Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020)
IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108-4

Transfer learning-based Object Detection by using Convolutional Neural Networks

Ms Bulbul Bamne
Department of Computer Science
Patel College of Science & Technology, Indore, India
bamne.bulbul@gmail.com

Ms Neha Shrivastava
Assistant Professor
Department of Computer Science
Patel College of Science & Technology, Indore, India
nehashriva@gmail.com

Mr Lokesh Parashar
Assistant Professor
Department of Computer Science
Patel College of Science & Technology, Indore, India
lokesh23324@gmail.com

Mr Upendra Singh
Assistant Professor
Department of Information Technology
Shri Govindram Seksaria Institute of Technology and Science, Indore, India
Upendrasingh49@gmail.com

Abstract: Object detection has become an important task for various purposes in our daily lives. Machine learning techniques have been used for this task from earlier but they are used for the classification of image-based species to extract the feature set. This task of deciding the feature set helps to decide the desired object detection. To overcome the object classification problem, this paper proposes a transfer learning-based deep learning method. The different convolutional neural networks (CNN) are studied in this work. Here for the improvement in the result, the majority voting scheme is used. The overall work is carried out on the CUB 200-2011 dataset. The results obtained have shown incredible improvement in the accuracy of the proposed work when compared to the different CNN models.

Keywords: Machine Learning (ML), CNN, Object Detection, Transfer Learning, Majority Voting

I. INTRODUCTION

In this digital era, there is a tremendous growth in the area of artificial intelligence and machine learning. The aim of doing such experiments is to build machines that mimic the detection capability of humans. This growth has provided various options to the researchers. Also, the introduction of machine learning started using deep learning approaches to provide a big field for research. The fields like artificial intelligence, speech recognition, face recognition, object detection are the various areas of the applications of deep learning [1]. The convolutional neural network has been widely used For object detection work. The pre-processing tasks needed to perform very much lower as compared to other classification algorithms. Object detection is used in various fields, especially for face recognition. Here in the proposed work, various types of birds are classified. CNN is used for the classification task.

Convolutional Neural Network: In the fields like image recognition and classification, the neural network employed here is the convolutional neural network. It is a category of the neural network, which worked efficiently in the field of image recognition or object detection [1]. CNN has worked well in identifying a face, objects and traffic signboards. The idea behind its good working is that it is an extension of the deep learning algorithms [2]. In the convolutional neural network, there is no need of providing high cost and large resources because the standard algorithm can work. By applying the relevant filters, the spatial and the temporal dependencies can be captures by the convolutional network. The architecture of the convolutional neural network is having the connectivity patterns as that of the neurons in the human brain and was inspired by the visual cortex.

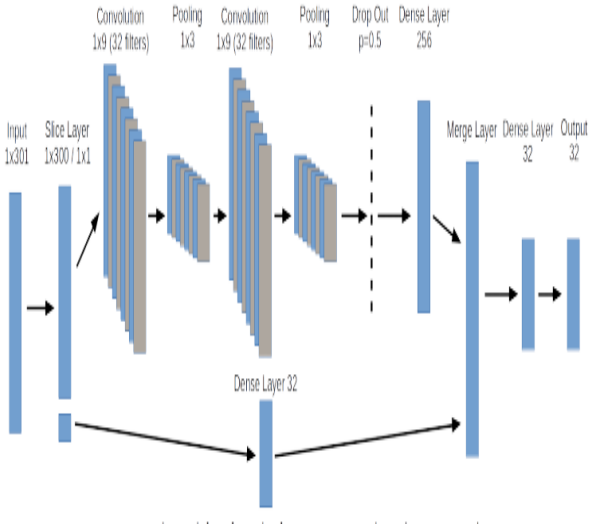

Figure 1. Convolutional Neural Network

II. LITERATURE REVIEW

Existing Work: The authors have worked with the CNN model for object detection. Here they have worked on the fine granularity for detection of objects. They have developed a model to detect almost 200 types of objects (birds). The dataset used by them is the CUB-200-2011. The results have shown that they achieved a higher mean average precision of up to 71.5%. [3]

The authors in this paper have used the region based on CNN. Here they have used a faster CNN method which works on the detection of based on the full image. The results have proved that the faster CNN works well as compared to other CNN models. [4]

Here the authors worked to improve the mostly used algorithms like YOLO and faster CNN. TO do this, they used two techniques that are separable convolution and the filter pruning. The results obtained have improved the results at a great extent. [5]

