ISSN: 0971-1260
Vol-22-Issue-16-August-2019

A Review on Google Landmark Recognition Challenge Using CNN

Pawan Kumar Sharma
Student
School of Computer Science and Engineering,
Lovely Professional University, Phagwara, Punjab
kpawansharma2@gmail.com

Savleen Kaur Chowdhary
Assistant Professor
School of Computer Science and Engineering,
Lovely Professional University, Phagwara, Punjab
iridiumsavleen@gmail.com

Abstract

It may happen we see our vacation photos and ask ourselves: What was the name of themonument I visited in China? Who created this temple I saw in India? Landmark recognition can help us to recognize the locations easily. A landmark is a natural or artificial feature which is recognizable and can be used for navigation purpose, it is a feature that stands out from its surrounding environment and is often visible from long distances. This technique can predict landmark using landmark labels directly from pixels of images, which will help person to better understand and organize thecollection of photos.Today, hindrancein landmark recognition research is the absence of enormous commented on datasets.

Keywords: Feature, Classification, Neural Network, CNN, Deep Learning, ReLU, Max-Pooling

1. Introduction

A landmark is a physical object, created by man or by nature, with a high recognition value. Examples of landmarks are buildings, monuments, statues, parks, mountains and other structures and places. Due to their recognition value, landmarks often serve as geographical points for navigation and localization. Big number of photos of landmarks contain only one landmark image, which in the most of the cases takes in 80% of the photo area, and very few cases are there in which it takes only a small part of the photo when it is taken from apart. A marginal part of photos shows two or more landmarks. Landmark recognition is a technology to classify the images. Many a times, it happens that we might forget name of monument, we once visited for the purpose of organization of photos in proper manner or to predict which place it is in front of which we are standing we can use landmark recognition software. It is a software system which can facilitate to predict landmark labels from image pixels directly.

An additional advantage of landmark recognition will help people to better understand and organize their collected photos. Too often people visit some place they take tons of photos and only realize a month later that they have no idea what they are standing in front of. By being able to train models that can recognize landmarks, we are able to recall exactly where the place is through the raw image pixels. Landmark recognition model is a model which can recognize various landmarks in an image, when we pass an image to this model or GUI, we get the landmark that was recognized in it, along with each landmark's geographic coordinates and the region of the image the landmark is found. In future, this can also be applied to travel guide recommendation applications as well as for geo-location visualization.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

In this project, the ideas are to build a model that recognize landmark in a dataset of challenging test images. The main challenge in the development of this application is to obtain a large annotated dataset. To add annotations manually for all images are extremely difficult. So, to build the model, we utilize the annotated data provided by Kaggle. This data has been constructed by clustering visually similar photos and matching them based on local features. One of the great obstacles to landmark recognition system is the absence of enormous annotated dataset. Google landmark recognition presents the biggest worldwide dataset to date and challenges to build models which recognize the landmarks from the test images. A technology where one can accurately predict landmark labels directly from image pixels and can broadly benefit applications in various areas such as photo management, maps, aviation, and satellite images.

We collected dataset on which we are going to work. The dataset on which we are working is derived from the Google Landmark Recognition Challenge that took place on Kaggle. The Project is to build models that classify the images provided in such a way that it matches the correct landmark with each unique image. After retrieving the data, we explored the data for better understanding. We performed data pre-processing and data cleaning, correct anomalies that may pose some issue later. After data extraction and data pre-processing step, we performed data visualization so that we can have more insight about what is happening under the hood and how data is distributed.

We divided the data in two parts one as training data and other as testing data. Through training data, we will be training or teaching our system what it must do and after it is trained, we will test the system by using testing data that it is performing its task correctly or not. Beginning with a training sequence of images, we identify that there are group of images corresponding to distinctive landmarks. Each group is characterized by a set of feature distributions. At run-time, the observed images are compared with the sets of models in order to recognize the landmarks in the input stream. The number of reasons is observed because of which recognizing landmarks in sequence of images is a difficult and challenging problem. To start with, the appearance of any of the given landmark varies essentially from one observation to the next. In addition to variation caused because to different aspects like illumination change, external clutter, and changing geometry of the imaging devices are some other factors affecting the variability of the recorded landmarks. In a training stage, the framework is provided with a set of images in sequence. The goal of the training is to organize the images into groups based on similar type of feature distributions. The size of the groups may be defined by the user, or by the system itself. In the further case, taking the global distribution of the images in account the system tries to find out the most relevant groups. In the second step, the system is designated with new set of images, which it tries to classify as one of the learned groups, or if it is in the category of unrecognized images. After that we downloaded the images from test and train datasets, after that we clean and remove the corrupted images. After the data is ready to use, we will be performing data modelling we will create a model to make the goal of landmark recognition work well.

When the model is fully deployed and created then we check by passing the testing data through it and check whether it is working properly or not. All of this is done using Python programming language and concepts of machine learnings.Python provides bunch of libraries and packages

2. Literature Review

Neural networks are set of algorithms which is designed based on working of human brain. It is mainly used for image classification [1]. Thesensory data are interpreted with the help of machine labelling or clustering. The machine work on pixels value which are numerical contained in vectors.

Deep Neural Networks are best at image classification, detection and segmentation. But due to its high computational complexity even powerful hardware may find it difficult to provide desired outcome. CNN gets more efficient with max-pooling layers [2].

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

CNNsbasically mimics the nature of complex cells in visual cortex of mammals. It consists of three layers with various combinations which are convolutional, max-pooling and fully connected layers. It may consist of one or more convolutional, pooling and fully connected layers.[3]

In CNNs the neurons which belongs to same layer share same weights. Weight sharing results in more efficient and better trained models [3].

