A Literature Survey on Classification of Images using State of the Art Machine Learning Techniques

Dr. Leena Patil[1], Miss. Anagha M. Pawar[2]
[1]*Professor, CSE Department,* [2]*Final Year M.Tech Student*

Abstract- Classification of objects into respective insightful groups have always been a very popular task in the area of machine learning and statistics. Yet another popular area of research these days is image processing and graphical learning. Most of the tasks in these domains of machine learning are always complex, seems conundrums in terms of classification objectives. There is a popular solution to these hard problems called artificial neural networks (ANNs). ANNs are always in discussions because of their complex structure and vast usefulness. In this project I will make use of convolutional neural network study to predict and classify different objects with the help of state of the art technology called TenserFlow. Convolutional neural network (CNNs) is the competitive image recognition method which has been deeply exploited in the last years. It essentially mimics the local receptive capacity of the neurons as seen in the human brain, with sophisticated weight sharing and the linkage methods, significantly decreases the training constraints in comparison with traditional neural network approach.

Keywords- Classification, ANNs, MobileNets, TenserFlow, Convolutional Neural Networks.

I. INTRODUCTION

With an advancement in the technology, the neural network has brought great convenience to people, where lots and lots of information have to be processed either in the textual, image, audio or video formats However, the information security has always been a concern in this technological development. There are many use cases around the efficient identification technologies in the areas of banking, e-commerce, social networks, and government activities. In addition to this image recognition problem, which I will be focusing on this project will have main advantage over to the specific use case such as the disharmony caused by terrorists in this world. Objects in the images can be predicted from the computer system which can extract the visual ciphers from the group of objects in the image and can assign the prediction scores to each of the identified objects in the particular image.
we can link the problem of object recognition to the problem of flower classification used in the paper [1]. Where the author made the use of CNNs to achieve higher accuracy in image classification. Complex use of CNNs have also been observed in the research works of [2, 3, 4, 5, 8], which truly reflects the utility value of the CNN in the image recognition domain. However, the complex structure of the model doesn't necessarily make it efficient, with respect to execution speed and the confidence interval of the classification accuracy values. Main motive of utilizing the CNNisto do the computationally intense task of image detection and classification with the limited resources resulting low latency in execution and higher prediction accuracy as suggested by [2, 3, 4].
Similarly, we can link the problem of object recognition with the recent increase in popularity of mobile devices, many apps have been introduced to record and analyze food intake automatically. These applications, mostly aimed at automatically assessing the nutritional content of a meal, typically require that the user take one (or more) image(s) of the meal using a mobile device's camera. These images are then analyzed by the system to automatically monitor diet and advice users about more healthy choices. One can describe food image analysis as a chain of three main tasks: image segmentation, food recognition, and quantity estimation. Most methods proposed for automatically analyzing food images include a step in which the type of food displayed in the image is automatically recognized. In fact, food recognition is probably the most popular topic in the literature about automatic diet monitoring.
Taking, inspiration from the above use cases of flower identification or the diet plan identification, in this effort of reviewing the papers, we will be studying the competent network design as shown in the paper [1]. We will also make study how to use small set of hyper-parameters in order to keep my predictive model small and portable to different execution environments over limited computational resources.

II. LITERATURE SURVEY

In the conventional image recognition methods, CNNs [7-6] uses the multilayer convolution to do the feature engineering and combine these features internally. It also uses the pooling layer, and the fully connected layer along with softmax. Google's TenserFlow an open source library [1-7] can be used for the arithmetic calculations, specialized for machine learning computations. Recently launched second generation of the Google's artificial intelligence learning system got more views and appreciation in the scientific community in the field of machine learning from all over the world. TenserFlow has many advantages such as high flexibility, high accessibility, and the great provision of coders and TenserFlow practitioners on the GithHub makes this implementation task smoother.

