Sensors and Materials, Vol. 34, No. 1 (2022) 47–54
MYU Tokyo

S & M 2792

Detection of Instruments Inserted into Eye in Cataract Surgery Using Single-shot Multibox Detector

Maina Sogabe,[1] Norihiko Ito,[2] Tetsuro Miyazaki,[1]
Toshihiro Kawase,[3,4] Takahiro Kanno,[5] and Kenji Kawashima[1*]

[1]Department of Information Physics and Computing, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[2]Department of Agriculture, Tottori University, 4-101 Koyama-cho Minami, Tottori 680-8550, Japan
[3]Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University
1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
[4]Institute of Innovative Research, Tokyo Institute of Technology,
4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8503, Japan
[5]Riverfield, Inc., Yotsuya Medical Bldg. 5th floor, 20 Samon-cho, Shinjuku-ku, Tokyo 160-0017, Japan

(Received December 1, 2021; accepted December 21, 2021)

Keywords: ophthalmic surgery, image processing, single-shot multibox detector, object detection, cataract surgery

Estimating the position of a surgical instrument and identifying its type are important first steps in developing surgical assistance and automation technology. However, there are few reports on object detection technology using a surgical microscope in the field of ophthalmology. Some of the major challenges in position estimation and tool recognition under microsurgery in this field are that the target area is narrow, the image may be distorted through refraction at the air–liquid interface, and the angle of view is often enlarged or reduced as needed by the surgeon. To address these challenges, we applied a single-shot multibox detector (SSD) technique to determine the position and type of an instrument during cataract surgery. SSD is an object detection technique with superior accuracy and processing speed to existing deep-learning-based detection methods. Using this method, we detected two major surgical tools in images during cataract surgery and obtained a mean average precision of 0.75. Our results show that it is possible to recognize instruments inserted in the eye and estimate their positions.

1. Introduction

Cataract accounts for a large proportion of the causes of premature blindness and is a serious problem worldwide.(1) Since cataract is directly linked to the deterioration of visual acuity, surgical treatment is an important technique that contributes to the improvement of patients' quality of life. The instruments and devices used in ophthalmic surgery are well established and the surgical procedure has been standardized.(2) However, current practice necessitates a human surgeon to control complicated equipment. In addition, the instruments and equipment are expensive, and it takes considerable time for a surgeon to acquire the necessary skills, making it difficult to provide high-quality and uniform surgery without limitations.

*Corresponding author: e-mail: kkawa729@g.ecc.u-tokyo.ac.jp
https://doi.org/10.18494/SAM3762

ISSN 0914-4935
https://myukk.org/

To improve the situation, there are great expectations for robot-assisted or autonomous surgery. Even on remote islands and in rural areas where an expert surgeon may not reside, high-quality ophthalmic medical care can be provided if we create an environment where a surgery robot is installed and a local human technician is employed to assist the robot.

Unlike thoraco-laparoscopic surgery, where robot-assisted surgery has already been applied, the introduction of robots to ophthalmic surgery should be rather straightforward because the instruments and devices are well established and the surgical procedure is standardized; the base platform may be relatively easily replaced by robot-assisted surgery. Cataract surgery with assistance using the da Vinci robot has already been realized.[3,4] Although cataract surgery requires a delicate operation that is extremely difficult for humans to learn, the realization of semi-autonomous robots can be expected if we take advantage of the remarkable advancements of precision machinery, control technology, and image recognition technology based on deep learning, and fuse them together. The first task will be to develop a technique that is capable of discriminating a wide variety of similar instruments used in ophthalmic intraocular surgery to recognize the position of their tips. As an enabling technology, object detection in the image processing field is the most promising.

