REAL-TIME MULTIPLE OBJECT RECOGNITION

Pradeep Doss[1], Aditya Rai[2], Shashank Rai[3]

[1]Assistant Professor, Dept. of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.

[2]Student, Dept. of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.

[3]Student, Dept. of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.

ABSTRACT

This paper manages the field of computer vision, mostly for the execution of ongoing article acknowledgment utilizing TensorFlow Object Detection API. TensorFlow Object Detection API is an open source structure based over TensorFlow, it makes it simple to build, prepare and convey Object Detection Models. The API likewise gives an accumulation of Detection Models pre-prepared on Microsoft COCO dataset. This paper will manage one of the Detection Models which is the mix of Single Shot Detectors (SSD) and MobileNets design, it is quick, proficient and does not require tremendous calculation to achieve the protest acknowledgment assignment.

Keyword: - *Real-time Object Recognition ,TensorFlow Object Detection API , COCO, Single Shot Detector, MobileNets, Detection Models*

1. INTRODUCTION

Object Recognition is a standout amongst the most essential applications in the field of computer vision. As of late, with the quick advancement of profound taking in, various research regions have accomplished great outcomes, and joined by the ceaseless change of convolution neural network, computer vision has touched base at another pinnacle. Convolution neural system has numerous applications in field of computer vision, for example, confront acknowledgment, question identification, protest following, semantic division, et cetera. Late advances in question identification are driven by the accomplishment of district proposition technique and locale based convolutional neural network (RCNNs).

Contrasted with picture order, protest discovery is an all the more difficult errand that requires more intricate strategies to explain. Because of this many-sided quality, current methodologies prepare models in multi-arrange pipelines that are moderate and inelegant. Question acknowledgment has developed from single protest location in a picture to various question recognition, and now we will perceive different protests progressively. TensorFlow Object Detection API utilizing Single Shot Detectors and MobileNets design and prepared on COCO dataset will enable us to accomplish continuous protest acknowledgment. One of the essential objectives of constant question acknowledgment is comprehension of visual scenes. Scene understanding includes various undertakings including perceiving what objects are available, restricting the items in 2D and 3D, deciding the articles' and scene's qualities, describing connections among articles and giving a semantic portrayal of the scen

1.1 TensorFlow

Released on the 15th of November 2015 by Google, TensorFlow of the most recent open source library written in Python for numerical computation. The reason to choose TensorFlow for this task was TensorFlow has great success in the Machine Learning community and in less than three years it also had a lot of support and development by Google itself, more over by many community projects, developed in any area of Deep Learning. The main advantage of TensorFlow is use of data flow graphs. Where nodes represent mathematical operations, edges represent the multidimensional data arrays called Tensors. Under the TensorFlow project, Google has developed an API called TensorFlow Object Detection API, this API can be used to detect, with bounding boxes, objects in images and/or video using either some of the pre-trained models made available or through models you can train on your own.

1.2 Microsoft COCO: Common Objects in Context

Microsoft presents a new dataset COCO with the goal of advancing the state-of-the-art in object recognition, by placing the question of object recognition in context of scene understanding. This is achieved by gathering images of complex everyday scenes containing common objects in their natural context. Objects are labeled using per-instance segmentations to aid in precise object localization. COCO claims to contain photos of 91 object types that even a 4 year old can recognize. With a total of 2.5 million labeled instances in 328k images.

