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

Scene Recognition using Convolution Neural Network

Tejas Prakash Gidde[#1], Shreyash Sadashiv Bokade[#2], Abhishek Sunil Hajare[#3]

Electronics and Telecommunication Engineering, Smt. Kashibai Navale College ofEngineering, Pune, India – 411041

ABSTRACT: Scene recognition is a task of computer vision. The project is all about to detect the scene with its attribute and category. Here, we are going to use convolutional neural network (CNN) to detect the scene. For image classification Convolutional Neural Network is effective. To get accurate results we must learn the deep features of image which is possible with the help of convolutional neural network. Convolutional neural network is a type of artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Neural network is a computational model that works in a similar way to the neurons in the brain. Each neuron takes an input, performs some operations then passes the output to the following neuron. Here, We are going to teach the computer to recognize the scene in image and classify them into one of the six categories such as Auditoriums, Bathroom, Bedroom, Farm, Forest and Swimming pool. This topic has received great deal of attention in computer vision because of its wide applications such as it's an important feature for driver-less car.

Keywords: Convolutional Neural Network, Image Classification, Prediction, Python.

1. INTRODUCTION

Scene recognition involves the detection of objects and scenes in an image. This topic has received a lot of attention in computer vision due to its wide applications, such as an important feature for driverless cars. We will use a convolutional neural network (CNN) for accurate scene detection. A neural network is a computational model that works in a similar way to neurons in the brain. Each neuron receives an input, performs some operations, and then passes the output to the next neuron. A convolutional Neural Network is a form of Artificial Neural Network whose connectivity pattern is inspired by the organization of an animal's visual cortex. The visual cortex has small areas of cells that are sensitive to specific areas of the visual field. Some individual neuronal cells in the brain only respond in the presence of edges of a certain orientation. For example, neurons fire when exposed to vertical edges and some when presented with horizontal or diagonal edges.

We will teach the computer to recognize the scene in the image and classify it into one of 6 categories such as auditorium, bedroom, bathroom, forest, farm and swimming pool. For example, we need to recognize a cat in a picture. To do this, we first need to teach the computer what a cat looks like before it can recognize a new object. The more cats the computer sees, the better it recognizes cats. With the help of CNN, the

computer begins to recognize patterns present in cat images that are absent from other images and begins to build its own cognition. Convolutional Neural Network (CNN) is one of the most popular techniques used in improving the accuracy of image classification. A CNN has a convolutional layer at the beginning that divides the image into a number of tiles, the machine then tries to predict what each tile is. Finally, the computer tries to predict what is in the picture based on the prediction of all the tiles. The computer extracts features from each tile, called in-scale features. In order to detect a scene, we need to combine all the features within a scale to create a multi-scale feature. To do this, we use the Fisher vector (FV) and Vector Linearly Aggregated Descriptors (VLAD) coding techniques from python, which summarize all within-scale to multi-scale features. Thecomputer predicts the scene based on multi-scale features.

