

# Recognition and Detection of Objects from Images: An Approach using CSLBP Feature Extractor and CNN

Mini Rani  
Research Scholar  
RIMT University  
Mandi Gobindgarh, Punjab

Ankur Gupta  
Assistant Professor  
RIMT University  
Mandi Gobindgarh, Punjab

## ABSTRACT

Recognition and identification for groups are two aspects of object recognition. Class recognition aims at classifying an object into one of many predefined categories. The detection goal is to distinguish objects from the background. There are growing difficulties in identifying objects including background removal and object detection etc. In this paper, a new approach is proposed for object recognition using the CNN. For feature extraction of the input data CSLBP and LPQ were used and then concatenated. These features were then used to train the convolutional neural network. Fifteen categories from Caltech101 dataset that contains 101 categories of images are considered in this research. The results states that the model achieved higher accuracy of 99% for almost all the 15 categories of the images that were considered. Thus, it can be said that the proposed model shows the efficiency of network in recognizing the objects correctly.

## General Terms

Object Recognition, Object Detection, Digital image processing, Feature Extraction, Image Processing

## Keywords

CNN, Recognition, MATLAB, CSLBP, LPQ

## 1. INTRODUCTION

Object recognition identifies the images or videos that contain the object. Identifying the objects from the images and differentiating them is very easy for the human being. Nevertheless, for the machines this method is quite complicated because they have no pre-defined capability to classify the artifacts before we teach them to do so. Normally, it uses the different algorithms such as machine learning algorithms to allow the machines to recognize and distinguish the different objects. The object recognition is used for various purposes in a number of fields. The technology like driverless cars, for example, uses the object recognition to identify the various objects on the road like traffic signs, persons, etc. Other applications, which use object recognition, are industrial inspection, robotic vision, and bio imaging disease detection [1].

Deep learning and neural networks gains more popularity in the number of systems aiming for object recognition [2]. Recognition of individual's objects poses different problems. Typically, artifacts are to be viewed against cluttered, noisy backgrounds and other objects in different lighting and contrast conditions. Proper feature representation is a crucial step in an object recognition system as it improves performance by separating the object from the background or other artifacts in various lightings and scenarios. The object recognition features were divided into two sparse and dense

representations of groups. Interest-point detectors are used to identify structures like corners and blobs to the target of incomplete feature representations [3]. The object recognition technique normally works by first identifying objects from the image and then recognizing them.

In number of object recognition applications, first the object is detected using the various approaches like frame differentiation, optical flow, background subtraction, and point detectors [4]. Object detection involves object recognition and recognizing patterns within the image or frame of a video sequence. In any tracking process, an object detection mechanism is needed either in-frame or when the object appears first in the video. The most popular object detection technique is to use data in a single image or frame. After that, feature extraction is done for the detected objects so that they can be identified based on the trained data. Objects usually have to be detected under different lighting and contrast conditions against cluttered, noisy backgrounds and other artifacts. In an object recognition system, proper feature representation is a critical step, as it enhances efficiency by distinguishing the object from the context or other objects in various lightings and scenarios [13].

## 1.1 Feature Extraction

### Local Binary patterns to extract texture features

LBP operator takes the center pixel's gray value as a threshold, and if the eight-neighborhood pixel's gray value is greater than the threshold, '1' will be assigned, then '0' will be assigned. Therefore, It will get eight bits and this pixel can be considered as a name. Each pixel on a gray-level image was assigned a label by the LBP operator. The interaction between this pixel and its eight pixel neighbors, affects the mark mapping to a line.

Several methods exist to retrieve the most useful features from (pre-processed) face images in order to perform image recognition. Local Binary Pattern (LBP) process is one of those methods of extraction of functionality. Ojala et al. presented this relative new method in 1996 [12]. With LBP the texture and shape of a digital image can be represented.



Figure 1: An image divided into multiple regions

Such features consist of binary patterns defining the pixel surroundings in the regions. The characteristics obtained from

the regions are concatenated into a single characteristic histogram which forms an image presentation. Images can then be compared between their histograms by calculating the similarity (distance) [12].

