International Journal of Engineering & Technology, 7 (2.24) (2018) 33-36

International Journal of Engineering & Technology

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

Research paper

Real Time Object Detection using CNN

Akash Tripathi [1*], T.V. Ajay Kumar[2], Tarun Kanth Dhansetty[3], J. Selva Kumar [4]

[1,2,3]Final Year B. Tech (ECE), [4]Associate Professor,
Department of ECE, SRMIST, Kattankulathur603 203
**Corresponding Author Email: akashtripathi29196@gmail.com*

Abstract

Achieving new heights in object detection and image classification was made possible because of Convolution Neural Network(CNN). However, compared to image classification the object detection tasks are more difficult to analyze, more energy consuming and computation intensive. To overcome these challenges, a novel approach is developed for real time object detection applications to improve the accuracy and energy efficiency of the detection process. This is achieved by integrating the Convolutional Neural Networks (CNN) with the Scale Invariant Feature Transform (SIFT) algorithm. Here, we obtain high accuracy output with small sample data to train the model by integrating the CNN and SIFT features. The proposed detection model is a cluster of multiple deep convolutional neural networks and hybrid CNN-SIFT algorithm. The reason to use the SIFT featureis to amplify the model‟s capacity to detect small data or features as the SIFT requires small datasets to detect objects. Our simulation results show better performance in accuracy when compared with the conventional CNN method. As the resources like RAM, graphic card, ROM, etc. are limited we propose a pipelined implementation on an aggregate Central Processing Unit(CPU) and Graphical Processing Unit(GPU) platform.

Keywords: *Convolution Neural Network(CNN), Scale-Invariant Feature Transform(SIFT), confidence value, object detection, proposed regions.*

1. Introduction

Human computer interaction is an important application of object detection using CNN which is a challenging and an interesting problem. It can be used to build more smatter and accurate robots with an ability of better understanding the objects. There are many other real life applications of this paper such as surveillance cameras used at highways to prevent over speeding, interactive game development and driverless cars. Object detection which includes location detection and detecting the categories of various objects present in one single image, nearly two thousand regions are proposed to contain an object in the image which are called the proposed regions. Several machine learning and feature extraction algorithms have been developed for object detection tasks. Numerous handcrafted feature extraction techniques for object detection such as SVM(Support Vector Machines), SGD(Stochastic Gradient Descent), Convolutional Neural Support Vector Machines (CNSVMs) or an integration of multiple features have been proposed. Due to the success of Convolutional Neural Network (CNN) in image classification tasks[3] it has also been used in object detection tasks. Contrasting to conventional Computer Vision(CV) systems and other machine learning tasks where each feature must be defined beforehand manually, here in CNN, in automatically learns to extract features from the predefined database of features. The CNN is combined with neural network classifiers which are feed forward to make the CNN network trainable on the dataset all over. CNN requires a very large number of trained data to speculate well enough. The capability of CNN to work on larger datasets even with a limited computational power increases the value of CNN and making it more likely to be used. However, this is not the case always, sometimes we need to detect objects with a limited number of features. Although SIFT[6] and other traditional detection techniques give a less accurate result than CNN[4],[5], they require only a small amount of datasets to speculate. The conventional methods have their own limitations such as their modeling capacities are bound which stay unchanged for different sources of data. In this paper we offer an imperial approach of the combination of both CNN and SIFT i.e. a hybrid having the best of both worlds. Here we are comparing individual CNN model and SIFT-CNN model.

2. Related Work

The present research in computer vision field is on object detection and classification. There are several existing object detection algorithms and techniques. Some of the techniques include algorithms like SVM‟s(Support vector machines), SGD(Stochastic Gradient Descent), Convolution Neural Support Vector Machines (CNSVMs) which can be found in [10],[11],[12],[13].In [11], the author suggests that support vector machines are used as learning algorithms that scrutinize data for image classification and reversion analysis. From several papers [11],[12] , when a data set is taken, SVM maps by separating comparative features and hence it also clears the gap and makes it as wide as possible. In addition, classification can be performed in two different ways i.e. linear and non-linear based on kernel trick. These kernel methods are used in machine learning, in which different algorithms can analyze the pattern from the images [12]. Supervised learning models are not possible when the data sets are not labeled and for clustering the data into groups, unsupervised learning models are required and then the new data is mapped to these formed groups [11]. The main limitation in SVM from [11]-[12], is speed and it leaves some of the points in the data set while dividing similar points. In [10] stochastic gradient descent (SGD)-based adaptive neural network is present, with a better tracking performance an also provides optimization algorithms for the weights. Stochastic gradient descent is generally an approximation

