OBJECT DETECTION USING CONVOLUTIONAL NEURAL NETWORK

[1]Kolluri Bhavesh., [2]Patade Omkar, [3]Rajapkar Adesh, [4]Prof. Saroja T.V.

[1]Student, [2]Student, [3]Student, [4]Associate Professor
[1]Computer Engineering Department,
[1]Shivajirao S. Jondhale College of Engineering, Dombivali(E), India

Abstract: Viewing systems are essential for building a mobile robot that will complete a task such as navigation, surveillance, and explosive disposal (EOD). This will let the robot's controller or operator know what's in nature and perform the following tasks. With the recent development of deep neural networks in image processing, distortion and precision detection are now possible. In this paper, Convolutional Neural Networks (CNN) is used to find natural objects. Two types of art models are compared for object acquisition, Single Shot Multi-Box Detector (SSD) with MobileNetV1 and Faster Region-based Convolutional Neural Network (Faster-RCNN) with InceptionV2. The result shows that one model is suitable for real-time use because of the speed and the other can be used to get something more accurate.

Index Terms - **Computer vision, convolutional neural networks, image classification, object discovery, transfer learning.**

I. INTRODUCTION

Object discovery is essential for computer viewing programs. It can be used for many applications such as video surveillance, medical imaging, and robotic navigation. Many algorithms can be used for this function such as background extraction, time variation, light flow, Kalman filtering, vector support mechanism, and mounting similarity. Apart from the stated algorithms, the new method used for object detection is called convolutional neural networks (CNN). The fragmentation of images began when Alex won the 2012 ImageNet competition using deep convolutional neural networks. They trained in deep CNN to distinguish high-resolution images of 1.2 million in the 1000N ImageNet competition. They have gained more accurate predictions than previous models of art models. In this regard, many researchers are interested in finding a new way to build more complex networks of convolutional neural network.

There are some recent methods of object acquisition, in paper, the authors have developed a low-speed transmitter using a flexible deep structure and a regular transfer learning framework to deal with the acquisition using a few training data. Another paper in proposed a regional selection network and an object collection network. The regional selection network serves as a guide for where to choose regions to learn features from them. On the other hand, the collection network acts as a local feature selector that converts feature maps. In convolutional neural networks are used to track targeted attention. They have created a large ad-hoc database with good and bad examples of framed objects from the ImageNet database and selected the most promising patch using AlexNet architecture. Another method used for object discovery is active learning. This is a segment of search algorithms that search for highly instructive samples to be included in a training database that works with image classification. Finally, in artificial neural networks are used to detect objects in the shape and form of a color pattern.

This paper is organized as follows; Phase II is a brief introduction to convolutional neural networks. Section III deals with the concept of transfer of learning. Section IV discusses the different TensorFlow models used for object acquisition. Section V shows the test set and information about the database. Section VI is a discussion of the results.

II. EXISTING SYSTEM

In the existing system, the purpose of detection was fulfilled but the network model was not so effective and accuracy was not up to the mark. Here we have added series of convolutional features layer in the network to make it more effective and accurate. Also, the old system was singular whereas now because of the use of CNN it can detect multiple objects in real time and is faster.

III. PROPOSED SYSTEM

Convolutional neural network (CNN) is a class of deep, advanced artificial neural network used to produce precise performance in computer visual functions, such as image separation and acquisition. CNN is similar to the traditional neural network, but with deeper layers. It has weights, prejudices and consequences through offline performance. CNN neurons are arranged in a volumetric manner such as, height, width and depth.
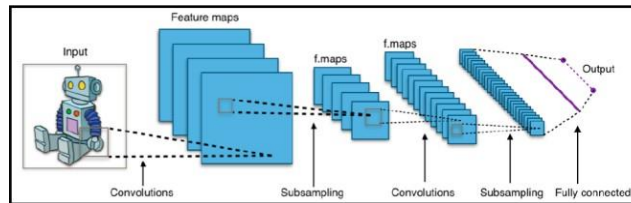
Fig. 1. CNN Architecture

Fig. 1 shows the structure of CNN, consisting of a convolutional layer, a composite layer and a fully connected layer. Conversion layer and merging layer are usually changed and the depth of each filter increases from left to right while the output size (length and width) decreases. A fully integrated layer is the last layer similar to the last layer of standard neural networks.

