Building Infrastructure Classification and Detection

By

Faisal Ahmed

Roll: 1607048

&

Md. Mahmudul Islam

Roll: 1607052

Supervisor

Dr. Sk. Md. Masudul Ahsan

Professor,

Department of Computer Science and Engineering,

Khulna University of Engineering & Technology,

Khulna, Bangladesh

Contents

1. Introduction

A building is a structure that has a roof and walls and stands permanently in one place. There are different types of building infrastructure such as houses, mosques, temples, shopping malls, etc. There are different sizes of building such as small, medium, and large. People live, work, and do activities in those places.

Classification of different types of building is essential for a wide variety of computer vision tasks such as augmented reality, graphics system, scene understanding, etc. this area has received great attention for several applications. Image processing and machine learning techniques can be used for building infrastructure classification and detection.

Figure 1: input image (source)

Figure 2: output building house

2. Related works

In computer vision, image classification is one of the fundamental problems. Another important problem is to detect the particular region from the image. Different types of algorithms are proposed to classify the image according to its categories and then detect the regions. Sometimes, a rectangle box is drawn over the image to detect the region.

Building infrastructure classification and detection is an active area of research. Mainly the task is divided into two parts. At first classification and then detection. Some researchers only classify the building and some researchers are only detecting the building.

1. Bappy et al. [1] proposed a method to classify the real state image. They proposed a high confidence image classification framework whose inputs are images and outputs are labeled. The core of the classification is long short-term memory (LSTM) and a fully connected neural network along with a substantial preprocessing using ‘contrast limited adaptive histogram equalization’ for image enhancement. They also provided Real Estate Image (REI) datasets for evaluating the algorithm.

2. Pouraseed et al. [2] proposed a vision-based real estate price estimation. They develop a method for estimating the luxury level of real estate photos. They evaluate the impact of the visual characteristics of a house on its market value. Using deep convolutional neural networks is to classify the luxury level and set automated values based on categories.

3. Vakolopoulou et all. [5] proposed building detection in very high-resolution multispectral data with deep learning features. Automated man-made object detection and building extraction from single satellite images.

4. Silva et al. [4] Proposed a deep learning-based image classification technique to detect concrete cracks in a building. This is a binary classification of crack detection. They use a transfer learning approach to make a VGG -16 architecture to develop the model.

3. Objectives

I. To learn how to classify different types of the building structure
II. To know how to extract a region from the image
III. To collect different types of building images to make an image dataset
IV. To develop some image processing techniques to enhance the image quality.
V. To develop our own image classification system to classify the building images.

4. Dataset collection

We introduce a new dataset for building infrastructure. The images collected from our smartphone camera and some images are collected from the internet. The dataset is split into three parts as a train set, validation set, and test set. The train set is used for training the model. During the training time, a validation set is also added for proper training. The test set is called the unseen data that the model has never seen before. The model is evaluated on the test set. Firstly, there are only four categories in our dataset. In the future, we will increase the categories.

The four categories are:

I. Building house
II. Mosque
III. Temple
IV. Tin shed house

Figure 3: building house (source)

Figure 4: mosque (source)

Figure 5: Temple (source)

Figure 6: tin shed house

Firstly, we collect 100 images for each class. Then we initially start our proposed work. We are increasing our dataset. Our target is at least 500 images for each class. Using data augmentation, the dataset can be increased to train the model.

5. Methodology

Our goal is to build a model architecture that can classify and detect the building infrastructure. The task is divided into many parts.

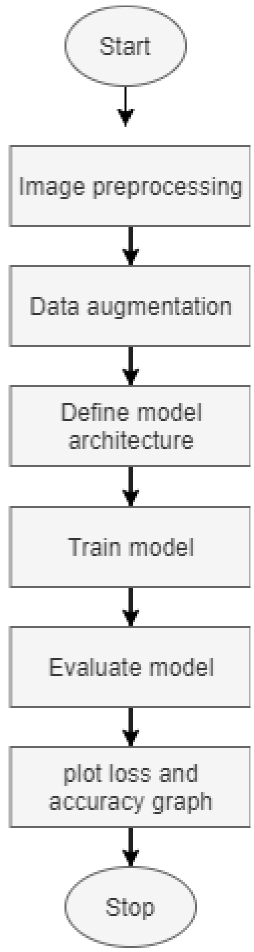


Figure 7: Flow chart of the methodology

5.1 Image preprocessing

As we collect our image data from different sources. We have to do some preprocessing tasks to remove noise from images, crop the image from the unwanted background, resize the image into a particular shape. Though the images are collected from various sources, the quality of images is varying. Some images are very much lower resolution images and some are higher resolution images. Some images are blurry and it is tough to preprocess those blurred images. Some images have problems in illumination and shadows. All this problem is solved in the image preprocessing step.

All the images are resized into the same shape to train the model properly. Though the collection of all images is different, we set the shape of height and weight of an image to be 500x500 pixels. In some cases, rotation is required to view the building like a human sees the building at the front side.