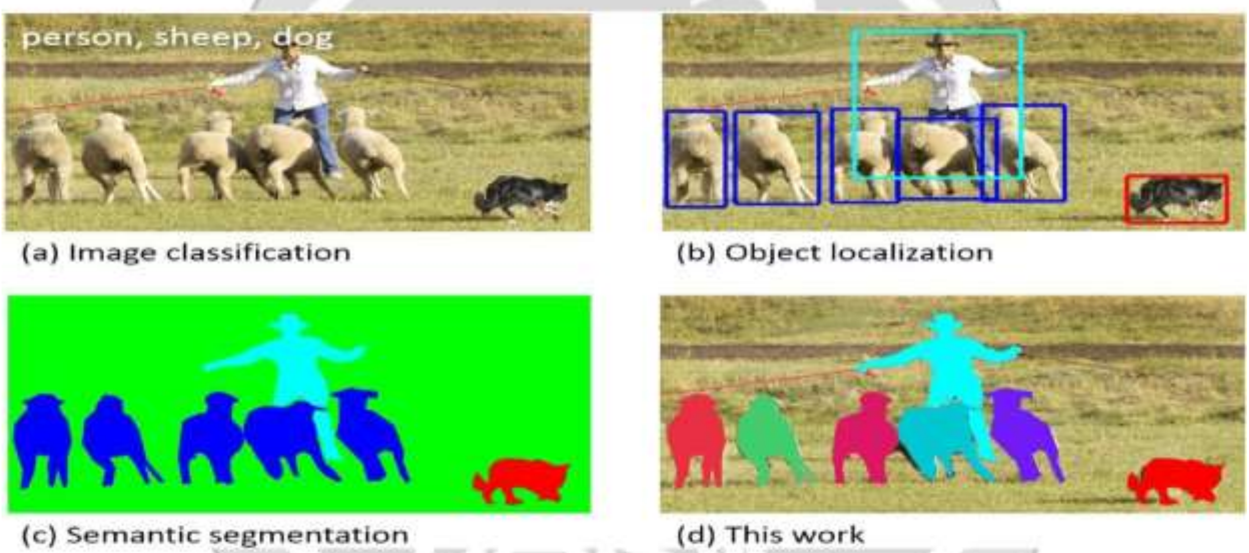

Fig -1: Other datasets have attempted (a) image classification, (b) object localization, (c) semantic segmentation Microsoft COCO on (d) segmenting individual object instances.

1.3 Fast R-CNN

Fast R-CNN proposes a new training algorithm that fixes the disadvantages of R-CNN and SPPnet, while improving on their speed and accuracy. Fast R-CNN is fast to train and test. This method has several advantages such as higher detection quality (mAP) than R-CNN, SPPnet, training is single-stage using a multi-task loss, training can update all network layers, no disk storage is required for feature caching. This algorithm is written in Python and C++ (Caffe). A Fast R-CNN network takes an entire image and a set of object proposals as input. The network then first processes the whole image with several convolutional (conv) and max pooling layers to produce a conv feature map. For each object proposal a RoI pooling layer extracts a fixed-length feature vector from the feature map.

Two sibling output layer is achieved when each feature vector is fed into a sequence of fully connected layers. One produces softmax probability estimates over K object classes and another outputs four real-valued numbers for each of the K object classes. Each set of 4 values encodes refined bounding-box positions for one of the K classes.

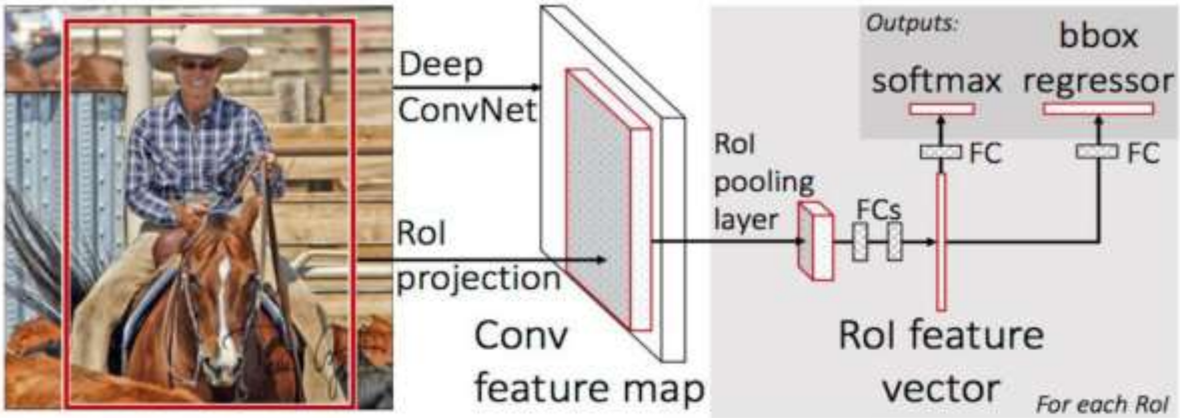

Fig -2: Fast R-CNN Architecture

1.4 Faster R-CNN

Microsoft presents a new dataset COCO with the goal of advancing the state-of-the-art in object recognition, by placing the question of object recognition in context of scene understanding. This is achieved by gathering images of complex everyday scenes containing common objects in their natural context. Objects are labeled using per-instance segmentations to aid in precise object localization. COCO claims to contain photos of 91 object types that even a 4 year old can recognize. With a total of 2.5 million labeled instances in 328k images.

