International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 01, 2019
ISSN: 1475-7192

Enumeration on the various tenets in Scene Recognition - Applications and Techniques

Bhavesh Shri Kumar, J. Naren, K. Prahathish and Dr.G. Vithya

Abstract--- *Scene Recognition is a task of great significance in computer vision. Certainly, it is not very easy due to various factors like cluttered image, poor separation of boundaries in between the scene objects, bad lighting, etc. Hence the topic receives a huge research attention. In this paper the various applications using Scene Recognition and the various techniques that are incorporated to classify, feature extract and to cluster scene images are reviewed.*

Keywords--- *Scene Recognition, Deep Learning, Artificial Neural Networks, CNN, RNN, RBM, DBN.*

I. INTRODUCTION

Currently under development are arrhythmia monitors for ambulatory patients whose Recognition of external environmental condition is one of the important tasks that any Artificial Intelligence (A.I) system must perform, especially if its functionalities depend on its interaction with the surrounding, like the Google's autonomous cars and Tesla's auto pilot. This identification of the external conditions comes under the scope of Scene Recognition. We human beings are experts in observing and identifying scenes. For our human brain because of its structured organization and neural architecture, Scene Recognition is an extremely simple task which we perform naturally without any training [11]. Scene Recognition is not limited to the above task alone its branches are Video Scene Recognition - where the input is a video, Audio Scene Recognition - where the inputs involve pre-recorded audio from the scene. Content based image retrieval (CBIR) is another field where scene object recognition is utilized, images stored in a database are queried and retrieved based on the contents of the image rather than using a keyword.

As the utility of Scene Recognition is large and irreplaceable, a lot of works are being done in the field, many methods are being used to classify, cluster and feature extract the scene images. The main intention to perform Scene Recognition is to classify the images containing the scenes into one or more categories based onsemantics

[12]. Methods used for classification of scene images include: PLSA, graph cut algorithm, bag-of-words method, probability distribution modelling, perceptual organization model, etc. Since the discovery of Deep Learning, the Deep Neural Network architectures are widely being deployed for various image recognition tasks. The advantage of DNN is that it has greatly simplified the task of feature extraction which otherwise has to be implemented by using traditional methods or manual extraction, which are very tedious tasks. In this paper we review the various methods that are already proposed for the scene classification and feature extraction. The second section contains a survey of the various applications dedicated to scene recognition; the third section is devoted for the review of the different

Bhavesh Shri Kumar, B.Tech Computer Science and Engineering School of Computing, SASTRA Deemed University, Thanjavur, India. E-mail: bhaveshshrikumar.n@gmail.com
J. Naren, Assistant Professor, School of Computing, SASTRA Deemed University, Thanjavur, India. E-mail: naren.jeeva3@gmail.com
K. Prahathish, B. Tech Computer Science and Engineering, School of Computing, SASTRA Deemed to be University, Thanjavur, Tamil Nadu, India. E-mail: prahathishk97@gmail.com
Dr.G. Vithya, Professor, School of Computing, KL University, Vijayawada, AP, India. E-mail: vithyamtech@gmail.com

Received: 14 Nov 2018 | Revised: 23 Dec 2018 | Accepted: 06 Jan 2019

methodologies and architectures that are incorporated to perform scene recognition. The fourth section gives a brief comparison of the works the method used and the dataset in which the method was validated.

II. APPLICATIONS DEDICATED TO SCENE RECOGNITION

Audio scene recognition

Recognition of the surrounding environment based on the audio information is referred as Audio Scene Recognition (ASR). The benefits of ASR are due to the fact that it is not affected due to variation in lighting or blur etc. It offers more privacy as in some scenes where picturization the scene content is restricted. Yan Leng et.al [1] in their work presented a topic model based ASR algorithm utilizing the document-event co-occurrence matrix to extract the topic distribution expressing a better audio document. Their method to obtain the document-event co-occurrence matrix is based on matrix factorization through topic model rather than creating a classification model, which is the usual procedure.

