A REVIEW PAPER ON OCR USING CONVOLUTIONAL NEURAL NETWORKS

Ujwala B.S, Sumathi K
Department of Electronics and Communication Engineering,
JNN College of Engineering, Visvesvaraya TechnologicalUniversity, Karnataka, India

Abstract— **this paper presents a literature review on OCR for different languages using convolutional neural network techniques. Optical Character Recognition is the process of converting an input text image into a machine encoded format. Different methods are used in OCR for different languages. The main steps of optical character recognition are pre-processing, segmentation and recognition. Recognizing handwritten text is harder than recognizing printed text. Convolutional Neural Network has shown remarkable improvement in recognizing characters of different languages. The novelty of the OCR is its robustness to image quality, image contrast, font style and font size. Common machine learning methods usually apply a combination of feature extractor and trainable classifier. The use of CNN leads to significant improvements across different machine-learning classification algorithms.**

Keywords— **Convolutional neural network, Deep learning, Optical Character Recognition, Feature Extraction.**

I. INTRODUCTION

Optical Character Recognition [6] is a process that can convert text, present in digital image, to editable text. It allows a machine to recognize characters through optical mechanisms. The output of the OCR should ideally be same as input in formatting .The process involves some pre processing of the image file and then acquisition of important knowledge about the written text.

Deep learning [1] Techniques has achieved top class performance in pattern recognition tasks. These include image recognition, human face recognition , human pose estimation and character recognition . These deep learning techniques have proved to outperform traditional methods for pattern recognition. Deep learning enables automation of feature extraction task. Traditional methods involve feature engineering which is to be done manually. This task of crafting features is time consuming and not very efficient. The features ultimately determine the effectiveness of the system. Deep learning methods outshine traditional methods by automatic feature extraction.

Convolutional Neural Networks (CNN) is a popular deep learning method and is state of the art for image recognition. Handwritten character recognition is a difficult task as the characters usually has various appearances according to different writer, writing style and noise. Researchers have been trying to increase the accuracy rate by designing better features, using different classifiers and combination of different classifiers. These attempts however are limited when compared to CNN. CNNs can give better accuracy rates but it has some problems that needs to be addressed.

II. MOTIVATION

In the last years the trend to digitize (historic) paper based documents such as books and newspapers, hasemerged. The aim is to preserve these documents and make them fully accessible, searchable and processable in digital form [6]. Knowledge contained in paper based documents is more valuable for today's digital world when it is available in digital form. The first step towards transforming a paper based archive into a digital archive is to scan the documents. The next step is to apply an OCR (Optical Character Recognition) process, meaning that the scanned image of each document will be translated into machine processable text. Due to the print quality of the documents and the error-prone pattern matching techniques of the OCR process, OCR errors occur. Modern OCR processors have character recognition rates up to 99% on high quality documents. Assuming an average word length of 5 characters, this still means that one out of 20 words is defect. Thus, at least 5% ofall processed words will contain OCR errors. On historic documents this error rate will be even higher because the print quality is likely to be of lower quality.

After finishing the OCR process several post- processing steps are necessary depending on the application,

e.g. tagging the documents with meta-data (author, year, etc.) or proof-reading the documents for correcting OCR errors and spelling mistakes. Data which contains spelling mistakes or OCR errors is difficult to process. For example, a standard full-text search will not retrieve misspelled versions of aquery string. To fulfill application's demanding requirements toward zero errors, a post-processing step to correct these errors is a very important part of the post-processing chain.

A post-processing error correction system can be manual, semi-automatic or fully automatic. A semi-automatic post-correction system detects errors automatically and proposes corrections to human correctors who then have to choose the correct proposal. A fully-automatic post-correction system does the detection and correction of errors by its own. Because semi-automatic or manual corrections require a lot of human effort and time, fully-automatic systems become necessary to perform a full correction.

III. DESIGN OF OCR

The OCR System consists of the following stages: Image is feed into the system as an input.

A. Pre-processing

In the preprocessing stage, the character image is processed for removing all the undesirable entities from an image to make the process of recognizing easier. The input images are resized to a suitable format. It must not be too large. Preprocessing helps to remove the noise from an image, suppresses unwanted distortions, enhances some image features and hence can play an important role in OCR.

B. Segmentation

Segmentation is one of the most important phases of OCR system. By applying good segmentation techniques we can increase the performance of OCR. Segmentation subdivides an image into its constituent regions and objects. Basically in segmentation, we try to extract basic constituent of the script, which are certainly characters. Thisis needed because our classifiers recognize these characters only.

