JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

OBJECT DETECTION WITH COMPUTER VISION USING VGG-19 ARCHITECTURE

[1]RISHABH SHUKLA, [2]ER.NANDINI SHARMA

[1]STUDENT, [2]ASSISTANT PROFESSOR

[1]DEPARTMENT OF INFORMATION TECHNOLOGY,

[1]SHRI RAMSEAROOP MEMORIAL COLLEGE OF ENGINEERING AND MANAGEMENT(SRMCEM), LUCKNOW, UP,INDIA

Abstract : The motivation behind this study is to perform the process of object detection with the help of computer vision technology in which we have insert the images in any format which are in .pdf form, .jpg form, .png form, .jpeg form . .gif format. Here we can use the VGG 19 algorithm or we can also use the YOLO algorithm. But VGG 19 is awesome algorithm as it is a convolutional neural network which has 19 deep layers. This study also gives the percentage of accuracy of predictions of object present in the image. The accuracy of this model is 95%.This model is based on transfer learning model in which we use the VGG-19 architecture model based on Convolutional Neural Network (CNN). VGG-19 is a convolutional neural network layer architecture which is mostly utilized for transfer learning process. It consists of 16 convolutional layer,19 learnable weights, 3FC layer and one output layer.

I. INTRODUCTION

Some years ago, the development of software and hardware was limited to only user interface. But after the advent of operating system , the whole scenario of computer got changed when the majority of the developers switched to solving the problems of image processing itself. However this has not led to solve many difficult and rigorous problems like recognizing the faces , recognizing the car numbers ,recognition of objects in the images, road sign, automatic weapon detection, detection of medical images and disease identification.

As modern technical solutions are turn out to be expensive the task of creating the software which are based on concept of automation , for solving intellectual problems is delineated and intensively solved abroad.

Object detection is to describe a collection a computer vision tasks that involves activities like identifying objects in images ,digital identification of photographs ,detection of objects in videos, real-time camera detection .

Computer vision is one of the most using technologies in today's world. It comes under the section of Artificial Intelligence (A.I.). In case of Object Detection, these two tasks comes in the category of Computer Vision-:

- **Image Classification**
  This is done to predict to type or class of an object in an image.
- **Object Detection**
  Object Detection is a computer vision technology that is related to computer vision and image processing that deals with detecting the instances of semantic objects of a particular class (such as humans, animals, tanks, fountains , vehicles) in digital images and videos .
  Face detection, Weapon detection system , Surveillance detection system, Gun detection system, Automatic lock system , Disease detection system.

Human can detect and identify the images more easily. But when we want to work on various no of images, it is recommended to use the model based on artificial intelligence technology . Human visual system is used in that perspective .Human visual system is fast ,more accurate and precise but in that we have to train our model with large no of datasets of objects which we want to find. The availability of large number of datasets , faster GPU(Graphics Processing Unit), and usage of better and advanced algorithm help computers to train easily and allow computers to give accurate and precise predictions.

II. OTHER RELATED WORKS

Image Segmentation and Image Classification are important parts of Machine Learning (M.L) and Artificial Intelligence(A.I.) and they are mostly used in object detection model to get the type of class of the object.

Here we use VGG architecture present in keras library of Machine Learning. In VGG architecture, all Convolutional Neural Network (CNN) uses filters of size of 3*3 with stride=1 and same padding. Filters of size of 2*2 with stride=2 are present in all max-padding layers .Here VGG model based on CNN performs the operations of image segmentation and image classification.

