Accepted Manuscript

Title: Deep learning for Arabic NLP: A survey

Author: Mahmoud Al-Ayyoub Aya Nuseir Kholoud Alsmearat Yaser Jararweh Brij Gupta

PII: S1877-7503(17)30375-7
DOI: https://doi.org/doi:10.1016/j.jocs.2017.11.011
Reference: JOCS 802

To appear in:

Received date: 2-4-2017
Revised date: 9-11-2017
Accepted date: 20-11-2017

Please cite this article as: Mahmoud Al-Ayyoub, Aya Nuseir, Kholoud Alsmearat, Yaser Jararweh, Brij Gupta, Deep learning for Arabic NLP: A survey, *<![CDATA[Journal of Computational Science]]>* (2017), https://doi.org/10.1016/j.jocs.2017.11.011

Deep Learning for Arabic NLP: A Survey

Mahmoud Al-Ayyoub[a,*], Aya Nuseir[a], Kholoud Alsmearat[a], Yaser Jararweh[a], Brij Gupta[b]

[a] *Jordan University of Science and Technology, Irbid, Jordan*
[b] *National Institute of Technology Kurukshetra, India*

Abstract

The recent advances in Deep Learning (DL) have caused breakthroughs in many fields such as computer vision, natural language processing (NLP) and speech processing. Many DL based approaches have been shown to produce state-of-the-art results on various tasks that are of great importance to Online Social Networks (OSN) and social computing such as sentiment analysis (SA) and Pharmacovigilance. NLP tasks are becoming very prominent in OSN and DL is offering researchers and practitioners exciting new directions to address these tasks. In this paper, we provide a survey of the published papers on using DL techniques for NLP. We focus on the Arabic language due to its importance, the scarcity of resources on it and the challenges associated with working on it. We notice that DL has yet to receive the attention it deserves from the Arabic NLP (ANLP) community compared with the attention it is getting for other languages despite the vast adoption of social networks in the Arab world. The majority of the early works on using DL for ANLP focused on OCR-related problems while the more recent ones are more diverse with the increasing interest in applying DL to SA, machine translation, diacritization, etc. This survey should serve as a guide for the young and growing ANLP community in order to help bridge the huge gap between ANLP literature and the much richer and more mature English NLP literature.

*Corresponding author
Email addresses: maalshbool@just.edu.jo (Mahmoud Al-Ayyoub), nuseir_aya@yahoo.com (Aya Nuseir), khlood1.smearat@live.com (Kholoud Alsmearat), yijararweh@just.edu.jo (Yaser Jararweh), gupta.brij@gmail.com (Brij Gupta)

Preprint submitted to Journal of Computational Science *November 27, 2017*

Keywords: Deep Learning, Arabic Natural Language Processing, Social Computing, Optical Character Recognition, Machine Translation, Text Categorization, Sentiment Analysis, Text Recognition, Speech Analysis.

1. Introduction

Over the past few decades, the amount of user generated content has grown rapidly. This is mainly due to the prevalence of Online Social Networks (OSN). Such surge of available raw data contributed to the rise of many new interesting fields such as social network analysis and big data in addition to increasing the interest in existing fields such as artificial intelligence and natural language processing (NLP).

social media analytics is closely related to NLP. The use of NLP techniques for processing the contents of OSN have gained a lot of attention from the academic world as well as from the industry. In fact, this myriad research and commercial interests have lead to/helped in the creation of entire events dedicated to this topic, such as SocialNLP,[1] in addition to the emergence of a large number of businesses [1].

NLP techniques can help in addressing many interesting problems in OSN such as geolocation identification, public opinion mining, sentiment/emotion analysis, trend analysis, event extraction, controversy detection, crowd monitoring, public health monitoring, disaster management, etc. [1, 2, 3, 4].

Recently, the NLP community has witnessed many breakthroughs due to the use of deep learning. Deep Learning (DL), a subfield of Machine Learning (ML), depends on a set of algorithms in order to learn multiple levels of representation with the aim of finding a model for high level abstractions in data. DL tries to mimic the human brain, by constructing an architecture that consists of an input layer and an output layer with many hidden layers (encoders) between them. These hidden layers are responsible for doing complex computations in

[1]`https://sites.google.com/site/socialnlp2016/`

order to extract features from the raw data in order to obtain a better representation. On the other side, there is the shallow learning, which is a type of learning that consists of at most of three levels, and it is widely used with linear problems [5]. Many techniques are used with DL, such as Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), etc., with several success stories across different domains.

In this paper, we survey the published works on applying DL techniques on NLP problems. We focus on the Arabic language due to its importance, the large number of Arabic speakers in OSN, the scarcity of resources on it and the challenges associated with working on it. This survey should serve as a guide for the young and growing Arabic NLP (ANLP) community in order to help bridge the gap between ANLP literature and the much richer and more mature English NLP literature.

The Arabic language is of undeniable importance. Not only it is being used by hundreds of millions of people, it has a quickly increasing online presence (in terms of the users and content). Moreover, Arabic has many unique characteristics that makes the automated handling and understanding of Arabic text challenging and interesting. Giving an overview of the Arabic language and a survey of the young field of Arabic NLP is beyond the scope of this paper. Interested readers are referred to [6, 7] for such coverage.

Among the many characteristics of the Arabic language, there are some characteristics that have significant effect on the general approaches to handle the different NLP problems. Examples include the language's complex nature (derivational, inflectional, etc.). Another example is the existence of many variations of the language such as the Classical Arabic (CA) (which is mainly used in ancient and theological texts but is still understood due to its use in the Holy Quran), the Modern Standard Arabic (MSA) (which is a modernized and simplified version of CA) and Dialectal/colloquial Arabic (DA) (where each region has its own dialect). The most widely-understood variation among Arabic speakers is MSA due to its wide adoption in education, media, and formal communication across the different Arabic speaking countries [8, 9].

