Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018)
IEEE Xplore Compliant - Part Number:CFP18J06-ART, ISBN:978-1-5386-0807-4; DVD Part Number:CFP18J06DVD, ISBN:978-1-5386-0806-7

Object Recognition in Images using Convolutional Neural Network

Mr.Sudharshan Duth P
Dept. of Computer Science
Amrita School of Arts and Sciences
Amrita Vishwa Vidyapeetham
Mysuru, Karanataka, India.
sudhisp@gmail.com

Ms.Swathi Raj
Dept. of Computer Science
Amrita School of Arts and Sciences
Amrita Vishwa Vidyapeetham
Mysuru, Karanataka, India.
swathiraj148@gmail.com

ABSTRACT:**Object detection from repository of images is challenging task in the area of computer vision and image processing in this work we present object classification and detection using cifar-10 data set with intended classification and detection of airplain images.So we used convolutional neural network on keras with tensorflow support the experimental results shows the time required to train,test and create the model in limited computing system. We train the system with 60,000 images with 25 epochs each epoch is taking 722to760 seconds in training step on tensorflow cpu system.At the end of 25 epochs the training accuracy is 96 percentage and the system can recognition input images based on train model and the output is respective label of images.**

Keywords: Convolutional neural network, Deep neural network, Tensorflow, Cifar-10, epoch, keras.

I.INTRODUCTION

The Deep Neural Network (DNN) is the latest advancement in the deep learning made task simple for image recognition by as deeply as possible learning. Deep learning is a subdivision of machine learning algorithms, which are excellent in identifying patterns, but usually require more data. The most popular technique used in improving the accuracy of image classification is the Convolutional neural network(CNN).CNN is a special type of neural network that works just like the normal neural network, which initially has a convolution layer. Rather than bolstering the whole pictures as a variety of numbers, the pictures is separated into various tiles, the machine at that point tries to predict what's in the image in view of the expectation of the considerable number of tiles. This enables the PC to parallelize the operations and recognize the object little heed to where it is located in the image. To begin with pre-processing, the pictures are extremely composed by the dataset and contain little to zero clamor, we have to include a tad of change to the information. We include misleadingly clamor utilizing the Python library imgaug. Following make an arbitrary mix of images:

•The crop parts of the image

•Flip the image on a level plane

•At last, alter the shading, difference, and saturation.

Second, partitioning takes a long time to calculate the model gradient using our dataset,so images of small batches will be used during each repetition of optimizer.Usually the batch size is usually 32 or 64.We use 64 because we have a large number of images.The dataset is then divided into a training set consisting of 50,000 images,and a test set with 10,000images.At longlast,to fabricate a convolutional neural system.2x2 Max-Pooling,a three-convolution layer with a maximum pooling,is an approach to diminish image dimensions by taking a lattice most extreme pixel value.This makes it all the more overfitting and test makes the most generic.the case beneath 2x2 max pool, we include completely associated layers.

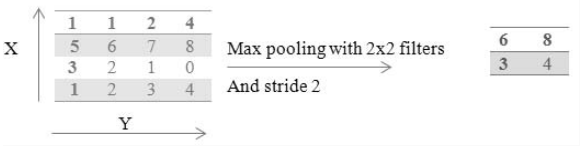

Fig1-2x2 max pooling

The completely associated layers info ought to be two dimensionals, and the output of convolution layer is four dimensional,we need a flattening layer between them.

