# A Blended Attention-CTC Network Architecture for Amharic Text-image Recognition

Birhanu Hailu Belay[1,3], Tewodros Habtegebrial[1], Marcus Liwicki[2], Gebeyehu Belay[3] and Didier Stricker[1,4]

[1]*Technical University of Kaiserslautern, Kaiserslautern, Germany*

[2]*Lulea University of Technology, Lulea, Sweden*

[3]*Bahir Dar Institute of Technology, Bahir Dar, Ethiopia*

[4]*DFKI, Augmented Vision Department, Kaiserslautern, Germany*

Keywords: Amharic Script, Blended Attention-CTC, BLSTM, CNN, Encoder-decoder, Network Architecture, OCR, Pattern Recognition.

Abstract: In this paper, we propose a blended Attention-Connectionist Temporal Classification (CTC) network architecture for a unique script, Amharic, text-image recognition. Amharic is an indigenous Ethiopic script that uses 34 consonant characters with their 7 vowel variants of each and 50 labialized characters which are derived, with a small change, from the 34 consonant characters. The change involves modifying the structure of these characters by adding a straight line, or shortening and/or elongating one of its main legs including the addition of small diacritics to the right, left, top or bottom of the character. Such a small change affects orthographic identities of character and results in shape similarly among characters which are interesting, but challenging task, for OCR research. Motivated with the recent success of attention mechanism on neural machine translation tasks, we propose an attention-based CTC approach which is designed by blending attention mechanism directly within the CTC network. The proposed model consists of an encoder module, attention module and transcription module in a unified framework. The efficacy of the proposed model on the Amharic language shows that attention mechanism allows learning powerful representations by integrating information from different time steps. Our method outperforms state-of-the-art methods and achieves 1.04% and 0.93% of the character error rate on ADOCR test datasets.

## 1 INTRODUCTION

Amharic is an official working language of the Federal Democratic Republic of Ethiopia and it is the second most widely spoken Semitic language in the world next to Arabic. Amharic is spoken by more than 100 million people in the country and it is also widely spoken in different countries like Eritrea, USA, Israel, Somalia and Djibouti (Meshesha and Jawahar, 2007; Amh, ; Mekuria and Mekuria, 2018).

In Amharic script, there are about 317 different alphabets including 238 core characters, 50 labialaize characters, 9 punctuation marks and 20 numerals which are written and read, like English, from left to right (Meshesha and Jawahar, 2007; Belay et al., 2019a; Belay et al., 2019b). There are multiple documents, containing religious and academic contents, written in Amharic script dated back from $12^{th}$ century (Meyer, 2006). Since then, these documents are stored in different places such as Ethiopian Orthodox Tewahdo Churches, public and academic libraries in the form of hardcover books. With a digitization campaign, many of these manuscripts are collected from different sources. However, they are still preserved in a manual catalog and/or scanned copies of them in Microfilm format (Wion, 2006).

The shape and structural formation of sample basic Amharic characters with their unique features are depicted in Figure 2.

Numerous works, in area of Optical Character Recognition (OCR) and Document Image Analysis (DIA), have been done and widely used for decades to digitize various historical and modern documents (Breuel et al., 2013; Maitra et al., 2015; Mondal et al., 2017; Martınek et al., 2020). Researchers achieved a high recognition accuracy and most scripts now have commercial off-the-shelf OCR applications. However, OCR often gives a better recognition result only



Belay, B., Habtegebrial, T., Liwicki, M., Belay, G. and Stricker, D.
A Blended Attention-CTC Network Architecture for Amharic Text-image Recognition.
DOI: 10.5220/0010284204350441
In *Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2021)*, pages 435-441
ISBN: 978-989-758-486-2

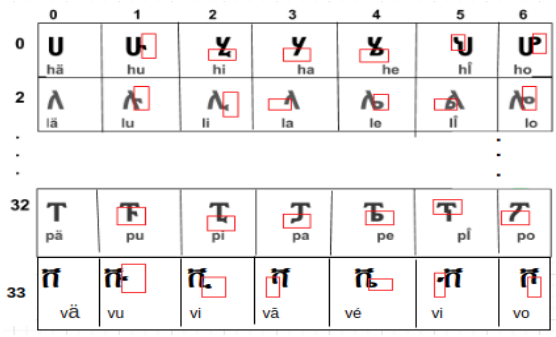


Figure 1: Shape formation of sample basic Amharic characters (Belay et al., 2020). Orders of consonant-vowel variants ($34 \times 7$). Characters in the first column are consonants and the others are derived variants. Vowels are derived by adding diacritics and/or remove part of consonants and the orthographic identities of each character vary across rows as marked with the red color.

for a specific use cases, moreover there are multiple indigenous scripts, like Amharic, which are underrepresented in the area of Natural Language Processing (NLP) and DIA (Belay et al., 2019b).