Object detection is a technology that captures an image and detects the position and category (class) of a detected object from the image. Among the various object detection methods, deep learning is a powerful machine learning technique that automatically learns the features of an image required for a detection task and is already used in the surgical field for discriminating tools. Various methods have been developed for object detection, where Region-based Convolutional Neural Networks (R-CNN)[5] and its associated methods are the most well known. R-CNN uses an object proposal algorithm called a selective search to detect candidate regions at the first stage, and then applies an image of the candidate regions to a CNN to extract features. The obtained features are then used for classification. Since the process requires two steps to acquire the area from which the features are extracted and the application of the extracted features to multiple SVMs, it has a disadvantage of a slow execution time. Later, Fast R-CNN[6] and Faster R-CNN[7] were developed, which had higher execution speeds. However, the improvement of the processing speed was limited because the algorithms employ a sequential processing configuration that moves the bounding box around the image to find a good place for object detection and perform identification. Meanwhile, regression-based object detection techniques such as You Only Look Once (YOLO)[8–10] and the single-shot multibox detector (SSD)[11,12] have been developed, which incorporate image recognition into regression problems and realize detection and identification at the same time. These methods are widely used for the detection of surgical instruments.[13,14]

On the other hand, there are few reports on object detection in the field of ophthalmology. The need for the detection of surgical instruments outside the eye has been reported only recently,[15] while research on the detection of surgical instruments inside the eye, which is the area where actual surgery is performed, is still at an early stage. In intraocular surgery, the surgical field is very narrow and multiple instruments may be positioned close together in the surgical field, with some of them overlapping. Moreover, images tend to be distorted owing to the liquid–air interface (outside air, cornea of the plaque, anterior chamber, lens layer). Also,

sometimes a surgical instrument may be hidden by a crushed crystalline lens. These factors make object detection in ophthalmology very challenging.

In this study, we tackle these challenges by using the SSD to detect an object inserted in the eyeball of a dog from an image obtained during actual cataract surgery. During cataract surgery, unlike the object detection problem for cars and road signs, it is assumed that only a small number of instrument tips are present in the eyeball. Thus, the SSD is employed as an object detection method, which yields relatively few false positives, instead of YOLO, which can reliably detect true positives.[16] The SSD also has an advantage of being able to achieve both accuracy and detection speed. For input images, the SSD generates network predictions from multiple layers of feature extraction networks of different sizes to collect and decode the prediction results, and creates a bounding box. Since multiple feature extraction network layers are used, the SSD has the advantage of accurate detection even for a relatively low resolution image.

We demonstrate that the SSD is capable of object recognition in microsurgery and simultaneously detecting two surgical instruments of different sizes. Specifically, learning is performed for 702 surgical images taken during the divide and conquer process in veterinary cataract surgery, and verification is performed for a total of 300 images obtained from three different surgical scenes involving different breeds of dog.

2. Materials and Methods

2.1 Cataract surgery

Cataract surgery consists of four major steps. First, a circular slit is made in the front of the lens capsule under a microscope. Next, an instrument is inserted through the slit in the lens capsule and ultrasound is used to break up the cloudy lens. In the third step, the lens is further shredded and removed by irrigation and aspiration. Finally, an intraocular lens is inserted into the empty lens capsule and fixed in place, completing the surgery. The process of crushing and suctioning out the lens is called divide and conquer. During this process, an ultrasonic handpiece and a hook to hold the lens in place are used (Fig. 1). This divide and conquer step, in which multiple instruments are used in cataract surgery, was chosen as the target scene for this study. The actual size of the surgical field is roughly 15–20 mm. A normal lens is about 10 mm in diameter and 5 mm thick, shaped like a convex lens, and almost transparent.

The surgical instruments in Fig. 2 are an ultrasonic handpiece [Fig. 2(a), Whitestar Signature ELLIPS Handpiece, Johnson & Johnson Surgical Vision, Inc., CA, United States] designed to facilitate lens extraction during cataract surgery and an M-hook [Fig. 2(b), Inami & Co., Ltd., Aichi, Japan] to capture and remove the lens. The ultrasonic handpiece uses longitudinal vibration to emulsify the cataract and aspirate the debris with a tip.

The microscope used for the surgery was a Leica M844 F40 series, with 1920 × 1080 RGB images acquired at 29 frames per second (fps). The images were cropped in advance to 300 × 300 pixel RGB images to maintain the input's aspect ratio. After that, each image was annotated with bounding boxes including an index representing the tip of each instrument together with the instrument type (Fig. 3).

