Out of vocabulary word detection and recovery in Arabic handwritten text recognition

Sana Khamekhem Jemni, Yousri Kessentini, Slim Kanoun

Version of Record: https://www.sciencedirect.com/science/article/pii/S0031320319301785
Manuscript_96b216fa416af6a565078deb48a8d765

Out of Vocabulary Word Detection and Recovery in Arabic Handwritten Text Recognition

Sana KHAMEKHEM JEMNI[a], Yousri KESSENTINI [a,b], Slim KANOUN [a]

[a] *University of Sfax, MIRACL Laboratory, Sfax, Tunisia*

[b] *Digital Research Center of Sfax, B.P. 275, Sakiet Ezzit, 3021 Sfax, Tunisia*

{sana.khamekhem, yousri.kessentini, slim.kanoun}@gmail.com

Abstract

Today's Arabic Handwriting recognition systems are able to recognize arbitrary words over a large but finite vocabulary. Systems operating with a fixed vocabulary are bound to encounter so-called out-of-vocabulary (OOV) words. The aim of this research is to propose a two-step approach that tackles the problem of OOV words in Arabic handwriting. In the first step, we exploit different types of sub-word units to detect the potential OOVs. In the recovery stage, a dynamic dictionary is built to extend the initial static word lexicon in order to cope with the detected OOVs. The recovery includes a selection step in which the best word candidates extracted from the external resource are kept. Experiments were conducted on the public benchmarking KHATT and AHTID/MW databases. The obtained results revealed that sub-word modeling could give cues for improving the detection and that the use of a dynamic dictionary significantly improves the recognition performance compared to one-step approaches that are based on a large static dictionary or the combination of different sub-word units. We achieve the state of the art results on the KHATT dataset.

Keywords

Arabic Handwriting recognition, Out of vocabulary detection and recovery, Static lexicon, Dynamic lexicon, Statistical language model, Deep learning, Multi-dimensional long short term memory network.

1. Introduction

Handwritten Text Recognition is an active research area in the field of pattern recognition, which aims at converting a text from an image format to an electronic one. Therefore, the text recognition engine remains the main component of a document processing system. In fact, the success of any document processing system involves a highly precise text recognition system. Several systems are commonly trained and used for handwritten and printed text recognition tasks. Various approaches have been involved to deal with handwritten documents for a large vocabulary recognition task. Specifically, the most widely used methods are based either on the hidden Markov's models (HMMs) [83] or on the recurrent neural networks (RNN).

These systems rely on the internal representations that are produced using the sliding window approach, in which features are extracted from the line image vertical frames, whose output is fed to a trainable classifier. This method transforms the problem to a sequence to sequence transduction one, while eventually encoding the two-dimensional image nature using convolutional neural networks [75] or defining the relevant features [82] [34].

While the issue of learning features has been a topic of interest for decades, substantial progress has been achieved with the development of deep learning methods during the last few years. Especially, deep learning methods allowed

building systems that can handle both the 2D aspect of the input image and the prediction sequential aspect. In particular, multidimensional long short-term memory recurrent neural networks (MDLSTM-RNNs) associated with the Connectionist Temporal Classification (CTC) [28] yield low error rates and have become the state of the art model for handwriting recognition [76] [78-80]. More recently, attention-based models have been applied to recognize handwritten text averting the paragraph to lines segmentation problems [77].

Traditional handwriting recognition research rely on linguistic resources including static word lexicons [17], referred to as In Vocabulary words (IV words) in this research study [15] [47-58] [25-26] [60-61]. Over the last three decades, several research works have taken into account the presence of words that do not belong to the used word lexicon [1-14] [62]. In our work, we referred to these words as Out-Of-Vocabulary words (OOV words). OOV words represent an important source of error in word spotting [14], speech and handwritten text recognition systems and thus several research works have been proposed to address this issue.

In the field of Automatic Speech Recognition (ASR), there has been significant work in OOV word detection and recovery. The methods addressing this problem can be grouped into two categories: OOV detection based approaches and lexicon selection based ones. The OOV detection based approaches [64-66] proceed by detecting the OOV words and/or locating the OOV regions in the ASR hypothesis, followed by a search process to match the phoneme sequence that constitutes the OOV word. Generally, these methods mainly involve hybrid language models thanks to their ability to model both in-vocabulary word and sub-word units. However, the OOV detection methods rely heavily on features taken from the speech recognition hypothesis such as posterior scores. Such features are not that reliable as they may reflect the correspondence between the word hypothesis and the signal input and not the presence of the OOV word. Besides, a hybrid LM may require careful selection of sub-word units that can sometimes lead to increased error rates [67]. Vocabulary selection based approaches propose a relevant vocabulary for speech recognition based on additional text data. The second category based approaches has been proposed to minimize the OOV rate for a domain specific corpus [68-69]. Moreover, they are more dynamic [70] as they can suggest context specific vocabulary.

For handwriting recognition, the methods addressing the OOV problem employ the same techniques used in ASR systems. These methods can be subdivided also into two categories. One-step approaches [1-4], [9-12] try to recover OOV words during the recognition process by increasing the vocabulary size, which generally increases the computational complexity and the confusability in the data. An alternative approach is to use sub-word units, either to estimate a full sub-word LM or to generate a hybrid LM that incorporates both words and sub-words. The performance of sub-word modeling approaches will however depend on the language model design and most importantly on the properties of the training corpus compared to those of the test corpus. In addition, they can produce some words that do not belong to the language and consequently their recognition performance drastically drops. The second category is based on two processing steps: OOV detection and OOV recovery [5-8], [13].

The detailed study of the existing handwriting recognition research works shows that most of the existing word and text recognition systems integrate only the OOV words recovery without any preliminary detection step. To our knowledge, only one existing system handles the OOV words detection in handwritten Latin script [5]. Such detection is essentially based on the comparison of the confidence scores of the recognized words with a heuristic threshold

whose value is determined through several experiments. In this framework, the used hypothesis is that the OOV words will, in most cases, have lower confidence scores than those of the IV words.

Considering the OOV words recovery, most of the proposed handwritten text recognition systems rely on the so-called sub-lexical units. These systems enrich the word lexicon by decomposing the words into different sub-word lexical units. These units can be letters or syllables for several scripts. They can be, Part of Arabic Words (PAWs) or morphemes, particularly, for the Arabic script. The PAWs result from the natural segmentation of words because of the presence of letters that do not connect to their successors in the words. A morpheme can be a prefix, which is added at the beginning of the word, or a suffix, which is added at the end of the word, or a stem. The morphemes result from the morphological structure of the Arabic vocabulary [16] [18] [59]. Other systems rely on a hybrid lexicon combining the different sub-word lexical units and words [1-3].

To increase the OOV words recovery rates, several systems have used, in addition to the statistical language models, a text corpus, freely available through the web [19-24]. Such text corpus are used to feed the used initial word lexicons with new words and to build new statistical language models whose states, transitions and transition state probabilities are determined on the basis of the ground truth texts of the training parts of the used image databases and the freely available text corpus.

Since this paper proposed, a new Arabic Handwritten Text Recognition system, called AHTR system for the detection and recovery of OOV words in Arabic text images, it was rather limited to a detailed critical analysis of the research works proposed in [1-3] that deals only with the recovery of these words.

In [1], hybrid LMs consisting of words and PAWs were used to recover OOV words. The authors have decomposed the less frequent words in the training corpus into PAWs in order to provide an opportunity for newer words to appear. The used recognition engine relies on the hybridization of HMMs and Multi-Dimensional Long Short Term Memory Networks (MDLSTMs), which directly exploit the pixel values of text line images in four different scan directions. A CTC is used during the training step. To generate the letter sequence hypotheses, the Viterbi algorithm combined to the Weighted Finite State Transducers (WFSTs) [29], is applied.

In [2], the Arabic word morphological decomposition is adapted in the handwriting recognition system. Unlike the PAW decomposition, the morphology based one uses the internal structure of the Arabic word (i.e. prefix, stem and suffix). This technique decreases the out of vocabulary words by including new words generated from the morphological decomposition process. This process allowed a 1 % improvement in the system performance. In addition to the use of this model, the authors exploit a text corpus collected from freely available newspapers and forums to direct the recognition stage. This study, therefore confirms that such text corpus exploitation has brought about a significant increase for the OOV words recovery rates and therefore decreased the word error rate significantly. The optical model is constructed using the Hidden Markov Models (HMMs) [27]. Each text line image is represented by a feature vector sequence extracted from a sliding window of size 9x30 with one pixel overlap.

More recently, the vocabulary augmentation was performed by decomposing the lexicon into morphemes and PAWs using a hybrid morphological decomposition [3]. Although theoretically interesting, this method results in a nominal improvement when compared to the PAWs or morphemes modeling.

References [1-3] analyze and compare different aspects of sub-word (PAWs and/or morphemes and words) LMs to handle the OOV issue. However, there are still some persisting relevant problems to be solved. These approaches

are able to recognize OOV words by the concatenation of sub-word hypotheses. Indeed, the concatenated sub-words can lead to an incorrect word that does not belong to the Arabic language, since no word lexicon is provided either during recognition step or in the post-processing stage. Moreover, the statistical language models estimation carried out on a decomposed text corpus can produce an increased statistical bias that may affect the vocabulary single items. Unlike the presented methods for OOV recovery in Arabic scripts, the study proposed in [5] precedes the OOV words recovery by a preliminary detection step. This detection relies on the confidence score feature. However, this proposal suffers the shortcoming that such a measure is useful to determine whether a word hypothesis is correct or not, but not whether it is an OOV or not. A more rigorous investigation on the existing methods is given in Table 1.

TABLE 1

A summary of existing OOV recovery methods in handwritten documents.