II.RELATED WORK

Christian Szegedy et al.,(2013)[1]"Deep Neural Networks for Object Detection" Deep Neural Networks(DNNs) have as of late indicated superb performances in image characterization operations.Take above and beyond in this paper author and resolve the issue of location of object recognition utilizing DNNs,not only the grouping but rather the accuracy of various classes.They show target location as an issue that counteracts covers as a case that restrains the question straightforward but effective plan. Portray a multi-scale basic leadership technique that produces minimal effort high-determination protest finders from some-arrange applications.The most developed execution of the strategy is appeared at Pascal VOC. Soren Goyal and Paul Benjamin(2014)[2] A Survey "Identification of objects utilizing Deep Neural Networks is a dynamic region of research and has made many advances in the previous couple of years.Paper compresses the historical backdrop of research on Neural Networks and clarifies the most recent advancements in this field. Recorded on the as of late created sensory system calculation on benchmark datasets. At long last, a few applications in this field are given. Samer Hijazi et al.,(2015)[3]Convolution Neural Networks (CNN) are generally utilized for pattern and image acknowledgment issues they have a few points of interest over different innovations.Samer Hijazi plot the difficulties of utilizing CNNs in installed frameworks and should demonstrate Cadence® Tensilica® Vision P5 Digital Signal Processor (DSP) Key Features in Imaging and Computer vision and software that make it appropriate for CNN applications in many image altering editing programs and related to the tasks of identification. Igor Sevo and Aleksej Avramovic(2016)[4] have observed a lot of every day procurement Aeronautical and satellite symbolism.Such a huge scale examination May be helpful for some useful uses of information.This Letter,present aerial content based analysis Symbolism to recognize and distinguish discretionary articles or zones High determination images For that purpose,they have proposed the method Detection of automatic object detection based on a convolutional neural network.A novel two-step method has been implemented for network trainingAnd aerial image classification functions and reviewedFind the object detection.At first,author have tried the proposed preparing technique Utilizing UCMerced Data Set of Aerial images and Accuracy Around 98.6%.Furthermore,the technique for robotized material Identification has been executed and checked.For usage GPGPU is the procedure time required for aerial measured size 5000 × 5000 pixels around 30s. Matias Valdenegro-Toro (2016)[5] Forward-looking sonars give high determination images that can be utilized for different submerged capacities Condition.In For this work, they have utilized Convolutional Neural Networks (CNN) Protest acknowledgment in the prospective Sonar images.They have appeared A sort of CNN surpasses the state of art images by accomplishing 99.2% precision. If there should arise an occurrence of state The art format has the precision between similarity techniques 92.4% and 97.6%. They have additionally think about the quantity of educated Parameters and layout similarity of CNN are required..To accomplish top achievement.results demonstrated that CNNs are required Low parameters to give better acknowledgment abilities Comprehend the missing information well. Somchai Pohtongkam and Jakkree Srinonchat.(2016)[6] This paper presents a recognition of images objects that are out of touch by capturing the texture of objects that existed in everyday life which sub-divided according to the different parts of the human palm. Data Vector Finding the summed value of each segment after setting ratio and converting to a binary image in each a series of tests using a segmentation of 15,20 and 26 Regions which will generate 15,20 and 26 Vectors respectively. Then, sorting the Vector data to perform Train process which consists of 300 series and another 300 series for testing. Militsyn A., Malykhina G and Malykhina G.(2016)[7] This paper suggests a scenario of aerial image processing, inclusive the neural networks based classification of images, filtering, binarization, searching for predetermined objects, detecting roads, and binding objects to the terrain on the basis of cross-correlation function.Neural network was applied for image classification as well as for objects classification. Classification errors did not exceeding 10% may be consider as a satisfied result because of diversity of objects shape, size and rotation.Meera MKTech and Shajee Mohan B S(2017) [8] student and Associate Professor of Shaji Mohan BS.The paper creators utilized image information from Coil-20p with 20 material classifications. There are 72 grey images in every classification, each image has a pixel value of 128x128, PNG and in the 80's - for the most part 8 classes. Every class has 10 subcategories. Fundamental classes are images of apple, auto, bovine, glass, pooch, steed, pear and tomato. Subcategories are imagesof various classifications of a similar query. Each subcategory has 41 pictures. Pictures are 256x256 in PNG design. The strategy utilized is a pictures of a picture recognizing from an arrangement of pictures, first utilizing the inquiry grouping to distinguish the picture class of the query. The principle phases of work incorporate extraction