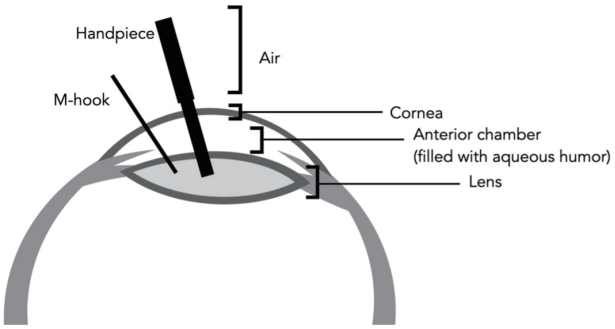
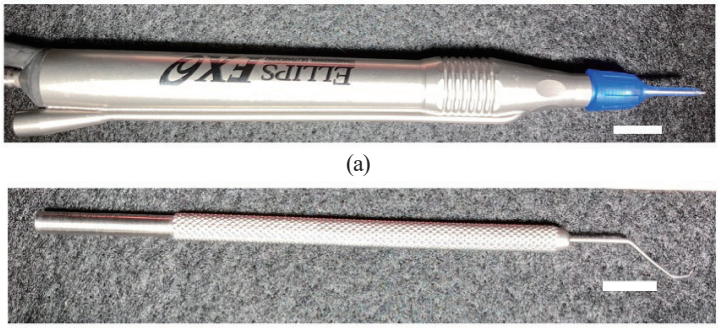


Fig. 1. Overview of the eye and areas approached in cataract surgery.



(a)

(b)

Fig. 2. (Color online) Surgical instruments. (a) Ultrasonic handpiece and (b) M-hook (scale bar = 1 cm).

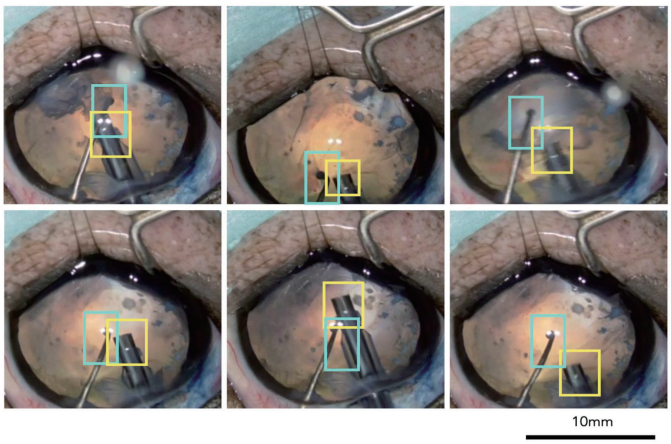


Fig. 3. (Color online) Illustrations of image annotation by drawing bonding boxes around tips of M-hook (cyan) and handpiece (yellow). The scale bar indicates 10 mm.

2.2 SSD architecture and data acquisition

Figure 4 shows the SSD architecture. For the SSD base network, ResNet50,[17,18] which had been pre-learned by ImageNet consisting of over 1.4 million images, was used. The detailed network using ResNet50 was based on the method of Chen[19] Fine tuning was performed using

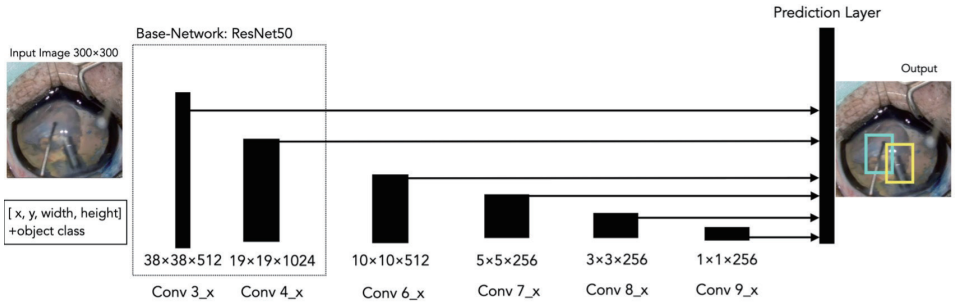


Fig. 4. (Color online) SSD architecture.

the surgical images (conquer phase). The eye used for the training data was that of a toy poodle. Data augmentation such as contrast change, horizontal translation, and random scaling to match the characteristics of the cataract surgery was applied to the images used for the training data.

