Machine and Deep Learning on PKlot Dataset

Furkan Atak*
†01.09.2022
furkan.atak@boun.edu.tr

***Abstract*—In today's world, one of the toughest challenge we face everyday is to find parking space in crowded cities. In this regard, surveillance on parking lots can mitigate the time spent on finding available lots. So, this paper aims to propose and explain a solution by using machine and deep learning methods applied on the well known PKLot dataset collected using live CCTV cameras. Namely, k-nearest neighbors(k-NN), support vector machine(SVM) and random forest classifiers used for traditional machine learning. Convolutional neural network(CNN) is build for deep learning. The code written in python is provided at the end of the paper in appendix section.**

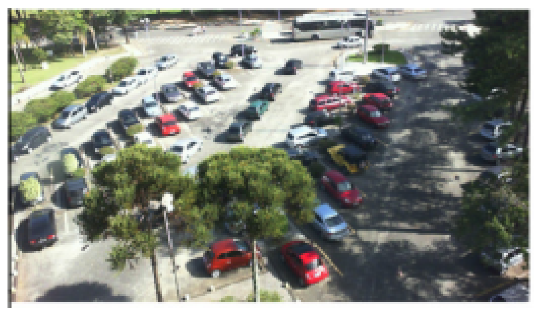
***Index Terms*—Machine learning, deep learning, PKLot, parking lot detection, SVM, k-NN, RF, CNN.**

I. Introduction

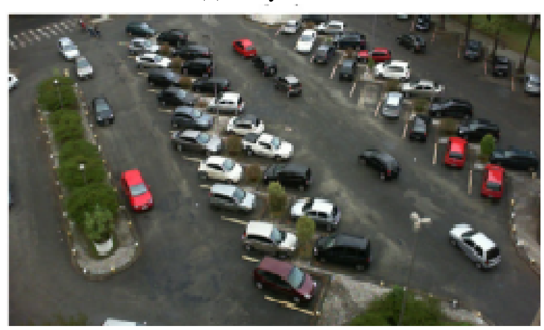
It takes a normal of 12 minutes to hang tight for a spot in city parking garages in Istanbul. This causes around 30 percent of the traffic stream in urban areas and causes clog during top hours. Drivers are impacted both mentally and monetarily during the time they spend tracking down stopping. The parking area issue, which arose with urbanization and keeps on expanding today, has turned into an examination subject. Different strategies have been created to beat this issue and save time and exertion in finding an unfilled parking spot. Sensor frameworks are utilized in many retail plazas to help drivers in this respect. These sensors illuminate drivers regarding the parking spot by illuminating green while parking spots are accessible, and red if not. Moreover, some vehicle stops and organizations have RF-ID frameworks. On account of this framework, both client charges are paid furthermore, hindrance control is given. A few frameworks permit drivers to save stopping ahead of time for explicit time spans. When we look at this large number of frameworks, we see that they are exorbitant arrangements. Sensors, RF-ID frameworks, SMS/Web reservation frameworks are frameworks laid out to fill a solitary need. Notwithstanding, CCTV cameras are accessible all over today. Albeit these cameras are utilized for security purposes, they can likewise be utilized to track down parking spots. In this review, we will inspect how we can find a parking garage with the picture taken from a CCTV camera. Picture based stopping inhabitance identification chiefly incorporates recognition of vehicle objects in parking spots. In the literature, object recognition is generally handled by extricating hand tailored visual elements like the Oriented Gradients Histogram (HOG) from pictures. What's more, fruitful results were gotten with deep learning and machine learning methods, which are among the order strategies. In this regard, we tried Support Vector Machine(SVM), k-nearest neighbors(k-NN) and Random Forest(RF) algorithms as traditional machine learning along with Convolutional neural network(CNN) as a deep learning method. Details, working principles of the algorithms explained vaguely to get a sense of how it works. Some metrics collected from the models are provided with comparisons. PKLot dataset that is used for training and testing also briefly introduced in the dataset section.

