International Journal of Engineering & Technology, 7 (2.24) (2018) 428-430

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Object Detection Using Support Vector Machine and Convolutional Neural Network - A Survey

Rishi Khosla[1], Yashovardhan Singh[2], T Balachander[3]

[1,2]UG Student, [3]Assistant Professor
Computer Science and Engineering, SRM Institute of Science & Technology Chennai, India
**Corresponding Author Email: rishi.khosla21@gmail.com*

Abstract

Mobile Technologies have been in trend for quite some time and with the advances in machine learning, they have become more powerful. Computer Vision, Computational Analysis and Computer Graphics have changed over the course of time. In this Project, our aim is to figure out the domains in which Machine Learning can be applied to enhance the capabilities of a Mobile Device which would lead to a better and sustainable mobile user experience. The models we would use are a convolutional neural network (CNN), support vector machine (SVM) and scale-invariant feature transform (SIFT). This project uses the real-time image from a mobile device and does the classification and detection with the help of Tensor Flow and provides the result with a confidence score.

Keywords: *Convolutional Neural Network (CNN), Mobile Device, Scale-Invariant Feature Transform (SIFT), Image Detection, Image Classification, Support Vector Machine (SVM), Machine Learning.*

1. Introduction

Machine Learning is implemented everywhere and with Mobile Technologies being in trend for quite some time with Computer Vision, Computational Analysis and Computer Graphics have changed over the course of time. In this Project, our aim is to figure out the domains in which Machine Learning can be applied to enhance the capabilities of a Mobile Device which would lead to a better and sustainable mobile user experience. We would use CNN and SVM to classify the different domains specific challenges. Mobile devices can be made more powerful by applying Machine Learning. We aim to gain as much accuracy as possible to overcome domain specific challenges. Convolutional neural network (CNN) is often used in visual detection because of its unmatched accuracy. Apart from classification, there are many applications of CNN in visual recognition such as object detection [2], sentence generation from the image [3], segmentation [4], and localization [5]. We focus only on detection, as it would have multiple applications in our lives. The main aim of this project is to identify objects in an image frame by providing the confidence score and to predict the corresponding bounding boxes for each detected object. We would use rCNN because region CNN (rCNN) [5] is the one that provides very high accuracy in object detection. We minimize the complexity of underlying rCNN. First, we would use edge box [7], a recently published algorithm to generate region proposals, in place of selective search used in rCNN. We were extremely careful in designing of the training data and fine-tuning the CNN. Fig. 1 is an shows our system.

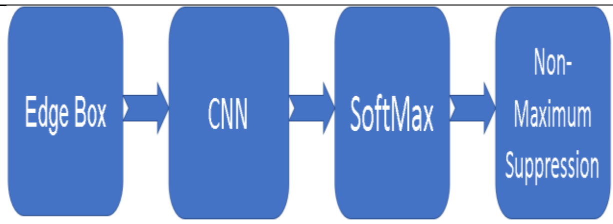

Fig. 1: Proposed Object Detection Model

1). Proposals generated by Edge Box, which is an algorithm designed purely for this purpose. 2). A fine-tuned CNN. 3) A Confidence score is generated by the deployment of Softmax on all the different classes for each and every proposal. 4). Non-maximum suppression (NMS) is used for all classes to merge the overlapping proposals.

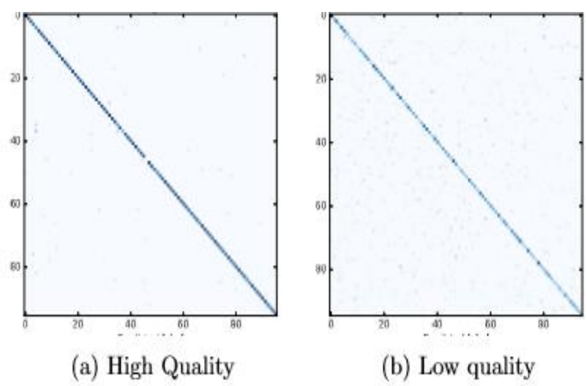

Fig. 2: Confusion matrix for SVM

2. Approach

Our object Detection model has a few vital steps. Details of our object detection model are mentioned below.