Even though OCR research for Amharic script started in 1997 (Alemu, 1997), it is still in its infancy and it is still an open area of research. Since then, attempts have been made to develop Amharic OCR (Meshesha and Jawahar, 2007; Alemu, 1997; Cowell and Hussain, 2003; Assabie and Bigun, 2009) using different statistical machine learning techniques. Recently, following the success of deep learning, other attempts are also made to develop a model for Amharic OCR and achieved relatively promising results (Belay et al., 2019a; Belay et al., 2019b; Belay et al., 2018; Reta et al., 2018; Gondere et al., 2019).

In literature, attempts to Amharic OCR neither shown results on large dataset nor considering all possible characters used in Amharic writing system. Recently publish work (Belay et al., 2019b), introduced an Amharic OCR database called ADOCR. We took a sample text-line image from ADOCR database whose word formation and character arrangements in a sample word are illustrated in Figure 2.

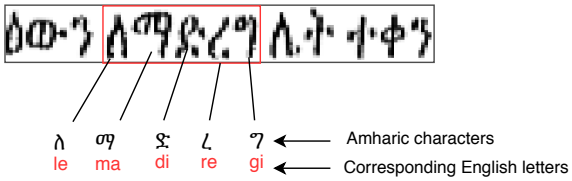


Figure 2: Sample Amharic text-line image from the dataset. A word marked by red box is composed of five individual Amharic characters and the corresponding sounds of each character is described with English letters using red color.

Convolutional and Recurrent networks have been used in Amharic OCR. In this paper we aim to push the limits of Amharic OCR models by introducing attention mechanism. Therefore, in this paper, we propose an attention based CTC network called Blended Attention CTC (BACTC) for Amharic text-line image recognition. BACTC has a CNN-LSTM layers as encoder followed by attention and CTC layers which used to pick the only important features of encoded inputs and transcription respectively.

The rest of the paper is organized as follows. Section 2 talks about related works. Our proposed model is presented in section 3, and section 4 presents the detail of datasets. In the last two sections, experimental results and conclusions are presented respectively.

## 2 RELATED WORK

Previous research on Amharic OCR was focused on the statistical machine learning techniques. Most of these techniques are segmentation based and character level OCR models (Meshesha and Jawahar, 2007; Belay et al., 2019a; Cowell and Hussain, 2003; Belay et al., 2018). The only exception work were Assabie (Assabie and Bigun, 2009) who proposed a segmentation free OCR based on HMM model for offline handwritten Amharic word recognition. Recently published works (Addis et al., 2018) and (Belay et al., 2019b) proposed a Bidirectional LSTM (BLSTM) network architecture with CTC for Amharic text-line image recognition. An end-to-end learning, that uses CNN, LSTM and CTC in a unified framework (Belay et al., 2020), is also proposed for Amharic OCR and achieved a better recognition performance.

Following the first Amharic OCR research, which was only able to recognize a character written with Washera font and 12 point type, attempted by Worku in 1997 (Alemu, 1997), other research works have been made including typewritten (Teferi, 1999), machine printed (Meshesha and Jawahar, 2007; Belay et al., 2018), Amharic document image recognition and retrieval (Meshesha, 2008), Ethiopic number (Reta et al., 2018) and handwritten (Gondere et al., 2019) recognition.

Based on recurrent neural network, several OCR techniques have been studied and demonstrated groundbreaking performances for multiple Latin and Non-Latin scripts. BLSTM with CTC for Amharic text-image recognition (Belay et al., 2019b), Convolutional Recurrent Neural Network (CRNN) for Japanese handwritten recognition (Ly et al., 2017), segmentation free Chinese handwritten text recognition (Messina and Louradour, 2015), a

hybrid Convolutional-LSTM for text-image recognition (Breuel, 2017), Multidimensional LSTM for Chinese handwritten recognition (Wu et al., 2017), combined Connectionist Temporal Classification (CTC) with Bidirectional LSTM for unconstrained online handwriting recognition (Graves et al., 2008).