III. METHODOLOGY-:

Here in our project we use the concept of Transfer Learning methodology which is the part of deep learning and Computer Vision. Transfer Learning models are said to those model which is already trained by other users and developers on various datasets and then we use the features of those already pre trained models in our problem sets for getting the required solutions.
In Transfer Learning , we do not train all the convolutional layers . Here we lock some of the layers and use only those trained weights in locked layers to extract particular features from our data. We do not need to retain some of lower layers ,we can choose to retrain some of lower layers because these lower layers will be specialized by our datasets.
Many challenges comes in the field of Computer Vision to overcome the former discussed challenges comes in the field of object /image classification .Although in previous decades ,the convolutional approaches such as Hand-Crafted Feature(HCF) were used but did not get much success.
Handicraft features included Histogram of Oriented Graph(HOG),geometric features ,Scale Invariant Feature Transformation(SIFT), Difference of Gaussian(DOG),Speeded-UP Robust Features(SURF) and texture features (HARLICK). These features proposed to exploit a hybrid set of features to get a better a representation of an object. By keep regard of these challenges, the concept of deep learning has been introduced to show the best performance against computational time. That's why a large number of convolutional neural network (CNN) pre trained models have been proposed like AlexNet ,VGG(VGG-16, VGG-19) ,GoogleNet, ResNet and Inception. All these models are based on ImageNet Dataset.
This process has given rise to concept of feature fusion whose definition is a process of combining several feature population into a single feature space, which has been adopted in various applications ranging from surveillance detection system to object classification. The concept of Feature Fusion provides perfect and concise classification accuracy that's in todays world it is mostly used by developers and scientists

REASON FOR USING TRANSFER LEARNING METHODOLOGY AND DATASETS-;

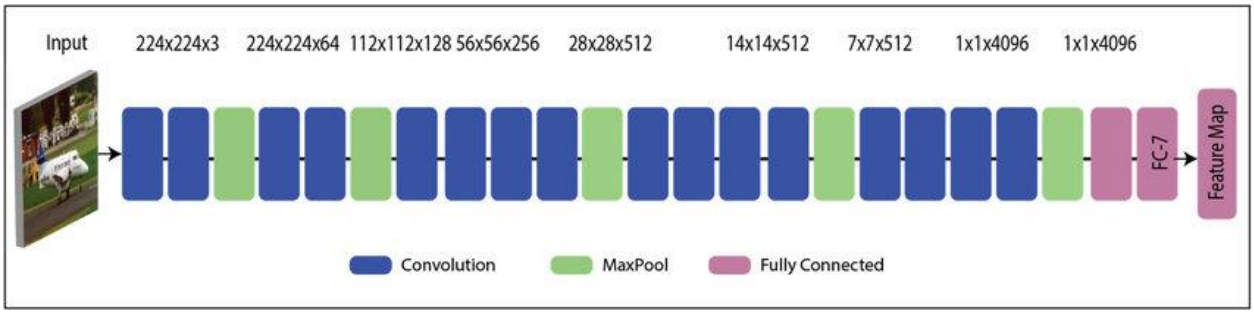
- Transfer Learning Methodology is used because if we did not use it then we have to train our models with a lot of dataset of images ,which we have to download either from kaggle or from OIDV4 toolkit to train our model.
- And during training we need to wait for 100 epochs to be completed for creating the weights with regard to that dataset.
- And different weight files get created after 1000 iterations.
- If a person has a graphics card(like NVIDIA) in his/her computer then the training process runs faster as training process involves GPU along with the CPU.
- If a person does not have a graphics card in his/her system then the training process can take a week or even more to be completed as process of creation of weight files gets slowed as in that case only CPU is involved not GPU.
- That's why we have to use the process of transfer learning so as to make our computing faster.

3.1 VGG-19 Architecture

VGG-19 is a convolutional neural network layer architecture which is mostly utilized for transfer learning process. It consists of 16 convolutional layer,19 learnable weights, 3FC layer and one output layer. The learning weights and bias of the first convolutional layer are 3*2*3*64 and 1*1*64. Total number of layers is 1792. For second layer, the total learnable layer is 3698 and this layer extracts the local feature of an image.
The learnable weights and bias of first FC layer are 4096*25088 and 4096*1 respectively and after this dropout layer is added between FC layers, where dropout rate is 50%. For FC layer 7 ,total learnable is 16,781,312 and learnable weights are 4096*4096 respectively. For last layer ,the total learnable is 4097000 and for learnable wrights are 1000*4096. When the activation is applied, it returns a feature map of dimension of 1*1*1000 . For fully connected layer 1 and 2, 1*1*4096 is the feature map dimensions.

