

Teaching Tool for Fun Learning of AI-based Banknote Detection Technology

Cheng-Yu Yeh,^{*} Chun-Cheng Lin, and Kuan-Chun Hsu

Department of Electrical Engineering, National Chin-Yi University of Technology,
No. 57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 41170, Taiwan

(Received November 18, 2020; accepted February 3, 2021)

Keywords: object detection, artificial intelligence (AI), deep learning, you only look once (YOLO)

This paper presents a teaching tool for schoolchildren to learn artificial intelligence (AI) technology through which a variety of banknotes can be recognized. This was done by first using a pretrained YOLOv3 object detection model. Secondly, transfer learning was conducted on the pretrained model using 11 collected banknotes, including US\$, Euro, Japanese Yen, and NT\$. The banknote detection model was experimentally validated to give an average precision (*AP*) of up to 99.09% if the threshold Intersection over Union (*IoU*) is not less than 0.8. Once a banknote was successfully recognized, the face value and the country name thereon were displayed, and schoolchildren can access suggested websites, i.e., Wikipedia, Google Maps, and the Bank of Taiwan, to learn more about the exchange rate between currencies and the history and location of the country that issued the banknote. Consequently, schoolchildren can have fun using this tool and acquire a more global outlook. Moreover, they may be motivated to become AI professionals in the future.

1. Introduction

Given their rapid advances, artificial intelligence (AI) technologies have been widely applied and already affect our daily lives. In AI-related technologies, deep learning is a hot issue, and considerable progress has been made in addressing image recognition issues in computer vision.^(1–4) A clear advantage of a deep learning model is that significantly improved recognition accuracy and robustness can be achieved. Furthermore, an input image can be directly applied to the model without conventional preprocessing.

Deep-learning-based image recognition models have been developed on the basis of convolutional neural networks (CNNs). Representative models include AlexNet,⁽⁵⁾ visual geometry group (VGG) Net,⁽⁶⁾ GoogLeNet, also referred to as Inception,⁽⁷⁾ and ResNet.⁽⁸⁾ Essentially, later-developed models are upgrades or modified versions of VGG Net, Inception, and ResNet. Object detection^(9–16) and face recognition^(17–19) are the most common techniques involved in image recognition tasks.

In an object detection task, it is necessary to first locate and then recognize specific objects in an image or a video. Object detection is an event in the ImageNet Large Scale Visual

^{*}Corresponding author: e-mail: cy.yeh@ncut.edu.tw
<https://doi.org/10.18494/SAM.2021.3193>

Recognition Competition (ILSVRC), hosted by ImageNet.⁽²⁰⁾ Today, commonly used object detection models include Regions with CNN features (R-CNN),⁽⁹⁾ Faster R-CNN,⁽¹⁰⁾ Single Shot MultiBox Detector (SSD),^(11,12) and You Only Look Once (YOLO).^(13–15) The COCO dataset⁽²¹⁾ can be trained to recognize up to 80 types of objects, including humans, vehicles, cats, and dogs, and can be widely applied to fields such as smart homes, smart security, smart traffic, and intelligent image analysis and retrieval.

However, the well-established object detection techniques have a limitation, that is, only the objects in a pretrained model can be detected, such as the 80 types of objects in the COCO dataset. In other words, transfer learning must be conducted so as to detect objects not contained in the COCO dataset. In this manner, object detection techniques can be applied to a wide variety of disciplines.

In light of this, we present in this paper a teaching tool for banknote detection, through which schoolchildren, especially those from seven to ten years old, can learn AI technology in a fun way. Once a banknote is successfully recognized, the face value and the country name are displayed instantly, and suggested websites, i.e., Wikipedia, Google Maps, and the Bank of Taiwan, are listed. Schoolchildren can access the listed websites to learn more about the country that issued the banknote, and consequently acquire a more global outlook. Hopefully, this tool will appeal to schoolchildren and encourage them to engage in the AI industry in the future.

The presented AI-based teaching tool was developed using a pretrained YOLOv3 object detection model.⁽¹⁵⁾ Transfer learning was conducted on the model using a variety of collected banknotes. The teaching tool can be used by schoolchildren to access the internet for teaching purposes. The YOLOv3 model is acknowledged as an efficient object detection model with a satisfactory mean average precision (mAP), a measure of object detection performance.^(21,22) This feature gives the YOLOv3 model a clear advantage over its counterparts.

