IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Study on Object Detection Using Convolutional Neural Networks and Various Pretrained Models

Chaitanya Kumar, Shiv Sharad Choudhary, Prashant Kumar, Gauri Shankar Mishra

Sharda University, Greater Noida, Uttar Pradesh, India

Abstract* — *Deep learning has evolved into a powerful machine learning technology that incorporates multiple layers of features or representations of data to get cutting-edge results. Deep learning has demonstrated outstanding performance in a variety of fields, including picture classification, segmentation, and object detection. Deep learning approaches have recently made significant progress in fine-grained picture categorization, which tries to discriminate subordinate-level categories. Due to strong intra-class and low inter-class variance, this task is highly difficult. Object detection and the identification of pedestrians is critical in autonomous driving applications. In real-time applications, approaches based on Convolutional Neural Networks have shown significant increases in accuracy and decision speed. In this research, authors present various state of the art deep learning algorithms i.e., VGG-16, VGG19, DenseNet-121, InceptionV3 and customized 3 layers CNN model for object detection. Model adopted is trained and validated on self-made five class of furniture dataset. After extensive experiments, highest accuracy obtained was 99.89% with VGG-19.

Keywords* — *Object Detection, Object classification, Convolutional Neural Network, VGG, Confusion Matrix, Accuracy.

I. INTRODUCTION

With the development of mobile internet and the popularization of various social media, the amount of image data on Internet has increased rapidly, but human beings cannot process efficiently so many image data. So, it is expected to carry out these data processing automatically with the aid of computer to solve large-scale visual problems. With a deeper understanding of image processing technology, comprehensive understanding of the image and accurate identification of the target object of the image becomes more and more important **[4]**. The people not only concern about the classification of images simply, but also want to accurately obtain the semantic category of object and the location in the image **[3]**, so the object detection technology had received wide attention **[4]**. Object detection technology tries to detect target items using image processing and pattern recognition theories and methods, determine semantic categories for these objects, and mark the target object's particular position in the image. **[1]**.

In this paper authors addressed this problem with five convolutional neural networks algorithms and various pretrained models which were successfully implemented. Out of the five models compared under study VGG-16 achieved better accuracy. Realtime images were used for the testing purpose and to detection of object.

Bed **Chair** **Sofa** **Swivel Chair** **Table**

Fig-1: Five Classes sample dataset for model training

Fig-1 is the sample of training dataset which self-made dataset with high resolutions and it consists of five classes as bed, chair, sofa, swivel chair and table images with size 1572 x 1548 pixels.

Remaining Contents of the paper hence follows: - Section II discusses the data structure and the types adopted. Section III commands on the working or implementation of the given algorithms. Section IV comprises the proposed work's results and discussion, and lastly, Section V extracts the paper with future directions.

II. DATASET AND METHODS

A. Dataset:

Five classes self-made dataset is used for this study which contains bed, chair, sofa, swivel chair and table images. We train our model on five class of furniture image classes in which bed has 9000 training and 100 validation images, chair has 9000 training and 100 validation images, sofa has 9000 training and 100 validation images, swivel chair has 9000 training and 100 validation images and table has 9000 training and 100 validation images. The total number of image samples in dataset is 5000. Table-1 below represent the structure of our dataset.

Table-1: Number of classes and total images in dataset

Classes	Training Images	Validation Images	Total Images
Bed	900	100	1000
Chair	900	100	1000
Sofa	900	100	1000
Swivel Chair	900	100	1000
Table	900	100	1000
Total Images	**4500**	**500**	**5000**

B. Image Pre-processing:

During image pre-processing our datatset images are compromised with the size of 256 x 256 pixel from 1572 x 1548 pixel to minimize the background area and 256 x 256 pixel is best fit for convolutonal neural networks. The compromised image is then utilised for validation and training. Dataset on DenseNet-121, InceptionV3, VGG-16, VGG-19 and CNN with three layers. During model training authors set various hyperparameters eg, training and validation split to 0.1 during image pre-processing, rescale to 1./255, shear range to 0.2, zoom range to 0.2 .

C. Methods:

Contemporary, Convolutional Neural Networks Because of its capacity to extract features from images without complex pre-processing, as well as transfer learning and fine-tuning parameters, it is a state-of-the-art approach. These types of study uses VGG-16, VGG-19, DenseNet-121, and InceptionV3, which make use of transfer learning often used in deep learning. We use transfer learning receive the quality vector for arranging furniture (object) using CNN and differentiate the results to decide which learning is the perfect for object detection. Table-2 below presents the various convolutional neural networks models over different criteria.