3.2 Feature Extraction Using Transfer Learning

In this step, we employ the transfer learning by which we retrain our VGG-19 architecture on selected datasets. For training we set a 60:40 approach with labeled data i.e. 60% for training process and 40% for testing process . Then, we perform preprocessing in which we resize the images according to input layer of each model. Then we choose the input convolutional and outer layers as feature mapping. In VGG-19, we choose the first convolutional layer as input layer and FC7 layer as output layer. Once the CNN architecture gets activated, we obtain the training and testing vectors.
On FC7 layer in which feature extraction process takes place, a resultant feature vector of 1*4096 is obtained.



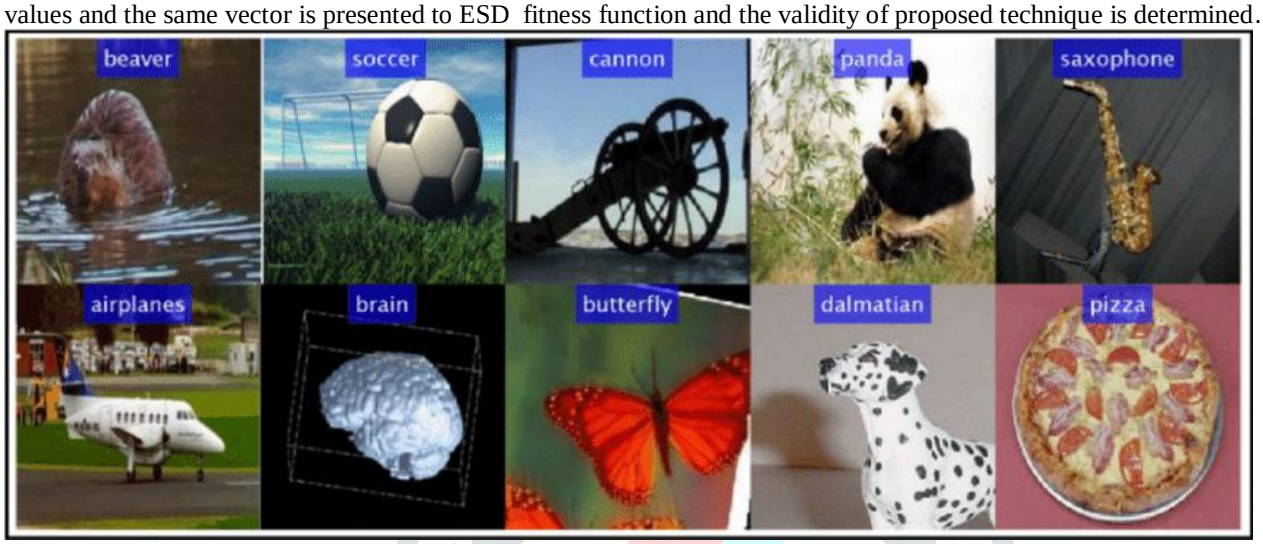
MODIFIED VGG-19 ARCHITECTURE FOR FEATURE EXTRACTION

3.3 Feature Fusion-:

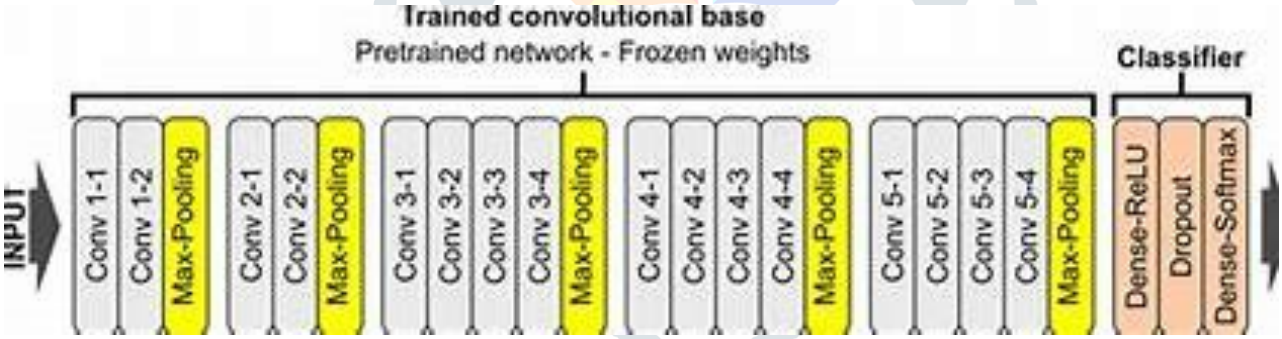
Feature Fusion is the latest research of feature recognition in which fusion of multiple features comes in one matrix. This process is done to obtain a stronger feature vector for classification. It has been found that the fusion process improves the overall accuracy. But its main disadvantage is high computational time. However, main priority is to improve classification accuracy. That's why, we implement a PARALLEL MAXIMUM COVARIANCE(PMC) approach for feature fusion. In this we need to equalize the length of both extracted feature vectors.

