ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/329146883

On Georgian Handwritten Character Recognition

Article *in* IFAC-PapersOnLine · January 2018

DOI: 10.1016/j.ifacol.2018.11.279

CITATIONS
5

READS
435

6 authors, including:

Davit Soselia
San Diego State University
6 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Magda Tsintsadze
Ivane Javakhishvili Tbilisi State University
22 PUBLICATIONS **40** CITATIONS
SEE PROFILE

Levan Shugliashvili
San Diego State University
3 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Irakli Koberidze
San Diego State University
4 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Wildfire Pattern Analysis using Machine Learning View project

All content following this page was uploaded by Davit Soselia on 26 November 2018.

The user has requested enhancement of the downloaded file.

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-30 (2018) 161–165

On Georgian Handwritten Character Recognition

Davit Soselia. Magda Tsintsadze* **.**
Levan Shugliashvili. Irakli Koberidze**. Shota Amashukeli**. Sandro Jijavadze**.**

**Iv. Javakhishvili Tbilisi State University, Tbilisi, Georgia*
(Tel: (+995) 577-20-4433; e-mail: magda.tsintsadze@tsu.ge).
***San Diego State University Georgia, USA*
(e-mail:{dsoselia,mtsintsadze,lshugliashvili,ikoberidze,samashukeli,sjijavadze}@sdsu.edu)

Abstract: The article addresses the issue of Georgian handwritten text recognition. As a result of the performed research activity, a framework for recognizing handwritten Georgian text using Self-Normalizing Convolutional Neural Networks (CNN) was developed. To train the CNN model, an extensive dataset was created with over 200 000 character samples. This framework has been deployed as a web service, as well as in the form of apps for Windows, Linux, and iOS.

© 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Artificial intelligence, CNN, SNN, OCR, Handwritten Character Recognition.

1. INTRODUCTION

Data processing is an inseparable part of daily routines and represents one of the top issues of contemporary research; the problem of computationally processing handwritten text remains an unresolved task for many languages and for Georgian in particular (Khachidze et al, 2017). A large portion of writings exists in the handwritten form exclusively, and significant manpower is expanded in manually typing such texts in the efforts for their preservation and transformation into an editable form. This problem is particularly relevant at office workplaces, as manually producing searchable and editable backups of partially or fully handwritten documents can be time- and resource-consuming. As a result, the creation of high accuracy handwritten text OCR (Optical Character Recognition) has been a priority challenge for decades. While significant progress has been made for the OCR systems dedicated to recognizing Latin characters (Patel et al., 2012), many languages' character sets, including those of Georgian language, do not yet have recognition models or datasets of acceptable quality. In today's interconnected world, with problems often needing global and long-term solutions, tackling non-latin scripts can be beneficial to technological development and cultural progress. Georgian handwriting script's cursive nature provides an unique technological challenge and existence of such a system can streamline access to a large number of scripts available in Georgian language only.

Due to the availability of large datasets and relatively less cursive nature of English handwritten language, most of the progress in the field of handwriting OCR has been done for English characters, with models achieving over 95% single character accuracies, with the errors minimized by high quality auto-correction systems.

Prior to this work, no large-scale handwritten characters set had been created for Georgian script. In our research process, we collected 200 000 Georgian character samples by scanning and segmenting handwritten texts by over 200 authors and deploying an online handwritten sample collector.

In our initial attempt we trained several tested models on the dataset, including VGG (Simonyan, 2014), ResNet (He et al., 2015), and Inception (Szegedy, 2015). However, due to the high complexity and variety of Georgian character written forms, we achieved the highest performance with a novel architecture based on based on the self-normalizing structure presented by Klambauer et al., (2017).

