Novel Deep Learning Based Small Object Detection Using Hebbian Principle

Avarnita Chauhan[1], Prof. Sapna Chaudhari[2]
[1]Dept of CSE
[2]Professor, Dept of CSE
[1, 2] Shri Ram Group of Institutions Jabalpur, Madhya Pradesh, India.

Abstract- *With the improving of the intelligent driving awareness, object detection as an important part of intelligent driving, has now become a research hotspot in the world. In recent years, convolutional neural network (CNN) has attracted more and more attention in the field of computer vision. CNN has made a series of important breakthroughs in the field of object detection.*

This paper introduces the object detection method based on deep learning. This paper mainly introduces the detection algorithm based on Deep Learning based on Hebbian principle, and analyses the advantages and disadvantages of the detection algorithm. On this basis, the public data sets and evaluation criteria related to small object detection are introduced.

Keywords- Detection of moving objects; tracking of moving objects; behavior understanding, Neural Network, Caffe model, CNN.

I. INTRODUCTION

Compared with other computer vision tasks, the history of small object detection is relatively short. Earlier work on small object detection is mostly about detecting vehicles utilizing hand-engineered features and shallow classifiers in aerial images [1, 2]. Before the prevalent of deep learning, color and shape-based features are also used to address traffic sign detection problems [3]. With the rapid advancement of convolutional neural networks (CNNs) in deep learning, some deep learning-based small object detection methods have sprung up. However, there are relatively few surveys and researches focusing only on small object detection. Most of the state-of-the-art methods are based on existing object detection algorithms with some modifications so as to improve the detection performance of small objects. To the best of our knowledge, Chen et al. [4] are perhaps the first to introduce a small object detection (SOD) dataset, an evaluation metric, and provide a baseline score in order to explore small object detection. Later, Krishna and Jawahar [5] build upon their ideas and suggest an effective upsampling-based technique that performs better results on small object detection. Different from the R-CNN (regions with CNN features) used in [4, 5], Zhang et al. [6] use deconvolution R-CNN [7] for small object detection on remote sensing images. Faster R-CNN [8] and single shot detector (SSD) [9] are two major approaches in object detection. Based on Faster R-CNN or SSD, some small object detection methods [10–11] are proposed.

1.1 Challenges in small object detection

Accurately detecting and tracking object from a video sequence is a challenging task because of the fact the object can have complicated structure and can change shape, size or orientation over subsequent video frames. Designing efficient and accurate system is always a big challenge. Some of the major challenges that arise while detecting small objects because of occlusions, short comings of capturing devices, variations in scenes or appearance, shadows of the objects appearing in the frame. These tend to degrade the performance of the developed algorithm and results of detections may be poor.

To overcome these challenges, the developed algorithm must take care of these issues while proposing solutions for specific applications. The major design issues are summarized in this section that can act as various challenges for researchers to focus upon.

1. Low quality of Image capturing tool results in generation of noisy or blurry image that in turn leads to false detection. The image could also be noisy because of weather conditions like rain, fog etc. The system shall be able to work in noisy images and able to detect the objects in videos with precise boundaries. The quality of camera deployed for capturing images need to be considered together with the weather conditions.

2. Camera jitter It makes the captured object look blurred with prolonged boundaries. The proposed system shall provide effective ways to handle **camera jitter** that occurs due to high velocity winds blowing at time of image capturing. The detection method shall be able to overcome this limitation.

3. **Video captured on moving cameras** like cameras installed on top of vehicles add another dimension to this already challenging area. The movement of camera needs to be simulated by the algorithm for effective and accurate object detections. The problem of moving object detection within moving camera is one of the most happening areas being explored by researchers.

4. Changes in illumination can occur because of presence or disappearance of a light source in background for eg bulb, tube light, sun etc Rapid Illumination changes in the scene of interest lead to false detections in consecutive frames or over multiple frames. The developed solution shall be able to work on different levels of illuminations.