In the rest of this paper, we discuss the papers using DL for ANLP tasks. The coverage is based on the addressed problems. We start with the most common application of DL for ANLP, which is for Optical Character Recognition (OCR) problems.

2. Optical Character Recognition (OCR)

In this section, we discuss the works employing DL techniques for Optical Character Recognition (OCR) problems in ANLP. We exclude the efforts focusing only on digit recognition including the very popular and highly cited MNIST dataset [10] as well as other similar datasets such as the CMATERdb 3.3.1 dataset.[2]

It is worth mentioning that Arabic scripts differ greatly from other languages and, thus, approaches that work well for Latin or Chinese scripts may fail horribly when applied to Arabic script. Reasons behind this include the use of cursive right-to-left style of writing, the fact that the shape of a character depends on its position within the word, the similarity between different characters (as they sometimes differ only by the number or positions of dots), etc. [11, 12].

As can be inferred from the size of this subsection, DL adoption for OCR of Arabic script has been the most popular among ANLP subfields, however, there is still a lot of room for improvement on OCR problems already addressed by DL. Moreover, there are other OCR-related problems that have yet to be addressed using DL approaches such as write identification [13].

2.1. Handwritten Text

We start our discussion with the one of the most frequently addressed OCR problems, which handwritten Arabic text recognition. In one of earliest works, Al-Jawfi [11] introduced an approach for Arabic handwritten text recognition problem. Based on the work of LeCun et al. [14], Al-Jawfi's approach used

[2] `https://code.google.com/archive/p/cmaterdb/`

LeNet, which is a convolutional neural network (CNN). The approach involves two phases: one for recognizing the body of the characters and one for recognizing the dots. It takes as input a set of 16×16 normalized images. They go through two hidden layers and the backpropagation algorithm was used to update the weights in the network. The experiments on this network was carried out using a dataset of 758 segmented images of Arabic characters (550 for training and 208 for testing) handwritten by different people. On average, the Mean Square Error (MSE) for the training set was 0.087 and, for the testing set, it was 0.42. The results are not very encouraging and the author do not compare them with the state-of-the-art results on the same dataset.

In [15, 16], the authors proposed to combine multidimensional RNN (MDRNN) and Connectionist Temporal Classification (CTC) to design an alphabet-independent offline handwriting recognition system that takes raw pixel data as input. Specifically, they propose to repeatedly compose multidimensional LSTM (MDLSTM) layers with feed-forward layers. They tested their network on the IFN/ENIT dataset [17], which consists of 32,492 images of Tunisian town names. They used 30,000 images for training and the remaining 2,492 images for testing. The reported error rate was 8.57%, which was better than the state-of-the-art at that time.

In [18], the authors compared MDLSTM learned features [15] with four state-of-the-art handcrafted feature sets. For recognition, each set of features was fed into bidirectional LSTM (BLSTM)-CTC network. For MDLSTM, the authors used the same network as the one used in [15] where the RNNLIB implementation [19] was used. The experiments were carried out using the IFN/ENIT dataset. The results showed that MDLSTM learned features produced the least Recognition Rates (RR) among all tested feature sets. However, the results also showed that merging the MDLSTM learned features with each one of the four handcrafted feature sets (through plurality voting) lead to improved RR. Finally, merging all five sets of features did not provide significantly better RR.

Building on [18], the same authors in [20] proposed a framework for evaluating feature sets for offline handwriting recognition. They proposed to train a RNN classifier on each feature and use a weighted vote combination of these classifiers. The goal is to measure the strengths and complementarity of different feature sets. They use their system to evaluate several feature sets (including handcrafted and automatically learned (using MDLSTM networks) feature sets) using an Arabic dataset (IFN/ENIT) and a Latine one (RIMES [21]).

Another work on Arabic handwritten text recognition problem is that of Porwal et al. [12], where the authors used Deep Belief Networks (DBN) with one input layer and three hidden layers. The training process had two phases. The first phase is the pre-training phase, which was done using Restricted Boltzman Machine (RBM). The output of each hidden layer was used as input parameters for the above hidden layer. The second phase was fine tuning, and DBN was used with features based approach. For the experiments, the authors used the Applied Media Analysis (AMA) Arabic Handwritten dataset of Parts of Arabic Words (PAW) classes.[3] The dataset had 6,464 training samples and 848 testing samples distributed among 34 classes. They compared their approach with those of [22, 23] and the results showed that using the proposed DBN approach on the raw images did not yield better results compared to [22, 23]. However, combining the DBN approach with the handcrafted features of [22, 23] gave the best accuracy of 89.4%.

In [24], the authors discussed combining Propabilistic Graphical Models (PGM) classifiers for Arabic handwritten words recognition. The considered classifiers include Hidden Markov Models (HMM) and DBN. They experimented with the IFN/ENIT dataset to show the feasibility of the proposed approach.

A large number of papers were published over the past few years by a group of researchers at the University of Gabes and Sfax University, Tunisia, on utilizing different DL networks such as DBN, DNN, CNN, RNN, LSTM, etc., for Arabic handwritten text recognition [25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37]. Similar to [12], Elleuch et al. [25] applied DBN on raw images of Arabic scripts. Specifically, they experimented with two datasets: the HACDB dataset for of-

[3] `http://appliedmediaanalysis.com/Datasets.htm`

fline Arabic character recognition [38] and the ADAB dataset for online Arabic word recognition [39]. The HACDB dataset contained 6,600 images, each normalized 28×28 pixels in gray scale. The images represent 66 classes: 58 classes representing the unique ways to write single Arabic characters and 8 classes representing the unique ways to write pairs of Arabic characters. The training part of the dataset contains 5,280 images while the testing part contained the remaining 1,320 images. For this dataset, experiments showed that the best performance reached a good error rate of 2.1% and it was achieved with two hidden layers, each with 1,000 units. According to the authors, this accuracy is comparable to what is reported in the literature. As for the ADAB dataset, it contained 946 different labels representing Tunisian towns names. It has a total of 33,164 sub-words and 174,690 characters written by 166 persons. For this dataset, experiments showed that the best performance reached a very poor error rate of 41% and it was achieved with two hidden layers, each with 400 units. The results for this dataset is way below what is reported in the literature.