Author(s)/ref.	OOV recovery method	Scope of application	Advantages	Drawbacks
Hamdani et al. [2]	A sub-word based language-modelling method exploiting the morphological structure of the Arabic vocabulary. A hybrid lexicon including words and morphemes directs the recognition process.	Arabic Handwritten script	Recover OOV words by the concatenation of morpheme hypotheses.	Appearance of words that do not belong to the Arabic language.
BenZeghiba et al. [1]	A sub-word based language-modelling method exploiting the Arabic script nature. A hybrid lexicon including words and Part-Of-Arabic words directs the recognition process.	Arabic Handwritten script	Recover OOV words by the concatenation of PAW hypotheses.	Appearance of words that do not belong to the Arabic language.
BenZeghiba [3]	A mixed sub-word based language-modelling method exploiting the morphological structure of the Arabic vocabulary and the Arabic script nature. The mixed lexicon is constructed by decomposing words into morphemes and PAWs. This lexicon is used to direct the recognition process.	Arabic Handwritten script	Recover OOV words by the concatenation of mixed sub-word hypotheses.	Appearance of words that do not belong to the Arabic language. IV words are misrecognized.
Oprean et al. [5]	Two stages were used to recover OOV words. Firstly, a detection method based on the confidence score measure generated using the BLSTM classifier is performed to identify OOV	Latin Handwritten script	OOV words are recovered and the overall system performance is improved compared to the use of large static dictionary.	The detection method, used in the first step, is not reliable as it is based on the word posterior probability generated using the BLSTM classifier.

	words. Secondly, dynamic lexicons created from Wikipedia were used for recovery.			
Swaileh et al. [4]	A statistical language model combining syllables and characters estimated on a Wikipedia corpus is used. The syllabic lexicon is constructed using a supervised spelling syllabification method.	Latin Handwritten script	The syllabic model ensures the coverage of a large proportion of OOV words.	Appearance of words that do not belong to the used language.
Kozielski et al. [9]	Character and word n-grams language model interpolation.	Arabic and Latin Handwritten scripts	The interpolated language model improves the recognition results.	Appearance of words that do not belong to the used language.
Bazzi et al. [19]	Hybrid word-Character n-grams.	Arabic and Latin printed scripts	Competitive results have been achieved by the hybrid system if compared to the word-based system.	The hybrid model allows the appearance of a character sequence that does not belong to the used language.

Different from the AHTR systems proposed in [1-3] which focus only on the OOV words recovery, the AHTR system proposed in this paper starts by the detection of the OOV words then proceeds with their recovery. In addition, while the AHTR systems, proposed in [1-3], rely directly either on handcrafted features, ours relies on learned features deduced automatically through a deep multi-dimensional network architecture. This architecture consists of MDLSTMs and Convolutional Neural Network (CNN) layers, along with max-pooling and arranged alternately.

Contrary to [5] that proposes an OOV detection method based on confidence score, we suggest different OOV words detection methods and demonstrate that sub-word lexical units (PAWs and morphemes) modeling could give cues for improving detection.

In addition, and contrary to the AHTR system proposed in [2] which increases the used word lexicon by the most frequent words from the text corpus, freely available through the web, a dynamic lexicon is built in this paper by selecting words from the text corpus based on their string similarity to the detected OOV words.

The first contribution of this paper concerns the OOV detection module where three different methods are proposed. The first method is based on the word confidence scores of the Word Lexicon Driven recognition method (WLD). The second method relies on the difference between the word hypotheses from Word Lexicon Driven (WLD), PAW Lexicon Driven (PLD) and Morpheme Lexicon Driven (MLD) recognition methods. The third method uses the word confidence scores of the three sub-word modeling approaches. We demonstrate that sub-word modeling could give cues for improving the detection and that the best detection method is the second, which is not based on the confidence score. The second contribution is the use of a dynamic lexicon that extend the initial lexicon in order to cope with the detected OOVs. It includes a selection step in which the best word candidates from the external resource are kept. Finally, the proposed OOV detection and recovery methods are generic and independent of the recognition engine. The obtained results reveal that the proposed method achieves state of the art results on KHATT dataset and

significantly improves the recognition performance over the use of reduced and large static dictionaries and the combination of different sub-word modeling approaches.

In the remaining of this paper, we first described the system proposed for AHTR and especially the methods suggested for the detection and recovery of the OOV words. In a second step, we detailed, the used text line images database and word lexicon. Thirdly, the obtained experimental results were revealed and discussed. Finally, the main conclusions were drawn and some future works were suggested.

2. Proposed Arabic handwritten Text recognition System

The fundamental objective of this paper was to propose an original method for OOV words detection and recovery in handwriting recognition. To achieve this objective, three different lexicon driven recognition methods were used: the first method is a Word Lexicon Driven (WLD), a second method is a PAW Lexicon Driven (PLD), and a third method is a Morpheme Lexicon Driven (MLD). For the first recognition method, the text line hypotheses construction is carried out relying on a Word Statistical Language Model WSLM (Fig 1. a). The second recognition method is based on a PAW Statistical Language Model PSLM (Fig 1. b) whereas the third recognition method uses a Morpheme Statistical Language Model MSLM (Fig 1. c). The three different recognition methods are based on the same letter recognition engine. A word or PAW or morpheme hypothesis consists of the letter hypotheses concatenation. In addition, the three lexicons used to direct the word and the PAW and the morpheme recognition methods, are, called, the Reduced Static Word Lexicon, the Reduced Static PAW Lexicon and the Reduced Static Morpheme Lexicon respectively. These are constructed from the ground truth texts training dataset of the image database. In the remainder of this section, the letter recognition engine that is being was described. In a second step, the proposed methods for the detection and recovery of the OOV words were detailed. For OOV words recovery, we used two other word lexicons constructed from text corpus, freely available, through the web, in addition to the ground truth texts of the used training image dataset.

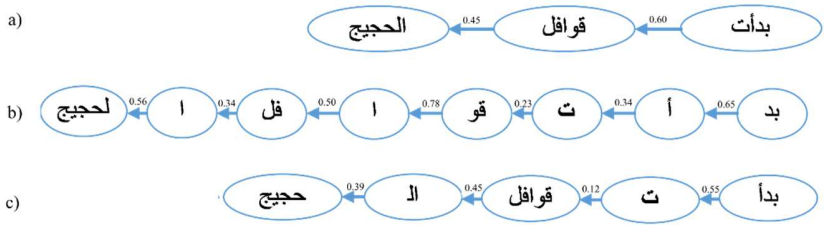


Fig 1. Example of each of the three used statistical language Models: a) WSLM, b) PSLM, and c) MSLM.

2.1 Letter recognition engine

As mentioned in section 1, it is worth reminding that the proposed AHTR system and therefore, the used letter recognition engine, is based on the segmentation free approach. Similar to the prior work [30], the proposed architecture consists of MDLSTMs and CNN layers, along with max-pooling and arranged alternately (Fig 2).

The choice of the MDLSTM network is justified by the fact that it represents a robust method that allows for a flexible modeling of the multidimensional context by establishing recurrent connections for all spatiotemporal dimensions that exist in the input data [36]. These connections provide the MDLSTMs with a high resistance to local distortions in an input image (for example, rotation, shear, etc.).

First, each original gray image is normalized to a fixed height of 96. Then, the image is presented to four MDLSTM layers, one for each scanning direction to generate a feature sequence from the input images. A convolutional layer, subsampling the feature maps, follows each MDLSTM layer. The two-dimensional sequence presented on the height axis and generated using the last MDLSTM layer is then reduced to a one-dimensional one. Thereafter, the resulted feature map is collapsed to a fixed height of one. The character posterior probabilities are predicted using a Softmax layer that already processed the collapsed feature map. Finally, a WFST [29] combined with the CTC is used to transform the input text line image into a sequence of word or PAW or morpheme hypotheses.

The architecture of our proposed network is composed of five alternating pairs of convolution and MDLSTM layers. Concerning the dropout technique, it is applied for forward connections of all CNN layers (the MDLSTM layers and the output layer) except for the first CNN layer.

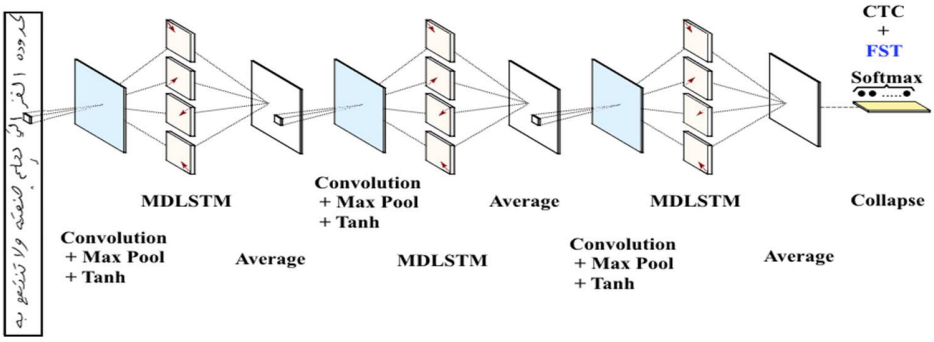


Fig 2. The cascade of MDLSTM and CNN layers used in the proposed AHTR system. (Figure adapted from Pham et al. [31]).

To prove the effectiveness of the proposed recovery and detection modules, we implemented a second recognition engine based on handcrafted HOG features [32] [33] combined with six other features, describing the pixel distribution inside each frame, and Bidirectional Recurrent Neural Networks (BLSTM) [63]. More details about the used BLSTM architecture can be found in our previous work [34].

