

Training a Neural Network to Segment, Detect and Recognize Numbers

Authors : Abhisek Dash

Abstract : This study had three neural networks, one for number segmentation, one for number detection and one for number recognition all of which are coupled to one another. All networks were trained on the MNIST dataset and were convolutional. It was assumed that the images had lighter background and darker foreground. The segmentation network took 28x28 images as input and had sixteen outputs. Segmentation training starts when a dark pixel is encountered. Taking a window(7x7) over that pixel as focus, the eight neighborhood of the focus was checked for further dark pixels. The segmentation network was then trained to move in those directions which had dark pixels. To this end the segmentation network had 16 outputs. They were arranged as "go east", "don't go east ", "go south east", "don't go south east", "go south", "don't go south" and so on w.r.t focus window. The focus window was resized into a 28x28 image and the network was trained to consider those neighborhoods which had dark pixels. The neighborhoods which had dark pixels were pushed into a queue in a particular order. The neighborhoods were then popped one at a time stitched to the existing partial image of the number one at a time and trained on which neighborhoods to consider when the new partial image was presented. The above process was repeated until the image was fully covered by the 7x7 neighborhoods and there were no more uncovered black pixels. During testing the network scans and looks for the first dark pixel. From here on the network predicts which neighborhoods to consider and segments the image. After this step the group of neighborhoods are passed into the detection network. The detection network took 28x28 images as input and had two outputs denoting whether a number was detected or not. Since the ground truth of the bounds of a number was known during training the detection network outputted in favor of number not found until the bounds were not met and vice versa. The recognition network was a standard CNN that also took 28x28 images and had 10 outputs for recognition of numbers from 0 to 9. This network was activated only when the detection network votes in favor of number detected. The above methodology could segment connected and overlapping numbers. Additionally the recognition unit was only invoked when a number was detected which minimized false positives. It also eliminated the need for rules of thumb as segmentation is learned. The strategy can also be extended to other characters as well.

Keywords : convolutional neural networks, OCR, text detection, text segmentation

Conference Title : ICCVIP 2018 : International Conference on Computer Vision and Image Processing

Conference Location : Mumbai, India

Conference Dates : February 22-23, 2018