1.5 skLearn

The scikit-learn project started as scikits.learn, a Google Summer of Code project by David Cournapeau. Its name stems from the notion that it is a "SciKit" (SciPy Toolkit), a separately-developed and distributed third-party extension to SciPy. The original codebase was later rewritten by other developers. In 2010 Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort and Vincent Michel, all from INRIA took leadership of the project and made the first public release on February the 1st 2010. Of the various scikits, scikit-learn as well as scikit-image were described as "well-maintained and popular" in November 2012.

2. IMAGENET

The ImageNet project is a large visual database designed for use in visual object recognition software research. Over 14 million URLs of images have been hand-annotated by ImageNet to indicate what objects are pictured; in at least one million of the images, bounding boxes are also provided. ImageNet contains over 20 thousand ambiguous categories; a typical category, such as "balloon" or "strawberry", contains several hundred images. The database of annotations of third-party image URL's is freely available directly from ImageNet; however, the actual images are not owned by ImageNet. Since 2010, the ImageNet project runs an annual software contest, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), where software programs compete to correctly classify and detect objects and scenes. The ImageNet Challenge uses a "trimmed" list of one thousand unambiguous classes.

3. IMPLEMENTATION

This paper communicates the significance of profound learning innovation applications and the effect of dataset for profound learning using the speedier r-cnn on new datasets. As of late, the innovation of profound learning in picture arrangement, protest recognition and face

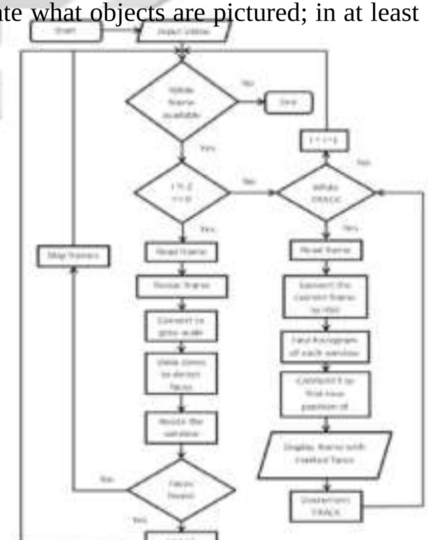
Fig -3: Data Flow Diagram

recognizable proof and numerous other PC vision errands have made awesome progress. Trial information demonstrates that the innovation of profound learning is a viable instrument to pass the man-made component depending on the drive of understanding to the getting the hang of depending on the drive of information. Substantial information is the base of the accomplishment of profound adapting, extensive information similarly as fuel to the rocket for profound learning. An ever increasing number of uses are persistently aggregating progressively rich application information, which is basic to the further improvement and utilization of profound learning. In any case, the nature of the information influences the profound learning in deed, obviously, notwithstanding these genuine information, possibly we can likewise consider some of engineered information to expand the measure of information in the further.

4. CONCLUSIONS

This paper expresses the importance of deep learning technology applications and the impact of dataset for deep learning through the use of the faster r-cnn on new datasets. In recent years, the technology of deep learning in image classification, object detection and face identification and many other computer vision tasks have achieved great success. Experimental data shows that the technology of deep learning is an effective tool to pass the man-made feature relying on the drive of experience to the learning relying on the drive of data. Large data is the base of the success of deep learning, large data just as fuel to the rocket for deep learning. More and more applications are continually accumulating increasingly rich application data, which is critical to the further development and application of deep learning. However, the quality of the data affects the deep learning in deed, of course, in addition to these real data, maybe we can also consider some of synthetic data to increase the amount of data in the further.

5. REFERENCES

1. A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS,2012.
2. M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. In ECCV,
3. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.
4. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CVPR, 2016.
5. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in CVPR, 2014.
6. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei. ImageNet: A large-scale hierarchical image database. In CVPR, 2009.
7. J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. FeiFei. ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012).
8. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results, 2007.
9. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. IJCV, 2010.
10. Lin, Tsung Yi, et al. Microsoft COCO: Common Objects in Context. Computer Vision – ECCV 2014. Springer International Publishing, 2014:740-755.