The authors introduced a cascaded convolutional neural network that does object detection. To do this, the authors have used airborne videos. The framework uses three deep level CNN that does the object detection in a coarse to excellent manner. They have proved that their model achieves netter results. [6]

Here the authors worked on the detection of the areas where the snow has been seen. For this, they have used the CCTV cameras. Here the authors have used the CNN with the transfer learning model. The CCTV footage was taken from the japan website, which is publically available. They have achieved an average sensitivity of 80.90% and the snow detection specificity of 98.54%. [7]

Here the authors proposed a model for object detection based on the CNN. They have used the transfer learning to provide the granularity to the system. They have used the image dataset of COCO. They have proved that the fine-tuned network is more effective. [8]

Here the Japanese comics (manga) are used for the evaluation. They evaluated the R-CNN and SSD for the evaluation of the manga objects. And the authors have shown that the fast RCNN works well for this object detection which detects the character face and the text. [9]

Here the authors have used the object detection for the 360-degree panoramic images. Here they have used the CNN model for the object detection also have applied the post-processing in the model to fine-tune the overall result. They have shown that the proposed model works well in every aspect and gets proper object detection. [10]

The authors here worked for object detection along with the counting of objects on the tensor flow API. They worked on the detection of harmful objects. Along with the tensor flow API, they have used the faster R-CNN algorithm for the experiment. [11]

The authors differentiated the two types of CNN based object detection. The authors have proposed an ensemble method of CNN to detect the objects. They have used the model selecting and the box voting method for the proposed model. At last, they introduced that the proposed and the original method can be combined to get more accurate results. [12]

The authors here worked on the moving object detection. They have worked on object detection by using the tensor flow API. They have used a novel CNN based object detection. They have shown that the model proposed by them works efficiently and achieves the accuracy of up to 90.88%. [13]

Here the authors have analysed the neural networks in the field of deep learning. They gave a detailed description of the deep learning models attached to the neural networks. They surveyed to classify the deep learning neural networks for various categories of applications. [14]

IV. PROPOSED WORK

As studied the literature, it can be said that there are many chances of improvement in the existing work. So the proposed model uses the convolutional neural network with some extra features. The proposed model uses CNN with transfer learning and the majority voting scheme. Both the transfer learning and the majority voting scheme helps in better object detection.

A. **Convolutional Neural Network:** Some of the neural networks nowadays uses a feed-forward

methodology, which includes convolutional neural network (CNN) are used as the models which provide both the classification and the feature extraction. The convolutional neural network comes with different layers that are convolution, pooling and fully connected. The very first layer consists of filters which deal with the width and height of the input for the forward system. In this paper, the initialisation of the weights for filter is made randomly, and then the iteration is performed to update the weights during the training process. The discretisation of the samples is done in the pooling layer. This is done to reduce the dimensions of the input data. The fully connected layer performs the classification.

CNN for image classification works as it takes input in the form of image and gives the output as a category of that image. The CNN convolves already learned feature along with the input data, and it uses two-dimensional convolution layer. That is ideal for the processing of the two-dimensional pictures. In image classification, the CNN has input, output and hidden layer. The hidden layer is having some layers like convolution layer, Relu Layer, pooling layer and the fully connected layer.

The CNN works based on the feature extraction within itself, and it does not require manual extraction of the feature. This feature of CNN makes it suitable for the image classification.

B. **Transfer Learning:** This is the learning technique where the pre-trained convolutional neural network is retrained in some other dataset by using the weights of that particular network. This method of training makes the CNN work fast, where the random initialisation is did. Some of the pre-trained CNN models are VGG16, VGG19, AlexNet, GOogleNet, ResNet this all are commonly used for transfer learning by researchers.

AlexNet: This model consists of 25 layers that are distributed as five are of convolution; three are of pooling and three for fully connected layers and seven for the rectified linear units.

VGG-16 and VGG-19: In this, the VGG-16 model having 41 layers and the matrix used by all the convolution layer is 3*3. Whereas the number of layers used by the VGG-19 model is far more than the VGG-16 model.

GoogleNet: As this module has the inception modules, so it becomes a complex architecture to handle. In this, the addition of a large number of filters is done and sequentially stacking of layers is done due to which the astronomical calculation and memory cost is there.

ResNet-50:In this, some transition layers were ignored by giving preference to lower layers. In this 177 layers are in total, which is having a direct connection which increases its performance.

