Multi-font Printed Amharic Character Image Recognition: Deep Learning Techniques

Birhanu Hailu Belay[1](✉), Gebeyehu Belay[1], Tewodros Amberbir Hebtegebrial[2], and Didier Stricker[2]

[1] Bahir Dar Institute of Technology, Bahir Dar, Ethiopia
birhanu.hailub@gmail.com, ge.be09@yahoo.com
[2] DFKI, Technical University of Kaiserslautern, Kaiserslautern, Germany
tedyhabtegebrial@gmail.com, didier.stricker@dfki.de

Abstract. In this paper, we propose a technique to recognize multi-font printed Amharic character images using deep convolutional neural network (DCNN) which is one of the recent techniques adopted from the deep learning community. Experiments were done on 86,715 Amharic character images with different level of degradation and multiple font types. The proposed method has fewer pre-processing steps and outperforms the standard approach used in classical machine learning techniques. We systematically evaluated the performance of the recognition model and achieved 96.02% of character recognition accuracy.

Keywords: Amharic script · Deep convolutional neural network · Deep learning · Printed Amharic character · Pattern recognition · OCR · OCRopus · Visual Geometry Group

1 Introduction

Amharic is Ethiopian widely used language, which is spoken in most federal states of the country. Most histories of the societies and governments has been written and documented in this language. The sample script for Amharic language is depicted below (see Fig. 1). Nowadays, Ethiopian national archive and library agency tries to collect historical Amharic and Ge'ez documents so as to make them available for public. However, Amharic script scanned image documentation and retrieval process is time taking and also is not easy to distribute it for users. Changing this scanned image into full and editable document is an additional technical challenge. Hence, adoption of existing tools and algorithms in the area of OCR to account for Amharic scripts are paramount. An effective OCR model has been developed for other languages, including Latin and non-Latin scripts [6–10]. For Amharic OCR, many researchers attempted using different traditional and statistical machine learning techniques [1–4] which needs further experimentation with state-of-the-art techniques.

F. A. Zimale et al. (Eds.): ICAST 2018, LNICST 274, pp. 322–331, 2019.
https://doi.org/10.1007/978-3-030-15357-1_27

Optical Character Recognition (OCR) is a process that allows printed or hand-written text to be recognized optically and converted into machine-readable code that can be accepted by a computer for further processing and can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech, key data and text mining [1,2,8,12,16]. It is a systematic approach used for Amharic character recognition by segmenting input image into lines, characters and feature extraction. The extracted features used as an input for a classifier which causes recognition lattice [6,7] and also need more time and techniques to preprocess and extract feature from the image as well.

To optimize OCR and efficient text image extraction, we propose deep learning techniques. It is an advanced and dynamic tool for multi-font text image recognition. So as to minimize the number of steps and computational resources used for statistical feature extraction, a modified version of Visual Geometric Group (VGG) net is selected [7,8,19]. Deep learning takes GPU-days of computation to train on large scale data sets and it also used for character recognition [6,20], numerals recognition [22] and large-scale image recognition [23].

In VGG net, the convolutional layers enable to automatically extract the salient features which are invariant, a certain degree of shift and shape distortions of the input characters [6,17]. In addition, shared weight and the same filter are used for each input in the layer. Using sharing weight across CNN layers, reduces the number of parameter and improves performance [6]. On the other hand, different traditional machine learning algorithms applied, in recognition of text images, on a small private database of handwritten and printed character images encountered many challenges, which leads to a collection of multi-font character image and different level of degradation [1–3,6].

Furthermore, to make Amharic script documents available in the format of electronic text as opposed to page image, we develop an Amharic OCR system which is capable to recognize documents written in Amharic script with multiple fonts and different level of degradation. It is a systematic approach to develop a character image recognition model for printed real-life Amharic documents. In this research work, deep convolutional neural network called VGG net is applied to design the recognition model.

The goal of this paper is to use a synthetic training Amharic character dataset with a printed testing dataset and then to apply deep learning algorithms with different architectures rigorously so as to achieve very low error rates compared to other standard systems and methods. Therefore, in this paper, we apply a modified VGG net architecture using both synthetic and printed Amharic data sets and achieved state-of-the-art performance.

