IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 2, 2019, accepted December 11, 2019, date of publication December 16, 2019,
date of current version January 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2959667

Posting Techniques in Indoor Environments Based on Deep Learning for Intelligent Building Lighting System

XIAOPING LIN[1], PEIYONG DUAN[1], YUANJIE ZHENG[1,3,4,5], WENJIAN CAI[2], AND XIN ZHANG[2], (Member, IEEE)

[1]School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China
[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
[3]Key Laboratory of Intelligent Computing and Information Security in Universities of Shandong, Shandong Normal University, Jinan 250358, China
[4]Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, Shandong Normal University, Jinan 250358, China
[5]Institute of Biomedical Sciences, Shandong Normal University, Jinan 250358, China

Corresponding author: Peiyong Duan (duanpeiyong@sdnu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61773246 and 61374187, and in part by the Major Program of Shandong Province Natural Science Foundation under Grant ZR2018ZB0419.

ABSTRACT Recently, with the rapid development of society, solutions to reduce energy consumption in the world have attracted a lot of attention, especial electric energy. In this regard, a system that can control light on and off by determining the location of the person to reduce the waste of electricity used in public buildings, called intelligent building lighting system. Following the practical requirements of the intelligent building lighting system, a technique for positioning in indoor environments is proposed, supporting the design of a positioning system based on deep learning and the Cerebellar Model Articulation Controller (CMAC), called Y-CMAC.This technique adopts YOLOv3 (the method in the paper of YOLOv3: An Incremental Improvement) for object detections and makes the coordinate of a person in the image. On the other hand, using CMAC to calculate the actual position of the person in the indoor environment. Moreover, massive surveillance video is used to reduce the cost of equipment and facilitate the promotion of applications. The average positioning error is controlled at around 1m in this paper.

INDEX TERMS Intelligent building lighting system, indoor positioning, monitor identification, CMAC, deep learning.

I. INTRODUCTION

With the continuous development of the social economy, the accelerating process of global industrialization and urbanization, and the scale of the construction industry in various countries has continued to expand. The problem of building high energy consumption has become more and more serious in the world, and the energy-saving environmental protection has aroused people's high concern [1]. The electricity energy consumption is more serious than the other energy consumption in buildings, especially for lighting. According to incomplete statistics, lighting electricity accounts for about 15% of the total building energy consumption. An intelligent building lighting system is proposed in this paper to reduce the waste of electricity used for lighting. The intelligent building lighting system used artificial intelligence technology to monitor and track the indoor personnel in real-time, and determine the position coordinates of the personnel. The wireless transmission method automatically and smoothly adjusts the circuit voltage and current amplitude and the power supply position. It is a lighting control system that improves the efficiency of electricity consumption and reduces the waste of building electricity. Determining the position coordinates of the indoor personnel is important to achieve this function. Therefore, this paper focuses on an indoor personnel positioning method.

In the age of information, the ability to located persons and devices indoor environments has become increasingly important for a rising number of applications. With the emergence of Global Positioning Systems(GPS), the performance of outdoor positioning has become excellent [2]. But the indoor environment is beyond the coverage of GPS [3]. Therefore the

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

TABLE 1. Mainstream indoor positioning technologies.

Technology	Accuracy	Limitations
Infrared	mm	Easy to be blocked by objects and high equipment costs
Supersonic	3cm-10cm	Signal attenuation is fast and susceptible to multipath effects
Bluetooth	2m-3m	Large base station density
Ultra-Wideband	2cm-30cm	Easy to be blocked by objects and high equipment costs
Radio Frequency Identification	10cm	The high equipment costs
Wi-Fi	3m-5m	Poor positioning accuracy; Need to collect large amounts of data in advance

technology of indoor positioning has become a focus of research during the past decade. Recently, there are several existing indoor personnel positioning technologies:

Infrared technology(IR) [4] is a relatively mature indoor positioning technology, which is divided into static and dynamic positioning according to the state of the object to be determined. The existing IR indoor positioning system [5] has several indoor positioning systems such as Active Badge [6] and Firefly. At present, the average positioning error of IR is mm, but it has disadvantages. On the one side, IR can only be transmitted by line-of-sight, and it cannot be positioned when encountering obstacle obstructions. On the other side, the equipment is not easy to be promoted due to high costs.

Radiofrequency identification technology(RFID) [7]can be generally divided into active and passive categories. The RFID indoor positioning technology can obtain centimeter-level positioning accuracy information within a few milliseconds, with a small logo size and a large transmission range. However, it has the disadvantage of poor anti-interference ability, and it is difficult to promote its application in practical systems.

The indoor positioning technology of ultrasonic wave [8] is calculated the distance between the two points by calculating the time between the known point and the to-be-determined point, and obtains the to-be-determined point between three or more known points. However, technology is easily affected by obstruction, reflection, refraction, etc.

Ultra-Wide Positioning Technology (UWB) is using pre-arranged anchor nodes and bridge nodes at known locations to determine position using triangulation or ''fingerprint'' methods. However, it has high costs and power consumption [9]. WIFI technology [10] is very mature, but for indoor positioning, the accuracy is too low, it is easy to be disturbed by the environment, and it is difficult to update the data. The Bluetooth positioning principle is similar to WIFI technology, and the most widely used iBeacon [11] technology is used. However, the reason why it cannot be widely used is that the positioning accuracy is low, generally 2-3 m, requiring more hardware, high deployment density, and high software cost, making the system expensive.

