# Low Resolution Arabic Recognition with Multidimensional Recurrent Neural Networks

Sheikh Faisal Rashid
Image Understanding and
Pattern Recognition (IUPR)
Technical University
Kaiserslautern, Germany
rashid@iupr.com

Marc-Peter Schambach
Siemens AG
Bücklestraße 1-5, 78464
Konstanz, Germany
marc-peter.schambach
@siemens.com

Jörg Rottland
Siemens AG
Bücklestraße 1-5, 78464
Konstanz, Germany
joerg.rottland@siemens.com

Stephan von der Nüll
Siemens AG
Bücklestraße 1-5, 78464
Konstanz, Germany
stephan.vondernuell
@siemens.com

## ABSTRACT

OCR of multi-font Arabic text is difficult due to large variations in character shapes from one font to another. It becomes even more challenging if the text is rendered at very low resolution. This paper describes a multi-font, low resolution, and open vocabulary OCR system based on a multidimensional recurrent neural network architecture. For this work, we have developed various systems, trained for single-font/single-size, single-font/multi-size, and multi-font/multi-size data of the well known Arabic printed text image database (APTI). The evaluation tasks from the second Arabic text recognition competition, organized in conjunction with ICDAR 2013[1], have been adopted. Ten Arabic fonts in six font size categories are used for evaluation. Results show that the proposed method performs very well on the task of printed Arabic text recognition even for very low resolution and small font size images. Overall, the system yields above 99% recognition accuracy at character and word level for most of the printed Arabic fonts.

## Keywords

Arabic Text Recognition, Recurrent Neural Networks, APTI Database

## 1. INTRODUCTION

Arabic recognition has been a focus of research from the last three decades in pattern recognition community. A lot of effort has been done to recognize printed and handwritten Arabic text due to various potential applications like mail sorting, electronic processing of bank checks, official forms and digital archiving of different printed documents.

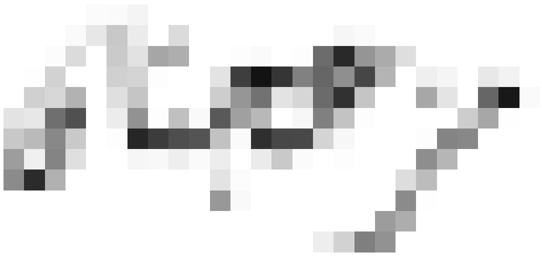


**Figure 1: Example of low resolution Arabic word image. Font: *Diwani Letter* at 6 point. The figure is rescaled for better visual appearance.**

Arabic is a widely used writing system in the world and is used for writing many languages like Arabic, Urdu, Pashto, Malay, Kurdish and Persian. Numerous amount of Arabic documents are available in the form of books, newspapers, magazines or technical reports. Arabic language appears in a variety of different fonts and font sizes in printed documents. Having a robust optical character recognition (OCR) strategy for multi-font Arabic text helps to provide digital access to this material.

It is interesting to provide an OCR system for very low resolution text [7]. Low resolution text usually appears in screen shots, video clips and text rendered at computer screens. OCR of screen rendered low resolution text is difficult due to anti-aliasing at small font sizes. The rendering process smoothes the text for better visual perception, and the smoothed characters pose difficulties for segmentation based text recognition systems. Figure 1 shows an Arabic word image sample, rendered in very small font size (6 point) at low resolution (72 dpi).

This work reports a segmentation free OCR method for low resolution digitally represented multi-font Arabic text. The method uses multidimensional recurrent neural net-

[1] diuf.unifr.ch/diva/APTI/competitionICDAR2013.html

*MOCR '13,* August 24 2013, Washington, DC, USA
Copyright 2013 ACM 978-1-4503-2114-3/13/08 ...$15.00.

works (MDRNNs), multi-directional long short term memory (MD-LSTM) and connectionist temporal classification (CTC) as described by Alex Graves in [1, 2]. However, in this work we apply different network topologies, height normalizations and, in addition, classifier selection for the development of different systems suited to the evaluation tasks. The recurrent neural network (RNN) models are trained and tested on a subset of Arabic printed text image (APTI) database [8]. Overall, we achieve more than 99% character and word level recognition accuracies for the evaluation tasks given bellow in section 1.2.