In LBP every pixel value  $p$  is compared with the radial distance  $r$  of its  $N$  neighbours. There are  $N$  comparisons for each pixel  $p$ , and the output for each. Pixel  $p$  is  $N$ -bit and the decimal number can be expressed as [14]:

$$LBP_{N,r}(p) = \sum_{i=1}^N s(|n_i| - |p|)2^{i-1} \quad (1)$$

$$s(j) = \begin{cases} 1 & j \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Where  $|p|$  is the pixel value and  $n_i$  are the radial distance  $r$  adjacent pixels. In action, the  $N$  and  $r$  values are 8 and 1, respectively. The histogram of LBP is calculated for the given picture or patch where the histogram length is  $2^N$ . Uniform patterns may reduce the duration of an LBP histogram. The LBP histogram is very long and can't be used efficiently for area descriptors [14]. LBP is invariant to changes in monotonic frequency. Hence, the lighting and contrast variations are robust. It is nevertheless prone to variations in noise and low pixel size [13]. Due to this, current research used CSLB (center symmetric local binary pattern) for the feature extraction purpose. These features are then used to train the neural network to classify and recognize the objects from different categories.

## 1.2 Center Symmetric Local Binary Patterns (CSLBP)

The CSLBP operator is quantized LBP operator representation as defined in Eq. 1. In CSLBP, instead of comparing each neighbor to  $p$ , only center-symmetrical neighbors are compared as shown in Fig. 1.

$$CSLBP_{N,r,T}(p) = \sum_{i=1}^{N/2} s\left(|n_i| - \left|n_i + \frac{N}{2}\right|\right)2^{i-1} \quad (2)$$

$$s(j) = \begin{cases} 1 & j \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The CSLBP histogram is measured for the given image or area, the histogram length is  $2^{N/2}$  which is very short as opposed to LBP. For  $N$ ,  $r$  and  $T$  the suggested values are 8, 1, 0.01, respectively [14]. To calculate the descriptor of CSLBP, the given patch  $P$  is divided into the spatial grid of  $G_x \times G_y$ , and the CSLBP histogram is calculated in each cell. Finally, they combine all histograms to form the descriptor. In CSLBP, there are four comparisons of eight neighbors for each pixel, as shown in Fig. 2.

Based on eight neighboring texture, Equation (2) uses 16 specific patterns, these patterns are represented by decimal numbers from 0-15, and the CSLBP histogram of these 16 patterns for all pixels is calculated for the given patch  $P$ .

$N_6$	$N_7$	$N_8$
$N_5$	$p$	$N_1$
$N_4$	$N_3$	$N_2$

Figure 2: Example of keypoint  $p$  with its  $N = 8$  [14]

The CSLBP as compared to the LBP [15]:

(a) The computational complexity is decreased, and its

calculation sum is half of the LBP;

(b) The histogram dimension of the LBP is  $2P$ , while the CSLBP requires only  $2P/2$ ;

(c) Because of the symmetry property, the CSLBP has a strong anti-noise capacity, such as the noise generated by the camera shakes edge deviation [15].

The CSLBP features are then selected for the environment modeling. The background model updating process is defined in the following subsections, and the background can be modified easily by using CSLBP features when the illumination unexpectedly changes [15].

## 1.3 Convolutional Neural Network

In the proposed model, it uses a convolutionary neural network. A neural convolution network resembles an ordinary neural network and includes neurons and weights for each neuron. While a standard neural network does not scale well to take full images as input, a convolutionary neural network may take large images as input and build their architecture accordingly. A convolutionary neural net uses three primary layers: the Convolutional Layer, the Pooling Layer, and the Fully-Connected Layer [11].