The above technology is not easy to promote in the building(Table1). Then, a technology is proposed, and we summarize our contributions as follows:

First, an indoor personnel positioning system that combines two neural networks, including convolutional neural network(YOLOv3)and Cerebellar Model Articulation Controller(CMAC). It is a new method of indoor personnel positioning.

Second, we use surveillance video data to achieve indoor personnel positioning. Compared with IR, RFID, and other technologies, we will reduce the cost of indoor personnel positioning that using existing surveillance video data, and this method is easy to promote.

Furthermore, this method is used to control the switch on or switch off of the light for the intelligent building lighting system. When we know the position of the person, we will transfer the position coordinates of the personnel to the light controller through the usbto433 interface, then we can control the switch of the light to reduce the waste of electricity used in lighting.

II. PRELIMINARY OF DEEP LEARNING AND CMAC

In recent years, deep learning technology can be applied to many fields [13], [32], [33], [35], which has attracted increasing interest from researchers. In this paper, we mainly explain two parts, one is object detection using convolutional neural network(CNN [12]–[15]) and the other is a coordinate approximation based on CMAC.

Object detection means when you get a video or an image, you will know the position of the object in the video or the image and know what the object is. Recently object detection technology has developed rapidly. The following are the more commonly used target detection methods:

DPM [16] as the best algorithm in the field of detection extracts the artificial features and classifies them with latetSVM. But the calculation about its speed is slow, and the detection effect is poor. Then Alex Krizhecsky proposed the Alexnet [17] with deep convolutional neural networks. This network can learn useful features by training many data. Zhang xian et.al proposed OverFeat [18] which uses the sliding windows to avoid the repetitive. But this network has a poor effect on the small object and the calculation is large. Ross B. Girshick of the University of California at Berkeley in 2014 proposed an Region-CNN (R-CNN) network for object detection [19]. The R-CNN inputs a fixed-size RGB image, and the final output is a 4096-dimensional feature vector. The candidate region is classified by a linear support vector machine. First, the feature vectors of all candidate regions are calculated for each image to be detected, and then the classification operation is implemented using a support vector machine. At the same time, it is sent to the fully connected network for coordinate position regression.

Ross Girshick made further improvements, and they promoted Fast R-CNN [20]. Its innovation is in RoI Pooling, and it uniformly samples the convolutional feature maps of different size candidate frames into fixed-size features. Because of the above operations, the training time of the network has been greatly reduced. Shaoqin Ren et al. proposed a Faster R-CNN [21] based on Fast RCNN and added RPN (Region Proposal Network) network in the backbone network. With the advent of the You Only Look Once (YOLO) [22] algorithm, the deep learning target detection algorithm begins with two-stage and single-stage. First, the algorithm needs to preprocess the data and crop the image to be detected to a uniform size. The image is divided equally into a fine-grained mesh to achieve the purpose of detecting objects at different locations. If the center of the target falls in a certain grid unit, the grid unit predicts the object. As the most advanced real-time object detection system, YOLO9000 can detect more than 9,000 object categories [23]. Compare with YOLOv1, that became more faster and stronger. YOLOv3, a faster network than YOLOv2, is quite good [24]. Single Shot MultiBox Detector (SSD) [25] adopts the method of meshing, and complete all operations in the same convolutional layer. It improves speed while ensuring accuracy. There are some methods of object detection that are mature, like Deconvolutional Single Shot Detector (DSSD) [26], Feature Fusion Single Shot Multibox Detector (FSSD) [27] et.al. As far as we know, although there are different methods to realize object detection one has to use the method which has high accuracy and fast speed if we want to recognize our purpose. In this paper, the method of YOLOv3 is used which has accuracy and can shorten the run time, and fine-tuning on this basis.

In 1975, Albus [28] proposed a new method using a controller called CMAC (Cerebellar Model Joint Controller) to solve the nonlinear coordinate approximation problem. It has the ability to store information about the input-output relationships of nonlinear dynamic systems. In 2007, S. D. Teddy proposed a novel CMAC architecture called hierarchical clustering adaptive quantization CMAC (HCAQ-CMAC) [29]. They use hierarchical clustering to non-uniformly quantify the input space to identify important input segments and apply HCAQ-CMAC to the automatic control of vehicle handling and modeling. Wang et al. Proposed neural embedding matching [34] based on representation learning to realize the conversion from source domain to target domain. This method can be used for personnel position conversion. In 2016, Xu and Jing [30] proposed the concept of association membership based on the analysis of the structural parameters of the CMAC neural network. He designed a novel Association Membership CMAC (AM-CMAC). Nonlinear functions based on industrial control problems can be solved using a fitting approximation. In the same year, a cerebellar model articulation controller (CMAC) network based on recurrent function link (RFL) was proposed to solve the problem of identification and prediction [31], and accurate recognition response and excellent dynamic performance are obtained.

In this paper, we will use the collected data to train CMAC, then we will calculate the actual position of the object.

