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

# A Novel Architecture based on Deep Learning for Scene Image Recognition

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

***Abstract---*** *Scene Recognition is a work of great significance in computer vision and artificial intelligence. Certainly, it is difficult 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 and a method based on DNN is proposed.*

***Keywords---*** *Scene Recognition, Artificial Neural Networks, Deep Neural networks, Convoluted Neural Networks.*

## I. INTRODUCTION

Due to the rapid advancements in technology, there is a need to develop more sophisticated Artificial Intelligent systems which possess the capabilities more than what we human beings have. This finally made us realize that the best architecture for intelligence was always within us- Our Brain. Starting from here people have tried to mimic the brain's neural architecture to the fullest extent possible and we have ended now in what is called Deep Learning or Deep Neural Networks. And especially this architecture based on our brain works undoubtedly very well in what we humans are exponents: Identifying patterns, Recognizing images, objects in images, speech, texts and natural languages. Nowadays this architecture is being used for building systems for various purposes like virtual assistants, recommendation systems, automatic translation of given text, playing games like chess, go etc.

The ultimate aim of our quest is to develop a General-purpose A.I system in which a single system could do anything a human counterpart can do and much more. A thing that makes human achieve all difficult feats easily is our visual system. It includes the organs taking inputs of visual data and the organs processing it to produce valuable knowledge out of it. Similarly, most A.I systems depending on the environment for their task accomplishments must possess a strong visual system which is termed as computer vision.

Thus, there is an ever growing need to develop systems having capabilities to accept the visual data and produce intelligent actions out of it. So Scene Recognition becomes an important task that needs to be performed to ulti-l the environment and requests a great attention. In this work we try to produce a novel architecture based on Deep Learning for scene image recognition.

*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*
*Dr.G. Vithya, Professor, School of Computing, KL University, Vijayawada. E-mail: vithyamtech@gmail.com*
*K. Prahathish, B.Tech Computer Science and Engineering, School of Computing, SASTRA Deemed University, Thanjavur, India. E-mail: prahathishk97@gmail.com*

**Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019** 

## II. RELATED WORKS

Researchers across the globe have always strived to improvise scene image processing and to encompass various state of art technologies into this feat. This section provides a brief note on the notable previous works related to this paper. Chang Cheng et.al [9] used perceptual organization model to segment the scene image and to detect the foreground boundary from the background using the fact the objects at the background of the scene image is cluttered with unstructured objects. The segmented scene image is classified using a Binary Adaptive Boost classifier. Jingyu Gao et.al used Convoluted Neural Network reducing the dimensionality of the features and Deep Boltzmann Machine for the classification of scene image recognition.

Andrew Payne et.al used the idea that outdoor scenes possess a lower amount of edges that are straight due to the unevenness in the outdoor scene objects. Their method the used for the classification was based on low-level edge analysis. Yuan Yuan et.al [10] presented a manifold regularization technique that used a kernel embedded Deep architecture for the structural information present in the data. A ulti-layered deep model was constructed by ulti-lay the correlations present in the data.

Jain Yao et.al proposed a model using the holistic understanding of the scene image. Contour detection methods were used to segment the scene image and to identify the presence of a particular class of objects in the scene. There are various neural network architectures that have been tested for accuracy performance like the Alex Net, GoogLe Net, ZF Net, VGG Net, Microsoft ResNet, etc and researchers have utilized this available model to perform the scene image recognition task by training them with different benchmark datasets.

Based on Alex Net DNN architecture, Jing Sun et.al [12] created a model for scene image classification. Alex Net has been used by them to extract features from the training scene images and an SVM is trained using the features and classify thetest images. They enhanced the model's generalization performance by near- suppression methodology.

Pengjie Tang et.al [14] presented a model based on Convoluted Neural Networks utilizing the GoogLe Net architecture. At each layer of a CNN the features represent different characteristics of the same image and the features generated at each stage have to be concatenated in order to get a complete feature vector for the task of scene detection. The enhanced feature vector is then normalized and used to train a SVM for the task of classification.

Jingyu Gao et.al [13] used a modified Restricted Boltzmann Machine to include the spatial details present in a scene image for the Scene Recognition task which a standard RBM lack. Their model was a centered convoluted RBM which has reduced instabilities as the information related to the adjacent patches are also being taken into consideration. The centered factors were obtained and a trained CCDBN with softmax regression performed the recognition task. For the recognition of words present in scene images Bolan Su et.al [4] proposed an approach at the word level that utilized two multilayered Recurrent Neural Networks trained with Long Short-Term Memory and Connectionist temporal Classification here every word is considered as a whole and no character wise is performed over the words. Two feature sets were created by converting the word image into sequences of column feature based on HOG. Then the RNNs were trained classify the feature sets.

**Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019** 

In the work [5] researchers have presented a Text-Attentional Convoluted Neural Network to detect the presence of text in the scene images.

The Text- Attentional CNN is applied to the image of the scene and a multitasking learning algorithm is devised to learn from the extracted deep features. Finally, the highly ambiguous components present in the scene are identified using CE-MSERs Mohammad Sabokrou et.al [15] in their proposed model used hierarchical representation of DNN. They deployed Deep Neural Networks to construct two cascaded classifiers. Speed of objects and irregular of the motion of an object taken as conditions to identify the anomalies in the scene.