In an extension of [25], the authors of [37] proposed to combine the DBN of [25] with Bottleneck Features (BNF) [40, 41]. They tested their approach on the LMCA dataset [42] padded with some manually segmented characters and ligatures extracted from the ADAB databset. The best error rate was 3.84%.

In another extension of [25], the authors of [26] proposed to use a Convolutional DBN (CDBN) combined with Support Vector Machine (SVM) classifier. Similar to [25], they considered two problems: character recognition, for which they used the HACDB dataset, and word recognition, for which they used the IFN/ENIT dataset. For the HACDB dataset, they showed that the proposed CDBN/SVM approach was better than the DBN approach of [25] with an error rate of 1.82% (compared with 2.1% error rate reported in [25]). As for the IFN/ENIT dataset, which consists of 26,459 words handwritten by more than 411 different persons representing 946 Tunisian town/village names, the reported error rate was 16.3%.

Another work by the same authors on CDBN is [32] in which the authors proposed to add offline hand-crafted features extracted using the BetaElliptical method. They tested their approach on the the ADAB and the LMCA dataset. The latter dataset contains images of digits, characters and words. However, the authors [32] only used the 100,000 images of 56 characters (with 70% used for training and 30% for testing). The reported error rates are 8.2% and 2.49% for the ADAB and LMCA datasets, receptively.

In another work [27] published in the same year as [25, 26], the authors compared DBN with CNN on the HACDB dataset. For DBN, they used the same architecture as [25, 26], but reported higher error rate of 3.64% compared with [25, 26]. As for CNN, the authors used two layers for convolution, each followed by a layer for sub-sampling to reduce dimensionality. Then, they had a fully connected layer for classification. The reported error rate for CNN was 14.71%.

In another extension of [25, 27], the authors of [33] experimented with using the dropout [43] and dropconnect [44] regularization methods to address the overfitting problem of DBN. They reported error rates of 2.73% and 2.27% when using dropout and dropconnect, respectively, which are better than the error rate of 3.64% reported in [27] when No-Drop was used. However, these results are still worse than the error rates reported in [25, 26].

Building up on [26, 27], the same authors proposed in [29, 30] to integrate CNN and SVM classifiers (while employing the dropout technique in [30]). The idea is to use SVM to alter the trainable classifier of the CNN and it was inspired by the work of Niu and Suen [45] on handwritten digit recognition. The CNN part is similar to the one proposed in [27], except that dropout is applied to the fully connected before feeding the output into an SVM classifier. They experimented on the HACDB and IFN/ENIT datasets. For the HACDB dataset, they reported error rates of 6.59% in [29] and 5.83% in [30], which are lower than the 14.71% error rate achieved by simple CNN in [27]. As for the IFN/ENIT dataset, which was only used in [30], the reported error rate was 7.05%.

Other works that tried to integrate SVM with DL was [28, 31], in which the same authors employed the idea of [46] to mimic DBN architecture in order to build a Deep SVM (DSVM) classifier by simply stacking several multi-class SVM classifiers (with Radial Basis Function (RBF) kernels). They experimented with the HACDB dataset and reported error rates of 8.64% in [28] and 5.68% in [31].

In another direction by the same group of researchers [34] considered using Deep Bidirectional LSTM (DBLSTM) networks [47], which is a special type of RNN. They trained it with CTC and used dropout to avoid overfitting. Using the RNNLIB library of [19], the authors built a network of 14 layers. Three of these layers are fully connected and hidden and they are composed of 20, 60 and 180 LSTM units. For the output layer, they used a softmax output layer of 46 units. They tested their network on the ADAB dataset and reported 26.28% error rate.

In [35], the authors built on the MDLSTM network of [15, 16]. They used 28 layers. Three of these layers are fully connected and hidden and they are composed of 2, 10 and 50 LSTM units. For the output layer, they used a softmax output layer of 121 units. The main contribution in [35] was to address the overfitting problem. To do so, the authors used the dropout technique. However, they only applied it on the feed-forward connections before the sub-sample layers which are not fully connected. They tested their approach on the the IFN/ENIT dataset to show its effectiveness compared with the original approach of [15, 16]. The error rate of [35]'s approach is 12.09% whereas the error of [15, 16]'s is 16.97%. In an extension of [35], the authors of [36] further studied applying the dropout technique in different places. Instead of only applying dropout after MDLSTM (as was done in [35]), they experiment with applying it before or inside MDLSTM. The results on the IFN/ENIT dataset showed reduced error rates of 11.62% and 11.88% for these two cases, respectively.

In [48], the authors described the systems they submitted to the Arabic handwriting recognition competition OpenHaRT 2013 [49]. These systems were based on an optical model using LSTM network trained to recognize the different forms of the Arabic characters directly from the image, without explicit feature extraction nor segmentation. Large vocabulary selection techniques and n-gram language modeling were used to provide a full paragraph recognition, without explicit word segmentation. Several recognition systems were also combined with the ROVER combination algorithm. The best system exceeded 80% of recognition rate.

Later, the same group showed in [50] how to improve the performance of RNN with LSTM cells on unconstrained handwriting recognition using dropout. They show how to carefully use dropout in the network so that it does not affect the recurrent connections, hence the power of RNNs in modeling sequences is preserved. The authors perform extensive experiments to confirm the effectiveness of dropout on deep architectures even when the network mainly consists of recurrent and shared connections. Among the datasets under consideration is the OpenHaRT 2013 dataset.

The Unsupervised and Transfer Learning Challenge (UTL) [51] consisted of five datasets from various domains including one with Arabic handwritten ancient manuscripts called AVICENNA.[4]. The authors of [52] discuss their DL approach for the UTL challenge datasets. Their approach combined and stacked different one-layer unsupervised learning algorithms, adapted to each one of the five datasets. For the AVICENNA dataset, they applied whitened Principle Component Analysis (PCA) on the raw data. In the second layer, they used Denoising Auto Encoder (DeAE) consisting of 600 hidden units. The last layer simply contained a transductive PCA.