The input size of the image was 300 × 300 × 3 and maximum 300 epochs of training were performed. The mini-batch size was 16 and the initial learning rate was 0.1. Adam[20] was used to optimize the hyperparameters. K-fold cross validation (k = 5) was used to detect and prevent over-fitting, and after 5 times training, the optimal model weights were defined. The machine used for training was an Intel® Xeon® Gold 6252 2.1 GHz CPU with an Nvidia GTX2080 Ti GPU. The calculations were implemented on MATLAB 2021a (MathWorks, MA, USA) and the learning time was 5 h. The threshold value was set as 0.25. Basically, there is no situation in which multiple instruments of the same type are inserted into the eye. Therefore, when multiple instruments of the same type exceeding the threshold were detected, the bounding box with the highest score was selected. The intersection over union (IoU) threshold was set to 0.25 for the accuracy evaluation.

Furthermore, we verified that the pre-learning of the detector can be applied to the images obtained in surgical series (Scenes 1–3) with different dog breeds. The dog breed of the eye used for the test data was the miniature dachshund. One hundred images randomly extracted from each video series (a total of 300 images) were used as the input to the pre-learned classifier for the evaluation.

3. Object Detection during Cataract Surgery

Figure 5 shows the results of detecting surgical instruments from the different surgical images using the classifier trained by the 702 image data sets. Table 1 shows the class and its average precision in each scene and Fig. 6 shows the precision–recall curves for Scene 2 as typical results of annotations in different surgical scenes. As shown in Fig. 5, we were able to detect the tips of two types of instrument even in the situation where the handpiece and M-hook overlapped and one of the instruments was hidden (Fig. 5, Scene 3), which was initially considered problematic. We were also able to detect the tip of the instrument even when the tip was hidden by a cloudy lens fragment (Fig. 5, Scene 2). The average precisions (APs) of handpiece and M-hook detection were 88.9 and 62.2%, respectively, and the mean average

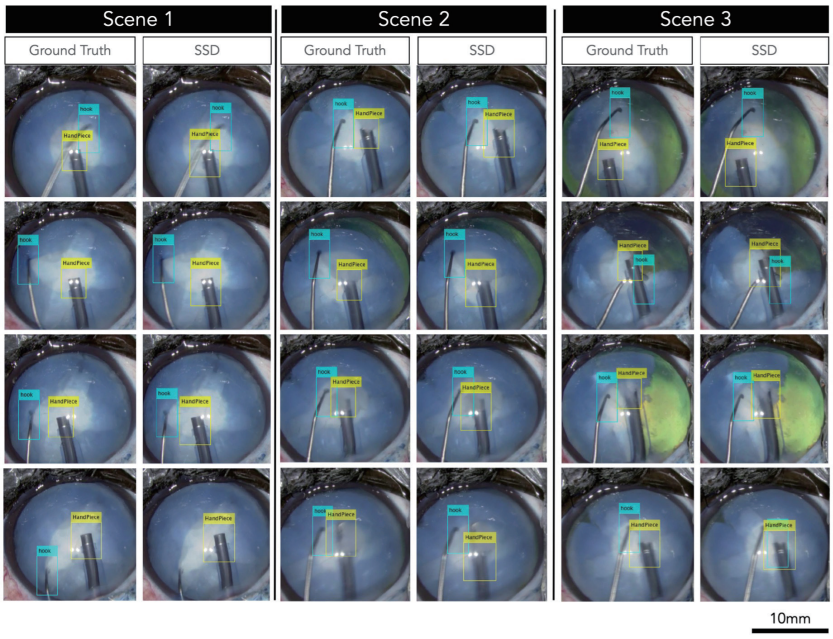


Fig. 5. (Color online) Results of surgical instrument detection. The scale bar indicates 10 mm.

Table 1
Result of AP in each scene.

