cerebral cortex. *Science* 270: 764-69

Meister M. 1996. Multineuronal Codes in Retinal Signaling. *PNAS* 93: 609-14

Mel B. 1997. SEEMORE: Combining color, shape and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation* 9: 777

Mely D, Linsley D, Serre T. 2018. Opponent surrounds explain diversity of contextual phenomena across visual modalities. *Psychological Review*

Meyers EM, Kreiman G. 2011. Tutorial on Pattern Classification in Cell Recordings In *Understanding visual population codes*, ed. N Kriegeskorte, G Kreiman. Boston: MIT Press

Miconi T, Groomes L, Kreiman G. 2016. There's Waldo! A Normalization Model of Visual Search Predicts Single-Trial Human Fixations in an Object Search Task. *Cerebral Cortex* 26: 3064-82

Mirpour K, Arcizet F, Ong WS, Bisley J. 2009. Been there, seen that: a neural mechanism for performing efficient visual search. *Journal of Neurophysiology* 102: 3481-91

Mumford D. 1992. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66: 241-51

Mutch J, Lowe D. *CVPR, New York, 2006*: 11-18.

Myerson J, Miezin F, Allman J. 1981. Binocular rivalry in macaque monkeys and humans: a comparative study in perception. *Behavioral Analysis Letters* 1: 149-59

Nassi J, Gomez-Laberge C, Kreiman G, Born R. 2014. Corticocortical feedback increases the spatial extent of normalization *Frontiers in Systems Neuroscience* 8

Nassi JJ, Lomber SG, Born RT. 2013. Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *Journal of Neuroscience* 33: 8504-17

Navalpakkam V, Itti L. 2007. Search goal tunes visual features optimally. *Neuron* 53: 605-17

Nielsen KJ, Logothetis NK, Rainer G. 2006. Discrimination strategies of humans and rhesus monkeys for complex visual displays. *Current Biology* 16: 814-20

Olshausen BA, Anderson CH, Van Essen DC. 1993. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience* 13: 4700-19

Oram M, Perrett D. 1994. Modeling object recognition from neurobiological constraints. *Neural Networks* 7: 945-72

Orban GA, Van Essen, D., Vanduffel, W. 2004. Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences* 8: 315-24

Panzeri S, Rolls ET, Battaglia F, Lavis R. 2001. Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Network* 12: 423-40

Poggio T. 1990. A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology* 55: 899-910

Poggio T, Anselmi F. 2016. *Visual cortex and deep networks*. Cambridge, MA: MIT Press.

Poggio T, Edelman S. 1990. A network that learns to recognize three-dimensional objects. *Nature* 343: 263-6

Poggio T, Kur G, Banburski A. 2019. Double descent in the condition number. *arXiv* 1912.06190

Poggio T, Smale S. 2003. The mathematics of learning: dealing with data. *Notices of the AMS* 50: 537-44

Pospisil DA, Pasupathy A, Bair W. 2018. 'Artiphysiology' reveals V4-like shape tuning in a deep network trained for image classification. *eLife* 7

Potter M, Levy E. 1969. Recognition memory for a rapid sequence of pictures. *Journal of experimental psychology* 81: 10-15

Rajalingham R, Issa EB, Bashivan P, Kar K, Schmidt K, DiCarlo JJ. 2018. Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks. *Journal of Neuroscience* 38: 7255-69

Rao R, Ballard D. 1997. Dynamic Model of Visual Recognition Predicts Neural Response Properties in the Visual Cortex. *Neural Computational* 9

Rao RP. 2004. Bayesian computation in recurrent neural circuits. *Neural Comput* 16: 1-38

Rao RP. 2005. Bayesian inference and attentional modulation in the visual cortex. *Neuroreport* 16: 1843-8

Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* 2: 79-87

Rao RPN, Olshausen BA, Lewicki MS, eds. 2002. *Probabilistic Models of the Brain: Perception and Neural Function*. Cambridge: MIT Press.

Recht B, Roelofs R, Schmidt L, Shankar V. 2019. Do ImageNet Classifiers Generalize to ImageNet? *arXiv* 1902.10811

Richmond B, Wurtz R, Sato T. 1983. Visual responses in inferior temporal neurons in awake Rhesus monkey. *Journal of Neurophysiology* 50: 1415-32

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nature Neuroscience* 2: 1019-25

Riesenhuber M, Poggio T. 2002. Neural mechanisms of object recognition. *Current Opinion in Neurobiology* 12: 162-8

Rolls E. 1991. Neural organization of higher visual functions. *Current Opinion in Neurobiology* 1: 274-78

Rumelhart DE, Hinton G, Williams RJ. 1986. Learning representations by back-propagating errors. *Nature* 323: 533-36

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, et al. 2014. *ImageNet Large Scale Visual Recognition Challenge.* Presented at CVPR,