3.4 Feature Selection-:

Feature Selection is an very interesting research topic in computer vision nowadays and this technique plays a vital role in development of machine learning segment, computer vison segment and artificial intelligence segment. This technique shows a significant improvement in classification accuracy. In this technique , as we are using the VGG-19 architecture so we propose a new technique for feature selection , namely MULTI LOGISTIC REGRESSION CONTROL ENTROPY VARIANCES (MRCEV). This technique exploits a partial derivative based activation to remove irrelevant features and remaining robust features are passed to entropy variances functions. Though the latter ,a new vector is obtained. This vector contains only positive values and the same vector is presented to ESD fitness function and the validity of proposed technique is determined.



PROPOSED SYSTEM'S PREDICTED lABELLED OUTPUT FOR CALTECH-101 DATASET

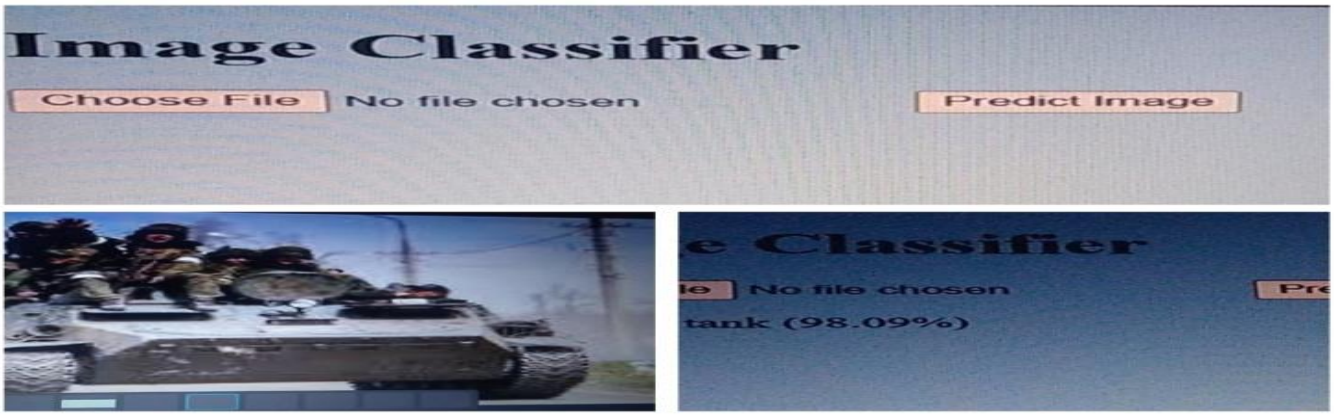


SCHEMATIC OVERVIEW OF PERSONALIZED VGG-19 NETWORK ARCHITECTURE

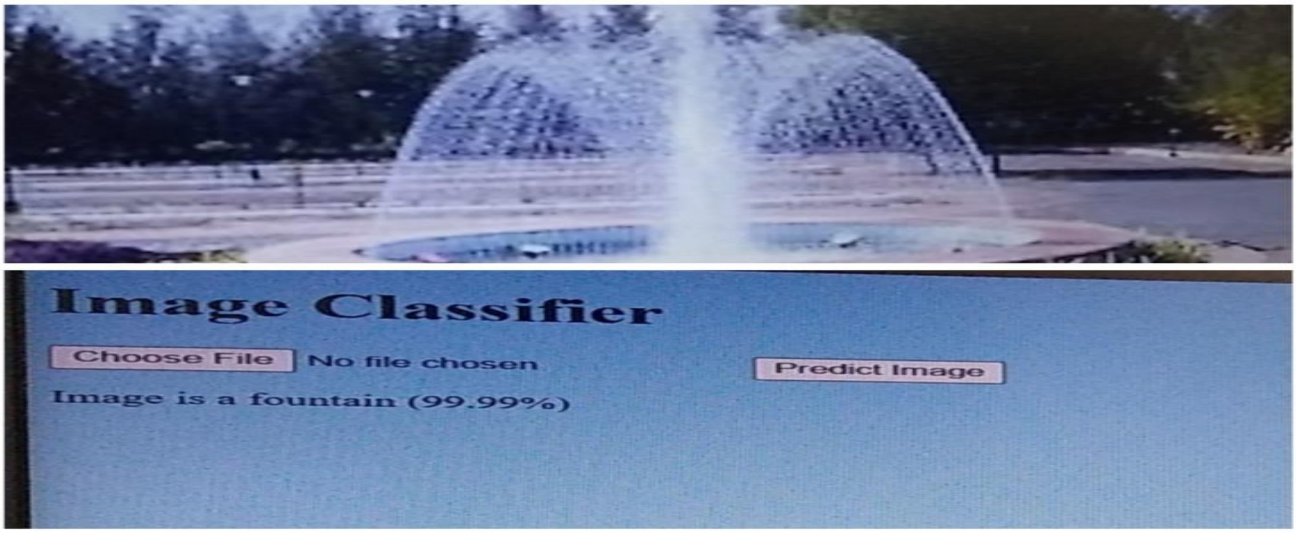
PROPOSED SYSTEM'S PREDICTED ABELLED OUTPUT FOR BIRD'S DATASET

IV. RESULT-:

This section only deals with detailed numerical analysis and predictions. As this architecture is already pre-trained as we are using the concept of Transfer Learning process. This architecture works on 60:40 approach i.e. 60% for training process and 40% for testing process . This architecture also uses various classifiers for various experimental process such as ensemble learning, SVM ,KNN and Linear Discriminant Classifier. This project can easily tell you the class and the type of the object present in the image. This research project also gives you the percentage of accuracy of prediction performed by VGG-19 architecture on images inputted in project.



PREDICTION AND ACCURACY ON INSERTING IMAGE OF TANK



PREDICTIONS ON INSERTING THE IMAGE OF FOUNTAINS

V. CONCLUSION-:

