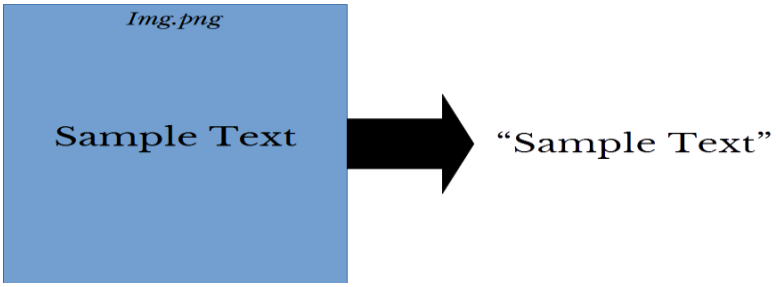
WEEK_4

Deep Learning for Image Search

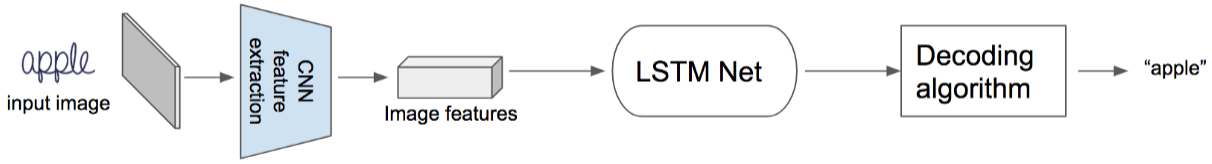
Introduction:

Optical Character Recognition is a process when images of handwritten, printed, or typed text are converted into machine-encoded text. Automated recognition of documents, credit cards, car plates and billboards significantly simplify the way we collect and process data.

Traditional OCR techniques are typically multi-stage processes. For example, first the image may be divided into smaller regions that contain the individual characters, second the individual characters are recognized, and finally the result is pieced back together.



How OCR works?

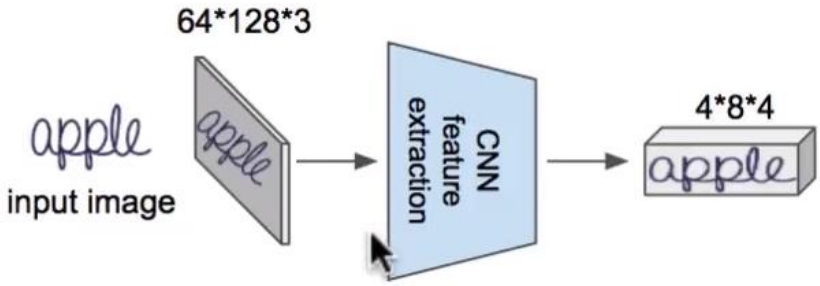


Firstly, image is feed to CNN to extract image features. The next step is to apply Recurrent Neural Network to these features followed by the special decoding algorithm. This decoding algorithm takes lstm outputs from each time step and produces the final labeling.

Step by Step Process:

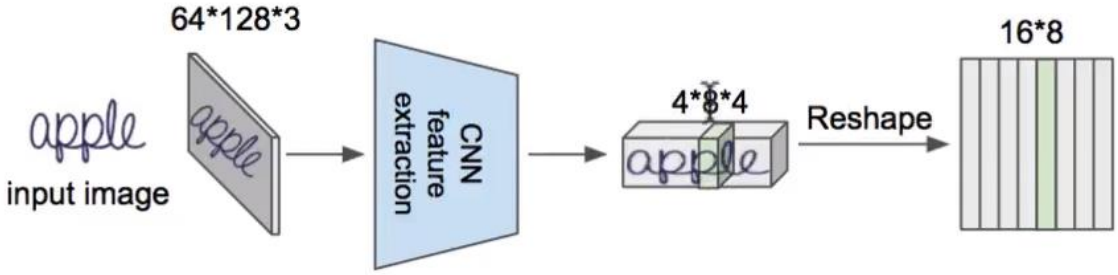
Step:1

We put image “apple” to the feature tensor so you can understand how to interpret it. Height equals to 4, width equals to 8 (These are spatial dimensions) and num channels equals to 4. Thus, we transform input image with 3 channels to 4 channel tensors.



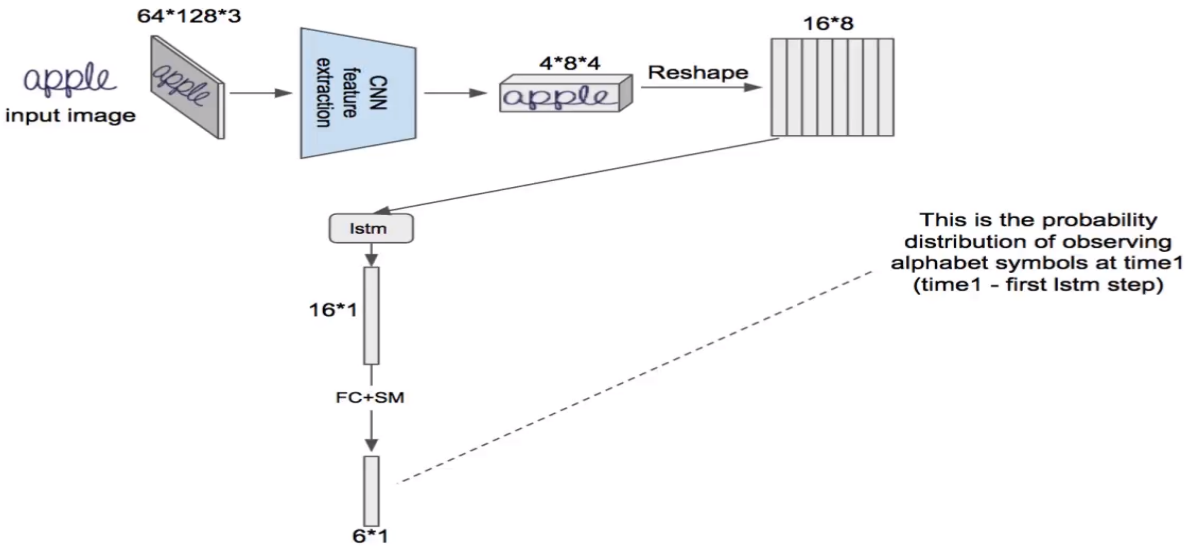
Step:2

Next, we do reshape operation. After that we obtain the sequence of 8 vectors of 16 elements. After that we feed these 8 vectors to the LSTM network and get its output — also the vectors of 16 elements.



Step:3

Then we apply fully connected layer followed by SoftMax layer and get the vector of 6 elements. This vector contains probability distribution of observing alphabet symbols at each LSTM step.



Step:4

As a result, we obtain the string of eight characters — one most probable letter at each time step. Then we have to glue all consecutive repeating characters into one. In our example two “e” letters are glued to single one. Special blank character allows us to split symbols that are repeated in the original labeling. We added blank symbol to the alphabet to teach our neural network to predict blank between such case symbols. Then we remove all blank symbols.