The algorithm adopts Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA) as the topic models. The proposed algorithm is executed in three parts: first, an audio vocabulary is created, then the document- event co occurrence matrix is generated and finally the ASR is performed based on the topic model. The proposed algorithm was tested in the IEEE AASP challenge dataset and DEMAND (contains recordings of noise from various environmental conditions) and was found to perform better than the method based on document-word co-occurrence matrix.

Animal recognition

Wildlife monitoring is done continuously to record the animal activities, their behavioral changes due to variations in climate, their movement and habitat pattern and to capture the human activities deteriorating the wild. Due to the fact that camera-traps are cost efficient and easy to use; they are widely used in detecting the wild. The basic working behind the camera-trap is: it will constantly monitor the area through the camera sensor and once it senses any movement in the forest scene it starts taking pictures of the scene and records other environmental parameters like temperature, humidity etc.

The challenge posed here is to correctly identify the motion of the animals or intruders and to neglect the motions due to shadow, movement of branches or ripples in water. Zhi Zhang et.al [2] proposed a method to detect the animals moving in the foreground from the cluttered background using a video object graph cut algorithm consisting of a background model and a foreground salience graph (FSG) which is constructed based on temporal salience and spatial salience and the graph cut minimization generates the foreground object separately. Now to verify that the segmented foreground acquired using the iterative embedded graph cut method is done using cross-frame animal object (image) verification method. The proposed method was tested using the CDnet dataset and is claimed to outperform Faster-RCNN method.

Lane detection

Detection of lane is an important task especially for creating driving automated vehicles which should be capable of identifying dangers on the road and to follow the correct lane on its own.

Received: 14 Nov 2018 | Revised: 23 Dec 2018 | Accepted: 06 Jan 2019

Jun Li et.al [3] in their work have taken the problem of structural prediction and detection of lanes in a traffic scene using Deep neural network. Automatic data representation is an advantageous feature of the DNN, which helps extracting features from the input automatically, but they failed in adopting task specific structures at the output end.

In [3] they have utilized two types of neural networks to perform the scene object recognition of the lane boundaries in traffic scene images. First a Deep Convoluted Neural Network (DCNN) is used identify the existence of lane marks and to estimate the position of the lanes. The identification part is done using a classifier and the estimation is done using a regressor. Secondly a Recurrent Neural Network is used over the CNN; this combination acts as a memory cell and enables learning of the output structures. The proposed model was trained over Inverse Perspective Mapped (IPM) images of real-world traffic scene image and is claimed to outperform the traditional classification methodologies utilizing SVM and Feedforward NeuralNetwork.

Words /Text detection

For the successful creation of autonomous cars it is necessary to provide them a mechanism to recognize words, though they mostly depend on the map for routes sometime there may be a take diversion or road closed sign which may not be shown in the map, these signs often involve words, the vehicles must hence be able to make instantaneous changes in the routes by recognizing the words in the scene. In any scene there is great chance of textual regions which often may contain the most valuable information. Many a time it is difficult to recognize words in scenes due to different font styling, character touching, distortions in the image and due toblurring.

In their work on recognition of words in scenes Bolan Su et.al [4] have devised a word level approach using two multi-layered RNN trained with Long Short-Term Memory (LSTM) and Connectionist temporal Classification (CTC) where each word is treated as a whole and no character segmentation is done. Two feature sets were created by converting the word image into sequences of column feature based on Histogram of Oriented Gradients. Then the two multi-layered RNNs were trained classify the two sets and CTC technique is used for calculating the score of each word for each RNN at last the scores are ensembled to get the output.

Tong Le et.al [5] in their work have presented a model: Text-Attentional Convoluted Neural Network (Text-CNN) to identify the regions involving textual information in a scene. The Text-CNN is applied to the scene image and deep features are extracted which are then learnt through a proposed multitask learning mechanism. Finally, the components that are highly ambiguous are detected using an extended Maximally Stable External Regions (MSERs) method called as CE- MSERs.

Crowded scene

Scene Recognition techniques are being used to monitor video scene images for any unusual happenings, this is termed as Anomaly Detection. Anomalies in scenes can be footage of a robbery or a murder; it can even be detection of medical errors. Mohammad Sabokrou et.al [15] in their work have proposed a model using a hierarchical representation of Deep Neural Networks. Two Deep Neural Networks are deployed to construct two cascaded classifiers.