In [53], the authors proposed a CNN network based on Fast Wavelet Transform (FWT) and Adaboost algorithm for Arabic handwritten character recognition. Based on Multi-Resolution Analysis (MRA) at different levels of abstraction, FWT is used to extract character's features. The authors evaluated their approach on the IESK-arDB dataset [54],[5] which includes 6,000 images, and the obtained accuracy was 93.92%.

Recently, El-Sawy et al. [55] discussed the use of CNN for recognizing Arabic handwritten characters. They used their won dataset consisting of 16,800 images

[4] `http://www.causality.inf.ethz.ch/ul_data/AVICENNA.html`

[5] `http://www.iesk-ardb.ovgu.de/`

and reported an error rate of 5.1%.

2.2. Printed Text

The authors of [56] addressed the problem of recognizing multi-font printed Arabic text in low-resolution images. The experimented with three types of networks inspired by [57, 15, 16]: MDRNN, MDLSTM, and CTC. The dataset they used is taken from the Arabic Printed Text Image (APTI) dataset [58] and it contained ten different fonts and six different font sizes. As for evaluation they used the tasks of the Arabic text recognition competition of ICDAR 2013.[6] The results showed that their system achieved recognition rates of 99% or above on most considered fonts and font sizes.

2.3. Video-Overlaid Text

The work of Yousfi et al. [59, 60, 61, 62] addressed the problem of detecting and recognition of Arabic text overlaid in videos. In [59], the authors proposed to use a CNN inspired by a prior work of the same authors [63]. They showed the superiority of the CNN approach compared with two approaches relying on multiexit asymmetric boosting cascade. The developed CNN consists of six layers with the first four layers being used for features extraction and the remaining two layers being used for classification. The authors used an in-house dataset collected from Arabic TV channels. The training set consisted of 30,000 text images and each method used bootstrapping to obtain negative examples. The dataset had two test sets: one has 201 images with 959 annotated text areas while the other has 164 images with 1017 annotated text areas. For the first test set, the CNN approach gave better F-measure and lower computation time compared with the other two approaches. However, its detection rate (DR) was lower. As for the second test set, CNN results were better in terms of F-measure and DR.

[6] `http://diuf.unifr.ch/diva/APTI/competitionICDAR2013.html`

In [60], the authors discussed their effort to build the ALIF dataset[7] for detecting Arabic text in TV broadcasts. It consists of 6,532 text images extracted from eight Arabic TV channels. The authors conducted experiments in the ALIF dataset using three approaches, one of which is a commercial OCR software (ABBYY Fine Reader 12). The remaining two approaches are RNN based ones inspired by [59]. Specifically, they used BLSTM coupled with CTC component. The difference between the two approaches is that one of them used CNN based features while the other used DBN based features. For the experiments, the authors used 4,152 images for training and the remaining images were distributed among three test sets of 900, 1,299 and 1,022 images. The evaluation metrics used were Character RR (CRR), Word RR (WRR) and Text RR (TRR). In all experiments, the BLSTM-CTC network with CNN based features produced the best results on all evaluation measures followed by the BLSTM-CTC network with DBN based features. The commercial software performed very poorly.

In a follow-up work, the authors of [61] extended the LSTM-CTC network of [59, 60] in order to apply it for recognizing Arabic text embedded in TV broadcasts. In addition to the two features methods of [59, 60], which are DBN based and CNN based, the authors of [61] added a third method based on Multi-Layer Perceptron (MLP) as Deep Auto Encoders (DAE). To evaluate these methods and compare them with commercial software as was done in [60], the authors used multiple datasets. The first one is the ArabCharSet, which is a set of 46,689 images of single letters and punctuations extracted from 17 hours of recordings from Arabic TV news channels. Two other datasets: ArabTrainText (consisting of 7,000 text images) and ArabValidText (consisting of 673 text images), were extracted from another 30 hours of recordings and were used for training and validating the BLSTM network. One more set, ArabTestText, consisting of 900 text images was used for testing. The experiments showed that the BLSTM-CTC network with CNN based features produced the best

[7] `https://cactus.orange-labs.fr/ALIF/`

results on all evaluation measures compared with the other two BLSTM-CTC networks, the commercial software and a BLSTM network based on the [64].

Finally, in [62], the authors proposed to use BLSTM-CTC network, but they integrate with it a large scale Arabic Language Model (LM). They employed a modified version of the Beam Search (BS) decoding algorithm that uses both responses from LM and LSTM. To control the decoding's effectiveness and efficiency, the authors introduced hyper-parameters and constraints. The decoding scheme is inspired by [65]'s scheme for speech recognition. As for the LM, the authors consider two approaches: RNN based LMs [66] and those combined with the Maximum Entropy (MaxEnt) models, in order to produce a character based Arabic LM. The approach is evaluated on the ALIF dataset of [60] where it showed significant improvement over a BLSTM baseline system in addition to a commercial software.

In [67], the authors addressed the problem of recognizing Arabic text in videos and natural scenes. They proposed to use a CNN-RNN hybrid architecture called Convolutional Recurrent Neural Network (CRNN). To evaluate their approach, they used several datasets. For the videos setting, they used the ALIF dataset [60] and the AcTiV dataset [68]. As for the natural scenes setting, they created their own datasets by collecting 500 word images containing Arabic text in different scenarios such as billboards, signs, etc. They called it the IIIT Arabic scene text dataset and they made it publicly available.[8] The results showed the superiority of the proposed approach compared with the existing approaches applied on the datasets under consideration including a commercial system (ABBYY) and an open source tool called Tesseract.[9]

2.4. Natural Scene Text

The last work discussed in the previous subsection ([67]) addressed the problem of detecting and recognizing text in natural scene images and mentioned

[8] `https://cvit.iiit.ac.in/research/projects/cvit-projects/arabic-text-recognition`

[9] `https://github.com/tesseract-ocr/tesseract`

the IIIT Arabic scene text dataset. In this section, we survey other papers addressing the same problem or problems relevant to it.

