International Journal of Science and Engineering Applications
Volume 13-Issue 04, 17 - 21, 2024, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1304.1004

Identification of Small Goods under the Adversarial Network

Feng Liu
School of Electric Information and Electrical Engineering Yangtze University
Jingzhou, China

Shunli Li
School of Electric Information and Electrical Engineering Yangtze University
Jingzhou, China

Abstract: Small targets are easily lost and misjudged in detection tasks due to their relatively low resolution in images. Aiming at the problems of large recognition error and weak generalisation ability of small goods in the commodity recognition task when facing complex conditions such as hand occlusion and poor lighting, this paper proposes an adaptive small goods recognition method based on adversarial networks. Firstly, an end-to-end adversarial process-adapted network model is designed using the adversarial and recognition process a priori, and the super-resolution network is used as a generator for reconstructing the semantic information of small commodities, and the recognition network is used as a discriminator for classification and authenticity judgement of the input image. In addition, a new target feature reconstruction function is designed for reconstructing discriminative feature information. In order to solve the problem of large recognition error caused by the difference in the appearance of goods, a feature map-based attention mechanism is proposed for enhancing the sensory field of the recognition network, so that the recognition of large and small goods achieves more accurate results. Tested on SVHN dataset, CIFAR10 dataset and homemade small goods dataset, the adversarial structure improves the recognition accuracy by 4.1% compared with the cascade structure. Meanwhile, ablation experiments are designed to verify the effectiveness of the target feature reconstruction function and the feature graph attention mechanism for feature reconstruction and small goods recognition, respectively. The experimental results show that the method improves the accuracy of small goods recognition and is robust to multi-scale goods recognition and partial occlusion.

Keywords: Small commodity identification; Countermeasures network; Super resolution; Multi scale recognition; Feature map attention mechanism

1. INTRODUCTION

In recent years, with the rapid development of artificial intelligence, target recognition, as a hot research technology in computer vision, has been widely used within the scope of many fields[1-3]. In the retail industry, intelligent retail containers have the advantages of low operating cost and high flexibility through the combination of unmanned intelligent service and mobile payment, and intelligent shopping gradually replaces the traditional retail method. However, affected by external environmental conditions such as occlusion and light, the small goods recognition technology cannot meet the consumer's application needs, so the small target recognition in intelligent retailing shows great research prospects.

Many scholars at home and abroad have conducted in-depth research on small target recognition task, until now a variety of small target recognition algorithms have been proposed, which can be summarised into the following three categories by collation.

Multi-scale recognition mechanism based on feature fusion[4-10]. It fuses the advantages of deep feature maps and shallow feature maps in order to achieve multi-scale recognition detection of small targets. Literature[11] builds on YOLOv3 by establishing feature fusion algorithms to alleviate the problem of insufficient features after multiple convolution of small targets.

One class is based on the combination of contextual semantic reasoning and visual information[13-15]. It enriches the feature representation of an image by fusing contextual semantic information to improve the stability and accuracy of target recognition. Literature[16] is based on the SSD structure by introducing a multilevel feature fusion approach with contextual information in order to improve the accuracy of recognition of small targets.

Another category is the algorithm that fuses the attention mechanism[18]. It improves the attention of the feature extraction network to small targets by selecting critical feature information. Literature[19] proposes a spatio-temporal neural network (called STNet) for small target recognition, where STNet fixes the region of interest using a super-resolution module and focuses on the distinguished region using a spatio-temporal attention module.

2. METHOD

As shown in Figure 1, there are three main parts of the research, including super-resolution image enhancement, image degradation, and image recognition. The specific research process is as follows:

Firstly, the small commodity image is fed into the super-resolution network, and more recognition-related feature information is recovered through up-sampling.1) In order to reconstruct the semantic features of the occluded part, the feature maps of the current layer of the generated image and the corresponding feature maps of the HR image are compared to obtain the content loss L_{cnt} .At the same time, in this paper, we use the probabilistic degradation model to model the stochastic factors in the degradation process, and better decouple the degradation from the image content decoupling, thus making the synthesis closer to the training samples in the test image domain; 2) The recognition module consists of two branches: the discriminator branch and the classification branch. Among them, the discriminator branch maps the output to [0,1] via Sigmoid to determine whether the samples are from true high-resolution images(HR); the classification branch maps the output to n-dimensions via

Softmax function to obtain the labels of the input samples and finally feeds the task goal to the super-resolution network. In addition, the recognition mechanism based on feature attention, FMAResNet, adaptively adjusts the size of the receptive field according to the input information, which improves the network's ability to express the model features and reduces the problem of large recognition errors.