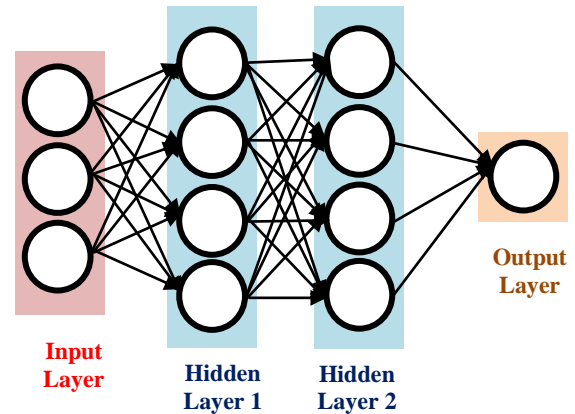


Figure 3: Basic Structure of CNN [11]

Convolutional Neural Networks (CNNs) are broad networks that have recently become an important and effective solution to computer vision issues such as image detection, object recognition, face recognition and human tracking. In addition to this, in order to obtain high accuracy results using CNNs, it is important to effectively train the models with large data sets may not always be attainable. For computer vision applications, data sets with a small number of samples can deteriorate the rate of performance recognition [5]. Features learned from deep Convolutional Neural Networks (CNNs) have been recognized as being more robust and expressive than those hand-made. In various computer vision tasks such as object detection, pattern recognition, and image understanding, they were used successfully.

## 2. RELATED WORK

Due to the popularity of object recognition application, in 2007, Yan, L. et al. [2] proposed a deep learning model for the recognition of iterative objects. The software can define a number of objects, based on user needs. During user interaction it will become ever smarter and identify more classes of objects. It has high performance by studying few instances, including five images of an object type. The level of recognition is getting higher and higher and it can recognize more and more categories of objects after studying

the new samples. Researchers have verified their powerful learning capacity, high recognition rates and wide range of applications by testing it on complex datasets. A flexible technique for object recognition was proposed by *J. Tang and G. Wen* in 2016 [6] through a multifunctional classifier interface. This shows the power of dealing with pose variation, occlusion and lighting changes by means of various characteristics associated with the classifier fusion. The experiment results indicate that in certain difficult situations the proposed system of object recognition can correctly and robustly identify the object. In 2016, *Guo, Q. et al.* [8] proposed the hybrid CNN-HMM to recognize the total number of street view building. This technique used CNN, which has been performing promisingly well in many computer vision problems recently, and HMM, which is a classic sequence model. Experiments show that by replacing GMM with CNN with HMM, the hybrid model will significantly increase the quality of reconnaissance. Another object recognition model using CNN was proposed by *Wu, H. et al.* in 2017 [7]. This approach, combined with optimized segmentation and CNN refinement model, could achieve practical recognition results. More importantly, the effectiveness and robustness of method is demonstrated by adequate experiments based on the database which conclude a large number of complicated images. Yet there are still some drawbacks to proposed method. On the one hand, it is difficult to separate the target object from the background for some complicated images, if there are obvious differences between foreground and background, which could have a negative impact on the results of the recognition. In comparison, more parameters need to be modified for CNN refinement.

In 2018, *Ciocca, G. et al.* [9] proposed an analysis of CNN-based food classification and retrieval functionality. They found various food databases publicly available to check the output of the proposed model and categorize them according to the representativeness of their food-domain which they expressed through the maximum number of images, number of domain categories and number of category examples. Research has also introduced a new food database, Food-475, which is a refining of previously proposed Food-524 database. The features learned from the new database outperform the ones learned from other food datasets and ImageNet's very large image database.

*Hao, W. et al.* (2018) [10] presented an advanced CNN-based image recognition model in conjunction with ideas for regions below. The key contributions were the choice of target area and weight enhancement. Based on the above improvements, effective tests are also used to demonstrate the efficiency and robustness of system using a greater number of images. However, further modifications are still required in the future. For example, if due to ineffective selection the target region is not selected correctly, the error could be increased. Therefore, the weight of the adjustment is based on subjective evaluation, but scope for subjective evaluation can be presented.