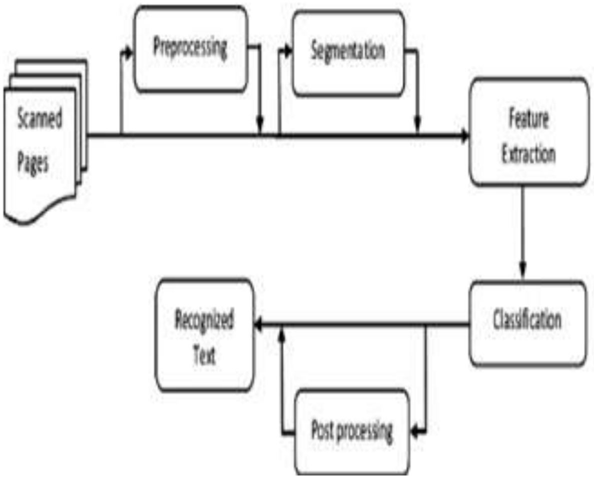

Figure I: Overall system Architecture.

C. Feature Extraction

The extraction features reduces the dimensionality of the representation and makes the recognition process computationally efficient. This method defines each character by the presence or absence of key features, including height, width, density, loops, lines, stems and other character traits. Feature extraction is a perfect approach for OCR of magazines, laser print and high quality images. Feature extraction can be defined as extracting the most representative information from the raw data, which minimizes the within class pattern variability while enhancing the between class pattern variability. For this purpose, a set of features are extracted for each class that helps distinguish it from other classes, while remaining invariant to characteristic differences within the class.

A CNN consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers. Description of the process as a convolution in neural networks is by convention. Mathematically it is a cross-correlation rather than a convolution. This only has significance for the indices in the matrix, and thus which weights are placed at which index..

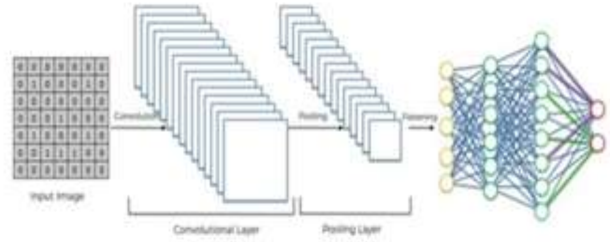

Figure 2: Operation of Convolutional Neural Networks

Convolutional layers apply a convolution operation to the input, passing the result to the next layer. The convolution emulates the response of an individual neuron to visual stimuli. Each convolutional neuron processes data only for its receptive field. Although fully connected feed forward neural networks can be used to learn features as well as classify data, it is not practical to apply this architecture to images. Convolutional networks may include local or global pooling layers which combine the outputs of neuron clusters at one layer into a single neuron in the next layer. For example, max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Another example is average pooling, which uses the average value from each of a cluster of neurons at the prior layer. Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional multi-layer perceptron neural network (MLP).

Each neuron in a neural network computes an output value by applying some function to the input values coming from the receptive field in the previous layer. The function that is applied to the input values is specified by a vector of weights and a bias (typically real numbers). Learning in a neural network progresses by making incremental adjustments to the biases and weights. The vector of weights and the bias are called a filter and represent some feature of the input. A distinguishing feature of CNNs is that many neurons share the same filter. This reduces memory footprint because a single bias and a single vector of weights is used across all receptive fields sharing that filter, rather than each receptive field having its own bias and vector of weights.

D. Classification

When an input image is feed into the character recognition system, all-important features are retrieved and inputted to a trained classifier like an artificial neural network. A comparison of the input features with stored patterns is done to find the appropriate match class for the input image. This is done with the help of classifiers. Correct labeled training data is required to classify test samples.

E. Post Processing

In this stage accuracy of recognition is further increased by connecting dictionary to the system in order to perform Syntax analysis, semantic analysis kind of higher level concepts, which is applied to check the recognized character. Post-processing stage is the last stage of the proposed recognition system. It prints the corresponding recognized characters in the structured text form. OCR accuracy can be improved by using post processing techniques and character is recognized.

IV. RELATED WORKS

Pranav P Nair et al [1] focuses on Convolutional neural network to extract features and uses CNN to extract and classify Malayalam characters. This is a method different from the conventional method that requires handcrafted features that needs to be used for finding features in the text. Hence provides the chance for giving higher accuracy rate for Malayalam characters.

CNN modeling means modeling the structure of CNN. The number of convolution layers, max pooling layers, ReLu layers and fully connected layers needs to be chosen. It is not possible to determine the exact number of layers that will yield the best outcome. Hence it is vital to try several configurations of the network and choose which network bestsuits. Size of Feature map=m*n+1, n <= m, where m is the height and width of the image, n is the height and width of the convolution layer. Max pooling has shown to be the most effective pooling strategy and hence will be used by our system. Back propagation using gradient decent will be used as the learning rule. The number of neurons in each convolutional layer was adjusted to get the maximum accuracy rate. An increasing pattern will be followed the number of neurons for each layer. A dropout layer will be used to aid in the training process. The dropout layer decreases the complexity and training time of the network.