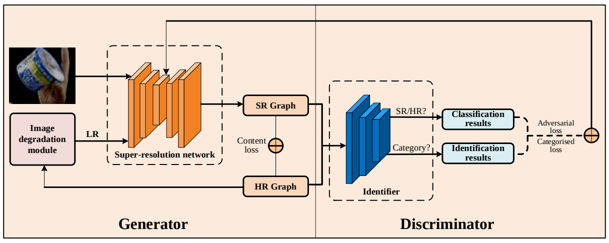

Figure.1 Overview of our method

3. ALGORITHM DESIGN

3.1 Super Resolution Module

The module mainly consists of a content extractor and a texture extractor. Firstly, the content extractor is used to extract the main semantic information of the input feature map P_2, and then the resolution of the content features is doubled by sub-pixel convolution. The texture extractor selects reliable texture regions from main and reference for small item recognition. Finally, the texture is fused with the super-resolution content features using residual concatenation, resulting in the output P_3:

$$P_3 = E_t(P_1 \square E_c(P_2)\uparrow_{2\times}) + E_c(P_2)\uparrow_{2\times} \tag{1}$$

Where $E_t(.)$ denotes the texture information extractor, $E_c(.)$ denotes the content information extractor, □ denotes that the extracted features are concatenated and then zoomed in two times.

3.2 Identify Network Modules

The image features extracted by convolutional neural networks are generally rich and relatively simple to compute. The commonly used feature extraction networks are ResNet50, which solves the phenomenon of disappearing or exploding network gradients brought about by the increase of network depth by taking the input features plus the learned residual features as the output features. However, the extracted features are still limited due to the spatial invariance and locality of the standard convolutional kernel, and the learned parameters are static. Based on this, this paper designs FMAResNet, a recognition mechanism based on feature map attention, on the basis of ResNet50 as a feature extraction network, and adopts the method of fusing the feature map attention mechanism to reduce the problem of large recognition error caused by the size of goods under complex conditions.

The specific steps are as follows:

(1) Decomposition. Given an input feature $X \in R^{H\times W\times C}$, set the convolution kernel sizes to 3×3 and 5×5 respectively, where the 5×5 convolution kernel is replaced by a 3×3 convolution kernel with dilation 2. Let it perform the convolution operation to get two outputs $F_1 : X \to U_1 \in R^{H\times W\times C}$ and $F_2 : X \to U_2 \in R^{H\times W\times C}$ respectively.

(2) Fusion. The outputs of different branches are fused using an element-by-element summation method to obtain:

$$U = U_1 + U_2 \tag{2}$$

Secondly, the global pooling operation is used on the consolidated information to get the global information, i.e.

$$s_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} U_c(i,j) \tag{3}$$

Where F_{gp} denotes the global average pooling operation function, s_c denotes the output of the c th channel, $U_c(i,j)$ denotes the coordinates of the c th channel, H is the height of the feature map, W is the width of the feature map, and i,j denote the values of the coordinates of the height and width of the feature map, respectively.

Finally, z is obtained by performing a dimensionality reduction operation on s_c through the fully connected layer with the following formula:

$$z = F_{fc}(s) = \delta(\beta(W_s)) \tag{4}$$

$$d = max(\frac{c}{r}, L) \tag{5}$$

where F_{fc} denotes the fully connected operator function, δ denotes the nonlinear activation function, β is the BN layer, d denotes the control of the fully connected layer with the reduction ratio r, and L is the minimum value of d, where $W \in R^{d\times c}$ and $z \in R^{d\times l}$.

(3) Selection. The attention of the channel is first generated and then used to self-adaptively select information of different sizes, expressed as follows:

$$\begin{cases} a_c = \dfrac{e^{A_c z}}{e^{A_c z} + e^{B_c z}} \\ b_c = \dfrac{e^{B_c z}}{e^{A_c z} + e^{B_c z}} \end{cases} \tag{6}$$

Where $A, B \in R^{c\times d}$, a_c, b_c denote the attention vectors corresponding to U and U, respectively, where A_c denotes the c th row, and a_c denotes the c th element of a.

Finally, the branch output features are weighted and fused to obtain V_c with the following formula:

$$V_c = a_c U_c + b_c U_c ; a_c + b_c = 1 \tag{7}$$

Where $V = [V_1, V_2, \cdots, V_C], V_C \in R^{H\times W}$ 。

3.3 Loss function

In order to reconstruct more texture detail information that is beneficial for recognition, the SR image and the real HR image generated by the generator are fed into the recognition network for feature extraction separately, and then the content loss L_{cnt} is obtained by using the root-mean-square error on the extracted feature maps:

$$L_{cnt} = \frac{1}{w_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j} G_{\theta G}((I^{LR}))_{x,y})^2 \quad (8)$$

Where , i, j denotes the i th convolution after the j th maximum pooling layer; $W_{i,j}$ and $H_{i,j}$ denote the width and height of each feature map within the recognition network, respectively, and ϕ corresponds to the feature map output from a certain convolutional layer in the recognition network after an activation function.