C. **Majority Voting:** The overall results are carried out by performing voting of the fine-tuned CNN models. An output is provided by every fine-tuned CNN model and among those outputs the majority is assigned to the class that specifies that the majority of the class label is the result of voting. When there are multiple CNN results, the final decision will be stated by the following equations:

$$L(i) = \underset{m}{\operatorname{argmax}} \{V_m(i)\} \quad (1)$$

$$V_m(i) = \sum_{n=1}^{N} \delta(y_n(i), m) \quad (2)$$

$$\delta(i) = \begin{cases} cat, & x < y \\ dog, & x > y \end{cases} \quad (3)$$

Here, for ith image label classification is L(i), for various category Vm(i) is vote results, for the nth result of CNN in(i) is used.

V. EXPERIMENTAL RESULTS

Dataset: The dataset used here in our experiment was the Caltech-UCSD Birds (CUB) 200-2011 dataset which contains images of birds [15]. The details of this dataset are as follows:

- Category of birds = 200
- Total Images = 11,788
- Annotations per image = Part locations:15, Binary Attributes:312, Bounding Box: 1

Experiment: Here, the dataset is applied on the proposed CNN model. This decreases the learning rate. The hardware used for the experiment consists of an Intel(R), CPU N3540 @2.16GHz, 64bit OS.

The rescaling of images is done as the pre-trained CNN works on a fixed-size image. The accuracy parameter is used to evaluate the proposed approach. The tabular form is used to present the numerical value obtained from the experiment.

Table1. Accuracy for various approaches with the presented model

Method	Training Time	Accuracy
AlexNet	0.17hr	94.78%
VGG-16	1.22hr	96.88%
VGG-19	1.54hr	97.05%
GoogleNet	0.32hr	96.56%
ResNet-50	1.15hr	96.44%
Majority Voting	**0.57hr**	**97.43%**

The table given for different schemes state that the obtained results for the proposed schemes are much better than the earlier methods. The accuracy and the training time is analysed and the values are stated. There is a clear vision of improvement in all the values for the presented majority voting scheme.

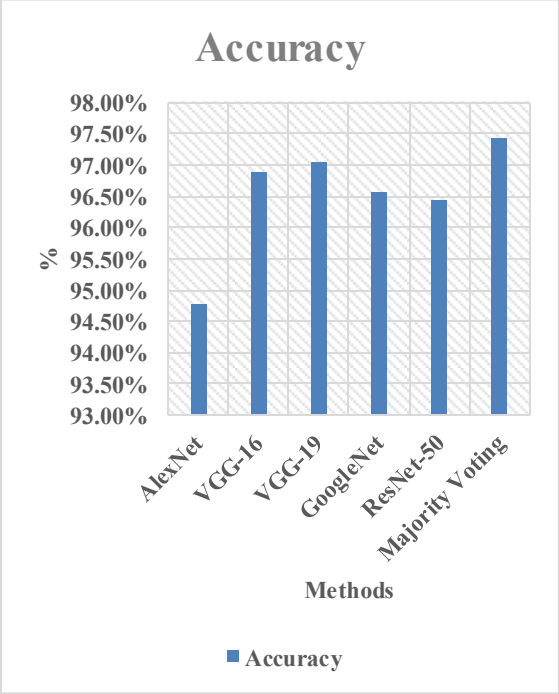

Figure 2. Accuracy with the Presented Models

Here figure 2 is the graphical representation of the accuracy obtained from various methods. The values obtained for the majority voting scheme 97.43% with very less time of training, i.e. .57 hour.

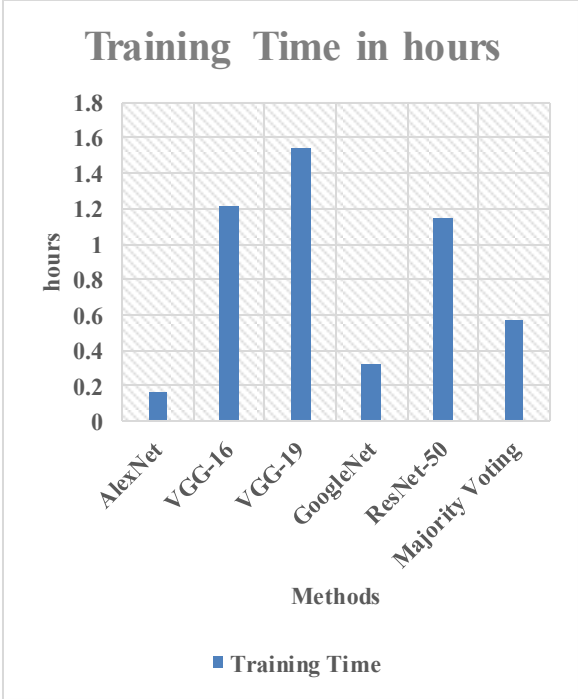

Figure 3. Execution time in Hours with the Presented Models.