A number of notable object detection methods have been published. These include many reviews on the detection under specific application scenarios, such as face detection, text detection, vehicle detection, traffic sign detection, and remote sensing target detection.

From the perspective of small object detection, Agarwal et al. [12] briefly introduce several methods for detecting small objects in the major challenges of the object detection task. However, they do not systematically analyze it in depth. Unlike these previous object detection surveys, we present a systematic and comprehensive review of deep learning-based algorithms that handle small object detection problems. Our survey is featured by in-depth analysis of small object detection. We summarize existing small object detection algorithms based on five different perspectives: multi-scale feature learning, data augmentation, training strategy, context-based detection and GAN-based detection. Furthermore, the performance of some typical algorithms on popular object detection datasets is carefully analyzed. Last the possible future research directions are discussed. We hope that our survey can provide researchers and practitioners with timely reviews and novel inspirations to facilitate understanding of small object detection and further catalyze research on detection systems.

To the best of our knowledge, most surveys research papers focus only on well known moving object detection and tracking approaches targeted at specific applications. There has been little mention of soft computing approaches in these surveys moreover these studies have been projected as supporting theories or as contributor towards specific applications. The contributions have not been fully acknowledged and this discussion is also randomly scattered. There has not been any survey to pit wonderful contributions that soft computing based approaches have been able to make in field of object detection and tracking in videos.

II. RELATED WORK

Generally, large images are nowadays very more common in many fields, e.g., medicine, digital microscopy and astronomy. Consequently, many basic computer vision tasks, such as object detection, now require a significant computational effort to deal with these big data [13]. Object detection is a relatively fast and simple task if the target image is small, multiple copies of the query object do not appear in the target image, and other objects are concisely dissimilar to the query object. However, this task becomes very computationally complex and challenging if the target image is very large and depicts several objects that are similar to the query one, or the query object is very small compared to the target image. To improve detection accuracy, several efficient methods based on local features [14], e.g., SIFT, and neural networks and FastRCNN have been proposed [15]. Methods based on local features require only an example of the query object, which is detected by matching its features with those extracted from the target image. Therefore, these methods can detect any object with no a priori training process. On the other hand, methods based on neural networks require several examples of the query object(s) for training [16]. Thus, they are restricted to detecting the object(s) provided in the training process.

Object detection, which not only requires accurate classification of objects in images but also needs accurate location of objects is an automatic image detection process based on statistical and geometric features. The accuracy of object classification and object location is important indicators to measure the effectiveness of model detection. Object detection is widely used in intelligent monitoring, military object detection, UAV navigation, unmanned vehicle, and intelligent transportation. However, because of the diversity of the detected objects, the current model fails to detect objects. The changeable light and the complex background increase the difficulty of the object detection especially for the objects that are in the complex environment. Usually, since small objects have low resolution and are near large objects, small objects are often disturbed by the large objects and it leads to failure in being detected in the automatic detection process.

2.1. Object detection method

Object detection can be defined as the task of predicting the location and class of instances belonging to a specific class from an input image. This task has been actively researched in the domain of computer vision for some time. Deep learning has attracted much attention from various fields since a method using deep learning [17] won overwhelmingly in the classification task at ImageNet Large-Scale Visual

Recognition Challenge (ILSVRC) e 2012 compared with its counterparts in the same year and in previous competitions. This trend has continued, and many object detection algorithms using deep learning have been presented since Regions with CNN features (RCNN) [18] was presented in 2014. Recent object detection algorithms can be classified into two main groups: ‘‘two-stage detection’’ such as RCNN and Faster-RCNN [19], in which object proposal and classification are done step by step, and ‘‘onestage detection’’ such as YOLO [20] and SSD [21], in which they are done simultaneously. A characteristic of detectors classified as ‘‘two-stage detection’’ is the high accuracy of the location of the resulting bounding boxes. In contrast, detectors classified in the ‘‘one-stage detection’’ group run much faster than those in the two-stage detection group. Some consumer-oriented drones can fly at a maximum speed of 72km=h (M 42.0S, 2019). For surveillance, a long processing time per frame leads to a delay in the detection of approaching drones. This means that countermeasures cannot be implemented promptly. Hence, object detection should preferably be performed quickly, such as at 10 fps which is the camera frame rate of the proposed system. For these reasons, detection processing speed is important for drone surveillance.