2. RELATED WORK

Optical Character Recognition has been a priority problem for much research, and significant progress has been made in the field throughout the past century. For high-quality scanned images of printed Latin-character text the problem was initially solved using Fourier descriptors; this solution was later improved using support vector machines. Much less progress had been made in the field of handwritten OCR due to the huge variety of handwritten characters; this problem had remained unsolved for decades even for the Latin character set, despite the existence of high-quality datasets of handwritten characters. Recent advances in computer vision, especially in methods utilizing deep convolutional neural networks, have improved the performance of handwriting recognition systems significantly (Yang et al., 2016).

Different types of character recognition techniques exist to target different kinds of OCR tasks. The most common form of OCR is the offline process that deals with the analysis of the static images of characters and texts; another form of OCR is the so-called online method, in which more information about the handwriting process is gathered by

2405-8963 © 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.11.279

capturing the handwriting motions such as the direction and the momentum of the input device "on fly"; this approach, while effective and far less resource intensive, can only function when the above-mentioned information is available, for example when writing involves a mouse or a stylus on digital devices (Keysers et al., 2017). Convolutional neural networks allow for an effective offline recognition of characters and text in static images. In addition to CNN-based architectures, RNN-based architectures have been used to facilitate OCR in cursive-based scripts, as per Ahmed et al. (2018), where the RNN architecture is used to facilitate character prediction by augmenting the network's decision process with the information about the overall sequence of characters.

3. APPROACH AND METHODS

We targeted the development of an offline handwriting recognition system capable of recognizing Georgian handwritten text and converting it to digital text. For this, we used a convolutional neural network architecture.

3.1 Artificial Neural Networks

Artificial neural networks, also known as multilayer perceptrons, are "a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs" (Reh, 2012). Artificial neural networks (ANN) are composed of nodes which imitate neurons in the organic brain.

ANN consists of several layers consisting of the so-called nodes, or neurons. The layers of ANN are the input layer, one or more hidden layers, and the output layer. The nodes of the input layer are linked with the nodes of the first hidden layer. Each node in the first hidden layer represents an activated sum of weighted values from the input layer. Each connection of a node of the input layer and that of the first hidden layer has an associated weight that determines how strongly that input affects the target node (Reh, 2012).

The first hidden layer acts as the input layer of the subsequent layer, and so on until the output layer's values are determined. To make an accurate prediction for the model, the weights need to be adjusted by training model to the data in such a way that when the model's inputs are set to be one of the entries of the dataset, the model's output matches the entry's associated label. Assuming the weights are set to the optimal value, predictions are made by propagating the values of the input data forward through the network and determining the values of the nodes in the output layer.

In order to give a neural network the ability to learn non-linear functions, the value of each node is "activated" with a non-linearity, meaning its value is passed through a non-linear function with properties beneficial to the network's training procedure. In addition, a single shared weight referred to as bias is learned for each layer of the network; this weight is summed with the pre-activation value of each node in the next layer.

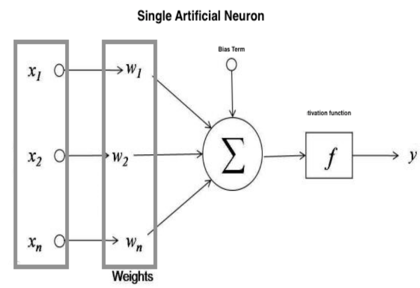

Fig. 1. Single node in ANN: sum of member-wise multiplications of vectors X (input) and W (Weights) is summed with the previous layer's bias, and the result is passed through activation function f, the output of which determines the node's value.

3.2 Convolutional Neural Networks

Convolutional neural network (CNN) is a type of artificial neural network that uses sliding-window weight structures to extract features from visual information. In 2012, CNNs showed state of the art performance in digit recognition for MNIST dataset with an error rate of 0.23%.

