International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 03 - Issue 05 || May 2018 || PP. 38-45

Comparative Analysis of Variegated Pre-Trained Models for Discrete Class-Labels

Keyur J. Thakkar
Information Technology, Charotar University of Science & Technology, India

Proff. Sonal P. Rami
Information Technology, Charotar University of Science &Technology, India

Abstract: Image classification is a sub-domain of Image detection; several methods are now proposed to detect particular sectors of an image. In these methods, Image classification is a supervised method to detect some peculiar parts of the image. Basically in a detection part you train a system to catch one or many class of objects (for instance, faces) according to the requirement and the approach would be easily found using supervised methods.

Classification itself is a vast term which includes training of data, testing of data , supervised as well as unsupervised learning. Also, concepts of Machine Learning and Deep Learning plays a vital role in classifying an image. Suppose that you can train your system to sort your images into different categories for eg, does a scene refer to a landscape or to an urban scene, to people etc. In that sense, object detection would be a classification reduced to two classes: the object, not the object.

General Terms: Classification, Identification, Image Detection, Recognition
Keywords: COCO, DataSets, Kitti, Pre-Trained Models, Tensorflow

I. Introduction

Gigantic masses of digital visual information are produced nowadays, both automatically and also by the people with access to easily usable tools for creating personal digital content online. Automatic image analysis techniques are called to analyse and organise these overwhelming sea of information.Particularly valuable would be techniques that could consequently examine the semantic substance of pictures and recordings as it is only the substance that decides the pertinence in a large portion of the images/recording. One important aspect of image classification is the object composition: the identities and positions of the objects the images contain. This paper discusses techniques for recognising and locating objects of some particular semantic class in images when the available test images are passed through some qualifies coding.

Quite often data analysis researchers work with data sets of their own. The problem is that the individual researchers try to solve or tries to get the effectiveness using their individual data sets. However, due to different data sets the relative potency of various approaches may be difficult to compare as the outcome may vary vividly.

Essentially object recognition is kind of quest for finding/identifying objects in an image or a video string and classifying the detected objects.

Object Detection & Image Classification formally includes:

- **Detection**– of separate objects
- **Description**– of their geometry and positions in 3D
- **Classification**– as being one of a known class
- **Identification**– of the particular instance
- **Understanding**-of spatial relationships between objects

II. Foundation Theory

Considerable lot of research demonstrated various approaches to enhance neural neural network based image classification systems. They initially indicate some new useful image transformations to extend the effectual size of the training set[1]. These were based on using more of the image to select training crops and additional color manipulations. They additionally demonstrated useful image transformation for creating testing predictions. They made expectations at various scales and created forecasts on various perspectives of the image[1].

Moreover few outcomes demonstrate that a large, deep convolutional neural network is capable of achieving record breaking results on a highly challenging dataset utilizing absolutely supervised learning[2]. It is denoted that their system's execution will scatter if a solitary convolutional layer is removed. For eg, expelling any of the center layers brings about lost around 2% for the top performance of the system. So the profundity truly is essential for accomplishing the results.[2]

Besides a framework has proposed a picture obscured image classification & recognition strategy that can naturally recognize obscured picture and reorders the image without either picture sharpening or part estimation.[3] They develop another blur metric: particular esteem highlight, and utilize it to distinguish the blurred part of a picture. They likewise dissect the alpha channel data and order the obscured picture locales into defocus blur or motion blur, respectively.[3]

Powerful rundown of an accumulation is described by some vital properties that the synopsis ought to possess[4]. In light of the work included, following properties were discovered much self-evident.

1) **Diversity**i.e Images in the summary should not be identical to each other by any means.
2) **Coverage**i.eThe summary should cover all fascinating and critical visual and semantic angles. Visual and semantic viewpoints with high probability ought to be available in the synopsis.
3) **Balance**i.eThe optical aspects should be present in a balanced way to avoid any variance. Other than this, the outline ought to be inadequate in the quantity of images[4].

III. Various Pre-Trained Models

There are piles of pre-trained detection models available on the COCO Dataset, the Kitti Dataset, and the Open Images Dataset. These models can be appropriate for intellection inference if you are interested in categories rendered by COCO (e.g., humans, cars, etc) or in Open Images (e.g., surfboard, jacuzzi, etc). They are also useful for initializing your models when training on novel datasets.