Received: 14 Nov 2018 | Revised: 23 Dec 2018 | Accepted: 06 Jan 2019

Irregular motion and speed of the motion of an object in the frame of the scene are taken as constraints to identify the anomalies in the scene. The model was validated using the UCSD and UMN benchmark datasets.

III. SCENE IMAGE CLASSIFICATION METHODS AND ARCHITECTURES

Edge analysis

Andrew Payne et.al [6] has proposed a method to classify the scene images as indoor or outdoor scene based on the straightness of the edges. The basic idea that they have proposed is: outdoor scenes have a lower proportion of edges that are straight when compared to indoor scenes due to more visual clutter and randomness in them. Using low-level edge analysis, features are obtained and the classification is done in two-stages: first, the scene image is divided into a large number of smaller regions and each smaller region is then classified as indoor or outdoor, finally the classification of the original bigger image is done based on the class assigned to the smaller regions.

CNNDBM

Jingyu Gao et.al [7] presented a model to recognize natural scene based on Convoluted Neural Network and Deep Boltzmann Machine (DBM). For the recognition of hand written digits the Deep Boltzmann Machine is claimed to have attained the best results but due to the involvement of large number of iterations and matrix operations in the scene image recognition the DBM faces higher computational complexity. Hence the scene images are pre-processed using the Convoluted Neural Networks for dimensionality reduction which is now given as input to the DBM for training. The training methodology utilizes the contrastive divergence algorithm and the scene image recognition part is done by SoftMax regression technique.

Holistic Understanding

In their work Jian Yao et.al [8] have proposed an approach to perform scene recognition based on a holistic understanding of the scene i.e., by knowing the semantics of the region, the type of the scene can be predicted efficiently. The task of segmentation of the image based on boundaries was done using the contour detection method using hierarchical segmentation of the image and to verify whether a class is present in the scene and to check the correctness, an object detector based on discriminatively trained part-based model is utilized.

Background recognition

Chang Cheng et.al [9] in their work presented a model to segment a scene image based on the perceptual organization model (POM). The boundary of the foreground is detected using the idea that background of the image usually contains unstructured objects with high amount of overlaps, clutter and objects often are random, and don’t have ordered structure like that of the foreground objects. The information regarding the object appearance is represented as textons, which are clustered to form a texton map. A histogram of textons represents each individual region in the scene image. To classify the scene region objects a Binary AdaBoost classifier is trained with the above data and background regions i.e. unstructured regions are identified. The scene image is segmented into the classified regions before applying the POM for boundary detection.

Received: 14 Nov 2018 | Revised: 23 Dec 2018 | Accepted: 06 Jan 2019

Manifold Regulation

Yuan Yuan et.al [10] presented a Deep Architecture based on Manifold Regularization to recognize scene images. The network proposed is a kernel embedded deep architecture which utilizes the data's structural information.

Locally Linear Embedding (LLE) is used to minimize the dimensions and to learn the data's structural information. By modelling the data's correlations that are of higher order, a multi-layered Deep Learning model was developed. The model was tested on the 15-scene data set, eight sports data set and the SUNdataset.

Alexnet

Based on Alex Net Deep Neural Network architecture, Jing Sun et.al [12] have worked on to build a model for scene image classification.

They have used the Alex Net to extract features from the training scene images and a Support Vector Machine is trained using the features derived from the last layer of the Alex Net model. The trained SVM is used to classify the test images. The model was tested using the ImageNet2012 dataset and NUS-WIDE data et and is found to be greater accuracy. The proposed model's generalization performance is enhanced by near- suppression technique.

Centered convoluted RBM

A standard Restricted Boltzmann Machine does not include the details regarding the spatial relationships and hence to reduce the instability caused due to the independencies in the features extracted from adjacent patches Jingyu Gao et.al

[13] proposed a model for Centered Convoluted Restricted Boltzmann Machine. They have reformulated the energy function to get the centered factors.

They performed the training of Centered Convolution Deep Belief Network layer by layer and softmax regression is used to classify the scene images. The model was tested on the MIT Indoor scene database, MIT Places 205 database, Caltech 101 dataset, sun 397 dataset.

