Digit Recognition using Fixed Feature Maps

Venkat Giridhar Mareedu[1] Shrinath Raju[2] Apurva Chintakunta[3]
[1,2,3]MVSR Engineering College

Abstract— Digit Recognition is a well-researched and a well-documented field. But the practical applications have been limited due to the error rates and the impact of every single wrong prediction of digit on the output process. This paper aims at designing a recognition model with a specific application using fixed feature maps. This paper does not implement a feature map extractor but with the input digit image taken in a fixed format it uses a CNN with pre-defined positions of feature maps to predict digits with highest possible accuracy rate.

Key words: Digit Recognition, CNN, Feature Maps

I. INTRODUCTION

With the growing amount of digitization around the world, almost everything is automated these days. The need to eliminate monotonous and redundant tasks has been identified so as to increase the speed of the process and remove the scope of human error. This automation has created numerous potential users with reduced skill threshold. It has also introduced several new problems and opportunities for development.

Powerful computers and reduced size devices have led to a new possibility of applications. Handwritten document processing is one such application which has been in use effectively since long. Alphabet recognition is relatively easier as they are done within their respective contexts. Whereas, digit recognition is much more convoluted and has 0 tolerance for error. With the use of inexpensive devices, the paper aims to predict digits narrowed down to a specific application.

Ideally, a webcam is used to capture the image containing digits. The image is then processed so as to break it down into individual digits and reduce it to a pre-defined form. Several other image processing techniques can be applied to eliminate noise and reduce it to grayscale. The designed system then predicts the individual digits based on the gray scaled images. This paper mainly focuses on recognition of digits of vehicle number plates.

II. METHODOLOGY

A. Image Processing:

Each picture captured is reduced to grayscale. It is then broken down into individual digits. Each individual image is then reduced to 29*29 pixels because of the considerations described by Dr. Simard [2]. These pixels form the input to the 5-layer convolutional neural network with layer 0 being the input layer.

Each layer in the CNN has fixed size feature maps with varying degree of detail. Layer 1 has fewer number of feature maps compared to layer 2 but with bigger size. Similarly layer 2 has fewer number of feature maps compared to layer 3 but bigger in size. This trend continues till layer 5 which is the output layer which has 10 neurons corresponding to the 10 digits.

Fig. 1.1: Original Captured Image

Fig. 1.2: Image converted to grayscale

Fig. 2: Image broken down into individual digits and padded into a square of 29*29 pixels

B. Activation Function

The activation function used is a scaled version of the hyperbolic tangent [1]. Scaling causes the function to vary between ±1.7159, and permits us to train the network to values of ±1.0. The relative ease of calculating the derivative of hyperbolic function makes it a good choice. The value of derivative can also be expressed in terms of the output value. More specifically, given that:

$$x = F(y) = \tanh(y) = \frac{\sinh(y)}{\cosh(y)}$$

where y is the input to the function (corresponding to the activation value of a neuron) and x is the output of the neuron.

Then

$$\frac{dF}{dy} = \frac{d}{dy}\left(\frac{\sinh(y)}{\cosh(y)}\right) = \frac{\cosh^2(y) - \sinh^2(y)}{\cosh^2(y)}$$

$$\frac{dF}{dy} = 1 - \tanh^2(y)$$

Which simplifies to

or since $x = \tanh(y)$, the result is

$$\frac{dF}{dy} = 1 - x^2$$

C. Fixed Feature Maps

The feature map is the output of one filter applied to the previous layer. As the input image is taken in a fixed format, the feature maps are fixed. As the vehicle number plates are printed in a standardized format, the features of a particular digit can be stored and for prediction, fixed feature maps are used to extract the features and then these features are compared with the stored features to get the output digit.

This compared against just pixel by pixel comparison gives better results as it allows certain amount of noise in the captured images. For example, a number plate wherein there is some amount of damage to it cannot be used for pixel by pixel comparisons as it cannot match the damaged part with any pre-defined images. By extracting features, a missing part of a digit can be ignored as for the most part, the features of the image match the stored features.

D. Requirements

- Computer with a camera peripheral.
- Any sequence of digits in a standardized format.

III. PROPOSED SYSTEM

The designed system enables the user to read digits in a particular format automatically without human intervention. More functionalities are proposed to be added to this system such as:

- Expanding beyond digits i.e. not just restricting to digits and including other characters as well.
- Using pre-defined sequence of digits to predict missing digits. For example, if a number plate reads "AP 1* CA 5765" where * is the missing digit. Then if the system has a repository of possible sequences then one of those sequences can be used to match the missing digit.
- Enabling auto-capture of images.
- Improving the scalability

IV. CONCLUSION

With the implementation of fixed feature maps, the accuracy is near perfect for images that perfectly comply with the standardized format for that particular application. Wide-scale usage of it can automate many slow procedures such as toll-gate process on highways. The scope of this paper is mainly to enhance the existing digit recognition models by introducing fixed feature maps restricted to their respective applications.

REFERENCES

[1] https://www.codeproject.com/Articles/16650/Neural-Network-for-Recognition-of-Handwritten-Digi
[2] Patrice Y. Simard, Dave Steinkraus, John Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis," External Link International Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society, Los Alamitos, pp. 958-962, 2003.
[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
[4] Y. LeCun, L. Bottou, G. Orr, and K. Muller, "Efficient BackProp," in Neural Networks: Tricks of the trade, (G. Orr and Muller K., eds.), 1998.