A. Proposal generation

We use edge boxes as the proposal generation algorithm. The algorithm used for this task generates proposals according to the image's edge map. Edge map is generated with the help of a structured edge detector with each pixel having the edge magnitude.

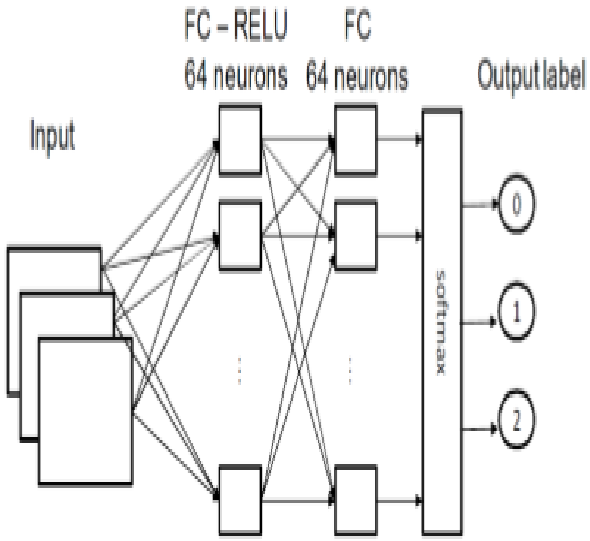

Fig. 3: The overview of CNN model

B. Training procedure

As mentioned in above, we will carefully design our training data. In CNN, bounding boxes with Intersection Over Union should be more than 0.6. If IoU is less than 0.3 then it is ignored. IoU as 0.6 is not to be used so as to differentiate between the data to be either positive or negative because it decreases the localization performance and would eventually give an inaccurate result.

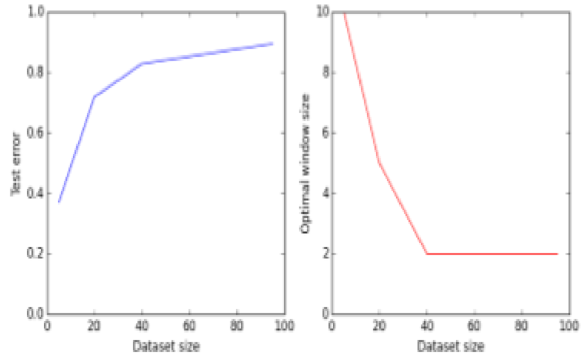

Fig. 4: Performance of SPM with varying Dataset Sizes

C. Testing procedure

We generate regional proposals with edge boxes, then performing the forward pass through the fine-tuned CNN. CaffeNet model input size is fixed to 227× 227 pixels. Non-Maximum Suppression (NMS) algorithm is applied to neglect the unwanted proposals. This algorithm sorts the proposals. This sorting is done on the basis of the confidence score and then neglecting the proposals overlapping with each other. This value of overlap is termed as the IoU. The performance of detection is calculated by the mean average precision (mAP).We use the ImageNet dataset for classification and detection. The dataset has more than 50000 images for image detection. Training dataset has 25000 images. 25000 images are used as the Testing dataset. Figure 8 and Figure 9 show the SVM and CNN outputs of a real-time image of a laptop. Figure 10 and Figure 11 show the outputs of SVM and CNN respectively for a real-time image of a bottle.

3. Discussion

The image classification into various categories is done with the help of machine learning. Computer vision also plays an important role in this. Therefore, to create a more precise system, we need a large and continuously growing data source. The accuracy of the model with CNN is much better as compared to the model made with SVM. All the values, such as precision, and F1 are high, which means high accuracy. The blue diagonal line is the proof of the claim. The confusion matrix and the values on the low-quality dataset are worse than the ones with a high-quality dataset. Our classifier misunderstands similar signs due to lack of enough features to distinguish in between these signs.

