IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 28, 2019, accepted February 12, 2019, date of publication February 22, 2019, date of current version March 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900369

Target Detection Method Based on Improved Particle Search and Convolution Neural Network

GUOWEI XU[1,2], XUEMIAO SU[1], WEI LIU[3], AND CHUNBO XIU[1,2]

[1]School of Electrical Engineering and Automation, Tianjin Polytechnic University, Tianjin 300387, China
[2]Key Laboratory of Advanced Electrical Engineering and Energy Technology, Tianjin Polytechnic University, Tianjin 300387, China
[3]School of Mechanical Engineering, Tianjin Polytechnic University, Tianjin 300387, China

Corresponding author: Guowei Xu (xuguowei@tpju.edu.cn)

This work was supported in part by the Natural Science Foundation of Tianjin China under Grant 18JCYBJC88300, Grant 18JCYBJC88400, Grant 17JCYBJC19400, and Grant 17JCYBJC18500.

ABSTRACT The border regression is a key technique of the regional convolution neural network (CNN) to locate the target. However, it relies on the border label information of a large number of sample data. Therefore, it is inefficient to generate the training sample set, and the location of the target is also inaccurate. For this, a novel target detection method based on the CNN and the particle search is proposed. A small number of probe particles are generated to roughly locate the target. The CNN is used to extract the image features, determine the target probability, and recognize the pattern of the target. A large number of searching particles are placed near the region where the target features are detected by the probe particles. The nearest neighbor clustering algorithm is used to classify the particles, which are recognized as the same category into different target sets. The positions of the targets can be determined by the bounding rectangle of the searching particles in the same target set. The method can be used to recognize and locate various kinds of targets. Furthermore, the method need not label the borders of the targets in the training samples, which enhance the generation efficiency of the samples. The simulation results show that the correctness of the recognition can be slightly improved, and the accuracy of the location can be significantly improved.

INDEX TERMS Particle filter, target detection, convolution neural network, target location.

I. INTRODUCTION

It is important to recognize or locate the image target from the complex background in many application fields, such as military, intelligent security, visual tracking, biomedicine, and industrial production [1], [2]. It has become a key technical issue to study on the automatic target recognition and precise location in the image processing based on the massive images. The task of the image target recognition can be decomposed into two subtasks: the target classification and the target location. The target classification is mainly used to determine whether there is a target in the image and classify the detected target. And the target location is used to determine the exact position of the detected target in the image [3]–[5]. In the process of the target detection, much matching calculation are needed, which increases the computation cost.

The particle filter shows good application in the target tracking fields [6], [7]. Making use of importance sampling techniques, the particle filter could sequentially approximate the posterior probability density function with any dynamic and measurement models [6]–[9]. As the number of particles increases, the particle filter could approach to the posterior probability density function more accurately. Therefore, the particle filter has been used widely to locate the target in the scene images [10], [11].

Although the study on the conventional target recognition and location methods has got great progress, complex operations, such as feature extraction, feature selection, and so on, restrict the recognition and location performances. Recently, the development of neural network provides the new reliable technological guarantee for the target recognition and location [12]–[14]. Especially, the deep learning network provides a new research idea for target recognition and location, and it is becoming increasingly attractive. For example, Convolution Neural Network (CNN) is a typical deep learning network. It can recognize and classify the targets in the images. VGGNet [15] is an efficient convolution neural network with 16 layers, consisting of five groups of small

The associate editor coordinating the review of this manuscript and approving it for publication was Choon Ki Ahn.

convolution filters with 3×3 size. It could show good recognition performances in the "ImageNet Large Scale Visual Recognition Challenge" (ILSVRC) [16], [17]. However, in a practical application, the proportion of the target in the image would affect the classification result. Generally speaking, the smaller the target is, the more difficult the correct classification is. In order to solve the classification problem of the targets, a regional convolution neural network (RCNN) [18] is proposed based on CNN. The principle of the method is to generate 1k$\sim$2k candidate regions for the detected image, and the deep network is used to extract the features from each candidate region. According to the features, a classifier is used to determine whether the candidate region is the detected category. When the detected category is determined, bounding box regression is performed to amend the position and the size of the candidate region. In this way, the target is located effectively. Although the RCNN can be used to perform the multi-targets detection, the computational cost is large. When the number of RCNN candidate regions is large, many candidate regions would be partially overlapped, and the features would be extracted from overlapping area. Thus, a mass of redundant computation would lead to the low calculation efficiency.