II. PKLot Dataset

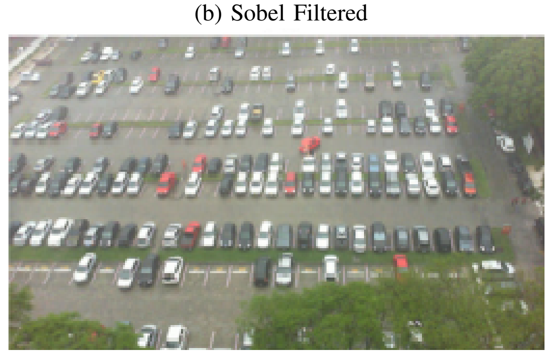
The PKLot dataset contains 12,417 images of parking lots and 695,899 images of parking spaces segmented from them, which were manually checked and labeled. All images were acquired at the parking lots of the Federal University of Parana (UFPR) and the Pontifical Catholic University of Parana (PUCPR), both located in Curitiba, Brazil.



(a) Gray-scaled



(b) Sobel Filtered



(c) Gray-scaled

However, in our machine learning methods that are SVM, k-NN and RF we used a toy data set that contains 5,400 image equally distributed as empty and occupied from the 695,899 segmented image. Also, CNN model is trained using 147,618 image distributed as 79,888 empty and 67,730 occupied parking lots from the 695,899 segmented image.

Fig. 2: Example of segmentation of parking space

III. LEARNING ALGORITHMS

In this section, algorithms used when training the models are explained to give an idea of how it works. This also helps to comprehend the performance difference between various models.

A. Machine Learning Algorithms

Machine learning can be defined as a subset of Artificial Intelligence that utilizes statistical learning calculations to fabricate frameworks that can naturally gain and improve from encounters without being expressly programmed. The problem at hand is solved using supervised learning methods which is a branch of machine learning. Since the images are labeled manually supervised learning algorithms can be applied. Various algorithms used in the work are explained in the following subsections.

1) k-nearest neighbors(k-NN): In statistics, the k-nearest neighbors algorithm (k-NN) is a non-parametric supervised learning strategy first created by Evelyn Fix and Joseph Hodges in 1951, and later extended by Thomas Cover. It is utilized for grouping and regression. In the two cases, the information comprises of the k nearest preparing models in a data set. Since in this work we dealed with k-NN classification only that will be explained.

Training examples contains multidimensional future space each contain class labels. In the training part, after splitting the data into training and validation groups, these feature vectors and their corresponding labels are stored. In classification, k is a manually selected constant value to search for the nearest k data points from the training group when a new validation data is inserted to be classified. After searching for the nearest k point in the multidimensional future space, the class of the new data point is set according to the majority voting which favours the class comes as the highest among the k nearest neighbours. To choose the closest k neighbours commonly used distance metric for continuous variables is Euclidean distance which is also the case for this work. A disadvantage of the fundamental ”majority voting” classification happens when the class conveyance is slanted. That is, instances of a more continuous class will generally rule the forecast of the new model, since they will more often than not be normal among the k nearest neighbors because of their huge number. One method for conquering this issue is to weight the order, considering the separation from the test highlight every one of its k closest neighbors. The class (or worth, in regression issues) of every one of the k closest focuses is duplicated by a weight corresponding to the inverse of the distance starting from that point the test point.

2) Random Forest: Decision trees are a famous strategy for different AI undertakings. Tree learning ”come[s] closest to meeting the requirements for serving as an off-the-shelf procedure for data mining”, says Hastie et al., ”since it is invariant under scaling and different changes of component values, is hearty to consideration of unessential highlights, and creates inspectable models. In any case, they are only from time to time accurate”

Specifically, trees that are become exceptionally profound will generally advance exceptionally unpredictable examples: they overfit their training sets, for example have low bias, yet exceptionally high variance. Random forests are an approach to averaging numerous profound decision trees, prepared on various pieces of a similar training set, fully intent on diminishing the variance. This comes to the detriment of a little expansion in the bias and some deficiency of interpretability, however for the most part enormously helps the presentation in the last model.