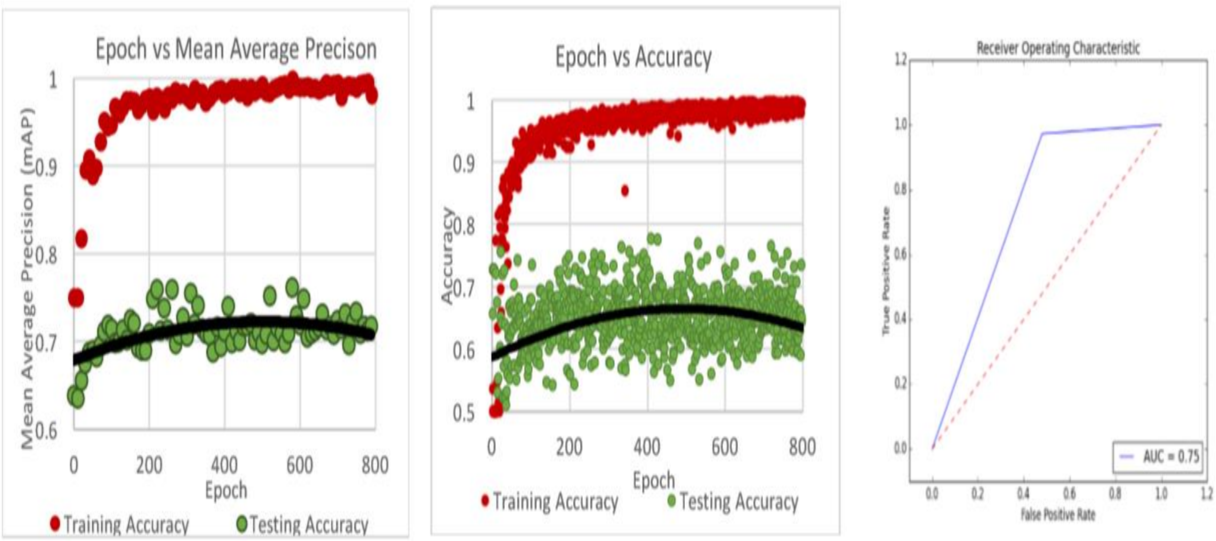

Fig. 5: The Accuracy Graph During Training and Testing

Table 1. SVM accuracy chart

	SVM	Log. Reg.	LSTM	SPM
Precision	0.566	0.444	0.109	0.075
Recall	0.550	0.444	0.091	0.076
F1	0.549	0.436	0.066	0.065
Training error	0.001	0.179	0.852	0.526
Testing error	0.450	0.556	0.908	0.895

4. Conclusion

The classification and detection of the image done through various methods bring down us to the point that fine-tuning the supervised dataset using Convolution Neural Network significantly increases the recognition accuracy than when it is done with support vector machines. We also monitored that the rate of error percentage is less when we used CNN. An extension to this project would be the possibility to classify and detect objects in a video frame. This could come handy in the field of surveillance and monitoring. We continue this project by adding more data to the data set or we can tune it further and use more pictures to run it with the unsupervised dataset which can be a concrete step in order to use technology in real time image classification and detection.

5. Appendix

This is how the Precise-recall curve for IoU=0.4 looks like.

