

Building Detection using Two-Layered Novel Convolutional Neural Networks

Dr. P. Karuppusamy,

Professor, Department of EEE,
Shree Venkateshwara Hi-Tech Engineering College,
Erode, India.
Email id: pkaruppusamyphd@gmail.com

Abstract: In the recent years, there has been a high surge in the use of convolutional neural networks (CNNs) because of the state-of-the art performance in a number of areas like text, audio and video processing. The field of remote sensing applications is however a field that has not fully incorporated the use of CNN. To address this issue, we introduced a novel CNN that can be used to increase the performance of detectors built that use Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). Moreover, in this paper, we have also increased the accuracy of the CNN using two improvements. The first improvement involves feature vector transformation with Euler methodology and combining normalized and raw features. Based on the results observed, we have also performed a comparative study using similar methods and it has been identified that the proposed CNN proves to be an improvement over the others.

Keywords: Building detection, CNN, remote sensing, histogram of oriented gradients, local binary pattern

1. Introduction

Over the past few years, there has been much development in the sensors that are attached to the satellites as they are built to capture images of high resolution. Thus the classification of land cover has become more advanced and has led to many methods for detecting and classifying the different structures and buildings on land. Some of the areas where the modeling and monitoring of urban changes will be effective are i) navigation, for constant update and development of mapping system that can be used for commute, ii) illegal construction for identifying illegal building activity, iii) protection of the city for assessment and analysis of the impact of natural disasters, flood and fire in a situation, iv) citizen

welfare, to plan the development of a city, v) sociology, to monitor the processes that take place in a urban complex environment. With the help of the reconstructed 3D representation as well as satellite images, it is possible to empower the building detection systems that are used in many applications such as urban change detection, urban mapping, city planning and other remote-sensing uses [1]. Simultaneously, urban decline and expansion is also found to be related to natural, economic and social factors apart from climatic changes that enables human prosperity. In recent years, 3D and 2D building detection using remote sensing is done with the help of Convolutional neural network infused machine learning. This methodology was initially introduced in the year 1990 but was later on discarded due to the introduction of better technologies like SVM. However, it was in the later part of the 1990s that the use of CNN was rekindled and it was found in a challenge on large scale visual recognition that CNN performed best when compared with its competitors [2-3].

In this proposed work, we have incorporated a CNN architecture that is used to detect the presence of building even with the data for training is low by accommodating all the models. Moreover, this methodology will also be useful to improve the performance because of its optimization quality and modular structure. The optimization simplicity of this work is seen to positively improve the accuracy by many folds. Here, we have used a combination of raw features and normalized features within the algorithm [4]. The process of normalization is used to make the system more robust and the addition of raw features will be used to further improve the reading of intensity variation. Similarly, we have also used Euler transformation in order to segregate the classes for classification based on the distance function of the features. The proposed methodology provides multiple advantages and solutions because the above said properties and have been experimented with in detail. It has been identified that based on the training performed, though the training process consumes time, it does not have an impact on the deployment performance.

2. Related Works

One of the challenging tasks in remote sensing applications is that of building detection. This is mainly because of the different texture, colour, 3D shape and size variation of different buildings observed [6]. Over the years, a number of building detection methodologies have been proposed and the level of success of these methodologies have been diagnosed [7] and determined with respect to their success level. In order to achieve good segmentation outputs,

