MJRET

OPEN ACCESS

Multidisciplinary Journal of Research in Engineering and Technology

www.mjret.in

eISSN:2348-6953

SCENE DETECTION USING CONVOLUTION NEURAL NETWORK (CNN)

Kushagra Jain, Rajot Saha, Merin Meleet, Rekha B S
Department of Information Science and Engineering
R V College of Engineering
Mysore Road, Bengaluru - 560059

Abstract: *Image classification is discipline of computer vision that deals with identifying and classifying objects from computer image based on certain properties. Image classification has a lot of applications such as robotics, smart traffic system, smart transportation and many more. The exponential growth of data in size and diversity cause major issue for image classification. Though, to detect and classify image from large dataset is require effective and efficient method of image classification. Therefore, machine learning algorithms are came into picture to achieve object detection. In this paper, scene detection is accomplished using machine learning algorithm named as convolution neural network (CNN). Scene detection is accomplished on Places Dataset. Places dataset is a repository of 10 million scene photographs, labeled with scene semantic categories, comprising a large and diverse list of the types of environments encountered in the world. The CNNs trained on Places shows that object detectors emerge as an intermediate representation of scene classification. With its high-coverage and high-diversity of examples, the Places Database along with the Places-CNNs offers a novel resource to guide future progress on scene recognition problems.*

Keywords: *Computer Vision, Image Classification, Scene Detection, Machine Learning, Convolution Neural Network, Places Dataset.*

1. INTRODUCTION

Computer vision is broad area which comprises of numerous disciplines [1]. Image classification is one discipline that deals with identifying and classifying objects from the computer image. Image classification [2] has several applications such as smart traffic control system, smart transportation, robotics, battlefield etc. Among many application, scene detection classify a scene image to one of the predefined scene categories (such as beach, kitchen, and bakery), based on the image's ambient content, objects, and their

layout. Visual scene understanding requires reasoning about the diverse and complicated environments that we encounter in our daily life. Identifying visual categories such as objects, actions and events is no doubt the indispensable ability of a visual system. Moreover, recognizing the scene where the objects appear is of equal importance for an intelligent system to predict the context for the recognized objects by reasoning “What is happening? What will happen next?” More about scene detection and convolution neural network, following sections presented.

A. Scene Detection

Scene detection is way to represent non sequential visual image [1]. There are some methods [3] have been advised for scene detection. One among of them focus the dc coefficients on an 8x8 block of a frame can form a sufficient representation of the image content as far as the scene cut detection is concerned.

B. Convolution Neural Network (CNN)

Convolution neural network is a type of deep neural network architectures [4]. It is also known as multilayer neural network that inspired by human perception. CNN has variety of application such as computer vision, natural language processing etc. The general and conceptual architecture of convolution neural network shows in figure 1 and 2.

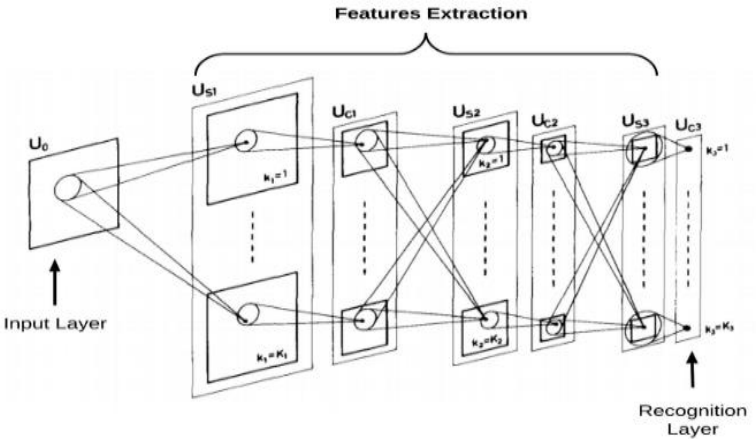


Fig.1 General layer architecure of CNN

CNN also named as ConvNet[4]. CNN is comprises of fininte set of layers which include input, convolution, pooling and output layer.

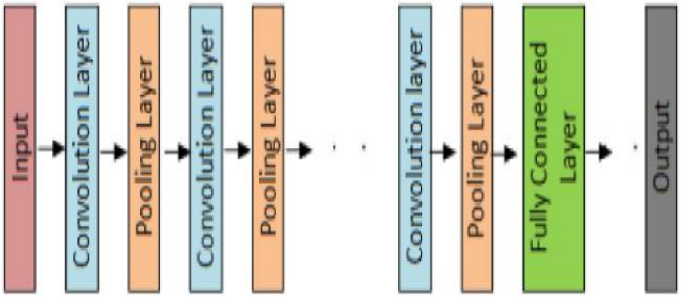


Fig.2 Conceptual Design of CNN

2. RELATED WORK

In every domain, research is continuous process. Though, some work on scene detection also suggested by researchers. This section explored some related work to scene detections.

