



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: V Month of publication: May 2019

DOI: <https://doi.org/10.22214/ijraset.2019.5040>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Object Detection for Accident Prevention using Convolution Neural Network

Gaurpriya Chodanker¹, Gajanan Gawde²

^{1, 2}Computer Science and Engineering, Goa College of Engineering, Farmagudi, Goa, India

Abstract: *The growth of automobile population and the dearth of appropriate infrastructure have been a serious issue causing road accidents on a large scale all around the globe. Detection of objects for assisting drivers is a fascinating but challenging dilemma in the arena of pattern recognition. It is the most intricate study that has attracted the eye of several researchers. Convolution Neural Network is achieving popularity due to its performance and ability to classify complex images. In this paper we have designed our own CNN architecture to detect an object from an image. This CNN model was tested with different sized datasets of images with and without noise. The results obtained using different random filter initialization functions are presented.*

Keywords: *Object detection, Pattern Recognition, Machine Learning, Deep learning, Convolution Neural Network*

I. INTRODUCTION

Objects have diverse characteristics and can be distinguished with respect to its different properties like color, texture, shape, motion etc. The proficiency of the systems to identify objects from a scene is referred to as object detection. The large changes caused by occlusion, background clutter, viewpoint variation, illumination, shadow in the visual display of objects are some of the challenges faced in detection of objects. Selection of appropriate features hence is a key role in the design of such systems. An extensive study is carried out and enormous approaches have been invented for detection of object from images. Jiawei Han and Gong Cheng in their survey of object detection have classified four main classes of methods based on template matching, machine learning, OBIA(Object Based Image Analysis), and based on knowledge^[3].

This field in pattern recognition of object detection is growing at a very rapid speed and has found wide applications in techniques like identification and detection of faces, vehicles, pedestrians, web images, security systems and driverless cars. The task of deriving information from an object in the image is very intrinsic due to the variation in size, image resolutions, complex background etc. Selection of appropriate features hence is pivotal in the design of such systems. The advancements in machine learning has lead to the growth of deep leaning.

Deep learning methods are well-known for their automatic feature extraction techniques and outstanding performance. Convolutional Neural Networks play a dominant role in the processing of natural language, videos and images, speech and pattern recognition etc.

India is well-known globally for the large number of street mishaps. Some of the major sources of accidents include distracted driving, drunk driving, over speeding, bad weather conditions, disobeying traffic rules etc. Latest researches have shown that driver's sleepiness represents up to 20% of deadly mishaps on motorways and dreary streets, which impede the driver's judgment and their capacity of controlling vehicles.

The alarming growth in automobiles on road and the lack of a smart road safety alert system has lead to rise in road accidents. Animals, pedestrians and other non living objects may be other targets of traffic accidents on a global scale, costing human lives, injuries and damage to properties. Driving safely is supposed to be significant concern of social orders everywhere throughout the world.

Deaths and injuries due to accidents is a very serious concern. The development and growth of intelligent systems for driver safety and assistance could be very beneficial to curtail the growing count of accidents that are happening on a large scale. The proposed work aims to employ deep learning to build a CNN model to detect objects in images. Datasets of different sizes are used for the purpose of training the model. This work aims to

- 1) To train a network using CNN and detect object from the visual image, and
- 2) To evaluate the model's performance.

II. ANALYSIS

A classification algorithm is used to match input a data sample to a particular class. The decision of choosing a correct classifier depends on the dataset type and the complexity of the problem. In deep learning techniques like Convolution Neural Network(CNN), Deep Belief Networks (DBN), Deep Boltzmann Machines (DBM) etc. categories are learned directly from its hidden layers. Convolution Neural Network work very well for complex visual images recognition containing data having spatially recurring patterns like images, speech etc. The intricacy of the model is increased by the addition of more layers. Good results are obtained when more data is made available. However the time taken to train such networks is more as more parameters are involved but testing time is less.

1) Comparision with Some Supervised, Non – Parametric Machine Learning Techniques Used To Classify Multiple Classes

Support vector machine(SVM) can be used on all kinds of data and have found wide applications. To improve the separation between classes the samples are projected on a feature space having high dimensions thus creating a lot of support vectors. It is not possible to increase the complexity of the model. It is more suitable for problems having datasets with fewer outliers. K-nearest neighbors (KNN) is a simple and easily implementable classifier. It performs very well with a lot of data points however remains sensitive to bad features therefore feature selection is very crucial. It achieves excellent performance when there are a lot of points in a low dimensional space. The test time increases on increasing data size. In decision trees complexity grows with the rise of number of trees and training samples. It can be used to solve problems on both classification and regression.