	Average precision (%)	
	Handpiece	Hook
Scene 1 (divide)	94.0	46.0
Scene 2 (divide and conquer)	83.0	73.5
Scene 3 (conquer)	89.6	67.1
All scenes	88.9	62.2
Mean average precision (mAP)	75.5	

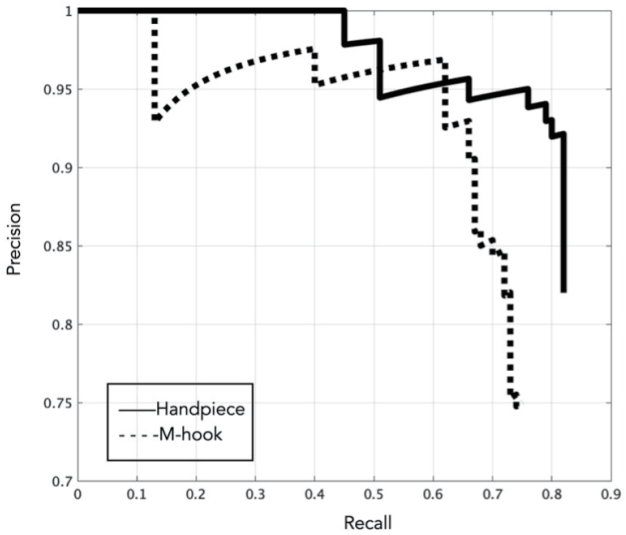


Fig. 6. Precision–recall curves in Scene 2.

precision (mAP) was 75.5%. The average object detection speed was 26.3 fps. There is still no clear detection accuracy benchmark for surgical instruments placed in the anterior chamber in cataract surgery. The detection accuracy of ophthalmic surgical instruments by SSD appears to be comparable to the detection accuracy of surgical instruments in conventional laparoscopic surgery images. [21]

In this study, the eyes used for learning were from a toy poodle, and the data used for validation were derived from a dachshund. Even in this case, the SSD was able to detect the object, suggesting that surgical instruments can be detected even when the training data set, the shape of the eyeball, and the color of the iris change. When applying the object detection technology to humans, it is necessary to train it using human cataract surgery images because the shape of the eye and its surroundings are different. The results suggests that SSD can be used to detect eye shape and iris color differences among individuals.

There were several patterns of situations where the detection failed. The first was when the area to be detected was at the edge of the field of view (Fig. 5, Scene 1, 4th row), in which case the detection accuracy was reduced. For another, if the tip of the surgical instrument was completely hidden by lens debris, the detection position could be drastically misaligned (Fig. 5, Scene 2, 4th row). In this case, since the tip of the correct instrument was buried in the lens debris, the visible area of the surgical tool was detected as the tip of the instrument. Also, blurred images (Fig. 5, Scene 3, 4th row) or image distortion caused by reflected light reduced the recognition accuracy of the M-hook, especially for thin shapes.

In this study, we tested whether we could detect two types of surgical instruments using the SSD, which is capable of object detection even with a relatively small number of data, for cataract surgery images that are affected by the liquid–air interface and the refraction caused by the lens itself.

In the future, we plan to increase the number of data sets and the variety of surgical scenes, eyes, and surgical instruments used for training. Along with the improvement of the data sets, we are considering the use of methods with higher detection accuracy, such as YOLOv4[10] and M2Det.[22]

4. Conclusion

In this study, by using the SSD technique, we successfully estimated the type and position of surgical instruments inserted in the anterior chamber of the eye from distorted images. The mAP was 75.5%, and the AP was 88.9% for an instrument with uniform morphology such as a handpiece. However, devices with distorted tips, such as M-hooks, are not always recognizable because the shape reflected by the camera varies greatly with the angle of inclination. In addition, they cannot be recognized when they are hidden in the shadows of scattered lenses or are affected by motion blur.

As a future improvement, time-series information from several previous frames, rather than a single-shot image, may be employed for more robust estimation even when the instrument is hidden by other tissues. Comparison with other methods and the improvement of accuracy are future issues.

The results of this paper are expected to serve as a basis for the development of semi-automated systems in cataract surgery, such as the automatic suction of lens fragments by the handpiece. Finally, these detection technologies will lead to the automation of cataract surgery. Also, the surgical instrument's detection is helpful in scoring the proficiency of ophthalmologists based on the timing of instrument insertion and the way it is moved.

Acknowledgments

This work was supported in part by MEXT/JSPS KAKENHI Grant Number 21K18074.