3.1 COCO

Common Objects In Context i.e MS COCO dataset contains photos of different 91 categories observed in our day to day life which could be clearly acknowledged by an infant. This data set has over 2 million images labeled manually. With over 25 lakhs labeled instances of more than 327K[8] images, the making of their dataset developed upon broad group labors input by means of novel UIs for class recognition, case spotting and occasion segmentation[8].

Model Name	Speed(ms)	Mean Average Precision	Output
ssd_mobilenet_v1_coco	30	21	Boxes
ssd_mobilenet_v2_coco	31	22	Boxes
ssd_inception_v2_coco	42	24	Boxes
faster_rcnn_inception_v2_coco	58	28	Boxes
faster_rcnn_resnet50_coco	89	30	Boxes
faster_rcnn_resnet50_lowproposals_coco	64		Boxes
rfcn_resnet101_coco	92	30	Boxes
faster_rcnn_resnet101_coco	106	32	Boxes
faster_rcnn_resnet101_lowproposals_coco	82		Boxes
faster_rcnn_inception_resnet_v2_atrous_coco	620	37	Boxes
faster_rcnn_inception_resnet_v2_atrous_lowproposals_coco	241		Boxes
mask_rcnn_inception_resnet_v2_atrous_coco	771	25	Boxes
mask_rcnn_inception_v2_coco	79	33	Masks
mask_rcnn_resnet101_atrous_coco	470	33	Masks
mask_rcnn_resnet50_atrous_coco	343	29	Masks

3.2 Kitti

ModelName	Speed(ms)	Mean Average Precision	Output
faster_rcnn_resnet101_kitti	79	87	Boxes

3.3 Open Images

The Open Images data set has been released by Google containing millions of annotated images which you can now use to train your own machine learning models using the same data Google has access to to train its own models.

ModelName	Speed(ms)	Mean Average Precision	Output
faster_rcnn_inception_resnet_v2_atrous_oid	727	37	Boxes
faster_rcnn_inception_resnet_v2_atrous_lowproposal_oid	347		Boxes

IV. Experimental Results

In here using different types of datasets which runs over a code in python in the SPYDER IDE, various outputs have been measured and the difference between the Accuracy and the Number of objects detected have been captured to provide a sure short comparative analysis of the used Data Sets.

4.1 Datasets & Class Labels

Distinct 4 types of datasets have been chosen particularly to point out the difference found in **1)Time Taken for Execution 2) Accuracy** and **3) Number of objects detected**, over the selected Test Images Set. Those DataSetsare :-

1) SSD MOBILENET VERSION 1 COCO

2) SSD INCEPTION VERSION 2 COCO

3) FASTER RCNN INCEPTION VERSION 2 COCO

4)MASK RCNN INCEPTION RESNET VERSION 2 ATROUS COCO

Here the Class Labels used are labelled on the bases of 80 object categories and 91 stuff categories observed in our routine life i.e animals, motorcycle, plants, furniture, vehicles, household items etc.

4.2 Result

Using the above mentioned datasets one by one over couple of randomly generated test images following results and differences were observed.