and characterization of four highlights. Grouping Optimization Authority has executed two kinds of classifiers - GIST includes that utilization Invariant Feature Transform (SIFT) and the Vector Machine (SVM) characterization of help for utilizing the k-nearest neighborhood (kNN) order. At that point the query image and the gathered datasets are the elements of highlights removed from coil20P and eth80.The reference work is utilized by the kNN order and takes a shot at the SIFT-based query network and primary datasets of reference symbolism. Inquiry portrayal is characterized by characterization. The greatest precision of the GIST + SVM loop 20P is 89.32% and 98.18%, separately, with 55.63% and 93.33% individually for the 80s. Touqeer Ahmad et al.,(2017)[9] The creators of this paper have gotten information from a freely accessible Road Identification Data set that contains 1,443 pictures of 800 × 600 determination, with notes of ground truth of 27 classes on the first informational index. Strategy approaches they consider five distinctive CNN structures with an assortment of convolution/MaxPool and completely associated layers and an alternate determination of street stamp patches. They have utilized a freely accessible Road Identification Data set and incorporate information upgrade to expand the span of this informational index important to prepare profound screens. The improved informational collection is arbitrarily part into 70% and 30% for preparing and testing. The best CNN arrange comes about are the normal ID rate of 99.05% for 10-class street markings in the test gathering. Nishani, E et al.,(2017) [10] Deep Neural Networks for Eralda Nashani Video Analysis. Profound models are more viable and ordinarily utilized for PC vision with the reassure structure. With the presentation of Graphics Processing Unit (GPU) to universally useful issues, GPU has been centered around abusing preparing power for profound learning calculations. Moreover, high-dosage information online can successfully prepare profound anxious systems.They propose three distinct kinds of research orders: to enhance momentum CNN supplies, utilize human repeat neural systems (RNNs) and at last rely upon the unsupervised learning model to prepare NNS. Observation of previous work the existing system such as OCR and number recognition uses less parameters compare to object recognition. The size of the image, number of layers, dimension is less in these systems. Extraction based biometric systems collects a set of sample images and store it in a repository. During test phase it accepts the bio-metric sample of the person and verifies with the available database and checks for similarity in these systems mathematical and computational model to simulate human intelligence is not applied and each.

III.DATASET:

We chose to utilize 60,000 pictures with a 32x32 pixel measure CIFAR-10 database. Each class comprises of 10 classes that are exclusively separate from every classification comprising of 6,000 pictures (don't overlap). The pictures are small, obviously labeled and have no clamor which makes the dataset perfect for this assignment with impressively significantly less pre-processing.here are few pictures taken from the dataset.32x32 image of training and testing dataset.

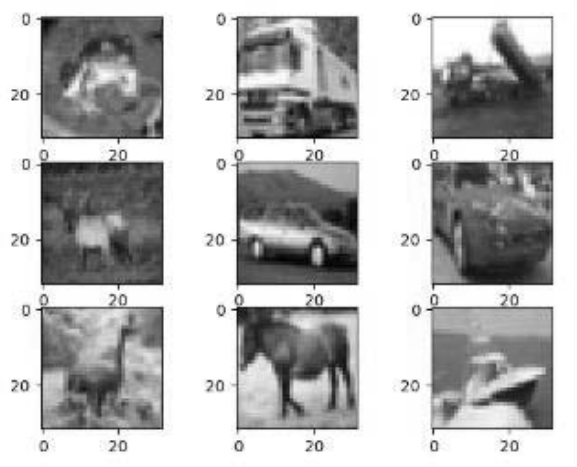

Fig2-Images from Dataset

IV.EXPERIMENTAL SETUP:

To compose the program we utilize Python and Tensorflow. Tensorflow is an open source deep learning structure that builds up the granular control of designers on each neuron(known as a "node" in tensorflow). So you can change the weight and accomplish the most ideal execution. The TensorFlow has many worked in libraries (some of which we use for image classification) and a superb group, so you can discover open source usage for any deep learning content. Computers can deal with calculations in numbers and cannot characterize images as we do. To comprehend the computers we have to change over images to numbers in any case.we consider following 5 features for feature extraction i. RGB- The colors can be represented by RGB values (going from 0 to 255, with red, green and blue). computer would then be able to extract the RGB estimation of every pixel and put the outcome in an array for interpretation.At the point when the