In order to make the generated images closer to the real image distribution, the adversarial loss L_adv is used in this paper:

$$L_{adv} = \sum_{n=1}^{N} log\, D_{\theta_D}(I^{HR}) + \sum_{n=1}^{N} log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (9)$$

where D_{θ_D} is the discriminator, θ_D is the weight of the discriminator, G_{θ_G} is the generator, θ_G is the weight of the generator, and N is the number of samples. $log\, D(x)$ is the probability that the discriminator determines the true data to be true, and $log(1 - D(G(z)))$ is the probability that the discriminator determines the generator-generated the probability of determining the false data as false data, and the total loss is the sum of the two, which should be guaranteed to be as large as possible.

For a given input image, the goal of the method in this paper is to classify it correctly. To achieve this condition the classification loss L_{cls} is defined:

$$L_{cls} = \sum_{n=1}^{N} log\, P(C = c \mid I^{HR}) + \sum_{n=1}^{N} log\, P(C = c \mid G_{\theta_G} I^{LR}) \quad (10)$$

Where $P(C = c \mid I^{HR})$ denotes the probability that the real sample is classified correctly, and $log\, P(C = c \mid G_{\theta_G} I^{LR})$ denotes the probability that the generated sample $G_{\theta_G}(I^{LR})$ is classified correctly.

In order to organically combine the super-resolution task with the recognition task, a new target feature reconstruction function is designed, which combines the content loss, the adversarial loss and the classification loss to guide the super-resolution network to reconstruct the detailed texture information that is favourable for recognition under the constraints of the content loss, and to further make the generated image more closely approximate the distribution of the real image through the adversarial loss and the classification loss. Based on the above derivation, the target feature reconstruction function can be expressed as:

$$L_G = L_{cnt} - \lambda_{adv} L_{adv} + \lambda_{cls} L_{cls} \quad (11)$$

The module is mainly used for target recognition and feeds the task target to the super resolution module, the recognition module loss function can be expressed as:

$$L_G = \lambda_{cls} L_{cls} - \lambda_{adv} L_{adv} \quad (12)$$

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of this paper's algorithm on the recognition accuracy of small goods, this paper sets up a comparison experiment to evaluate the algorithm performance based on the experimental data.

4.1 Environment Configuration

All the comparison experiments were conducted under the same hardware conditions, and the relevant experimental configurations are shown in Table 1.

Table.1 Main configuration module

Operating system	Windows 10
CPU	12th Gen Intel(R) Core(TM) i5-12490F
GPU	NVIDIA GeForce RTX 3060 12G
Python	3.8
Pytorch	1.12.0
CUDA	11.2

4.2 Results

In this section, the method is further compared with the mainstream small target recognition methods in recent years. In the experiments, the recognition and detection of small goods under different degrees of occlusion are carried out respectively, and finally the average value is taken, and the experimental results are obtained as shown in Table 2 below.

Table.2 Comparison between mainstream recognition algorithms and methods in this paper

Algorithm	mAP%
Faster-RCNN	91.51
SSD	83.62
YOLOv3	85.43
Ours	94.67

As can be seen from Table 2, in terms of recognition accuracy, comparing the average recognition accuracy of multiple algorithms, it can be found that this paper's algorithm has the best recognition effect, and this paper's algorithm improves the recognition accuracy by 3.3% compared with Faster-RCNN algorithm, which has a better recognition effect, thus further proving the advancement of this paper's algorithm.

Figure 2 shows the recognition effect, in the experiment, by changing the angle of holding the goods to simulate the various situations that may occur, from the experimental results, it can be seen that, for different degrees of small goods obscured, this paper's algorithm can achieve good recognition effect.