4.2.1 Figure 1

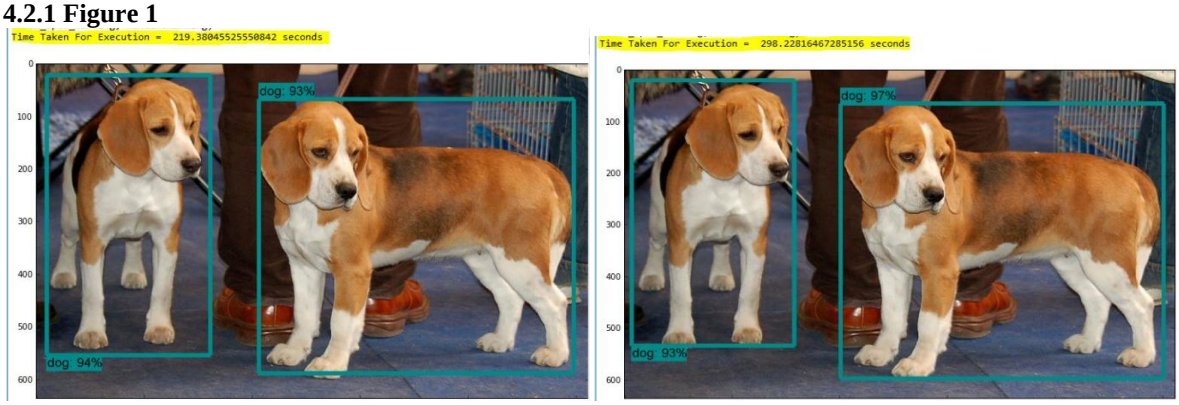

Fig. 1.1 ssd_mobilenet_v1_cocoFig. 1.2 ssd_inception_v2_coco

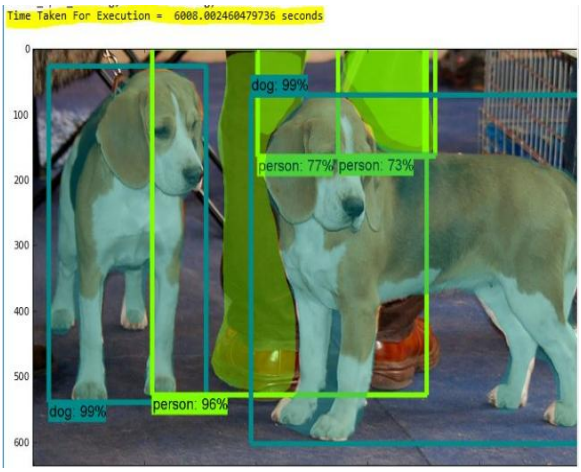

Fig. 1.3 faster_rcnn_inception_v2_cocoFig. 1.4 mask_rcnn_inception_resnet_v2_atrous

Observing the following images, one can easily classify the difference between those images. Furthermore to make it effortless, have a glance at the Table below.

Disclaimer: The Time Taken For Execution column here shows the time taken by the dataset to provide the output classification of the entire Test Images consisting of 27 Images.

Model Name	**Time Taken for Execution (sec)**	**Highest Accuracy**	**No. of objects detected**
ssd_mobilenet_v1_coco	219.58	94%	2
ssd_inception_v2_coco	298.22	97%	2
faster_rcnn_inception_v2_coco	420.71	99%	3
mask_rcnn_inception_resnet_v2_atrous	6008.02	99%	5

4.2.2 Figure 2

Fig. 2.1 ssd_mobilenet_v1_coco

Fig. 2.2 ssd_inception_v2_coco

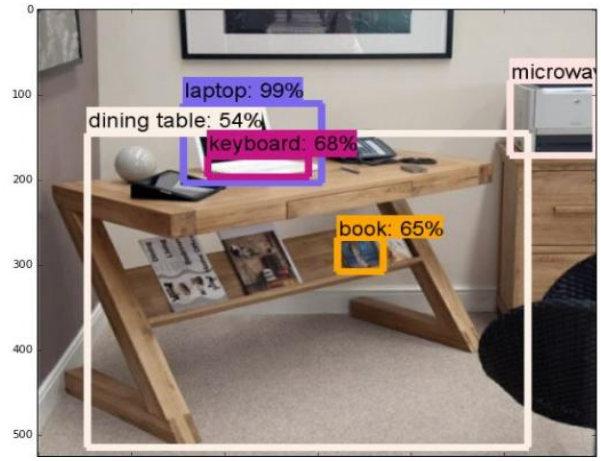

Fig. 2.3 faster_rcnn_inception_v2_coco

Fig. 2.4 mask_rcnn_inception_resnet_v2_atrous

Model Name	No. of objects detected
ssd_mobilenet_v1_coco	1
ssd_inception_v2_coco	1
faster_rcnn_inception_v2_coco	5
mask_rcnn_inception_resnet_v2_atrous	3