This paper is outlined as follows. Section 2 refers to the YOLOv3 model used for object detection, Sect. 3 details the operation of the presented AI-based teaching tool, Sect. 4 gives a discussion of experimental results, and Sect. 5 concludes the paper.

2. YOLOv3 Object Detection

YOLO, short for you only look once, is a real-time convolution-based object detection algorithm. In reality, real-time detection can be carried out well using YOLO but at the cost of an acceptable degradation of precision. As its name indicates, YOLOv3 is the third version of YOLO⁽¹⁵⁾ and has a high speed. Moreover, as experimentally validated in Ref. 15, YOLOv3-320 has a slightly higher mAP and runs three times faster than SSD321. Major improvements in YOLOv3, as compared with earlier versions, are detailed as follows.

Firstly, YOLOv3 employs Darknet-53 as the backbone,⁽¹⁵⁾ which is an upgraded version of Darknet-19 used in YOLOv2. In addition to more layers in the backbone, ResNet and Feature Pyramid Networks (FPN) were introduced into Darknet-53 for the following reasons. Firstly, the vanishing gradient problem due to more layers in the backbone can be resolved using ResNet, and small objects can be well detected using the FPN structure, which was a major problem in the earlier versions. Similarly to the earlier versions, YOLOv3 lacks a fully connected (FC)

layer, and consequently, there is no limitation on input image dimensions, except that they must be multiples of 32.

In YOLOv3, multiscale detection is carried out using FPN. More precisely, multiscale refers to 3-scale here. For example, three feature maps of sizes 13×13 , 26×26 , and 52×52 are employed to detect an input image of size 416×416 . A small feature map is used to detect a large object, and vice versa. Moreover, three anchor boxes are employed for object detection in each layer, that is, a total of nine anchor boxes are used to detect nine bounding boxes. As a consequence, $13 \times 13 \times 3 + 26 \times 26 \times 3 + 52 \times 52 \times 3 = 10647$ bounding boxes in total are required in this case, which is more than 12 times as many as that in a YOLOv2 counterpart.

Object detection generates two quantities: object localization and classification. The former was referred to as the bounding box prediction in Ref. 15, which predicted the coordinates of the bounding boxes and the confidence scores of an object, and the latter was referred to as the class prediction therein. For training purposes, a loss function is defined as the sum of the loss of bounding box offsets, the loss of object confidence, and the loss of class prediction, formulated as

$$Loss = L_{box} + L_{conf} + L_{cla}, \quad (1)$$

where

$$L_{box} = \lambda_{coord} \sum_{i=1}^{N_b} O_i^{obj} \left[(t_{i,x} - \hat{t}_{i,x})^2 + (t_{i,y} - \hat{t}_{i,y})^2 + (t_{i,w} - \hat{t}_{i,w})^2 + (t_{i,h} - \hat{t}_{i,h})^2 \right], \quad (2)$$

$$L_{conf} = - \sum_{i=1}^{N_b} \left[O_i^{obj} \log(\sigma(t_{i,o})) + \lambda_{noobj} (1 - O_i^{obj}) \log(1 - \sigma(t_{i,o})) \right], \quad (3)$$

$$L_{cla} = - \sum_{i=1}^{N_b} \sum_{j=1}^{N_c} O_i^{obj} \left[l_{i,j} \log(p_i(c_j)) + (1 - l_{i,j}) \log(1 - p_i(c_j)) \right]. \quad (4)$$

Here, N_b and N_c are the number of predicted bounding boxes and the number of types of recognizable objects, respectively. $N_b = 10647$ for an input image of size 416×416 , and $N_c = 80$ for the COCO dataset; λ_{coord} and λ_{noobj} are two constants defined in Ref. 15; σ is the sigmoid function; $\{\hat{t}_{i,x}, \hat{t}_{i,y}, \hat{t}_{i,w}, \hat{t}_{i,h}, O_i^{obj}, l_{i,j}\}$ are all ground truth parameters; $\{t_{i,x}, t_{i,y}, t_{i,w}, t_{i,h}, t_{i,o}, p_i(c_j)\}$ are outcomes predicted by YOLOv3. Finally, note that L_{box} is expressed as the sum of squared error losses, while L_{conf} and L_{cla} are expressed as the binary cross-entropy loss.