## 1.1 Arabic Printed Text Image Database

The Arabic printed text image (APTI) database has been built at the Document, Image and Voice Analysis (DIVA) research group, Department of Informatics, University of Fribourg, Switzerland [8]. The objective of this database is to provide a benchmarking corpus for evaluation of open-vocabulary, multi-font, multi-size and multi-style Arabic recognition systems. The database provides challenges in font, font size and font style variations, as well as rendering at very low resolution. The low resolution rendering adds more difficulty due to presence of noise and very small font sizes. The word images are rendered in ten fonts at different font sizes. They are equally distributed to six sets, with fair distribution of characters in each set. More details about the database can be found in [8].

Arabic script has 28 basic character shapes in isolated form. However, these characters are usually connected to each other and have different appearances due to their positions in a ligature or word. These character shape variations increase the number of potential classes and, in APTI database, 120 character classes for recognition are distinguished.

## 1.2 Evaluation Tasks

We use the evaluation tasks as proposed in the second Arabic printed text recognition context organized in conjunction with ICDAR 2013. The tasks are summarized below:

- Task 0: Six single-font/single-size systems for *Arabic Transparent* font.
- Task 1: One single-font/multi-size system for *Arabic Transparent* font.
- Task 2: One single-font/multi-size system for *DecoType Naskh* font.
- Task 3: One multi-font/multi-size system for *Andalus, Arabic Transparent, Advertising Bold, Diwani Letter, DecoType Thuluth, Simplified Arabic, Tahoma, Traditional Arabic, DecoType Naskh* and *M Unicode Sara* fonts.

We work with six font sizes 6, 8, 10, 12, 18, and 24 in "plain" font style for all the above tasks. In addition to these evaluation tasks, we also evaluate the proposed method on font size 6 for *Andalus*, *Simplified Arabic*, *Traditional Arabic* and *Diwani Letter* fonts.

Single-font/single-size systems are trained by using data from one particular font in one font size, single-font/multi-size systems were trained by using data from one particular font in multiple font sizes, and the multi-font multi-size system is trained by using data from multiple fonts in multiple font sizes.

# 2. PROPOSED METHOD

The presented method is based on a multi-dimensional recurrent neural network, long-short term memory cells and a connectionist temporal classification (CTC) output layer. The method utilizes a hierarchical network architecture to extract suitable features at each level of network hierarchy. A similar kind of approach has been successfully applied in recognition of off-line Arabic handwritten text by Graves and Schmidhuber [2]. However, in this work we use a slightly different network topology best suited for our problem and, in addition, a classifier selection procedure for the selection of the appropriate recognition model for final recognition. We train different RNN models depending on the evaluation tasks as explained above. The method consists of preprocessing, hierarchical features extraction, RNN training, and recognition steps.

## 2.1 Preprocessing

Preprocessing is one of the basic steps in almost every pattern recognition system. It is applied to transform the input data to a uniform format that is suitable for the extraction of discriminative features. Preprocessing normally includes noise removal, background extraction, binarization or image enhancement etc. However, here we are working on very clean, artificially generated word images and we do not need any sophisticated noise removal or image enhancement techniques. We directly operate on gray scale images and just normalize the gray values to mean 0 and standard deviation 1.

Additionally, image height normalization is performed in case of multi-size recognition tasks: We rescale every image to a specific height by adjusting the image width accordingly. This is necessary because, in each font size, images have variable pixel heights, and it is difficult to build a multi-size system without height normalization. The normalized image heights are selected based on statistical analysis of all image heights present in different fonts and font sizes. For training, images are normalized by using Mitchell [4] interpolation.

