

## **Chapter IX Towards a world with intelligent machines that can interpret the visual world**

### **Further reading:**

(Goodfellow et al 2016, Hassabis et al 2017, Nielsen 2019)

### **IX.1. The Turing test for vision**

(Good 1966, Turing 1950)

### **IX.2. Computer vision everywhere**

(Kay et al 2017, Lichtman et al 2014, Lin et al 2015, Redmon et al 2016, Soomro et al 2012)

### **IX.3. Incorporating temporal information using videos**

(Ben-Yosef et al 2020, Donahue et al 2017, Feichtenhofer et al 2016, Foldiak 1991, Giese & Poggio 2003, Hara et al 2017, Hochreiter & Schmidhuber 1997, Karpathy et al 2014, Kundu et al 2016, Lotter et al 2015, Lotter et al 2016, Lotter et al 2017, Lotter et al 2020, Olson & Pettigrew 1974, Simonyan & Zisserman 2014a, Tran et al 2015, Tran et al 2017, Wiskott & Sejnowski 2002)

### **IX.4. Major progress in object classification**

(Borenstein et al 2004, Borji & Itti 2014, Girshick 2015, He et al 2018, He et al 2015, Hochreiter & Schmidhuber 1997, Karpathy et al 2014, Kim et al 2019, Krizhevsky et al 2012, Li et al 2004, Lin et al 2015, Russakovsky et al 2014, Simonyan & Zisserman 2014b, Szegedy et al 2015, Tan & Le 2019, Zhang et al 2019, Zoph et al 2018)

### **IX.5. Real-world applications of computer vision algorithms for object classification**

(Elmore et al 2009, Kim & Brunner 2017, Linsley et al 2018, Lotter et al 2017, McKinney et al 2020, Nguyen et al 2016, Phillips et al 2018, Poplin et al 2018, Reardon 2020, Shallue & Vanderburg 2017, Taigman et al 2014, Van Horn et al 2018)

### **IX.6. Coming soon to a computer vision algorithm near you**

### **IX.7. Computer vision to help people with visual disabilities**

## **IX.8. Deep convolutional neural networks work outside of vision too**

## **IX.9. Image generators and GANs**

(Dosovitskiy & Brox 2016, Goodfellow et al 2014)

## **IX.10. DeepDream and XDream**

(Bashivan et al 2019, Dosovitskiy & Brox 2016, Erhan et al 2009, Ilyas et al 2018, Mordvintsev et al 2015, Ponce et al 2019, Walker et al 2019, Xiao & Kreiman 2020)

## **IX.11. Reflections on cross-validation and extrapolation**

(Geirhos et al 2018, Jacquot et al 2020, Recht et al 2019)

## **IX.12. Adversarial images**

(Szegedy et al 2014, Xie et al 2019, Zhou & Firestone 2019)

## **IX.13. Deceptively simple tasks that challenge computer vision algorithms**

(Fleuret et al 2011, Johnson et al 2016, Kim et al 2018)

## **IX.14. Challenges ahead**

(Kreiman & Serre 2020, Serre 2019)

## **References**

- Bashivan P, Kar K, DiCarlo JJ. 2019. Neural population control via deep image synthesis. *Science* 364
- Ben-Yosef G, Kreiman G, Ullman S. 2020. Minimal videos: Trade-off between spatial and temporal information in human and machine vision. *Cognition* In Press
- Borenstein E, Sharon E, Ullman S. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, 2004.*
- Borji A, Itti L. 2014. *Human vs. computer in scene and object recognition.* Presented at CVPR,
- Donahue J, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, et al. 2017. Long-term recurrent convolutional networks for visual recognition and description. *IEEE transactions on pattern analysis and machine intelligence*
- Dosovitskiy A, Brox T. 2016. Generating Images with Perceptual Similarity Metrics based on Deep Networks. *arXiv* 160202644

- Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, et al. 2009. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology* 253: 641-51
- Erhan D, Bengio Y, Courville A, Vincent P. 2009. Visualizing higher-layer features of a deep network, University of Montreal
- Feichtenhofer C, Pinz A, Wildes R. 2016. *Spatiotemporal residual networks for video action recognition*.
- 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
- Geirhos R, Temme C, Rauber J, Schütt H, Bethge M, Wichmann F. 2018. Generalisation in humans and deep neural networks. *arXiv* 1808.08750
- Giese MA, Poggio T. 2003. Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4: 179-92
- Girshick R. 2015. *Fast R-CNN*.
- Good I. 1966. Speculations Concerning the First Ultraintelligent Machine. *Advances in Computers* 6: 31-88
- Goodfellow I, Bengio Y, Courville A. 2016. *Deep learning*. Cambridge, Massachusetts: The MIT Press. xxii, 775 pages pp.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, et al. 2014. Generative Adversarial Nets.
- Hara K, Kataoka H, Satoh Y. 2017. Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet. *arXiv* 1711.09577
- Hassabis D, Kumaran D, Summerfield C, Botvinick M. 2017. Neuroscience-Inspired Artificial Intelligence. *Neuron* 95: 245-58
- 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
- Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Computation* 9: 1735-80
- Ilyas A, Engstrom L, Athalye A, Lin J. *International Conference on Machine Learning (ICML)2018*, 80: 2137-46.
- Jacquot V, Ying Z, Kreiman G. 2020. *Can deep learning recognize subtle human activities?* Presented at CVPR,
- 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
- Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. 2014. Large-Scale Video Classification with Convolutional Neural Networks. *IEEE conference on Computer Vision and Pattern Recognition*