For this, based on the RCNN, improved new recognition methods, FAST-RCNN [19] and FASTER-RCNN [20], have been proposed. FAST-RCNN is constructed by adding the pyramid pool layer into the RCNN [21], [22]. In the RCNN, each candidate region must be scaled to the default size, and the features must be extracted separately from each candidate region, so it has a larger calculation cost. In FAST-RCNN, the features in the whole image could be extracted by the CNN, so the calculation amount is reduced. However, the selective search is still used to generate the candidate regions in FAST-RCNN, so its computational speed is still slower. Therefore, based on FAST-RCNN [19], FASTER-RCNN [20] is proposed by introducing the Region Proposal Network to generate the candidate regions. FASTER-RCNN can perform the image processing with faster computation speed [20], [23]. However, FASTER-RCNN has a high false positive rate. Furthermore, All of the methods above are two-stage methods. Generally, two-stage methods have more computational cost than one-stage methods.

For this, YOLO detection method which is a one-stage method is proposed [24]. YOLO is a recognition network which can detect the target based on end-to-end. It can once predict a plurality of border positions and categories, and the false detection rate of the background can be effectively reduced [25], [26].

All the methods above which are based on RCNN use border regression principle to detect and locate the target. According to the border regression principle, each target in the training sample images must be classified and labeled the exact position of the border before training. This process makes it difficult to generate the training samples. Furthermore, the border regression results of the each candidate region maybe are not satisfied. The accuracy of the region regression is related to the many factors, such as background interference, the size and the shape of the target. Therefore, the regression results sometimes may be unstable.

In order to enhance the accuracy of target location, an improved CNN method based on the particle search is proposed to detect and locate the target. VGGNet is used to train the samples and extract the features, and the particle search is used to detect the target. The position of the target is determined according to the detection results of the particles. The difference between this method and the conventional target location methods based on RCNN is that the exact position of the target border in the sample set need not be labeled. Thus, tedious labeling task could be omitted. It is easier to generate the training samples. It not only increases the efficiency of target detection, but also improves the accuracy of target location.

II. PARTICLE FILTER

The standard particle filter algorithm is also called the serialized Monte Carlo method. Its main idea is that the prior probability density function of the target is represented by a group of particles with different weights, the likelihood of each particle is calculated, the approximate particle set which can describe the posterior density function of the estimated state can be got by fusing the prior data and the new likelihood data of each particle [6], [7].

Set the target state to be x_{t-1}, and the posterior probability density $p(x_t|z_t)$ can be used to estimate its subsequent state x_t. The posterior probability density $p(x_t|z_t)$ can be described by the weighted posterior sample set:

$$p(x_t|z_t) \approx \sum_{i=1}^{N} \overline{\omega}_t^{(i)} \delta(x_t - x_t^{(i)}) \tag{1}$$

where, N is the number of particles, $x_t^{(i)}$ is the target state, δ is the Kroneck function, and $\overline{\omega}_t^{(i)}$ is the normalized weight:

$$\overline{\omega}_t^{(i)} = \omega_t^{(i)} \Big/ \sum_{j=1}^{N} \omega_t^{(j)} \tag{2}$$

where, $\omega_t^{(i)}$ is the particle weight before normalization:

$$\omega_t^{(i)} \propto p(z_t|x_t^{(i)}) \tag{3}$$

Thus, the state estimate $\hat{x}_t$ at time t can be described as:

$$\hat{x}_t = E\left[x_t|z_t\right] \approx \sum_{i=1}^{N} \overline{\omega}_t^{(i)} x_t^{(i)} \tag{4}$$

Particle filter is often applied in the target tracking field. For the target detection, combining the convolution neural network particle search is used to enhance the detection efficiency.

VGGNet is used to extract the target information as the features. According to the similarity between the features and the information in the particle region, the particles are updated and resampled. That is, the particles with smaller weights would be discarded, and more particles would be placed near the regions which contain the target features. The target position would be located by searching for the external border of the particles.