An attempt happened to solve one of the most common problems of scene recognition [5]. The dataset called PLACES2 was used for the purpose of experimentation which contains nearly 7 million labeled pictures of various scenes to rain the model. The convolutional Neural Network was used to design a neural architecture for our machine to store and classify the images as desired for the recognition of scene in the dataset.

A new end-to-end scene recognition framework was proposed which called a Recurrent Memorized Attention Network (RMAN) model [6]. It performs object-based scene classification by recurrently locating and memorizing objects in the image. Based on the proposed framework, paper introduce a multi task mechanism that contiguously attends on the different essential objects in a scene image and recurrently performs memory fusion of the features of object focused by an attention model to improve the scene recognition accuracy.

A model [7] was proposed to understand and recognize a scene using depth data to make machines capable of interpreting the real time scenes like humans.

A new scene-centric database have introduced [8] called Places with over 7 million labeled pictures of scenes. Paper propose new methods to compare the density and diversity of image datasets and show that Places is as dense as other scene datasets and has more diversity. Using CNN, we learn deep features for scene recognition tasks, and establish new state-of-the-art results on several scene-centric datasets.

A network for Congested Scene Recognition called CSRNet [9] was proposed to provide a data-driven and deep learning method that can understand highly congested scenes and perform accurate count estimation as well as present high quality density maps. The proposed CSRNet is composed of two major components: a convolutional neural network (CNN) as the front-end for 2D feature extraction and a dilated CNN for the back-end, which uses dilated kernels to deliver larger reception fields.

3. PROPOSED METHODOLOGY

To build a system that takes an image as input, identify the objects present in it and predict the scene using CNN algorithm. The CNN algorithm helps to extract the high level features such as edges, color, and gradient from the input image. CNN helps to reduce the image into a form which is easier to process, without losing features which are critical for getting a good prediction. Objective of the system is to classify a scene image to one of the predefined scene categories (such as beach, kitchen, and bakery), based on the image's ambient content, objects, and their layout. To take an image as input, identify the objects present in the image, analyze the image and give the output displaying what the image is about.

Use CNN (Convolution Neural Network) to take in an input image, identify and assign importance (learnable weights and biases) to various objects in the image and be able to differentiate one from another. CNN results in a lower amount of pre-processing required as compared to other classification algorithms.

A. System Architecture

For the scene detection system, architecture shown in figure 3 is proposed that consist of training and test module. Training module deals with training of model or detector based on image dataset. After the training, Test module invoke that predict scene category for given input image. Test module comprises of input image, feature extraction, Trained classifiers etc.

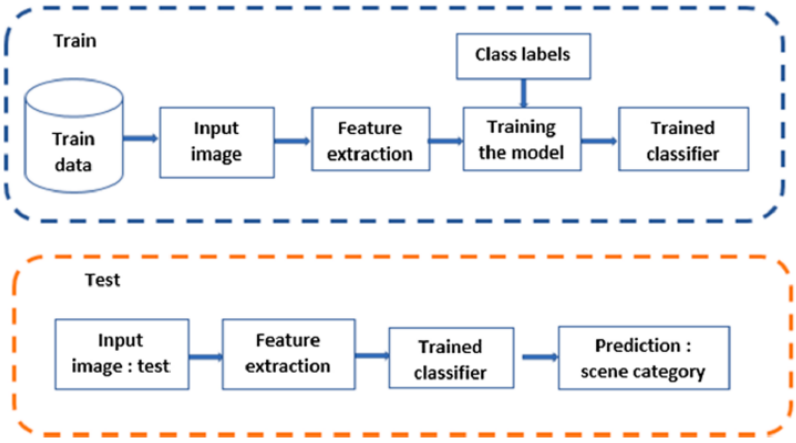


Fig.3 System architecure

B. System High Level Design

Apart from the architecture, high level and low level design of scene detection system also presented here. High level system design shown in figure 4 that represent broad steps of detector.