Deep learning is a transformative technology in machine learning, especially in computer vision. With the advantages of Convolutional Neural Networks in extracting high-level features of images, deep learning has achieved great success in the field of target detection. In 2014, Girshick et al. first proposed the deep learning model R-CNN based on convolutional neural network on CVPR 2014. In 2015, Girshick and Ren et al. proposed Fast-RCNN and Faster R-CNN algorithms. In 2016, Redmon et al. proposed the YOLO3algorithm. The detection speed is 10times faster than Faster R-CNN4. The extracted candidate regions are directly classified and regressed in the main network structure, and it will classify and locate the integrated thoughts provide new ideas for subsequent research. Based on the YOLO algorithm, in 2016, LIU and Redmon proposed the SSD algorithm. The SSD algorithm simplifies the entire process of object detection, integrates target determination and recognition, and greatly improves the operating speed. In 2020, YOLO v4 and YOLO v5 have been released one after another; introducing new data enhancement methods, and the detection speed and accuracy are greatly enhanced.

In this section, we will extensively review the methods of detecting small objects from five aspects, including multi-scale feature learning, data augmentation, training strategy, context-based detection and GAN-based detection.

2.2 Multi-scale feature learning

Handling feature scale issues is of crucial importance for small object detection. There are seven main paradigms addressing multi-scale feature learning problem: featurized image pyramids, single feature map, pyramidal feature hierarchy, integrated features, feature pyramid network, feature fusion and feature pyramid generation, and multi-scaled fusion module.

A simple operation is to resize input images into many different scales and to learn multiple detectors, each of which is in charge of a certain range of scales. Most of previous object detectors are based on hand-engineered features [22] utilizing featurized image pyramids to detect objects at various scales. However, the handcrafted features have been replaced with features computed by CNNs because deep CNNs have improved significantly the performance of object detection. Liu et al. [23] proposed a scale-aware network to resize images so that all objects were in a similar scale and then trained a single scale detector. Through conducting comprehensive experimental study on small object detection, Singh and Davis [24] held that training a single scale-robust detector to cope with all scale objects was harder than training scale-dependent detectors with image pyramids. Thus, they designed a new framework called scale normalization for image pyramids (SNIP). SNIP trained multiple scale-dependent detectors and each of which was in charge of a specific scale objects. Wang et al. [25] presented a cascade mask generation framework, which efficiently combined multi-scale inputs for fastly detecting small objects. Nevertheless, these works are computationally expensive considering rapid increase of memory consumption and inference time.

Recent detection algorithms such as Fast R-CNN [26], Faster R-CNN [27], SPPNet [28] and R-FCN [29] utilize the top-most feature maps computed by CNNs on a single input scale to predict candidate bounding boxes with different aspect ratios and scales. Nevertheless, the top-most feature maps conflict with objects at different scales in images due to their fixed receptive field. There is little information left on the top-most features especially for small objects, so it may compromise detection performance of small objects. Deep CNNs learn hierarchical features in different layers which capture information from different scale objects. MSCNN used de-convolutional layers on multiple feature maps to improve their resolutions and then predictions were made by using these refined feature maps. Based on the above analysis, we can find that multi-scale feature learning is useful for detecting small objects.

III. PROPOSED WORK

Our model is designed on the Hebbian principle which states that neurons that are coupled together activate together. Basically Hebbian learning is used for training the lower or higher layers of the neural network. Hebbian learning is effective in re-training the network of a pre-trained neural network. We address common issues that are overlooked in previous works regarding a new deep learning model design. We solve overfitting problems of wider network by introducing a sparse structure of convolutional blocks in our model. We engineer the model to solve sketch object iconic and abstract nature by using large number of training samples.