Inserted image that will capture pixel values. It has three dimensions such as width, height and depth (RGB channels) for example [50 x 50 x 3]. The convolutional layer will calculate the release of neurons connected to local regions in the input. Layout parameters are made up of a set of readable filters (or characters), which meet the width and height of the input volume that extends to its depth, combining the product of the dots between the input and filter filters. This displays a 2-sided working map of this filter and as a result, the network reads the first filter when it detects a specific type of feature in a specific input area. The function called the Rectified Linear Unit (ReLU) will perform the elementwise activation function. ReLU is defined in (1),

$$f(x) = \max(0, x) \qquad (1)$$

This activity is not zero in negative values and grows in the order of positive values. This will not affect the volume size. The integration layer produces high performance in the region. This floor is a sample size of area similar to width and height. Layer to release a fully connected layer similar to the last layer of the neural network. This layer has used the most common softmax performance to exclude possible distribution with a number of output classes.

IV. TRANSFER LEARNING

Transfer learning is a powerful and in-depth learning method where previously trained models can be used for feature extraction and fine adjustment. This method can be used for image classification such as vehicle classification, object detection and classification. The advantage of using this method saves time training the network from scratch and requires less data to get better results. Training can be done and using the central processing unit (CPU) even without the computer graphics processing unit (GPU). Instead of creating a new model from scratch, pre-trained models can be used. These species are trained from large database such as ImageNet and COCO database. Some of these types are AlexNet, VGG16, VGG19, ResNet50, InceptionV2, InceptionV3, Xception, DenseNet and MobileNet. This paper focuses on only two variants, InceptionV2 and MobileNet.

V. TENSORFLOW OBJECT DETECTION MODELS

TensorFlow is an open-source software source for high value calculation of price performance. The following types of TensorFlow are used to detect unauthorized quadrotor (UAV) aircraft with the person in the photo and video.

A. SSD With MobileNetV1

The Single Shot Multi-Box Detector (SSD) is a method for detecting an object in images using a deep neural network. It produces binding boxes and class scores for something obtained faster than the previous method such as You Only Look Once (YOLO). On the other hand, MobileNet is an in-depth learning model that can be used for a variety of recognition functions such as object detection, finegrain detection, remarkable recognition, and facial expressions. The benefits of MobileNet are accurate, fast, minimal and easy to tune. It is based on a simplified structure that uses deeply differentiated conversations to create light weight in deep neural networks. The SSD models used by MobileNet for object detection are simple so that they cannot be used on mobile devices and can be operated in real time.

B. Faster-RCNN with InceptionV2

The Single Shot Multi-Box Detector (SSD) is a method for detecting an object in images using a deep neural network. It produces binding boxes and class scores for something obtained faster than the previous method such as You Only Look Once (YOLO). On the other hand, MobileNet is an in-depth learning model that can be used for a variety of recognition functions such as object detection, finegrain detection, remarkable recognition, and facial expressions. The benefits of MobileNet are accurate, fast, minimal and easy to tune. It is based on a simplified structure that uses deeply differentiated conversations to create light weight in deep neural networks. The SSD models used by MobileNet for object detection are simple so that they cannot be used on mobile devices and can be operated in real time.

VI. FLOWCHART

1. <u>Capture Image</u>: - Image Capture refers to a process in which an electronic device resembles a scanner used to create digital image representation.
2. <u>Image Acquisition</u>: - Described as the act of retrieving an image from a source, it is usually a Hardware-based source for processing the object.
3. <u>Image Pre-processing</u>: - It is a common term for image processing at a very low level of both input and output is the intensity of the acquisition images.
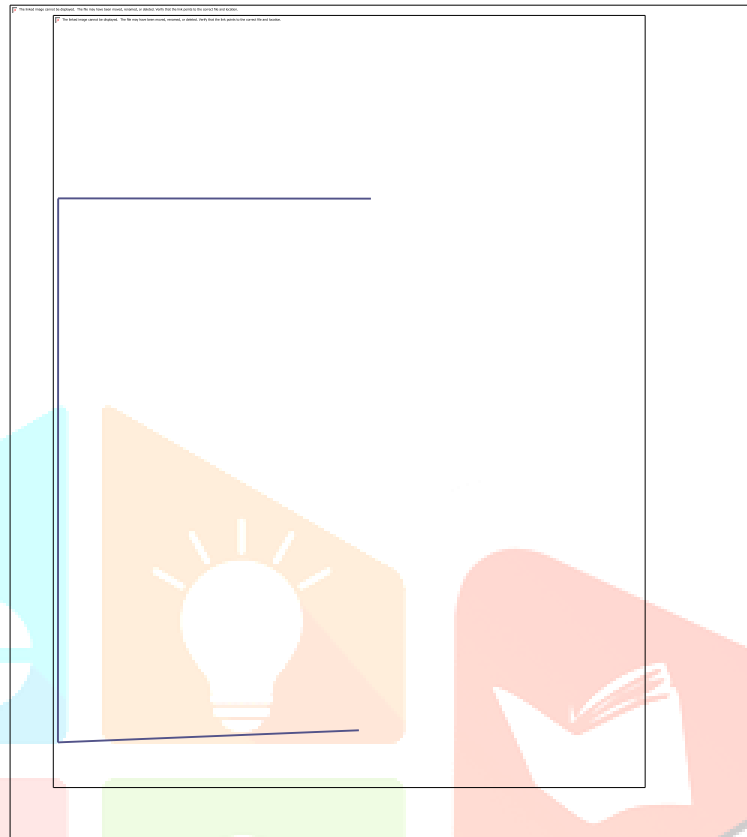
Fig. 2. Flowchart