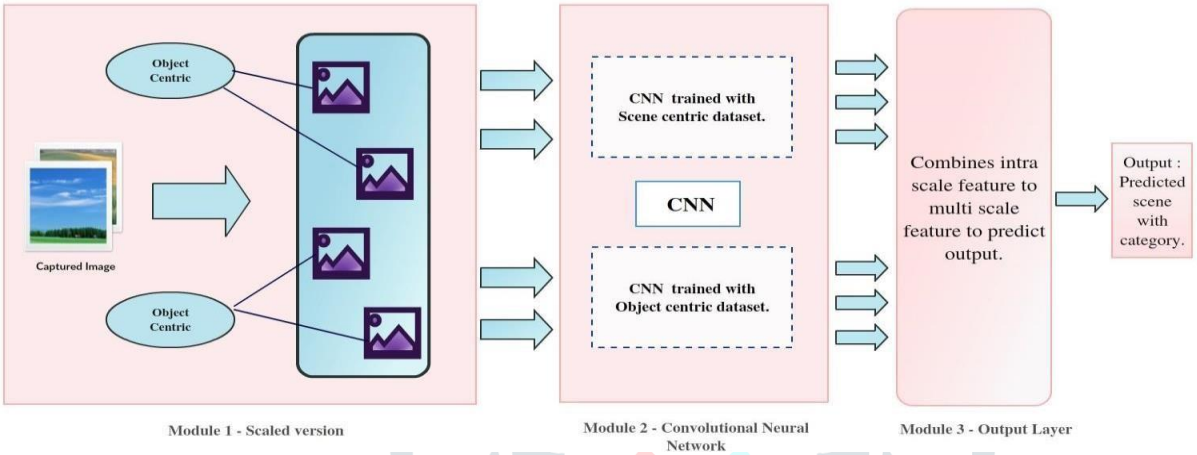
2. Literature Survey

Authors	Title	Publication & Year	Focus
Bolei Zhou1, Agata Lapedriza1, Jianxiong Xiao, Antonio Torralba1 and Aude Oliva1	**Learning Deep Features for Scene Recognition using Places Database**	Massachusetts Institute of Technology, Princeton University. (2015)	They introduced PLACES as a new dataset containing 7 million images from 476 places
Luis Herranz, Shuqiang Jiang, Xiangyang Li	**Scene recognition with CNNs: objects, scales and dataset bias**	IEEE Conference on Computer Vision and Pattern Recognition. (2016)	They implemented dataset bias of IMAGENET and PLACES to increases the accuracy up to 70%.
Bavin Ondieki	**Convolutional Neural Networks for Scene Recognition**	Stanford University . (2016).	They stated that, Convolutional neuralnetwork helps us to simulate humanvision which is amazing at scene recognition.

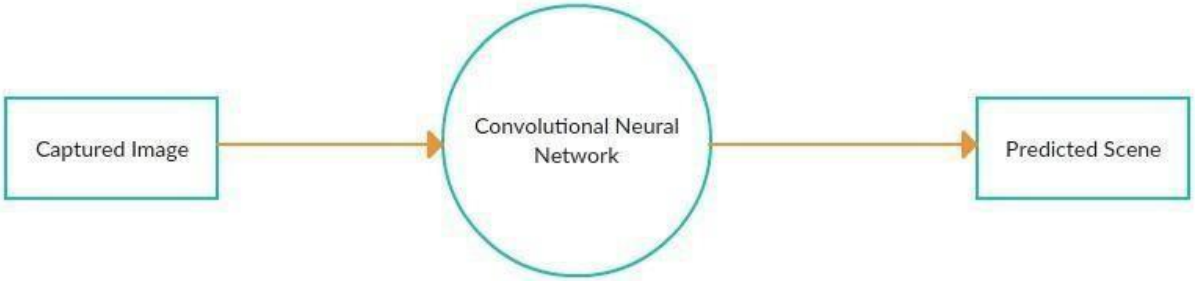
Reagan L. Galvez, Argel A. Bandala, Elmer Dadios, Ryan Rhay Vicerra, Jose Martine Maningo	**Object Detection Using Convolution Neural Network**	IEEE conferenceon proceeding of TENCON (2018).	They improved the PLACES dataset with adding extra 3 million images, containing 900 different categories