III. METHOD OF Y-CMAC

Video-based indoor location is one of the most important tasks in smart building lighting systems. The efficient positioning algorithm can not only determine the position of the personnel, but also achieve the intelligent control effect, and also reduce the cost and achieve the promotion effect. Our proposed indoor personnel positioning system is called Y-CMAC. Based on the Darknet network and YOLOv3 and CMAC algorithms, real-time indoor personnel identification and localization algorithm are constructed to achieve equalization of speed and precision. The first algorithm is a convolutional neural network for detecting objects, and the second algorithm is CMAC for coordinate approximation. The entire system is a single, unified end-to-end system for personnel positioning (Figure 1). In Section A, we introduced the network design for object detection. In Section B, we introduced the method of coordinate approximation.

A. CONVOLUTIONAL NEURAL NETWORK

In this part, we adopt the fine-tuned YOLOv3 [24] algorithm. Based on the YOLOv1 [22] and YOLOv2 [23] networks, the YOLOv3 model performs some application improvement methods, such as multi-label classification and multi-scale detection. In this method, the author frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. But in our work, we only used the bounding boxes and the coordinates of people in the image (the image is intercepted from the monitor video streams). Because we just need to detect a person in our work. The model of YOLOv3 uses the improved Darknet-53 network based on the residual neural network as the feature extractor. Why choose this network? There are two main purposes: a. This network structure can have good classification results on ImageNet. b. Initialize the subsequent test model. Because this network is strong. Compared to the ResNet-152 and ResNet-101 network architectures, the Darknet-53 has fewer network layers, requires less time to run, and results are more accurate. Darknet-53 means that there are 53 convolutions. First, YOLOv3 scales the original image to a size of 416 x 416 using a scaled pyramid structure similar to the FPN network. Then, the original image is divided into $S \times S$ equal-sized pixels according to the scale of the feature map. The detection is performed on three scales with feature map sizes of 13×13, 26×26, and 52×52, and the feature map is transmitted on two adjacent scales using 2 times upsampling. Each cell predicts three bounding boxes, where the bounding box has three anchoring boxes. The predictions correspond to:

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \tag{1}$$

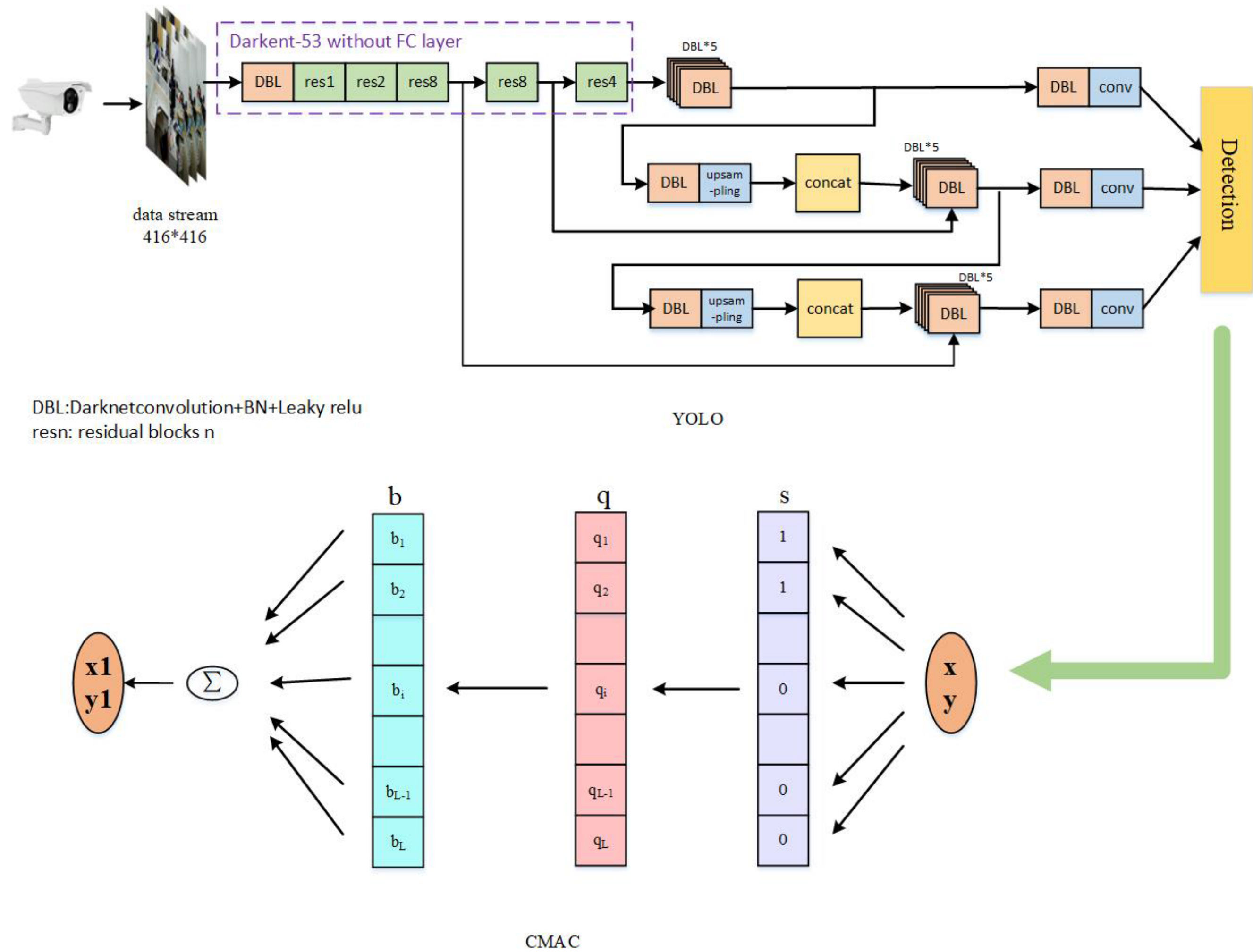


FIGURE 1. **Y-CMAC is a single, unified system for personnel positioning. It adds coordinate approximation to the end of YOLOv3, which can local person position.**