The authors of [69, 70] addressed the problem of detecting text in natural scene images, which can be more difficult than text printed on documents or embedded in videos due to the different levels of perspective distortion, illumination, etc. The problem's difficulty is increased when considering multi-script text on complex backgrounds. Following a patch based approach, in [69], they proposed to combine single layer CNN with a Naive-Bayes Nearest Neighbor (NBNN) classifier. For experimentation, they introduced their own MLe2e dataset,[10] which consists of 711 scene images covering four different scripts (not including Arabic). They also experimented with the Video Script Identification Competition (CVSI-2015) dataset [71],[11] which consists of 10,000 pre-segmented video words equally distributed among the ten scripts it covers (including Arabic). The experiments on both datasets showed the superiority of the proposed approach compared with several state-of-the-art approaches. However, for the CVSI-2015 dataset, the approach of the Google team produced slightly higher results.

In [70], the same authors extended their work by considering deeper CNN and training them using an Ensemble of Conjoined Networks. In addition to the MLe2e and CVSI-2015 datasets, the authors considered the SIW-13 dataset [72],[12] which consists of 16,291 pre-segmented text lines in 13 scripts (including Arabic). They also used the ICDAR2013 [73] and ALIF [60] datasets to evaluate the misclassification error of their approach. The results showed that the enhanced approach of [70] outperformed all other approaches (including [69]) expect that, for the CVSI-2015 dataset, the approach of the Google team produced slightly higher results.

Another work on the same problem is that of [74], in which the authors

[10] https://github.com/lluisgomez/script_identification

[11] http://www.ict.griffith.edu.au/cvsi2015/

[12] http://mclab.eic.hust.edu.cn/~xbai/mspnProjectPage/

proposed a patch based approach that assigns weights to patches using intra-cluster information entropy. Then, they employ what they call bag of CNN words. The authors used the MLe2e and SIW-13 datasets to evaluate their approach and the results showed that the proposed approach yielded competitive results.

In [75] proposed a simple approach of using CNN for recognizing text in scene images. For experiments, they used their own English-Arabic Scene Text (EAST) dataset, which consists of 2,450 training images and 250 testing image. The results showed that the proposed system produced encouraging results.

3. Caption Generation

The problem of automatically generating a description for a given image is an interesting and practical one . Unfortunately, it is almost untouched in the ANLP community. In [76], the author addressed this problem for Arabic by associating fragments of the image with root words. The fragments were extracted using a DNN pretrained on the ImageNet dataset [77]. DBN were used to find the most suitable words to be associated with the image and the different root words to be associated with the fragments. Finally, the caption sentences are generating by using dependency tree relations. For evaluation, the author created two datasets. The first one was created by taking 10,000 images from the ImageNet dataset and have human experts add a caption to each image. As for the second dataset, it was created by taking 10,000 images from the Al-Jazeera news website (which has both Arabic and English content). The results were very encouraging as they represent the first approach for Arabic caption generation. Moreover, they were much higher than the simple approach of generating English captions and automatically translating them into Arabic.

4. Automatic Speech Recognition (ASR)

In one of the earliest works employing DL for ASR, the author of [78] used RNN to recognize spoken Arabic digits. The RNN used was based on [79] and it consisted of two hidden layers. The proposed model performed well on an in-house dataset for both multi-speaker and speaker-independent settings.

The authors of [80] describe a distributed NN training algorithm, based on Hessian-free optimization, that scales to deep networks and large datasets. For the state-level minimum Bayes risk (sMBR) criterion, the authors state that this training algorithm is faster than stochastic gradient descent by a factor of 5.5 and yields a 4.4% relative improvement in word error rate on a 50-hour broadcast news task. The authors also state that distributed Hessian-free sMBR training yields relative reductions in word error rate of 713% over cross-entropy training with stochastic gradient descent on two larger tasks: Switchboard and DARPA RATS noisy Levantine Arabic. Their best Switchboard DBN achieved a 16.4% WER.

In [81], the authors proposed using feature-rich DNN language models (DNN-LM) on Egyptian Arabic. The inputs to the network are a mixture of words and morphemes along with their features. The authors state that significant Word Error Rate (WER) reductions were achieved compared to the traditional LM, which are based on words.

In [82], the authors described their efforts invested by the MIT team to tackle the 2016 Arabic Multi-genre Broadcast (MGB) Challenge,[13] which consisted of 1,200 hours of transcribed audio. In these efforts, the authors experimented with various DNN models including: Feed-forward NN, CNN, Time-Delay NN (TDNN), Recurrent LSTM, Highway LSTM (H-LSTM), and Grid LSTM (G-LSTM). The latter produced the best results with a WER of 18.3%. Previous works from the same group used different DNN and CNN models to build Arabic ASR systems [83, 84, 85].

Another team participating in the 2016 MGB Challenge was from QCRI. In [86], the authors describe their team effort, which included using the Lattice free Maximum Mutual Information (LF-MMI) modeling framework to generate purely sequence trained acoustic models. Another interesting aspect of this work

[13] `http://www.mgb-challenge.org/arabic.html`

is using four-gram and RNN with MaxEnt connections (RNNME) LM to do re-scoring. Finally, they used Minimum Bayes Risk (MBR) decoding criterion for system combination. The proposed approach achieved slightly better WER (14.2%) compared with the MIT team's approach of [82].

The final team in our discussion to participate in the 2016 MGB Challenge using a DL approach was from LIUM lab of the University of Le Mans, France. In [87], the authors describe this approach which employed GMM-derived (GMMD) to train a DNN, combined with the use of TDNN for acoustic models. The best reported results in [87] indicated that this approach had a WER of 15.7%.