4.2.3 Figure 3

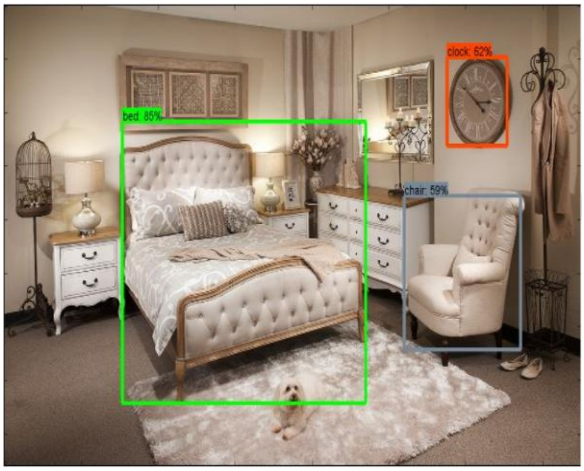

Fig. 3.1 ssd_mobilenet_v1_coco

Fig. 3.2 ssd_inception_v2_coco

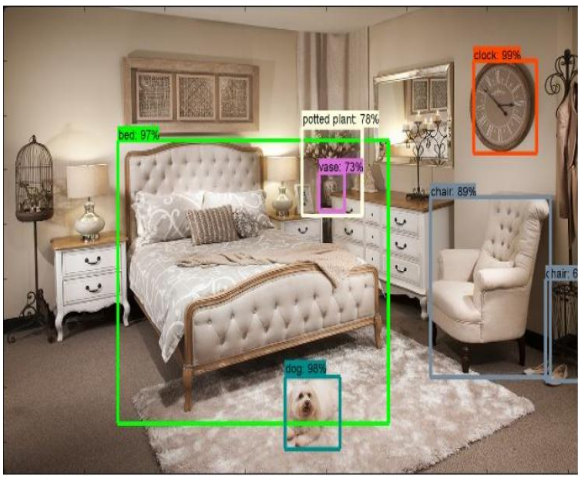

Fig. 3.3 faster_rcnn_inception_v2_coco

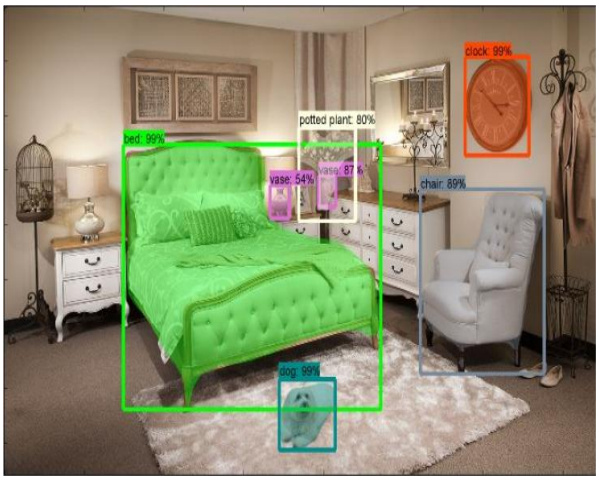

Fig. 3.4 mask_rcnn_inception_resnet_v2_atrous

Model Name	No. of objects detected
ssd_mobilenet_v1_coco	3
ssd_inception_v2_coco	3
faster_rcnn_inception_v2_coco	6
mask_rcnn_inception_resnet_v2_atrous	7