where t_x, t_y, t_w, t_h is the convolutional neural network predicts 4 coordinates for each bounding box on each cell, include the height h, width w and the coordinates (x, y). Where c_x, c_y is the offset of the target center in the top left corner of the image. Where p_x, p_y is the width and height of the predicted bounding box. Here, we use the logistic regression method to predict the bounding box. Each time a prediction is made, (t_x, t_y, t_w, t_h) is output, and then the corresponding coordinates are obtained through the above formula. The role of logistic regression is to perform a target score, determine the probability that the surrounding location is the target, and also reduce the amount of calculation.

This part of the object detection model can be divided into two network modules. The first part is the darknet which includes that the smallest unit of the YOLO network structure and the other block. The smallest unit is composed of a convolution layer, a BN layer and a Leaky Relu layer. We adopt a consistent approach, adding batch normalization after the convolution operation, and using Leaky relu as the activation function. When the input is less than 0, a small gradient can still be given without causing complete information loss. The other is composed of a padding layer, the smallest unit and a compose combination block. Compose is used to stack the network layer. It consists of a residual network composed of N * the smallest unit, and N represents the number of residual blocks. Residual block is used to solve the problem of vanishing gradient and improve network performance. The second part is the output part of a network which mainly samples and connects the features and outputs them. The output of darknet is the input of this part. Due to Darknet-53 architecture, it is easy to detect small objects. On the other side, Darknet-53 used residual layer jump connections, that can reduce each time convolution, and seep up the operation.

B. CEREBELLAR MODEL ARTICULATION CONTROLLER

In this part, we based on CMAC [22] and automatically transfer coordinates to CMAC to get the location of a person. Before training, we first standardize the data to facilitate better use in training. There are several common normalization methods: Min-Max Scaling, Z-score standardization and Nonlinear Scaling et.al. In our work, we used Min-Max Scaling to convert the raw data value to [0,1]:

$$x_k = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

where X is the abscissa of the coordinates from fine-tuned YOLOv3; X_{min} is the minimum value of the abscissa; X_{max} is the maximum value of the abscissa. We implemented a data normalization operation to speed up network training and make it easier to converge to the optimal solution.

Let A be a normalized input space, and divide the A into equal meshes. The divided mesh intersection number are recorded as P_j $(j = 1, 2, \ldots L)$, and the corresponding weights is q_j. Then the super-closed sphere is defined by P_j as:

$$U = \left\{x \middle| \left\|x - P_j\right\| \le R_b,\ x \in A\right\} \tag{3}$$

where R_b is the radius of super-closed sphere, and the value is 0.1.

Define the Gaussian function on the super-closed sphere as:

$$b_j(x_k) = \begin{cases} exp\left[\dfrac{\left\|x_k - P_j\right\|^2}{\sigma^2}\right], & x_k \in U \\ 0, & x_k \notin U, \end{cases} \tag{4}$$

where X is a (any) point, σ is 2.5.

Let $S = \{(x_k, y_k)\}\ (k = 1, 2, \ldots, N)$ be the learning sample whose output is a linear combination of the basis functions on the hyper-closed sphere centered on the active node, the error is:

$$\widehat{y_k} = S_k^T B(x_k)\, q \tag{5}$$

where $B(x_k) = diag\,[b_1(x_k), b_2(x_k), \ldots, b_L(x_k)]$; L is the number of grid intersections, $q = [q_1, q_1, q_1]^T$ is the weight coefficient vector, $S_k = [S_k, L]_{L\times 1}$ is the weight coefficient selection vector.

Then modify the weight coefficient until the output error meets the requirements:

$$\triangle q_{k-1} = \frac{\alpha\, e_{k-1} B(x_{k-1})\ S_{k-1}}{\beta + S_{k-1}^T B(x_{K-1})\ B^T(x_{K-1})\ S_{k-1}} \tag{6}$$

where α, β are constant, the value are $0 < \alpha < 2$, $\beta > 0$ and the training ends.

The first half of the model we proposed implements the detection of the position of the person in the video, and then realizes the mapping of the person's location to the three-dimensional space through the CMAC network to obtain the true position of the person. Our approach minimizes energy consumption.

IV. EXPERIMENTS

Following the structure of the methods, we provide our experimental results to analyze our work with respect to the different techniques and parameters. We set the learning rate in the system to 0.001, and the training time is about 3 hours. The training process mainly involves personnel detection and mapping of personnel positions to three-dimensional space. To speed up the calculation speed, we use GPU for training. The experiments were performed on a work station with intel(R) Core(TM) i7-8700K CPU, 16 GB RAM, and GeForce GTX1080Ti GPUs. We evaluate the performances of Y-CMAC under the classroom. We consider a monitor which is occupied by a placed at the (any) corners of the ceiling or the center of the wall. In our work, we chose to place the monitor in the center of the north wall. The potential objection is assumed to be distributed in the whole region with a mean density (Figure2).

