

The Social Impact of Computer Vision

Vanya Valindria, Imperial College London, v.valindria15@imperial.ac.uk

ICVSS Essay

1 Introduction

The scope of computer vision is as broad as everything that computer can do when it owns its intelligent vision.

There was a story about a man, born with completely functioned eyes. He owned his vision until his 21st, when suddenly an accident took away his vision. As he woke up, his world became dark. He was lost in a void. He was extremely despair, how could he continue his life without his vision? [1]

When God takes away a human's vision ability, he feels weak and lost. But, what happen when humans put the vision ability to a machine? It is the other way around. The machine can become more powerful, as it can help humans with vision impairment, can recognize every faces in our social media, can determine every places in this world [2], or even can teach kids to write [3].

Then, how does the story of a machine with vision ability can assist people? In the 1990s, we could not imagine that a computer can be as close as today in our social life. Commonly, we could only find machine vision in radiology, in industry (to detect crack in manufacture products or to classify the types of fish) and in astronomy [4]. Nowadays, the machine's vision is becoming smarter and smarter. Thanks to the rapid development of computer vision, which is one of the main gates to realize AI (Artificial Intelligence).

2 Ten years of Technological Transfer

In the past decade, researchers started to solve object recognition tasks via Pascal VOC challenge [5]. Globally, they competed to get the best accuracy in detecting and classifying ten objects from thousands of images in the database [5]. In general, most people applied classification methods using hand-crafted features to solve the tasks. The most common feature descriptor used is SIFT (Scale-invariant feature transform) [6], which is known to be powerful and invariant to scale as well as rotation.

In 2010, the journey of Pascal VOC was continued by ILSVRC (ImageNet Large Scale Visual Recognition Challenge) that scales the challenge into more-than-one-million images with thousands objects [7]. However, it was not until 2012, when deep neural networks hit a major breakthrough as the winner in ILSVRC [8]. With the availability of processing a large-scale data (via crowd sourcing annotation and an efficient implementation with GPU), that moment was the turning point that has evolved the visual recognition strategy; from

hand-crafted features into large-scale convolutional neural networks (CNN). Since then, the trends in computer vision tasks are dominated by CNN [7].

As an initiative from universities, ILSVRC shows a positive impact to bring the big players in industries, such as Google, IBM, Adobe enter the competition along with other teams from research intuitions or academics. Contributions on deep-learning and computer vision research from industries, like Google [9] or Microsoft [10], can be seen as a technological transfer to the community. Indeed, with those deep neural networks, image recognition accuracy can be hugely improved.

Furthermore, over the past few years there have been a lot of focus on large-scale recognition in the computer vision community. The term 'big-data' is paired with deep-learning, resulting in a buzz-word for the market worldwide. This big leap in computer vision has triggered industries to apply cutting-edge research into their products.

2.1 From Research to Assistive Technology

In 2006, we had no idea about iPad, smart-phones, or drone camera, which have been widely used like today. With a fast growth of hi-tech gadgets, computer vision is adjusted to follow their needs. For example, the depth cameras that can capture 3D motion tracking and body gesture (like Kinect, leap motion, or Google Tango), together with the wearable VR headset have shaped a new face of VR/AR.

Nowadays, it is common to see VR/AR applications in our everyday life, from children playing Xbox with Kinect, toddlers reading books with AR, industries selling their products based on VR/AR (such as AR.co in Asia [11] or Surreal-Vision in UK [12]), to an 'X-ray vision' assisting medical surgery [13]. VR development leads to a new social sensation, in such as way we can switch from real world to virtual world within seconds, and interact inside it. Hence, computer vision enables VR/AR as a unique selling point for brands and marketing purpose [11]: to enhance the experience of a product.

Not only for fun, research and industries are trying to offer their solutions for assisting people with visual impairment. OrCam MyEye [14] attaches a smart camera to the glasses and integrates computer vision/machine learning technology into it to help visually impaired people to read anything (with text-to-speech), to recognize products, and to identify people.

Despite many tools developed for the blinds, sometimes they are too complicated, and the blind people end up seeking help from nearby. Yet, being able to describe the surrounding can help visually impaired people to navigate. Research project Seeing AI behinds the Pivthead [15] is one of the real products that uses computer vision to describe a person's surroundings, to read text, to answer questions, and even to identify emotions. With all of the benefit of computer vision technology in assisting the vision-impaired, they may no longer depend on their white canes or guided dogs.

Meanwhile, another Pivthead technology for eye-wear [15] also turns the spy-gadgets into reality with computer vision. That eye-wear can capture images, videos, even life-stream our view within a simple gesture touch on the side of the glasses. Hence, this breakthrough makes our vision has no boundary to the world and brings new social issues as well (how do we aware of people

taking our photo/video and broadcast it to the world without showing a visible gesture?).

3 Future Prospects

From this point, we have seen an amazing development of computer vision applied in the real-world within a decade. In the next decade, it seems no wonder that the technology can be far beyond from what can we imagine today.