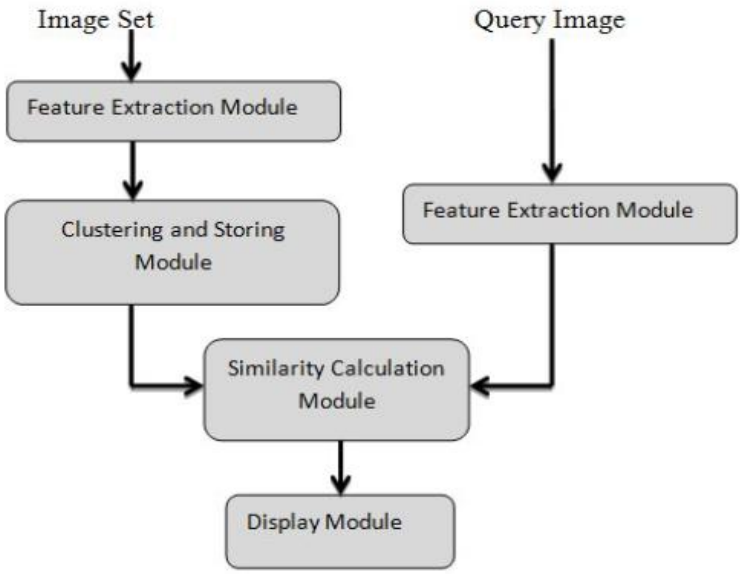


Fig.4 High level system design

C. System Low Level Design

Figure5 shows low level system design which presents detail working of each phase of detection system.

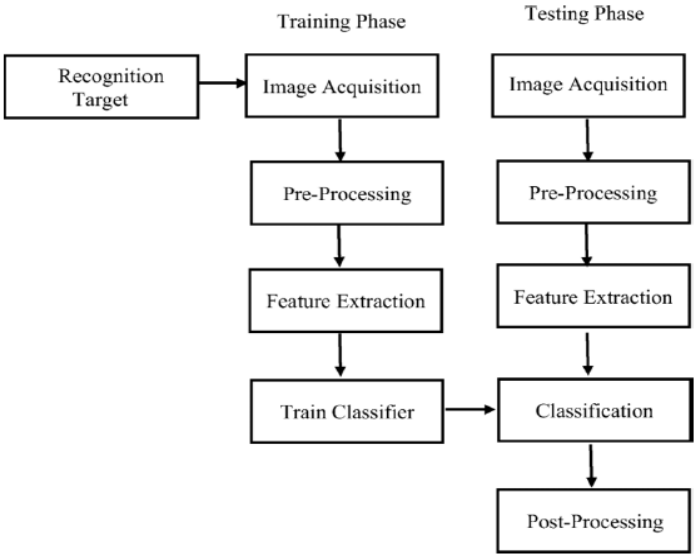


Fig.5 Low level system design

4. TESTING & RESULT ANALYSIS

Various tests were applied where the system was evaluated. The tests will be of various degrees of priorities ranging from high to low. It is very important that the system passes all the suites of higher priority, failure in doing so will result in revamp of the developmental approach. Passing of test cases of low priority may not be evaluated very strictly. Black-box testing is a method of software testing that examines the functionality of an application without peering into its internal structures or workings. This method of test can be applied virtually to every level of software testing: unit, integration, system and acceptance White-box testing is a method of software testing that tests internal structures or workings of an application, as opposed to its functionality. In white-box testing an internal perspective of the system, as well as programming skills, are used to design test cases.

A. Test Cases

Test scenario	Input Image File Format					
Test case description	Since the the input file for images should only be jpg format so we are testing with other formats too.			**Test priority**		High
Prerequisite	The web page should be opened and running.			**Post-requisite**		Material uploaded on portal
S.No.	**Action**	**Input**	**Expected output**	**Actual output**	**Test result**	
1	Launch the site	Visit the site where the uploading of image is done	Opened successfully	Opened successfully	**PASS**	
2	Upload image	The Image of any scene is uploaded	Image Uploaded	Image Uploaded	**PASS**	
3	Submit (FAIL)	Submit buuton is clicked for input of other file format like jpeg, png.	Unable to upload	**Unable to Upload**	**PASS**	
4	Submit (SUCCESS)	Submit buuton is clicked for input of jpg	Unable to upload	Unable to upload	**PASS**	