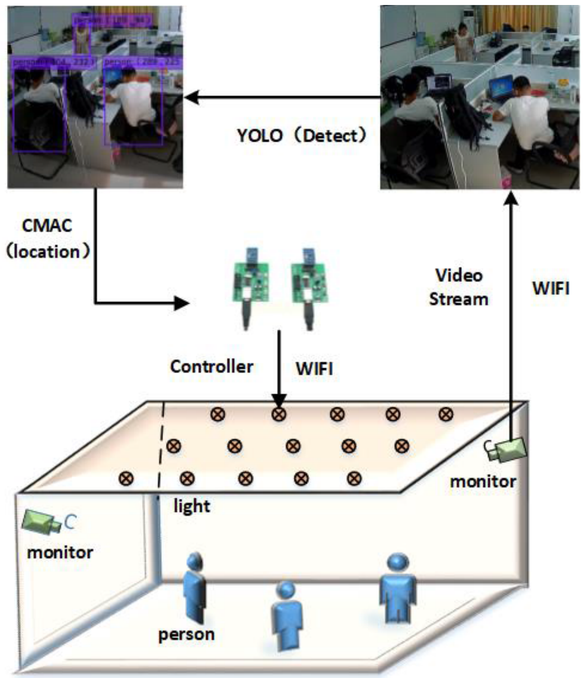


FIGURE 2. The system of intelligent building lighting system.

(1) Capture images from surveillance video streams using the protocol of RTSP.

(2) Set the video size to 1000*1000, and use the fine-tuned YOLOv3 model to detect the person in the image, and output the coordinate.

(3) Determine the product space and normalize it, then determining the spatial node, the initial value of the selection weight coefficient.

(4) Select relevant parameters, such as σ, ρ et.al.

(5) Given a sample point, find the super-closed sphere containing the point, it means to determine the selection matrix. Then estimating the estimated value and repeat it until the error reaches our expectations.

In this paper, we selected a small sample for testing. For future applications, we choose a classroom in which the length is 6.4m and the wide is 8.8m. We divide the room into 11*8 grids. When we training our system, we need people to move around the four vertices and centers of the grid. The result as follows(Figure 3 and Figure 4):

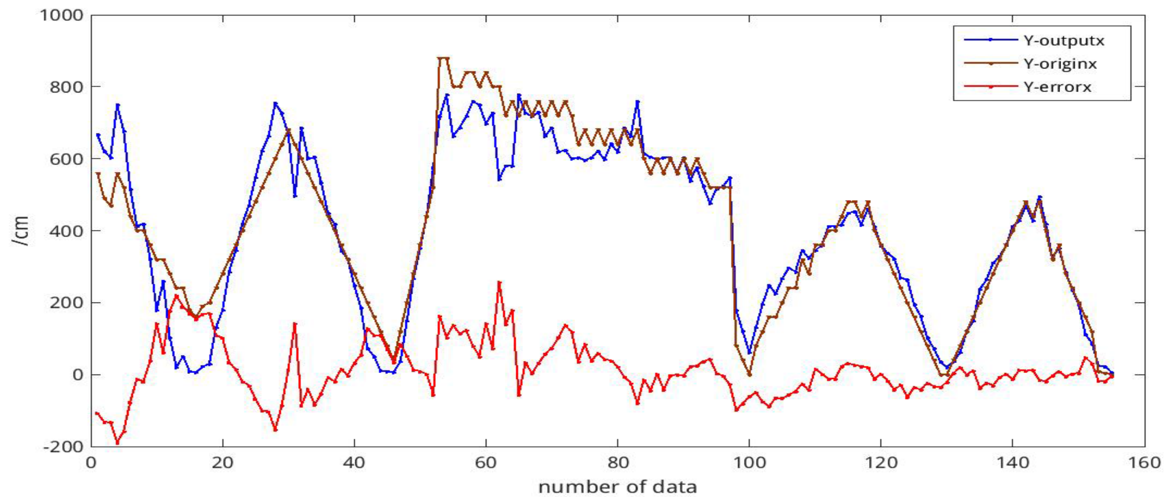


FIGURE 3. The result of YOLOv3-CMAC-x.

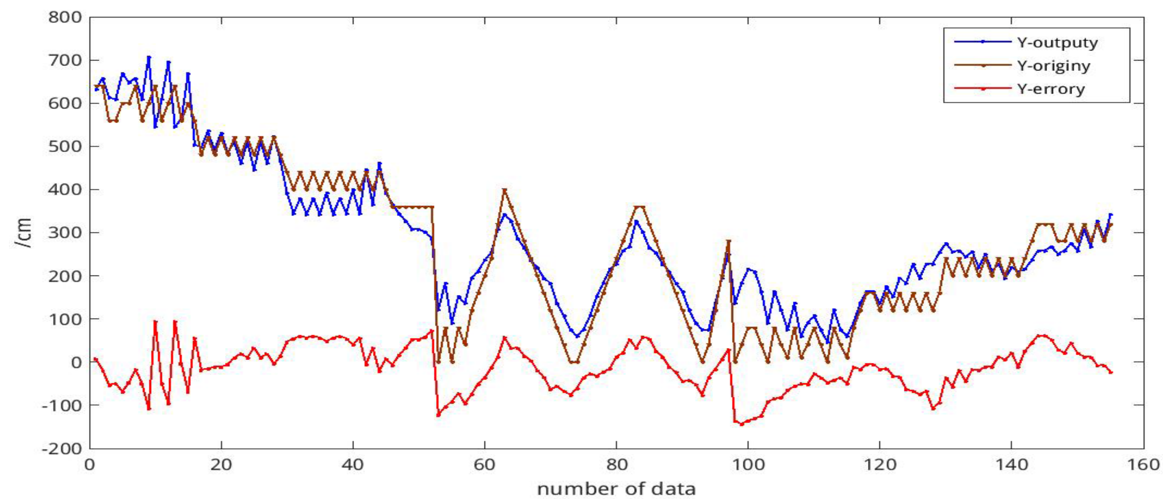


FIGURE 4. The result of YOLOv3-CMAC-Y.