5.2 Data augmentation

If the datasets have not contained enough image data, data augmentation is very much useful to increase the image data. It is a technique to create new data from the existing training data. It is mainly applied to the training dataset and not for the validation and test set. Data augmentation is different from data preparation such as image resizing, scaling, cropping. There are many types of data augmentation techniques

I. Horizontal and vertical shift
II. Horizontal and vertical flip
III. Random rotation
IV. Random brightness
V. Random zoom

As our dataset is small, we do these augmentation techniques on the dataset and save the augmented images into a new directory. Approximately 500 images are added to the training set for each category.

5.3 Define model architecture

In machine learning, a model is a file that has been trained to recognize a particular pattern. The model is trained over a training set. The model is evaluating on testing set data that the model has never seen before. The model makes predictions over testing data.

A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. A convolutional neural network (CNN) is a neural network that has one or many convolutional layers. Convolutional layers are the main important parameter for image classification, segmentation, and also for many other image processing tasks. a convolutional is a sliding window over an image with a fixed kernel. There are other layers in CNN called the pooling layer, the fully connected layer. There are many hidden layers that perform multiplication or dot product between the other layers. The multiplication process depends on the activation functions.

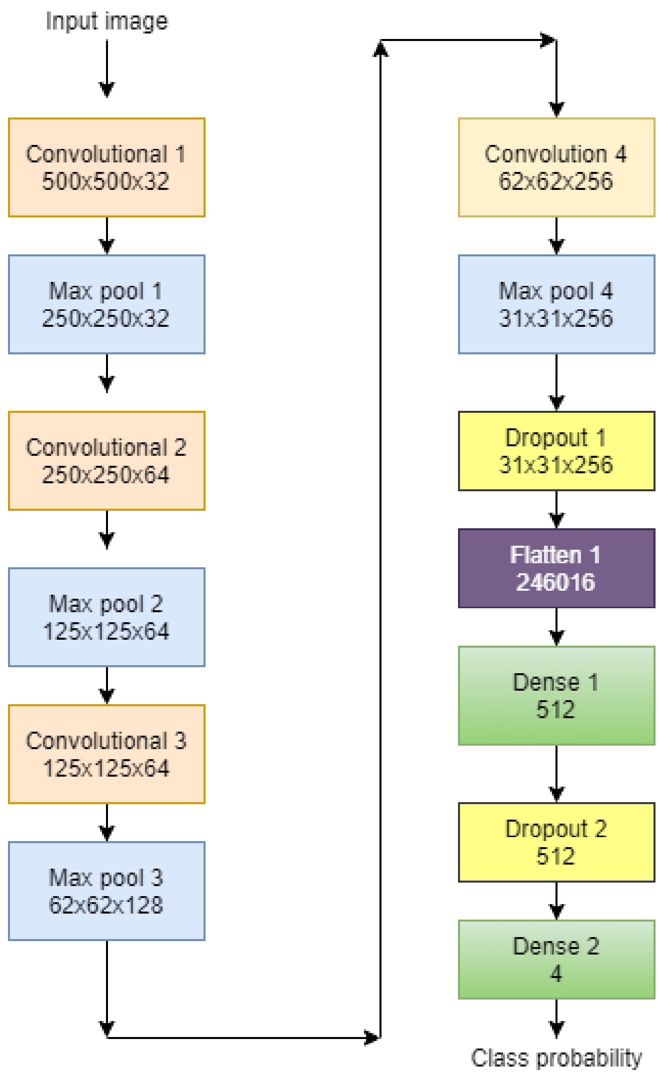


Figure 8: proposed CNN architecture

This is our proposed CNN architecture. There are 4 convolutional layers, 4 max-pooling layers, 2 dropout layers, 2 dense layers, the activation function is ReLU instead of the last layer. The softmax activation function is used in the last layer. Categorical cross-entropy is the loss function and the optimizer is RMSProp.

5.4 Train model

Machine learning models are trained on a particular dataset to predict new things. As the dataset is split into train set, validation set, and test set. The model is trained on a training set. For the training purpose, we have to define batch size, epochs, callbacks function, steps per epoch, and many more hyperparameters.

One epoch is the total dataset that is passed forward and backward through the neural network once. Since the dataset is too large, then one epoch is divided into small batches. Initially, we set epoch 50 for training the model. Batch size is a term that refers to the number of training examples in one iteration. We set step per epoch 3.

5.5 Evaluate model

Since the model is fit on the training set, the model is evaluated on a testing set. The accuracy of the model is found 87.27 % and the loss is 0.36. if the hyperparameters are changed the accuracy and loss values are also changed. The model can make predictions on new data.

5.6 Plot accuracy and loss graphs

Matplotlib is very much useful for data visualization. Different types of graphs are drawn using matplotlib. When training the model, every epoch the accuracy and loss values are changed. Rate of changes of accuracy and loss function values saves into a variable to plot the graphs.

There are four graphs found from the history variable from the fit model.