building properties based energy functions have been analysed and found to have higher accuracy levels. Similarly, building detection was also made by using line detection to detect the building shapes. Several methodologies that use shadow detection have also been used by using the tall structures of building that could be passed on as candidate buildings [8]. Texture features and corners were also used to determine the presence of building based on the maxima of distribution. HOG features to detect the regions where buildings were present were also introduced. In the early 2000, the feature-based building detector was introduced that used SVM classifier [9] in order to improve the robustness of the system. When using CNN, a number of modifications have been introduced over the years to improve the reliability and performance of the system. As an alternative to hyperbolic tangent and sigmoid, the rectified linear unit was introduced. It was observed that this methodology gives a better environment to train the CNN [10] such that it is quicker to train and also easy to represent sparse information. To prevent overfitting, dropout is another technique that is proposed while training the deep neural networks. Adaptive and maxout dropouts are small modifications of the dropout method that were also experimented later. To avoid overfitting and to further elevate model discriminability, the Network in Network model is proposed that involves a deep neural network with multi-layer perceptrons [11]. This further led to the involvement of variable filter sizes with respect to the varying sizes of patterns captured. Further, to tackle the issue of problem degradation, residual learning was proposed by researchers in [12] that resulted in excellent performance metrics. On the other hand, ETL and NL layers were introduced as the alternative means to improve robustness and accuracy of the CNNs. Without the need for additional training, these methodologies performed simple transformation techniques. The drawback with this methodology is that the inputs are doubled in the next layer because of the methods proposed. Because of the reduction in training weights, ETL and NL layers [13] are said to be more efficient, especially when they are combined. Moreover, in order to establish a connection with an operation, and its accuracy, the proposed work is found to be more effective.

3. Proposed Work

This work takes into account the improvements involved in using LBP and HOG in detecting buildings. Using our novel CNN, we have improved and extended the methodology of detecting buildings. Fig.1 shows a flowchart of the proposed work where the image to be

tested is first split into multiple scales and overlapping windows and are used as input to the novel CNN methodology to detect buildings. As the first step, feature-based algorithms are used to get the accurate values of image blocks. The buildings detected in the first stage will serve as the input to the novel CNN. Further, it is also possible to determine the height of the building with the aid of 3D reconstruction methodology. The proposed methodology will further narrow the result towards accuracy by eliminating false detections. Thus both LBP and HOG features can be used in a positive note to generate descriptive features using the CNN. This will not only speed up the process of detection but will also suppress new false alarms.

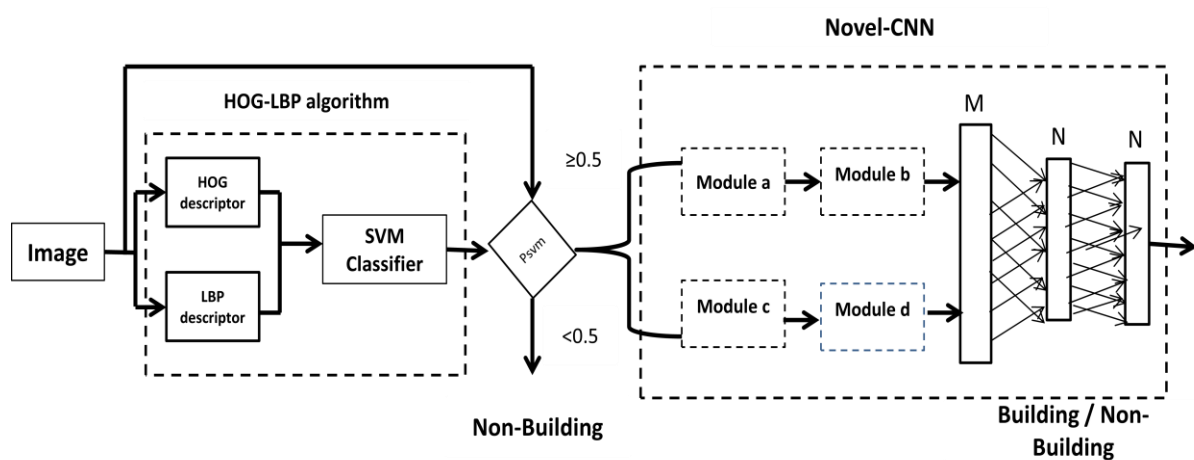


Fig.1. Architecture of CNN for Detecting Buildings

In the proposed methodology, there are only two units or modules that are arranged in parallel or sequential pattern. Here every module denotes a CNN that combines pooling layers, activation functions and convolution layers. This architecture holds 2 linear layers that are completely connected. The first layer is used to decrease the features count and will also optimize the parameters involved. The second layer completes linear mapping without affecting the vector dimensions of the features. The final layer is used for classifying the feature vectors based on their characteristics. One methodology used for optimizing the novel CNN is to add modules to the architecture of the CNN, one module at a time thereby optimizing the hyper parameters of similar modules in an independent manner. The process for optimization is performed till the destined width or depth occurs or there is an increase in the performance of the CNN. There are two additional layers involved in this methodology

namely Euler Transform and Normalisation layer which ensure that the robustness and performance of the CNN improves at a large level.