4.2.4 Figure 4

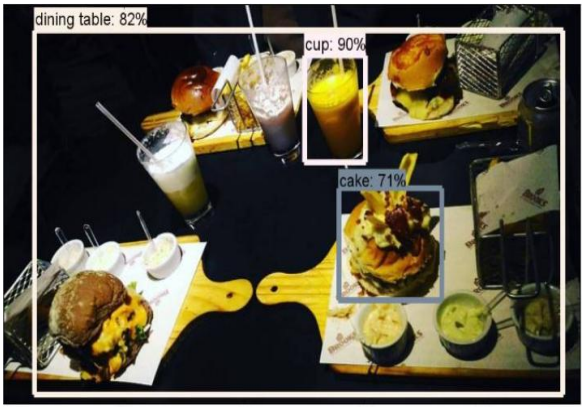

Fig. 4.1 ssd_mobilenet_v1_coco

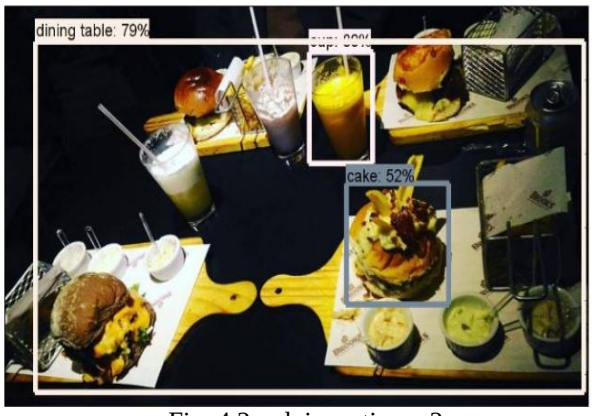

Fig. 4.2 ssd_inception_v2_coco

Fig. 4.3 faster_rcnn_inception_v2_coco

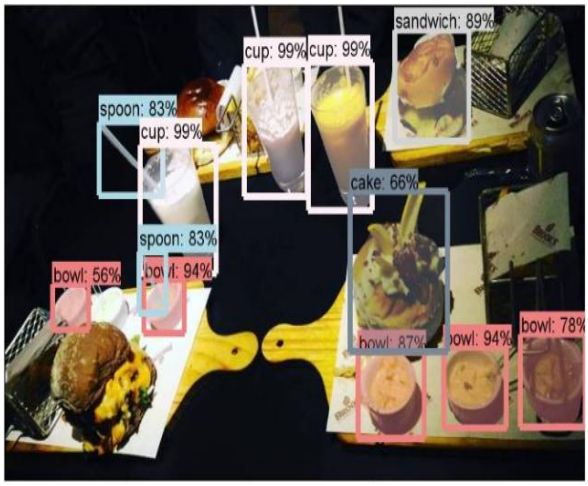

Fig. 4.4 mask_rcnn_inception_resnet_v2_atrous

Model Name	No. of objects detected
ssd_mobilenet_v1_coco	3
ssd_inception_v2_coco	3
faster_rcnn_inception_v2_coco	10
mask_rcnn_inception_resnet_v2_atrous	11

V. Conclusion

All the projects that have been working on image classification and object recognition have just used the coco dataset i.e.**ssd_mobilenet_v1_coco** and no proceedings and findings were made about how many different datasets can also be used to find the discrete variance in the output of the images regarding Time of Execution, Accuracy and No. of Objects detected. Here we have used 4 distinct models out of all the above mentioned datasets and provided the findings.

References

Journal Papers:

[1] Howard, Andrew G."Some improvements on deep convolutional neural network based image classification." arXiv preprint arXiv:1312.5402(2013)

[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[3] Su, Bolan, Shijian Lu, and Chew Lim Tan. "Blurred image region detection and classification." Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011.

[4] Mahajan, Dhruv, et al. "A classification based framework for concept summarization." Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, 2012.

[5] Sonntag, Daniel, et al. "Fine-tuning deep CNN models on specific MS COCO categories." arXiv preprint arXiv:1709.01476 (2017)

[6] Käding, Christoph, et al. "Fine-tuning deep neural networks in continuous learning scenarios." Asian Conference on Computer Vision. Springer, Cham, 2016

[7] https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md

[8] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014

[9] https://github.com/tensorflow/models/tree/master/research/object_detection/samples/configs

[10] Wilson, D. Randall, and Tony R. Martinez. "The general inefficiency of batch training for gradient descent learning." Neural Networks 16.10 (2003): 1429-1451.

[11] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In International Conference on Computer Vision, pages 2146–2153. IEEE, 2009.

[12] A. Krizhevsky. Learning multiple layers of features from tiny images. Master's thesis, Department of Computer Science, University of Toronto, 2009.

[13] A. Krizhevsky and G.E. Hinton. Using very deep autoencoders for content-based image retrieval. In ESANN, 2011.

[14] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, et al. Handwritten digit recognition with a back-propagation network. In Advances in neural information processing systems, 1990.

[15] Y. LeCun, F.J. Huang, and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, pages II–97. IEEE, 2004.

[16] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pages 253–256. IEEE, 2010

[17] SupriyaDeshmukh, Leena Ragha, "Analysis of Directional Features - Stroke and Contour for Handwritten Character Recognition", IEEE International Advance Computing Conference, pp.1114-1118, 6-7 March, 2009,India

[18] AmrithaSampath, Tripti C, Govindaru V, Freeman code based online handwritten character recognition for Malayalam using Back propagation neural networks, Advance computing: An international journal, Vol. 3,No. 4, pp. 51-58, July 2012.