2 Related Works

Various methods have been proposed for character recognition and high recognition rates are reported for different Latin and non-Latin scripts such as the OCR of English [6,7,25], Fraktur [8], Arabic [9], Devanagari [11], Malayalam [12] and

Chinese [13–15]. For the last two decades [1–5] many researchers attempted to address problems in Amharic character recognition.

Dereje [1] proposed Binary Morphological filtering algorithm for OCR of Amharic typewritten text. He recommended that adopting recognition algorithms which are not very sensitive to the features of the writing styles of characters helps to enhance recognition rate of Amharic OCR.

Million [2], conducted research to investigate and extract the attributes of Amharic characters written in different fonts and then generalize previously adopted recognition algorithms. Using different test case, 49.38%, 26.04% and 15.75% recognition accuracy were registered for WashRa, Agafari and Visual Geez fonts respectively. In addition, Yaregal [3] noted that, font variation is also a problem in designing OCR system for printed Ethiopic documents.

Hailu et al. [24], attempted to develop a CNN based model so as to recognize Amharic character images with different level of degradation and achieved 92.71% testing accuracy. Wondowsen [4], attempted to develop OCR engine for special type of handwritten Amharic text, which is traditionally called "Yekum Tsehuf" using multilayer perceptron with backpropagation and also Nigussie [5], presented a Neural Network (NN) architecture for a recognition of handwritten Amharic characters of bank cheque.

On the other hand, Shatnawi and Abdallah [18] model a real distortion in Arabic script using real handwritten character examples to recognize characters. They used these distortion models to synthesize handwritten examples that are more realistic and achieved 73.4% recognition accuracy. Zhao et al. [19], proposed a deep learning method based on the convolutional neural networks to recognize scanned characters from real life with the characteristics of illumination variance, cluttered backgrounds, geometry distortion and achieved promising result.

In summary, the above studies noted that written style of characters including font variation, size and level of degradation are important factor for developing OCR model [1–3,18]. And also preprocessing, feature extraction and classification algorithms were the major steps for constructing an effective recognition model [2,3,19].

In addition, to best of the researchers' knowledge, not attempts has been made for Amharic character recognition using deep learning techniques due to scarcity of training dataset. Even though, many researchers have been attempted to recognize a small set of Amharic character image employing hand crafted features which is usually not robust and are computationally intensive.

Therefore, there is still room for improvement of the printed Amharic character recognition using state-of-the-art techniques that can reduce preprocessing steps and computational resource as well.

3 Materials and Methods

This study is concerned with the development of multi-font Printed Amharic character image recognition using deep learning techniques. For experimentation, two datasets; one synthetically generated dataset and the other is scanned image of printed Amharic document, were used.

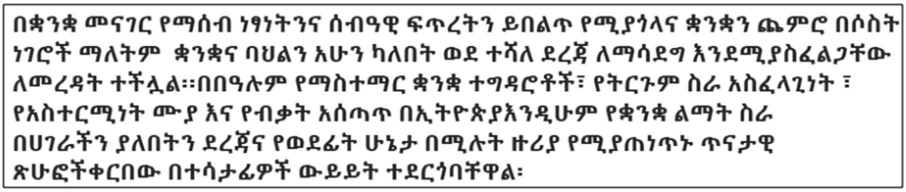

Fig. 1. Sample Amharic script.

All experiments were performed on a desktop computer with an 8 GB of NVIDIA GPU memory and LUNIX operating system. During experimentation, we have used 80,000 synthetically generated Amharic character images and 6715 scanned character images of printed Amharic documents. The synthetic dataset used in our experimentation were prepared by Hailu et al. [24]. And we prepared the printed character image data sets using Kyocera TASKalfa 5501i flatbed scanner at a resolution of 300 DPI. Preprocessing steps including binarization, skew detection and correction, noise reduction and character image segmentation has been done using the OCRopus [21]. Once we segmented each scanned Amharic document in to character images, we manually labeled a ground truth set of each character image in a text file. In addition, VGG net takes the character images as an input and then multiple convolutional layers have been adopted to extract automatic discriminating features.

3.1 Visual Geometry Group (VGG) Network

For recognition, we use a VGG net architecture. The advanced architecture of this network may consist stacks of sixteen and nineteen convolutional layers followed by max-pooling layers followed by fully connected and soft-max layer [22].