3.1 Euler Transform Layer

The initial use of l_2 norm was replaced by a new distance function in order to determine the computational displacement that occurs in frames of a video. The l_2 norm methodology was known to be highly influenced by noise attributes. Hence in order to overcome this discrepancy, the new distance function was introduced and was considered to be more robust to noise and other distractions. Consider y_i and y_j to be two feature vectors that have values within $[0,1]$ range such that the dissimilarity can be measured as follows:

$$d(y_i, y_j) = \sum_{p=1}^p \sum_{l=1}^L b_p (1 - \cos(a_p \pi (y_j(l) - y_i(l))))$$

where L represents the length, P denotes the limit of the sinusoidal terms in the approximated value. If only one sinusoidal value is considered, then the equation can be further condensed to represent:

$$d(y_i, y_j) = \sum_{l=1}^L (1 - \cos(a_p \pi (y_j(l) - y_i(l))))$$

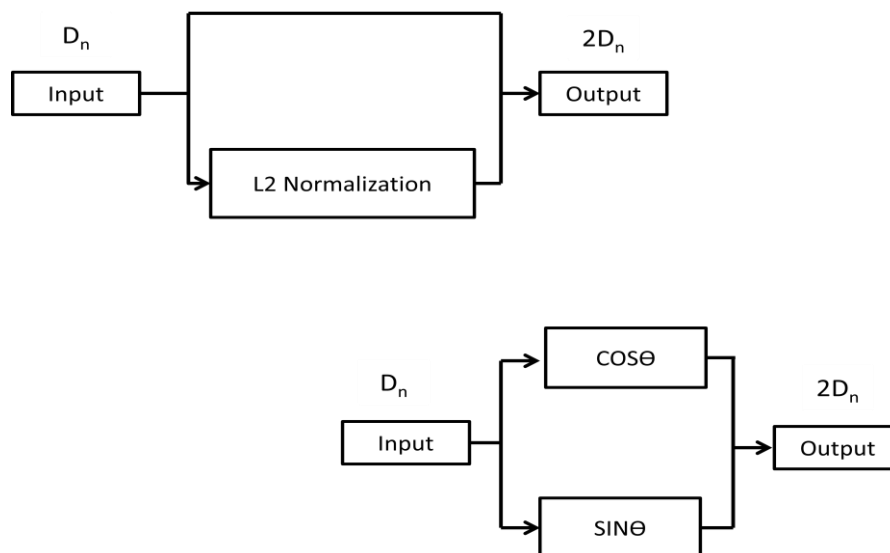


Fig.2. Normalization layer (Top) and Euler Layer (Bottom)

The redescending m-estimator using Andrew's M-Estimate is used for outlier rejection methodology. This is because, the function calculated in eq (2) does not have an increasing

function such as that of the l_2 norm. However, for values that have large differences, it will redescend smoothly towards zero. This will give the function easier mechanism to eliminate the difference leading to reduction in outliers and noise.

$$\psi(r) = \begin{cases} \sin(\pi r), & \text{if } -1 \leq r \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

3.2 Normalization Layer

A very common pre-processing methodology that is applied on the input data is known as normalization. This step is generally used to improve the performance of a system, when it is based on the gradient optimization algorithm. Because of the scaled features and their equal weighing, this is required. Else, there is possibility for the two values of input to saturate the neurons such that they reach a state of inactivity and will further affect the following layers in a neural network structure. But, since the CNN output is designed to map the inputs normalized with non-linear mapping, normalization effect tends to get lost or fades away. Hence it is important to perform normalization operation before the classification of the neurons. However, since the normalized data already provide raw features, this type of mapping will result in the introduction of new normalization with respect to feature space. According to this approach, we introduce the use of feature vector classification such that the input feature vector is taken and fed as twice the size of the initial vector. The initial vector is taken to be the first half of the new feature vector and the next section can be normalized with a l_2 norm copy.