### 3. METHODOLOGY

In the proposed research work, CALTECH101 dataset is used [15] that has 101 categories of various objects. Current research uses 15 categories namely faces, faces easy, leopards,

motorbikes, airplanes, bonsai, brain, butterfly, carside, chandelier, grand piano, hawkbill, ketch, laptop, and watch. The selection of categories was based on the large number of images present in each category. Because to train the CNN for particular category, large number of images are required to obtain high accuracy of recognition.

To train the convolution neural network, first the pre-processing is done on the images that includes below steps:

1. Read the image (see fig. 5) Resize the images and convert it to grayscale color by using `rgb2gray` function.
2. Apply WLS filter on images to preserve the edges and smooth the images (see fig. 6). This is done to enhance the edges as mentioned by H. Wu et al. in 2017 [7].
3. After that, CSLBP features of the images were extracted and stored (see fig. 7). In CSLBP, each pixel is known as the center pixel, the same as LBP. Center symmetric pixel differences are measured which are independent of the center pixels.
4. Then, LPQ (Local Phase Quantization) features were extracted (see fig. 8). LPQ is another commonly used histogram-based feature extractor from the local texture descriptor family that performs phase assessment at the pixel location in a local window.
5. After the feature extraction, the GLCM (Gray level co-occurrence matrix) was estimated. The co-occurrence gray level matrix was used for extraction of features from the pattern map of CSLBP and LPQ. Different combinations of distances and angles (0, 45, 90, and 135 degrees) were observed and related accordingly in the present study.
6. Therefore, feature vector of four elements for each angle from each of CS-LBP and LPQ feature image resulted in final feature vector of eight elements after concatenation.

This feature vector of each image is then used to train the CNN containing 12 layers that are image input, two convolutional layers, two batch normalization, two relu layers, two max pool, one fully connected layer, one softmax layer, and one output layer.

### 4. RESULTS AND DISCUSSION

This section describes the results obtained from the proposed design of system for object recognition. MATLAB 2018b software is used for the research work. Figure 5 to figure 8 shows the pre-processing steps of the images that are done to extract the required features from the images. These features are then fed into the convolutional neural network for training. Figure 5 shows the input image of Ketch class. Figure 6 shows the image after converting it to grayscale and applying the WLS filter. Figure 7 and 8 shows the feature extraction result using the CSLBP and LPQ feature extractor.

After the training of network is done, the images are tested using the trained network to obtain the performance parameters of the network like precision, sensitivity, f-score and accuracy (see fig. 9-fig. 12).