III. THE STRUCTURE OF VGGNet

CNN is used to extract the features from different categories of samples in the database. CNN has many forms. Different networks are suitable for different tasks. Such as, the LeNet5 [27] has obtained great success in the field of handwriting identification. AlexNet [28] which has the better the image classification ability could be constructed by modifying the LeNet5. Based on the AlexNet, VGGNet could be constructed by using a smaller convolution kernel and further increasing the layer number. And VGGNet has the faster convergence speed and the better classification accuracy [29]–[31]. Therefore, VGGNet is chosen to recognize and locate the target in this paper.

The parameters of VGGNet are shown in Table 1.

TABLE 1. Parameters of VGGNet.

Layer	operator	number	size	Layer	operator	number	size
1	conv	64	1×1	12	conv	512	1×1
2	conv	64	1×1	13	conv	512	1×1
3	pool	64	2×2	14	pool	512	2×2
4	conv	128	1×1	15	conv	512	1×1
5	conv	128	1×1	16	conv	512	1×1
6	pool	128	2×2	17	conv	512	1×1
7	conv	256	1×1	18	pool	512	2×2
8	conv	256	1×1	19	fc-4096		
9	conv	256	1×1	20	fc-4096		
10	pool	256	2×2	21	fc-C		
11	conv	512	1×1	22	Softmax		

The network model is composed of 13 convolution layers, 5 pooling layers and 3 fully connected layers. The size of the convolution kernel is [3,3], the step size of the convolution operator is 1×1, the size of the pooling kernel is [2,2], and the step size of pooling operator is 2×2.

When the image is input into the VGGNet, the features would be extracted by multiple convolution and pooling layers. And the fully connected layers are used to perform the image classification. There is a loss function Softmax which is used to calculate the recognition probability on the last layer. The Softmax function is defined as:

$$S_i = \frac{e^{V_i}}{\sum_i^C e^{V_i}} \tag{5}$$

where, V_i is the i-th input of the Softmax function, C is the number of the classes, and S_i is the recognition probability of the input image for the i-th class. Nontargets are regarded as the C-th class. That is, the first $(C-1)$-th classes are the target classes, and the C-th class is the nontarget class. The input image would be recognized as the class with the maximal recognition probability. If the input image is recognized as the C-th class, it means that there is not any target in the image. Thus, VGGNet could perform the image recognition for the multiple classes.

IV. TARGET LOCATION BASED ON PARTICLE SEARCH

Target recognition and location from the given image can be performed by combining VGGNet and the improved particle search. VGGNet plays the role of the feature extraction and target classification. And the particle search plays the role of locating the target. In this way, the accurate position of the target could be determined. Furthermore, spatial pyramid pooling layer is added into VGGNet to improve the scale-invariance, restrict the over-fitting and save the computation cost. The method can be described as:

Step1: VGGNet is trained by the training samples in the database to extract the features which would be used to recognize and classify the target.

Step2: The image is evenly divided into $s \times s$ regions. Each region is recognized by CNN, and the probability for each class can be got as: $\{S_1, S_2, \ldots, S_C\}$. The maximum probability is determined as the target recognition probability of the region. That is, suppose $S_l = \max\{S_1, S_2, \ldots, S_C\}$. If $l \neq C$, it means that the image contains the target features of the l-th class with the probability S_l. Otherwise, i.e., $l = C$, it means that no target is detected.

Step3: Initialize N probe particles as $\left\{x_0^i, \varpi_0^{(i)}\right\}_{i=1}^N$, and the initial probe particles are randomly distributed in the regions which contain targets. The normalized weight of the i-th particle is set as $\varpi_0^{(i)} = 1/N$. The bigger size of the particle region is chosen as $M \times M$.