Attention based networks have been intensively applied in the area of NLP tasks and came up with successive results in neural machine translation (Bahdanau et al., 2014; Luong et al., 2015; Ghader and Monz, 2017) and speech recognition (Das et al., 2019; Watanabe et al., 2017). The most works so far with attention mechanism has focused on neural machine translation. However, researchers have recently applied attention in different research areas. Therefore, it becomes popular and a choice of many researchers in the area of OCR.

Attention mechanism is now also widely applied for recognizing handwritten texts (Poulos and Valle, 2017; Chowdhury and Vig, 2018), characters in the wild (Lee and Osindero, 2016; Huang et al., 2019; Huang et al., 2016), and handwritten mathematical expression (Zhang et al., 2018; Li et al., 2020). Attention mechanism assists the network in learning the correct alignment between the input image pixels and the target characters. In addition, it improves the ability of the network in extracting the most relevant feature for each part of the output sequence. Inspired by the success of attention in sequence to sequence translation, we continue to focus on OCR tasks, and we integrate the capability of attention mechanism in the CTC network so as utilize the benefits from both techniques.

## 3 OVERVIEW OF THE PROPOSED MODEL

The proposed blended attention-CTC model, as shown in Figure 3, consists of three modules. The encoder module takes the input features $x$ and maps them to a higher-level feature representation $h^{enc} = (h_1{}^{enc}, h_2{}^{enc}, ..., h_T{}^{enc})$. The attention module takes the output features $h^{enc}$ of the encoder module and computes the context vector from each hidden features. The output of the attention module, attention context information, is passed to the Soft-max layer in order to produce a probability distribution, $P(y_0, y_1, ..., y_{t-1}|x)$ over the given input sequence $x$. Finally, the probability distribution $P$ goes to the CTC-decoder for transcription.

The proposed method integrates attention mechanism, LSTM and CTC layers. The intuition for the use of the attention layer is to infer a more powerful hidden representation through a weighed a context vector $(C_{\text{vec}})$. The attention based weighting offers a powerful way to aggregate inputs from different time steps. The weighted context vector is computed using Equation (1). The training objective of the proposed model follows the same CTC training objective explained in (Graves et al., 2008).

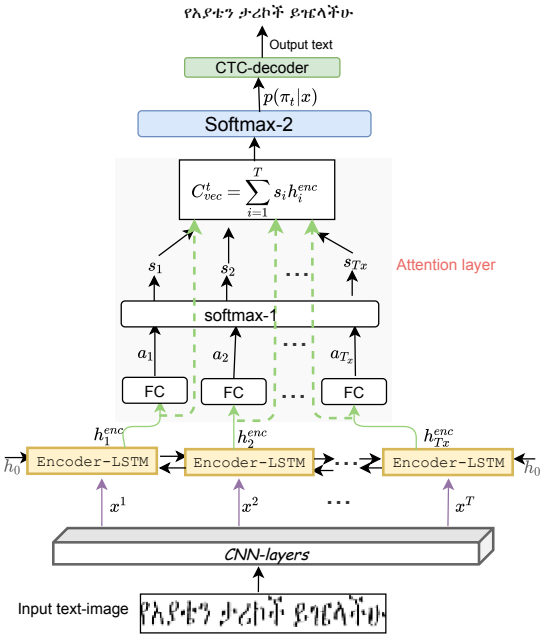


Figure 3: The proposed blended attention-CTC model. Alignment score $(a_i)$ of encoder hidden state $(h_i^{enc})$ at each time-step is computed using a scoring function described in (3), just by propagating the $h_i^{enc}$ through fully connected network (FC). Attention distribution, called attention weights $(s_i)$ are computed by running all alignment scores over a soft-max (*Soft-max-1)* layer. To compute the alignment vector , we multiply each $h_i^{enc}$ with its corresponding soft-maxed score $(s_i)$. Then, the sum of all alignment vectors produced the context vector $(C_{\text{vec}})$, which is an aggregated information. Once the $C_{\text{vec}}{}^t$ is obtained, it passes through the second soft-max layer *(Soft-max-2)* for probability distribution over the n possible characters in the ground-truth (GT). The output of *Soft-max-2* is a sequence of T time steps of $(n+1)$ characters which is then decoded using CTC-decoder.