In this architecture, the three convolutional layers are used consecutively with a rectified linear unit (ReLU) as activation function formulated as follows:

$$f(X) = max(0, X) \tag{1}$$

Where X is the input. We did experiments with different architectures of convolutional neural network and the result reported, in this paper, employed a modified version of the VGG net architecture depicted as follows (see Fig. 2) and the pseudocode of the proposed model is presented in Algorithm 1 below.

In each block, the outputs of each convolutional layers pass through an activation function followed by single 2×2 pixel window called max-pooling layer and illustrated as below (see Fig. 3). And then two fully connected layers with 2048 neurons of each were used at the final layer. The 3×3 convolutional filters

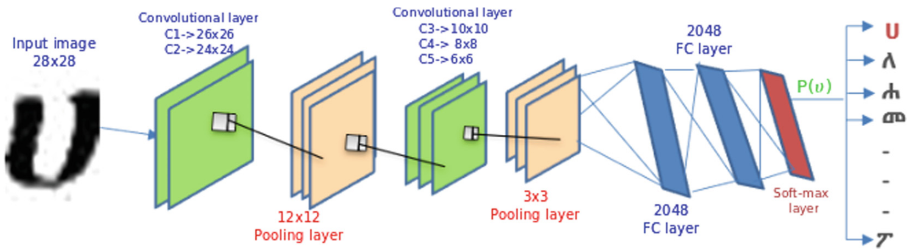

Fig. 2. The proposed VGG network architecture for Amharic character image recognition.

Algorithm 1. Pseudocode for training the proposed model

1: *Input*: Model, T, t // Model= Keras sequential model, T=training dataset, t=test dataset
2: *Output*:Trained model
3: *Start* : // adding convolutional and pooling layers (in this paper, in every
  //three consecutive convolutional layers we apply max-pooling
  Model.add(xConvy), // x= number of filter, y= kernel, Conv=convolution
  Model.add(xPy)//x= is pooling size, y= stride, P=pooling layer
  //here adding the convolutional and pooling layers as many as we want
  Model.add(xFC)//x=number of neurons, FC= fully connected layer
  Model.add(NC)//NC=number of class
  Model.add(softmax)
  Model.compile(Optimizer, accuracy measure)
  Model.fit(T, epochs, validation-split)
4: *Stop*:

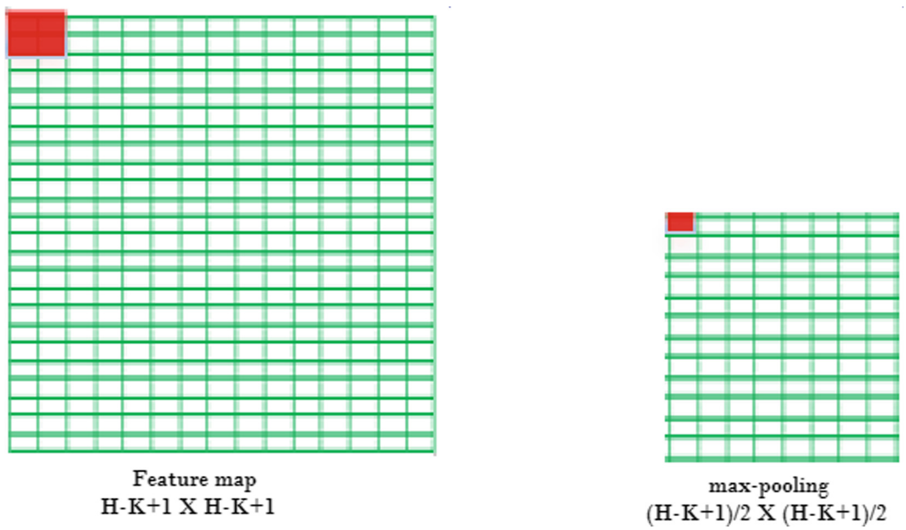

Fig. 3. Pooling process on the feature map that produce a new feature map.

with stride 1 was applied for performing filtering and sub-sampling operations [22]. The total number of parameters used during training the VGG net can be computed as:

$$P = L^2 \times (F^2 \times C^2) \tag{2}$$

Where P is number of parameters, L is the number of consecutive convolutional layers, F is number of filter and C is channel of the image. We considered an input size of 28×28 pixel Amharic character images. The output of each convolutional block is computed using:

$$O = (\frac{W - F + 2 \times P}{S}) + 1 \tag{3}$$

Where O is output of each convolutional block, W is input volume, F is kernel size, P is zero padding and S is number of strides. And then the output of each block is fed to the next network block until the probability of the last layer is calculated. The general process of convolutional layer is depicted in (Fig. 4).