## 2.2 Feature extraction

Selection of appropriate features is a difficult task. In most pattern recognition applications, features are manually described and vary from application to application. In this work, we adopt a hierarchical network topology to extract suitable features from raw data. This kind of hierarchical structures have been applied in different pattern recognition and computer vision applications [3, 5, 6]. Here, we adopt the hierarchical network structure as used in [1]. The network hierarchy is built by composing the MD-LSTM layers with feed-forward layers in repeated manner. First, the image is decomposed into small pixel blocks (illustrated in section 2.3.2), and MD-LSTM layers scan across the blocks in all four directions. The activations of the MD-LSTM layers are collected into the blocks which are later passed to the next feed-forward layer. We use three MD-LSTM layers with 2, 10 and 50 LSTM cells, two feed forward layers with 6 and 20 feed forward units, followed by first and second MD-LSTM layers respectively. The output of the last MD-LSTM layer is passed to the CTC output layer containing 120 output units.

## 2.3 Recurrent Neural Network Training

In this work, we train various recurrent neural network models in reference to specific Arabic recognition tasks. The details of data used for training and network parameters are explained in following sections.

### 2.3.1 Data

We use sets 1–4 of the APTI database for training and set number 5 for final evaluation of the systems. The single-font/single-size systems are trained by using 75, 550 Arabic word images for each font and font size by taking all the available data from the first four sets. We randomly select 10, 000 word images from training data for validation. The validation set is used as a stopping criteria during network training only.

The single-font/multi-size systems are trained on the same 75, 550 images, but this time we use only 17% from each font size. Similarly, we randomly select 10, 000 word images for validation.

The multi-font/multi-size system is trained with 453, 300 randomly selected word images from all ten fonts and six font sizes. This is 10% of the total data for all the fonts and font sizes. In addition, we use 45, 330 word images for validation.

### 2.3.2 Network parameters

As mentioned above, the approach is based on MDRNNs and LSTM layers with a hierarchical network topology. Careful selection of topology parameters as well as the size of the LSTM layers are very essential for optimal recognition performance. The network uses block structures to pre-process the images for suitable feature extraction and feeds the information forward to higher layers for further processing. Input block size, hidden block size, subsample size and LSTM size are the interesting parameters; mostly, recognition performance of the trained models is dependent on these parameters.

We select appropriate values for these parameters by taking advantage of previous knowledge about training these kinds of networks, and we achieve good enough recognition performance for all the different evaluation tasks in the first attempt. We use input block size $1 \times 2$, hidden block sizes $1 \times 2$ and $1 \times 2$, subsample sizes 6 and 20, and hidden sizes 2, 10 and 50, corresponding to our network topologies in all the experiments. The only exception is made for single-font/single-size systems with font sizes 18 and 24. For these two systems, we used slightly bigger input blocks $2 \times 3$, leaving the rest of the configurations the same.

### 2.3.3 RNN Models

We train fifteen RNN models in total. For task 0, six separate models are trained for each font size without height normalization. For task 1, we train two models with target heights 15 and 21. Similarly, for task 2, we use two different target heights 23 and 28. However, for task 3 we use only one target height 24. In these three tasks, the input images are normalized to the target heights before training the models. In addition to these models, we build four more models exclusively for font size 6 for some other fonts.

## 2.4 Recognition

Recognition is performed by using one of the existing trained model for a given evaluation task. In case of single-font/single-size systems, input images are directly passed to the recognition system, and recognition is done by using the specifically trained model for that particular task.

In case of multi-size recognition systems, we have more than one trained model, and we choose a suitable model with the help of a classifier selection procedure. This procedure specifies the appropriate trained model and target image height by analyzing the input image height. If the input image height is within certain thresholds, the image is rescaled to a specific height and recognition is performed by using an appropriate trained model or classifier.