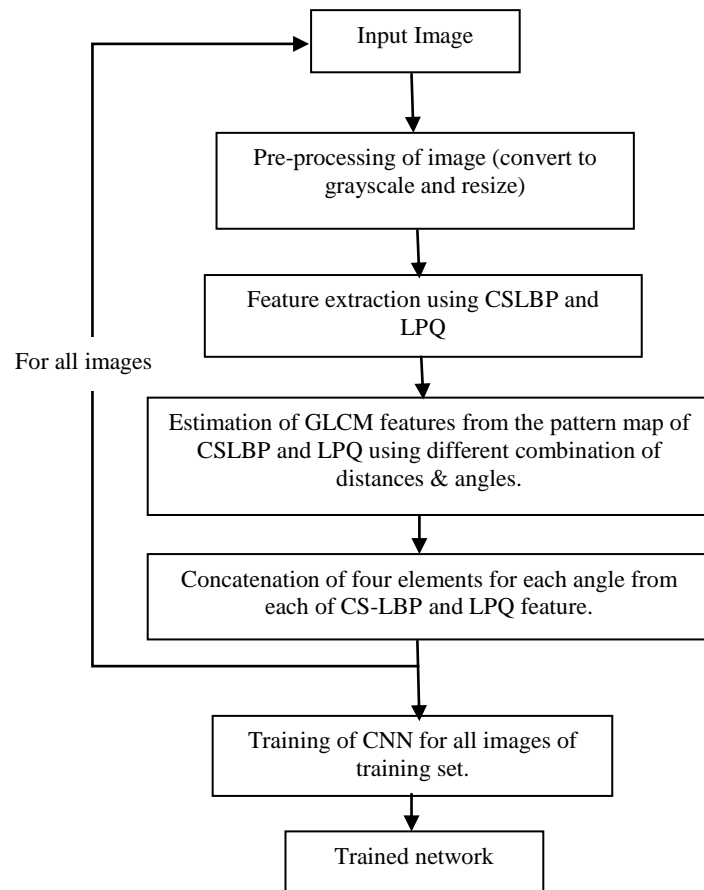


Figure 4: Flowchart of proposed network design

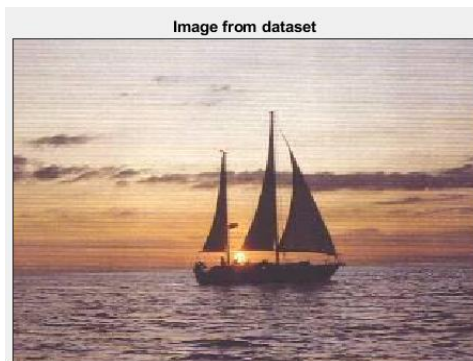


Figure 5: Input image of class Ketch



Figure 6: Image showing results of pre-processing and filtering using WLS filter

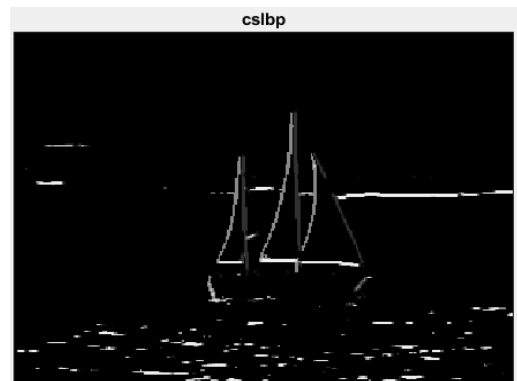


Figure 7: Image after CSLBP feature extraction

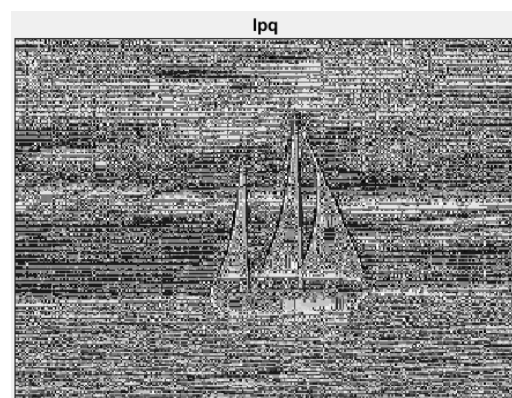


Figure 8: Image after feature extraction using LPQ

Figure 9 shows the results of precision obtained from the proposed system of object recognition. It is the ratio of correctly predicted objects to the total of predicted positive objects. It can be defined as precision = true positives/(true positive + false positive). From the results it can be seen that the system achieved correct prediction value for most of the predictions of objects.