The final output of the model is determined by a Soft-max function, which tells the probability that any of the classes are true, is defined as follows:

$$f(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{4}$$

Where z is the vector of the inputs for output layer and j is the indexes which runs from 1 to K and K is output label. To train the proposed network model, we used mini-batches size of 128 and 0.001 learning rate. A dropout rate of 0.25 was used to regularize the network parameters. The ADAM optimizer was used as the optimizing function whereas categorical cross entropy is used to calculate the loss and then found the parameter of connected layers that minimize the prediction loss.

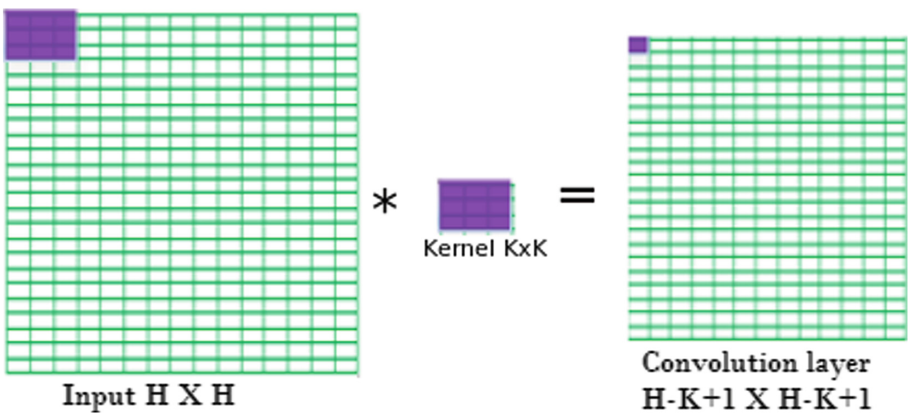

Fig. 4. Illustration of convolutional process.

4 Experimental Results

The implementation is based on deep learning framework using python programing on Keras Application Program Interface (API) along with TensorFlow backend. Experiments were run following the network architecture and system settings introduced in Sect. 3. In this paper, we only consider the 231 basic Amharic character sets and the results observed during experimentation are reported as follows (see Figs. 5 and 6).

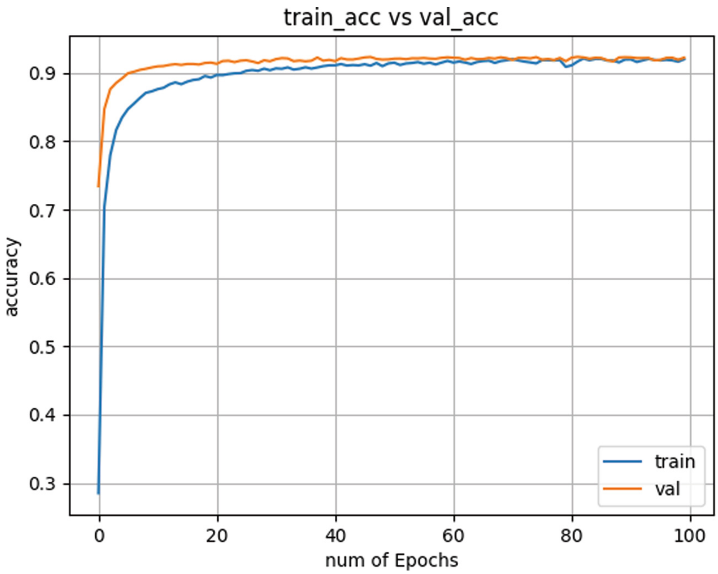

Fig. 5. Training versus validation accuracy for 100 epochs.