One interesting case is the Pivthead [15] or Google-glass like product that may extent in the future into an optical implant. Recently this year, Samsung, SONY, and Google [16] are putting an effort to implant a camera inside people's eyes, so that people can record and capture images with just one blink. Consequently, the social issues of privacy can become harder to control than the Pivthead [15] case. Although it sounds scary, but when it is real, this jargon wins it: *Your vision is our vision*.

On the other hand, the realm of AI will also be more immerse. As stated by Facebook AI [17] that their goal is to make AI systems that are better than humans at our primary senses: vision, listening, etc. Within the next 10 years, Facebook hopes to deliver systems that can recognise everything and understand the context of an image/video [17]. If that is what their wish, then this is very close to be real and can happen sooner than expected.

In addition, with the advancement of deep neural networks in computer vision we have today, added with deep compression [18], it is not impossible to realize a smart bionic eye in the future. Deep compression technique [18] may open an effort to deploy deep neural networks in a smaller hi-tech-gadget (mobile/wearable/watches). Perhaps in the future, together with optical implants [16], this technology may lead to a smart bionic eye, which can assist the blinds to recognise everything, to read anything, to understand emotions/facial gestures, to capture every precious moment, or even to go further, such as fast object counter or ability to zoom in/zoom out their vision.

4 Conclusion

To summarize, computer vision has brought social impacts in the real world, as we find it in a wide range of applications, from children gaming, security, AR/VR products, retail experience, assistive tools, to medical imaging. This phenomenon shows that our senses are getting integrated to AI in everyday life.

Humans are becoming more comfortable to share their experiences with the machines. Nowadays, people tends to feel more secure to see themselves being watched in real-time by smart home monitoring system (with computer vision technology to detect unfamiliar presence) [19] inside their house. In the context of professionalism, with the availability of a large-scale medical imaging and clinical data, computer vision can also increase the doctor's confidence to deliver diagnosis [20].

Thus, computer vision makes humans and machines share the same sense: vision. However, we should argue that the social impact in the real world is still evolving as the AI improves and the machine gets smarter. What we are sure

is that, our expectation from computer vision research is always the same: to make the world a better place to live.

References

- [1] Disability Youth Center Indonesia Article. URL <http://tv.liputan6.com/read/2370742/pantang-menyerah-sikdam-aktivis-tunanetra-peduli-disabilitas>, May 2016.
- [2] T.Weyand, I.Kostrikov, J.Philbin. PlaNet - Photo Geolocation with Convolutional Neural Networks, arXiv preprint 1602.05314, Feb 2016.
- [3] CoWriter - Learning to write with a robot, URL <http://chili.epfl.ch/cowriter/>, June 2016.
- [4] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, USA, 2006.
- [5] M. Everingham, L. Gool, A. Zisserman. *The Pascal Visual Object Classes (VOC) Challenge*. International Journal of Computer Vision, Vol.88, no.2, pp.303, 2010.
- [6] D.G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, vol.60, no.2, pp. 91-110, 2004.
- [7] O. Russakovsky, A. Karpathy, Li Fei-Fei, et.al. *ImageNet Large Scale Visual Recognition Challenge*. International Journal of Computer Vision, vol.115, no.3, pp. 211-252, 2015.
- [8] A. Krizhevsky, I. Sutskever, G. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. Advances In Neural Information Processing Systems, pp. 1-9, 2012.
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. *Going deeper with convolutions*. In CVPR, 2015.
- [10] K.He, X.Zhang, J.Sun. *Deep Residual Learning for Image Recognition*. arXiv preprint 1512.03385, 2015.
- [11] ARco Asia, URL <http://www.ar-innovation.com/en/home/>, accessed in June 2016.
- [12] Surreal Vision UK, URL <http://surreal.vision/>, accessed in June 2016.
- [13] Microsoft X-ray vision, URL <http://research.microsoft.com/en-us/news/features/touchlessurgery-060712.aspx>, 2012 (accessed in June 2016).
- [14] Orcam MyEye, URL <http://www.orcam.com/>, accessed in June 2016.
- [15] SeeingAI in PivotHead, URL <http://www.pivthead.com/build/>, accessed in June 2016.

- [16] A.Conrad. *Intra-ocular device*. United States Patent 20160113760, issued April 28, 2016
- [17] FB Telepaty tech, URL <https://www.theguardian.com/technology/2015/jul/01/facebook-mark-zuckerberg-telepathy-tech>, accessed in June 2016.
- [18] S.Han, H.Mao, B.Dally. *Deep Compression: Compressing Deep Neural Network with Pruning, Trained Quantization and Huffman Coding*. Deep Learning Symposium, NIPS 2015.
- [19] Smart Home Monitoring NETATMO, URL <https://www.netatmo.com/en-US/product/camera>, accessed in June 2016.
- [20] Zebra Medical Vision, URL <https://www.zebra-med.com/>, accessed in June 2016.