Next, $O_i^{obj} \in \{0, 1\}$ is used to determine whether there exists an object in the i th bounding box. $O_i^{obj} = 0$ represents a negative case, while a value of 1 represents a positive case. Similarly, $l_{i,j} \in \{0, 1\}$ is used to determine whether there exists an object of the j th type in the i th bounding box. $l_{i,j} = 0$ represents a negative case, while a value of 1 represents a positive case. $\sigma(t_{i,o})$ is a confidence score predicted by YOLOv3, and is used to indicate the probability that

there is an object in the i th bounding box. $p_i(c_j)$ is used to indicate the probability that it is an object of the j th type. Finally, $\{t_{i,x}, t_{i,y}, t_{i,w}, t_{i,h}\}$ and $\{\hat{t}_{i,x}, \hat{t}_{i,y}, \hat{t}_{i,w}, \hat{t}_{i,h}\}$ are the coordinate offsets of the predicted and ground truth bounding boxes, respectively. The compensated coordinates of the bounding box are given by

$$b_x = \sigma(t_x) + c_x, \quad (5)$$

$$b_y = \sigma(t_y) + c_y, \quad (6)$$

$$b_w = P_w e^{t_w}, \quad (7)$$

$$b_h = P_h e^{t_h}, \quad (8)$$

where (b_x, b_y) represents the coordinates of the centroid of the bounding box, (c_x, c_y) represents the offset between the top-left corner of an image and that of the top-left grid cell, (b_w, b_h) and (P_w, P_h) represent the widths and heights of the bounding and ground truth boxes, respectively.

3. Proposed System

Our aim was to develop a teaching tool for schoolchildren to learn AI-related technologies. Schoolchildren are expected to have fun using the tool, become interested in AI technologies, and may even become more interested in being AI professionals in the future. Once a banknote is successfully recognized, the face value and the country name thereon are listed immediately. Schoolchildren can access suggested websites, i.e., Wikipedia, Google Maps, and the Bank of Taiwan, for more information, e.g., the exchange rate between currencies and the history and location of the country that issued the banknote. Hopefully, this will help schoolchildren to acquire a more global outlook.

Illustrated in Fig. 1 is the flow of the AI-based banknote detection tool. As can be seen therein, an image is captured using a webcam as the first step. Subsequently, the captured image is input into an AI-based banknote detection model. The image, identified as a banknote, is framed and the information thereon, i.e., the value and country name, is then displayed. A number of suggested websites, as mentioned previously, are also listed for users. Otherwise, the detection tool waits for the next input image. In this way, schoolchildren can familiarize themselves with the use of this AI-based teaching tool.

This work was developed using a pretrained YOLOv3 model, whereon transfer learning was carried out. There were two tasks before conducting transfer learning. The first was to collect training data, that is, a variety of banknotes having different denominations and issued by different countries. The second was to label the collected training data, including bounding boxes and classifications.

Table 1 lists the development environment in which the presented banknote detection system was developed. As can be seen therein, the codes were written in Python, and libraries including Keras, TensorFlow, OpenCV, and numpy were used. The hardware consists of a PC, a web camera, and a GeForce GTX 1060Ti graphics card.

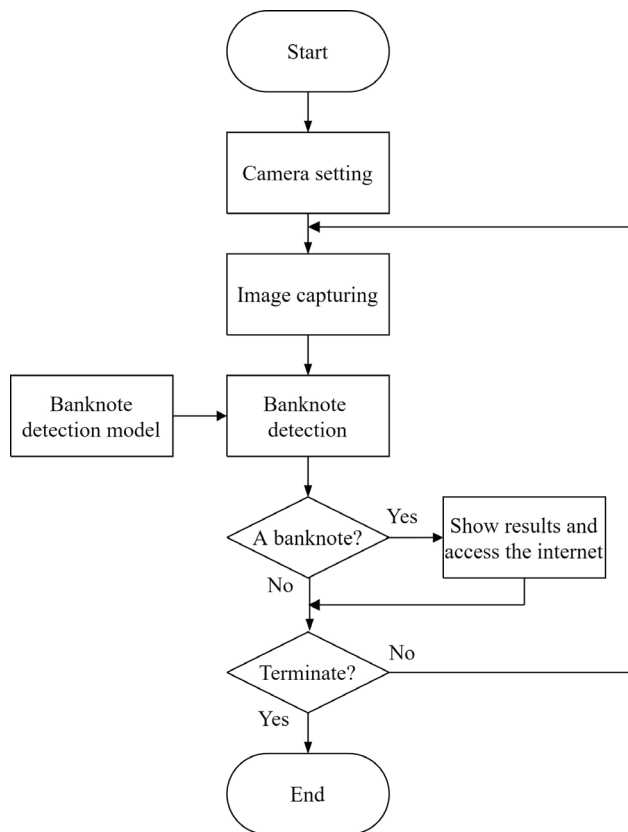


Fig. 1. Operation flow of the AI-based teaching tool.