Step4: According to the recognition result of the image in the particle region obtained by CNN, $p\left(z_1|x_1^{(j)}\right)$ could be calculated as:

$$p\left(z_1|x_1^{(j)}\right) = \begin{cases} S_l, & l \neq C \\ 0, & l = C \end{cases} \tag{6}$$

The recursion particle weight could be calculated as:

$$\omega_1^{(j)} = \omega_0^{(j)} p\left(z_1|x_1^{(j)}\right) \tag{7}$$

The normalized weight could be calculated as:

$$\varpi_1^{(j)} = \frac{\omega_1^{(j)}}{\sum_{i=1}^{N} \omega_1^{(i)}} \tag{8}$$

Step5: Discard the low-weight particles, and generate N' new search particles near the high-weight particles according to the weight probability, where $N' > N$, and the smaller size of the search particles is chosen as $m \times m$, that is, $m < M$. Label the new search particles according to the recognization results. The nearest neighbor clustering algorithm is used to classify the particles with the same labels into different target sets. The external border of the particles in the same target set could be determined as the position of one target, and the labels of the particles indicate the category of the target.

The operation of particle generation is similar to the choosing candidate region operation in RCNN. The given image is evenly divided several regions, which could ensure that the target could not be missed. The image features of the region could be extracted by VGGNet to determine whether the region contains a target. Particles are randomly generated on the given image according to the target probabilities, and the weights of the particles could be calculated according to the recognition results. In this way, the target location could be quickly performed by the particle search.

The improved method proposed in this paper uses the less initial detection particles with the bigger sizes. Thus, the method could quickly complete the coarse location of the target. In the resampling process, the number of searching particles is increased and the size of the searching particles becomes smaller. Therefore, the fine positioning of the target could also be determined with the high accuracy.

Furthermore, the cluster analysis of the search particles with the same labels is done by the nearest neighbor clustering algorithm. That is, if there are multiple targets in the image, the nearest neighbor clustering algorithm could classify the search particle into multiple target sets. Each target could be located according to the external border of particles only in the same target set. Therefore, the method could manage multiple objects in the same category.

Considered from the computational cost, both the proposed method and RCNN are two-stage methods, and the proposed method uses two-level particle search to detect the targets, that is, the targets could be detected from an image by two iterations. So it needs less computation cost than RCNN. However, YOLO is one-stage method, so it has less computation cost than two stage methods. That is, the proposed method needs more computation cost than YOLO. However, YOLO needs to label the borders of the targets for the training samples, which needs huge labor costs to make the training samples. Meanwhile, it has low efficiency when the recognized task is changed. In contrast, the proposed method need not label the borders of the targets for the training samples. Therefore, it could save more labor costs than YOLO.

V. SIMULATION EXPERIMENTS

A. COMPARISON EXPERIMENTS

A training sample set which contains 5 kinds of targets is made, and each class has 800 samples. The proposed method and the YOLO are separately used to recognize and locate the targets.

The location results of the single target in the simple background obtained by the two methods above are shown in Figs. 1 and 2.

In Figs. 1 and 2, the plane is the detected target. Fig. 1 shows the location results obtained by YOLO method, and Fig. 2 shows the location process of the proposed method.

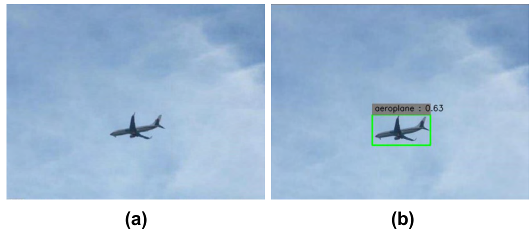


FIGURE 1. Detection results of YOLO detection algorithm. (a) Original image. (b) Location results.

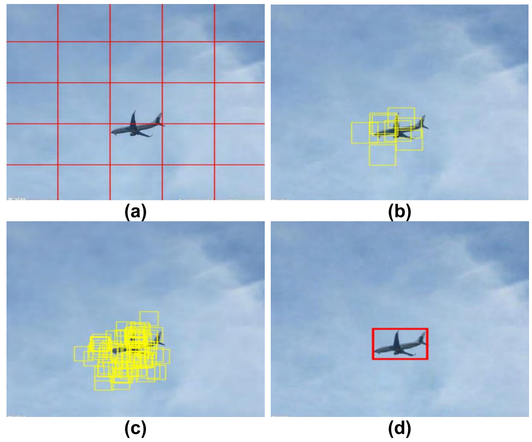


FIGURE 2. Detection results of the improved method. (a) Regional division. (b) Probe particles. (c) Searching particles. (d) Location results.

