Road Signs Classification

Itamar Ziv-On
Ori Ben-Hamo
Boaz Sharabi
Ariel University, Ariel, Israel

Keywords: Road Signs Classification, Convolutional Neural Network, Deep Learning

Abstract: Road signs classification play an important role in the unmanned automatic driving. In this paper, a road signs classification method is presented based on Convolutional Neural Network. In this method, each image is represented by grayscale value pixels as features and a number related to its road sign name as the image label. The method is used for classification based on the extracted features. The experimental results show that the method achieves a very good performance on road signs classification, with a high 99.69% accuracy rate.

INTRODUCTION

Nowadays, unmanned automatic driving technology has attracted increasing attentions from research and industry communities. The road signs classification play an important role in this field. For this purpose, it is possible to use several tools.

This paper will show the use in Logistic Regression with SoftMax, Neural Network (NN) and Convolutional Neural Network (CNN), In addition a comparison between this three tools has been described.
This article expands and show the different results depending different models.

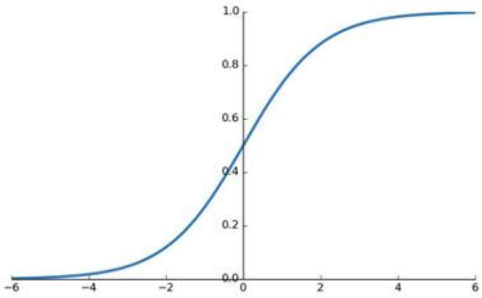


Figure 1: SoftMax/Sigmoid activation function

RELATED WORK

There is a lot of research and methods on this topic, we do not pretend to innovate anything on the subject. The goal of this project is our learning how to build a model that will give the best results to the problem of classifying road signs.

REQUIRED BACKGROUND

Logistic Regression – a statistical model that in its basic form uses a logistic function to predict the class of given sample, The result given by logistic regression is suppose to relate to the probability of the instance belonging to the class p(y= $Class_i$ | X) for every $Class_i$.

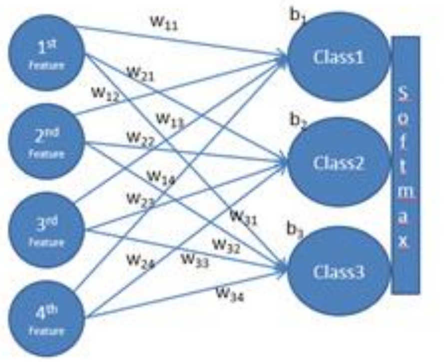


Figure 2: A Logistic Regression model

Neural Network – An extended model that contain layer (or layers) of neurons that is learned from the previous features layer and with their help it is possible to learn better about the sample.

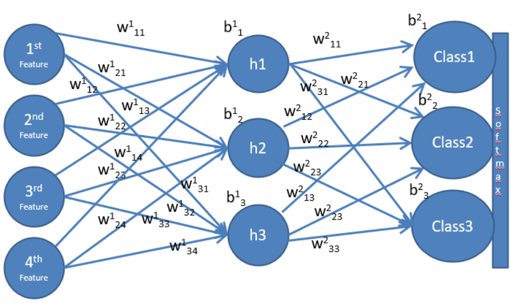


Figure 4:A Neural Network model

Convolutional Neural Network – A complex model used especially for images that learn the relation between close pixels. This model consists of more several layers which precede to the Neural Network layers. The additional layers are Convolutional layers and pooling layers.

Convolutional layers

Each Convolutional layer contain number of kernels, the kernel consists of several weights. The kernel goes over the whole image and from each area of pixels outputs one value, from each kernel comes a matrix of values. Convolutional layers convolve the input and pass its result to the next layer, each convolutional neuron processes data only for its receptive field (depending on kernel's size). Each neuron learned from a image's pixel and its surrounding pixels. In each convolutional layer of CNN model, each neuron takes input from a larger area of pixels in the input image than previous layers. This is due to applying the convolution over and over.

