COMPUTER VISION

From active perception to deep learning

Danica Kragic is the Director of the Centre for Autonomous Systems and a Professor of Computer Science at the KTH Royal Institute of Technology, Stockholm, Sweden. Email: dani@kth.se

When the area of computer vision started almost half a century ago (early 1970), it envisioned the development of artificial vision systems with human-level abilities. In 1966, according to a well-known story (*1*), Marvin Minsky at MIT tasked his student Gerald Jay Sussman to "spend the summer linking a camera to a computer and getting the computer to describe what it saw." However, building seeing machines was more difficult than anticipated. In early 1980, David Marr stated that "in the 1960s, almost no one realized that machine vision was difficult. The reason for this misperception is that we humans are ourselves so good at vision" (*2*). Most of us will agree that Marr made an important point, given that we still do not have artificial vision systems that demonstrate the flexibility, scalability, and adaptability of the human visual system.

There are many areas of computer vision that have been receiving a lot of attention recently and where important progress has been made, especially during the past decade. Most of this progress is because of deep learning and includes tasks such as object detection, image segmentation, recognition, categorization, image generation, Internet-scale image search, video search, three-dimensional human pose estimation, computational photography, and scene understanding.

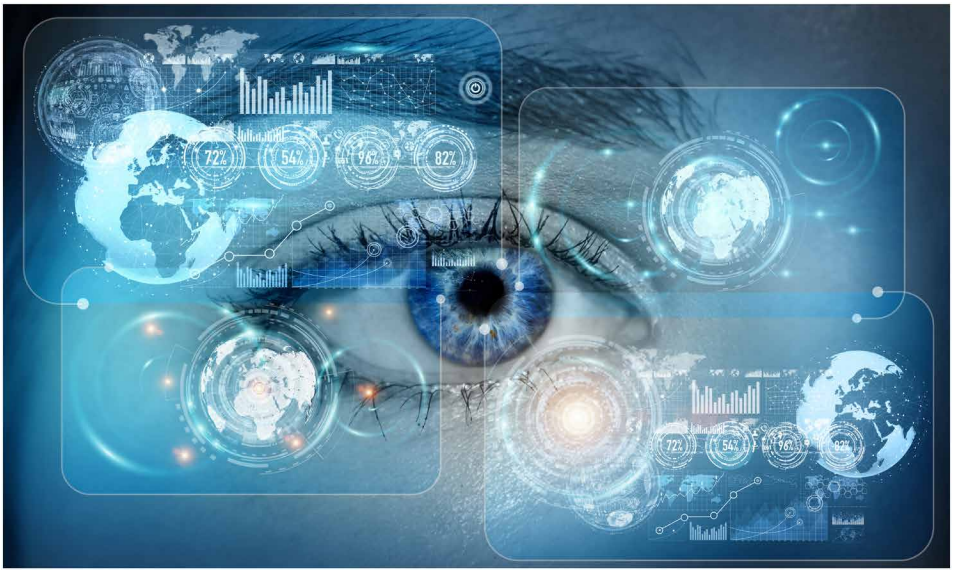
Many of these individual problems are also integral to and promoted by competitions that are directed at both academia and industry. Competitions and contests are important tools to increase the knowledge for open problems in the community while, at the same time, being an opportunity for researchers to evaluate, compare, and promote their solutions.

Related to the above, the availability of high-quality labeled data is essential for enabling and evaluating state-of-the-art academic research. The ImageNet (*3*) challenge, or the Large Scale Visual Recognition Challenge (LSVRC), is an annual contest including individual categories such as object classification, detection, and localization.

ImageNet includes about 14 million images with more than 20,000 image tags. The LSVRC has gained a lot of attention since the 2012 presentation of AlexNet (*4*), which reduced the error rate on images to 15.7%. What followed focused on and succeeded in reducing errors even further. For the purposes of image segmentation and detection, the Common Visual Data Foundation, Microsoft, Facebook, and Mighty AI have jointly released a data set named COCO (*5*), and there are several other examples.

It is interesting to note, though, that the research that followed Minsky's initiative during 1980s and built on the initial ideas of "computer vision systems" was in the area of active vision and active vision systems, envisioning systems that can manipulate the viewpoint of the camera(s) to investigate the environment and to get better information from it (*6*). This was an important area of research because it brought computer vision and robotics closer together.

Unfortunately, during the years that followed, these two communities drifted apart, and many of the computer



CREDIT: SDECORET/SHUTTERSTOCK.COM

vision solutions available today are not directly applicable in robotics. Robotic applications bring some important challenges that are not addressed appropriately or at all in the computer vision community: demonstrating performance in large-scale and unconstrained environments, real-time aspects, robustness, and the ability to recover from failure, to name some.

However, there are problems and methods today that show the potential of bringing the communities closer together again. Image captioning building on natural language processing may be a way of building systems that can communicate the content between images. The use of methodologies such as reinforcement learning and generative adversarial networks (GANs) (*7*) is important for future development in these areas. GANs made the basis for new adversarial models that increase the level of realism of synthesized data and may also provide important insights into the performance of the methods. In addition, adversarial learning techniques may provide a framework for use of unlabeled data to train machine learning models, thus supporting future development of unsupervised learning methods. This is one of the open areas of research, and there is still much to be discovered regarding both theoretical and practical attributes of these methodologies, as well as expanding their applications to address problems with real-world complexity.

A recent article in *Science Robotics* (*8*) identified 10 grand research challenges, which included navigation and exploration in extreme environments as well as fundamental aspects of AI that include perception and action. As stated, these will require innate capabilities to adapt, learn, and recover. We are now accepting expressions of interest for a special issue on computer vision titled "From active perception to deep learning." We seek works that present the latest advances in computer vision in the most relevant subareas, give a historical perspective on the development of some of the subareas, or bridge the gap between related areas such as computer vision and robotics. The special issue may include presentation of new data sets, regular Research Articles, Review or Perspective articles, or short Focus articles that provide insight into the biggest challenges in a specific field.

– Danica Kragic

REFERENCES

1. M. Boden (*9*) cites (*10*) as the original source. There is also a Vision Memo (1966) authored by S. Papert.
2. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (MIT Press, 1982).
3. http://Image-net.org.
4. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proceedings of the 25th International Conference on Neural Information Processing Systems-Volume 1 (NIPS'12)*, F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger, Eds. (Curran Associates Inc., 2012), pp. 1097–1105.
5. http://cocodataset.org.
6. J. Aloimonos, I. Weiss, A. Bandopadhay, Active vision. *Int. J. Comput. Vis.* **1**, 333–356 (1988).
7. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks (2014); arXiv:1406.2661.
8. G.-Z. Yang, J. Bellingham, P. E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, R. Merrifield, B. J. Nelson, B. Scassellati, M. Taddeo, R. Taylor, M. Veloso, Z. L. Wang, R. Wood, The grand challenges of *Science Robotics*. *Sci. Robot.* **3**, eaar7650 (2018).
9. M. Boden, *Mind as Machine: A History of Cognitive Science* (Oxford Univ. Press, 2006).
10. D. Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence* (Basic Books Inc., 1993).

10.1126/scirobotics.aav1778

Citation: D. Kragic, From active perception to deep learning. *Sci. Robot.* **3**, eaav1778 (2018).

Science Robotics

From active perception to deep learning

Danica Kragic

Sci. Robotics **3**, eaav1778.
DOI: 10.1126/scirobotics.aav1778