Table-2: CNN Models over different Criteria

Model Name	Size (MB)	Parameters (Millions)	Depth
VGG-16	528	138.3	23
VGG-19	549	143.6	26
DenseNet-121	33	8	121
Inception V3	92	23.8	159

1). 3 Layer CNN: 3-layer CNN consists of a convolutional layer, a pooling layer, and a fully connected layer. The CNN's main building block is the convolution layer **[3]**. It accounts for most of the computational load of the network. The pooling layer uses the summary statistics of neighboring outputs to transform the outputs of the network at specific locations. This minimizes the spatial size of the representation, thereby reducing the amount of computation and load required. As in a conventional fully convolutional neural network, fully connected layers have full

connection with all neurons in the preceding and subsequent layersThe completely linked layer aids in the mapping of the input and output representations.

2). DenseNet-121: A DenseNet is a type of convolutional neural network that employs dense connections between layers through dense blocks, with all layers directly connected to each other. DenseNet was created to address the problem of decreased accuracy caused by the longer path between the input and output layers, where information evaporates before it reaches its goal **[6]**.

3). VGG-16: Visual Geometry Group is a convolutional neural networks architecture They concentrated on having 3x3 convolution layers because it has a large number of hyper-parameters. filter size **[2]**. The VGG-16 convolutional neural network architecture is a simple and extensively used convolutional neural network design. VGG-16 is used in many deep learning image classification techniques and is popular due to its ease of implementation. VGG-16 is extensively used in learning applications due to the advantage that it has. In VGG-16 the number 16 defines the layers and depth. This CNN network has very large network approx 138 million parameters.

4). VGG-19: VGG-19 is a convolutional neural network that is a variant of the VGG model, with a total of 19 layers (16 convolution layers, 5 MaxPool layers, 3 fully connected layers and 1 softmax layer) **[7]**. Only 33 convolutional layers are placed on top of each other in increasing order of depth in this convolutional neural network architecture. This is a very large network, it has approx 143 million parameters.

5). Inception V3: The image recognition model Inception V3 is very popular. Convolution, Average Pooling, Max Pooling, Concats, Dropouts, and Fully Linked Layers are some of the symmetric and asymmetric building elements that make up the model **[1]**. Activation inputs are subjected to batch normalisation, which is employed throughout the model. The Softmax method is used to calculate the loss.

III. MPLEMENTATION WORK

To train our images dataset authors used 3 convolutional neural networks and various pretrained models (VGG-16, VGG-19, DenseNet-121, InceptionV3) which is better performed on the real time object detection data set earlier. During model training, the author specified various hyperparameters (learning rate (lr) to 0.001, batch size to 32, starting function ReLu and Softmax, epoch size 10) and employed batch normalization during parameter building to avoid overfitting and underfitting. For model training, the ReLu and Softmax activation functions were utilised, and the Tensorflow and Keras frameworks were used for implementation. The first layer is a convolution layer with 64 filters, followed by a stack of three CNN blocks, each containing 32, 32 and 128 filters, with dimension reduction performed for each layer in the CNN block. In VGG-16, VGG-19 and DenseNet-121 having one filter with size 1024. After the data pre-processing and model building, we trained our model over 10 epochs utilizing the TensorFlow and Keras frameworks with CPU processing itself. Fig-2 demonstrates the implementation scenario.