Table 1

Development environment of AI-based teaching tool.

Programming language	Python
Library	Keras, TensorFlow, OpenCV, numpy, etc.
Detection model	YOLOv3
Hardware	PC, web camera, graphics card (GeForce GTX 1060Ti)

Table 2 lists the collected training data, that is, a total of 11 banknotes, each including the obverse and reverse sides. Since the image on the obverse side is very different from that on the reverse side, there are 22 banknote images in this work, numbered 1–22, as shown below the “image number” field in Table 2. Finally, transfer training was conducted using the collected training data in the pretrained YOLOv3 model, and the model was validated using the testing data. Table 3 gives all the numbered banknote images.

4. Experimental Results

An object detection model not only needs to recognize an object, but also has to determine a bounding box thereof. Precision is used as a performance measure of a detection model and was tested for our teaching tool. As listed in the rightmost column of Table 2, each collected banknote image was assigned 20 items of testing data, that is, there were 440 pieces of testing data in total. For unbiased testing, no items of training data were reused as items of testing data.

As its name indicates, the Intersection over Union (*IoU*) refers to the intersection area between two objects divided by the union area, expressed as

Table 2
Collected banknote images.

Banknote				Total number	Total number
Currency	Value	Side	Image number	of training data	of testing data
NTD	100	+	1	214	20
		-	2	214	20
	500	+	3	248	20
		-	4	248	20
	1000	+	5	245	20
		-	6	245	20
USD	5	+	7	214	20
		-	8	214	20
	20	+	9	228	20
		-	10	228	20
	50	+	11	238	20
		-	12	238	20
EUR	5	+	13	242	20
		-	14	242	20
	10	+	15	217	20
		-	16	217	20
	20	+	17	216	20
		-	18	216	20
JPY	100	+	19	242	20
		-	20	242	20
	1000	+	21	212	20
		-	22	212	20
Total				5032	440

Note that “+” and “-” identify the obverse and reverse sides of a banknote, respectively.

$$IoU(A, B) = \frac{A \cap B}{A \cup B}, \quad (9)$$

where A and B represent the predicted and ground truth bounding boxes, respectively. A high value of IoU indicates that there is a good match between A and B . Therefore, the precision was evaluated using IoU as a threshold. For example, IoU was set to 0.5 in the PASCAL VOC challenge.⁽²²⁾

The precision for the j th classification is defined as

$$precision(c_j) = \frac{TP(c_j)}{TP(c_j) + FP(c_j)}, \quad 1 \leq j \leq N_c, \quad (10)$$

where $N_c = 22$ represents the number of classifications. $TP(c_j)$ is the number of true positives (TP s) in the recognition of the object of the j th type. TP means that the predicted object type matches the ground truth type and IoU is greater than a default threshold. Otherwise, the predicted outcome is classified as a false positive (FP).

Table 4 lists the values of precision with IoU as a parameter. Note that 100% precision is achieved for all the images in the case of $IoU \geq 0.7$. This observation also applies to the case of

Table 3
Numbered banknote images.


















Banknote image (image number)			
 (1)	 (2)	 (3)	 (4)
 (5)	 (6)	 (7)	 (8)
 (9)	 (10)	 (11)	 (12)
 (13)	 (14)	 (15)	 (16)
 (17)	 (18)	 (19)	 (20)
 (21)	 (22)		

Table 4
IoU dependence of the precision across banknote images.

