Institution: Oxford Brookes University

Unit of Assessment: 11, Computer Science and Informatics

Title of case study: Artificial Intelligence for saliency detection and action recognition in mobile phone cameras and performance analytics in sport

Period when the underpinning research was undertaken: 2011–2020

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Fabio Cuzzolin	Professor of Artificial Intelligence, Director of Visual AI Lab	[text removed for publication]
Nigel Crook	Professor, Associate Dean Research & Knowledge Exchange	
Philip Torr	Professor	
Ming-Ming Cheng	Research Fellow	

Period when the claimed impact occurred: Application of saliency detection (2018–2020), application of action recognition (2019–2020)

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact

Seminal work carried out at Oxford Brookes University, to create a global cue video processing algorithm that identifies the most significant regions of an image, has laid the foundation for development of a significant product upgrade for a major phone handset manufacturer. Huawei has deployed this technology as a 'selfie' app that is built into tens of millions of Huawei smartphones sold globally. Oxford Brookes University's technical expertise on computer vision and action recognition has also been instrumental in producing the algorithms that have been selected for use in other popular Huawei smart devices. In addition, the pioneering technology has enabled Sportlight, a sports analytics start-up, to create state of the art intelligence tools for professional sports teams, including English Premier League football clubs.

2. Underpinning research

The Computer Vision Research Group at Oxford Brookes University, which was founded in 2005 by Professor Philip Torr, has made many outstanding contributions to the field, especially in scene understanding within images. One of these contributions arose from an EPSRC funded project 'Scene Understanding Using New Global Energy Models' (EP/I001107/1, GBP439,228, 2011-2013) in which the Global Cue (GC) method was developed for automatically detecting which areas of an image are foreground and which are background. The foreground is often the most visually salient region of an image, representing the person or object that is the primary subject of a photograph.

This work was carried out by Ming-Ming Cheng. During his 17 months at Oxford Brookes University, Cheng, working with Torr and Crook, substantially revised his original GC

method, which had been developed during his doctoral research at Tsinghua University, China. This improved its efficiency so that it could be applied to scene understanding in video sequences. The GC method collects samples from across the image based on their appearance, grouping similar samples together that are close to each other in the image. This approach avoids the erroneous hard boundaries often created with the super pixel approach **[R1]**.

This implementation of the GC method was compared experimentally with 18 alternative algorithms on the largest publicly available dataset of images (1000 images) with pixel accurate 'ground truth' annotations **[R2]**. The experiment showed that the GC method's salient object region detection results were 25% better than the previous best results produced by alternative methods **[R1]**. Fig 1 below shows a visual comparison of GC with some of the alternative methods. The results also demonstrate that the refined GC method out-performs the alternative algorithms in terms of processing speed.

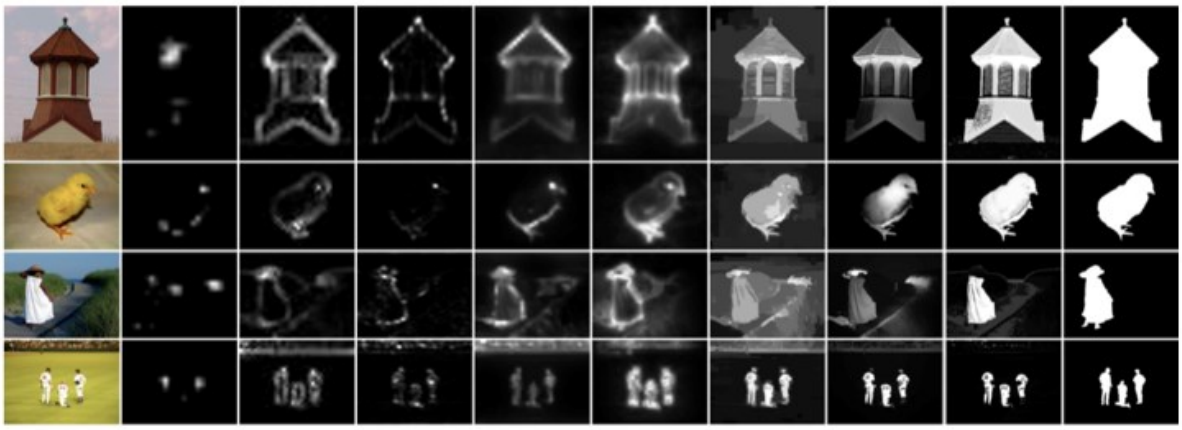
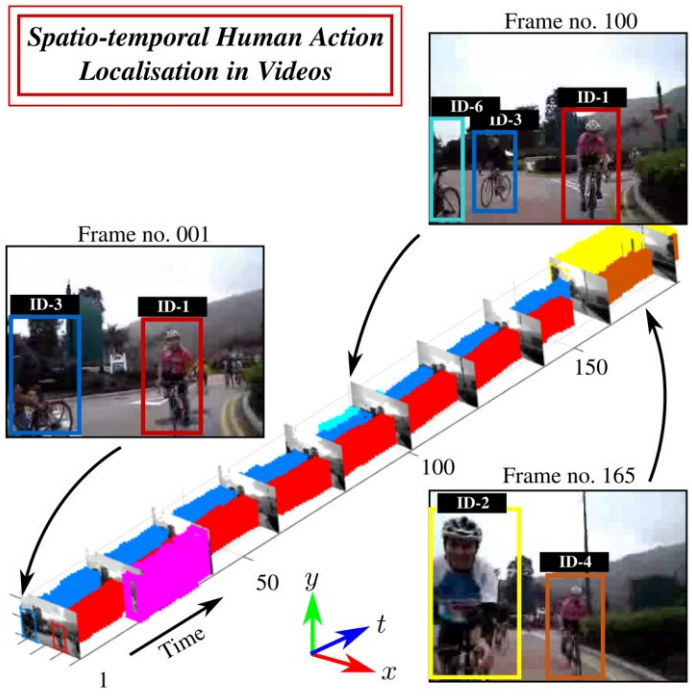