## III. PROPOSED SYSTEM ARCHITECTURE

Our proposed framework is as shown in fig.1. In the first level the dataset that is to be used to train the model has to be pre-processed. There are varieties of datasets available for the scene recognition task. Some benchmark datasets include MIT Places 205 dataset, Scene 15 dataset Caltech 101 dataset, Sun397 dataset, ImageNet 2012 dataset, MIT 67 dataset, SIFT flow dataset, MIT Indoor dataset, etc. These datasets have been specially created for the training of scene recognition models.

The images are the data points for this process and pre- processing enables them get fitted into our Deep Learning model.

The images are reduced to the prescribed dimension that maximizes the performance of the model. The images may also possess portions of noise which will deteriorate the accuracy and performance of the model. The noise may be present in the images in various forms: it may be damaged pixel information or unwanted information. The noise present in the image has to be cleared off at the outset. Adaptive filtering technique has been utilized for this purpose as this method produce better images than the linear filtering techniques. The adaptive filter when compared with the linear filters is more selective and hence preserves the edges and other parts of the image with high frequency.

Next the model is trained to learn for the Deep features. The features are extracted by the model and since the features at each layer of the network are different and they all must be included in the final vector of features. The features at each layer have been made into a feature vector and are used for the training of our proposed model. The Deep neural network classifier is deployed to classify the scene images and to verify the accuracy performance with the benchmark dataset.

The accuracy of the proposed methodology is better than the native traditional techniques that were used for this purpose previously.

### *A. Algorithm for Scene Image Recognition*

Step 1: Input the complete scene Images. Step 2: Pre-process the images using

Step 3: Report the corresponding Patterns for the Image.

Step 4: Classify and analyse the pattern using Deep Neural Classifier. Step 5: Obtain the Split Images

Fig 1.3.1. explains the overview of scene/image recognition and prediction based on the algorithm mentioned above.

**Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019** 

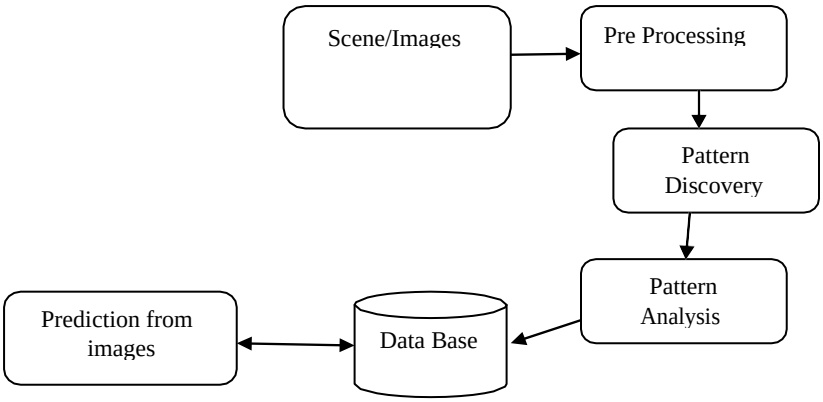


Fig 1.3.1: Overview of scene/image recognition and prediction

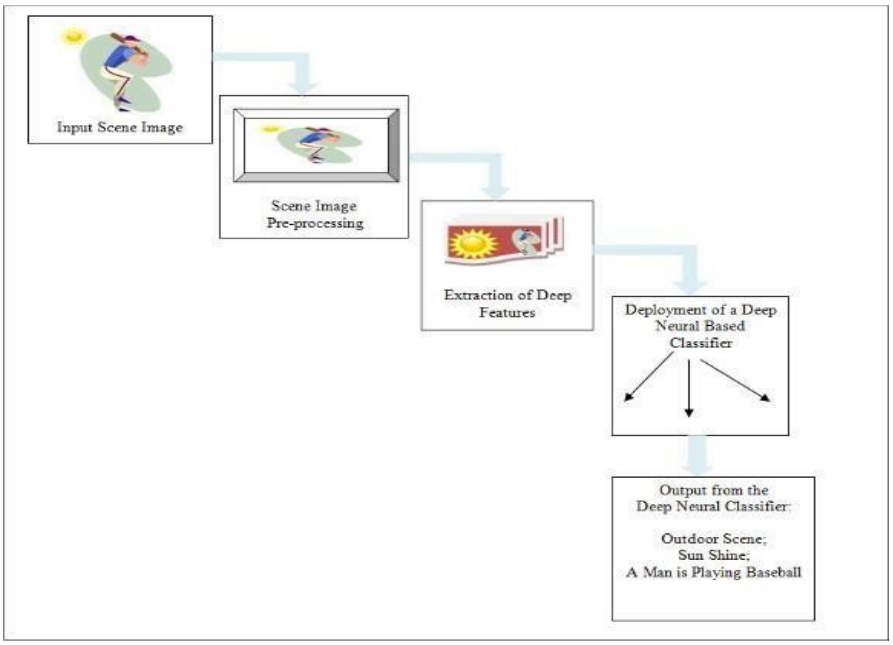


Fig.1.3.2: Diagrammatic representation of the proposed Deep Neural Network based classifier for Scene image recognition

## IV. CONCLUSION

Thus in this work we have proposed a method and architecture for Scene recognition task based on Deep Neural Network. 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 scene recognition is not very easy due to the various threats posed to it at various levels like 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 etc. Due to the recent discoveries in Neural Network architectures like the Recurrent Neural 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.

**Received: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019** 

## 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. 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. Multimed., 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] 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] Payne and S. Singh, "Indoor vs. outdoor scene classification in 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] Cheng, A. Koschan, 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] 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] 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] 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] 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: 16 Nov 2018 | Revised: 25 Dec 2018 | Accepted: 10 Jan 2019** 