Here figure 3 states the training time of the fine-tuned models of CNN. The VGG-19 is providing the highest results of 97.05%. The results of all CNN models are presented to the majority voting scheme to get the impact of all on the classification problem. The training time was obtained is .57 hours for the majority of votes.

VI. CONCLUSION

So the proposed work has the convolutional neural network with transfer learning and the majority voting scheme works well in comparison with the actual work. The proposed work gives fine granularity to work done earlier. The results have shown an extreme improvement in the field of object detection. Here, the CUB-200-2011 dataset is used to detect different types of birds. The 5 different pre-trained models of CNN were used, and their results are added with the majority voting scheme, and the results obtained a high accuracy of 97.45%.

Also, it is known that there is always a chance for improvement. So in our work, there may present some limitations that can be analysed further. In future, some other classification methods will be applied. Also, the presented model can be tested on different datasets for evaluation.

Reference:

1. Roy, S. K., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. (2019). *HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters, 1–5.* DOI:10.1109/lgrs.2019.2918719
2. He, R., Wu, X., Sun, Z., & Tan, T. (2018). *Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1.* DOI:10.1109/tpami.2018.2842770
3. Ümit BUDAK, Abdulkadir ŞENGÜR, Aslı Başak DABAK and Musa ÇIBUK, Transfer Learning-Based Object Detection and Effect of Majority Voting on Classification Performance, 2019 International Artificial Intelligence and Data Processing Symposium (IDAP)
4. Wang, C., & Peng, Z. (2019). *Design and Implementation of an Object Detection System Using Faster R-CNN. 2019 International Conference on Robots & Intelligent System (ICRIS).* DOI:10.1109/icris.2019.00060
5. Qian, L., Fu, Y., & Liu, T. (2018). *An Efficient Model Compression Method for CNN Based Object Detection. 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS).* DOI:10.1109/icsess.2018.8663809.
6. Zhang, W., Li, J., & Qi, S. (2018). *Object Detection in Aerial Images Based on Cascaded CNN. 2018 International Conference on Sensor Networks and Signal Processing (SNSP).* DOI:10.1109/snsp.2018.00088
7. Parintorn Pooyoi; Punyanuch Borwarnginn; Jason H. Haga; Worapan Kusakunniran…Snow Scene Segmentation Using CNN-Based Approach With Transfer Learning, 2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON).
8. Wendi Cai; Jiadie Li; Zhongzhao Xie; Tao Zhao; Kang LU, Street Object Detection Based on Faster R-CNN…2018 37th Chinese Control Conference (CCC)… 10.23919/ChiCC.2018.8482613
9. Hideaki Yanagisawa; Takuro Yamashita; Hiroshi Watanabe…..A study on object detection method from manga images using CNN…. 2018 International Workshop on Advanced Image Technology (IWAIT)
10. Yiming Zhang; Xiangyun Xiao; Xubo Yang…..Real-Time Object Detection for 360-Degree Panoramic Image Using CNN… 2017 International Conference on Virtual Reality and Visualization (ICVRV)
11. B N Krishna Sai; T. Sasikala…Object Detection and Count of Objects in Image using Tensor Flow Object Detection API… 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)
12. Jinsu Lee; Sang-Kwang Lee; Seong-Il Yang…An Ensemble Method of CNN Models for Object Detection, 2018 International Conference on Information and Communication Technology Convergence (ICTC)
13. Shraddha Mane; Supriya Mangale, Moving Object Detection and Tracking Using Convolutional Neural Networks…. 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)
14. D. A. Bashar, "Survey on Evolving Deep Learning Neural Network Architectures," *J. Artif. Intell. Capsul. Networks*, vol. 2019, no. 2, pp. 73–82, 2019, DOI: 10.36548/jaicn.2019.2.003.
15. Wah C., Branson S., Welinder P., Perona P., Belongie S. "The Caltech-UCSD Birds-200-2011 Dataset." Computation & Neural Systems Technical Report, CNS-TR-2011-001.