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in machine learning for linear classification under convex loss functions such as logistic regression and support vector machines. In the paper [13], for minimizing an objective function, an iterative method is used which is written as sum of differentiable functions. In this iteration, maxima and minima of the functions can also be found out by SGD. The scarce machine learning problems which occur in text classification and natural language processing have been fortunately applied by SGD‟s. When the data is sparse, the linear classifiers used in the SGD‟s easily scale to issues with more than 10^5 training examples and more than 10^5 features [10]. SGD classifier is used for classification of different objects given in the data set. This classifier supports different loss functions and penalties for classification. The main advantage of using SGD is efficiency and the ease of its implementation [10]. The limitations of using SGD are, it requires many number of hyper-parameters and iterations which can be found in [10],[13]. Convolutional neural support vector machines (CNSVMs) is a combination of CNN and SVM, which is used for visual pattern recognition and learning models.

In the paper [14], CNSVM methods are done on the color FERET data collection and also in CNSVM‟s, the fully-connected layer of CNN acts as an input to SVM and the output layer of CNN is substituted by SVM.

3. Design and Architecture

In this paper, we are using a webcam of eight mega pixels, for the system to be more robust against the noise. Each image trained in the system is amplified by using different transformations linearly. These transformations include rotated and scaled trained images. In this model we have built a characteristic CNN network architecture from square one. This network contains6 layers, of which there are 2max-pooling layers followed by **a** fully connected layer. Each time there‟s a max-pooling layer added the number of filters get doubled. The window size of the filter considered here is 3x3 and then comes the max pooling layer with a size of 2x2 is stationed past every 2 convolutional layers. Max-Pooling is accustomed to reduce the computation for the deeper layers and also helps in presenting of translation invariance. The SIFT algorithm is activated only at the times we intend to process small features. The SIFT [7] algorithm is used to extract the key points from the object image. After detecting features like key points, magnitude and direction of the object image, using key points of neighboring pixels we calculate the image gradient. Another example of SIFT and CNN is proposed in [8].

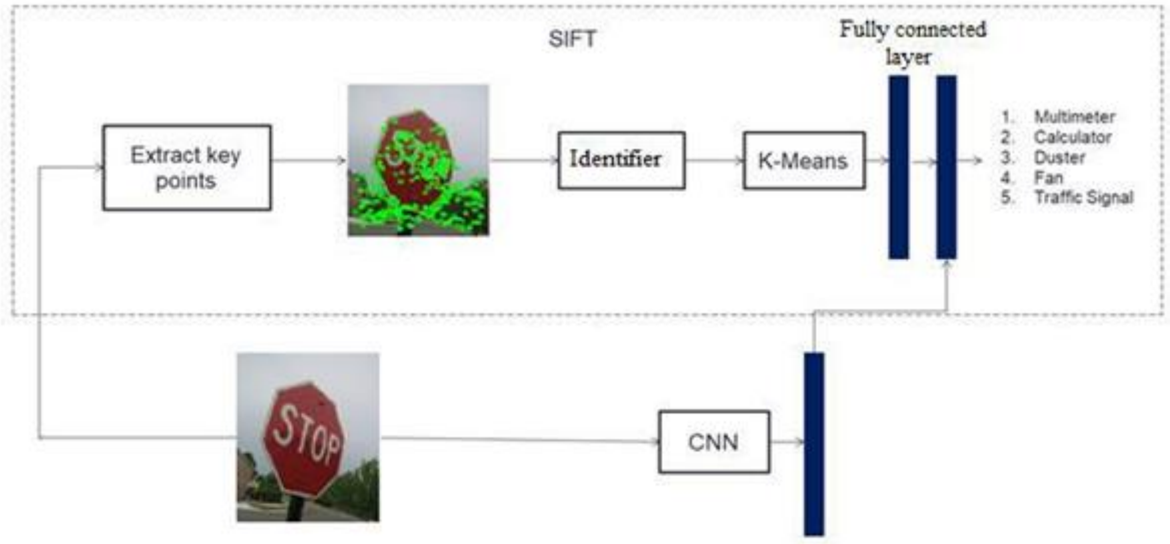


Fig. 1: Overview of the proposed methods

The SIFT algorithm is robust with respect to scale. Which means if we calculate the SIFT descriptor for the detected points we can use the Euclidean distance to match them regardless of the key points' scale. SIFT first uses a reference image to extract the key points of the object and then stored in the database. Now each new image feed for test is compared individually to the database and Euclidian distance of their feature vector is used to find candidate matching feature. From the full set of suitable key point matches a subset of key points that consent with the object and its angle, location, scale, orientation and texture of the testing or the new image are identified to filter out good matches. Images are transformed into a large collection of feature vectors in Lowe's method for image feature generation [9], each of these features is unaltered to image scaling, rotation and translation. It is also partially invariant to changes in illumination and local geometric distortions are robust. Similar to the properties of neurons in visual cortex that are encoding basic texture, colour, form and movement for object detection. Here the identifier is also called as the descriptor, which identifies and describes the detected key points detected from the previous step. The identifier basically searches the image for over all scales and image locations. The „K‟ means is used to detect „K‟ neighbouring key points. The aim of K mean clustering is to partition and observe into the „K‟ clusters in which each observation belongs to the cluster with the nearest mean, thereby serving as a prototype for the cluster. Finally there‟s a fully connected layer where the highest level of reasoning in the neural network is done.