One of the traditional methods of image recognition is the use of elastic graphs, in which the grid structure is used to create templates with which different images are compared within the bound of grids. In this approach specifically the recognition performance is better, compromising the execution speed and execution time.There has been several other investigations on object recognition using arithmetic mean based Hidden Markov Model (HMM) techniques where the probabilistic estimation of the single value feature of objects are transformed into vectors and then HMM is used on the top to predict the image. There is an another method as well in the scope of machine learning optimization called the Eigenface method, which is based on Principal Component Analysis (PCA), using PCA transform to reduce the original image processing, and then to classify and identify. Alsothe Fisherface method based on Linear Discriminant Analysis (LDA), the features of the image are reduced and then LDA is used to transform and extract the features of the principal components after the dimension reduction. It is expected that large inter-class divergence and small intra-class divergence.

Talking about the MobileNets [3] is one of the already trained models on the TensorFlow. It is regarded as routine development to the initial structure of computer graphics learning after Inception-v1 [6], Inception-v2 [5], Inception-v3 [4] in the year 2015. The MobileNet [3] model is trained on the popular ImageNet datasources, consisting of the 1000s of categories in ImageNet, the fault percentage of top-5 is upto 3.5% leaving behind average accuracy portion of ~90% in desired cases, the fault percentage of top-1 dropped to 17.3% [2] [12]. Following section we will be discussing the different state of the art convolutional neural learning approaches.

III. TYPES OF CONVOLUTIONAL NEURAL TRAINING

In this section we will be analyzing different state of the art convolutional neural networks, and later in the last section we will be discussing their performances.

A. Region based convolutional neural network. (R-CNN)

In R-CNN, the **selective search** method is suggested for effective image recognition by [13]. This method is strong alternative to the already existing exhaustive search in an image to capture object location. In this approach author initialized small regions in an image and merges them with anordered grouping set. Ultimately the final grouping set contains the overall image.

Fig.1: Selective Search application, top: visualisation of the segmention results of the algorithm, down: visualization of the region proposals of the algorithm. [13]

The CNN model explained by the authors in [13] is trained on the 2012 ImageNet dataset of the original challenge of image classification tasks. It is then adjusted using the region proposals corresponding to an IoU>0.5 with the help of ground-truth boxes. The best R-CNNs models have a 62.4% mean average precision score over the PASCAL VOC 2012 test dataset.

B. Fast region based convolutional neural network (Fast R-CNN)

The main functionality of the fast-region based CNN (Fast R-CNN) is explained by [14] is to reduce the time complexity related to the high number of model complexities necessary to analyze all region proposals in the training model. In Fast region CNN the multiple convolutional layers are taking the entire image as input fed to the training phase instead of using a CNN for each region (R-CNN). The best Fast R-CNNs have reached mean average precision scores of 70.0% for the 2007 PASCAL VOC test dataset.

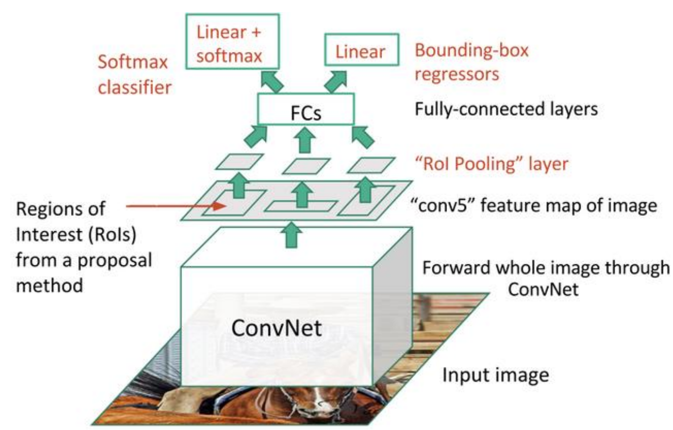

Fig.2: The entire image feeds a CNN model to detect RoI on the feature maps training phase [14].