3.1 Framework Details

Our model is trained on the TU-Berlin sketch dataset which consists of 20,000 objects from 250 categories. We apply data-augmentation techniques on the dataset to elastically increase its size. Our model achieves ground breaking recognition accuracy.

The dataset consist of some pre-defined objects image as well as some sketched images.

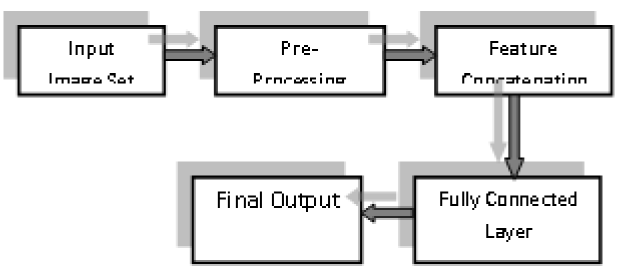

Figure 4.1: Block Diagram of Proposed Model.

In this system, the image is feed to the system as an input. The image file is loaded with different image category list. Here we are using more than 250 image categories with some predefined images as well as sketched images. Then label encoders are applied to encode the images. One hot image encoding technique is applied for encoding process. The one hot encoder will generate a sparse integer. Then we will convert the sparse integer into text label format like true for 1 and false for 0. Then a confusion matrix will be generated based on predicted labels. Then we will select the top five prediction and we will print the accuracy.

3.2 Proposed Pre-Processing Steps

Following are the pre-processed applied to generate ROI for the pooling layers.

Step 1: Input Image Dimension.
Step 2: Select number of Channels.
Step 3: Create a matrix to store all flattened images.
Step 4: Re-shape the image and create a train data.
Step 5: Split the data into training and testing set. Train dataset is created at batch size of 32 and epoch 200.
Step 6: Convert class vectors to binary class matrix.
Step 7: Send the binary class matrix to next layer.

3.3 Proposed Processing Steps

Following are steps in object detection Improved Fast-RCNN.

Step-1: Input Imageset.
Step-2: Take following layers for Pre-Processing Function. We select 15 convolution layers and 3 Max Pooling Layers. Again 5 Convolution layers and 3 Max Pooling layers. The process is repeated untill a drop out of 0 (Zero) is generated.
Step-3: Sparse Structure is generated from all the input layers of first step.
Step-4: All the Features are concatenated after Sparse Structure is generated.
Step-5: Again Max Pooling is applied to select maximum pixels value.
Step-6: The output of Max Pooling will be send as input to Fully Connected Layer.
Step 7: The final prediction will be generated with model prediction Accuracy.

By successfully solving sketch recognition, we can now move towards solving multi-object recognition, sketch object segmentation, image retrieval based on sketch query and the most popular current trend in computer vision, the use of Generative Adversarial Networks to synthesis sketch objects or use a sketch object to synthesis a complete photo realistic image.

IV. RESULTS WORK

The dataset contains sketch objects of a given category of 20,000 unique sketches evenly distributed over 250 object categories. With this dataset we perform a perceptual study and find that humans can correctly identify the object category of a sketch. We compare human performance against computational recognition methods. Also it contains some jpeg images of small objects like Cifar-10 dataset.

Table 6.1 Comparison of Existing and Proposed Model.

Models	Prediction Accuracy
Fast-RCNN	45%
Faster-RCNN	65%
Multiscale RCNN	71 %
Our Model	85 %

We have tested the models on the Tu-Berlin Sketch datasets. From the above result we can conclude that our model performs better than existing models.

REFERENCES

[1] A. Kembhavi, D. Harwood, L.S. Davis, Vehicle detection using partial least squares, IEEE Trans. Pattern Anal. Mach. Intell. 33 (6) (2011) 1250–1265.