CNNs augment the traditional ANN architectures through additional layer structures used in place of hidden layers. Convolutional layers use learned weigh kernels of small sizes to generate visual features from the input by "sliding" the kernel over the input and associating each location of the input with the dot product of the kernel and the input's kernel-sized section placed at the current location (the current location acting as the "origin point" of the section); the resultant associations form a so-called feature map of the input. In most practical implementations, a single convolutional layer contains multiple kernels, and thus a feature map is produced for each of the layer's kernels, and the layer's output contains a vector for each location on the input, where the mth element of the vector in the nth location is the value of the nth location on the feature map produced through the convolutional layer's mth kernel.

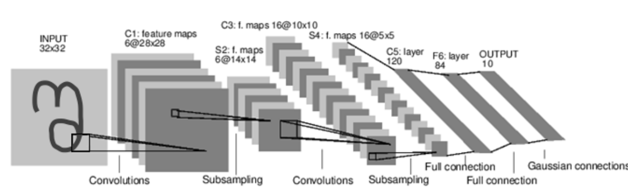

Fig. 1. A Georgian character being passed through a CNN.

Mathematically, the output of the convolutional layer with a 2D input and a single 3-by-3 kernel at the location (i,j) can be represented as

$$[F(i,j) = \sum_{m=0}^{2} \sum_{n=0}^{2} I(i+m, j+n) K(m,n)] ,$$

where I(x,y) represent's the value of the input at location (x,y) and K(x,y) is the value inside the kernel at the location (x,y).

As with the hidden layers of ANNs, the values on the feature maps produced by convolutional layers are normally passed through an activation function. In our model we use the SeLU activation function introduced in Klambauer et al. (2017).

$$f_{t_l,a_l t_r a_r}(x) = \begin{cases} t_l + a_l(x - t_l) & \text{for } x \leq t_l \\ x & \text{for } t_l < x < t_r \\ t_l + a_l(x - t_l) & \text{for } x \leq t_l \end{cases}$$

Commonly, the feature maps produced by the convolutional layer subsequently have their dimensionality reduced through the so-called max-pooling, or subsampling layers, which divide the feature maps into sections of given size and only output the highest value in each section. After the input passes through multiple convolutional and max-pooling layers, it is most commonly 'flattened' (the tensor's elements are placed in a single vector) and passed through fully-connected hidden layers, followed by the output layer.

3.3 Dataset

The creation or the selection of a dataset is a crucial step in machine learning. The publicly available dataset of letters we found had scarce and unbalanced data. In respect to mentioned problem the implementation of OCR system for Georgian handwriting scripts required solution of the following tasks:

- Development of dataset of Georgian optical characters
- Segmentation and pre-processing of optical information
- Development of neural network and its training

Due to lack of dataset with labelled Georgian characters, we had to develop our own system from scratch to collect data and create a dataset. In frames of the research appropriate Web and Windows applications in Python and C# respectively were implemented later on used to collect the required data from volunteers in a quick and efficient manner.

We manually gathered, scanned, and pre-processed images of over 120 000 handwritten Georgian characters as well. In total, we managed to acquire over 200,000 samples of characters used for recognition model training later.

Choosing the model, which would work well for our problem, was a challenging task. We tested several classical architectures such as VGG models, Inception, ResNet and a more recent SNN (Klambauer et al., 2017) architectures. Testing of the trained architectures was performed on a cross-validation portion of the dataset, specifically set aside for testing, as well as by allowing a number of volunteers to test the application in the real world setting. One of the most efficient results were acquired using ResNet50, which accepts 64x64 pixel pictures and output layer consists of 33 nodes which correspond to the numeric value of characters in Georgian alphabet.

To achieve the best possible accuracy for the dataset we have trained on all the leading CNN networks, modifying them for our use, as well as creating a new CNN model based on Self-Normalization. We have trained modified VGG16, inception, models for Georgian character recognition. VGG16 is a high-depth model that uses smaller kernels for weight preservation through layers. Is often considered goto CNN. GoogLeNet developed in 2015 adds parallel structure to basic sequential models Patel et al (2012).