2.2 Proposed methods for OOV words detection

In this section, the three proposed OOV word detection methods were detailed in the framework of the proposed AHTR system. The first method is based on the confidence scores of the word hypotheses generated using the WLD recognition method. The second is based on the differences between the word hypotheses from WLD, PLD and MLD recognition methods. The third method applies, in a first step, the first method based on the confidence scores of the word hypotheses from the three recognition methods: WLD, PLD and MLD. It then applies, in a second step, a majority vote process.

It is noted that the confidence score of a word or a PAW or a morpheme hypothesis reflects a similarity degree between the image corresponding input (represented by a feature vectors sequence) and the trained letter models (see section 2.1). It is, therefore, obvious that the higher a confidence score is, the closer the word or PAW or morpheme associated hypothesis will be to the correct one (that of the ground truth).

In the framework of the proposed AHTR system, the word, PAW, or morpheme confidence score is derived from posterior probabilities. These probabilities are estimated over the word or PAW or morpheme graphs (WGD, PGD and MGD), which are compact representations of the word or sub-word hypotheses. Posterior probabilities in the lattices were computed using the Minimum Bayes Risk Decoding algorithm [37]. Since, lattices could be potentially very large, with thousands of nodes and edges, the graph densities were adjusted to restrict the search space. The word, PAW or morphemes graph densities obviously have an impact on the confidence error rates [72]. In other words, if the graph density becomes very low, there is a significant leak in the score performance. Consequently, WGD, PGD or MGD fitting was performed, which enables us to get the best confidence error rates. This adjustment would be performed by pruning the word, PAW and morpheme lattices, in such a way that those, which do not reach a threshold, were removed. If no pruning was performed, the lattice could be highly accurate but also exorbitantly large. We selected the threshold that had a reasonable lattice density while keeping the best confidence error rates.

2.2.1 First method: OOV words detection based on confidence score hypotheses from RDWL recognition

Having a text line image as input, the WLD recognition method forwards a text line hypothesis formed by the word hypotheses succession. Each word hypothesis has its own confidence score. In this first method, a word is considered as an OOV if its confidence score is lower than a heuristic threshold whose value is determined following several experiments.

2.2.2 Second method: OOV words detection based on the differences between the word hypotheses from WLD, PLD and MLD recognition methods

Initially, it is useful to remind that the first method is based only on the WLD recognition method. Thus, the main characteristic of this method is that such confidence measures are good only at predicting whether the hypothesized word is correct or not. Therefore, it cannot distinguish between the errors due to OOV words and those caused by other phenomena such as degraded writing conditions. Consequently, this second method relied not only on the word hypotheses from the PLD and the MLD recognition methods, but also on the word hypotheses generated using the WLD recognition method. This is justified by the fact that the PLD and the MLD recognition methods can forward different word hypotheses from those of the WLD recognition method, on the one hand, and the fact that the PLD and the MLD recognition methods lead to textual entity hypotheses smaller than the word, on the other. Thus, it can be noticed that a word hypothesis is built by the PAW hypotheses concatenation and the morpheme hypotheses concatenation for the PLD recognition method and the MLD recognition method, respectively.

Having a text line image as input, each recognition method, whether WLD, PLD or MLD, emits a text line hypothesis made up of the word hypotheses succession. Thus, three text line hypotheses are at our disposal. The three hypotheses for each word were compared for these three text line hypotheses. Therefore, an alignment process is required at this stage to compare the three text line recognition hypotheses. To this end, the dynamic programming algorithm which

is implemented in Kaldi's toolkit [71] was used for this purpose. A word is finally considered an OOV if two of its three hypotheses are different in at least two text line hypotheses.

2.2.3 Third method: OOV words detection based on sub-word modeling and confidence score

In this third method, in addition to the word confidence scores obtained from the WLD recognition method, we exploited those associated with the word hypotheses from PLD and MLD recognition methods. The confidence score of a word hypothesis from these recognition methods is equal to the sum of the confidence scores of, respectively, the PAW hypotheses or the morpheme hypotheses, which constitutes the corresponding word hypothesis, divided by the number of PAWs or the number of morphemes, respectively.

Having a text line image as input, each recognition method whether the WLD, PLD or MLD, forwards a text line hypothesis formed by the word hypotheses succession. Thus, three text line hypotheses are at our disposal. Each word hypothesis in each text line hypothesis has its own confidence score. Here, the first method described in 2.2.1, was applied on each of these text line hypotheses in a first step, which allows us to establish the OOV word hypotheses for each one and the located OOVs are labeled as « OOV » in the three-text line hypotheses. In a second step, a majority vote is applied in order to decide whether a word is an OOV or not. This process is based on dynamic programming implemented within the ROVER algorithm. Finally, a word hypothesis is considered as an OOV if it is detected as an OOV in at least two hypotheses.

Figure 3 illustrates the three proposed methods for OOV words detection. For the first method, a heuristic threshold is used to classify the words into OOV or not OOV. Thus, the words 'أثر', 'حاجة' and 'نوافل' are considered as OOVs. For the second OOV detection method, the three systems (PLD, MLD, WLD) were explained in section 2.2.2. Therefore, the words 'تلبية', 'أثر', 'حاجة' and 'نوافل' are detected as OOVs in the final text line transcription. Considering the third detection method, the OOV words are firstly detected in the three text line hypotheses generated using the PLD, MLD, WLD systems. The detected OOVs are labeled as « OOV » in the three text line hypotheses. An alignment process that relies on a dynamic programming is performed. As a result, the words 'تلبية', 'أثر', 'حاجة' and 'نوافل' are finally considered as OOVs.

Fig 3. Illustration of the proposed methods for the OOV words detection.

2.3 OOV words recovery proposed method

In this section, we described our proposed recovery method based on the Dynamic Word Lexicon Driven recognition method (DWLD). It relies on the use of a dynamic lexicon built by selecting words from large a text corpus freely available on the web, based on their string similarity with a reference OOV. The reference OOV is defined as the word that will be used as requested for the construction of the dynamic lexicon.

For each detected OOV word in the text line hypotheses generated using the WLD recognition method, we identified their equivalent words in the text line hypotheses generated by the PLD and the MLD recognition methods using the ROVER alignment method [38]. The reference OOV was then identified as the word with the higher confidence score among the PLD and MLD recognition method hypotheses. This choice of a reference OOV is justified by the fact that the WLD recognition method is constrained to return the most similar word from the lexicon to the detected OOV, which does not necessarily exist. Contrarily, the words returned by the PLD and the MLD recognition methods are likely to be the most similar words to the detected OOV. The reference OOVs' identification is shown in Fig 4 (references OOVs are written in red).

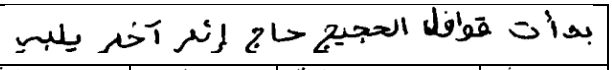
Text line image input							
WLD recognition output	تلبية	آخر	أثر	حاجة	الحجيج	نوافل	بدأت
OOV Detection results	OOV	آخر	OOV	OOV	الحجيج	OOV	بدأت
PLD recognition output	تلبى	آخر	اثر	حاج	الحجيج	قوافل	بدأت
	0.89	0.96	0.91	0.92	1.00	0.93	0.94
MLD recognition output	تلبية	آخر	أثر	حاج	الحجيج	نوافل	بدأت
	0.88	0.99	0.92	0.91	1.00	0.89	0.95
reference OOV	تلبى	آخر	اثر	حاج	الحجيج	قوافل	بدأت

Fig 4. Reference word identification for the OOV recovery step.

Once the reference OOV is identified, a dynamic lexicon is built by selecting words from a large text corpus freely available on the web, based on their string similarity with the reference OOV word character string using the Levenshtein distance [39]. A word is considered similar to another if the difference between them is lower than a heuristic threshold. The extracted words are used to extend the initial word lexicon with new words. This extension was performed for all the identified reference OOVs in the text line hypothesis. Hence, for each text line image, a dynamic lexicon was built to drive the recognition process. The initial word lexicon was extended by new words and the word statistical language model was adapted by readjusting the transition probabilities between the existing words. The described OOV recovery method is iterated for each text line image of the used image database test set.

Fig 5 illustrates the OOV word recovery method based on a dynamic lexicon. As shown below, for the given input text line image, a first recognition was performed using the WLD recognition method. Then, the words (نوافل, حاجة , أثر, تلبية) were identified as OOV using the proposed OOV detection method. After the reference OOVs identification, a lexicon search was performed to find the nearest words in the large external text corpus. For instance, considering

the reference OOV (تلبى), the words "يلبي,تلبي, تلبية , طلبي, etc." were added to the initial word lexicon. Then, the extended lexicon was used to drive the recognition process.