Figure 3 shows the absolute error of the person' abscissa which is used the fine-tuned YOLOv3; figure 4 shows the absolute error of the person' ordinate which is used the fine-tuned YOLOv3. The number of lights in the classroom where we eperimented is 3*5. The distance between each light is 2m. As shown in the figure3 and figure4, we can see that the small sample positioning error is about 1 m. The results meet the requirements of our work in the future.

We also train Y-CMAC to use fine-tuned SSD [25]. The SSD model is based on a feedforward convolution network and is capable of generating a fixed-size set of bounding boxes and scoring the instances of the object classes that exist in the bounding box. The model is based on a standard architecture which is VGG-16 network and adds convolutional feature layers to the end of the truncated based network. The added feature layer has the effect of producing a fixed detection prediction. For each of the default boxes, it should be fixed relative to the location of its corresponding feature map cell. When they training this model, the ground truth information needs to be assigned to specific outputs in the fixed set of detector outputs by matching each ground truth box to the default box with the best jaccard overlap. Then the loss function and backpropagation are applied end-to-end. The result as follows(Figure 5 and Figure 6):

Figure 5 shows the absolute error of the person' abscissa which is used the fine-tuned SSD; figure 6 shows the absolute error of the person' ordinate which is used the fine-tuned SSD. As shown in the figure5 and figure6, we can see that the small sample positioning absolute error is about 1 m in the same experimental environment. Comparing the experimental results of the two models, it is not difficult to find that the model trained by fine-tuned YOLOv3 is more accurate and significantly faster than fine-tuned SSD. When we calculated the relative error of the two models, we found that the highest

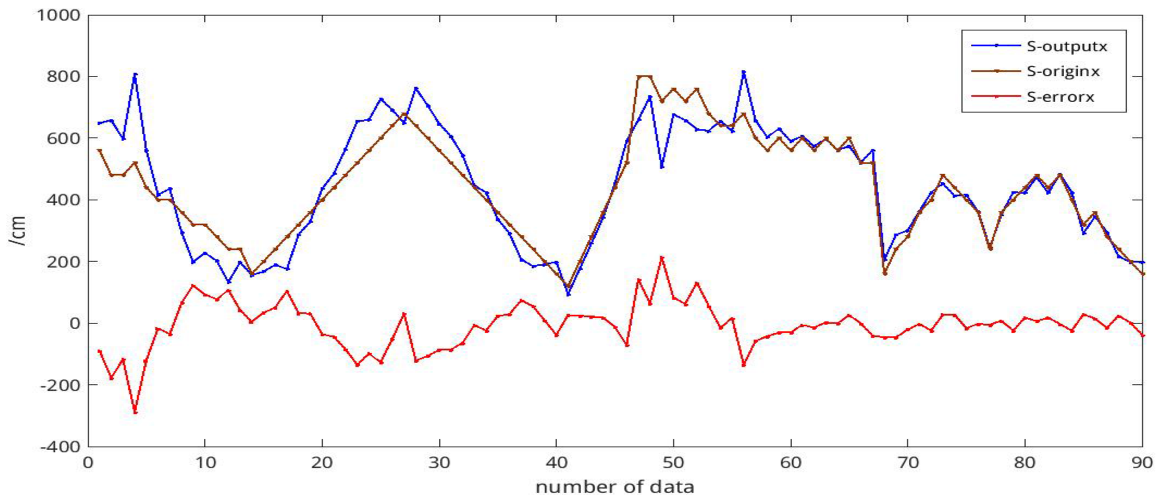


FIGURE 5. The result of SSD-CMAC-x.