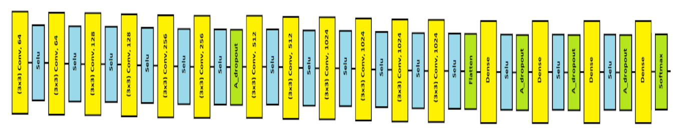

Fig. 2. VGG-based convolutional neural network with SeLU activations; our proposed architecture.

To increase accuracy, we used the latest advances in Self Normalizing Neural Networks (SNN), that have shown good promise on dense layers. And through the process of repeated testing, parameter modification, and pre-processing adjustment we developed new SNN model. The model is based on structure of VGG but changes convolution and max pooling arrangement, but most importantly using SeLU activation, Alpha-dropout and Increased neuron count in Dense layer.

The model has demonstrated the optimal compromise between processing time and accuracy. Offering Complexity and adjustable parameter size that is less than VGG, while offering +3% performance on datasets tested.

4. RESULTS

The training accuracy for single character based on different CNN architecture is represented in Table 1:

Table 1. Accuracy

Model	Test Accuracy
Inception (Imagenet pre-trained)	63%
VGG16 (Cifar100 pre-trained)	54%
CNN English dataset	99.1%
VGG16(Chinese characters from Liu et al., (2011))	74%
VGG16	89%
ResNet	95%
SNN	94%

We got maximum accuracy in cross-validation between training and testing sets of 94.3% after training both VGG and GoogleNet models, with and without trained convolutional weighs. This result was achieved using adopted weights for Convolutional layers from training on ImageNet dataset. The worst result was demonstrated by VGG model, from the ground up training.

After 12 hours of training for each model, VGG achieved 78 % accuracy using ImageNet trained Convolutional Layers and 50% accuracy on the ground up training. This can mostly be attributed to small sample size that resulted in over fitting of the model thus leading to low accuracy generalization for the testing image set.

At the same time, GoogleNet achieved 93.8 % accuracy on ImageNet trained model and 23% accuracy on the ground up model. The last one is due to the depth of the GoogleNet, typical ground-up calculation time of which using a GTX Titan X system is around 1 week, while the 94.3% accuracy can be considered sufficient for the handwritten recognition system.

SNN has required training time of under 6 hours, while achieving over 94% accuracy and the least training and prediction times.

The SNN model was also tested on English language Dataset (Cohen et all., 2017) achieving State if the art accuracy of over 99%.

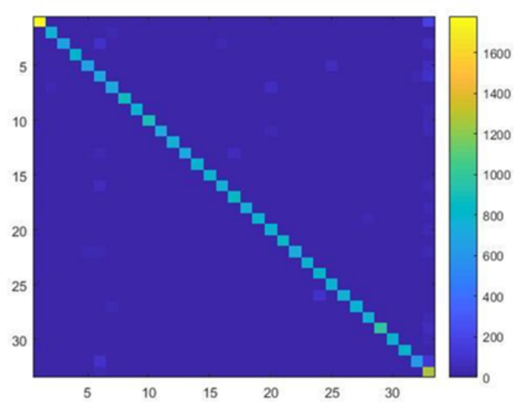

Fig.4.Confusion Matrix

The results were validated by generating Confusion Matrix and calculating Precision and Recall for each character.

Figure 4 describes the confusion matrix of the classifier for each character class, figure 5 provides the accuracy and figure 6 respectively recall results in percentage on the testing set.