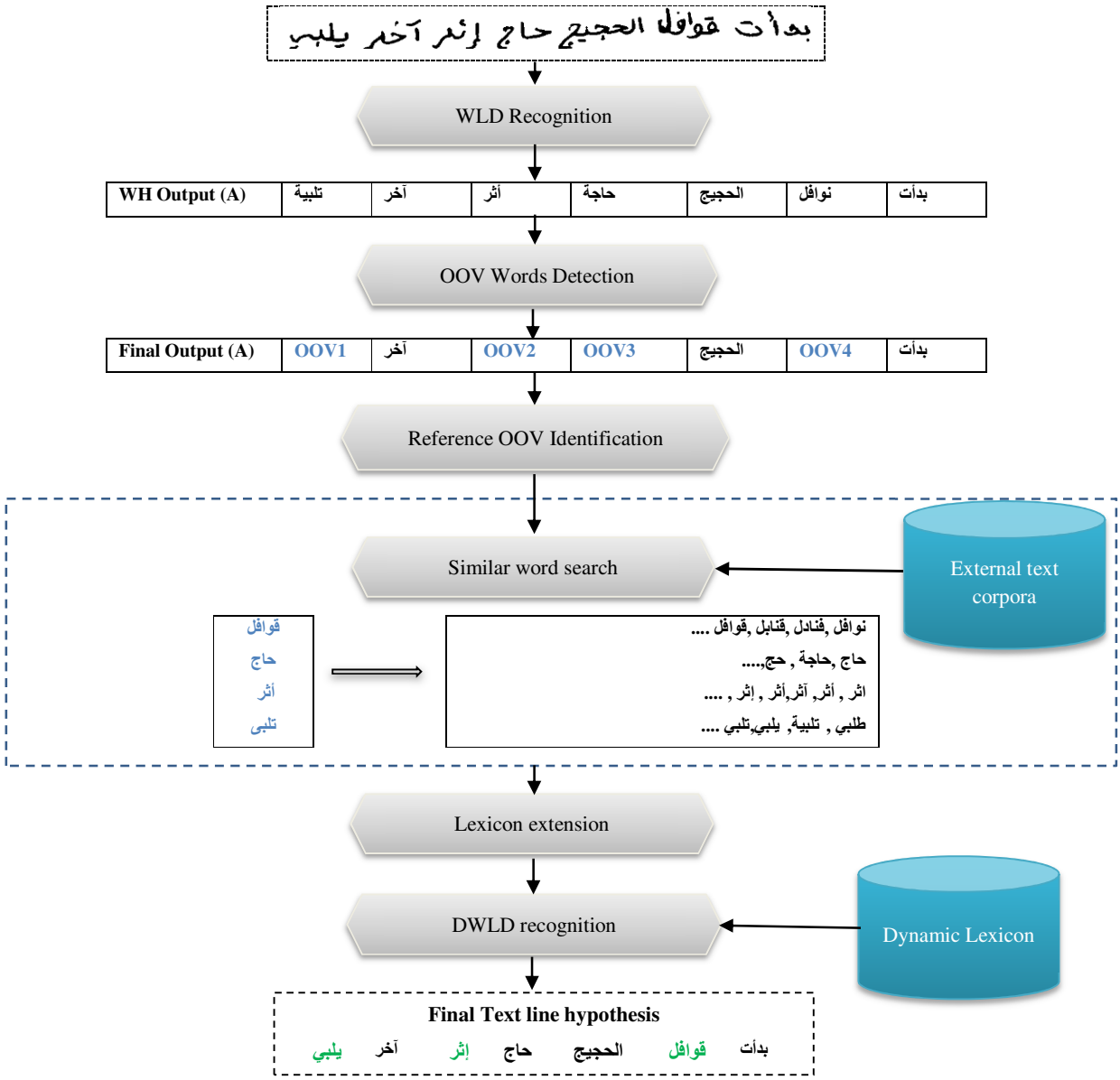


Fig 5. OOV Recovery method based on a dynamic lexicon.

3. Arabic databases description

Performances of the proposed OOV detection and recovery methods are evaluated on two benchmarking Arabic databases, namely KHATT and AHTID/MW databases.

3.1 KHATT database

The KHATT [40] is more challenging database than several available and well-known Arabic databases such as IFN / ENIT database [41]. It is more appropriate for our study since it is composed of line texts and not isolated words as showing in Fig 6.

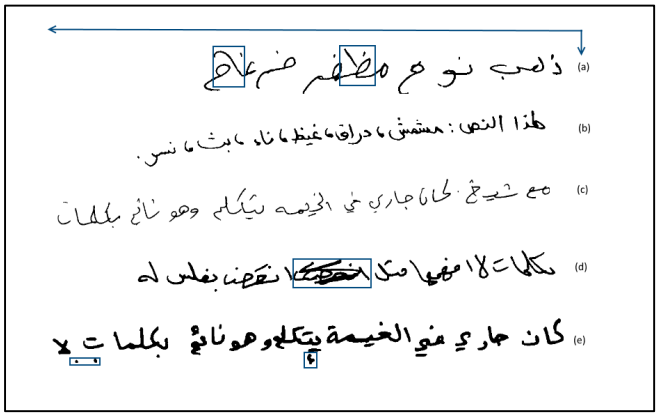


Fig.6. Samples of KHATT text line images : a) deslanted letters, b) deskewed text, c) writing style variation, d) noisy writing, e) noise introduced from adjacent lines.

TABLE 2

Statistics in terms of number for the KHATT dataset

	TRAIN	TEST	Validation
Pages	690	141	148
Lines	9475	2007	1902
Words	129 826	26 449	26 142
Characters	605 537	122 757	121 433
PAWs	246 762	49 781	49 285
Morphemes	190 757	38 617	38 493
OOV words	--	3067	2998
OOV PAWs	--	975	1 017
OOV Morphemes	--	2 957	2 426

The KHATT database is an offline handwritten text image database that includes 4000 paragraphs written by 1000 distinct authors. The text images are scanned at multiple resolutions. The text line images are automatically extracted

from the proposed paragraph images. In this work, the experiments were conducted on all the line images scanned at 300 dpi resolution. This database consists of three subsets: training, validation and test. Table 2 reports some statistics on the KHATT database. The default word lexicon is composed of 18933 words corresponding to 7885 distinct PAWs and 13422 distinct morphemes (17 prefixes, 37 suffixes, and 13368 stems). Table 2 reports also the number of OOV words in the test and validation subset.

3.2 AHTID/MW database

Additional experiments were performed using the AHTID/MW handwritten text line database [81]. It is an offline handwritten database that includes 3710 line images written by 43 individuals. The handwritten texts were scanned in grayscale with the resolution of 300 dpi. We used 887 line images for the test and 2823 for the training. It should be noted that the test set presents an OOV word rate of 12.49%. The statistics describing AHTID/MW database are presented in Table 3.

TABLE 3

AHTID/MW handwritten text image database statistics.

Subset	TRAIN	TEST
Lines	2819	887
Words	25 884	6786
Morphemes	37 653	11 267
PAWs	54 972	15 602
Characters	106 946	31 874
OOV words	--	848
OOV Morphemes	--	847
OOV PAWs	--	636

4. Experimental results

The experimental results of the proposed AHTR system are presented in the terms of word error rates (WER). We used the Levenshtein edit distance between the recognized text and the reference one. The editing distance is calculated by computing the number of editing operations (insertions, substitutions and deletions), required to transform a source character string into a target character string.

For the evaluation of the MDLSTM network, we used the RETURNN framework [42]. All the experiments were performed on the Tesla 80 GPUs. We fix the network parameters as presented in Table 4. We used a learning rate of 0.0005 that is then decreased to 0.0001 in the 35th epoch. During the training step, the Character Error Rate (CER) was computed on a sub-set of 10% of the training data. This measure is evaluated without using a lexicon or a language model. The training is stopped if the CER does not improve for 20 epochs.

TABLE 4

Parameters values for training the MDLSTM network.

Parameters	Values
Hidden Units	15n
Learning Rate	0.0005
Momentum	0.9
Filter size	3*3
Pooling Block	2*2
Dropout	0.25
Batch size	600k pixels

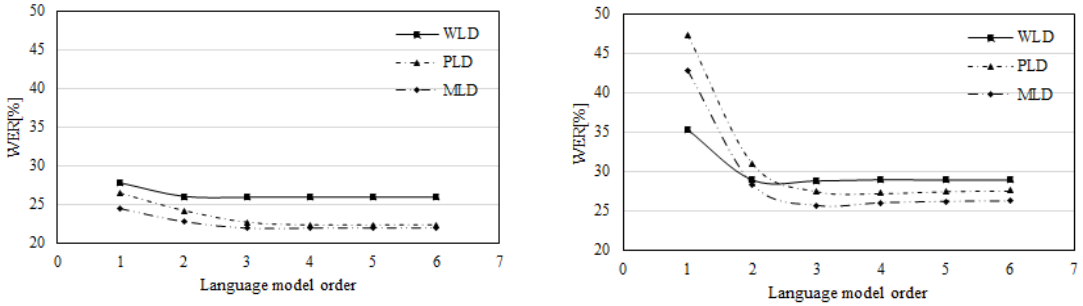
A second recognition engine based on handcrafted features was implemented for comparison. We used the BLSTM classifier with the following learning protocol. The BLSTM learning rate was set to 10^{-4} with a momentum equal to 0.9. The training stops if there is no enhancement of the character error rate on the validation set after 15 epochs. The choice of these parameters is justified by the fact that they provided a good accuracy for similar tasks [35]. We used the EESEN framework of the BLSTM [43]. In this framework, the CTC output layer was limited to the character labels of the ground truth texts of the KHATT training set. Since the Arabic script consists of 28 letters and each letter may have from one to four shapes, a set of 150 character shapes can be generated. As the number of character models has an impact on the system performance, we grouped the similar character shapes within the same class. A set of 108 character models were finally considered for the KHATT database, including punctuation, symbols and digits.

4.1 WLD, PLD and MLD recognition results

In this section, we presented the experimental results of the WLD, PLD and MLD recognition methods. For the MLD recognition method, the morphological decomposition was performed using the toolkit introduced in [44]. The decoding stage is based on a beam search in a FST, with a token passing algorithm. The EESEN speech recognition toolkit introduced in [43] was used for this purpose. This method is a variation of the technique explained in [29]. In these experiments, we restricted the search space to a sub-set of hypotheses using a defined beam. Different beam values were tested (10, 20 and 30). A beam value equal to 10 was chosen in such a way that keeps a low graph density and the best recognition performance as explained in section 2.2.

The WSLM, the PSML and the MSLM statistical language models were the standard n-gram trained using the SRILM toolkit [45]. In the training step, OOV words are mapped in a single special word. Thus, the decoding output was always limited to the intersection of the words in the language model with those in the lexicon.