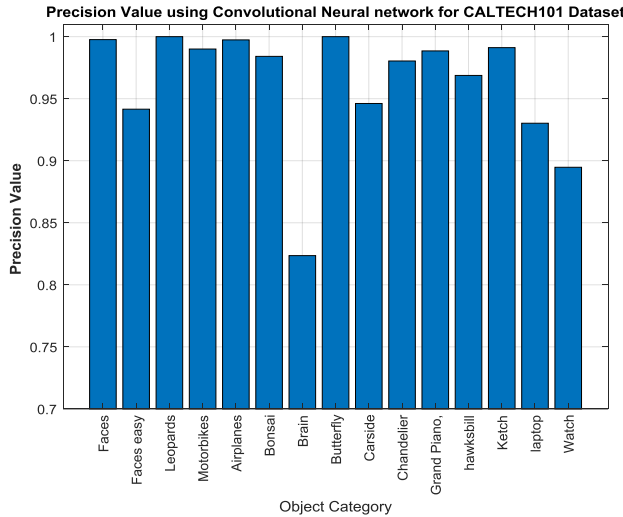


Figure 9: Result of precision for each class using CNN

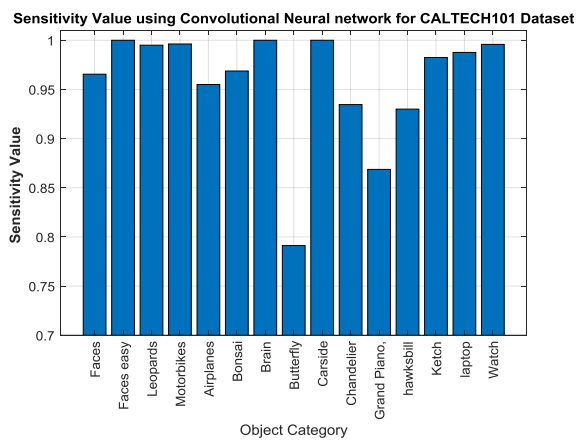


Figure 10: Result of Sensitivity for each class using CNN

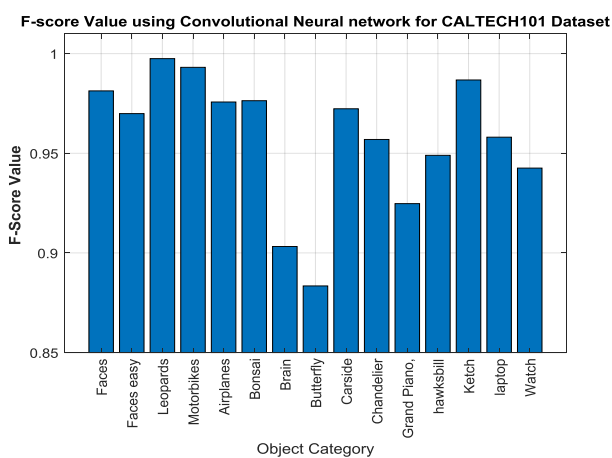


Figure 11: Result of F-score for each class using CNN

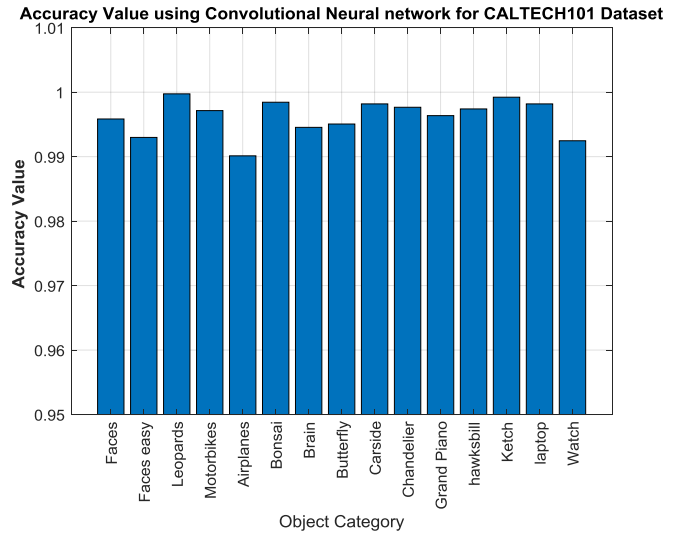


Figure 12: Result of Accuracy for each class using CNN

The proposed model achieved overall 99% accuracy in recognizing the objects as per 15 different classes that were considered in this research work.

## 5. CONCLUSION

This research presented a new approach for detecting and recognizing the object from images. In this paper, a technique based on LBP i.e. CSLBP (center symmetric local binary pattern) is used. Along with this, for extracting the features of the dataset, local phase quantization is used. The extracted features from CSLBP and LPQ are then concatenated and used to train the convolution neural network in MATLAB 2018b. The techniques were evaluated using the caltech101 dataset's fifteen classes. The results shows the efficacy of proposed technique in terms of identifying the objects effectively from the images of different types. Almost all the performance metrics showed the good results. The overall achieved accuracy of the model is 99%. It can be said that the proposed technique effectively recognize the objects with high accuracy.