It uses Sample generation and CNN modeling which are time consuming tasks and the later also requires a CUDA enabled GPU for parallel processing.

Khaled S. Younis et al [2] have discussed a new implementation for the automatic recognition of handwritten characters that implements deep neural networks with some regularization techniques, namely Dropout and batch normalization and apply it using Tensor Flow framework. The additional regularization parameters lead to better outcomes with regard to the accuracy. The accuracy reached by DNN with Dropout in two consecutive layers is almost 1% higher than that when only dropout was used in one layer only. Adding another hidden layer of ReLu activation function also improved the performance by almost 1%.

DNN model consisted initially of 5 layers, the first input layer has 784 neurons that are essentially the input pixels of each MNIST digits image for each successivehidden layer, the input is multiplied by the weights, the products are added to the biases and the results are the input to ReLu activation function. This is cascaded till the output layer.

The first hidden layer consisted of 500 neurons, the second had 1000 neuron, and the third had 250 neurons. To improve the generalization ability of the model for predicting the class of unseen samples, dropout technique at the end of the last layer is used with keep probability parameter of 0.5 where at each epoch training iteration half of the neurons of the last layer get activated while the other half activation will be set to zero. This tends to prevent our network from over fitting by not building a model so is so tightly bonded to training samples. Finally, the output layer had 10 neurons to match the number of classes. However, before applying the activation function in each layer, they found it useful to apply a technique called Batch Normalization as a final layer, this help to fit the data when it fed into the activation function and make the training process faster and effective using back propagation.

The implemented model can be still improved by adding regularization, tweaking parameters or by applying more advanced CNN.

Meduri Avadesh et al [3] discusses about OCR for Sanskrit using convnets as classifiers for Indic OCRs and convnets are more suitable than SVMs and ANNs, for multi-class image classification problems. It showed that designed OCR is ideal for digitizing old and poorly maintained material as it robust to font size and style, image quality and contrast. Convnet are more suitable than SVMs and ANNs, for multi-class image classification problems. Categorical Cross-Entropy as a cost function that is the appropriate cost functions for multi-class classification problems. They used Adam Optimizer to find the minima ofthe cost function with a varying learning rate between 0.03 - 0.0001, that recalculate its value after each batch. As we shall see in the Results section, minor modifications were made to this architecture to improve the performance.

Learning ability can still be improved by introducing letter segmentation and identification. This could be achieved with the help of a selective search algorithm followed by an R-CNN.

Shailesh Acharya et al [4] have described about the use of new system architecture deep Convolutional neural networks and image datasets to detect and classify text with the use of Dropout and dataset increment approach to improve test accuracy.

In this paper Convolution layer consists of the raw pixel values from the 32X32 grayscale image and has no trainable parameters. The first convolution layer has 4 feature maps with 784 units/neurons each (28 x 28). Each feature map is shown in figure as 2D planes and they have different set of weights. All the units in a feature map share the same set of weights and so they are activated by the same features at different locations. This weight sharing not only provides invariance to local shift in feature position but also reduces the true number of trainable parameters at eachlayer.

Each unit in a layer receives its input from a small neighborhood at same position of previous layer. So the number of trainable weights associated with each unit in a convolutional layer depends on the chosen size of the neighborhood of previous layer mapped to that unit. Since all the units are activated only from the input taken from a local neighborhood they detect local features such as corners, edges, end-points. For a 5x5 kernel, the number of input weights for each unit is 25. In addition the units also have a trainable bias. The total number of units in a layer depends upon the size of kernel in the previous layer and overlap between the kernels. The convolution layer is followed by a sub sampling/pooling layer. Sub sampling layer reduces the resolution of the feature map from convolution layer by averaging the features in the neighborhood or pooling for a maximum value. Because the exact position of features vary for different images of the same character, it is more desirable.

Khaled s. younis[5] have presented a deep neural network for the handwritten Arabic Character recognition problem that uses Convolutional neural network (CNN) models with regularization parameters such as batch normalization to prevent over fitting.

Here Batch Normalization technique is used which is a normalization and regularization technique to address the following issues that appear during the training process of deep neural networks: 1. Internal Covariate Shift: which refers to the change in the distribution of input of each layer (features) that is affected by parameters in all input layers in which a small change in the network can significantly affect the entire network; and 2. Vanishing Gradient in saturating nonlinear functions: such as tanh and sigmoid, which are prone to get stuck in the saturation region as the network grows deeper despite the proposed solutions to carefully initialize the network, using small learning rate or replacing these functions by ReLU function.