4. Result

In this paper, we have trained multiple objects and tested them on Python platform. The maximum count is set to 33 and a threshold of 26.4 is set so that the name of the object is displayed only after this threshold is crossed.

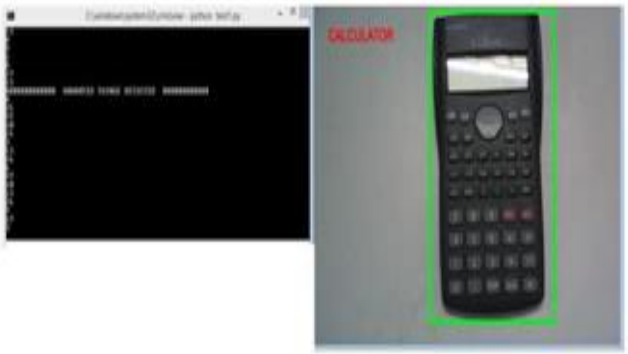


Fig. 2: Simulation of real time object 1(Calculator)

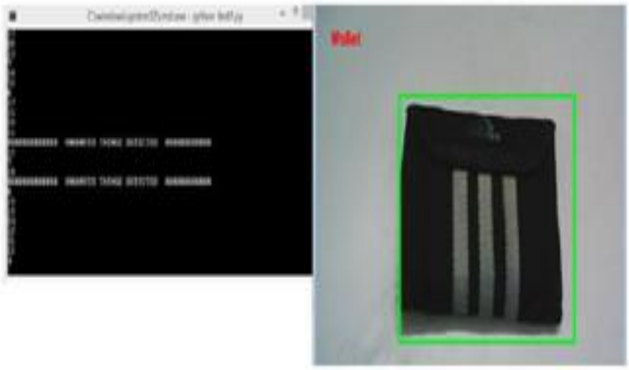


Fig. 3: Simulation of real time object 2(Wallet)

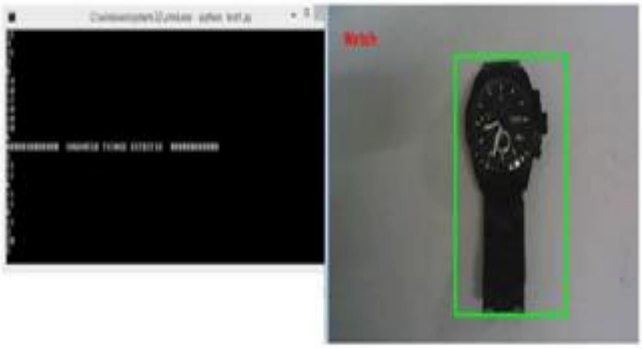


Fig. 4: Simulation of real time object 3(Watch)

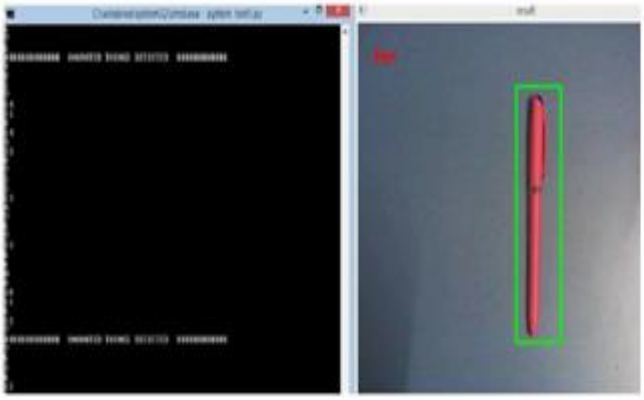
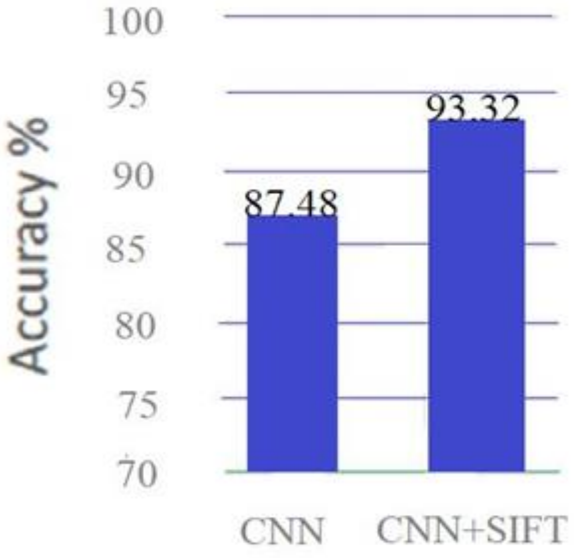