Forests resemble the arranging of decision tree calculation endeavors. Taking the collaboration of many trees hence working on the exhibition of a solitary random tree. However not exactly comparative, woods give the impacts of a k-fold cross validation.

One of the technique used in random forest algorithm is bootstrap aggregating, bagging for short. Bagging repeatedly selects a random sample with replacement of the training set and fits trees to these samples. The bagging method leads to better performance since it decreases sensitivity to noise in training data.

3) Support Vector Machine: Support Vector Machine is a supervised learning algorithm that identifies the best hyperplane to divide the dataset. Support vectors are the points that are closest to the hyperplane. Hyperplane is a subspace with a dimension 1 lower than its ambient space. It serves to divide the space into multiple sections. For linear SVM if the training data is linearly separable, one can select two parallel hyperplanes that separate the two classes of data, so that distance between them is maximum. This is called hard margin. if the training data are not linearly separable then by the help of hinge loss function classification can be done. This is called soft-margin. Also different kind of kernels which transforms the training data can be used to make the data linearly separable. Common kernel types are: polynomials, Gaussian radial basis and sigmoid function.

B. Deep Learning Algorithms

Deep learning is a class of AI algorithms that utilizes various layers to extricate more elevated level highlights from the crude information logically. For instance, in image processing, lower layers might recognize edges, while higher layers might distinguish the ideas pertinent to human-like digits or letters or faces. Most modern deep learning models are based on ANNs, Artificial Neural Network, and specially CNNs, Convulitional Nueural Networks. In this work, we also used those models since they are common and high-performance methods for image classification.

1) Convulitional Neural Network: Convolutional neural networks are specific kinds of artificial neural networks that utilize a numerical activity called convolution instead of general framework duplication in no less than one of their layers. They are explicitly intended to deal with pixel information and are utilized in picture acknowledgment and handling.

IV. Methodology

A. Methodolgy used in Machine Learning Algorithms

In Machine Learning Algorithms, feature engineering is an important part to be considered. In this regard, two different feature extraction methods which are Grayscale and Histogram of Oriented Gradients(HOG) applied.

1) Grayscale: As a common practice for dimensionality reduction grayscale features are extracted from the 64x64 rgb images according to formula :

$$Y' = 0.2989 * R + 0.5870 * G + 0.1140 * B$$

So which in total gives 64x64(4096) feature values to get weighted during training using machine learning classifiers.

2) Histogram of Oriented Gradients: The fundamental idea behind the histogram of oriented gradients descriptor is that neighborhood object appearance and shape inside a picture can be depicted by the appropriation of power slopes or edge headings. The picture is separated into little associated locales called cells, and for the pixels inside every cell, a histogram of gradient directions is incorporated. The descriptor is the link of these histograms. For further developed exactness, the neighborhood histograms can be contrast-standardized by working out a proportion of the power across a bigger district of the picture, called a block, and afterward utilizing this worth to standardize all cells inside the block. This normalization brings about better invariance to changes in enlightenment and shadowing.

The HOG descriptor enjoys a couple of key upper hands over different descriptors. Since it works on nearby cells, it is invariant to mathematical and photometric changes, aside from object direction. Such changes would just show up in bigger spatial locales. Also, as Dalal and Triggs found, coarse spatial testing, fine direction examining, major areas of strength for and photometric standardization allows the singular body development of people on foot to be disregarded insofar as they keep a generally upstanding position. The Hoard descriptor is in this manner especially appropriate for human recognition in pictures.

The hog features that are gotten in this work uses 1 cell per block and (2,2) pixels per cell with 9 orientations bin. So for a 64x32 resized image. 32X16X9 = 4608 feature values are obtained to get weighted when training using machine learning classifiers.