framework translated another image, it changes a range into the image utilizing a similar strategy, at that point looks at the examples of numbers against objects that it definitely knows. The framework at that point offers certainty scores for every class. The class with the most astounding certainty score is generally the anticipated one ii. Grayscale- The image is changed over to grayscale (white shading to the dim shades of dark) the computer assigns value in view of how dark every pixel is. Every one of the numbers are put into a cluster and the computer does calculations on that exhibit. iii. ZCA/PCA- Principle component analysis it helps to fetch relevant data,tries to change over information to as meager as could be allowed, so each row must be more like each other with basic basis functions (images with one active pixel). Furthermore, it is conceivable to accomplish, in light of the fact that the correlation in characteristic images are relatively neighborhood (so de-correlation channels can likewise be nearby). iv. EDGE-The image can essentially diminish the measure of information expected to process the edge detection calculation and thus filter information that can be considered less relevant, and additionally safeguarding the key basic attributes of the image. Edge detection is one of the necessary steps in image preparing,analysis,model recognition, and computer vision techniques. v. HOGP1-The quantity of pixels in f(with powers [0,L-1]) is a discrete capacity h(rk) = nk ,Where rk is the kth intensity value and nk is the quantity of pixels in f with intensity rk.The general routine with regards to normalizing the histogram is to partition the segments through the aggregate pixels in the picture and accept the MxN picture, which yields p (rk) = nk K = 0,1,2,...,L-1 – p(rk) The likelihood of event of intensity level RK in/mn, essentially, Σ p (rk) = 1 for L-1 - p (rk),in HOG p1.It is Straightforward, budgetary tool usage for histogram Image enhancements,Image statistics, Image compression,Image segmentation and computing software that are a prevalent tool for ongoing image processing.Keras used to bind tensorflow runtime with numpy array which are multidimensional. It manages the input image batches then transformed image batches and finding the image with the respective class.By using keras library it uses sequential and functional model Keras is capable of running on a high level neural system site API and Tensorflow written in Python. It has been created to empower quick trial. It is important to have good research if possible to get out of the imagination to cause at least possible delays. Keras Runs flawlessly on CPU and GPUs.CONV2D is used, example of CONVD2 is spatial convolution over images. This layer makes a convolution kernel, which is refined with layer input that produces items tensor. In the event that Use_bias is True, an incomplete vector is added to the structure and items. At long last, if there is no actuation, it applies to items. When utilizing this layer as the primary layer in the example, the catchphrase contention is input_shape (tuple of numbers, does exclude the sample axis) e.g. data_format = input_shape = (32,3,3) for 32x32 RGB pictures in data_format: "channels_last",and we consider i. kernel_size- An integer or tuple / list of 2 integers, indicates the width and stature of the 2D convolution window. It can be a integer number to indicate a similar value for every single spatial dimensions.ii. Padding- "valid" or "identical" (case-insensitive). iii. data_format- a string, channels_last (default) or channels_first one. The dimensions of the dimensions in the input will correspond to the inputs with the shape (batch, height, width, channels) but the channels will correspond to inputs with the core shape (batch, channels, height, width). It will be the default for the image_data_format value found in your keras configuration file in ~ / .keras / keras.json. If you do not set it up, it will be "channels_last". And finally iv. Activation- Function to use (see activations). If you do not specify anything, any activation does not apply (ie "linear" activation: a (x) = x).

V.PROPOSED SYSTEM:

The extraction based object recognition systems are with limitation such as it can't be scalable for unknown data. A machine learning algorithms are used to simulate human intelligence to classify, recognize and group the data. This features modeled in object recognition with images .Given an image the system accepts and process the image and gives it's category as an output. Extraction based systems are not associated with computational intelligence but proposed system is modeled with intelligence.

VI.ARCHITECTURE:

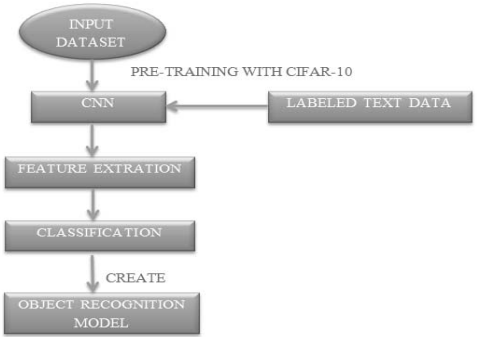

Fig3-Training Dataset

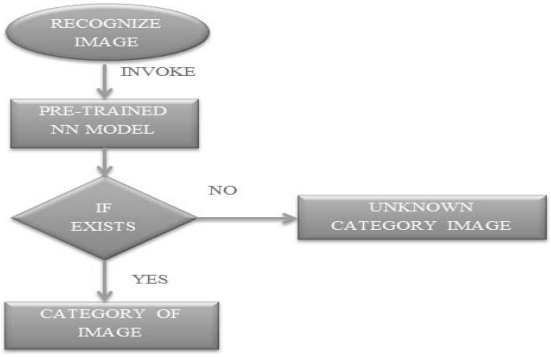

fig4- Testing Dataset

VII.RESULTS:

LAYER(TYPE)	OUTPUT SHAPE	PARAM#
conv2d_1(Conv2D)	(None,32,32,32)	896
dropout_1(Dropout)	(None,32,32,32)	0
Conv2d_2(Conv2D)	(None,32,32,32)	9248
Max_pooling2d_1(MaxPooling2	(None,32,16,16)	0
flatten_1(Flatten)	(None,8192)	0
Dense_1(Dense)	(None,512)	4194816
Dropout_2(Dropout)	(None,512)	0
Dense_2(Dense)	(None,10)	5130
Total params:4,210,090		
Trainable params:4,210,090		
Non-trainable params:0		

Fig5-Trained and Tested Tensorflow

VIII.CONCLUSION:

In this work we trained and tested object recognition in images using cifar-10 dataset and tested for single class such as aeroplain. We identified the issue of computing resources plays a major role in developing tensorflow based neural network systems. Since we experimented on cpu version of tensorflow it takes maximum time to create model with 25 epochs.We are planning to conduct experiment on tensorflow gpu version to improve the processing time and analyze the object recogniton system for more number of categories.

IX.REFERENCES:

[1]"Deep Neural Networks for Object Detection" Szegedy, C., Toshev, A., & Erhan, D. (2013). Deep neural networks for object detection. In *Advances in neural information processing systems* (pp. 2553-2561).,2013.

[2]"Object Recognition Using Deep Neural Networks: A Survey"Goyal,S.,&Benjamin, P. (2014). Object recognition using deep neural networks:A survey. *arXiv preprint arXiv:1412.3684*.,2014.

[3] "Using CNN for Image Recognition", S., Kumar, R., & Rowen, C. (2015). *Using convolutional neural networks for image recognition*. Tech. Rep., 2015.[Online]. Available: http://ip. cadence. com/uploads/901/cnn-wp-pdf.,2015.

[4]"Convolutional Neural Network Based Automatic Object Detection on Aerial Images" Ševo, I., & Avramović, A. (2016). Convolutional neural network based automatic object detection on aerial images. *IEEE Geoscience and Remote Sensing Letters*, *13*(5), 740-744.,2016.

[5] "Object Recognition in Forward-Looking Sonar Images with Convolutional Neural Networks"Valdenegro-Toro, M. (2016, September). Object recognition in forward-looking sonar images with Convolutional Neural Networks. In *OCEANS 2016 MTS/IEEE Monterey* (pp.1-6). IEEE.,2016.

[6]" Object Recognition from Human Tactile Image Using Artificial Neural Network",Somchai Pohtongkam and Jakkree Srinonchat. Pohtongkam, S., & Srinonchat, J. (2016, June). Object recognition from human tactile image using artificial neural network. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2016 13th International Conference on* (pp. 1-6). IEEE.,2016.

[7]" Application of Dynamic Neural Network to Search for Objects in Images", Militsyn A., Malykhina G. Militsyn, A., & Malykhina, G. (2016, May). Application of dynamic neural network to search for objects in images. In *Industrial Engineering, Applications and Manufacturing (ICIEAM), International Conference on* (pp. 1-3). IEEE.,2016.

[8]"Object Recognition in Images",S Meera M K M.Tech Student and Shajee Mohan B S Associate Professor. Lowe, D. G. (1987). Three-dimensional object recognition from singletwo-dimensionalimages. *Artificial intelligence*, *31*(3), 355-395.,2017.

[9]"Symbolic Road Marking Recognition Using Convolutional Neural Networks", Touqeer Ahmad .Ahmad, T., Ilstrup, D., Emami, E., & Bebis, G. (2017, June). Symbolic road marking recognition using convolutional neural networks. In *Intelligent Vehicles Symposium (IV), 2017 IEEE* (pp. 1428-1433). IEEE.,2017.

[10]"Computer Vision Approaches based on Deep Learning and Neural Networks"Nishani, E., & Çiço, B. (2017, June). Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation. In *Embedded Computing (MECO), 2017 6th Mediterranean Conference on* (pp.1-4).IEEE.,2017.