3. Methodology

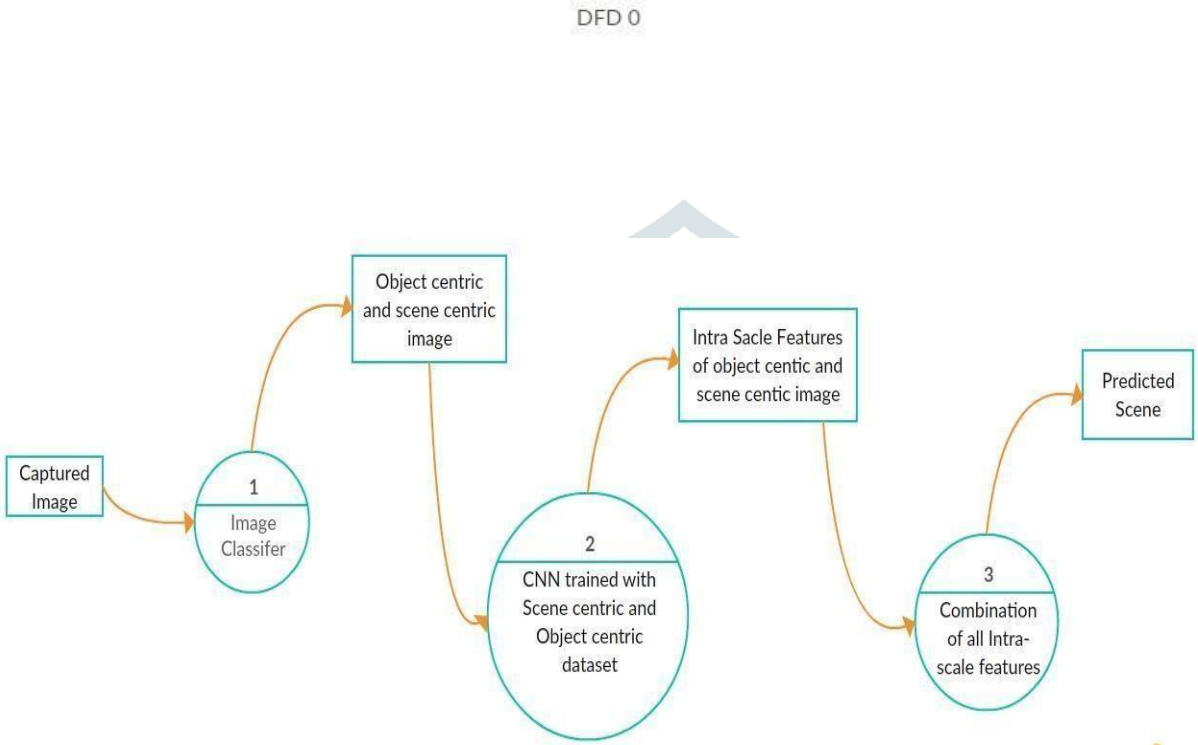
3.1 System Architect



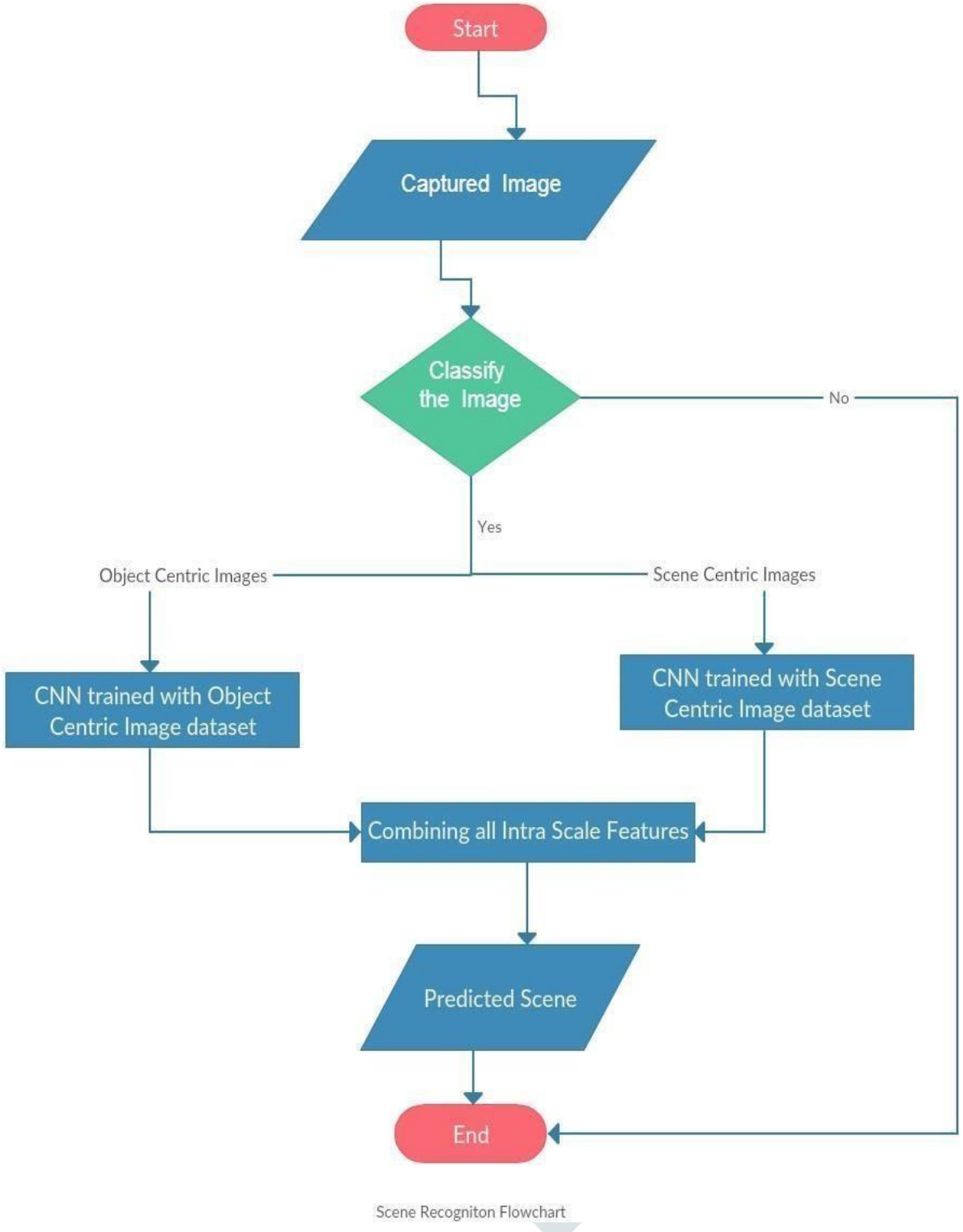
3.2 System Modeling Design



DFD Level-1



3.3 Flowchart



Scene Recogniton Flowchart

3.4 Algorithm

- Input Layer: Input will hold the raw pixel values of image. In this case an image of width 32 and height 32 and with three color channel (RGB).
- Convolutional Layer: In this layer compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a Input small region they are connected to in the input volume.
- Normalization (RELU): This layer will apply an element wise activation function such as max(0, x) thresolding at zero. This leaves the size of the volume unchanged[32*32*12].
- Max Pooling: The POOL layer will perform a down-sampling operation along spatial dimensions (width, height) resulting in volume such as [16*16*12]
- Fully Connected Layer: This layer will compute the class scores, resulting in volume of size (1*1*10) where as each of 10 numbers correspond to a class score.