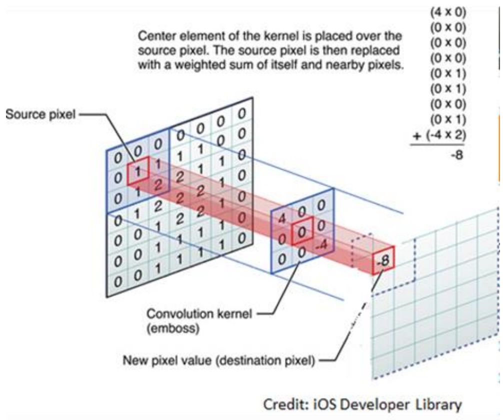


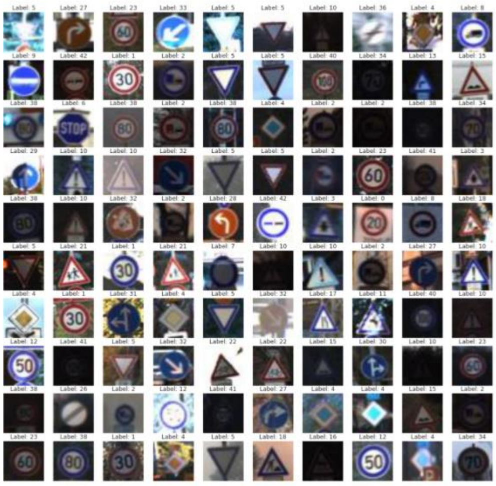
Figure 3: A Convolutional layer

Pooling layers

Convolutional networks use pooling layers to reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer, combines small clusters, typically 2 x 2. There are two common types of pooling: max and average. Max pooling uses the maximum value of each cluster of neurons at the prior layer, while average pooling instead uses the average value. Although fully connected feedforward neural networks can be used to learn features and classify data, this architecture is impractical for images. It would require a very high number of neurons, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10,000 weights for *each* neuron in the second layer. Instead, convolution reduces the number of free parameters, allowing the network to be deeper. For example, regardless of image size, tiling 5 x 5 region. Using regularized weights over fewer parameters avoids the vanishing gradient and exploding gradient problems seen during backpropagation in traditional neural networks.

EXPERIMENTAL DATA

Forty-three different road signs classes, each class contains about three thousand images. The images size is 32x32 pixels. 60% from the data used for training the model, 20% for validation and the rest 20% for test.



PREVIOUS ATTEMPTS

The first model examined for the classification is simple logistic regression with mini batch of size 50 and Gradient Decent optimizer. The accuracy received is 93.55%. The second model is Neural Network with one neurons layer of size 100, ReLU activation function, mini batch of size 50 and Gradient Decent optimizer. The accuracy received is 90.59%. The third model is Neural Network with two neurons layers, the first of size 100 and the second of size 50, ReLU activation function for both layers, mini batch of size 50 and Adam Decent optimizer. The accuracy received is 96.48%.

PROJECT DESCREPTION

Convolutional Neural Network

The model contains two convolutional layers, two Max-Pooling layers and two fully connected layers. Every image represents by 32x32 matrix of features that represent by grayscale value. Each convolution layer contains multiple convolution kernels. These convolution kernels are able to scan the image features via different expressions, based on this, we can acquire various feature maps in different locations. The Max-Pooling layer, following the convolution layer, is mainly used to reduce the resolution of the feature map, to extract the existing image features and to determine the features' relative location. The fully connected layer is give further learning, do the model more complex in order to improve learning.
The output from convolution layers and the first fully connected layer is pass through ReLU activation function which give the model non-linear combination of the features, simplify the gradient calculations and avoid vanishing gradient in the Back propagation updates (related to Sigmoid activation function). The output of the last Max-pooling layer is reshaped from matrix to vector for the input of the first fully connected layer. The model use Dropout regularization on the first fully connected layer to avoid overfitting. The output of last fully connected layer pass through SoftMax activation function because that case of multi-class classification. The prediction of image based on the maximum value in the output vector of SoftMax.

Table 3: CNN Constructing Parameters.