The statistical language models were optimized through several experiments as shown in Figure 7. The best performances are obtained using 3-gram LMs for the WLD and MLD recognition methods and using a 4-gram for the PLD recognition method.



a) b)

Fig 7. WER as a function of the order of the word, PAW and Morpheme language models, computed on the test data of the KHATT database for the two recognition engines: a) CNN-MDLSTM recognition engine (OOV = 11 .46%), b) HOG-BLSTM recognition engine (OOV = 11 .46%).

TABLE 5

WER (%) performance for the WLD, PLD and MLD recognition methods performed on the Test set of the KHATT database.

OOV Rate (%)	WLD recognition method		PLD recognition method		MLD recognition method	
	CNN and MDLSTM	HOG and BLSTM	CNN and MDLSTM	HOG and BLSTM	CNN and MDLSTM	HOG and BLSTM
11.46	25.93	28.80	22.34	27.25	21.96	25.67
20	29.70	35.49	23.65	32.20	22.29	29.52
30	38.39	43.68	29.81	38.04	29.39	35.35
40	48.16	53.10	31.75	39.85	35.60	41.74
50	58.17	62.74	33.53	41.59	37.78	43.94
Recognition (With OOV and No LM)						
11.46	**34.36**	**54.07**	43.37	75.66	46.42	77.32
Baseline system (No OOV and No LM)						
0	**15.91**	**22.97**	24.84	45.70	21.01	39.01

Table 5 presents the performance of the WLD, PLD and MLD recognition methods with different proportion of OOVs. Starting with the default OOV rate of 11.46%, words from the testing set are arbitrary eliminated from the static dictionary, each time by 10%, until OOVs represent approximately 50% of the testing set.

The results are shown using two recognition engines. They prove that the CNN-MDLSTM architecture significantly improves the recognition performance compared to the HOG-BLSTM architecture. For an OOV rate of 11.46%, the relative improvements using the WLD, PLD and MLD recognition methods are 2.87%, 4.91% and 3.71%, respectively. It can also be observed that the PLD and MLD recognition methods systematically outperform the WLD recognition method thanks to their ability to recognize OOV words. PLD and MLD approaches show better behaviour in terms of accuracy as the proportion of OOVs increases.

An additional set of experiments was conducted to evaluate the performance of the optical models independently of the LMs. The obtained results (7th row of Table 5) confirm the superiority of CNN-MDLSTM over the HOG-

BLSTM recognition engine. These results reveal also an interesting efficiency of LMs, especially for sub-word LM. In fact, an improvement by 22.34% and 48.41% is observed using the PAW LM for the CNN-MDLSTM and the HOG-BLSTM systems, respectively.

Further experiments (8th row of Table 5) were carried out by closing the vocabulary to include the test set words in the LMs as a vocabulary. These experiments were performed without using n-gram LMs. These results show the performance of the proposed system in optimal conditions (without OOVs), and will serve for further comparison.

4.2 The OOV words detection results

This section describes the experimental results of the proposed OOV words detection methods presented in section 2.2. The OOV words detection results are provided using two recognition engines based on learned and handcrafted features and using two OOV rates (11.46% and 40%). The OOV detection results are reported using the recall-precision curve by varying the confidence threshold. The recall designs the fraction of correctly detected OOV words over the OOV number in reference texts, while the precision designs the fraction of correctly detected OOV words among the retrieved ones.

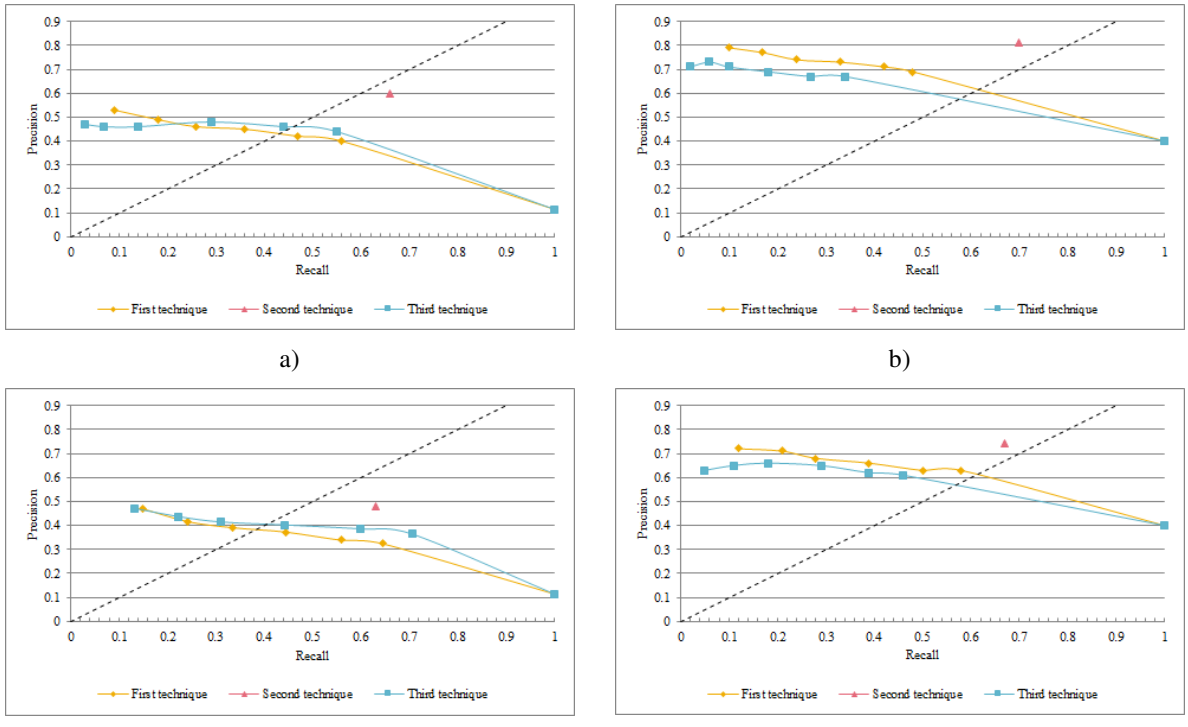


Fig 8. The three proposed OOV words detection methods experimental results performed on the Test set of the KHATT database: a) Detection using CNN-MDLSTM recognition engine (OOV = 11 .46%), b) Detection using CNN-MDLSTM recognition engine (OOV = 40%), c) Detection using HOG-BLSTM recognition engine (OOV = 11 .46%), d) Detection using HOG-BLSTM RECOGNITION ENGINE (OOV = 40%).

Figure 8 shows the precision vs. recall performance of the three detection methods on the same test set. For the first and the third detection methods, which are based on the confidence score, we display the recall and precision values for multiple threshold values (These thresholds are 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 1). The second detection method does not rely on confidence measure, which explains that only one recall-precision value is displayed (red triangle). It is clear from Figure 8 that the second detection strategy gives better detection performance and offers a good trade-off between precision and recall for the two recognition engines. Using the CNN-MDLSTM recognition engine, a recall of 65% and a precision of 60% are obtained with an OOV rate of 11.46%. With an OOV rate of 40%, the detection performance was improved and a recall of 70% and a precision of 81% are achieved.

TABLE 6

OOV words detection statistics [%]

using the CNN-MDLSTM Recognition Engine (OOV = 11 .46%).

	First Method	Second Method	Third Method
Detected as OOV, really OOV	62.21	69.09	60.22
Detected as OOV, really recognition error	33.94	20.93	33.02

Table 6 presents additional results to compare the quality of the different OOV detection methods. These results are provided for an OOV rate equal to 11.46% using the MDLSTM classifier. Given all detected OOVs, we present in Table 6 the percentage of words that are really OOVs and that correspond to recognition errors. It is clear that the second detection method gives the best result where 69.09% of the detected OOV words are really OOV. We conclude that the confidence score measures are good at predicting whether the hypothesized word is correct or not, but they are unable to distinguish between errors due to OOV words and errors caused by other phenomena such as degraded writing conditions. In the following recovery experiments, we used the second OOV words detection method, which gave the best detection performance.

4.3 The OOV words recovery results

We present in this section the result of the proposed two-step approach for the detection and recovery of OOV words. As presented in section 3.2, the proposed recovery method build a dynamic dictionary that extend the initial lexicon in order to cope with the detected OOVs. To build the dynamic dictionary, we include a selection step in which the best word candidates extracted from the external resource are kept. We design this method by Dynamic Word Lexicon Driven recognition method (DWLD).

The proposed OOV words recovery method is compared to two additional recovery methods that do not use an OOV word detection preliminary step. The first method consists in decoding the text line image using a large static lexicon made up of the initial lexicon and all the words of the text corpus which is freely available on the web. This recognition is designed by Word Wide Static Lexicon Driven recognition method (WWSLD). The second method combines the output hypotheses of the three WLD, PLD and MLD recognition methods using the ROVER algorithm.

The wide static lexicon is built from the WATAN corpus [46] that is freely available on the web. This text corpus contains 20 291 text documents composed of 272 638 distinct words. For both DWLD and WWSLD methods, KHATT and WATAN text corpora are combined to train the language model which is subsequently used to validate the most appropriate word hypothesis sequence.

It is clear from Table 7 that the DWLD method improves the recognition performance compared to the WWSLD method. For an OOV rate of 11.46%, an improvement of 0.64% and 2.77% were obtained using the CNN-MDLSTM and the HOG-BLSTM recognition engines, respectively. Significant improvement was also observed when the OOV rate increases.

TABLE 7

Experimental results (WER %) for OOV words recovery using DWLD and WWSLD recognition methods on the test set of the KHATT database.