B. Methodology used in Deep Learning Algorithms

Since the deep learning methods has no need for feature engineering. Only image prepossessing was applied to the images, namely resizing all to 64x64, before forwarding the data into the models. The layers and methods used in the models are explained in the subsections.

1) Convolutional Neural Network: Convolutional layers are the core building blocks of CNNs. Layer parameters consist of trainable filters(kernels). During the forward pass, each filter is convolved with its inputs. According to element-wise multiplication between filters and the input. This produces a 2-D convolutional activation map. So that network learns a filter that activates specific features. Since Keras was used in implementing the model. Conv2d layer is used. After each layer max pooling is applied(2,2). Which reduces the size of the matrix by selecting the maximum value from the (2,2) submatrix. after convolution and max pooling steps input is flattened and put into dense layers. A summary of the model is provided in the below figure.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 62, 62, 32)        896

 max_pooling2d (MaxPooling2D  (None, 31, 31, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 29, 29, 64)        18496

 max_pooling2d_1 (MaxPooling  (None, 14, 14, 64)       0
 2D)

 flatten (Flatten)           (None, 12544)             0

 dense (Dense)               (None, 64)                802880

 dense_1 (Dense)             (None, 2)                 130

=================================================================
Total params: 822,402
Trainable params: 822,402
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3: CNN model summary

V. Evaluation and Metrics

After training, each model is evaluated on the validation test dataset which is 0.2 of the total dataset used (5400 for machine learning models and 147,618 for deep learning methods). Also, machine learning accuracy results are 3-fold cross-validated for a more reliable result.

Accuracy Results	
Model name	Accuracy %
k-NN / gray-scale ft	95.2
SVM / gray-scale ft	98.5
RF / gray-scale ft	98.7
k-NN / HOG ft	99.5
SVM / HOG ft	98.2
RF / HOG ft	94.0
CNN	99.9

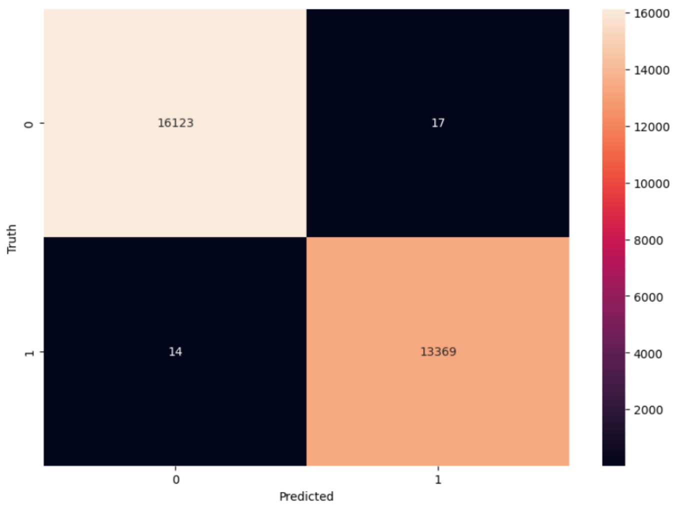


Fig. 4: Confusion matrix of the CNN model

VI. Conclusion

Taking everything into account, the issue of finding parking spots has been taken care of in various ways today. These have been sensor-based based, RF-ID-based, and picture-based arrangements. Sensor and RF-ID-based arrangements are just in a neighborhood very costly arrangements. Our answer, which is an image-based approach, we expect to utilize CCTV cameras that are as of now accessible. As a methodology, we have offered an answer that can utilize the pictures we have gotten from the live picture, mark the regions we are keen on and recognize the pixel changes in those areas. Some of the machine learning algorithms and deep algorithms are applied with high accuracy which is high and quite successful compared to related solutions in the literature

References

1 Almeida, P., Oliveira, L. S., Silva Jr, E., Britto Jr, A., Koerich, A., PKLot – A robust dataset for parking lot classification, Expert Systems with Applications, 42(11):4937-4949, 2015.