As this paper tested the approach for limited set of categories, so, in future, complete dataset can be considered to train the convolution neural network. Other than this, more algorithms for feature extraction can be considered like RLBP and MLBP along with new segmentation techniques as well so that the accuracy of the model can be increased and more categories can be predicted.

## 6. REFERENCES

- [1] Rani, M., & Gupta, A. 2019. Recognition and Detection of Multiple Objects from Images: A Review. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 8(12).
- [2] L. Yan, Y. Wang, T. Song, and Z. Yin. 2017. An incremental intelligent object recognition system based on deep learning. Chinese Automation Congress (CAC), pp. 7135–7138.
- [3] R. Bhuvaneshwari and R. Subban. 2018. Novel object detection and recognition system based on points of interest selection and SVM classification. Cognitive Systems Research, vol. 52, pp. 985–994.
- [4] C. M. Sukanya, R. Gokul, and V. Paul. 2016. A survey on object recognition methods. International Journal of

- Science, Engineering and Computer Technology, vol. 6, no. 1, pp. 48–48.
- [5] A. Cheddad, H. Kusetogullari, and H. Grahn. 2017. Object recognition using shape growth pattern. Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, pp. 47–52.
- [6] J. Tang and G. Wen, “Object recognition via classifier interaction with multiple features,” 8th International Conference on Intelligent Human- Machine Systems and Cybernetics (IHMSC), vol. 2, pp. 337–340, 2016.
- [7] H. Wu, R. Bie, J. Guo, X. Meng, and C. Zhang, “CNN refinement based object recognition through optimized segmentation,” *Optik*, vol. 150, pp. 76–82, 2017.
- [8] Q. Guo, F. Wang, J. Lei, D. Tu, and G. Li, “Convolutional feature learning and Hybrid CNN-HMM for scene number recognition,” *Neurocomputing*, vol. 184, pp. 78–90, 2016.
- [9] G. Ciocca, P. Napoletano, and R. Schettini, “CNN-based features for retrieval and classification of food images,” *Computer Vision and Image Understanding*, vol. 176, pp. 70–77, 2018.
- [10] W. Hao, R. Bie, J. Guo, X. Meng, and S. Wang, “Optimized CNN based image recognition through target region selection,” *Optik*, vol. 156, pp. 772–777, 2018.
- [11] Kumar, Sandeep, Aman Balyan, and Manvi Chawla. "Object detection and recognition in images." *Int. J. Eng. Dev. Res.(IJEDR)* 5, no. 4, pp. 1029-1034, 2017.
- [12] Rahim, Md Abdur, Md Shafiu Azam, Nazmul Hossain, and Md Rashedul Islam. "Face recognition using local binary patterns (LBP)." *Global Journal of Computer Science and Technology* (2013).
- [13] Satpathy, Amit, Xudong Jiang, and How-Lung Eng. "LBP-based edge-texture features for object recognition." *IEEE Transactions on Image Processing* 23, no. 5 (2014): 1953-1964.
- [14] Baber, Junaid, Maheen Bakhtyar, Ihsan Ullah, Mariam Rehman, and Muhammad Khalid. "Effective Compression of Center Symmetric Local Binary Pattern." *Mehran University Research Journal of Engineering and Technology* 36, no. 2 (2017): 209-224.
- [15] Zhang, E. H., Y. C. Li, and J. H. Duan. "Moving object detection based on confidence factor and CSLBP features." *The Imaging Science Journal* 64, no. 5 (2016): 253-261.
- [16] Caltech101. (2006). *Computational Vision: [Home]*. [https://www.vision.caltech.edu/Image\\_Datasets/Caltech101/#Description](https://www.vision.caltech.edu/Image_Datasets/Caltech101/#Description)