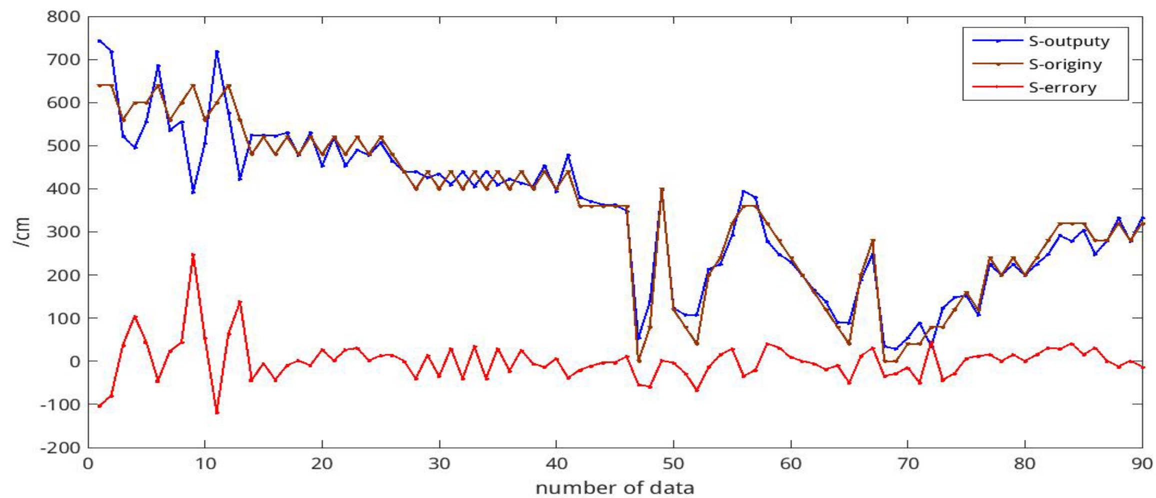


FIGURE 6. The result of SSD-CMAC-y.

relative error of the Y-CMAC was 0.22, and the average was 0.14. The fine-tuned SSD model had the highest relative error of 0.36, and the average relative error of 0.18. It has a relative advantage over other SSD-dependent systems, but it is slightly slower than real-time systems. At the same time, the other import problem is that when we used the model which detected by fine-tuned SSD, the object of distant did not be detected. So we choose YOLOv3 to detected objects at last.

V. CONCLUSION AND FUTURE WORK

This paper mainly proposes a new indoor positioning scheme based on the deep learning algorithm, called Y-CMAC, to achieve cost-effective indoor positioning. The CMAC algorithm is used in data processing to obtain a more accurate result, and an indoor personnel location data set is designed and designed. The experimental results show that the YOLOv3 and CMAC algorithms used in this paper are feasible and efficient for indoor personnel positioning. Although the current positioning technology is not accurate enough to be millimeters, it is sufficient for our future intelligent building lighting system. In particular, the cost is small and it is easy to promote in the future.

In the future, we will further optimize our personnel positioning model, improve the speed and accuracy of the model, reduce the energy consumption of the intelligent building lighting system, and hope to deploy our model based on the development board.

REFERENCES

[1] D. Wu, "Analysis on zero energy consumption strategy for office buildings lighting in Lianyungang area," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 111, no. 1, 2018, Art. no. 012035.

[2] R. Mautz and S. Tilch, "Survey of optical indoor positioning systems," in *Proc. IEEE Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, 2011.

[3] Z. Yin, X. Jiang, Z. Yang, N. Zhao, and Y. Chen, "WUB-IP: A high-precision UWB positioning scheme for indoor multiuser applications," *IEEE Syst. J.*, vol. 13, no. 1, pp. 279–288, Mar. 2019.

[4] T. Arai, T. Yoshizawa, and T. Aoki, "Evaluation of indoor positioning system based on attachable infrared beacons in metal shelf environment," presented at the IEEE Int. Conf. ICCE, Las Vegas, NV, USA, Jan. 2019.

[5] I. AL-Furati, A. T. Rashid, and A. Al-Ibadi, "IR sensors array for robots localization using K means clustering algorithm," in *Proc. Int. Conf. UKSim-AMSS*, Cambridge, U.K.:, Mar. , 2019.

[6] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.*, vol. 10, no. 1, pp. 91–102, 1992.

[7] L. M. Ni, Y. Liu, and Y. C. Lau, "LANDMARC: Indoor location sensing using active RFID," in *Proc. PerCom*, Fort Worth, TX, USA, Mar. 2003, pp. 407–415.

[8] M. Hazas and A. Ward, "A novel broadband ultrasonic location system," in *Proc. Int. Conf. Ubiquitous Comput.*, Berlin, Germany, Sep. 2002.

[9] M. Ridolfi *et al.*, "Experimental evaluation of UWB indoor positioning for sport postures," *Sensors*, vol. 18, no. 1, p. 168, 2018.

[10] C. Figuera, J. L. Rojo-Alvarez, I. Mora-Jimenez, A. Guerrero-Curieses, M. Wilby, and J. Ramos-Lopez, "Time-space sampling and mobile device calibration for WiFi indoor location systems," *IEEE Trans. Mobile Comput.*, vol. 10, no. 7, pp. 913–926, Jul. 2011.

[11] M. Koühne and J. Sieck, "Location-based services with iBeacon technology," in *Proc. 2nd Int. Conf. Artif. Intell., Modelling Simulation*, Madrid, Spain, Nov. 2014, pp. 315–321.

[12] Z. Cai, Q. Fan, and R. S. Feris, "A unified multi-scale deep convolutional neural network for fast object detection," in *Proc. ECCV*, 2016, pp. 354–370.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[16] P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. Int. Conf. CVPR*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[18] P. Sermanet, D. Eigen, and X. Zhang, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*. [Online]. Available: https://arxiv.org/abs/1312.6229

[19] R. Girshick, "Fast R-CNN," in *Proc. ICCV*, Santiago, Chile., Dec. 2015, pp. 1440–1448.

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR*, Columbus, OH, USA, Jun. 2014, pp. 580–587.

[21] S. Ren, K. He, and R. Girshick, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2015, pp. 91–99.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. CVPR*, Las Vegas, Nevada, USA, Jun. 2016, pp. 779–788.

[23] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 7263–7271.