5. Language Modeling

In previous sections, we discussed the use of LM in different problems (e.g., [62] used them for OCR and [81] used them for ASR). However, building LM can be studied independently. In [88], the authors proposed a character-level LM that can work on English as well as MRL such as Arabic. The proposed model applies CNN on input characters before feeding them into LSTM RNN-LM. The results for the Arabic language showed that the proposed LM outperformed various baselines working on word level or morpheme level. The authors have made the implementation of their approach publicly available.[14]

6. Automatic Machine Translation (ATM)

In this section, we discuss the efforts invested in developing DL based ATM systems for translating text from/to the Arabic language. We have noticed a scarcity of publications on this subject despite the popularity of DL approaches in ATM systems (especially, the commercial ones such as Google's). There are other relevant problems that have yet to be addressed using DL such as the translation between Arabic dialects [89].

[14] `https://github.com/yoonkim/lstm-char-cnn`

The authors of [90] proposed to address the Arabic-English machine transliteration problem using DBN, which contains multiple layers of RBM. The proposed approach has three important parts. The first one is the source encoder, which deals with source words by converting them to dimensional binary vectors, then feeding them into first layer in the source encoder, the output of each layer is considered as an input to the next layer. The second part called joint layer. This layer uses the output of the source encoder as an input in order to get a state of hidden neurons, and infer an output state to use as input to the top level of the output encoder. The third part is the target encoder. Within this part, the output vector is decoded by traversing down words through the output encoder. Because RBMs are bidirectional models, the translation process could be from Arabic to English, and vice versa. During the experiments, the authors studied the effects of network structure, by making the number of layers and the size of the bottom layers in the source and target encoders fixed, but the joint layer could be in different layers and different sizes.

Another relevant work is [91], where the authors explored the use of embeddings obtained from different levels of lexical and morpho-syntactic linguistic analysis. They showed that such an approach improved machine translation evaluation into morphologically rich languages (MRL) such as Arabic.

7. Dialect Detection

In [92], the authors discuss Discriminating between Similar Languages (DSL) shared tasks, which included a task for identifying Arabic dialects. The authors note that high-order character n-grams were the most successful feature, and the best classification approaches included traditional supervised learning methods such as SVM, logistic regression, and language models, while DL approaches did not perform very well.

In [93], the authors describe their character-level NN for the Arabic dialects identification task of the DSL challenge [92]. Given a sequence of characters, their model embeds each character in vector space, runs the sequence through multiple convolutions with different filter widths, and pools the convolutional representations to obtain a hidden vector representation of the text that is used for predicting the language or dialect. The implementation of their approach is publicly available [15] and the obtained F-measure is 48.3%.

Another deep learning approach for the dialect identification task of the DSL challenge [92] is [94]. The author used CNN and LSTM networks and obtained 43.29% weighted F-measure using CNN approach using default network parameters.

8. Dialectal Arabic (DA) Segmentation

Segmenting text written in DA can be a very challenging problem due to the lack of rules, standards and resources on DA. In [95], the authors built on the work of Yao and Huang [96] and proposed to use a character-level BLSTM network combined with the Conditional Random Field (CRF) algorithm to build a segmenter for the Egyptian dialect. Without relying on any additional resources, the proposed system produced results comparable to the state-of-the-art tools in text segmentation such as MADAMIRA and Farasa.

In another work by the same group [97], the authors entertained the idea of building a segmenter for one dialect and using it for another dialect. They also proposed to build a single segmenter for all dialects by joining training data for all dialects, thus, eliminating the need for dialect detection as a pre-processing step for segmentation. The authors used the BLSTM-CRF model discussed in the previous paragraph. They considered four dialects and the datasets they used is publicly available.[16]

9. Text Categorization (TC)

TC is a fundamental problem in information retrieval and NLP. However, not many works have proposed to use DL to address it. The only exceptions

[15] https://github.com/boknilev/dsl-char-cnn
[16] http://alt.qcri.org/resources/da_resources/

are [98, 99].

The authors of [98] developed an approach for Arabic text categorization that is based on three stages. The first stage is preprocessing, where the punctuation marks, auxiliary verbs, pronouns, etc., were removed. Then, the remaining words are represented in root patterns using a letter weight and order scheme. The second stage is clustering. It is done into two phases using Fuzzy C-Means (FCM) clustering and Markov clustering. The third stage is training DBN for each cluster. They evaluated their approach using two datasets: Al-Jazeera dataset (consisting of 10,000 articles) and the Saudi Press Agency dataset (consisting of 6,000 articles). The author reported an F-measure of 91.02%, which is significantly higher than the results reported in the literature.

10. Sentiment Analysis (SA)

In [100], the authors focused on building a word embeddings using a 3.4 billion words corpus from a collected 10 billion words web-crawled corpus. They used a CNN trained on top of pretrained Arabic word embeddings for SA to evaluate the quality of these word embeddings. The experiment results show that their scheme outperforms existing methods on several publicly available datasets [101, 102, 103, 104, 105].

Relying on word embeddings as the main source of features for SA, the authors of [106] presented their system which consists of the following steps. First, they compile a large Arabic corpus from various sources to learn word representations. Second, they train and generate word vectors (embeddings) from the corpus using the word2vec tool [107]. Third, they use the embeddings in their feature representation for training several binary classifiers to detect subjectivity and sentiment in both Standard and Dialectal Arabic. The implementation of this approach is publicly available.[17] The authors compare their results with other methods in literature. Their approach, which does not employ hand-

[17] https://github.com/iamaziz/ar-embeddings

crafted features, achieved a slightly better accuracy than the top hand-crafted methods.

The authors of [108] investigated the results of applying DL for sentence level Arabic SA. Four types of DL algorithms were used: DBN, DAE, DNN, and Recursive Auto Encoder (RAE). The input features for the first three models were extracted based on the ArSenL lexicon [109], while the features for the fourth model were the raw words indices, which are drawn from a known vocabulary obtained from a separate and independent training set. For evaluation, the authors used the Linguistic Data Consortium Arabic Tree Bank (LDC ATB) dataset, which consists of 1180 sentence, split into 80% training and 20% testing. The results showed the superiority of RAE compared with the three models under consideration in addition to state-of-the-art classifiers applied on the same dataset.