The regional division is shown in Fig.2(a), and VGGNet could be used to find the regions which contain the target. The probe particles in Fig.2(b) are generated to complete the coarse location of the target. And the searching particles in Fig. 2(c) are generated to complete the fine location of the target. Finally, Fig. 2(d) gives the location result of the target obtained by searching for the external border of the particles with bigger weights.

Because the background is simple, both the two methods could give the accurate location results. But the method proposed in this paper need not label the training samples. Therefore, it is convenient and efficient to use the method to locate the target in the actual application.

FIGURE 3. Detection results of YOLO detection algorithm. (a) Original image. (b) Location results.

The location results of the single target in the complex background are shown in Figs. 3 and 4.

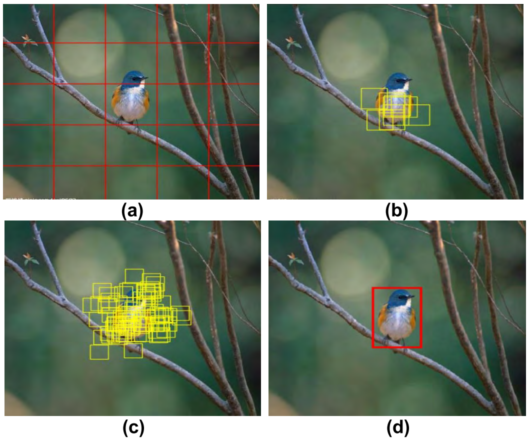


FIGURE 4. Detection results of the improved method. (a) Regional division. (b) Probe particles. (c) Searching particles. (d) Location results.

The bird in Figs. 3 and 4 is the detected target. The location results obtained by YOLO method are shown in Fig. 3. Although the target could also be located, some background has been divided into the target area. Thus, the position of the target is not accurate enough. That is, in the complex background, the size of the target area located by the YOLO is bigger than the actual size, which causes to the inaccurate location results. The location process obtained by the proposed method is shown in Fig. 4. The target could be accurately located by the probe particles and the searching particles. Comparatively, it can be seen that the target location method combining particle search and VGGNet could gives more accurate location result.

The detection of multiple targets in a simple background are done by the methods above, and the experimental results are separately shown in Figs. 5 and 6.

FIGURE 5. Detection results of YOLO detection algorithm. (a) Original image. (b) Location results.

Both the cat and the dog in Figs. 5 and 6 are the detected targets. The process of multiple targets detection is similar to that of single target detection. When VGGNet recognizes the target features, the resampling process would be redone near the regions which contain the target features to perform the fine location of the target. From the experiment results, the improved method could correctly recognize the two targets and locate their accurate positions. However, YOLO method mistakenly recognizes the two adjacent targets as one target, which leads to the location failure.

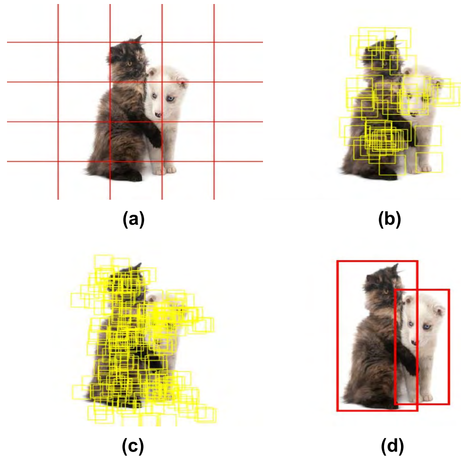


FIGURE 6. Detection results of the improved method. (a) Regional division. (b) Probe particles. (c) Searching particles. (d) Location results.

The detection results of multiple targets in the complex background obtained by the two methods above are shown in Figs. 7 and 8.

FIGURE 7. Detection results of YOLO detection algorithm. (a) Original image. (b) Location results.