- Output layer: This layer predicts the output.

4. Result & Discussion

We are going to build three modules. In **first module** the given input image will be biased into image centric and scene centric. The **second module** is CNN. We are going to build two specific CNN, one of the CNN will be specifically trained for extracting intra scale features of objects in image, the other CNN will be specifically trained for extracting intra scale features of scene in the image. The **last module** will combine all the intra scale features and it will predict the accurate scene.

Module 1: Scaled Versions:

The captured images is given as the input and will be biased into image centric andscene centric.

Module 2: Convolutional neural network (CNN).

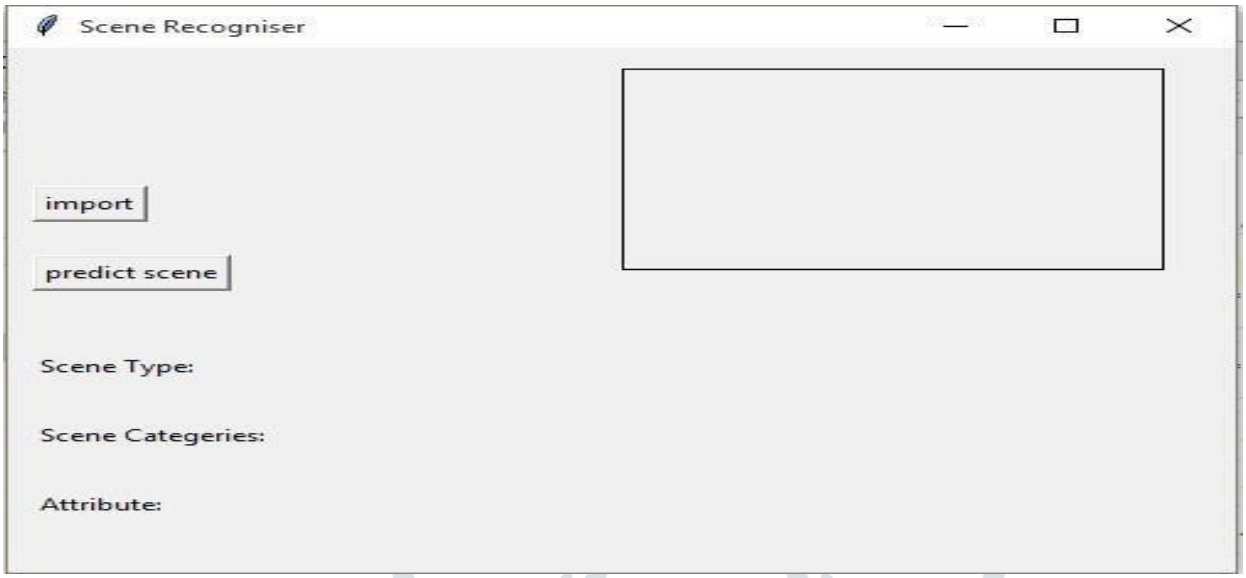
The biased images are classified with the help of the Convolutional neural network.

1.Input layer 2. Convolutional layer 3.Normalization 4.Max pooling 5.Output layer .