The main motive of this research is to perform object detection process using two techniques, one is n=multi layer deep feature fusion technique and second one is selective based technique. The vital contribution of this research stresses in combination of deep learning model, VGG-19 architecture and then selection of robust features for this project. Three major important steps are used in this research: Feature Extraction using transfer learning, Feature Fusion of deep learning model VGG-19 using PMC, selection of the powerful features using MULTI LOGISTIC REGRESSION CONTROLLED ENTROPY-VARIANCE method.
The performance of MRcEV method is enhanced by the ESDA classifier used for classification process. The combo of two different deep learning features gives a very good impact on classification accuracy

REFERENCES

- Arshad, H.; Khan, M.A.; Sharif, M.I.; Yasmin, M.; Tavares, J.M.R.; Zhang, Y.D.; Satapathy, S.C. A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition. Expert Syst. 2020, e12541
- Arshad, H.; Khan, M.A.; Sharif, M.I.; Yasmin, M.; Tavares, J.M.R.; Zhang, Y.D.; Satapathy, S.C. A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition. Expert Syst. 2020, e12541
- Khan, M.A.; Akram, T.; Sharif, M.; Javed, M.Y.; Muhammad, N.; Yasmin, M. An implementation of optimized framework for action classification using multilayers neural network on selected fused features
- Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Improved inception-residual convolutional neural network for object recognition. Neural Comput. Appl. 2018, 32, 1–15.
- Li, Q.; Peng, Q.; Yan, C. Multiple VLAD encoding of CNNs for image classification. Comput. Sci. Eng. 2018, 20, 52–63
- Mahmood, A.; Bennamoun, M.; An, S.; Sohel, F. Resfeats: Residual network based features for image classification. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017
- Ly, H.-B.; Le, T.-T.; Vu, H.-L.T.; Tran, V.Q.; Le, L.M.; Pham, B.T. Computational hybrid machine learning based prediction of shear capacity for steel fiber reinforced concrete beams. Sustainability 2020