TABLE I
Comparision Of Supervised Classifiers

Classifier	Type	Time to train	Accuracy
CNN	Non-linear	High	High
SVM	Linear	High	Moderate
K - NN	Non-Linear	Less	Low
Decision Tree	Non-Linear	Less	Low

III.CNN ARCHITECTURE

CNNs are similar to the human visual processing system, being highly optimized in structure for processing 2D and 3D images, and are effective at learning and extracting abstractions of 2D features. CNN was first discovered by Fukushima^[1] but was rarely used due to limits of computation hardware for training the network. However, later good results were found when gradient based learning algorithm were applied to CNN by LeCun^[2] for classification of handwritten digits. CNNs work efficiently even using few preprocessing techniques.

CNN comprises of an input layer, multiple hidden layers and an output layer. It involves extracting features and then classification. Each layer has a set of neurons. The neurons from the top layer are fully connected to the neurons of the layer below. The input low-level feature vectors is fed into first layer of the model and, than high-level features vectors are obtained at the end as we move from layer to layer. The output contains the probability of the input vector belonging to a particular category. No prior knowledge is required by these networks.

A. Some key elements of convolution neural network

- 1) *Convolution Layer:* A set of filters, containing a matrix of $k \times k$ of weights that contains parameters of the model to be learned is convolved with the entire image matrix. The speed at which the filter moves across the image, or the number of units it shifts every time is known as the stride. This produces a convoluted image that acquire low level features that can find preliminary visual information, like oriented edges, points at the ends, corners which are then integrated at the higher layers to get high level distinguishable features. The application of the filter on the input image yields an output matrix with a size smaller than the original image. Thus padding is done to get the same size output.
- 2) *Activation Functions:* They are non-linear functions that decide the actual output. As compared to other activation functions like tanh, sigmoid etc. that suffer due to vanishing gradients, ReLU(Rectified Linear Unit) can be used train networks with multiple layers hence is popularly used in CNN's. Leaky ReLU a variation of ReLU known to fix the problem of "dying ReLU" issue, as it does not have parts with zero-slope. It speeds up the training process thus making training faster.

- 3) **Pooling Layer:** This sub sampling layer involves dividing the image into rectangular sets that do not overlap over each other. Pooling is performed in different ways using functions like max, average and L2-norm. Max pooling is most commonly used as it computes the maximum value as the output for each sub region and is known to outperform the other functions. It makes the features invariant against noise and distortions. It reduces the spatial size of the representation and also controls over-fitting. It is very usefull in absorbing shape variations.
- 4) **Fully Connected Layer:** This last layer of the CNN model is connected to all the activation functions and is combined into a feature vector of one dimension thus taking a flattened form of the features obtained from the top layers. It has an output for each category in the task of recognition. A Softmax function is used to calculate the probability the individual target category with respect to all the possible target categories.
- 5) **Dropout Layer:** When a trained model learns every detail along with noise in the data such that the performance of the model is affected negatively, overfitting occurs. Dropout layer is thus added to curb this problem of overfitting.

B. Configuration of our CNN model

The model consists of a total of three layers, containing two convolutional layers containing convolution, activation and maxpooling function and one fully connected dense network. The convolution layer in the first layer has 32 number of filters of size of 3×3 and 64 in the second layer. LeakyReLU is used as the activation function. The max pooling layer has a window size of 2×2 and a stride equal to 2. The flattened layer flattens the output and passes it as an input to the dense layer which in turn passes its output as input to the output layer. Categorical crossentropy is used as the cost function to measure the overall loss. The gradient descent algorithm is used to decrease the cost function to approach the minimum point^[6]. Adam optimizer is the best gradient descent algorithm having a batch of size 256.

IV. EXPERIMENT AND RESULTS

We have trained the CNN model in Anaconda environment with Python version of 3.6. A python API of higher level, Keras can build and train neural network models very swiftly using Tensorflow library as the back-end. 2-D images of objects belonging to three categories namely bikes, cars and humans were collected and used to prepare datasets of different sizes. These images of different colors and sizes were converted to grayscale and rescaled to a size of 50×50 to reduce computation costs. Since data has different formats, we have applied normalization to make this data formats uniform.