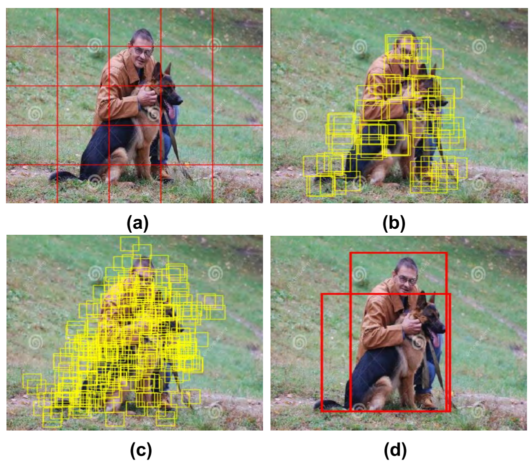


FIGURE 8. Detection results of the improved method. (a) Regional division. (b) Probe particles. (c) Searching particles. (d) Location results.

In Figs. 7 and 8, both the man and the dog in the complex background are the targets to be detected. From the experimental results, it can be seen that the improved method uses the rough probe particles and the fine searching particles to correctly locate the two targets in the complex background. Comparatively, although the YOLO method could also recognize the two targets, the position of the dog is obviously devious. Thus, the proposed method has the better location precision than YOLO.

The target recognition and location experiments for five classes of targets are done by the methods above. The correct rate c, the location accuracy rate p and the overlapping rate d are used to evaluate the performances of the two methods. The correct rate c is defined as the proportion of the targets recognized correctly.

$$c = \frac{N}{S} \tag{9}$$

where, N is the number of the targets recognized correctly, and S is the total number of the targets.

The location accuracy rate p is defined as the ratio of the area difference between the detection target area l and the actual target area s to the actual target area s when the detection target area could completely cover the actual target area. That is,

$$p = \frac{l - s}{s} \tag{10}$$

Furthermore, the overlapping rate d is determined as:

$$d = \frac{s - h}{s} \tag{11}$$

where, h is the overlapping area of the detection target area and the actual target area.

The statistical results of the correct rate c, the location accuracy rate p and the overlapping rate d are shown in Table 2.

TABLE 2. Experimental comparison results.

Target		Target classes				
		Plane	Bird	Cat	Dog	Human
c	This method	92.3%	88.7%	86.3%	87.8%	90.5%
	YOLO	90.1%	86.5%	84.2%	84.6%	88.7%
p	This method	17.6%	18.1%	17.3%	19.6%	18.4%
	YOLO	31.5%	47.2%	45.5%	41.7%	38.8%
d	This method	98.7%	96.5%	96.1%	96.4%	94.3%
	YOLO	97.2%	96.1%	94.8%	90.4%	93.5%

From Table 2, the correct rate of the method in this paper is slightly higher than that of YOLO method. However, it is well to be reminded that the location accuracy rates of various targets of the method in this paper are obviously higher than those of YOLO method. The average location accuracy rate of each category is enhanced more than 10%. The method has the better overlapping rate, which can also prove the validity of the method.

The single frame average computational time of the method is about 80ms. The particle filter could supply the stronger heuristic information to reduce the number of the candidate regions, so this method could have a faster computational speed than RCNN which needs about 150ms for an image. Of course, YOLO which is a one-stage method needs about only 50ms for an image. Though the computation cost of the proposed method is larger than that of YOLO, the main advantage of this method is that the border labeling of the target in the training sample images is omitted. Therefore, the generation efficiency of the training samples is significantly improved.

B. TARGET DETECTION BASED ON ImageNet

ImageNet is a large database. Five kinds of objects, potted plant, boat, cattle, cup and orange, are chosen as the recognition targets. The detection results of the improved method are shown in Fig. 9.