$$C^t_{\text{vec}} = \sum_{i=1}^{T} s_{\text{ti}} h_{\text{i}}{}^{enc} \tag{1}$$

where the $s_t$ is an attention weight of each annotation $h_{\text{i}}{}^{enc}$ computed by soft-maxing its corresponding attention score using Equation (2),

$$s_{\text{i}} = \frac{exp(a_{\text{i}})}{\sum_{k=1}^{T} exp(a_{\text{k}})} \tag{2}$$

where $a_{\text{i}}$ is the alignment score of $h_i^{enc}$ at each time

step $t$ and it can be computed using Equation (3).

$$a_{\mathrm{i}} = f(h_{\mathrm{i}}^{enc}), \text{for } i = 1, ..., T_x \tag{3}$$

The function $f$ in Equation (3) is a feed-forward neural network with tanh function. The intuition of this scoring function is to let the model to learn the alignment weights together with the translation while training the whole model layers.

During training the blended attention-CTC model, a CTC loss function ($l_{\mathrm{CTC}}$) is used to train the network from end-to-end. For training data D, the function $l_{\mathrm{CTC}}$ can be defined as in Equation (4).

$$l_{\mathrm{CTC}} = -log\left(\prod_{(x,z)\in D} p(z|x)\right) \tag{4}$$

where $x = (x_1, x_2, ...., x_T)$ is the input sequence with length T, and $z = (z_1, z_2, ..., z_C)$ is the corresponding output sequence in ground-truth for $C < T$ in every pair of $x$ and $z$. The $P(z/x)$ is computed by multiplying the probability of labels along the path $\pi$ that contains output label over all time steps $t$ as shown in Equation (5).

$$p(\pi|x) = \prod_{t}^{T} p(\pi_t, t|x) \tag{5}$$

where t is the time step and $\pi_t$ is the label of path $\pi$ at t.

A target label in path $\pi$ is obtained by mapping reduction function B, using the example explained in (Belay et al., 2019b), that convert a sequence of Softmax output for each frame to a label sequence by removing repeated labels and blank tokens from the sequences of character (C) with the highest score (i) generated using Equation (6).

# 4 DATASET

To the best of our knowledge, ADOCR (Belay et al., 2019b) database is the only publicly available Amharic OCR dataset with the benchmark experimental results. Therefore, to train and evaluate the proposed model, we use the ADOCR database introduced by (Belay et al., 2019b) and which is freely available at http://www.dfki.uni-kl.de/~belay/.

The original Amharic OCR database composed of 337,337 Amharic text-line images, each with multiple word instance collected from different sources. In this database there are about 40,929 printed text-line images written with power Geez font 197,484 and 98,924 text-line images synthetically generated with Power Geez and Visual Geez fonts respectively.

In addition, the ADOCR database contains 280 unique Amharic characters and punctuation marks which are mutually exist in both the training and test samples. All images a 48 by 128 pixels gray-scale, and the maximum string length of the ground-truth text is 32 characters.

Then the images are normalized into 32 by 128 pixels as stated in (Belay et al., 2020). and sample Amharic text-line images from ADOCR database are given in Figure 4.


Figure 4: Sample Amharic text-line images from ADOCR database: **a**) Printed text-line images written by power Geez font type. (**b**) Synthetic text-line images generated with Power Geez font type. (**c**) Synthetic text-line images generated with the Visual Geez font.

# 5 EXPERIMENTS

We train and test our model on the ADOCR database which contains 318,706 training and 18,631 test samples. Training details and experimental results are presented below.

## 5.1 Training

Our model is trained with the ADOCR database. Similar to (Belay et al., 2020), images are scaled to 32 by 128 pixels so as to minimize computations. Since there is no explicitly stated validation data, in ADOCR dataset, we randomly selected 7% of the training samples for validation samples. In this study, we first implement and train an attention based encoder-decoder network proposed by Bahdanau (Bahdanau et al., 2014) and then the blended attention-CTC network is formulated. In the later model, the main contribution of this paper, we directly taking the advantage of attention mechanism and CTC network as an integrated framework and trained in an end-to-end fashion.

