INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 4)
Available online at: www.ijariit.com

Convolutional Neural Network (CNN) based indoor scene recognition model

Harsimran Pal Singh
singh.harsimranpal1991@gmail.com
Ramgarhia Institute of Engineering and Technology, Phagwara, Punjab

Dr. Vikrant
singh.harsimranpal1990@gmail.com
Guru Nanak Dev Engineering College, Ludhiana, Punjab

Dr. Balwinder Rai
singh.harsimranpal1992@gmail.com
Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, Punjab

ABSTRACT

The indoor scene recognition is the process of recognizing the scene in focus, which can be either an office or home. The indoor scene recognition has many real-time applications, which includes the automatic camera mode selection, robotic mobility, automatic CCTV coverage and alert setting based upon region visiting frequency, etc. The proposed model is designed for the indoor scene recognition based upon the multiple features, which primarily involves the combination of textural and color-pattern based features. The scene recognition is entirely based upon the pattern & texture recognition, which includes the speeded up robust features (SURF), HOG and scale invariant feature transformation (SIFT) features to differentiate the different indoor scenes. The proposed model has coupled with Classify Neural Network (CNN) classification algorithm. The proposed model has achieved the perfect accuracy of nearly 94%, which has outperformed all other data models with individual features of SIFT, SURF or HOG. This shows the effectiveness of the proposed model based upon multi-feature combination in comparison with the individual feature based applications.

Keywords— *SIFT, SURF, HOG, SVM, Indoor scene recognition, Color analysis, Textural analysis*

1. INTRODUCTION

Scene classification is aimed at labeling an image into semantic categories (room, office, mountain etc). It is an important task to classify, organize and understand thousands of images efficiently. From an application point of view, scene classification is useful in-content based image retrieval. As the accurate classification of an image, as better as it helps in better organization and browsing of the image data. Scene classification is highly valuable in remote navigation also.

Indoor scenes are cluttered with many objects. So classification techniques simply based on color, texture, and intensity are not very effective to classify indoor scenes. Pioneering works used SIFT, SURF etc in combination with supervised learning. But these techniques fail to distinguish many indoor scenes. One way to bridge the semantic gap between image representation and image recognition is to make use of more and more sophisticated models, but good learning and inference is an extremely difficult task for such models. Alternatively the semantic gap between low-level features like color, intensity, texture etc. and high-level category label can be reduced by introducing object-based representation as an intermediate representation. As the performance of scene recognition is heavily dependent on feature representation, this object-based intermediate representation proves to be useful in enhancing classification results. Recently objects-based techniques for indoor scene classification have proven to be showing promising performance over other state-of-art techniques.

In this work, we will review the recent and significant techniques that have been used for indoor scene classification. Besides we will identify the key approaches being used in indoor scene classification. The major contributions made by each significant work and the challenges posed to efficient classification will also be discussed.

2. EXPERIMENTAL SETUP

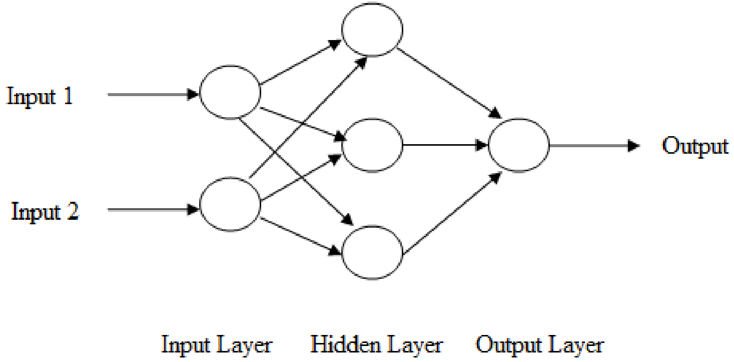
The following is the feature matching and classification algorithm for matching the extracted indoor scene image with the different images of the same scene, which are taken at different times, from different viewpoints, or by different sensors.

Algorithm 1: Feature Extraction Algorithm

```
1. Load the 3-D (colored) test image as test object matrix $\mathbf{T_m}$
2. Convert the test image matrix to grayscale image matrix $\mathbf{G_m}$
3. Apply Genetic Algorithm to enhance and recover the image
```

4. Define the Gaussian Filter **Gf**
5. Apply Gaussian Filter **Gf** on $\mathbf{G_{m}}$ to produce the de-noised $\mathbf{G_{md}}$
6. Calculate $\mathbf{G_{md}}$ into the front-ground estimation feature $\mathbf{F_{EF}}$
7. Define the dilation object of adequate color, shape, and size $\mathbf{S_{E1}}$
8. Dilate the image $\mathbf{G_{md}}$ with respect to $\mathbf{S_{E1}}$to produce the image object $\mathbf{A_{I}}$
9. Perform morphological closing of the image object $\mathbf{A_{I}}$
10. Subtract the $\mathbf{F_{EF}}$ from $\mathbf{G_{m}}$ to produce **F(BG)**
11. Return the **F(BG)**

Convolutional Neural Networks (CNN) are computational structures that are motivated by the natural nervous arrangement of the human brain. CNN comprises a huge number of neurons that work in a dispersed way. There are three types of layers – the input layer, at least one hidden layers and an output layer. There are connections between neurons of back to back layers. Nonetheless, there is no interconnection between neurons of a single layer.