The proposed model is evaluated using 80% images from the dataset for the purpose of training while testing was performed on the remaining 20% images of the dataset. A comparative study was performed on datasets of different sizes using different filter initialization functions for convolution layer. The total time to train, accuracy and loss of the model were recorded when the model was tested on different datasets of images with noise and without noise. Each function has a different outcome and same was compared with others for performance analysis.

TABLE 1 Performance Analysis Of Designed Cnn Model

Sr. No	Total No. of Images in the Dataset	Filter Initialisation	Total Training Time		Accuracy		Loss	
			Without Noise	With Noise	Without Noise	With Noise	Without Noise	With Noise
1.	5000	Uniform Distribution	66.36s	60.04s	81.33%	74.46%	0.45	0.60
		Normal Distribution	64.89s	62.09s	86.06%	74.6%	0.37	0.59
		Glorot Normal Initializer	63.92s	59.94s	83.33%	74.33%	0.4	0.59
		Glorot Uniform Initializer	59.16s	59.56s	83.2%	76.03%	0.42	0.57
2.	10000	Uniform Distribution	113.01s	112.57s	96.8%	96.03%	0.09	0.12
		Normal Distribution	113.84s	112.08s	98.23%	95.16%	0.05	0.14
		Glorot Normal Initializer	118.80s	111.93s	98.01%	96.3%	0.06	0.13
		Glorot Uniform initializer	115.03s	117.46s	97.75%	95.73%	0.07	0.95
3.	21000	Uniform Distribution	229.46s	231.72s	95.8%	92.66%	0.12	0.19
		Normal Distribution	298.39s	236s	96.69%	92.33%	0.09	0.20s
		Glorot Normal Initializer	240.68s	238.9s	95.8%	91.66%	0.11	0.21s
		Glorot Uniform Initializer	234.39s	255.19s	97.1%	91.18%	0.08	0.24s

The model achieves highest accuracy of 98.23 using normal distribution for initializing convolutional filter when trained for 113.84s while an accuracy of 96.3 was obtained using glorot normal initialiser to train noisy images for 111.93s on a dataset having 10000 images.

V. CONCLUSIONS

Road accidents continue to be on rise and have been a concern worldwide. Efficiently designed systems for object detection can be very helpful to drivers in reducing accidents. We have proposed CNN model with three layers consisting of two convolutional layers and a fully connected dense layered network in order to detect objects. Initially, we trained the CNN model with datasets containing images with and without noise. Then different random convolutional filter initialization functions were used to test the effect on the performance. The training time of CNN model was recorded and it was small value. CNN model is capable of capturing complex features from the image. Classification of the test samples using CNN is very efficient and can outperform humans. We have achieved maximum accuracy of 98.23 without noise and 96.3 with noisy images.

REFERENCES

- [1] Fukushima and Kunihiko, "Neocognitron: A hierarchical neural network capable of visual pattern recognition", Neural networks 1.2, 1988
- [2] LeCun and Yann, "Gradient-based learning applied to document recognition" Proceedings of the IEEE, vol. 86, no. 11, 1998.
- [3] Gong Cheng and Junwei Han, "A survey on object detection in optical remote sensing images", ISPRS Journal of Photogrammetry and Remote Sensing, 2016
- [4] Andrej Karpathy, George Toderici and Sanketh Shetty, "Large-scale Video Classification with Convolutional Neural Networks", IEEE Conference on Computer Vision and Pattern Recognition, 2014
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Communications Of The ACM, 2017
- [6] Yann LeCun, Leon Bottou, Genevieve B. Orr and Klaus-Robert Muller, "Efficient BackProp", Springer 1998
- [7] Zeiler, M. and Fergus, R. "Visualizing and understanding convolutional networks", In ECCV, 2014
- [8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", ICLR, 2014.
- [9] Kaiming He and Jian Sun, "Convolutional neural networks at constrained time cost.", In CVPR, 2015.
- [10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions" in IEEE Conf. on CVPR, 2015.
- [11] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," in BMVC, 2014.
- [12] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," NeuroImage, vol. 108, pp. 214–224, 2015.
- [13]



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)