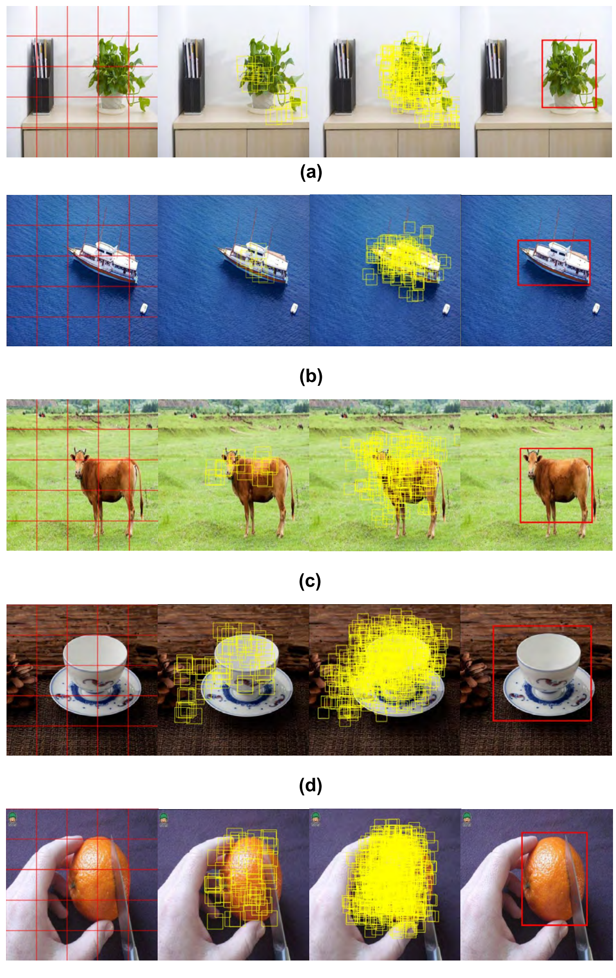


FIGURE 9. Detection results based on the CNN and particle filter. (a) The detection process of the potted plant. (b) The detection process of the boat. (c) The detection process of the cattle. (d) The detection process of the cup. (e) The detection process of the orange.

From the detection results shown in Fig. 9, the particle search could generate some candidate regions, and VGGNet could recognize targets in the regions and give the target probabilities which guide the search direction of the particles. In this way, the method has the better target detection performance. The detection performances based on ImageNet are shown in Table 3.

From Table 3, the recognition correct rate could be further enhanced by fine tuning VGGNet based on ImageNet. That is,

TABLE 3. Detection performance.

Target	Potted plant	Boat	Cattle	Cup	Orange
c	92.4%	90.6%	93.4%	92.1%	95.8%
p	16.3%	15.9%	18.4%	14.7%	11.5%
d	96.5%	95.1%	95.6%	96.7%	97.0%

the large data sets help to improve the recognition correctness. And the detection performances of the improved method would also be improved.

VI. CONCLUSION

Aiming at the problem that the bounding box regression used in the process of the target detection has low efficiency to generate the training samples and inaccurate location results, a new detection method combining CNN and the particle search is proposed. VGGNet is used to extract features and classify the targets. The particle search is used to complete the coarse and the fine location of the target. The location accuracy of the target could be improved obviously. The experimental results show that the correct recognition rate of the target could be slightly enhanced, and the average location accuracy rate could be significantly enhanced more than 10% for each category. And above all, this method can omit the border labeled operation of the targets in the training sample images, so the training samples are simplified and the generation efficiency of the training samples is significantly enhanced.

REFERENCES

[1] Y. M. Zhang, B. Du, Y. Zhang, and L. Zhang, "Spatially adaptive sparse representation for target detection in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1923–1927, Nov. 2017.

[2] M. Tong, Y. Pan, Z. Li, and W. Lin, "Valid data based normalized cross-correlation (VDNCC) for topography identification," *Neurocomputing*, vol. 308, pp. 184–193, Apr. 2018.

[3] D. Chimura, R. Toh, and S. Motooka, "Spectrum compensation for time-reversal method on ultrasonic target detection using pulse compression," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 64, no. 12, pp. 1874–1883, Dec. 2017.

[4] X. Wang, Z. Peng, P. Zhang, and Y. He, "Infrared small target detection via nonnegativity-constrained variational mode decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1700–1704, Oct. 2017.

[5] Y. Pang, M. Sun, X. Jiang, and X. Li, "Convolution in convolution for network in network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 9, no. 5, pp. 1587–1597, Feb. 2018.

[6] J. Carpenter, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," *IEEE Proc.-Redar, Sonar Navigat.*, vol. 146, no. 1, pp. 2–7, Feb. 1999.

[7] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc. F (Radar Signal Process.)*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[8] J. M. Hammersley and K. W. Morton, "Poor man's Monte Carlo," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 16, no. 1, pp. 23–38, Jan. 1954.