During experimentation, we train our model for different epochs, compared the results, and rerun our model on more epochs with selected parameters and the most promising validation accuracy was recorded at 100^{th} epoch using 70%, 10% and 20% training, validation and test of synthetic Amharic character image dataset respectively. Once the recognition model is developed, we achieved 92.71% of average recognition accuracy using the synthetic Amharic character images [24].

We also run our experiment using printed Amharic character images, as a test set, with the most popular fonts such as PowerGeez, VisualGeez, Agafari and WashRa (that are mostly used for typing purposes), and 3.98% of character recognition loss was recorded. An investigation of the experimental result indicated that some characters were incorrectly recognized due to similarity while the others were incorrectly recognized even with no similarity between them.

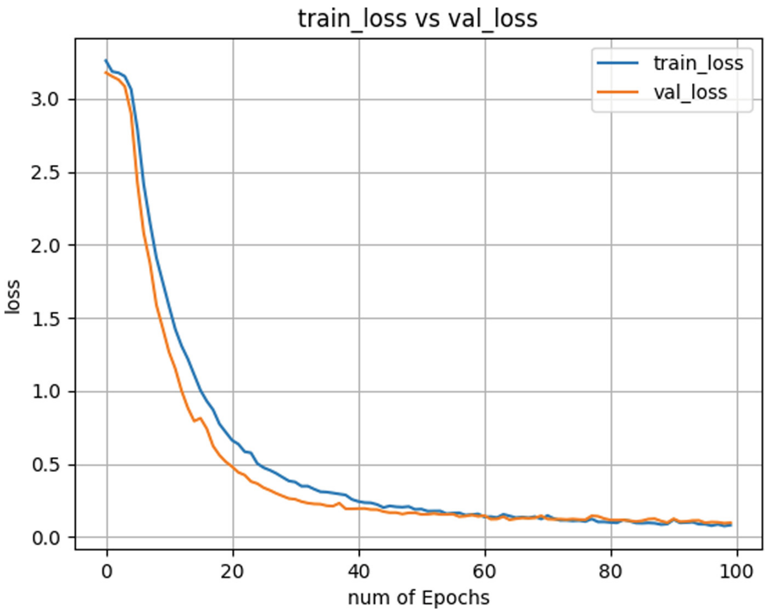

Fig. 6. Training versus validation loss for Amharic character images.

5 Performance Comparison

To the best of the researchers' knowledge, there is no any publicly available dataset for Amharic script recognition and all researchers have been done experiments by preparing their own dataset. Therefore, it is sometimes hard to compare, because previous work has not experimented with large database. However, in order to show the progress of Amharic character image recognition, the result of the current Amharic character recognition model, and the results of previous researches [1–3, 24] are shown in Table 1.

Table 1. Performance comparison of related works.

Researchers	Dataset size	Classifier	Accuracy (100%)
Dereje [1]	5172-character images	ANN	61%
Million [2]	7680-character images	SVM	90.37%
Yaregal [3]	1010-character images	ANN	73.18%
Hailu et al. [24]	80,000-character images	CNN	92.71%
Our method	86,715-character images	CNN	96.02%

6 Conclusion

We experimentally developed an Amharic character images recognition model using a modified version of VGG net that outperforms the state-of-the-art methods. The deep convolutional neural network tends to work better with raw input pixels rather than features or part of an image. During dataset preparation we consider only 231 basic Amharic character images. In our dataset each character exists 376 times on average. Once we developed the recognition model with the synthetically generated Amharic character images, we used a printed test set of Amharic character images (about 10% of the total dataset) and achieved an average recognition accuracy of 96.02%. As an extended research, the OCR technology towards the recognition of Amharic document image can be further referenced as word and sentence level, in parallel, large corpus of historical and hand-written document images will be prepared with benchmark experimental result. In addition, an extensive experiment will be done with a syntactic data so as to solve shortage of training dataset and recognize better for printed images.