Fig 1: Visual comparison of previous approaches (b) – (h) with the GC algorithm (i). (a) original image. (j) ground truth

In 2012, the Computer Vision Group was re-named the Visual Artificial Intelligence Laboratory and, under Professor Fabio Cuzzolin, the group has achieved a prominent position in the field of human action recognition from videos. This work received funding from an EPSRC First Grant on 'Tensorial modeling of dynamical systems for gait and activity recognition' (EP/I018719/1, GBP98,364, 2011-2014). Work carried out under this funding included a method for detecting actions in videos by representing them as graphs of parts **[R3]**. In 2016 the Visual AI Lab was among the first in the world to employ deep neural networks for detecting actions performed by people in videos, represented as 'action tubes' as illustrated below, where, frame by frame the actions (in this case, cycling) are identified and tracked over time.

The approach exploits an original dynamic programming algorithm which form 'tubes' linking rectangular areas of a continuous series of frames together to form 'action tubes' in real time and which can anticipate future events **[R4]**.



Visualising predicted action tubes in a "Biking" video

In 2017, the team was the first to propose a deep-learning-based approach for detecting multiple events in real time in a streaming video **[R5]**. The method was further developed into the first causal 3D CNN architecture (Recurrent Convolutional Neural Network), achieving state-of-the-art performance in action recognition **[R6]**. The accuracy of this detection technique was recognised with 2nd place awards at in the action detection challenges at both the Computer Vision and Pattern Recognition (CVPR) Charades (http://vuchallenge.org/charades.html) and ActivityNet (http://activity-net.org/challenges/2016/program.html#leaderboard). These activities challenge computer vision groups from around the world to compete on a task.

3. References to the research

[R1] Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H. and Hu, S.M. (2014) 'Global contrast based salient region detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3), pp.569-582. DOI: 10.1109/TPAMI.2014.2345401

Over 3,000 citations.

[R2] Cheng, M.M., Warrell, J., Lin, W.Y., Zheng, S., Vineet, V. and Crook, N. (2013) 'Efficient salient region detection with soft image abstraction', *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2013),* pp.1529-1536 DOI: 10.1109/ICCV.2013.193

Over 500 citations.

[R3] Sapienza, M., Cuzzolin, F. and Torr, P. (2014) 'Learning discriminative space-time action parts from weakly labelled videos', *International Journal of Computer Vision* 110(1): pp.30-47. DOI: 10.1007/s11263-013-0662-8

[R4] Saha, S., Singh, G., Sapienza, M., Torr, P. and Cuzzolin, F. (2016) 'Deep learning for detecting multiple space-time action tubes in videos', *Proceedings of British Machine Vision Conference (BMVC 2016),* 58, pp.1-58.13 DOI: 10.5244/C.30.58

This paper was cited 120 times in the period Sep 2016 – May 2020.

[R5] Singh, G., Saha, S., Sapienza, M., Torr, P. and Cuzzolin, F. (2017) 'Online real-time multiple spatiotemporal action localisation and prediction', *Proceedings of the IEEE*

*International Conference on Computer Vision (ICCV 2017), pp.*3657-3666.
DOI: 10.1109/ICCV.2017.393

This paper was cited 143 times in the period October 2017–May 2020.

[R6] Singh, G., and Cuzzolin, F. (2019) 'Recurrent Convolutions for Causal 3D CNNs', *IEEE International Conference on Computer Vision Workshop - Holistic Video Understanding (ICCVW 2019),*pp.1456-1465. DOI: 10.1109/ICCVW.2019.00183

4. Details of the impact

The world leading research on artificial intelligence for computer vision and action recognition at Oxford Brookes University has led to innovative new products for both a major existing international company and a start-up company.

Artificial Intelligence for saliency detection and action recognition in Huawei mobile phone cameras

The research that Cheng completed whilst at Oxford Brookes University resulted in clear evidence that the GC method for saliency detection was capable of producing state-of-the-art image analysis, with substantial commercial potential. It laid the foundation for the development of a saliency detection algorithm that has now been deployed in tens of millions of Huawei mobile phones, from the Mate 10 model onwards.