Fig. 5: Simulation of real time object 4(Pen)

Table 1: Accuracy versus average confidence value

OBJECT NAME	AVERAGE CONFIDENCE VALUE	ACCURACY
MULTIMETER	31	93.93%
WATCH	32	96.96%
WALLET	30	90.90%
GANESH IDOL	30	90.90%
CALCULATOR	31	93.93%



Graph 1: Accuracy percentage for CNN and CNN+SIFT methods

In this paper we have trained the model with numerous objects some of which are shown in figures 2,3,4 and 5.Usually the number of trained images is greater than the number of tested images. This table displays the accuracy of various object images with respect to the confidence value obtained in CNN model and the integration of CNN with SIFT model.

5. Conclusion

In this paper, we propose a hybrid Convolutional Neural Network and SIFT aggregator for efficient object detection. We have shown how the integration of CNN with SIFT can enormously increase the efficiency with a noticeable reduction in processing time. This paper represents a pipelined system of CNN and SIFT where the tasks at hand are divided accordingly. For large number of features in a proposed region CNN is used to detect whereas in case of small feature analysis SIFT is used. One of the major drawbacks of CNN is it requires a large amount of data sets, which is subdued by using SIFT algorithm. With this integration model the gain performance is better compared to any other state of art techniques.

References

[1] Mundher Ahmed Al-Shabi, Wooi Ping Cheah, Tee Connie, “facial Expression Recognition Using a Hybrid CNN-SIFT Aggregaator”Aug 10, 2016

[2] Huizi Mao, Song Yao, TianqiTang,Boxun Li, Jun Yao and Yu Wang , Towards Real-Time Object Detection on Embedded Systems. August 2016 , IEEE Transactions on Emerging Topics in Computing

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” ArXiv151203385 Cs, Dec. 2015.

[4] J. Li and E. Y. Lam, “Facial expression recognition using deep neural networks,” in 2015 IEEE International Conference on Imaging Systems and Techniques (IST), 2015, pp. 1–6.

[5] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, and T. Brox, “Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks,” ArXiv14066909 Cs, Jun. 2014.

[6] D. G. Lowe, “Object recognition from local scale-invariant features,” in The Proceed-ings of the Seventh IEEE International Conference on Computer Vision, 1999, 1999, vol. 2, pp. 1150–1157 vol.2.

[7] Kim, B.-K. et al.: Hierarchical Committee of Deep Convolutional Neural Net-works for Robust Facial Expression Recognition. J. Multimodal User Interfaces. 10, 2, 173–189 (2016).

[8] Sun, B. et al.: Facial Expression Recognition in the Wild Based on Multimodal Texture Features. J. Electron. Imaging.25, 6, 061407–061407 (2016).

[9] Wikipedia. Link of the reference-“ https://en.wikipedia.org/wiki/Scale-invariant_feature_transform”.

[10] 4.Girshick, Ross, et al. "SGD-Based Adaptive NN Control Design for Uncertain Nonlinear Systems”. IEEE Transactions on Neural Networks and Learning Systems (Volume: PP, Issue: 99). 30 January 2018

[11] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Fuzzy SVM for 3D facial expression classification using sequential forward feature selection” Computational Intelligence and Communication Networks (CICN), 2017 9th International Conference. 16-17 Sept. 2017

[12] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.

[13] R. G. J. Wijnhoven and P. H. N. de With, "Fast Training of Object Detection Using Stochastic Gradient Descent," 2010 20th International Conference on Pattern Recognition, Istanbul, 2010, pp. 424-427.

[14] J. v. d. Wolfshaar, M. F. Karaaba and M. A. Wiering, "Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, 2015, pp. 188-195.

[15] 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.

[16] S.V. Manikanthan , T. Padmapriya “An enhanced distributed evolved node-b architecture in 5G tele-communications network” International Journal of Engineering & Technology (UAE), Vol 7 Issues No (2.8) (2018) 248-254.March2018.

[17] S.V. Manikanthan, T. Padmapriya, Relay Based Architecture For Energy Perceptive For Mobile Adhoc Networks, Advances and Applications in Mathematical Sciences, Volume 17, Issue 1, November 2017, Pages 165-179.