[9] J. E. Handschin, "Monte Carlo techniques for prediction and filtering of non-linear stochastic processes," *Automatica*, vol. 6, no. 4, pp. 555–563, Jul. 1970.

[10] B. Ristic, J. Houssineau, and S. Arulampalam, "Robust target motion analysis using the possibility particle filter," *IET Radar, Sonar Navigat.*, vol. 13, no. 1, pp. 18–22, Jan. 2019.

[11] Z. Musa, M. Z. Salleh, R. A. Bakar, and J. Watada, "GbLN-PSO and model-based particle filter approach for tracking human movements in large view cases," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 8, pp. 1433–1446, Aug. 2016.

[12] C.-M. Lin, C.-Y. Tsai, Y.-C. Lai, S.-A. Li, and C.-C. Wong, "Visual object recognition and pose estimation based on a deep semantic segmentation network," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9370–9381, Nov. 2018.

[13] Z. X. Zhang, H. J. Liang, C. W. Wu, and C. K. Ahn, "Adaptive event-triggered output feedback fuzzy control for nonlinear networked systems with packet dropouts and actuator failure," *IEEE Trans. Fuzzy Syst.*, to be published. doi: 10.1109/.TFUZZ.2019.2891236.

[14] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 690–703, Mar. 2017.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–14.

[16] X. Liu, M. M. Chi, Y. F. Zhang, and Y. Q. Qin, "Classifying high resolution remote sensing images by fine-tuned VGG deep networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Valencia, Spain, Jul. 2018, pp. 7137–7140.

[17] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2019, pp. 248–255.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 580–587.

[19] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 1440–1448.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2015.

[21] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 985–996, Apr. 2018.

[22] X. Li *et al.*, "A unified framework for concurrent pedestrian and cyclist detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 269–281, Feb. 2017.

[23] C. Qi, C. Shi, C. Wang, and B. Xiao, "Logo retrieval using logo proposals and adaptive weighted pooling," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 442–445, Apr. 2017.

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.

[25] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 507–517, Feb. 2018.

[26] B. Walker, "Comparison of the birds point-new Madrid floodway, Mississippi river and the Yolo bypass, Sacramento river," *J. Earth Sci.*, vol. 27, no. 1, pp. 47–54, Feb. 2016.

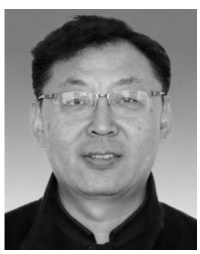
[27] F. Lauer, C. Y. Suen, and G. Bloch, "A trainable feature extractor for handwritten digit recognition," *Pattern Recognit.*, vol. 40, no. 6, pp. 1816–1824, 2007.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 1–9, Oct. 2017.

[29] I. Ha, H. Kim, S. Park, and H. Kim, "Image retrieval using BIM and features from pretrained VGG network for indoor localization," *Building Environ.*, vol. 140, pp. 23–31, Sep. 2018.

[30] Q. Dong, H. Wang, and Z. Hu, "Statistics of visual responses to image object stimuli from primate AIT neurons to DNN neurons," *Neural Comput.*, vol. 30, no. 2, pp. 447–476, Feb. 2018.

[31] J. Wang, J. Lin, and Z. Wang, "Efficient hardware architectures for deep convolutional neural network," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 6, pp. 1941–1953, Jun. 2018.



GUOWEI XU received the M.Sc. degree in motor and electrical professional from the Shenyang University of Technology, Shenyang, China, in 2000, and the Ph.D. degree in textile engineering from Tianjin Polytechnic University, Tianjin, China, in 2015. His research interests include sliding mode control, neural networks, and artificial intelligence.

XUEMIAO SU received the B.S. degree in control science and engineering from Jiamusi University, Jiamusi, China, in 2016. Her research interests include neural networks and image processing.

WEI LIU received the M.Sc. degree in computer application technology from Tianjin Polytechnic University, Tianjin, China, in 2010. Her research interests include detection technology, intelligent control, and computer applications of control methodologies.

CHUNBO XIU received the Ph.D. degree in navigation, guidance and control from the Beijing Institute of Technology, Beijing, China, in 2005. His research interests include neural networks, system modeling, and chaos control.