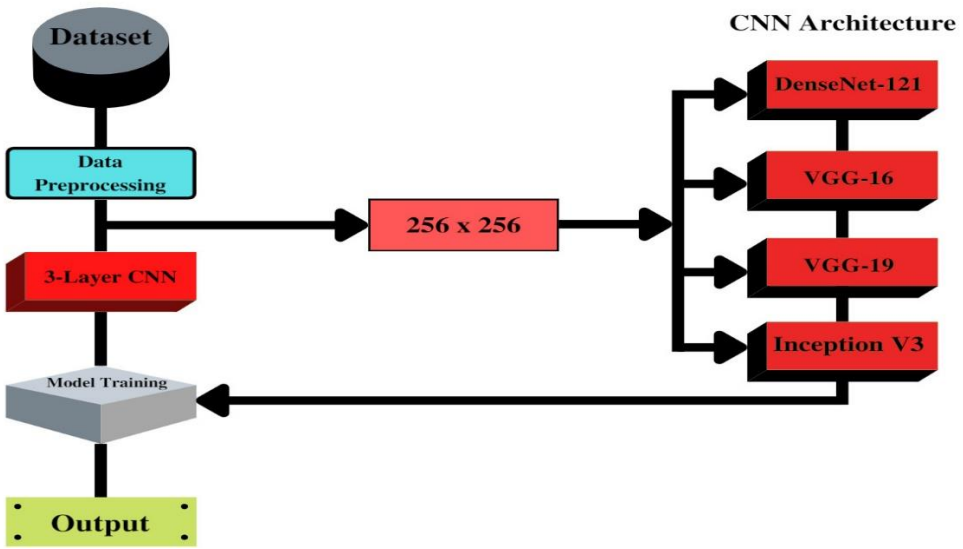

Fig-2: Implementation Scenario

IV. RESULTS

The accuracy, F1 score, and confusion matrix of our dataset were calculated by the authors to evaluate our suggested technique. The degree to which the estimated findings reflect the ground reality is referred to as accuracy. The performance and accuracy of our proposed model were compared to [4], where the author estimated their performance and accuracy value is 98%. Our proposed approach uses DenseNet-121, VGG-16, VGG-19, InceptionV3 andthree layers To identify the photos of the CNN, which is a more efficient neural network design. leaf disease dataset and VGG-16 gives the highest accuracy 99.89%. Our proposed model's (VGG-19) training and validation loss, as well as the training and validation accuracy, are plotted on a graph., Confusion matrix is shown in figure.

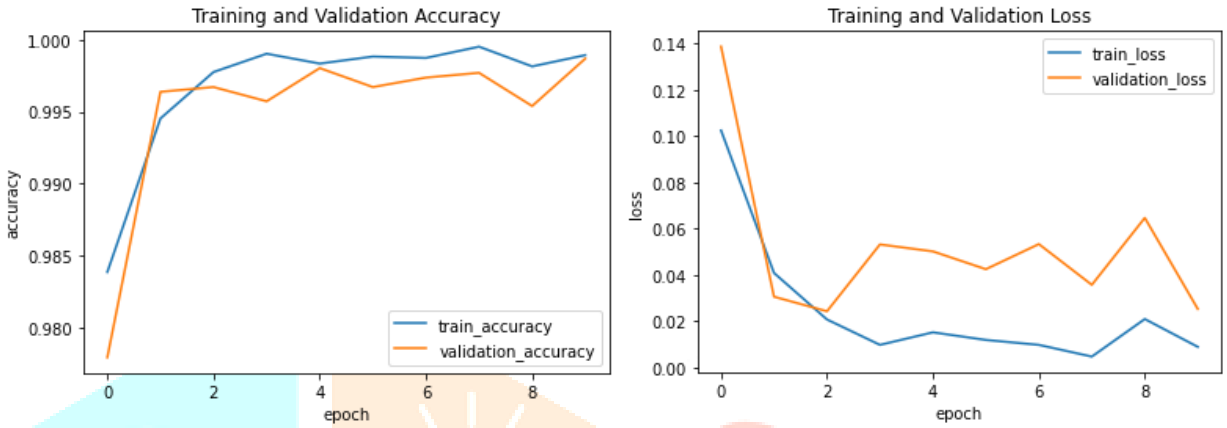

Fig-3 and Fig-4: Graphical Result

Table-3: Comparative Analysis of Different Models based on Training and Validation Accuracy.

Model	Train Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)
3-Layer CNN	94.64	91.23	90.36
VGG-16	98.16	95.22	95.69
VGG-19	99.89	99.87	98.47
DenseNet-121	92.49	91.78	91.18
Inception V3	90.75	88.96	85.60

