George Joseph, Shobhik Bhadraray and Shahnawaz Ahmed

Cyclops : A CNN based number recognition tool

I. INTRODUCTION

Identification of text in natural environments is a difficult problem. Convolution neural networks have shown promising trends towards prediction of text and this work will explore the prediction of digits of various lengths using the SVHN dataset. We will use TensorFlow as the backend for building the convolution neural network with Keras.

II. DATA

The dataset we use is The Street View House Numbers (SVHN) dataset. It consists of images with labeled digits and we have bounding box information for each digit.

A. Description
We have the following counts for numbers in the data set.

Digits	Number
1	5137
2	18130
3	8691
4	1434
5	10

B. Preprocessing images

The preprocessing involves stitching together individual bounding boxes and generating numbers of various lengths. We assume that the largest sequence in this data set is of length 5. We use vectors of length 6 as the output for each image with the first 5 elements denoting the digits and the last element denoting the length of sequence. All images are re-sized to 50x50. This was performed using the Python Image Library using a cubic spline interpolation.

C. One hot encoding of target

The target vectors are initially of dimension 6 with the first 5 elements representing the number and the 6th element representing the sequence length. '0' is represented by 10 and a 0 in the target vector represents that the digit is not present.

$$[6, 10, 6, 0, 0, 3]$$
$$[4, 5, 4, 0, 0, 3]$$

The target vectors are one hot encoded and the vectors of length 6 are now converted into matrices of shape 6 x 11. Each digit is now represented by a 11 dimensional vector.

D. Data augmentation

Since the proportion of data was skewed, we generate more data for 3, 4 and 5 digit numbers for training

Digits	Number
1	5137
2	18130
3	26821
4	10124
5	1434

III. MINI GUN: BASELINE MODEL

We use a baseline model to predict all values in the output vector - digits and the length of the sequence. The data used is the original data where we have not performed any augmentation.

A.Architecture

B.Results

IV.   NUMBER TOWER

Number tower is the predictor for the number of digits in the image

A.Architecture

B.Results

V. ION CANNON: INDEPENDENT TRAINING FOR DIGITS

proportion of images of each sequence length.

A.Architecture

A.Results

VI. THE COMPLETE PIPELINE

A.Image preprocessing

B.Number tower prediction

C.Digit prediction

D.Results on validation set
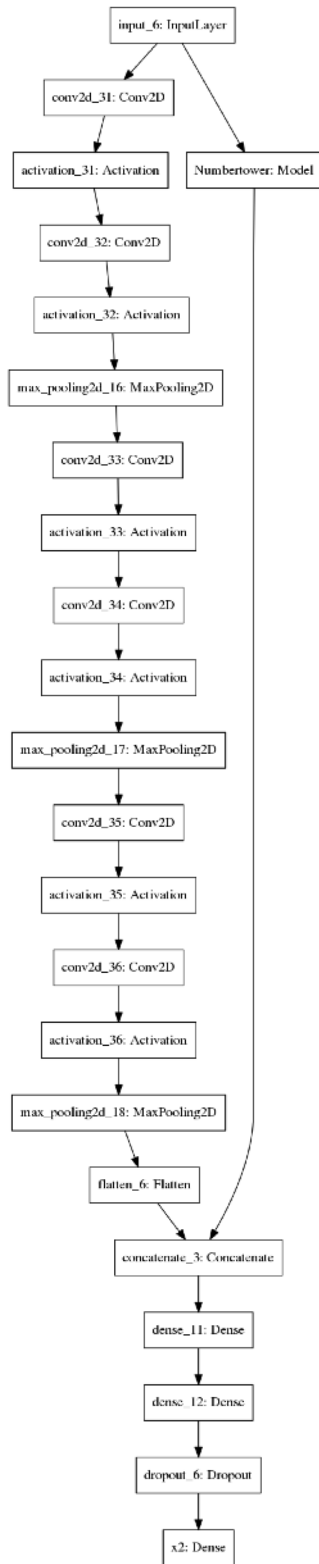
VII. CONCLUSION

We

ACKNOWLEDGMENT

A

REFERENCES

1. Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet (2013, Dec). Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. Available: https://arxiv.org/abs/1312.6082

input_6: InputLayer

conv2d_31: Conv2D

activation_31: Activation

Numbertower: Model

conv2d_32: Conv2D

activation_32: Activation

max_pooling2d_16: MaxPooling2D

conv2d_33: Conv2D

activation_33: Activation

conv2d_34: Conv2D

activation_34: Activation

max_pooling2d_17: MaxPooling2D

conv2d_35: Conv2D

activation_35: Activation

conv2d_36: Conv2D

activation_36: Activation

max_pooling2d_18: MaxPooling2D

flatten_6: Flatten

concatenate_3: Concatenate

dense_11: Dense

dense_12: Dense

dropout_6: Dropout

x2: Dense

acc

input_1: InputLayer

digit1: Model

digit2: Model

digit3: Model

digit4: Model

digit5: Model