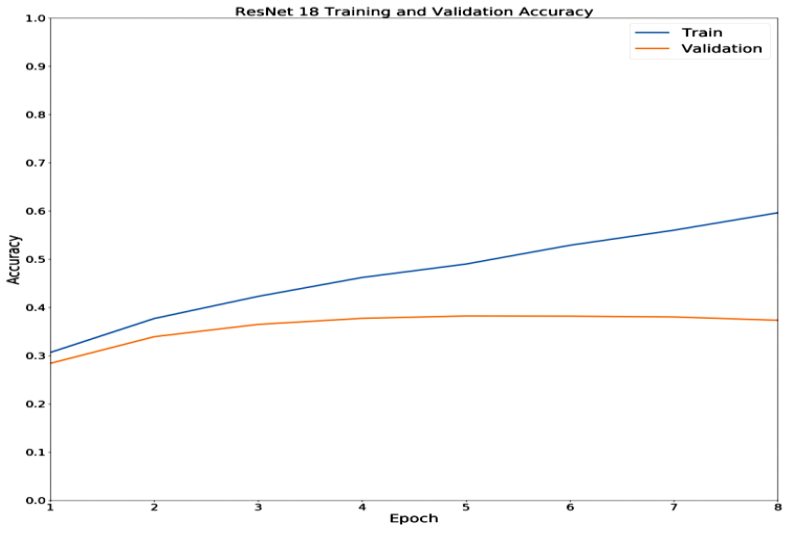
Test scenario	Single Object Detection					
Test case description	Try to detect single object			Test priority		Low
Prerequisite	The web page should be opened and running.			Post-requisite		Material uploaded on portal
S.No.	Action	Input	Expected output	Actual output	Test result	
1	Launch the site	Visit the site where the uploading of image is done	Opened successfully	Opened successfully	PASS	
2	Upload image	The Image of any scene is uploaded	Image Uploaded	Image Uploaded	PASS	
3	Submit (Normal Operation)	Submit button is clicked	Object gets detected	Object gets detected	PASS	
4	Submit (Exceptional Operation)	Submit button is clicked	Object does not gets detected rather other scene related to object gets detected	Object does not gets detected rather other scene related to object gets detected	PASS	

B. Result Analysis

Classification accuracy on the validation set and test set of Places365 shows in below table. We have used the class score averaged over 10-crops of each testing image to classify the image.

	Validation Set of Places365		Test Set of Places365	
	Top-1 acc.	Top-5 acc.	Top-1 acc.	Top-5 acc.
Places365-AlexNet	53.17%	82.89%	53.31%	82.75%
Places365-GoogLeNet	53.63%	83.88%	53.59%	84.01%
Places365-VGG	**55.24%**	84.91%	**55.19%**	85.01%
Places365-ResNet	54.74%	**85.08%**	54.65%	**85.07%**

The below graph shows Top-1 Training Accuracy per Epoch for 18 layer ResNet.



5. CONCLUSION

Deep convolution neural networks are designed to benefit and learn from massive amounts of data. The Places365 dataset is used that contains 10 million labeled images to detect and represent places and scenes found in the real world. We have identified the environment type (indoor or outdoor), attribute in the image and predict the scene in the input image according to the images in the dataset. Scene Detection project can also be applied to videos, where it will accept videos from user and identify the attributes, environment type and predict the scene. It could also be used for video summarization (Creating and publishing a shorter version of the video rather than the full video), video surveillance applications, handwriting recognition and many other applications.

REFERENCES

[1] *Y. Avrithis, A. Doulamis, N. Doulamis and S. Kollias, "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases," Computer Vision and Image Understanding, Vol. 75, No. 1/2, pp. 3-24, July/August 1999.*

[2] *A. Pentland, R. W. Picard and S. Sclaroff, "Photobook: Content- Based Manipulation of Image Databases," Int. J. Comput. Vision, Vol. 18, No. 3, pp. 233-254, 1996.*

[3] *Boon-Lock Yeo and Bede Liu, Fellow, IEEE, CSVT, "Rapid Scene Analysis on Compressed Video," Vol. 5, No. 6, Dec 1995.*

[4] *Ghosh, Anirudha & Sufian, A. & Sultana, Farhana & Chakrabarti, Amlan & De, Debashis. (2020). Fundamental Concepts of Convolutional Neural Network. 10.1007/978-3-030-32644-_36.*

[5] *Sarfaraz Masood, Umer Ahsan, Fatima Munawwar, Danish Raza Rizvi, Mumtaz Ahmed, Scene Recognition from Image Using Convolutional Neural Network, Procedia Computer Science,Volume 167, 2020, Pages 1005-1012, ISSN 1877-0509.*

[6] *Shao, Xi & Zhang, Xuan & Tang, Guijin & Bao, Bingkun. (2020). Scene Recognition Based on Recurrent Memorized Attention Network. Electronics. 9. 2038. 10.3390/electronics9122038.*

[7] *Ahmed, A.; Jalal, A.; Kim, K. A Novel Statistical Method for Scene Classification Based on Multi-Object Categorization and Logistic Regression. Sensors 2020, 20, 3871. https://doi.org/10.3390/s20143871.*

[8] *Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2014. Learning deep features for scene recognition using places database. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'14). MIT Press, Cambridge, MA, USA, 487–495.*

[9] *Li, Yuhong & Zhang, Xiaofan & Chen, Deming. (2018). CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. 1091-1100. 10.1109/CVPR.2018.00120.*