The input plane receives images which are normalized. Neurons can extract features such as shapes, edges etc. These combined features stored in features map [4]. A max-pooling layer is usually followed by max-pooling layer [3].

ReLU is basically an activation function which is used to remove negative value, it is linear for all positive value and zero for all negative values. It is used in neural networks. It does not have the problem of vanishing gradient which other activation functions (sigmoid, tanh) generally have. It can improve model efficiency and give better output. ReLU is considered as standard activation function for neural networks [5].

In CNN pooling is used for down sampling, after computing all convolutions. It takes input feature maps and after computation on sub-area of input it gives low-dimension feature maps as output. Image stack is reduced to smaller size bypassing throughpooling layer. Pooling optimizes for further computing. [2]

Fully Connected Layer is an essential part of CNNs. The fully connected layer also called classification layer is equivalent to a multilayer Perceptron (MLP). MLP is a class of feedforward ANN (Artificial neural network). In neural networks neurons connects inputs from one layer to each activation unit of next layer. After repetition of pairs of convolutional and pooling layers, the features extracted are integrated through fully connected layers [5].

The Fully connected layer is the last layer in the CNNs, and it is used for the final decision. Usually function such as soft-max is used in the classification layer.

3. Algorithms and Techniques

In Deep Learning Algorithm, CNN(Convolutional Neural Network) is one of the main categories of neural network in which image classification, image recognition, object detection, etc are done.

In this project, using framework of TensorFlow with Keras library CNN model is built.A schematic structure of a CNN is shown in Fig. 1 with an example of image as input and then output the predictions.

Convolutional Neural Network have following layers:

- Convolution
- ReLU (Rectified Linear Unit)
- Pooling
- Fully Connected

A computer understands an image using number at each pixel. In the convolutional layer, feature maps are created for the input image for which multiple kernels are used. Each kernel has feature to produce a convolved image. A pooling layer down samplesconvolved images and reduces the dimensionality of the feature maps. This down samplinghelps in overcoming overfitting issues and reduce computational cost.We stack up 3 layers- Convolution, ReLUandPooling. These 3 layers are then repeated. As the convolutional network becomes deeper, the feature map becomes smaller, more

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

abstract and loses spatial information. These reduced feature maps are then fed into the fully connected layers where the high-level reasoning is performed. The neurons in the fully connected layer have full connections to all activation in the previous layer. At the end the fully connected layer converts 2D feature map to 1D feature vector which later classified. In summary, the convolutional, ReLU, pooling layers act as feature extractors from the input image and the fully connected layers act as an image classifier.

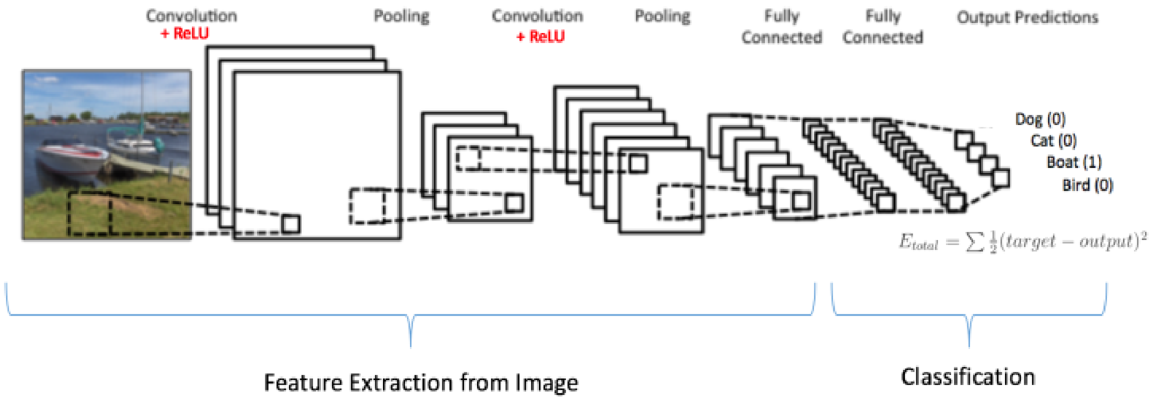

Fig. 1: A schematic of CNN model

3. Conclusion

The input image with shape of (192, 256) is supplied to the image extraction network, which consists of five sequential combination of convolutional and max pooling layers. The extracted features are supplied to three fully connected layers with dropout layer to identify 200 landmark ids. The accuracy of final CNN with data augmentation was 66.6%.

4. Future Scope

We can enhance or improve this project by adding some more features so that it is more useful and helpful. It can be used in satellite imaging, aviation, GPS and in extraction of nearby information such as cafe, restaurants, libraries, etc.

5. References

[1] Wang, C. and Xi, Y., Convolutional Neural Network for Image Classification. *Johns Hopkins University Baltimore, MD, 21218.*

[2] Giusti, A., Cireşan, D.C., Masci, J., Gambardella, L.M. and Schmidhuber, J., 2013, September. Fast image scanning with deep max-pooling convolutional neural networks. In *2013 IEEE International Conference on Image Processing* (pp. 4034-4038). IEEE.

[3] Hu, W., Huang, Y., Wei, L., Zhang, F. and Li, H., 2015. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, *2015*.

[4] LeCun, Y. and Bengio, Y., 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), p.1995.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

[5] Ide, H. and Kurita, T., 2017, May. Improvement of learning for CNN with ReLU activation by sparse regularization. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 2684-2691). IEEE.

[6] https://www.kaggle.com/c/landmark-recognition-2019,8:32 PM, 10/12/2019

[7] https://medium.com/@abhinaya08/google-landmark-recognition-274aab3c71ae,8:32PM,10/12/2019

[8] https://edureka.co/blog/convolutional-neural-network/ ,8:32PM, 10/12/2019