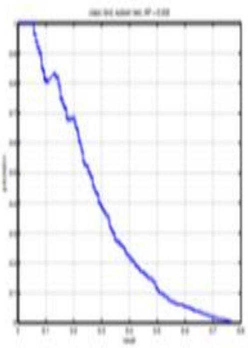

Fig. 6: laptop

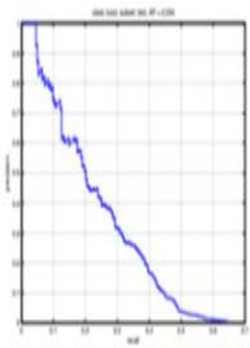

Fig. 7: Bottle

Fig. 8: SVM for laptop

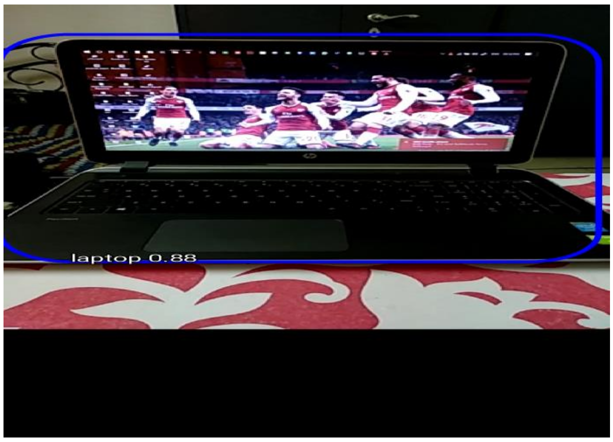

Fig. 9: CNN for laptop

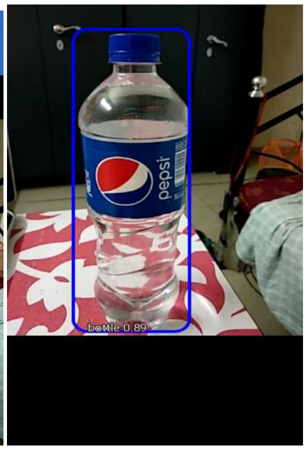

Fig. 10: SVM for bottle **Fig. 11:** CNN for bottle

References

[1] Yasutake, Taizo. "Graphical input controller and method with rear screen image detection." U.S. Patent 5,483,261, issued January 9, 1996.

[2] Chum, Ondrej, James Philbin, and Andrew Zisserman. "Near Duplicate Image Detection: min-Hash and tf-idf Weighting." In BMVC, vol. 810, pp. 812-815. 2008.

[3] Wang, Bin, Zhiwei Li, Mingjing Li, and Wei-Ying Ma. "Large-scale duplicate detection for web image search." In Multimedia and Expo, 2006 IEEE International Conference on, pp. 353-356. IEEE, 2006.

[4] Hambly, N. C., M. J. Irwin, and H. T. MacGillivray. "The Super COSMOS Sky Survey—II. Image detection, parametrization, classification and photometry." Monthly Notices of the Royal Astronomical Society 326, no. 4 (2001): 1295-1314.

[5] Sugiyama, Susumu, Ken Kawahata, Masakazu Yoneda, and Isemi Igarashi. "Tactile image detection using a 1k-element silicon pressure sensor array." Sensors and Actuators A: Physical 22, no. 1-3 (1990): 397-400.

[6] Hsu, Yu-Feng, and Shih-Fu Chang. "Image splicing detection using camera response function consistency and automatic segmentation." In Multimedia and Expo, 2007 IEEE International Conference on, pp. 28-31. IEEE, 2007.

[7] Miyasaka, Tsutomu, Koichi Koyama, and Isamu Itoh. "Quantum conversion and image detection by a bacteriorhodopsin-based artificial photoreceptor." Science255, no. 5042 (1992): 342-344.

[8] S.V.Manikanthan and T.Padmapriya "Recent Trends In M2m Communications In 4g Networks And Evolution Towards 5g", International Journal of Pure and Applied Mathematics, ISSN NO:1314-3395, Vol-115, Issue -8, Sep 2017.

[9] T. Padmapriya and V.Saminadan, "Improving Performance of Downlink LTE-Advanced Networks Using Advanced Networks Using Advanced feedback Mechanisms and SINR Model", International Conference on Emerging Technology (ICET), vol.7, no.1, pp: 93, March 2014.