OOV Rate%	DWLD		WWSLD	
	CNN-MDLSTM	HOG-BLSTM	CNN-MDLSTM	HOG-BLSTM
11.46	20.83	21.57	23.74	24.03
20	20.90	23.34	23.74	24.04
30	23.51	25.35	23.74	24.04
40	24.97	26.15	25.93	26.06
50	25.80	26.74	26.01	26.14

Table 8 presents the experimental results of the OOV words recovery method based on the combination of the output hypotheses of the WLD, PLD and MLD recognition methods via the ROVER algorithm. As expected, the overall recognition accuracy increases when combining the WLD with PLD or/and MLD recognition methods compared to the use of each method separately (see Table 5). For an OOV rate of 11.46%, the best performance is obtained by combining the PLD and MLD recognition methods where the word error rate is reduced to 21.51% using the CNN-MDLSTM recognition engine. Nevertheless, the DWLD method gives a better recognition accuracy with both CNN-MDLSTM and the HOG-BLSTM recognition engines.

The results of Tables 7 and 8 confirm the interest of the proposed two-step detection and recovery proposed approach compared to the one-step approaches based on a large static lexicon and on the combination of the outputs generated using different sub-word modelling methods.

TABLE 8

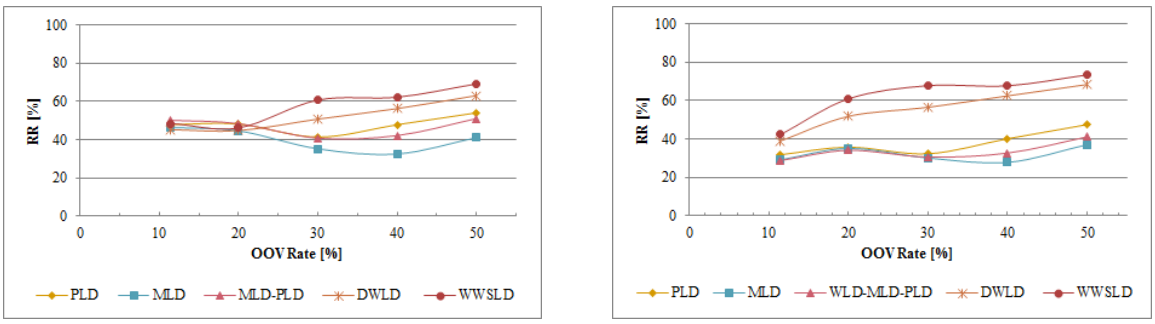
The experimental results of the OOV word recovery method based on different combination schemes (WER %) and performed on the Test set of the KHATT database.

	WLD and PLD recognition methods		**WLD and MLD recognition methods**		**PLD and MLD recognition methods**		**WLD and PLD and MDL recognition methods**	
OOV %	CNN-MDLSTM	HOG-BLSTM	CNN-MDLSTM	HOG-BLSTM	CNN-MDLSTM	HOG-BLSTM	CNN-MDLSTM	HOG-BLSTM
11.46	22.97	26.35	23.21	25.80	**21.51**	25.50	21.63	**24.28**
20	23.11	31.07	23.38	29.17	**21.46**	29.38	21.56	**28.26**
30	29.60	37.59	30.71	35.87	**27.80**	35.10	28.11	**34.16**
40	32.02	40.07	37.58	42.87	**31.87**	38.87	33.07	**38.78**
50	34.19	42.13	40.20	45.51	**32.90**	40.10	35.35	**40.87**

Figure 9 plots the recovery rate (RR) as a function of OOV proportion in the testing set. The recovery rate (RR) measures the proportion of OOV words that are successfully recovered and it is defined as:

$$RR = \frac{\text{Number of Correctly recognized OOV words}}{\text{Number of OOV words in Reference}} * 100$$

The DWLD recognition method presents the best OOV word recovery rate compared to the PLD, MLD and the combination of WLD, PLD and MLD recognition methods. Although the use of a wide static lexicon (WWSLD) yields a better recovery rate, the recognition rate of the dynamic lexicon based-recognition (DWLD) method remains the best. This result confirms that using the wide static lexicon increases the confusions with similar words which considerably decreases the recognition rate of In Vocabulary (IV) words.



a) b)

Fig 9. OOV Words Recovery results performed on the Test set of the KHATT database:

a) OOV Recovery using CNN-MDLSTM system,

b) OOV Recovery using HOG-BLSTM system.

Figure 10 shows examples of recognition errors using the different recovery methods. The illustrations are provided for an OOV rate of 30%. It can be shown that the output texts of the PLD and MLD approaches are complementary. For instance, the word "وكذلك" is correctly recognized using the PLD approach. However, using the MLD approach, this word is recognized as "وذلك". Reciprocally, the word "مكتسب" is misrecognized using the PLD approach, but correctly recognized using the morpheme based system.

The use of a dynamic lexicon further improves the recognition performance. For instance, the words "مجتهد", "مفتاح" and "العقل" are not recognized using the WLD approach. Yet, the use of the dynamic lexicon helps to correctly recover these words.

An analysis of the most likely recognition errors shows that some errors occur because of the misclassification of elongated Alif (آ) and Alif without Hamza (ا). For instance, the word "الآخره" is recognized as "الاخره" using the PLD approach. Besides, the misplacement of diacritics leads to many confusions like those caused by the words "مكنونةٌ" and "مكنونه".

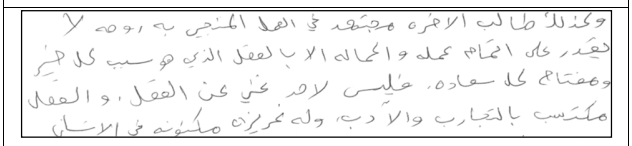
Handwritten Text Image

Ground Truth Text
وكذلك طالب الآخره مجتهد في العمل المنجي به روحه لا يقدر على اتمام عمله و اكماله إلا بالعقل الذي هو سبب كل خير ومفتاح كل سعادة فليس لأحد غني عن العقل والعقل مكتسب بالتجارب والأدب وله غريزةٌ مكنونةٌ في الانسان
WLD Recognition
وذلك طاب الاخرى مجهر ففي الها المنحر بة وجه تقدر علي تمام عملة الحاله الا بالعمل اللذي وسبب لكل غير فتاة هل سادة ليس لحد ي نحن الحقل والعمل يكتسب بالتجارة الادب له حريز مكونة ففي الإنسان
PLD Recognition
وكذلك طالب الاخره مجاهد في الها المنحر به وجه لا يقدر على اتمام عملة والحماله الابالعمل الذي هو سبب كل خير وفتاة كل سعاده وليس لاحد ني عن الحقل والعمل كتسب بالتجارب والادب وله حريزن مكنونه في الانسان
MLD Recognition
وذلك طالب الاخرى مجهر في الهل المنحر به وجه لا يقدر عل تمام عمله والحاله الا بالعمل الذي وسبب كل خير وفتاة كل سادة فلس لأحد في عن العمل والعمل مكتسب بالتجارب والادب وله حريزيزن مكنونه في الإنسان
Dynamic dictionaries
وكذلك طالب الاخرى مجتهد في الحل المنحم به روحه لا يقدر على اتمام عملة و الحاله الا بالفعل الذي هو سبب كل خير مفتاح كل سعاده فليس لاحد في عن العقل والعمل مكتسب بالتجارة الادب وله تمريرة مكونة في الانسان

Fig. 10. Example of recognition error extracted from the KHATT dataset.

In order to compare the achieved results with the most recent works presented in the literature, we report in Table 9 some recent results that were obtained on the test set of the KHATT database. We notice that our proposed AHTR achieves state of the art results with a WER of 20.83%.

TABLE 9

Comparison between the proposed AHTR system and the existing AHTR systems.

System	Used Features	Classifier	Method	WER%
Our proposed AHTR system	**Learned Features**	**CNN-MDLSTM**	**Dynamic lexicons (DWLD)**	**20.83**
System of Hamdani et al. [2]	Image pixels values+ Principal Component Analysis	HMM	Hybrid language model estimated on a large text corpus (morphological decomposition of Arabic words : morphemes)	26.80
System of BenZeghiba et al. [1]	Image pixels values	HMM/Artificial Neural Network	Hybrid language models (Word/ PAWs) estimated on the Train and Dev. sets of the ground-truth texts	30.90
System of BenZeghiba [3]	Image pixels values	MDLSTM-RNNs	Mixed hybrid language models (words, PAWs and morphemes)	34.30

An additional set of experiments was carried out using the AHTID/MW database in order to validate our proposed OOV detection and recovery method. In this work, the PAW, morpheme and word statistical language models were estimated on the training ground-truth texts of the AHTID/MW database.

The OOV recovrey results performed on the AHTID/MW database are summarized in Table 10, which shows the word error rate and recovery rate using the WLD, WWSLD and the DWLD recognition methods. The expermiments are performed using the CNN-MDLSTM recognition engine. As seen below, the best performance (WER =18.13%) is obtained by using the OOV detection which is based on the sub-word modeling combined with dynamic dictionaries (DWLD). These results are in accordance with those obtained on the KHATT database and they confirm the robustness of the proposed OOV recovery method.

TABLE 10

Experimental results performed on the AHTID/MW database

for OOV words recovery methods using the CNN-MDLSTM recognition engine.

Recognition Method	WER%	RR%
WLD	25.63	--
WWSLD	31.61	41.51
DWLD	**18.13**	**24.88**

4.4 Discussion

We propose in this paper a two-step approach for the detection and recovery of OOVs to improve the dictionary coverage in Arabic handwriting recognition systems. In the literature, OOV recognition issue can be handled using different approaches. One way to deal with OOVs is to build open vocabulary systems such as sub-word-based models [1] [2] [3]. Sub-word modeling methods can reach a coverage rate of nearly 100% of words belonging to the test datasets. Their performance will however depend on the language model design and most importantly on the properties of the training corpus compared to those of the test corpus. In addition, they can produce some words that do not belong to the language and consequently their recognition performance drastically drops.