When training both models, we use two bi-directional LSTM, each with 128 hidden units and

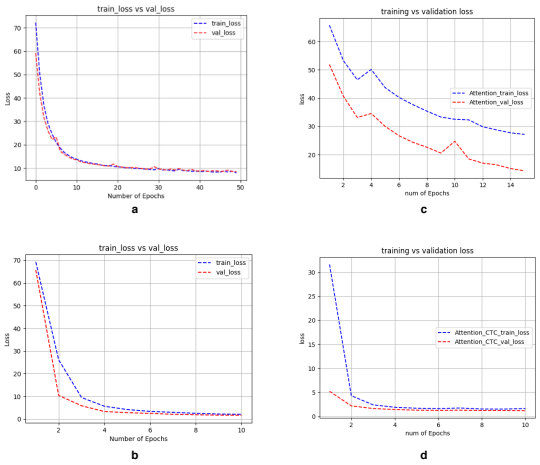


Figure 5: Training & validation losses of model training with different network settings: (**a**). CTC loss with an LSTM-CTC model (Belay et al., 2019b). (**b**) CTC loss with a CNN-LSTM-CTC model (Belay et al., 2020). (**c**) CE loss of attention based encoder-decoder model. (**d**) CTC loss of the proposed blended attention-CTC model.

dropout rate of 0.25, on top of seven convolutional layers that are stacked serially with ReLU activation function as an encoder. In the attention based encoder-decoder model, the decoder has a unidirectional LSTM with 128 hidden units while in the blended attention-CTC approach, the decoder LSTM is removed from the previous model and then the CTC objective function that are blended with attention mechanism is in place.

The convolutional layers of the encoder module is composed of seven convolutional layers which have a kernel size of $3 \times 3$, except that the one on top is with a $2 \times 2$ kernel size, four max-pooling layers with pool sizes of 2 for the first pooling layer, and $2 \times 1$ for the remaining pooling layers. Strides are fixed to one, and the 'same' padding is used in all convolutional layers. The number of feature maps are 64, 128, 256, 256, 512, 512, 512 from bottom to top layers.

We use a batch size of 128 with Adam optimizer and each model, Attention based Encoder-Decoder (AED) and BACTC, is trained for 10 and 15 epochs respectively. The attention based encoder-decoder model minimizes a categorical-cross entropy (CE) loss while the blended attention-CTC model tried to minimize the CTC-based loss described in Equation (4). Once the probability of labels obtained from the trained models, we use best path decoding (Graves, 2008) to generate a character $(C_i)$ that has the maximum score at each time step $t$.

$$C_i = \arg\max_i (y_i^t | x), \text{for } t = 1, 2, ..., T \qquad (6)$$

We implement both model with Keras Application Program Interface (API) on a TensorFlow backend. The learning loss of the proposed model and other models trained using different network settings are depicted in Figure 5.

## 5.2 Results

The performance of the proposed blended-attention model is evaluated against three test datasets, and compared it with state-of-the-art approaches. Table 1, presents the details of experimental results with Character Error Rate (CER) and the proposed model improves the recognition performance by 0.67–7.50% from the original paper (Belay et al., 2019b) and by 0.16–0.52% from the recent published paper which use the same test dataset (Belay et al., 2020). CER is computed, using Equation (7), by counting the number of characters inserted, substituted, and deleted in each sequence and then dividing by the total number of characters in the ground truth (Belay et al., 2019b).

$$CER(P,T) = \left( \frac{1}{q} \sum_{n \in P, m \in T} D(n,m) \right) \times 100, \qquad (7)$$

where $q$ is the total number of target character labels in the ground truth, $P$ and T are the predicted and ground-truth labels, and $D(n,m)$ is the edit distance between sequences $n$ and $m$.

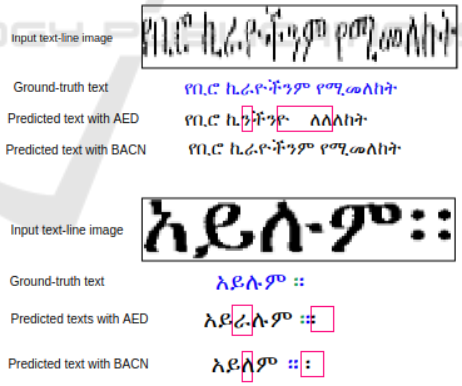


Figure 6: Sample text-line image with the corresponding predicted and ground-truth texts. Characters, in the predicted text, marked with red rectangle are wrongly predicted by both Attention based Encoder-decoder(AED) and Blended Attention-CTC( BACN).

We also implemented the attention based encoder-decoder model, without the CTC network, and the performance of this model is evaluated with the three ADOCR test datasets. The blended attention-CTC model outperforms the attention based encoder decoder model by 24.09%.