Image	Precision (%)			Image	Precision (%)		
	<i>IoU</i> ≥ 0.7	<i>IoU</i> ≥ 0.8	<i>IoU</i> ≥ 0.9		<i>IoU</i> ≥ 0.7	<i>IoU</i> ≥ 0.8	<i>IoU</i> ≥ 0.9
1	100	100	70	12	100	100	45
2	100	100	70	13	100	100	95
3	100	100	45	14	100	100	70
4	100	100	55	15	100	100	70
5	100	100	65	16	100	100	65
6	100	95	75	17	100	95	70
7	100	95	65	18	100	100	50
8	100	100	50	19	100	100	60
9	100	100	60	20	100	100	65
10	100	100	60	21	100	95	40
11	100	100	50	22	100	100	55

IoU ≥ 0.8, except that 95% precision is obtained for images 6, 7, 17, and 21. As compared with the previous two cases, the precision plunges across all the images in the case of *IoU* ≥ 0.9. The precision in each case is averaged and listed in Table 5. As can be seen therein, there is poor average precision (*AP*) in the case of *IoU* ≥ 0.9, that is, *AP* = 61.36%.

Table 5
AP over banknote images

<i>IoU</i> threshold	<i>IoU</i> ≥ 0.7	<i>IoU</i> ≥ 0.8	<i>IoU</i> ≥ 0.9
AP (%)	100	99.09	61.36

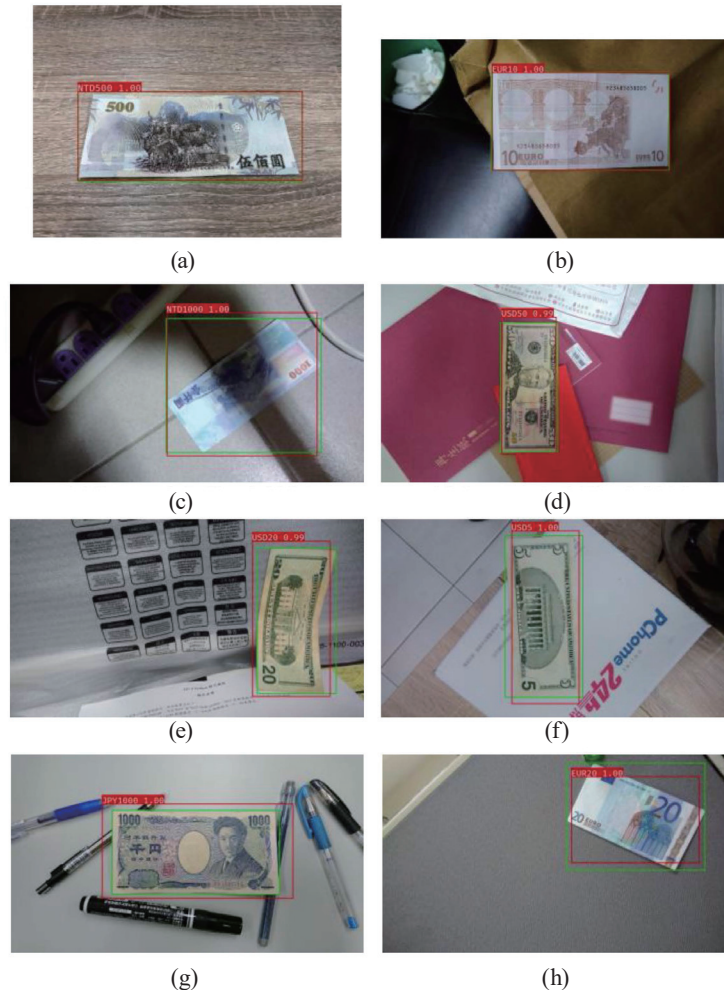


Fig. 2. (Color online) Detected images using the presented banknote detection model. (a) Image 4, *IoU* = 0.9837. (b) Image 16, *IoU* = 0.9836. (c) Image 6, *IoU* = 0.9086. (d) Image 11, *IoU* = 0.9016. (e) Image 10, *IoU* = 0.8183. (f) Image 8, *IoU* = 0.8173. (g) Image 21, *IoU* = 0.7811. (h) Image 17, *IoU* = 0.7459.

Figure 2 shows predicted and ground truth bounding boxes in red and green, respectively, for comparison purposes in each test case. The banknote images in Figs. 2(a) and 2(b) were detected with the highest and second highest values of *IoU*, respectively; in Figs. 2(c) and 2(d), they were detected with an *IoU* of approximately 0.9; in Figs. 2(e) and 2(f), they were detected with an *IoU* slightly higher than 0.8; and in Figs. 2(g) and 2(h), they were detected with an *IoU* below 0.8. The recognized currency, face value, and confidence score are also presented above the upper-left corner of the predicted bounding box in Figs. 2(a)–2(h).