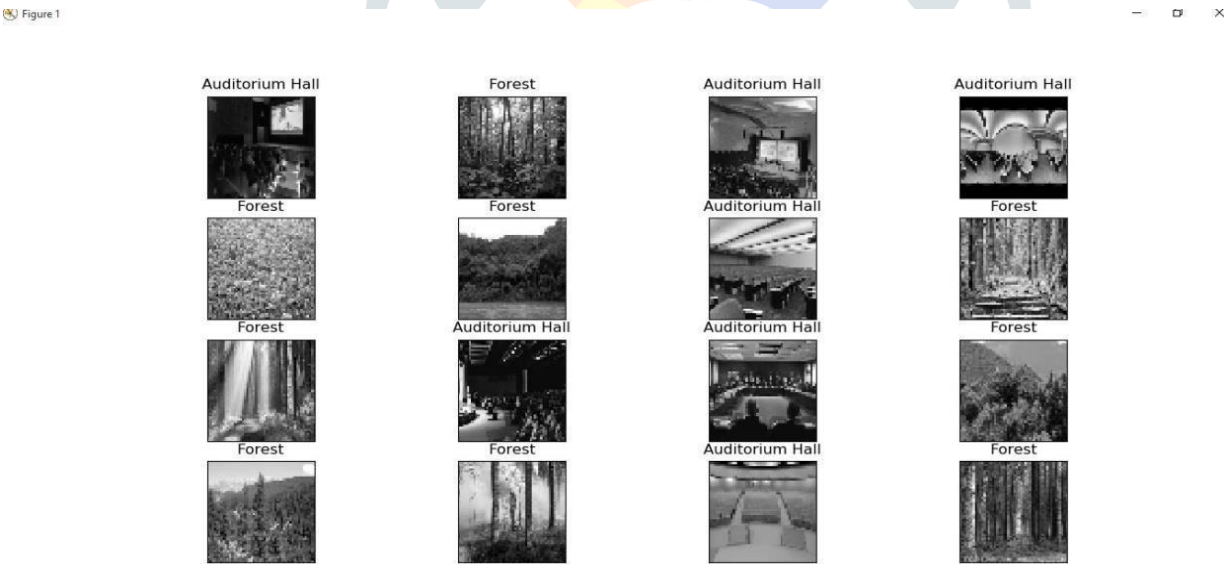
Module 3: Output Layer:

The last module i.e. output layer will combine all the intra scale features and it will predict the accurate scene. Intra-scale feature: The output given by the max pooling (CNN) is considered as the intra-scale output. Multi-scale features: It's combines all intra-scale feature and predicted the accurate scene.

This is the UI for our input prediction.



This is the UI for iur output prediction.



5. Conclusion

Convolutional neural networks (CNNs) have achieved astounding results in a variety of disciplines, including medical studies, and there is growing interest in radiology. Although deep learning has emerged as the dominant method for a wide range of complex tasks such as image classification and object detection, it is not a panacea. Knowing the main ideas and benefits of CNN, as well as the limitations of deep learning, is vital for using it in radiology research with the objective of increasing radiologist performance and, eventually, patient care.

6. Reference

[1] Bolei Zhou1, Agata Lapedriza1, Jianxiong Xiao, Antonio Torralba1 and Aude Oliva1, "**Learning Deep Features for Scene Recognition using Places Database**". Massachusetts Institute of Technology, Priceton University. (2015)

[2] Luis Herranz, Shuqiang Jiang, Xiangyang Li,"**Scene recognition with CNNs: objects, scales and dataset bias**". IEEE Conference on Computer Vision and Pattern Recognition. (2016)

[3] Bavin Ondieki,"**Convolutional Neural Networks for Scene Recognition**" Stanford University.(2016) .

[4] Reagan L. Galvez, Argel A. Bandala, Elmer Dadios, Ryan Rhay Vicerra, Jose Martine Maningo **"Object Detection Using Convolution Neural Network".** IEEEconference on proceeding of TENCON (2018).

[5] P. Agrawal, R. Girshick, and J. Malik. **Analyzing the performance of multilayer neural networks for object recognition.** In *Proc. ECCV*. 2014.

[6] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: **A deep convolutional activation feature for generic visual recognition**. 2014.

[7] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. **Cnn features off-the- shelf: an astounding baseline for recognition.** *arXiv preprint arXiv:1403.6382*, 2014.

[8] J. Sanchez, F. Perronnin, T. Mensink, and J. Verbeek. **Image classification with the fisher vector: Theory´ and practice.** *Int'l Journal of Computer Vision*, 2013.

[9] A. Wong, M. J. Shafiee, F. Li, et al." Tiny SSD: **A Tiny Single-shot Detection Deep Convolutional Neural Network for Real-time Embedded Object Detection**", arXiv preprint arXiv:1802.06488, 2018.C.