Table 1: Comparison of test results (CER).

| | #test-set | image-type | Font-type | CER (%) |
|---|---|---|---|---|
| Addis (Addis et al., 2018) * | 12 pages | printed | - | 2.12% |
| Belay (Belay et al., 2019b) | 2,907 | Printed | Power Geez | 8.54% |
| Belay (Belay et al., 2019b) | 9,245 | Synthetic | Power Geez | 4.24% |
| Belay (Belay et al., 2019b) | 6,479 | Synthetic | Visual Geez | 2.28% |
| Belay (Belay et al., 2020) | 2,907 | Printed | Power Geez | 1.56% |
| Belay (Belay et al., 2020) | 9,245 | Synthetic | Power Geez | 3.73% |
| Belay (Belay et al., 2020) | 6,479 | Synthetic | Visual Geez | 1.05% |
| Ours | 2,907 | Printed | Power Geez | **1.04**% |
| Ours | 9,245 | Synthetic | Power Geez | **3.57**% |
| Ours | 6,479 | Synthetic | Visual Geez | **0.93**% |

* Denotes methods tested on different datasets.

As we observed the empirical results, the attention based encoder-decoder model implemented without CTC, becomes poor when the sequence length increases. In most cases, the first 4 to 6 characters are always correctly predicted while the rest errors have no any patterns. Such character errors are not observed in the blended attention-CTC model. In summary, the proposed blended attention-CTC model outperforms all the state-of-the-art models on the ADOCR test datasets.

## 6 CONCLUSION

In this paper, we have introduced a blended attention-CTC network called BACTC for Amharic text-line image recognition. The proposed method consists of a Bidirectional LSTM, stacked on top of CNN layers, as an encoder and a CTC layer as a decoder. To enhance the hidden layer feature representation, the attention mechanism is embedded between the LSTM and CTC network layers without changing the CTC objective function and the training process. All the encoder, attention and CTC modules are trained jointly from end-to-end.

We evaluated our model with both synthetically generated and printed Amharic text-line images and a significant improvement is achieved on all the three ADOCR test datasets compared with state-of-the-art model results. Thus, we can conclude that the blended attention-CTC network is more effective for Amharic text image recognition than widely used attention-based encoder-decoder and CNN-LSTM-CTC based networks as well. This work can be potentially extended and applied for handwritten Amharic text recognition.

## REFERENCES


Will Amharic be AU's lingua franca? https://www.press.et/english/?p=2654#. Accessed: 2020-01-14.

Addis, D., Liu, C.-M., and Ta, V.-D. (2018). Printed ethiopic script recognition by using lstm networks. In *2018 International Conference on System Science and Engineering (ICSSE)*, pages 1–6. IEEE.

Alemu, W. (1997). The application of ocr techniques to the amharic script. *An MSc thesis at Addis Ababa University Faculty of Informatics*.

Assabie, Y. and Bigun, J. (2009). Hmm-based handwritten amharic word recognition with feature concatenation. In *2009 10th International Conference on Document Analysis and Recognition*, pages 961–965. IEEE.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Belay, B., Habtegebrial, T., Liwicki, M., Belay, G., and Stricker, D. (2019a). Factored convolutional neural network for amharic character image recognition. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2906–2910. IEEE.

Belay, B., Habtegebrial, T., Liwicki, M., Gebeyehu, and Stricker, D. (2019b). Amharic text image recognition: Database, algorithm, and analysis. In *2019 The 15th International Conference on Document Analysis and Recognition (ICDAR 2019)*. IEEE.

Belay, B., Habtegebrial, T., Meshesha, M., Liwicki, M., Belay, G., and Stricker, D. (2020). Amharic ocr: An end-to-end learning. *Applied Sciences*, 10(3):1117.

Belay, B., Habtegebrial, T., and Stricker, D. (2018). Amharic character image recognition. In *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, pages 1179–1182. IEEE.

Breuel, T. M. (2017). High performance text recognition using a hybrid convolutional-lstm implementation. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 11–16. IEEE.

Breuel, T. M., Ul-Hasan, A., Al-Azawi, M. A., and Shafait, F. (2013). High-performance ocr for printed english

and fraktur using lstm networks. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 683–687. IEEE.

Chowdhury, A. and Vig, L. (2018). An efficient end-to-end neural model for handwritten text recognition. *arXiv preprint arXiv:1807.07965*.