An alternative way to deal with OOVs is to use large static dictionaries [19], [24] that consists in increasing the size of the working lexicon. However, large lexicons increase both computational complexity and confusions with similar words.

The two-step approach presented in this paper combines the advantages of these two previous approaches. We propose a new OOV detection approach that exploit the sub-word models to detect the potential OOVs. Then we build dynamic lexicons that extend the initial lexicon in order to cope with the detected OOVs. To build the dynamic dictionary, we include a selection step in which the best word candidates from the external resource are kept. The obtained results reveal that the proposed method significantly improves the recognition performance compared to the use of large static dictionaries and the combination of different sub-words models.

5. Conclusion

In this paper, we proposed a novel OOV word detection and recovery method, which exploits the modeling of different lexical entities such as words, PAWs and morphemes. The proposed OOV detection and recovery method is generic and independent of the letter recognition engine. It was validated using two different recognition architectures based on learned features and handcrafted ones. The obtained experimental results reveal that the proposed method significantly improves the recognition performance compared to the use of a large static lexicon and the combination of different sub-lexical units.

This work can be further extended by introducing Arabic mixed morpheme-PAWs language modeling into the detection schema. Furthermore, the selection of similar words using the Levenshtein distance can be optimized using a tree based implementation , or the integration of parallel processing on GPU. As perspectives to the use of sub-word units, we can mention the use of connectionist language models. This would remove the constraint of using the fix length dependency modelling introduced by the n-gram modelling approach. Finally, we may extend the proposed approach by including more languages, possibly with different scripts such as Latin, or Hindi languages to explore more in depth the advantages of using sub-word units for the recognition of multi-language handwriting.

References

[1] M. F. BenZeghiba, J. Louradour and C. Kerrnorvant, Hybrid Word/Part-of-Arabic-Word Language Models For Arabic Text Document Recognition, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2015, pp. 671 - 675.

[2] M. Hamdani, A. E. Mousa and H. Ney, Open Vocabulary Arabic Handwriting Recognition Using Morphological Decomposition, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2013, pp 280 - 284.

[3] M. F. BenZeghiba, Arabic Word Decomposition Techniques for Offline Arabic Text Transcription, in: Proceeding of International Workshop on Arabic Script Analysis and Recognition, ASAR, 2017, pp. 31 - 35.

[4] W. Swaileh, K. Ait Mohand and T. Paquet, A syllable based model for handwriting recognition, in: Proceeding of International Francophone Symposium on Writing and Document, CIFED, 2016, pp. 23-37.

[5] C. Oprean, L. Likforman-Sulem, A. Popescu and C. Mokbel, Handwritten Word Recognition using Web Resources and Recurrent Neural Networks, International Journal on Document Analysis and Recognition, Vol 18 (2015) 287 - 301.

[6] C. Oprean, L. Likforman-Sulem, A. Popescu and C. Mokbel, BLSTM-based handwritten text recognition using Web resources, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2015, pp. 466 - 470 .

[7] C. Oprean, L. Likforman-Sulem, A. Popescu and C. Mokbel, Using the Web to create dynamic dictionaries in handwritten out-of-vocabulary word recognition, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2013, pp. 989 - 993.

[8] C. Oprean, L. Likforman-Sulem, A. Popescu and C. Mokbel, Using the World Wide Web for the recognition of out of vocabulary handwritten words, in: Proceeding of International Francophone Symposium on Writing and Document, CIFED, 2014, pp. 217 - 232.

[9] M. Kozielski, M. Matysiak, P. Doetsch, R. Schloter and H. Ney, Open-Lexicon Language Modeling Combining Word and Character Levels, in: Proceeding of International Conference on Frontiers in Handwriting Recognition, ICFHR, 2014, pp. 343 - 348.

[10] M. Cai, W. Hu, K. Chen, L. Sun, S. Liang, X. Mo and Q. Huo, An Open Vocabulary OCR System with Hybrid Word-Subword Language Models, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2017, pp. 519 - 524.

[11] R. Messina and C. Kermorvant, Over-Generative Finite State Transducer n-gram for Out-Of-Vocabulary Word Recognition, in: Proceeding of IAPR International Workshop on Document Analysis Systems, DAS, 2014, pp. 212 - 216.

[12] M. Kozielski, D. Rybach, S. Hahn, R. Schluter and H. Ney, Open Vocabulary handwriting Recognition Using combined word-level and character-level language models, in: Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2013, pp. 8257 - 8261.

[13] S. Quiniou, M. Cheriet and E. Anquetil, Handling out-of-vocabulary words and recognition errors based on word linguistic context for handwritten sentence recognition, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2009, pp. 466 - 470.

[14] J. Puigcerver, A. H. Toselli and E. Vidal, Querying out-of-vocabulary words in lexicon-based keyword spotting, Neural Computing and Applications, (2016) 2373 - 2382.

[15] Y. Hilal, Tahlil Sarfi Lil Arabia, in: Proceeding of Computer Processing of Arabic Language, Kuwait, 1985.

[16] A. Ben Hamadou, A Compression Technique for Arabic dictionaries: The Affix analysis, in: Proceeding of Conference on Computational linguistics, COLING, 1986, pp. 286 - 288.

[17] J. Dichy and M. Hassoun, The DIINAR.1-«معالي» Arabic Lexical Resource, an outline of contents and methodology, The ELRA Newsletter, Vol. 10, n°2, April-June 2005, pp. 5 - 10.

[18] I. A. Al-Sughaiyer and I. A. Al-Kharashi, Arabic Morphological Analysis Techniques: A Comprehensive Survey, Journal of the American Society for Information Science and Technology, Vol 55 (2004) 189 - 213.

[19] I. Bazzi, R. Schwartz and J. Makhoul, An omnifont open-vocabulary OCR system for English and Arabic, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 21 (1999) 495 - 504.

[20] A. Brakensiek, D. Willett and G. Rigoll, Unlimited vocabulary script recognition using character N-Grams, in: Proceeding of Deutsche Arbeitsgemeinschaft für Mustererkennung, DAGM, 2000, pp. 490 - 504.

[21] F. Farooq, D. Jose and V. Govindaraju, Phrase-based correction model for improving handwriting recognition accuracies, Pattern Recognition, Vol 42 (2009) 3271 - 3277.

[22] T. Bluche, J. Louradour, M. Knibbe, B. Moysset, M. F. Benzeghiba and C. Kermorvant, The A2iA Arabic Handwritten Text Recognition System at the OpenHaRT2013 Evaluation, in: Proceeding of IAPR International Workshop on Document Analysis Systems, DAS, 2014, pp. 161 - 165.

[23] W. Swaileh, T. Paquet, Y. Soullard and P. Tranouez, Handwriting Recognition with Multigrams, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2017, pp. 137 - 142.

[24] U.-V. Marti and H. Bunke, On the influence of vocabulary size and language models in unconstrained handwritten text recognition, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2001, pp. 260 - 265.

[25] A. L. Koerich, R. Sabourin, C. Y. Suen, Large vocabulary off-line handwriting recognition: A survey, Pattern Analysis & Applications, Vol 6 (2003) 97 - 121.

[26] J. F. Pitrelli and A. Roy, Creating word-level language models for large-vocabulary handwriting recognition, International Journal on Document Analysis and Recognition, Vol 5 (2003) 126 - 137.

[27] T. Plötz and G. A. Fink, Markov models for offline handwriting recognition : A survey, International Journal on Document Analysis and Recognition, Vol 12 (2009), 269 - 298.

[28] A. Graves and F. Gomez, Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks, in: Proceeding of Proceedings of the 23rd international conference on Machine learning, ICML, 2006, pp. 369 - 376.

[29] M. Mohri, F. Pereira, and M. Riley, Weighted finite-state transducers in speech recognition, Computer Speech & Language, Vol. 16 (2002) 69 - 88.

[30] P. Voigtlaender, P. Doetsch and H. Ney, Handwriting Recognition with Large Multidimensional Long Short-Term Memory Recurrent Neural Networks, in: Proceeding of International Conference on Frontiers in Handwriting Recognition, ICFHR, 2016, pp. 228 - 233.

[31] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour, Dropout improves recurrent neural networks for handwriting recognition, in: Proceeding of International Conference on Frontiers in Handwriting Recognition, ICFHR, 2014, pp. 285 - 290.

[32] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, 2005, pp. 886 - 893.

[33] J. Rodriguez and F. Perronnin, Local gradient histogram features for word spotting in unconstrained handwritten documents, in: Proceeding of International Conference on Frontiers in Handwriting Recognition, ICFHR, 2008, pp. 19 - 21.

[34] S. K. Jemni, Y. Kessentini, S. Kanoun and J. Ogier, Offline Arabic Handwriting Recognition Using BLSTMs Combination, in: Proceeding of IAPR International Workshop on Document Analysis Systems, DAS, 2018, pp. 31 - 36.

[35] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke and J. Schmidhuber, A novel connectionist system for unconstrained handwriting recognition, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 31 (2009) 855 - 868.

[36] A. Graves and J. Schmidhuber, Offline handwriting recognition with multidimensional recurrent neural networks, in: Proceeding of International Conference on Neural Information Processing Systems, NIPS, 2008, pp. 545-552.

[37] H. Xu , D. Povey , L Mangu and J. Zhu, Minimum Bayes Risk Decoding and System Combination Based on a Recursion for Edit Distance, Computer Speech and Language , Vol 25 (2011) 802 - 828.

[38] J. G. Fiscus. A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER), in: Proceeding of IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU, 1997, pp. 347 - 354.

[39] V. I. Levenshtein, Binary codes capable of correcting deletions insertions and reversals. In Soviet physics doklady, Vol 10 (1966) 707.