[2] V.I. Morariu, E. Ahmed, V. Santhanam, D. Harwood, L.S. Davis, Composite discriminant factor analysis, IEEE Conference on Applications of Computer Vision 2014, pp. 564–571.

[3] T.T. Le, S.T. Tran, S. Mita, T.D. Nguyen, Real time traffic sign detection using color and shape-based features, Asian Conference on Intelligent Information and Database Systems 2010, pp. 268–278.

[4] C. Chen, M.-Y. Liu, O. Tuzel, J. Xiao, R-CNN for small object detection, Asian Conference on Computer Vision 2016, pp. 214–230.

[5] H. Krishna, C.V. Jawahar, Improving small object detection, Asian Conference on Pattern Recognition 2017, pp. 340–345.

[6] W. Zhang, S. Wang, S. Thachan, J. Chen, Y. Qian, Deconv R-CNN for small object detection on remote sensing images, IEEE International Geosciences and Remote Sensing Symposium 2018, pp. 2483–2486.

[7] R.B. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, IEEE Conference on Computer Vision and Pattern Recognition 2014, pp. 580–587.

[8] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, Adv. Neural Inf. Proces. Syst. (2015) 91–99.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S.E. Reed, C.-Y. Fu, A.C. Berg, SSD: single shot multibox detector, European Conference on Computer Vision 2016, pp. 21–37.

[10] C. Eggert, S. Brehm, A.Winschel, D. Zecha, A closer look: small object detection in faster R-CNN, International Conference on Multimedia and Expo 2017, pp. 421–426.

[11] C. Cao, B. Wang, W. Zhang, X. Zeng, X. Yan, Z. Feng, Y. Liu, Z. Wu, An improved faster R-CNN for small object detection, IEEE Access 7 (2019) 106838–106846.

[12] S. Agarwal, J.O.d. Terrail, F.e.e. Jurie, Recent Advances in Object Detection in the Age of Deep Convolutional Neural Networks, CoRR abs/1809.03193 2018.

[13] J. Kremer, K. Stensbo-Smidt, F. Gieseke, K. S. Pedersen, and C. Igel, "Big universe, big data: Machine learning and image analysis for astronomy," IEEE Intelligent Systems, vol. 32, no. 2, pp. 16–22, Mar 2017.

[14] P. Azad, T. Asfour, and R. Dillmann, "Combining harris interest points and the sift descriptor for fast scale-invariant object recognition," in 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct 2009, pp. 4275–4280.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, June 2017.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 2014, pp. 580–587.

[17] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097e1105.

[18] Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proc. Of the IEEE Conf. On Computer Vision and Pattern Recognition, pp. 508e587.

[19] Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91e99.

[20] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: Proc. Of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 779e788.

[21] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. Ssd: single shot multibox detector. In: European Conf. On Computer Vision. Springer, pp. 21e37.

[22] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition 2005, pp. 886–893.

[23] Y. Liu, H. Li, J. Yan, F.Wei, X.Wang, X. Tang, Recurrent scale approximation for object detection in CNN, IEEE International Conference on Computer Vision 2017,pp. 571–579.

[24] B. Singh, L.S. Davis, An analysis of scale invariance in object detection SNIP, IEEE Conference on Computer Vision and Pattern Recognition 2018, pp. 3578–3587.

[25] G.Wang, Z. Xiong, D. Liu, C. Luo, Cascade mask generation framework for fast small object detection, IEEE International Conference on Multimedia and Expo 2018.

[26] R. Girshick, Fast R-CNN, IEEE International Conference on Computer Vision 2015, pp. 1440–1448.

[27] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, Adv. Neural Inf. Proces. Syst. (2015) 91–99.

[28] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.

[29] J. Dai, Y. Li, K. He, J. Sun, R-FCN: object detection via region-based fully convolutional networks, Advances in Neural Information Processing Systems 2016, pp. 379–387.