4. Results and Discussion

To determine the overall functioning of the proposed ETL and NL layers in the novel CNN is considered such that there are two architectures examined- with and without the layers. An examination is done on both the architectures and the output thus obtained is analyzed. The size of the features map has increased filters from 16 to 64 while the size of feature map is reduced from D to $D/4$. Fig.1. represents the optimal performance of the proposed CNN along with CNN+NL+ETL layers. The examination is done for the values of α within the range of 0 and 1.9.

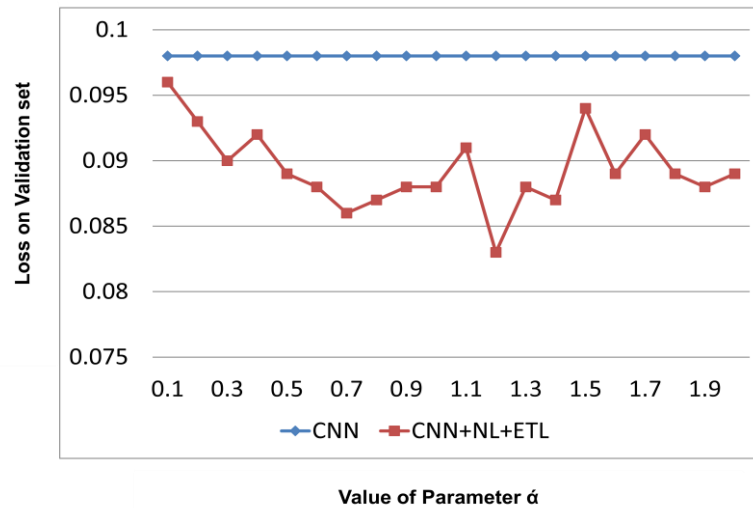


Fig.3. World View dataset- Performance Validation with respect to average loss

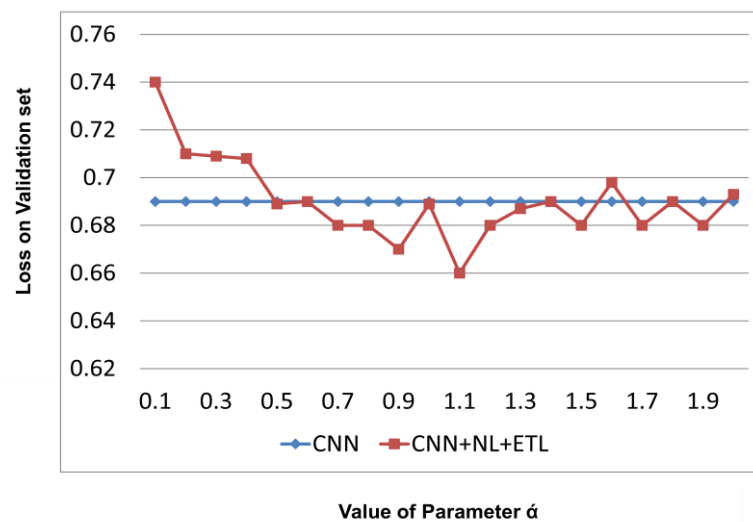


Fig.4. CIFAR-10 dataset- Performance Validation with respect to average loss

Fig.4 and Fig.5 further validate the improvement in performance using MNIST and CIFAR datasets. The size of MINST is 28x28 pixels and that of CIFAR-10 is 32x32 pixels.