[40] S. A. Mahmoud, I. Ahmad, M. Alshayeb, W. G. Al-Khatib, M. T. Parvez, G. A. Fink, V. Margner, and H. El Abed, KHATT: Arabic offline handwritten text database, in: Proceeding of International Conference on Frontiers in Handwriting Recognition ,ICFHR, 2012, pp. 449 - 454.

[41] R. Hussain, A. Raza, I. Siddiqi, K. Khurshid and C. Djeddi, A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation, EURASIP Journal on Image and Video Processing, Vol 46 (2015).

[42] P. Doetsch, A. Zeyer, P. Voigtlaender, I. Kulikov, R. Schluter and H. Ney, RETURNN: The RWTH Extensible Training framework for Universal Recurrent Neural Networks, in: Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2017, pp. 5345 - 5349.

[43] Y. Miao, M. Gowayyed and F. Metze, EESEN: End-to-End Speech Recognition using Deep RNN Models and WFST-based Decoding, in: Proceeding of IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU, 2015, pp. 167 - 174.

[44] A. Abdelali, K. Darwish N. D. H. Mubarak, Farasa: A Fast and Furious Segmenter for Arabic, in: Proceeding of Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, NAACL, 2016, pp. 11 - 16.

[45] A. Stolcke, SRILM – An Extensible Language Modeling Toolkit, in: Proceeding of International Conference on Spoken Language Processing, ICSLP, 2002, pp. 901 - 904.

[46] M. Abbas, K. Smaili, D. Berkani, TR-Classifier and kNN Evaluation for Topic Identification Tasks, the International Journal on Information and Communication Technologies, Vol 3 (2010) 65 – 74.

[47] S. Procter, J. Illingworth and F. Mokhtarian, Cursive handwriting recognition using hidden Markov models and a lexicon-driven level building algorithm, in: Proceeding of IEE Vision, Image and Signal Processing, VISP, 2000, pp. 332 - 339.

[48] S. Garcia-Salicetti, B. Dorizzi, P. Gallinari and Z. Wimmer, Maximum mutual information training for an online neural predictive handwritten word recognition system, International Journal on Document Analysis and Recognition, Vol 4 (2001) 56 - 68.

[49] A. L. Koerich, R. Sabourin and C. Y. Suen, Lexicon-driven HMM decoding for large vocabulary handwriting recognition with multiple character models, International Journal on Document Analysis and Recognition, Vol 6 (2003) 126 - 144.

[50] A. El-Yacoubi, M. Gilloux, R. Sabourin and C.Y. Suen, Fellow, An HMM-Based Approach for Off-Line Unconstrained Handwritten Word Modeling and Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 21 (1999) 752 - 760.

[51] Y. Kessentini, T. Paquet and A. B. Hamadou, A multi-stream approach to off-line handwritten word recognition, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2007, pp. 317 - 321.

[52] A. B. Bernard, F. Menasri, R. El-Hajj, C. Mokbel, C. Kermorvant, and L. Likforman, Dynamic and contextual information in HMM Behaviour for handwritten word recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 99 (2011) 2066 - 2080.

[53] S. Kanoun, A. M. Alimi, and Y. Lecourtier, Natural language morphology integration in off-line Arabic optical text recognition, IEEE Transactions on Systems, Man, and Cybernetics, Vol 41 (2011) 579 - 590.

[54] H. Xue and V. Govindaraju, On the Dependence of Handwritten Word Recognizers on Lexicons, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 24 (2002) 1553 - 1564.

[55] A. Esfahani, F. Farahnak, A. Katanforoush, A stroke-level wordnet for Farsi handwriting recognition, in: Proceeding of Iranian Conference on Machine Vision and Image Processing, MVIP, 2015, pp. 232 - 235.

[56] J.J. Hull, Incorporating Language Syntax in Visual Text Recognition with a Statistical Model, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 18 (1996) 1251 - 1255.

[57] S. Madhvanath and V. Govindaraju, The role of holistic paradigms in handwritten word recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 23 (2001) 149 - 164.

[58] G. A. Abandah, F. T. Jamour and E. A. Qaralleh, Recognizing handwritten Arabic words using grapheme segmentation and recurrent neural networks, International Journal on Document Analysis and Recognition, Vol 17 (2014) 275–291.

[59] M. Mustafa, A. S. Eldeen, S. Bani-Ahmad and A. O. Elfaki, A Comparative Survey on Arabic Stemming: Approaches and Challenges, Intelligent Information Management, Vol 9 (2017) 39-67.

[60] M. T. Parvez and S. A. Mahmoud, Offline arabic handwritten text recognition: A Survey, ACM Computing Surveys, Vol 45 (2013) 23 - 35.

[61] B. M. Al-Helali and S. A. Mahmoud, Arabic Online Handwriting Recognition (AOHR): A Survey, ACM Computing Surveys, Vol 50 (2017) 1 - 35.

[62] I. Ahmad, S. A. Mahmouda, G. A. Fink, Open-vocabulary recognition of machine-printed Arabic text using hidden Markov models, Pattern Recognition, Vol 51 (2016) 97–111.

[63] M. Schuster and K. K. Paliwal, Bidirectional recurrent neural networks, IEEE Transaction on Signal Processing, Vol 45 (1997) 2673 - 2681.

[64] S. R. Young, Detecting misrecognitions and out-of-vocabulary words, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 1994, pp. 21- 24.

[65] A. Rastrow, A. Sethy, B. Ramabhadran, and F. Jelinek, Towards using hybrid word and fragment units for vocabulary independent LVCSR systems, in: Proceedings of Annual Conference of the International Speech Communication Association, ISCA INTERSPEECH, 2009, pp. 1931-1934.

[66] W. Chen, S. Ananthakrishnan, R. Prasad, and P. Natarajan, Variablespan out-of-vocabulary named entity detection, in: Proceedings of Annual Conference of the International Speech Communication Association, ISCA INTERSPEECH, 2013, pp. 3761-3765.

[67] M. A. B. Shaik, A. E. D. Mousa, S. Hahn, R. Schluter, and H. Ney, Improved strategies for a zero OOV rate LVCSR system, in IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, 2015, pp. 5048-5052.

[68] A. Allauzen and J.-L. Gauvain, Diachronic vocabulary adaptation for broadcast news transcription, in: Proceedings of the 9th European Conference on Speech Communication and Technology, INTERSPEECH, 2005, pp. 1305-1308.

[69] A. I. R. M. Sun, Y. Chen, Learning OOV through semantic relatedness in spoken dialog systems, Proceedings of Annual Conference of the International Speech Communication Association, ISCA INTERSPEECH, 2015, pp. 1453-1457.

[70] O. S. S. Seneff, A two-pass strategy for handling OOVs in a large vocabulary recognition task, Proceedings of Annual Conference of the International Speech Communication Association, ISCA INTERSPEECH, 2005, pp. 1669-1672.

[71] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz and and G. Stemmer, The kaldi speech recognition toolkit, in Proceeding of Workshop on Automatic Speech Recognition and Understanding, ASRU, 2011, pp. 1-4.

[72] T. Fabian, Confidence Measurement Techniques in Automatic Speech Recognition and Dialog Management, dissertation, 2007.

[73] M. Pechwitz and V. Margner, HMM-Based Approach for Handwritten Arabic Word Recognition Using the IFN/ENIT Database, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2003, pp. 890-894.

[74] Y. Kessentini, T. Paquet, A. B. Hamadou, Off-line handwritten word recognition using multi-stream hidden Markov models, Pattern Recognition Letters, Vol 31 (2010) 60-70.

[75] T. Bluche, H. Ney, and C. Kermorvant, Feature Extraction with Convolutional Neural Networks for Handwritten Word Recognition, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2013, pp. 285-289.

[76] A. Graves, J. Schmidhuber, Offline handwriting recognition with multidimensional recurrent neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, NIPS, 2009, pp. 545-552.

[77] T. Bluche, J. Louradour and R. Messina, Scan, Attend and Read: End-to-End Handwritten Paragraph Recognition with MDLSTM Attention, in: Proceeding of International Conference on Document Analysis and Recognition, ICDAR, 2017, pp.1050-1055.

[78] R. Maalej, N. Tagougui and M. Kherallah, Recognition of Handwritten Arabic Words with Dropout Applied in MDLSTM, in: Proceedings of the International Conference on Image Analysis and Recognition, ICIAR, 2016, pp. 746–752.

[79] R. Yan, L. Peng, G. Bin, S. Wang and Y. Cheng, Residual Recurrent Neural Network with Sparse Training for Offline Arabic Handwriting Recognition, in: Proceedings of International Conference on Document Analysis and Recognition, ICDAR, 2017, pp. 1031-1037.

[80] S. Rawls, H. Cao, S. Kumar and P. Natarajan, Combining Convolutional Neural Networks and LSTMs for Segmentation-Free OCR, in: Proceedings of International Conference on Document Analysis and Recognition, ICDAR, 2017, pp. 155-160.

[81] A. Mezghani, S. Kanoun, M. Khemakhem, A Database for Arabic Handwritten Text Image Recognition and Writer Identification, in: Proceeding of International Conference on Frontiers in Handwriting Recognition, ICFHR, 2012, pp. 399-402.

[82] Y. Chherawala, P. P. Roy , M. Cheriet, Combination of context-dependent bidirectional long short-term memory classifiers for robust offline handwriting recognition, Pattern Recognition Letters, Vol 90 (2017) 58-64.

[83] Y. Kessentini, T. Paquet and A. B. Hamadou, "A Multi-Lingual Recognition System for Arabic and Latin Handwriting," *10th International Conference on Document Analysis and Recognition*, Barcelona, 2009, pp. 1196-1200.