I. Accuracy
II. Loss
III. Validation accuracy
IV. Validation loss

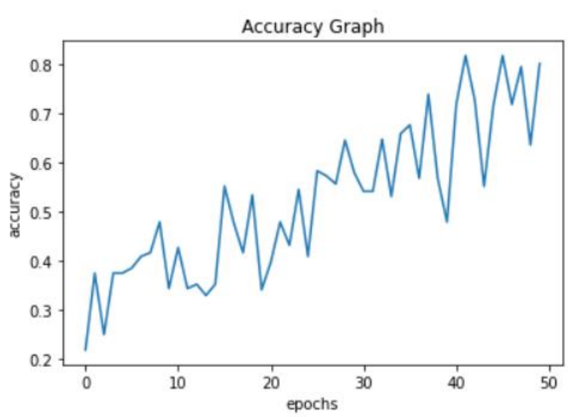


Figure 9: Accuracy graph

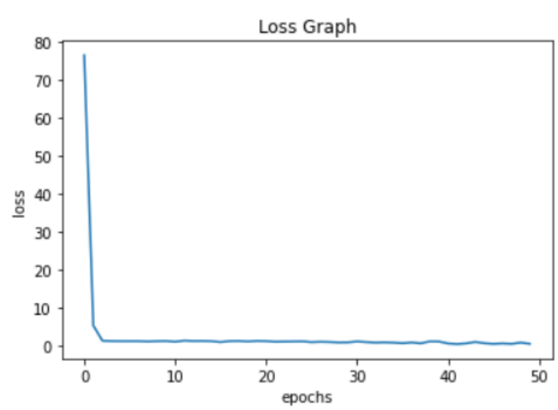
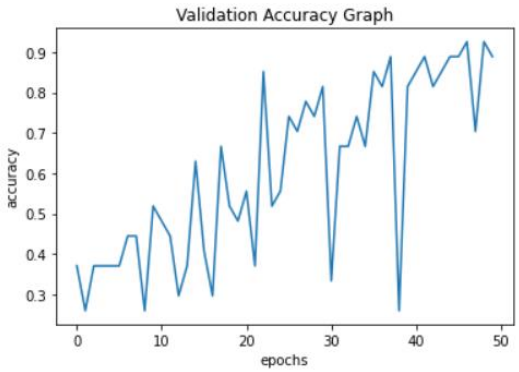


Figure 10: Loss graph



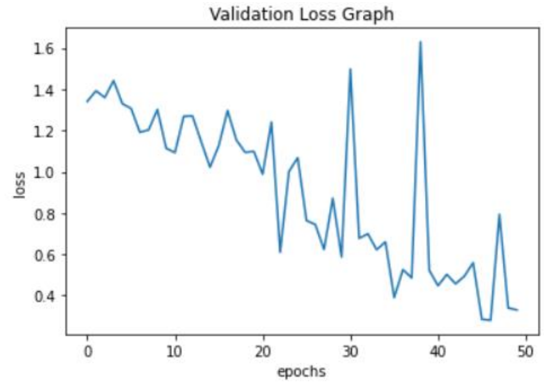


Figure 11: Validation accuracy graph

Figure 12: Validation loss graph

6. Experiment with different architecture

We change our architecture at different times, then we get different types of accurate results. At first, we choose only one convolutional layer for classification and then gradually increase the convolutional layers in our model. We found that if we increase our convolutional layers then the accuracy of the model is also increased at a certain rate. We choose four convolutional layer.

7. Work progress

As discussed previously, we want to develop an architecture for image classification. We proposed a model architecture to do the task. We build the maximum portion of our model. At first, we developed our model for image classification then we developed our model to localize the building into the image. We will find region of interest from the image.

8. Conclusion

Developing an image classification architecture for building infrastructure is an active area of research. Several researchers proposed different types of algorithms to solve the problem. In recent years, most researchers do image classification using convolutional neural networks. Recently, transfer learning is one of the most popular tools for doing this kind of thing. Transfer learning achieves a state-of-the-art accuracy for this type of thing. VGG-16, ResNet-50, Inception-v3 are the popular transfer learning architecture. We developed our own architecture to complete our proposed work.

9.References

1. Jawadul H. Bappy, Joseph R. Barr, Narayanan Srinivasan, Amit K. Roy-Chowdhury. Real Estate Image Classification, IEEE Winter Conference on Application of Computer Vision (WACA), 2017
2. Omid Pouraseed, Tomas Matera, Serge Belongie. Vision-based Real Estate Price Estimation, arXiv: 1707.05489v3, 2018
3. Neha Sharma, Vibhor Jain, Anju Mishra. An Analysis of Convolutional Neural Networks For Image Classification, International Conference on Computational Intelligence and Data Science, 2018
4. Wilson Ricardo Leal da Silva, Diogo Schwerz de Lucena. Concrete Cracks Detection Based on Deep Learning Image Classification, MDPI, 2018
5. M. Vakalopulo, K.Karantzalos, N. Komodakis, N.Paragios. Building Detection in Very High- Resolution Multispectral Data with Deep learning Features, IGARSS,2015