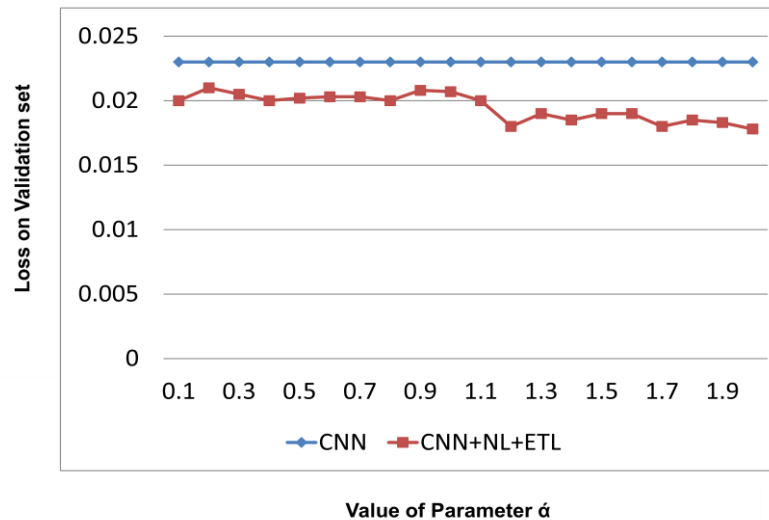


Fig.5. MNIST dataset- Performance Validation with respect to average loss

5. Conclusion

In this proposed work, we have introduced a novel CNN architecture that operates in line with the feature based classifier, to enhance the performance of detecting building using 3D and 2D data. To further improve the robustness and discrimination ability of the CNN architecture, two novel layers have been introduced in the proposed methodology. The novel layers and their effects have been analyzed to determine the beneficial impact on comparison with other novel methodologies that are used on remote sensing images to detect buildings. As future scope for this paper, the proposed methodology can be used to video and image classification which will improve the robustness and accuracy of the deep networks. Another area of high interest that falls in line with this work is the implementation of Euler Transform in CNN which can be examined in deep networks that involve multiple ETL layers.

References

- [1] Zhang, Q., Wang, Y., Liu, Q., Liu, X., & Wang, W. (2016, July). CNN based suburban building detection using monocular high resolution Google Earth images. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 661-664). IEEE.
- [2] Sun, L., Tang, Y., & Zhang, L. (2017). Rural building detection in high-resolution imagery based on a two-stage CNN model. *IEEE Geoscience and Remote Sensing Letters*, 14(11), 1998-2002.

- [3] Alidoost, F., & Arefi, H. (2018). A CNN-based approach for automatic building detection and recognition of roof types using a single aerial image. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 86(5), 235-248.
- [4] Shirley, D. R. A., Janeera, D. A., Padmini, J. J., Banu, S. M. A., & Abirami, T. (2018). MODELLING AND ANALYSIS OF MODIFIED BAUGH-WOOLEY MULTIPLIER USING GATE DIFFUSION INPUT AND IMPROVED SHANNON ADDER. *International Journal of Pure and Applied Mathematics*, 118(22), 773-777.
- [5] Chen, C., Gong, W., Hu, Y., Chen, Y., & Ding, Y. (2017). Learning oriented region-based convolutional neural networks for building detection in satellite remote sensing images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 461.
- [6] Sungheetha, A., & Sharma, R. (2020). A Novel CapsNet based Image Reconstruction and Regression Analysis. *Journal of Innovative Image Processing (JIIP)*, 2(03), 156-164.
- [7] Vakalopoulou, M., Karantzalos, K., Komodakis, N., & Paragios, N. (2015, July). Building detection in very high resolution multispectral data with deep learning features. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1873-1876). IEEE.
- [8] Chen, C., Gong, W., Chen, Y., & Li, W. (2019). Learning a two-stage CNN model for multi-sized building detection in remote sensing images. *Remote Sensing Letters*, 10(2), 103-110.
- [9] Hamaguchi, R., & Hikosaka, S. (2018). Building detection from satellite imagery using ensemble of size-specific detectors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 187-191).
- [10] Han, Q., Yin, Q., Zheng, X., & Chen, Z. (2021). Remote sensing image building detection method based on Mask R-CNN. *Complex & Intelligent Systems*, 1-9.
- [11] Zhou, Z., & Gong, J. (2018). Automated residential building detection from airborne LiDAR data with deep neural networks. *Advanced Engineering Informatics*, 36, 229-241.
- [12] Smys, S., Basar, A., & Wang, H. (2020). CNN based Flood Management System with IoT Sensors and Cloud Data. *Journal of Artificial Intelligence*, 2(04), 194-200.
- [13] Vijayakumar, T. (2019). Neural network analysis for tumor investigation and cancer prediction. *Journal of Electronics*, 1(02), 89-98.