This system suggests the use of Batch Normalization as a part of the network architecture and it was experimentally proven to cause an improvement in terms of speed and accuracy. The Batch Normalization layer is added just before the nonlinearity and especially after the convolutional layers to limit its output away from the region of saturation using the mean and variance. The Pooling or sub sampling layer reduces the dimensionality of each filtered image, but preserves the most important features in the previous layer. Pooling can be of different types: Maximum, Average Sum, etc. The output will have the same number of images, but they will each have fewer pixels. However, it is argued that max-pooling can be redundant and could be replaced by purely using convolutional layer with increased stride without loss in accuracy.

Dropout layers are also used in convolutional neural networks with the aim of reducing over fitting. This layer "drops out" a random set of neurons in that layer by setting their activation to zero. Finally, the fully connected layers are the basic building blocks of traditional neural networks. They treat the input as one vector instead of two dimensional arrays. Full connection implies that every neuron in the previous layer is connected to every neuron in the next layer. The output from convolutional and pooling layers represents high level features and fully connected layers used to classify material (input images) into the appropriate class based on the training of the dataset. They used the Softmax activation function to output probabilities between 0 and 1 for each class representing the confidence that a certain character belongs to a specific class. For updating the weights during training, they used the Categorical Cross-Entropy as a cost function which is the appropriate cost function for multi-class classification problems.

By the literature review it can be concluded that a novel algorithm for OCR using Convolutional neural networks can be designed to improve learning ability, accuracy and computational time and to Explore, investigate deep neural network to recognize the characters in Indic languages.

V. CONCLUSION

This paper has presented a related work on OCR for different CNN techniques and also various available techniques are studied to find a best technique. But it is found that the techniques which provide better results are slow in nature while fast techniques mostly provide inefficient results. It is found that OCR techniques based on neural networks provide more accurate results than other techniques.

VI. ACKNOWLEDGEMENT

This work was supported by the JNNCE research centre; Shivamogga. Ujwala B S, Sumathi K gratefully acknowledges the support from JNNCE, Shivamogga.

VII. REFERENCES

[1]. Pranav P Nair, Ajay James C Saravanan(2017). "Malayalam Handwritten Character Recognition Using Convolutional Neural Networks" , International Conference on inventive Communication and Computational Technologies .

[2]. Khaled S. Younis, Abdullah A. Alkhateeb(2017) "A New Implementation of Deep Neural Networks

for Optical Character Recognition and Face Recognition'', Proceedings of the New Trends in Information Technology (NTIT-2017). The University of Jordan, Amman, Jordan. 25-27 April.

[3]. Meduri Avadesh , Navneet Goyal (2017) "Optical Character Recognition for Sanskrit using Convolution Neural Networks" IAPR International Workshop on Document Analysis Systems.

[4]. Shailesh Acharya, Ashok Kumar Pant ,Prashnna Kumar Gyawali (2015) "Deep Learning Based Large Scale Handwritten Devanagari Character Recognition'' 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA).

[5]. Khaled S. Younis(2017) "Arabic Handwritten Character Recognition based on Deep Convolutional Neural Networks" Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 3, No. 3, December . [6]Sukhpreet Singh (2013) "Optical Character Recognition Techniques: A survey"International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 6, June

[6]. Rahul Khadse "Survey of Research on Optical Character Recognition using Artificial Neural Network, Genetic Algorithm, Fuzzy Logic and Vedic Mathematics"

[7]. MamtaKadyan, Deepti Ahlawat(2017) "A Review on Character Recognition Using OCR Algorithm" Journal of Network Communications and Emerging Technologies (JNCET Volume 7, Issue 5, May .

[8]. [9]A. Krizhevsky, I. Sutskever, and G. E. Hinton. (2012) "Imagenet classication with deep convolutional neural networks", In Advances in neural information processing systems, pages 1097–1105.

[9]. V.J. Dongre, V.H.Mankar (2010) "A Review of Research on Devnagari Character Recognition," International Journal of Computer Applications" Vol.12,No 2,pp.0975-8887,November.

[10]. M. Z. Alom, P. Sidike, T. M. Taha, and V. K. Asari, "Handwritten bangla digit recognition using deep learning," arXiv preprint arXiv:1705.02680.

[11]. M. Alex and S. Das, (2016) "An Approach towards Malayalam Handwriting Recognition Using Dissimilar Classifiers," Procedia Technology, vol. 25, pp. 224-231.

[12]. I. M. Keshta, (2017) "Handwritten Digit Recognition based on Output- Independent Multi-Layer Perceptrons," HAND, vol. 8, 2017.

[13]. L. M. Seijas, R. F. Carneiro, C. J. Santana, L. S. Soares, S. G. Bezerra, and C. J. Bastos-Filho, (2015) "Metaheuristics for feature selection in handwritten digit recognition," in Computational Intelligence (LA-CCI), 2015 Latin America Congress on, pp. 1-6.

[14]. V. L. Sahu and B. Kubde,(2013) "Offline handwritten character recognition techniques using neural network: a review," International journal of science and Research (IJSR), vol. 2, pp. 87-94.