References

1. Dereje T.: Optical Character Recognition of Typewritten Amharic Text. Master thesis, School of Information studies for Africa, Addis Ababa University, Addis Ababa, Ethiopia (1999)
2. Million, M.: A Generalized Approach to Optical Character Recognition of Amharic Texts. Master thesis, School of Information studies for Africa, Addis Ababa University, Addis Ababa, Ethiopia (2000)
3. Yaregal, A.: Optical Character Recognition of Amharic Text: An Integrated Approach. Master thesis, School of Information studies for Africa, Addis Ababa University, Addis Ababa, Ethiopia (2002)
4. Wondwossen, M.: Optical Character Recognition for Special Type of Handwritten Amharic Text ("Yekum Tsifet"): Neural Network Approach. M.Sc. thesis, School of Information Studies for Africa, Addis Ababa University, Addis Ababa, Ethiopia (2004)
5. Negussie, T.: Handwritten Amharic Text Recognition Applied to the Processing of Bank Cheques. Master thesis, School of Information studies for Africa, Addis Ababa University, Addis Ababa, Ethiopia (2000)
6. Bai, J., Chen, Z., Feng, B., Xu, B.: Image character recognition using deep convolutional neural network learned from different languages. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 2560–2564. IEEE (2014)
7. Yuan, A., Bai, G., Jiao, L., Liu, Y.: Offline handwritten English character recognition based on convolutional neural network. In: 2012 10th IAPR International Workshop on Document Analysis Systems (DAS), pp. 125–129. IEEE (2012)
8. Breuel, T.M., Ul-Hasan, A., Al-Azawi, M.A., Shafait, F.: High-performance OCR for printed English and Fraktur using LSTM networks. In: 2013 12th In ternational Conference on Document Analysis and Recognition (ICDAR), pp. 683–687. IEEE (2013)
9. ElAdel, A., Ejbali, R., Zaied, M., Amar, C.B.: Dyadic multi-resolution analysis-based deep learning for Arabic handwritten character classification. In: 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 807–812. IEEE (2015)

10. Shaw, B., Parui, S.K., Shridhar, M.: Offline handwritten deanagari word recognition: a holistic approach based on directional chain code feature and HMM. In: International Conference on Information Technology, ICIT 2008, pp. 203–208. IEEE (2008)
11. Ghosh, D., Dube, T., Shivaprasad, A.: Script recognition-a review. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2142–2161 (2010)
12. Anil, R., Manjusha, K., Kumar, S.S., Soman, K.P.: Convolutional neural networks for the recognition of Malayalam characters. In: Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 1041–1045 (2015)
13. Yang, W., Jin, L., Xie, Z., Feng, Z.: Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge. In: Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 551–555 (2015)
14. He, M., Zhang, S., Mao, H., Jin, L.: Recognition confidence analysis of handwritten Chinese character with CNN. In: Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 61–65 (2015)
15. Zhong, Z., Jin, L., Xie, Z.: High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps. In: Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 846–850 (2015)
16. Baird, H.S., Tombre, K.: The evolution of document image analysis. In: Doermann, D., Tombre, K. (eds.) Handbook of Document Image Processing and Recognition, pp. 63–71. Springer, London (2014). https://doi.org/10.1007/978-0-85729-859-1_43
17. Shin, H.C., et al.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging **35**, 1285–1298 (2016)
18. Shatnawi, M., Abdallah, S.: Improving handwritten Arabic character recognition by modeling human handwriting distortions. ACM Trans. Asian Low Resour. Lang. Inf. Process. **15**, 1–12 (2015)
19. Zhao, H., Hu, Y., Zhang, J.: Character recognition via a compact convolutional neural network. In: 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 29 November –1 December. IEEE (2017)
20. Lavin, A., Gray, S.: Fast algorithms for convolutional neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27–30 June 2016
21. Breuel, T.M.: The OCRopus open source OCR system. In: Document Recognition and Retrieval XV, vol. 6815, p. 68150F. International Society for Optics and Photonics (2008)
22. Ashiquzzaman, A., Tushar, A.K.: Handwritten Arabic numeral recognition using deep learning neural networks. IEEE (2017)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
24. Hailu, B., Amberbir, T., Stricker, D.: Amharic character image recognition. In: 18th IEEE International Conference on Communication Technology, 8–11 October 2018. (Accepted for Publication)
25. Tkachenko, I., Gomez, P.: Robustness of character recognition techniques to double printed and scanned documents. In: 14th IEEE International Conference on Document Analysis and Recognition (2017)