References

1 D. Lam, S. K. Rao, V. Ratra, Y. Liu, P. Mitchell, J. King, M. J. Tassignon, J. Jonas, C. P. Pang, and D. F. Chang: Nat. Rev. Dis. Primers **1** (2015) 15014. https://doi.org/10.1038/nrdp.2015.14
2 R T. J. Hassani, O. Sandali, A. Ouadfel, M. Packer, F. Romano, G. Thuret, P. Gain, M. D. de Smet, and C. Baudouin: J. Fr. Ophtalmol **43** (2020) 929. https://doi.org/10.1016/j.jfo.2020.05.006 [Article in French]
3 T. Bourcier, J. Chammas, P. H. Becmeur, A. Sauer, D. Gaucher, P. Liverneaux, J. Marescaux, and D. Mutter: J. Cataract Refract Surg. **43** (2017) 552. https://doi.org/10.1016/j.jcrs.2017.02.020
4 T. Bourcier, J. Chammas, D. Gaucher, P. Liverneaux, J. Marescaux, C. Speeg-Schatz, D. Mutter, and A. Sauer: Translational Vision Sci. Technol. **8** (2019) 26. https://doi.org/10.1167/tvst.8.3.26
5 R. Girshick, J. Donahue, T. Darrell, and J. Malik: Proc. IEEE Int. Conf. Computer Vision (IEEE, 2013). https://doi.org/10.1109/CVPR.2014.81.
6 R. Girshick: Proc. IEEE Int. Conf. Computer Vision (IEEE, 2015) 1440. https://doi.org/10.1109/ICCV.2015.169
7 S. Ren, K. He, R. Girshick, and J. Sun: Advances in Neural Information Processing Systems (2015) 91. https://doi.org/10.1109/TPAMI.2016.2577031
8 J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: Proc. 2016 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (IEEE, 2016) 779–788. https://doi.org/10.1109/CVPR.2016.91
9 J. Redmon and A. Farhadi: YOLOv3: An Incremental Improvement (2018). http://arxiv.org/abs/1804.02767
10 A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao: YOLOv4: Optimal Speed and Accuracy of Object Detection (2020). http://arxiv.org/abs/2004.10934
11 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, and A. C. Berg: Eur. Conf. Computer Vision (2016) 21. https://doi.org/10.1007/978-3-319-46448-0_2
12 G. Singh, S. Saha, M. Sapienza, P. H. S. Torr, and F. Cuzzolin: IEEE Int. Conf. Computer Vision (IEEE, 2017) 3637–3646. https://doi.org/10.1109/ICCV.2017.393
13 K. S. Peng, M. Hong, J. Rozenblit, and A. J. Hamilton: 2019 Spring Simulation Conf. (2019) 1–12, https://doi.org/10.23919/SpringSim.2019.8732863
14 C. Yang, Z. Zhao, and S. Hu: Comput. Assist. Surg. **25** (2020) 15. https://doi.org/10.1080/24699322.2020.1801842.
15 W. Xu, R. Liu, W. Zhang, Z. Chao, and F. Jia: IEEE 13th Int. Conf. Computer Research and Development (IEEE, 2021) 11–15. https://doi.org/10.1109/ICCRD51685.2021.9386349
16 Á. Morera, Á. Sánchez, A. B. Moreno, Á. D. Sappa, and J. F. Vélez: Sensors **20** (2020) 4587. https://doi.org/10.3390/s20164587
17 K. He, X. Zhang, S. Ren, and J. Sun: Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2016) 770–778. https://doi.org/10.1109/CVPR.2016.90
18 J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy: Proc. IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2017). https://arxiv.org/abs/1611.10012
19 G. Chen: E3S Web of Conf. (2021). https://doi.org/261.01011.10.1051/e3sconf/202126101011
20 D. Kingma and B. Jimmy: Adam – A Method for Stochastic Optimization (2014). https://arxiv.org/abs/1412.6980.
21 K. Jo, Y. Choi, J. Choi, and J. W. Chung: Appl. Sci. **9** (2019) 2865. https://doi.org/10.3390/app9142865
22 Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling: Proc. AAAI Conf. Artificial Intelligence (AAAI, 2019) 9259–9266. https://doi.org/10.1609/aaai.v33i01.33019259