In a follow up work by the same group [110], the authors addressed some limitations of using RAE for Arabic SA, such as its poor handling of Arabic's morphological complexity. So, they proposed A Recursive Deep Learning Model for Opinion Mining in Arabic (AROMA), which starts by morphologically tokenizing the input before using semantic and sentiment embedding models. Finally, the structure of automatically generated syntactic parse trees is used to determine the order of the model's recursion. AROMA was evaluated using several datasets and the results showed that it outperformed baseline RAE as well as other well-known ML approaches for Arabic SA.

Baly et al. [111] proposed to use Recursive Neural Tensor Networks (RNTN) for SA of Arabic text. In order to use RNTN, the authors created the first Arabic Sentiment Treebank (ArSenTB). ArSenTB is enriched with orthographic and morphologic information. The authors made use of other tools such as word2vec [107] and MADAMIRA [112]. The input for the RNTN The proposed approach outperforms well-known classifiers such as SVM, RAE and LSTM.

The same group in [113] showed how to adapt state-of-the-art English SA systems in order to use it for Arabic SA. Then, they compared this approach with the RNTN approach they proposed in [111]. The first approach used surface, syntactic and semantic features fed into a SVM classifier. As for the second approach, they used their approach from. For evaluation, they used the Arabic Sentiment Twitter Data (ASTD) [102].

In [114], the authors present their Twitter dataset consisting of opinions on health services. After discussing the data collection and filtration process, the authors discussed the pre-processing and annotation processes. For classification, they conduct several experiments using various machine leaning algorithms including DNN and CNN. DL approach showed promising results, but they were outperformed by other classifiers such as SVM.

The authors of [115] focused on SA of tweets datasets that are highly imbalanced. They compared several classifiers fed word embedding features. For evaluation, they used the Syria Tweets dataset[18]

In a follow-up work [116], the same authors proposed to use CNN and LSTM networks with word embeddings features. The proposed systems were evaluated on free datasets and the results showed that the LSTM model is the best.

In a very recent work, the authors of [117] addressed the Aspect based SA (ABSA) problem. They used the Arabic subset of the SemEval-2016 Task 5 [118, 119], which consists of 2,291 hotel reviews. There were mainly three tasks associated with this dataset: aspect category identification, aspect opinion target expression (OTE) extraction, and aspect sentiment polarity identification. For these different tasks, the authors extracted lexical, morphological, syntactic, and semantic features and used them to train a SVM classifier. They compared this classifier with a RNN trained on the same features in addition to word embedding features. The results showed that the SVM classifier produced more accurate results, however, it took much longer time to do so.

SemEval 2017 witnessed more papers employing DL approaches for Arabic tasks. In [120], the authors addressed three subtasks of SemEval-2017 Task 4 [121]. Specifically, they addressed Subtask A (Message Polarity Classification) using an enhanced version of the sentiment analyzer developed by the

[18] http://saifmohammad.com/WebPages/ArabicSA.html

same authors [122]. They also addressed Subtask B (Topic-Based Message Polarity classification) and D (Tweet quantification), for which they proposed to use three classifiers and combine their outcomes using voting. These classifiers are: CNN (trained on word2vec word embeddings), MLP and logistic regression (LR). The results showed that the proposed approaches for the three subtasks under consideration outperformed all of their competitors.

Other works using DL approaches to address the SemEval-2017 Task 4 include [123, 124]. In the former, the authors addressed all five subtasks of the task in question. They used three CRNN, whose inputs were out-/in-domain embeddings and sequences of words polarities. The outputs of the three networks were concatenated and fed into a fully-connected MLP network. The results of the proposed system were very promising.

Like [123], Baly et al. [124] addressed all five subtasks of the task in question. For Subtask A, they employed state-of-the-art approaches for English tweets to analyze Arabic tweets and showed that the results for such an approach were solid. As for Subtasks B-E, they introduce a topic based approach and showed that it was ranked first in Subtasks C and E, and second in Subtask D. One of the systems proposed by [124] uses the RAE of [108, 110].

11. Question Answering (QA)

In [125], the authors addressed the question ranking problem in Arbic community QA (cQA) forums. They used the Farasa toolkit[19] to build a UIMA[20] based processing pipeline that involves a Tree-Kernel (TK) based ranker. Then, they used LSTM networks to choose the question fragments to be used in the ranker. They complement the features they compute with word embeddings features computed using the word2vec tool[107]. They evaluated their approach on the CQA-MD corpus, a medical Arabic corpus that is part of the SemEval 2016

[19]http://farasa.qcri.org/
[20]https://uima.apache.org

Task 3-D [126]. The results show strong performance by the proposed pipeline, which is further enhanced by the use of LSTM networks.

12. Automatic Diacritization

In [127], a deep RNN (specifically, DBLSTM) is used to for automatic diacritization of Arabic text. This approach requires no preprocessing steps such as lexical, morphological and syntactical analysis. However, in order to achieve better results, post processing error correction techniques are employed. The network is built by placing multiple RNN hidden layers on top of each other, where the outputs of Layer L1 in the stack is considered as an input to the above layer (L2). Since it is bidirectional, this means that each hidden layer gets its input from both the forward and backward layers. RNN has been trained in two different ways; One-to-one network and One-to-many network. One-to-one makes sure to encode every possible diacritization of each single letter, and the input and the output sequences have the same length. On the other hand, within the One-to-many network, the input and the output sequences have different lengths (the output is longer than the input). The results showed that one-to-one outperforms one-to-many. So, the authors continued their experiments using one-to-one network. The experiments made using one-to-one network as an improvement of the performance of RNN to get better accuracy. The experiments tested the impact of the following: Weight noise distortion, network size, data size, and influence of the post-processing step. The authors tested their system on 11 books and reported average diacritic and word error rates of 2.09% and 5.82%, respectively. The experiments were conducted using the RNNLIB library [19].