[24] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: https://arxiv.org/abs/1804.02767

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. ECCV*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[26] C. Y. Fu, W. Liu, and A. Ranga, "Dssd: Deconvolutional single shot detector," 2017, *arXiv:1701.06659*. [Online]. Available: https://arxiv.org/abs/1701.06659

[27] Z. Li and F. Zhou, "FSSD: Feature fusion single shot multibox detector," 2017, *arXiv:1712.00960*. [Online]. Available: https://arxiv.org/abs/1712.00960

[28] J. S. Albus, "A new approach to manipulator control: The cerebellar model articulation controller (CMAC)," *J. Dyn. Syst., Meas., Control*, vol. 97, no. 3, pp. 220–227, 1975.

[29] S. D. Teddy, E. M. K. Lai, and C. Quek, "Hierarchically clustered adaptive quantization CMAC and its learning convergence," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1658–1682, Nov. 2007.

[30] S. Xu and Y. Jing, "A new CMAC neural network algorithm based on association membership," in *Proc. Int. Conf. CCDC*, Yinchuan, China, May 2016, pp. 1857–1860.

[31] Y.-L. Liao, Y.-C. Hu, and Y.-F. Peng, "Application of recurrent functional-link-based CMAC networks for identification and prediction," in *Proc. Int. Conf. IIAI-AAI*, Kumamoto, Japan, Jul. 2016, pp. 620–621.

[32] Y. Wu, B. Jiang, and N. Lu, "A descriptor system approach for estimation of incipient faults with application to high-speed railway traction devices," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 10, pp. 2108–2118, Oct. 2019.

[33] Y. Wu, B. Jiang, and Y. Wang, "Incipient winding fault detection and diagnosis for squirrel-cage induction motors equipped on CRH trains," *ISA Trans.*, to be published.

[34] Z. Wang, B. Du, and Y. Guo, "Domain adaptation with neural embedding matching," *IEEE Trans. neural Netw. Learn. Syst.*, to be published.

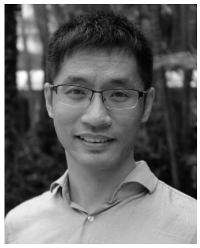
[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

XIAOPING LIN was born in Shandong, China, in 1994. She received the bachelor's degree from the School of Information Science and Engineering, Qufu Normal University, Rizhao, China, in 2017. She is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong Normal University, Jinan, China. Her current research interests include deep learning and intelligent building optimization algorithm.

PEIYONG DUAN received the B.S. degree from the Shandong University of Technology (merged into Shandong University, in 2000), Shandong, China, in 1996, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 1999. From 1999 to 2014, he was with the School of Information and Electrical Engineering, Shandong Jianzhu University, where he was appointed as an Associate Professor and a Full Professor, in 1999 and 2002, respectively. From 2017 to 2019, he was with the School of Information Science and Engineering, Shandong Normal University, Shandong. He has authored over 60 refereed articles. His current research interest includes discrete optimization and scheduling.

YUANJIE ZHENG was a Senior Research Investigator with the Perelman School of Medicine, University of Pennsylvania. He is currently a Professor with the School of Information Science and Engineering, Shandong Normal University, and a Taishan Scholar of the People's Government of Shandong Province of China. He is also serving as the Vice Dean of the School of Information Science and Technology and the Institute of Life Sciences, Shandong Normal University. His research interests include medical image analysis, translational medicine, computer vision, and computational photography. His ultimate research goal is to enhance patient care by creating algorithms for automatically quantifying and generalizing the information latent in various medical images for tasks, such as disease analysis and surgical planning through the applications of computer vision and machine learning approaches to medical image analysis tasks and development of strategies for image-guided intervention/surgery.

WENJIAN CAI received the B.Eng. degree from the Department of Precision Instrumentation Engineering, Harbin Institute of Technology, China, in 1980, the M.Eng. degree from the Department of Control Engineering, Harbin Institute of Technology, China, in 1983, and the Ph.D. degree from the Department of Electrical Engineering, Oakland University, USA, in 1992. Since 1999, he has been with the School of EEE, Nanyang Technological University. He has more than 20 years of experience in industry and research, in the areas of mechanical design, system modeling and simulation, energy and environmental system automation, and process control. He participated in many industry-related research projects, and published more than 100 technical articles, three books, and holds two patents. His areas of expertise are system modeling, control and optimization, multivariable system identification and control, sensor and instrumentation, mechanical system simulation and design, and intelligent systems. His current research works focus on industry applications in building HVAC processes, renewable energy processes, and environmental processes. He received three national awards. He has been often invited as a technical committee member or referee and a reviewer for a number of premier conferences and journals, including the *Journal of Process Control* and *Energy Conversion and Management*.



XIN ZHANG (M'15) received the Ph.D. degree in automatic control and systems engineering from The University of Sheffield, U.K., in 2016, and the Ph.D. degree in electronic and electrical engineering from the Nanjing University of Aeronautics and Astronautics, China, in 2014. He is currently an Assistant Professor of the EEE School and the Cluster Director of the ERIAN of Nanyang Technological University. He is also the General Manager of Air T&D, Singapore. He has published more than 100 international journal/conference papers and holds more than ten patents. He is generally interested in power electronics, power systems, advanced control theory, and smart green building. He has received the highly-prestigious Chinese National Award for Outstanding Students Abroad, in 2016. He is the Associate Editor of the IEEE TIE/JESTPE/ACCESS and *IET Power Electronics*.