Fig-5: Confusion Matrix

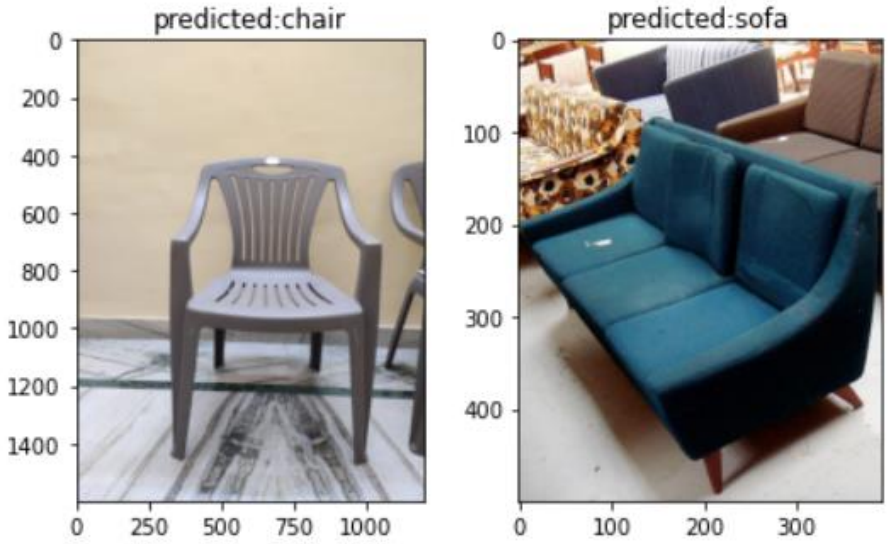

Fig-6: Proposed model prediction

V. CONCLUSIONS AND FUTURE WORK

Accurately locating an object in a surveillance video is one of the most important research areas in computers imaginable and has a wide range of cutting-edge programs in modern times. In the present day it is very difficult to cut modern day leaves such as low resolution, models of lights, moving objects beyond the ancient, small adjustments in the historical past, due to subsequent gadget photographs obtained from a surveillance video. We have presented a top degree visual development in item detection strategies. The detection approach takes place in background modelling, item detection, and object categories. In this paper, all available item detection strategies are classified into history subtraction, optical float and spatial-temporal filter out techniques and the advantages and drawbacks of today's techniques implemented in many modern-day datasets are referenced. Object type techniques are further categorized into strategies based on form-based thoroughness, movement-based and texture-based altogether.

In this paper, authors presented a comparative study of five Convolutional Neural Network Models classification of object detection. Out of the models under study VGG-19 outperformed all other in terms of accuracy. VGG-19 achieved an accuracy of 99.89% on training dataset, 99.87% accuracy on validation dataset and after testing on different furniture images, model obtained 98.47% accuracy.

REFERENCES

[1] Li, K., Wan, G., Cheng, G., Meng, L., & Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159, 296-307.

[2] Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, 9(2), 85-112.

[3] Chandan, G., Jain, A., & Jain, H. (2018, July). Real time object detection and tracking using Deep Learning and OpenCV. In 2018 International Conference on inventive research in computing applications (ICIRCA) (pp. 1305-1308). IEEE.

[4] Zhiqiang, W., & Jun, L. (2017, July). A review of object detection based on convolutional neural network. In 2017 36th Chinese control conference (CCC) (pp. 11104-11109). IEEE.

[5] Jia, B., Pham, K. D., Blasch, E., Wang, Z., Shen, D., & Chen, G. (2018, March). Space object classification using deep neural networks. In 2018 IEEE Aerospace Conference (pp. 1-8). IEEE.

[6] Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. Neurocomputing, 396, 39-64.

[7] Yanagisawa, H., Yamashita, T., & Watanabe, H. (2018, January). A study on object detection method from manga images using CNN. In 2018 International Workshop on Advanced Image Technology (IWAIT) (pp. 1-4). IEEE.

[8] Prabhakar, G., Kailath, B., Natarajan, S., & Kumar, R. (2017, July). Obstacle detection and classification using deep learning for tracking in high-speed autonomous driving. In 2017 IEEE region 10 symposium (TENSYMP) (pp. 1-6). IEEE.

[9] Uçar, A., Demir, Y., & Güzeliş, C. (2016, August). Moving towards in object recognition with deep learning for autonomous driving applications. In 2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA) (pp. 1-5). IEEE.

[10] Gao, H., Cheng, B., Wang, J., Li, K., Zhao, J., & Li, D. (2018). Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment. IEEE Transactions on Industrial Informatics, 14(9), 4224-4231.

[11] Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., & Zou, H. (2018). Multi-scale object detection in remote sensing imagery with convolutional neural networks. ISPRS journal of photogrammetry and remote sensing, 145, 3-22.

[12] Sun, X., Wu, P., & Hoi, S. C. (2018). Face detection using deep learning: An improved faster RCNN approach. Neurocomputing, 299, 42-50.

[13] Ji, Y., Zhang, H., Zhang, Z., & Liu, M. (2021). CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. Information Sciences, 546, 835-857.

[14] Li, K., Wan, G., Cheng, G., Meng, L., & Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159, 296-307.