GoogLeNet

Pengjie Tang et.al [14] present a CNN based model for Scene Recognition utilizing the GoogLe Net architecture. At each layer of a Convoluted Neural Network the features represent different characteristics of the same image and hence this work is based on the fusion the features generated at each stage to enhance the task of scene image identification.

They have used three groups of convolutions to generate multi-stage feature vector using the product rule and the feature vectors produced by the three parts are concatenated to produce a vector of size 3072. Then the feature vectors are normalized and fed to a Support Vector Machine for classification. The model was evaluated using the places 205 dataset; SUN 397 dataset, Scene 15 dataset and MIT 67 datasets.

Received: 14 Nov 2018 | Revised: 23 Dec 2018 | Accepted: 06 Jan 2019

Comparative Study of Various Works In Scene Recognition

Table 1.1: Comparative Studies of various works in scene recognition

S.no	Title of paper	Author/s	Technique used	Data Set
1	Animal Detection From Highly Cluttered Natural Scenes Using Spatiotemporal Object Region Proposals and Patch Verification	Zhi Zhang Zhihai He Guitao Cao Wenming Cao	Spatiotemporal o bject r egion proposal Patch verification	CDnet dataset
2	Accurate recognition of words in scenes without character segmentation using recurrent neural network	Bolan Su Shijian Lu	Recurrent Neural Network	ICDAR 2003,2011,2013 Google street view
3	G-MS2F: GoogLeNet Based multi-stage feature fusion of deep CNN for scene recognition	Pengjie Tang Hanli Wang SamK wong	Convoluted Neural Network	Places 205 SUN397 Scene15 MIT67
4	Deep Neural Network For Structural Prediction and Lane Detection in Traffic Scene	JunLi, Xue Mei Danil Prokhorov, Dacheng Tao	Deep Neural Network	IPM images of real world traffic Scene image
5	Text-Attentional Convolutional Neural Network for Scene Text Detection	Tong He Weilin Huang, Yu Qiao Jian Yao	Text-CNN	ICDAR 2013 data set
6	Natural Scene Recognition Based on Convolutional Neural Networks and Deep Boltzmannn Machines	Jingyu Gao Jinfu Yang Jizhao Zhang Mingai Li	Convoluted Neural Network Deep Boltzmann Machine	SIFT flow 15 scene dataset
7	Indoor vs. outdoor scene classification in digital photographs	Andrew Payne, Sameer Singh	Straightness of edges	872 consumer digital photograph
8	Describing the Sceneas A Whole: Joint Object Detection, Scene Classification and Semantic Segmentation	Jian Yao Sanja Fidler Raquel Urtasun	Holistic Scene Understanding	MSRC-21 dataset
9	Outdoor Scene Image Segmentation Based on Background Recognition and Perceptual Organization	Chang Cheng, Andreas Koschan Chung-Hao Chen, David L.Page, Mongi A.Abidi	Perceptual Organization Methods ,	Gould dataset Berkeley segmentation data set
10	Scene Recognition by Manifold Regularized Deep Learning Architecture	YuanYuan Lichao Mou, Xiaoqiang Lu	Deep Learning	15 scenes data set Eight sports dataset SUN 397
11	Audio scene recognition based on audio events and topic model	Yan Leng Nai Zhou Chengli Sun Xinyan Xu QiYuan Chuanfu Cheng Yunxia Liu	Event Topic Model	IEEE AASP challenge DEMAND

Received: 14 Nov 2018 | Revised: 23 Dec 2018 | Accepted: 06 Jan 2019

IV. CONCLUSION

Thus, in this work we have reviewed the various methodologies and architectures that have been used for Scene recognition task. The scene recognition has great importance in achieving general purpose AI especially in computer vision, thence it has become the need of the hour to create newer technologies and methods which enhance the identification task and achieve perfection comparable to that of human level or more. Certainly it is not an easy task due to the various threats posed to it at various levels: distortion in the scene image, lack of proper separation of boundaries in between the scene objects - this happens especially in outdoor scene images, bad lighting, falling shadows are some of the threats to scene image recognition. Due to the recent discoveries in Neural Network architectures like the RecurrentNeural Networks, Convoluted Neural Networks, Convoluted Restricted Boltzmann Machines and the various Deep learning architectures the job of extraction of features has become simpler and easier.