- Kay W, Carreira J, Simonyan K, Zhang B, Hiller C, et al. 2017. The kinetics human action video dataset. *arXiv*: arXiv:1705.06950v1
- Kim E, Brunner R. 2017. Star–galaxy classification using deep convolutional neural networks *Monthly Notices of the Royal Astronomical Society* 464: 4463-75
- Kim J, Linsley D, Thakkar K, Serre T. 2019. Disentangling neural mechanisms for perceptual grouping. *arXiv* 1906.01558
- Kim J, Ricci M, Serre T. 2018. *Not-So-CLEVR: learning same–different relations strains feedforward neural networks*. . Presented at Interface Focus theme issue on “Understanding images in biological and computer vision.”,
- Kreiman G, Serre T. 2020. Beyond the feedforward sweep: feedback computations in the visual cortex. *This Year in Cognitive Neuroscience*
- Krizhevsky A, Sutskever I, Hinton G. 2012. *ImageNet Classification with Deep Convolutional Neural Networks*. Presented at NIPS, Montreal
- Kundu A, Vineet V, Koltun V. 2016. Feature space optimization for semantic video segmentation. *Computer Vision and Pattern Recognition*
- Li F-F, Fergus R, Perona P. *CVPR 2004*.
- Lichtman JW, Pfister H, Shavit N. 2014. The big data challenges of connectomics. *Nature Neuroscience* 17: 1448-54
- 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, Kim J, Berson D, Serre T. 2018. Robust neural circuit reconstruction from serial electron microscopy with convolutional recurrent networks. *arXiv* 1811.11356
- Lotter W, Kreiman G, Cox D. *International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2015*.
- Lotter W, Kreiman G, Cox D. *International Conference on Learning Representations (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
- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, et al. 2020. International evaluation of an AI system for breast cancer screening. *Nature* 577: 89-94
- Mordvintsev A, Olah C, Tyka M. 2015. DeepDream - a code example for visualizing Neural Networks. In *Google Research*. Mountain View: Google
- Nguyen A, Dosovitskiy A, Yosinski J, Brox T, Clune J. *NIPS2016*.
- Nielsen M. 2019. Neural networks and deep learning.
- Olson CR, Pettigrew JD. 1974. Single units in visual cortex of kittens reared in stroboscopic illumination. *Brain Res* 70: 189-204
- Phillips PJ, Yates AN, Hu Y, Hahn CA, Noyes E, et al. 2018. Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences of the United States of America* 115: 6171-76

- Ponce CR, Xiao W, Schade PF, Hartmann TS, Kreiman G, Livingstone M. 2019. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell*
- Poplin R, Varadarajan A, Blumer K, Liu Y, McConnell M, et al. 2018. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* 2: 158-64
- Reardon S. 2020. Rise of robot radiologists. *Scientific American*
- Recht B, Roelofs R, Schmidt L, Shankar V. 2019. Do ImageNet Classifiers Generalize to ImageNet? *arXiv* 1902.10811
- Redmon J, Divvala S, Girshick R, Farhadi A. 2016. *You only look once: Unified, real-time object detection*. Presented at CVPR,
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, et al. 2014. *ImageNet Large Scale Visual Recognition Challenge*. Presented at CVPR,
- Serre T. 2019. Deep learning: the good, the bad and the ugly. *Annual Review of Vision* 5: 399-426
- Shallue C, Vanderburg A. 2017. Identifying Exoplanets with Deep Learning: A Five Planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *arXiv* 1712.05044
- Simonyan K, Zisserman A. 2014a. Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*
- Simonyan K, Zisserman A. 2014b. Very deep convolutional networks for large-scale image recognition. *arXiv* 1409.1556
- Soomro K, Zamir A, Shah M. 2012. UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. *arXiv*: 1212.0402
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. 2015. Rethinking the inception architecture for computer vision. *arXiv* 1512.005673v3
- Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, et al. 2014. *Intriguing properties of neural networks*. Presented at International Conference on Learning Representations,
- Taigman Y, Yang M, Ranzato M, Wolf L. 2014. *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*. . Presented at IEEE Conf. on Computer Vision and Pattern Recognition (CVPR),
- Tan M, Le Q. 2019. *EfficientNet: Rethinking model scaling for convolutional neural networks*. Presented at arxiv: 1905.11946v3,
- Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. *IEEE International Conference on Computer Vision2015*.
- Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. 2017. A Closer Look at Spatiotemporal Convolutions for Action Recognition. *arXiv* 1711.11248
- Turing A. 1950. Computing Machinery and Intelligence. *Mind* LIX: 433-60
- Van Horn G, Oisin M, Song Y, Cui Y, Chen S, et al. 2018. *The inaturalist species classification and detection dataset*. Presented at CVPR,
- Walker E, Sinz F, Froudarakis E, Fahey P, Muhammad T, et al. 2019. Inception in visual cortex: in vivo-silico loops reveal most exciting images. *Biorxiv*
- Wiskott L, Sejnowski TJ. 2002. Slow feature analysis: unsupervised learning of invariances. *Neural Comput* 14: 715-70

- Xiao W, Kreiman G. 2020. XDream: Finding preferred stimuli for visual neurons using generative networks and gradient-free optimization. *PLoS Computational Biology* 16: e1007973
- Xie C, Tan M, Gong B, Wang J, Yuille A, Le Q. 2019. Adversarial Examples Improve Image Recognition. *arXiv* 1911:09665
- Zhang M, Tseng C, Kreiman G. 2019. Putting visual object recognition in context. *arXiv* 1911.07349
- Zhou Z, Firestone C. 2019. Humans can decipher adversarial images. *Nature Communications* 10
- Zoph B, Vasudevan V, Shlens J, Le Q. 2018. *Learning Transferable Architectures for Scalable Image Recognition*. Presented at arXiv:1707.07012v4,