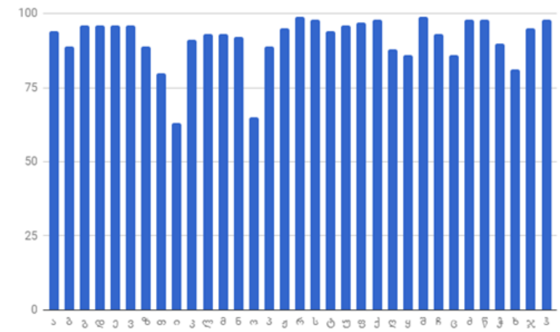

Fig.5. Precision

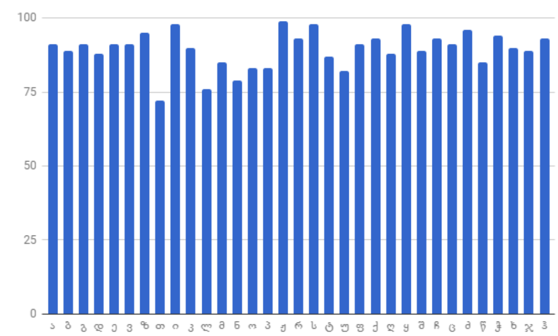

Fig.6. Recall

5. CONCLUSIONS

The presented work proposes a convolutional neural network to enable recognition of Georgian handwritten characters. While testing different architecture for CNN, best performance was demonstrated by Self Normalizing Neural Network achieving over 94% accuracy, the least training epochs, and the shortest prediction time. During the research appropriate framework was implemented for web (python), Desktop (c#), and ios(swift), offering high accuracy recognition for short handwritten texts. Such a software solution can serve as a tool in old handwritten text preservation, as well as increase efficiency and accessibility of Georgian handwritten texts.

Proposed model can be applied to other languages as demonstrated by showing state-of-the-art performance on the English handwritten character dataset. Therefore, the framework has wide academic and industry applications

While significant progress has been made in terms of creating datasets, choosing and training CNN models, there remain multiple topics for further studies as well. More prominent of which looks to be an image segmentation, to enable scanning of words rather than letters, introducing auto-correction option, expanding datasets to decrease over fitting, and using newest custom layers to find the best possible ANN model (Kim et al., 2016).

REFERENCES

Ahmed, S., Naz, M., Razzak, S., Rashid, M., Afzal M., and Breuel, T. (2018). "Evaluation of cursive and non-cursive scripts using recurrent neural networks".

Cohen, G., Afshar, S., Tapson, J. and van Schaik, A., 2017. Emnist: an extension of mnist to handwritten letters. *arXiv preprint arXiv:1702.05373*.

He, K., Zhang, X., Ren, S. and Sun, J. (2015). *Deep Residual Learning for Image Recognition*. [online] Arxiv.org. Available at: https://arxiv.org/abs/1512.03385 [Accessed 28 May 2018].

Keysers, D., Deselaers, T., Rowley, H.A., Wang, L.L. and Carbune, V., 2017. Multi-language online handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), pp.1180-1194.

Kim Y., and Cha, E. (2016) "Streamlined GoogLeNet algorithm based on CNN for Korean character recognition", *Journal of the Korea Institute of Information and Communication Engineering,* vol. 20, no. 9, pp. 1657-1665.

Khachidze M., Tsintsadze M., Archuadze, M. (2016), "Natural Language Processing Based Instrument for Classification of Free Text Medical Records", *BioMed Research International,* Vol 2016, Article ID 8313454, 10 pages

Klambauer, G., Unterthiner, T., Mayr, A. and Hochreiter, S. (2017) "Self-normalizing neural networks"

Liu, C.L., Yin, F., Wang, D.H. and Wang, Q.F., 2011, September. CASIA online and offline Chinese handwriting databases. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on* (pp. 37-41). IEEE.

Patel, C., Patel, A., and Patel, D., (2012) "Optical character recognition by open source OCR tool tesseract: a case study", *International Journal of Computer Applications*, vol. 55, no. 10, pp. 50-56

Reh, (2012) "OCR — optical character recognition", *Orthopädie & Rheuma*, vol. 15, no. 1, pp. 58-58.

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015, June. Going deeper with convolutions. Cvpr.

Yang, W., Jin, L., Tao, D., Xie , Z. and Feng, Z. (2016). "DropSample : a new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition", *Pattern Recognition*, vol. 58, pp. 190-203.

View publication stats