# 3. EXPERIMENTAL RESULTS

Evaluation is performed by computing the recognition accuracies at word and character level. Tables 1, 2, 3, 4 and 5 provide the recognition performances of the proposed method for evaluation tasks 0 to 3; in addition to this, Table 6 shows results for font size 6 at low resolution. The mean recognition rate (Mean RR%) in these tables gives the average performance of each system. Recognition rates are computed at character and word level without using language modeling. We achieve more than 99% word recognition accuracies for all evaluation tasks, and above 98% for low resolution Arabic text. Figures 2 and 3 compare the performance of different systems in recognition of *Arabic Transparent* and *DecoType Naskh* Arabic fonts for six font sizes. Similarly, figure 4 compares the performance of systems for font size 6, rendered in different fonts.

**Table 1: Task 0: Word and character recognition rates for *Arabic Transparent*, single-size systems.**

| Font size | 6 | 8 | 10 | 12 | 18 | 24 | **Mean** |
|---|---|---|---|---|---|---|---|
| WRR% | 99.82 | 99.93 | 99.92 | 99.93 | 99.96 | 99.94 | 99.92 |
| CRR% | 99.96 | 99.99 | 99.98 | 99.99 | 99.99 | 99.99 | 99.98 |

**Table 2: Task 1: Word and character recognition rates for *Arabic Transparent*, multi-size system.**

| Font size | 6 | 8 | 10 | 12 | 18 | 24 | **Mean** |
|---|---|---|---|---|---|---|---|
| WRR% | 99.51 | 99.82 | 99.90 | 99.83 | 99.95 | 99.87 | 99.81 |
| CRR% | 99.91 | 99.97 | 99.98 | 99.97 | 99.99 | 99.97 | 99.97 |

**Table 3: Task 2: Word and character recognition rates for *DecoType Naskh*, multi-size system.**

| Font size | 6 | 8 | 10 | 12 | 18 | 24 | **Mean** |
|---|---|---|---|---|---|---|---|
| WRR% | 98.71 | 99.39 | 99.33 | 99.70 | 99.68 | 99.34 | 99.36 |
| CRR% | 99.76 | 99.89 | 99.88 | 99.95 | 99.94 | 99.88 | 99.88 |

# 4. CONCLUSIONS

In this paper we describe a low resolution, multi-font, open vocabulary system for printed Arabic text. The system is based on multidimensional long-short term memory, recurrent neural network architecture with connectionist temporal classification layer. The proposed method is trained and evaluated using the APTI database. We report above 99% word recognition rate for multi font, multi size digitally represented Arabic text recognition task.

Table 4: Task 3: Word recognition rates (WRR%), multi-font/multi-size system.

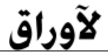
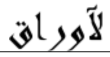
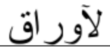
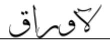
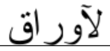
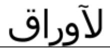
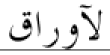
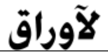
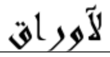
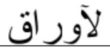
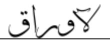
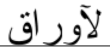
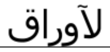
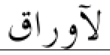
| Font | Sample Image | 6 | 8 | 10 | 12 | 18 | 24 | **Mean multi-size** |
|---|---|---|---|---|---|---|---|---|
| Advertising Bold |  | 99.85 | 99.95 | 99.96 | 99.95 | 99.95 | 99.93 | 99.93 |
| Andalus |  | 99.0 | 99.88 | 99.93 | 99.96 | 99.89 | 99.81 | 99.74 |
| Arabic Transparent |  | 99.44 | 99.94 | 99.96 | 99.97 | 99.96 | 99.94 | 99.86 |
| DecoType Naskh | | 97.23 | 99.33 | 99.42 | 99.56 | 99.51 | 99.17 | 99.03 |
| DecoType Thuluth |  | 96.35 | 99.18 | 99.49 | 99.51 | 99.33 | 99.93 | 98.79 |
| Diwani Letter | | 91.68 | 97.57 | 98.12 | 98.43 | 98.20 | 96.71 | 96.78 |
| M Unicode Sara | | 95.95 | 97.82 | 97.94 | 97.82 | 97.96 | 97.92 | 97.56 |
| Simplified Arabic |  | 99.27 | 99.92 | 99.94 | 99.95 | 99.94 | 99.78 | 99.8 |
| Tahoma |  | 99.64 | 99.94 | 99.95 | 99.96 | 99.96 | 99.95 | 99.90 |
| Traditional Arabic |  | 95.98 | 99.32 | 99.69 | 99.72 | 99.78 | 99.58 | 99.01 |
| **Mean multi-font** | | 97.44 | 99.29 | 99.44 | 99.48 | 99.45 | 99.17 | |
| **Overall Mean RR%** | | | | | | | | **99.04** |