Schwartz EL. 1997. Computational studies of the spatial architecture of primate visual cortex. *Cerebral Cortex* 10: 359-411

Serre T. 2019. Deep learning: the good, the bad and the ugly. *Annual Review of Vision* 5: 399-426

Serre T, Kouh M, Cadieu C, Knoblich U, Kreiman G, Poggio T. 2005a. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex, MIT, Boston

Serre T, Kreiman G, Kouh M, Cadieu C, Knoblich U, Poggio T. 2007a. A quantitative theory of immediate visual recognition. *Progress In Brain Research* 165C: 33-56

Serre T, Oliva A, Poggio T. 2007b. Feedforward theories of visual cortex account for human performance in rapid categorization. *PNAS* 104: 6424-29

Serre T, Wolf L, Poggio T. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, 2005b*. IEEE Computer Society Press.

Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv* 1409.1556

Singer J, Kreiman G. 2011. Introduction to Statistical Learning and Pattern Classification In *Visual Population Codes*, ed. N Kriegeskorte, G Kreiman. Boston: MIT Press

Soomro K, Zamir A, Shah M. 2012. UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. *arXiv*: 1212.0402

Standing L. 1973. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* 25: 207-22

Stringer SM, Perry G, Rolls ET, Proske JH. 2006. Learning invariant object recognition in the visual system with continuous transformations. *Biol Cybern* 94: 128-42

Sutherland NS. 1968. Outlines of a theory of visual pattern recognition in animals and man. *Proc R Soc Lond B Biol Sci* 171: 297-317

Sutton R. 1988. Learning to predict by the methods of temporal differences. *Machine Learning* 3: 9-44

Sutton R, Barto A. 2017. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. 2015. Rethinking the inception architecture for computer vision. *arXiv* 1512.005673v3

Tanaka K. 1996. Inferotemporal cortex and object vision. *Annual Review of Neuroscience* 19: 109-39

Tang H, Buia C, Madhavan R, Madsen J, Anderson W, et al. 2014. Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron* 83: 736-48

Tang H, Lotter W, Schrimpf M, Paredes A, Ortega J, et al. 2018. Recurrent computations for visual pattern completion. *PNAS* 115: 8835-40

Tank D, Hopfield J. 1987. Collective computation in neuron-like circuits. *Scientific American* 257: 104-&

Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature* 381: 520-22

Tsotsos J. 1990. Analyzing Vision at the Complexity Level. *Behavioral and Brain Sciences* 13-3: 423-45

Tsotsos JK. 2011. *A computational perspective on visual attention*. Cambridge, Mass.: MIT Press. xvi, 308 p. pp.

Ullman S. 1984. Visual routines. *Cognition* 18: 97-159

Ullman S. 1995. Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex* 5: 1-11

Ullman S. 1996. *High-Level Vision*. Cambridge, MA: The MIT Press.

Valiant LG. 2000. *Circuits of the Mind*. New York: Oxford University Press.

van der Maaten L, Hinton G. 2008. Visualizing High-Dimensional Data Using t-SNE. *J. Machine Learning Res.* 9: 2579-605

Virga A, Rockland, KS. 1989. Terminal Arbors of Individual "Feedback" Axons Projecting from Area V2 to V1 in the Macaque Monkey: A Study Using Immunohistochemistry of Anterogradely Transported Phaseolus vulgaris-leucoagglutinin. *The Journal of Comparative Neurology* 285: 54-72

Vogels R, Biederman I, Bar M, Lorincz A. 2001. Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *J Cogn Neurosci* 13: 444-53

Wallis G, Rolls ET. 1997. Invariant face and object recognition in the visual system. *Progress in Neurobiology* 51: 167-94

Walther D, Itti L, Riesenhuber M, Poggio T, Koch C. 2002. Attentional selection for object recognition - a gentle way. *Biologically Motivated Computer Vision* 2525: 472-79

Walther DB, Koch C. 2007. Attention in hierarchical models of object recognition. *Prog Brain Res* 165: 57-78

Wilson HR. 2003. Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences of the United States of America* 100: 14499-503

Winston P. 1975. Learning structural descriptions from examples In *The psychology of computer vision*, ed. P Winston, pp. 157-209. London: McGraw-Hill

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* 111: 8619-24

Yang G, Joglekar M, Song H, Newsome W, Wang X. 2017. Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience* 22: 297-306

Yuille A, Kersten D. 2006. Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci* 10: 301-8

Zelinsky GJ. 2008. A theory of eye movements during target acquisition. *Psychol Rev* 115: 787-835

Zhang C, Bengio Y, Hardt M, Recht B, Vinyals O. 2017. Understanding deep learning requires rethinking generalization. *arXiv* 1611.03530

Zhang M, Feng J, Ma KT, Lim JH, Zhao Q, Kreiman G. 2018. Finding any Waldo: zero-shot invariant and efficient visual search. *Nature Communications* 9: 3730