4. <u>Image Segmentation</u>: - It is a process of dividing a digital image into several segments to make it easier to analyze an image.
5. <u>Feature Extraction</u>: - It is one of the most researched areas in the field of image analysis as it is a prerequisite for representation.
6. <u>Classification</u>: - Refers to the task of extracting object information classes from raster multiband in the image.
7. <u>Show Result</u>: - Displays the output on the system screen.

VII. RESULT AND DISCUSSION

First do "Right Click" then click on "Open PowerShell Window here" as shown in fig.3.

Fig. 3. Open WindowsPowershell

Then type "file_name.py" and then it will simultaneously detect the objects and show the data sets as shown in Fig.4 & Fig.5 & Fig.6.

Fig. 4. Powershell Window Screen
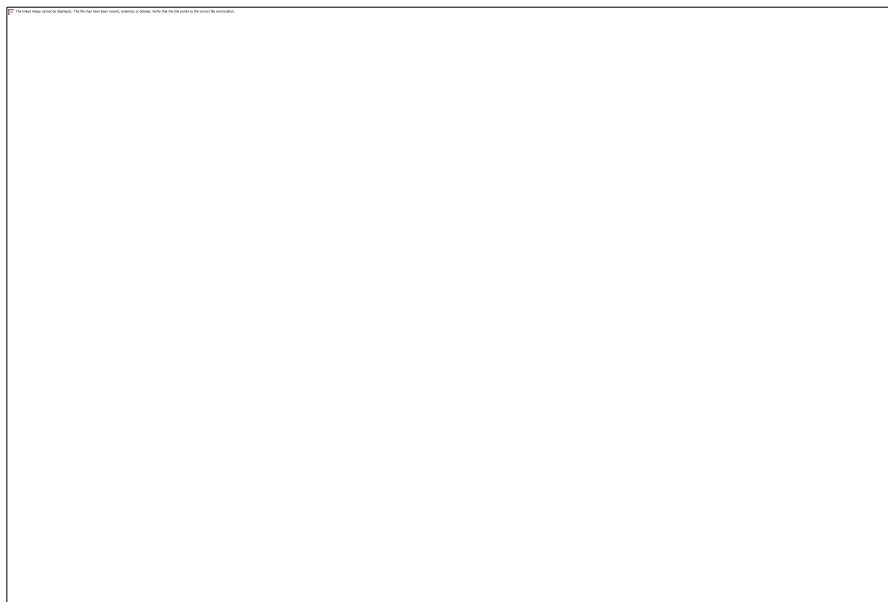

Fig. 5. Datasets of Objects

Fig. 6. Detection of Book

Fig.7. Represents successful detection of multiple objects

Fig. 7. Detection of Multiple Objects

VIII. FUTURE SCOPE

The object recognition system is often applied within the space of closed-circuit television, face recognition, fault detection, character recognition etc. the target of this thesis is to develop Associate in Nursing object recognition system to acknowledgement the 2D and 3D objects within the image. The performance of the item recognition system depends on the options used and also the classifier used for recognition.

IX. CONCLUSION

The discovery potential of both art models on CNN has been demonstrated successfully. It shows that the SSD with MobileNetV1 has high speed detection but low accuracy compared to Faster-RCNN with InceptionV2 which is lower speed but more accurate. Based on test results, there is a trade-off between accuracy and speed.

X. REFERENCES

[1]   V. Gajjar, A. Gurnani and Y. Khandhediya, "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach," in 2017 IEEE International Conference on Computer Vision Workshops, 2017.

[2]   M. Adel, A. Moussaoui, M. Rasigni, S. Bourennane and L. Hamami, "Statistical-Based Tracking Technique for Linear Structures Detection: Application to Vessel Segmentation in Medical Images," IEEE Signal Processing Letters, vol. 17, no. 6, pp. 555-558, June 2010.

[3]   X.-T. Truong, V. N. Yoong and T.-D. Ngo, "RGB-D and Laser Data Fusion-based Human Detection and Tracking for Socially Aware Robot Navigation Framework," in IEEE Conference on Robotics and Biomimetics, Zhuhai, China, 2015.