In 2014 Cheng returned to China to take up an academic position at Nankai University, where he continued to work on the GC method, developing a 'deep-learning' implementation of it, in which the global and local cues were learned rather than hand crafted. This implementation was published in a top-ranking international conference (*IEEE CVPR 2017*). Huawei recognised that this 'deep-learning' implementation of the GC method had clear application potential in their product line and Nankai University and Huawei jointly applied for a national invention patent for this technology **[S1]**. In 2017 Cheng and his research group won an award for their outstanding collaboration with Huawei **[S2]**.

Huawei used the technology to develop a selfie app which featured in their flagship mobile phones (Mate 10 series onwards), and which provided support for the smart camera effect of Huawei products. This effect is used to support auto-adjustment modes for 'selfies', to automatically improve portrait photographs. The smart camera then adjusts the contrast settings to optimise the sharpness of the image to produce a high quality self-portrait. This is a significant selling point for Huawei handsets because the ability to take high quality self-portrait images is a highly desirable feature for many phone users. This saliency feature was presented in an online demonstration at the MATE 10 product launch show in March 2018 **[S3]** and described in the marketing brochure for the Y6 phone in 2018 **[S4].** The AI features in this phone received favourable reviews **[S5]**. It was reported that 17,000,000 Mate 10 phones, which included this technology, were sold in the first year after the launch in October 2017 **[S6]**. The technology was also included in the Mate 20 phone which sold 10,000,000 units in the first year after its launch in October 2018 **[S7]**.

Cuzzolin is a consultant at Huawei's Vancouver IC Lab, which occupies 20% of his time. Through this consultancy, the Visual AI Lab's research on deep learning for human behaviour analysis has informed numerous projects. Details are commercially sensitive, but a number of computer vision algorithms developed there, under Cuzzolin's overall technical consultancy, have been selected to be installed in high-end smartphones, such as the Mate 20 Pro and the P20 Pro, based on Huawei's Kirin 970 and 980 chipsets. Vancouver IC Lab's

Director confirms that there are three projects "adopted, or in process to be integrated into the products" **[S8i]**. He also recognises that Cuzzolin's technical expertise on computer vision and action recognition "through the consultancy has had a significant impact, and [has] been instrumental in producing the algorithms that can be adopted into smart devices" **[S8ii]**. Cuzzolin has also had impact on the researchers in the Huawei team. The Director states, "Your comments and feedback for their project help them finish their projects, as well the growth along their career path" **[S8ii]**.

Artificial Intelligence for action recognition applied to performance analysis and sports analytics.

A Knowledge Transfer Partnership (KTP) with Createc Technologies (GBP190,000, 2019-2021) was launched in April 2019 to use action recognition technologies for monitoring players' physical movements on the football pitch. CEO of Sportlight (a business within Createc, which uses new technology to provide advanced intelligence for sports teams), states that the action recognition technology has "helped set Sportlight's products apart from current legacy ones (based on GPS or optical tracking)". He confirms that, "at least 3 recent Premier League connections were facilitated by our KTP work" and funding for development has been received from one premier league club **[S9i]**. The novelty provided by this technology has also helped to secure a new round of funding for Sportlight. The CEO states: "While the core product features are important for our key contacts (sports scientists and performance experts at elite football clubs), the potential 'new' insights, inspired by Fabio Cuzolin's research work on action recognition, allow us to work with more departments within clubs (physio/medical team, tactical analysts, coaches…) and also with more senior strategic decision makers (Director of Insights, Head of Research, Head of Analytics)." **[S9ii]**. The ability to analyse the cause of injuries in addition to tracking player performance is of considerable interest to football clubs.

In summary, the new technologies developed at Oxford Brookes University have contributed to the introduction of Huawei's world leading series of Mate 10 and 20 mobile phones, which has underpinned Huawei's rise to become the world's number one phone manufacturer since April 2020. Oxford Brookes University's technologies have also enabled Createc to expand its start-up business and excite the interest of top football clubs.

5. Sources to corroborate the impact

[S1] Huawei Application Certificate to Nankai University, confirming the application of Cheng's technologies to Huawei products, and joint application for three patents.

[S2] Huawei Award to Cheng's group at Nankai University for Outstanding Collaboration, 2017

[S3] Evidence of salient object detection used in Huawei Mate 10 mobile phone

[S4] Huawei marketing materials: '2018 Smartphone, best Selfie Toning Flash'

[S5] Huawei Mate 10 review 'Amazingly Smart Camera, Epic Battery Life', Tom's Guide, May 2018

[S6] Press report on Huawei Mate 10 sales figures, newzoo.com, October 2018

[S7] Press report on Huawei Mate 20 sales figures, from GSMARENA, 11 March 2019

[S8] Two email statements: Director, Vancouver IC Lab, (i) December 2019 and (ii) September 2020

[S9] Two email statements: Director, Sportlight, December 2019 and September 2020