REFERENCES

[1] Y. Leng*et al.*, "Audio scene recognition based on audio events and topic model," *Knowledge-Based Syst.*, vol. 125, pp. 1–12, 2017.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Z. Zhang, Z. He, G. Cao, and W. Cao, "Animal Detection from Highly Cluttered Natural Scenes Using Spatiotemporal Object Region Proposals and Patch Verification," IEEE Trans. Multimedia., vol. 18, no. 10, pp. 2079–2092, 2016.K. Elissa, "Title of paper if known,"unpublished.

[3] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep Neural Network for Structural Prediction and Lane Detection in Traffic Scene," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 3, pp. 690–703, 2017.

[4] B. Su and S. Lu, "Accurate recognition of words in scenes without character segmentation using recurrent neural network," *Pattern Recognit.*, vol. 63, no. June 2016, pp. 397–405, 2017.

[5] T. He, W. Huang, Y. Qiao, and J. Yao, "Text-Attentional Convolutional Neural Networks for Scene Text Detection," *arXiv Pre-print*, vol. 25, no. 6, pp. 1–10,2015.

[6] A. Payne and S. Singh, "Indoor vs. outdoor scene classificationin digital photographs," *Pattern Recognit.*, vol. 38, no. 10, pp. 1533–1545, 2005.

[7] J. Gao, J. Yang, G. Wang, and M. Li, "A novel feature extraction method for scene recognition based on Centered Convolutional Restricted Boltzmann Machines," *Neurocomputing*, vol. 214, pp. 708–717, 2016.

[8] J. Yao, T. Chicago, S. Fidler, and R. Urtasun, "Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation," 2012.

[9] C. Cheng, A. oschan, C. H. Chen, D. L. Page, and M. A.Abidi, "Outdoor scene image segmentation based on background recognition and perceptual organization," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1007–1019, 2012.

[10] Y. Yuan, L. Mou, and X. Lu, "Scene Recognition by Manifold Regularized Deep Learning Architecture," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 10, pp. 2222–2233, 2015.

[11] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning Deep Features for Scene Recognition using Places Database," *Adv. Neural Inf. Process. Syst. 27*, pp. 487–495, 2014.

[12] J. Sun, X. Cai, F. Sun, and J. Zhang, "Scene image classification method based on Alex-Net model," *2016 3rd Int. Conf. Inf. Cybern. Comput. Soc. Syst. ICCSS 2016*, pp. 363–367,2016.

[13] J. Gao, J. Yang, G. Wang, and M. Li, "A novel feature extraction method for scene recognition based on Centered Convolutional Restricted Boltzmann Machines," *Neurocomputing*, vol. 214, pp. 708–717,2016.

[14] P. Tang, H. Wang, and S. Kwong, "G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition," *Neurocomputing*, vol. 225, no. November 2016, pp. 188–197, 2017.

[15] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, "Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes," *IEEE Trans. Image Process*, vol. 26, no. 4, pp. 1992–2004, 2017.

[16] D. Khosla, R. Uhlenbrock, and Y. Chen, "Automated scene understanding via fusion of image and object features," *2017 IEEE Int. Symp. Technol. Homel. Secur. HST 2017*, pp. 15–18, 2017.

[17] B. Ondieki, “Convolutional Neural Networks for Scene Recognition,” pp. 2–8.
[18] T. Zhang and Q. Wang, “Deep Learning Based Feature Selection for Remote Sensing Scene Classificatio,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 1–5,2015.
[19] A. Helou and C. Nguyen, “Unsupervised Deep Learning for Scene Recognition,” pp. 1–10, 2011
[20] W. Zhong, “Movie scene recognition with Convolutional Neural Networks,”2015.
[21] X. Lu, X. Li, and L. Mou, “Semi-supervised multitask learning for scene recognition,” *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1967–1976, 2015.

Received: 14 Nov 2018 | Revised: 23 Dec 2018 | Accepted: 06 Jan 2019