Table 5: Task 3: Character recognition rates (CRR%), multi-font/multi-size system.

| Font | Sample Image | 6 | 8 | 10 | 12 | 18 | 24 | **Mean Multi-size** |
|---|---|---|---|---|---|---|---|---|
| Advertising Bold |  | 99.97 | 99.99 | 99.99 | 99.99 | 99.99 | 99.99 | 99.98 |
| Andalus |  | 99.82 | 99.98 | 99.99 | 99.99 | 98.98 | 99.97 | 99.95 |
| Arabic Transparent |  | 99.90 | 99.99 | 99.99 | 99.99 | 99.99 | 99.99 | 99.97 |
| DecoType Naskh | | 99.50 | 99.88 | 99.90 | 99.92 | 99.91 | 99.85 | 99.82 |
| DecoType Thuluth |  | 99.33 | 99.85 | 99.90 | 99.91 | 99.88 | 99.80 | 99.77 |
| Diwani Letter | | 98.44 | 99.56 | 99.66 | 99.71 | 96.66 | 99.40 | 99.40 |
| M Unicode Sara | | 99.27 | 99.62 | 99.64 | 99.62 | 99.64 | 99.63 | 99.57 |
| Simplified Arabic |  | 99.86 | 99.98 | 99.99 | 99.99 | 99.99 | 99.96 | 99.96 |
| Tahoma |  | 99.94 | 99.99 | 99.99 | 99.99 | 99.99 | 99.99 | 99.98 |
| Traditional Arabic |  | 99.25 | 99.87 | 99.94 | 99.95 | 99.96 | 99.92 | 99.81 |
| **Mean Multi-font** | | 99.53 | 99.87 | 99.9 | 99.91 | 99.9 | 99.83 | |
| **Overall Mean RR%** | | | | | | | | **99.82** |

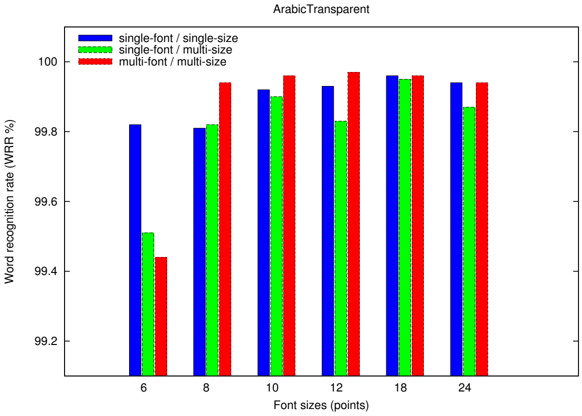


**Figure 2: Word recognition rates for *ArabicTransparent*. Systems have been trained for single-font/single-size, single-font/multi-size and multi-font/multi-size.**

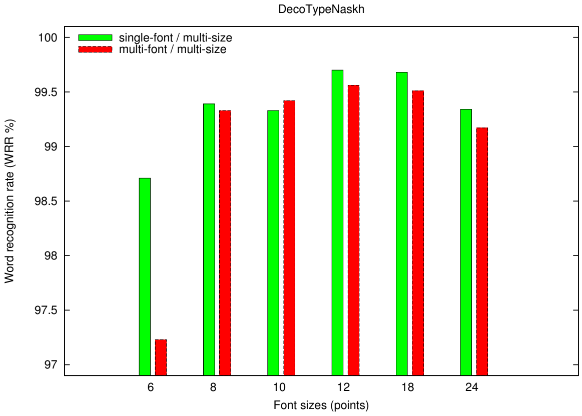


**Figure 3: Word recognition rates for *DecoTypeNaskh*. Systems have been trained for single-font/multi-size and multi-font/multi-size.**