Cowell, J. and Hussain, F. (2003). Amharic character recognition using a fast signature based algorithm. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, pages 384–389. IEEE.

Das, A., Li, J., Ye, G., Zhao, R., and Gong, Y. (2019). Advancing acoustic-to-word ctc model with attention and mixed-units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(12):1880–1892.

Ghader, H. and Monz, C. (2017). What does attention in neural machine translation pay attention to? *arXiv preprint arXiv:1710.03348*.

Gondere, M. S., Schmidt-Thieme, L., Boltena, A. S., and Jomaa, H. S. (2019). Handwritten amharic character recognition using a convolutional neural network. *arXiv preprint arXiv:1909.12943*.

Graves, A. (2008). Supervised sequence labelling with recurrent neural networks [ph. d. dissertation]. *Technical University of Munich, Germany*.

Graves, A., Liwicki, M., Bunke, H., Schmidhuber, J., and Fernández, S. (2008). Unconstrained on-line handwriting recognition with recurrent neural networks. In *Advances in neural information processing systems*, pages 577–584.

Huang, W., He, D., Yang, X., Zhou, Z., Kifer, D., and Giles, C. L. (2016). Detecting arbitrary oriented text in the wild with a visual attention model. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 551–555.

Huang, Y., Luo, C., Jin, L., Lin, Q., and Zhou, W. (2019). Attention after attention: Reading text in the wild with cross attention. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 274–280. IEEE.

Lee, C.-Y. and Osindero, S. (2016). Recursive recurrent nets with attention modeling for ocr in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2231–2239.

Li, Z., Jin, L., Lai, S., and Zhu, Y. (2020). Improving attention-based handwritten mathematical expression recognition with scale augmentation and drop attention. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 175–180. IEEE.

Luong, M.-T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.

Ly, N.-T., Nguyen, C.-T., Nguyen, K.-C., and Nakagawa, M. (2017). Deep convolutional recurrent network for segmentation-free offline handwritten japanese text recognition. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 7, pages 5–9. IEEE.

Maitra, D. S., Bhattacharya, U., and Parui, S. K. (2015). Cnn based common approach to handwritten character recognition of multiple scripts. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, pages 1021–1025. IEEE.

Martınek, J., Lenc, L., and Král, P. (2020). Building an efficient ocr system for historical documents with little training data.

Mekuria, G. T. and Mekuria, G. T. (2018). Amharic text document summarization using parser. *International Journal of Pure and Applied Mathematics*, 118(24).

Meshesha, M. (2008). *Recognition and retrieval from document image collections*. PhD thesis, IIIT Hyderabad, India.

Meshesha, M. and Jawahar, C. (2007). Optical character recognition of amharic documents. *African Journal of Information & Communication Technology*, 3(2).

Messina, R. and Louradour, J. (2015). Segmentation-free handwritten chinese text recognition with lstm-rnn. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 171–175. IEEE.

Meyer, R. (2006). Amharic as lingua franca in ethiopia. *Lissan: Journal of African Languages and Linguistics*, 20(1/2):117–132.

Mondal, M., Mondal, P., Saha, N., and Chattopadhyay, P. (2017). Automatic number plate recognition using cnn based self synthesized feature learning. In *Calcutta Conference (CALCON), 2017 IEEE*, pages 378–381. IEEE.

Poulos, J. and Valle, R. (2017). Character-based handwritten text transcription with attention networks. *arXiv preprint arXiv:1712.04046*.

Reta, B. Y., Rana, D., and Bhalerao, G. V. (2018). Amharic handwritten character recognition using combined features and support vector machine. In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 265–270. IEEE.

Teferi, D. (1999). Optical character recognition of typewritten amharic text. Master's thesis, School of Information studies for Africa, Addis Ababa.

Watanabe, S., Hori, T., Kim, S., Hershey, J. R., and Hayashi, T. (2017). Hybrid ctc/attention architecture for end-to-end speech recognition. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1240–1253.

Wion, A. (2006). The national archives and library of ethiopia: six years of ethio-french cooperation (2001-2006).

Wu, Y.-C., Yin, F., Chen, Z., and Liu, C.-L. (2017). Handwritten chinese text recognition using separable multi-dimensional recurrent neural network. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 79–84. IEEE.

Zhang, J., Du, J., and Dai, L. (2018). Track, attend, and parse (tap): An end-to-end framework for online handwritten mathematical expression recognition. *IEEE Transactions on Multimedia*, 21(1):221–233.