Layer	Type	Output	Kernel
0	Input	32×32	
1	Convolutional	$32\times32\times32$	5×5
2	Max-Pooling	$16\times16\times32$	2×2
3	Convolutional	$16\times16\times64$	5×5
4	Max-Pooling	$8\times8\times64$	2×2
5	Flatten	1×4096	
6	Fully Connected	1×1024	
7	Dropout	1×1024	
8	Fully Connected	1×43	

CNN training

In order to train the model, we used mini batch of size 50. The reason of using mini batch is that in large dataset it's takes long time to go through the entire dataset and perform one update of the weights and with the mini-batch method the updated occurred after 50 samples only.

In addition, the Batch Gradient Descent (all data passing) requires a lot of memory in the back propagation step.

Another reason is that the mini batch helps avoid being stuck at a local minimum position.

In the learning process the weights and the bias's values is updated by Back Propagation using Adam optimizer and learning rate α of 0.001.

Adam optimizer

The Adam optimizer is combination of other two optimizers (RMSProp and Momentum SGD) to achieve better updates of weights and bias's value

$$m := \beta_1 m + (1-\beta_1)\frac{\partial L}{\partial w} \qquad \textit{RMSProp}$$
$$w := w - \alpha m$$

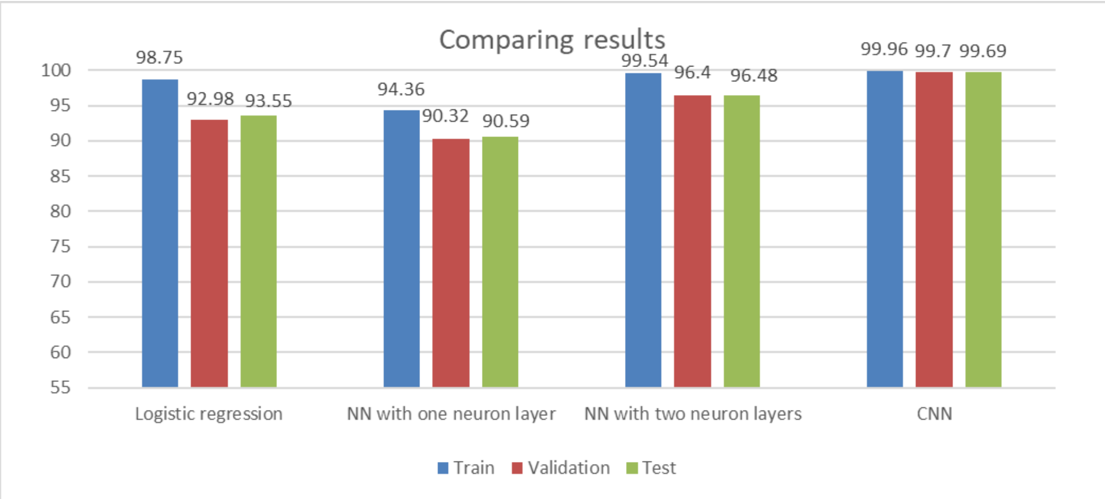
$$r := \beta_2 r + (1-\beta_2)\left(\frac{\partial L}{\partial w}\right)^2 \qquad \textit{Momentum SGD}$$
$$w := w - \alpha\frac{\frac{\partial L}{\partial w}}{\sqrt{r+\epsilon}}$$

β_1 and β_2 is the decay factors.

The update of w according to Adam optimizer:

$$\hat{m} = \frac{m}{1-\beta_1^t} \text{ and } \hat{r} = \frac{r}{1-\beta_2^t}$$
$$w := w - \alpha\frac{\hat{m}}{\sqrt{\hat{r}+\epsilon}}$$

t – is the round of update the model



Conclusion

In this paper, a classification method based on CNN has been proposed. In the training process, the deep image features are extracted by CNN in the Grayscale values. The different between simple logistic regression and neural network with one and two neurons layers was not significant while the different between those models to CNN was widely significant. In addition, the accuracy with Adam optimizer was much better then Gradient Descent optimizer. The accuracy of road signs classification is very important due to the fact that its use is for real time systems.

With 99.69% accuracy of the CNN model described in this paper , the model can be trusted for using in practice. In order to use this model for autonomous vehicles, a model for recognizing road signs compared to other objects is required first. In the future, we plan to add this model and complete the mission.