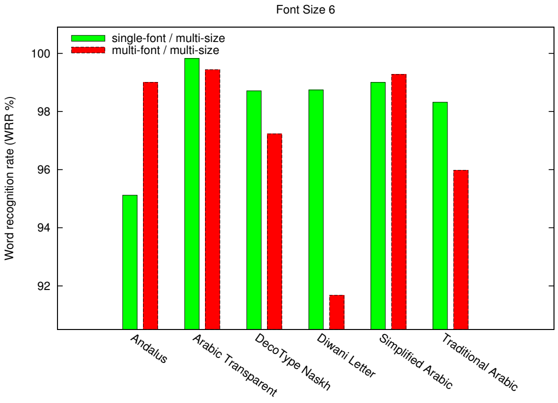


**Figure 4: Word recognition rates for various fonts with point size 6. Systems have been trained for single-font/multi-size and multi-font/multi-size.**

**Table 6: Word and character recognition rates for various fonts size 6. Single-font/single-size systems.**

| Fonts | WRR% | CRR% |
|---|---|---|
| Andalus | 95.12 | 99.14 |
| Arabic Transparent | 99.82 | 99.96 |
| Diwani Letter | 98.74 | 99.76 |
| Simplified Arabic | 98.93 | 99.80 |
| Traditional Arabic | 99.00 | 99.79 |
| **Mean RR%** | 98.32 | 99.69 |

The interesting finding is the application of the proposed method to very low resolution and small font size Arabic text. Despite of the challenges posed by the anti-aliased low resolution text, the proposed method yields above 98% average word recognition rate for most of the fonts. Analysis of figures 2, 3, and 4 reveals that, in few cases, the multi-font/multi-size system provides better recognition performance in comparison with single-font/single-size and single-font/multi-size systems. It seems that single-font/single-size systems are over-adaptive to the data, whereas the multi-font/multi-size system generalizes well due to training on a mixture of different fonts and font sizes.

## 5. REFERENCES

[1] A. Graves, S. Fernández, and J. Schmidhuber. Multi-dimensional Recurrent Neural Networks. *Artificial Neural Networks–ICANN 2007*, pages 549–558, 2007.

[2] A. Graves and J. Schmidhuber. Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks. *Advances in Neural Information Processing Systems*, 21:545–552, 2009.

[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[4] D. P. Mitchell and A. N. Netravali. Reconstruction filters in computer-graphics. *SIGGRAPH Computer Graphics*, 22(4):221–228, June 1988.

[5] S. Rashid, S. Bukhari, F. Shafait, and T. Breuel. A discriminative learning approach for orientation detection of urdu document images. In *IEEE 13th International Multitopic Conference, 2009. INMIC 2009.*, pages 1–5, 2009.

[6] S. Rashid, F. Shafait, and T. Breuel. Discriminative learning for script recognition. In *17th IEEE International Conference on Image Processing (ICIP), 2010*, pages 2145–2148, 2010.

[7] S. F. Rashid, F. Shafait, and T. M. Breuel. An evaluation of HMM-based Techniques for the Recognition of Screen Rendered Text. In *11th International Conference on Document Analysis and Recognition (ICDAR'11), 2011*, pages 1260–1264. IEEE, 2011.

[8] F. Slimane, R. Ingold, S. Kanoun, A. M. Alimi, and J. Hennebert. A new arabic printed text image database and evaluation protocols. In *10th International Conference on Document Analysis and Recognition (ICDAR'09), 2009*, pages 946–950. IEEE, 2009.