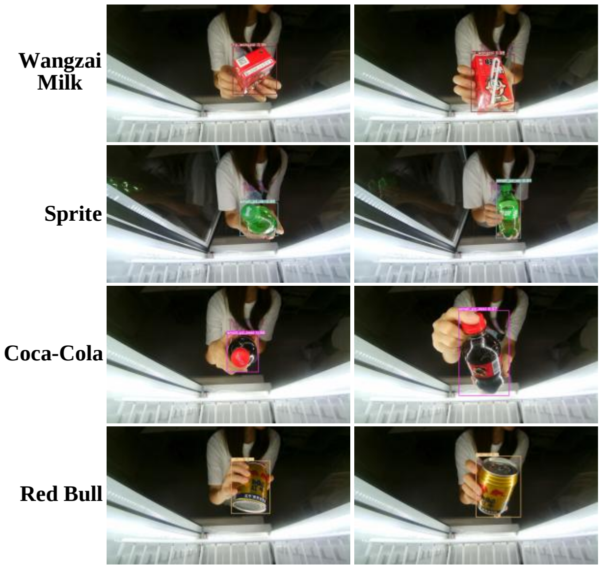

Figure.2 Identification result

5. CONCLUSION

In this paper, an adaptive small goods recognition method under adversarial process is proposed to improve the prediction accuracy in complex environments such as partial occlusion and multi-scale recognition. The interference due to environmental factors such as hand occlusion is reduced by the constraint of the reconstruction loss of target features, and at the same time, an adaptive dynamic selection mechanism is added on the basis of ResNet50 for optimisation, and the problem of large recognition error caused by the size of the goods is solved by adjusting the size of its receptive field, and the multi-scale recognition of goods is achieved. In this paper, the effectiveness of this paper's method in the task of recognition of occluded small commodities is proved by performing ablation comparison experiments on public datasets and self-acquisition datasets. In future work, on one hand, we hope to improve the robustness of this paper's algorithm by optimising the network parameters, and on the other hand, we hope to design a small target recognition method with better performance and apply it to small target recognition in more complex scenes.

6. REFERENCES

[1] Yin H F. Image micro-target recognition method based on multi-layer empirical mode decomposition algorithm[C]//2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE, 2022: 380-383.

[2] Mao R. Real-time Small-size Pixel Target Perception Algorithm Based on Embedded System for Smart City[C]//2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS). IEEE, 2021: 505-511.

[3] Wang Y, Cui W, Yang H. A small target recognition algorithm based on improved SSD[C]//2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). IEEE, 2019: 234-237.

[4] Peng H, Li X. Multi-scale selection pyramid networks for small-sample target detection algorithms[C]//2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). IEEE, 2021: 359-364.

[5] Hu J, Chen Z, Yang M, et al. A multiscale fusion convolutional neural network for plant leaf recognition[J]. IEEE Signal Processing Letters, 2018, 25(6): 853-857.

[6] Ye K, Fang Z, Huang X, et al. Research on small target detection algorithm based on improved yolov3[C]//2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE). IEEE, 2020: 1467-1470.

[7] Chen Y, Wang J, Dong Z, et al. An Attention Based YOLOv5 Network for Small Traffic Sign Recognition[C]//2022 IEEE 31st International Symposium on Industrial Electronics (ISIE). IEEE, 2022: 1158-1164.

[8] De Langlard M, Al Saddik H, Lamadie F, et al. A multiscale method for shape recognition of overlapping elliptical particles[C]//2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016: 692-697.

[9] Xu X, Wang W, Liu Q. Medical Image Character Recognition Based on Multi-scale Neural Convolutional Network[C]//2021 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC). IEEE, 2021: 408-412.

[10] WANG Z, GONG W, JIAO Y, et al. Fully Convolutional Network for Recognition of Small Buildings in Aerial Images[C]//2018 Fifth International Workshop on Earth Observation and Remote Sensing Applications (EORSA). IEEE, 2018: 1-5.

[11] Lu H, Chen T, Shi L. Research on Small Target Detection Method of Traffic Signs Improved Based on YOLOv3[C]//2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). IEEE, 2022: 476-481.

[12] Liu Y, Zeng J, Shan S, et al. Multi-channel pose-aware convolution neural networks for multi-view facial expression recognition[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, 2018: 458-465.

[13] Zhu C, Chen F, Ahmed U, et al. Semantic relation reasoning for shot-stable few-shot object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 8782-8791.

[14] Xiang W, Zhang D Q, Yu H, et al. Context-aware single-shot detector[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 1784-1793.

[15] Fang P, Shi Y. Small object detection using context information fusion in faster R-CNN[C]//2018 IEEE 4th International Conference on Computer and Communications (ICCC). IEEE, 2018: 1537-1540.

[16] Cao G, Xie X, Yang W, et al. Feature-fused SSD: Fast detection for small objects[C]//Ninth International Conference on Graphic and Image Processing (ICGIP 2017). SPIE, 2018, 10615: 381-388.

[17] Shu X, Liu R, Xu J. A Semantic Relation Graph Reasoning Network for Object Detection[C]//2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS). IEEE, 2021: 1309-1314.

[18] Sun S, Yin Y, Wang X, et al. Multiple receptive fields and small-object-focusing weakly-supervised segmentation network for fast object detection[J]. arXiv preprint arXiv:1904.12619, 2019.

[19] Liang Z, Liu S, Shi W, et al. Small Object Recognition Using a Spatio-Temporal Neural Network[C]//2021 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2021: 1-6.