Every one of these neurons, on the whole, find out about the inputs with the goal that they can enhance the yields. An input is presented to the input layer as a vector which additionally disseminates it to the hidden layers. Hidden layers make a decision and weigh the vectors to check how these progressions are corrupting and enhancing the output. Networks that have an assemblage of hidden layers in their architecture are called Deep Networks. There are two methodologies utilized for preparing – Supervised and Unsupervised Learning. In the former approach, both the inputs and required outputs are defined. Subsequently, both the provided sets are compared and error is computed. But in the latter approach, required outputs are not defined. The network needs to settle on its own how to assemble the info information. In this paper, the CNN model has been used to determine the indoor scene type in the target image. The CNN classifier is coupled with multiple features including textural and pattern features. The following section describes the flow of the proposed algorithm:

Algorithm 2: CNN based scene recognition model

Read the source image, and Extract the features from the source indoor scene image. Feature descriptor will be the sub-image and will describe smaller details than the original Target image.

1. Perform pre-processing step to validate the feature descriptor set and arrange all of the feature descriptors in the single feature sets as the training set.
2. Prepare the group data by adding the group IDs corresponding with all of the samples or feature descriptors in the training set.
3. Run CNN training on the feature descriptor training set and return the weight and bias information for all feature descriptors in the training set.
4. Run CNN classifier by submitting the CNN weight and bias data, group data and the testing feature descriptor vector.
5. Return the matching CNN classification information.
6. Evaluate the CNN classification information and return the decision logic.

3. RESULT ANALYSIS

The results have been obtained from the various experiments conducted over the proposed model. The training dataset has been classified into two primary classes, which primarily defines the airport, bakery, bedroom, closet, dining room, garage, kitchen, living room, lobby, and office categories of the indoor scenes. The given image is verified and the decision logic is returned with the detected category type.

3.1 Analysis of 50 test cases

The first experiment has been conducted over the 50 test samples, which has been randomly selected out of the given image set. The randomizer module generates the random index containing the fifty image ids, which are acquired from the given dataset. Such randomly selected samples are further processed and analyzed under the proposed model for the result evaluation.

The type 1 and type 2 errors have been evacuated from the testing of the input test samples. The table obtained from the values of the statistical type 1 and 2 errors have been presented in table 1. Table 1 has been obtained from the fifty testing samples and all of the samples show the equal statistical errors from the first evaluation. Table 1 shows the values obtained for different images.

Table 1: Type 1 and type 2 errors for 10 test cases

	COMBINED	SIFT	HOG	SURF
True Positive	48	8	49	38
False Positive	2	42	1	12
True Negative	0	12	0	0
False Negative	0	2	0	0

Table 1 explains the true positive, true negative, false positive and false negative errors. The statistical errors have been evacuated from the input fifty samples.

3.2 Result of Accuracy, precision, and recall

Table 2 shows the performance measures calculated over the obtained 50 samples. The proposed model and other models based on SIFT, SURF and HOG have been recorded with the variable percentage measured for all of the performance measures.

Table 2: Performance measures for 50 test cases

	COMBINED	SURF	HOG	SIFT	Previous work
Accuracy	94	74	96	14	63%
Precision	96	76	98	16	
Recall	100	100	100	100	
F1-Measure	98	86	99	28	

Table 2 shows the performance measures calculated from the randomly selected fifty test cases. The proposed model has been recorded with the highest values as per the other feature descriptors with support vector machine classification.

4. CONCLUSION

The proposed model has been designed to perform quickly and has been used to describe the color and texture oriented features of the images. The probabilistic multi-class based classification algorithm of classifying neural network (CNN) has been utilized to classify the matching images from the imaging database under the proposed indoor scene recognition system. The proposed model has undergone several experiments, whom performance is tested using the various performance indicators, which includes the statistical type 1 and 2 errors (includes true positive, true negative, false positive and false negative parameters), precision, recall, and accuracy. The proposed model designed with the multi-feature combination is learned to outperform the existing models with singular features.

5. REFERENCES

[1] Xinggang Wang, Baoyuan Wang, Xiang Bai, Wenyu Liu, and Zhuowen Tu. Max-margin multiple-instance dictionary learning. In ICML, 2013.

[2] Yang Wang and Liangliang Cao. Discovering latent clusters from geotagged beach images. In MMM, 2013.

[3] Yang Wang and Greg Mori. Max-margin hidden conditional random fields for human action recognition. In CVPR, 2009.

[4] Yang Wang and Greg Mori. A discriminative latent model of the image region and object tax correspondence. In NIPS, 2010.

[5] Zhengxiang Wang, Shenghua Gao, and Liang-Tien Chia. Learning class-to-image distance via large margin and l1-norm regularization. In ECCV, 2012.

[6] Zhengxiang Wang, Yiqun Hu, and Liang-Tien Chia. Image-to-class distance metric learning for image classification. In ECCV, 2010.

[7] Jianxin Wu and James M. Rehg. CENTRIST: A visual descriptor for scene categorization. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 33(8):1489– 1501, 2011.

[8] Jianxiong Xiao, Krista A. Ehinger, James Hays, Antonio Torralba, and Aude Oliva. SUN database: Exploring a large collection of scene categories. International Journal of Computer Vision (IJCV), pages 1–20, 2014.