C. You only look once (YOLO)

The YOLO i.e. you only look once predictive model [15]helps in directly predicting the underlying bounding boxes and inter class probabilities with a singular network working under singular evaluation. This simplicity of the YOLO model allows real-time predictions.Initially the model takes an input Y. It divides input Y into an YxY grid. Each cell of this input grid predicts B number of different bounding boxes with anassociated confidence score. This confidence is simply the probability to detect the object multiply by the intersection over union between the predicted and the ground truth boxes.

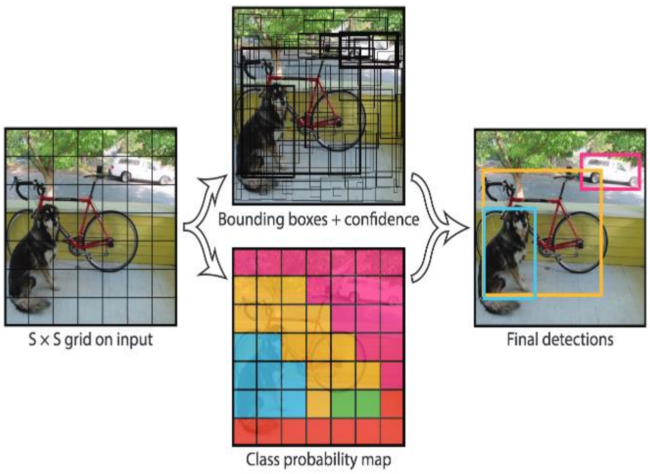

Fig.3: Example of application of YOLO[15].

IV. PERFORMANCE ANALYSIS

Performance metrics as given by [15], following table represents the performance analysis given by mean average precision parameter over real time systems on PASCAL VOC 2007 dataset. Comparison on computation speed is also given below.

Table-1 Precision performance analysis over CNN algorithms.

Model	Mean Average Precision	Frames Per Second
R-CNN	62.4%	0.5
Fast R-CNN	70%	0.5
YOLO	63.4%	45

V. CONCLUSION

There are many state of the art techniques to achieve efficient image classification and object recognition. We have studied the effectiveness of the convolutional neural networks and their subordinate advantages achieving the classification task. Out of them the convolutional neural network with YOLO approach out par the current static processing paradigm of image classification rendering the real time streaming analysis. Considering accuracy performance metrics fast region-based CNN serves the best among all with ~10% more prediction rate than R-CNN and YOLO algorithms.

VI. REFERENCES

[1]. Gavai, N.R., Jakhade, Y.A., Tribhuvan, S.A. and Bhattad, R., 2017, December. MobileNets for flower classification using TensorFlow. In *Big Data, IoT and Data Science, 2017 International Conference on* (pp. 154-158). IEEE.

[2]. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[3]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. “Rethinking the inception architecture for computer vision”. arXiv preprint arXiv:1512.00567, 2015.

[4]. C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261, 2016

[5]. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.

[6]. Angelova A, Zhu S, Efficient Object Detection and Segmentation for Fine-Grained Recognition[C]. Computer Vision and Pattern Recognition: 8 1 1-818, IEEE, 2013.

[7]. Peter Goldsborough FakultätfürInformatik, “A Tour of TensorFlowarXiv:1610.01178v1 [cs.LG] 1 Oct 2016.

[8]. Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M. and Cagnoni, S., 2016, October. Food image recognition using very deep convolutional networks. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 41-49). ACM.

[9]. K. Jarrett, K. Kavukcuoglu, Y. Lecun, et al. What is the best multi-stage architecture for object recognition? In 2009 IEEE 12th International Conference on Computer Vision, pages 2146–2153, Sept 2009.

[10]. Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. Nature, 521(7553):436–444, 2015.

[11]. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 25, pages 1097–1105. Curran Associates, Inc., 2012.

[12]. S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou. Food recognition for dietary assessment using deep convolutional neural networks. In New Trends in Image Analysis andProcessing–ICIAP 2015 Workshops, pages 458–465. Springer, 2015.

[13]. Girshick, R., 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

[14]. Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

[15]. Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[16]. https://medium.com/comet-app/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852