There is a satisfactory match between the predicted and ground truth bounding boxes if $IoU \geq 0.8$, that is, an AP of up to 99.09%, as listed in Table 5. It must be stressed that the presented banknote detection model was developed as a teaching tool for schoolchildren and not as a counterfeit money detector. Therefore, an error in banknote recognition does not result in any loss. It is even possible that schoolchildren will be motivated to correct the error as the first step to becoming a young AI engineer.

5. Conclusions

This paper presented an AI-based teaching tool for schoolchildren. A variety of banknotes can be well recognized using the teaching tool, through which schoolchildren can obtain hands-on experience in AI technologies. A pretrained YOLOv3 model for object detection played a key role in this tool. Transfer learning was conducted on the pretrained model using collected banknote images. The banknote detection model was experimentally validated to perform well if $IoU \geq 0.8$, that is, an AP of up to 99.09%. Finally, the model was implemented as a teaching tool.

Once a banknote was successfully recognized, relevant websites, i.e., Wikipedia, Google Maps, and the Bank of Taiwan, were displayed instantly, and schoolchildren can access the websites to acquire a more global outlook through the recognized banknote, e.g., exchange rates between currencies and the history and location of the country that issued the banknote. Hopefully, this teaching tool will appeal to children and motivate them to become AI engineers in the future.

Furthermore, a more efficient model, such as YOLOv4, will be employed in the near future so as to upgrade the performance of banknote recognition. In addition, another interesting teaching tool for schoolchildren is also planned.

Acknowledgments

This research was financially supported by the Ministry of Economic Affairs, Taiwan, under grant number 108-EC-17-A-02-S5-008.

References

- 1 L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille: IEEE Trans. Pattern Anal. Mach. Intell. **40** (2018) 834.
- 2 V. Sze, Y. H. Chen, T. J. Yang, and J. S. Emer: Proc. IEEE **105** (2017) 2295.
- 3 W. G. Hatcher and W. Yu: IEEE Access **6** (2018) 24411.
- 4 E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long: IEEE Access **6** (2018) 39501.
- 5 A. Krizhevsky, I. Sutskever, and G. E. Hinton: Commun. ACM **60** (2017) 84.
- 6 K. Simonyan and A. Zisserman: CoRR (2014). <http://arxiv.org/abs/1409.1556>
- 7 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich: Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2015) 1. <https://doi.org/10.1109/CVPR.2015.7298594>
- 8 K. He, X. Zhang, S. Ren, and J. Sun: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2016) 770. <https://doi.org/10.1109/CVPR.2016.90>

- 9 R. Girshick, J. Donahue, T. Darrell, and J. Malik: Proc. 2014 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2014) 580. <https://doi.org/10.1109/CVPR.2014.81>
- 10 S. Ren, K. He, R. Girshick, and J. Sun: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 1137.
- 11 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg: CoRR (2015). <https://arxiv.org/abs/1512.02325>
- 12 S. Zhai, D. Shang, S. Wang, and S. Dong: IEEE Access **8** (2020) 24344.
- 13 J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2016) 779. <https://doi.org/10.1109/CVPR.2016.91>
- 14 J. Redmon and A. Farhadi: Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2017) 6517. <https://doi.org/10.1109/CVPR.2017.690>
- 15 J. Redmon and A. Farhadi: CoRR (2018). <https://arxiv.org/abs/1804.02767>
- 16 K. He, G. Gkioxari, P. Dollár, and R. Girshick: CoRR (2017). <https://arxiv.org/abs/1703.06870>
- 17 F. Schroff, D. Kalenichenko, and J. Philbin: Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2015) 815. <https://doi.org/10.1109/CVPR.2015.7298682>
- 18 H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu: Proc. 2018 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2018) 5265. <https://doi.org/10.1109/CVPR.2018.00552>
- 19 J. Deng, J. Guo, N. Xue, and S. Zafeiriou: Proc. 2019 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2019) 4685. <https://doi.org/10.1109/CVPR.2019.00482>
- 20 ImageNet Large Scale Visual Recognition Competition (ILSVRC): <http://www.image-net.org/challenges/LSVRC> (accessed January 2018).
- 21 COCO dataset: <http://cocodataset.org> (accessed May 2018).
- 22 PASCAL VOC challenge: <http://host.robots.ox.ac.uk:8080/pascal/VOC> (accessed May 2018).