Another work on diacritization is [128], where the authors focus on the use of RNN to build a language-independent system for automatic diacritization. They experimented with different hidden layer types ranging from a single feed-forward layer to DBLSTM layers. They add a linear projection after the input layer and use softmax at the output layer. For the experiments, they use the dataset of [129] which consists of hundreds of thousands of words and millions of letters. On this dataset, the MaxEnt approach of [129] (which utilizes language-dependent resources such as segmenter and part of speech (POS) tagger) produced an error rate of 5.1% whereas the approach of [128] produced an error rate of 4.85%.

In [130, 131], the authors addressed the problem of automatically restoring diacritics. The presented the Confused Subset Resolution (CSR) along with Arabic Part-of-Speech (PoS) tagging using DNN. The proposed system was evaluated on several dataset and the results showed its superiority over state-of-the-art systems.

13. Distributed Word Representations

Across the previous sections, we have mentioned many works [100, 106, 120] employing different distributed word representations techniques such as word2vec [107]. The use of such techniques has been shown to produce interesting results especially with DL based approaches [99]

In [132], the authors presented Polyglot, a word embedding technique for Multilingual NLP. The authors apply their technique on more than 100 languages (including Arabic) based on the Wikipedia pages of each language. In [132], the authors evaluated their word embeddings on part of speech (POS) tagging of few languages (not including Arabic). In [133], they showed how to use Polyglot to address the Named Entity Recognition (NER) problem.

Another effort in the same direction is that of Zahran et al. [134], in which the authors consider different techniques to build vectorized space representations for Arabic. Specifically, the authors built three models: continuous bag of word (CBOW), Skipgram (SKIP-G), and GloVe [135, 136, 137]. They published these mode for public use.[21] The authors studied the impact of these representations on two NLP tasks: Query Expansion for Information Retrieval

[21] https://sites.google.com/site/mohazahran/data

and Short Answer Grading, and the results were encouraging.

The authors of [99] evaluated several Arabic word embedding techniques using their own benchmark,[22] instead of relying on translated benchmarks like previous works. They consider several NLP tasks such as TC and NER.

14. Conclusion

In this survey, we surveyed the literature related to applying DL for ANLP tasks. Despite the huge impact DL has had on the general NLP community, this is yet to happen for ANLP. We expect this to change as more researchers become aware and convinced of the power of DL.

References

[1] A. Farzindar, D. Inkpen, Natural language processing for social media, Synthesis Lectures on Human Language Technologies 8 (2) (2015) 1–166.

[2] K. Stowe, M. J. Paul, M. Palmer, L. Palen, K. Anderson, Identifying and categorizing disaster-related tweets, in: Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media, 2016, pp. 1–6.

[3] K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, Quantifying controversy in social media, in: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, ACM, 2016, pp. 33–42.

[4] A. Nikfarjam, A. Sarker, K. OConnor, R. Ginn, G. Gonzalez, Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features, Journal of the American Medical Informatics Association 22 (3) (2015) 671–681.

[22] `http://oma-project.com/res_home`

[5] D. Yu, L. Deng, D. Yu, Deep learning methods and applications, Foundations and Trends in Signal Processing.

[6] N. Y. Habash, Introduction to arabic natural language processing, Synthesis Lectures on Human Language Technologies 3 (1) (2010) 1–187.

[7] A. Farghaly, K. Shaalan, Arabic natural language processing: Challenges and solutions, ACM Transactions on Asian Language Information Processing (TALIP) 8 (4) (2009) 14.

[8] A. Alwajeeh, M. Al-Ayyoub, I. Hmeidi, On authorship authentication of arabic articles, in: The fifth International Conference on Information and Communication Systems (ICICS 2014), IEEE, 2014, pp. 1–6.

[9] K. Alsmearat, M. Al-Ayyoub, R. Al-Shalabi, G. Kanaan, Author gender identification from arabic text, Journal of Information Security and Applications 35 (2017) 85–95.

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

[11] R. Al-Jawfi, Handwriting arabic character recognition lenet using neural network, Int. Arab J. Inf. Technol. 6 (3) (2009) 304–309.

[12] U. Porwal, Y. Zhou, V. Govindaraju, Handwritten arabic text recognition using deep belief networks, in: Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE, 2012, pp. 302–305.

[13] A. Durou, I. Aref, S. Al-Maadeed, A. Bouridane, E. Benkhelifa, Writer identification approach based on bag of words with obi features, Information Processing & Management.

[14] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, L. D. Jackel, Handwritten digit recognition with a back-propagation network, in: Advances in neural information processing systems, 1990, pp. 396–404.

[15] A. Graves, J. Schmidhuber, Offline handwriting recognition with multidimensional recurrent neural networks, in: Advances in neural information processing systems, 2009, pp. 545–552.

[16] A. Graves, Offline arabic handwriting recognition with multidimensional recurrent neural networks, in: Guide to OCR for Arabic scripts, Springer, 2012, pp. 297–313.

[17] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, H. Amiri, et al., Ifn/enit-database of handwritten arabic words, in: Proc. of CIFED, Vol. 2, 2002, pp. 127–136.

[18] Y. Chherawala, P. P. Roy, M. Cheriet, Feature design for offline arabic handwriting recognition: handcrafted vs automated?, in: 2013 12th International Conference on Document Analysis and Recognition, IEEE, 2013, pp. 290–294.

[19] A. Graves, Rnnlib: a recurrent neural network library for sequence learning problems (2010).
URL http://sourceforge.net/projects/rnnl/

[20] Y. Chherawala, P. P. Roy, M. Cheriet, Feature set evaluation for offline handwriting recognition systems: application to the recurrent neural network model, IEEE transactions on cybernetics 46 (12) (2016) 2825–2836.

[21] E. Grosicki, M. Carre, J.-M. Brodin, E. Geoffrois, Results of the rimes evaluation campaign for handwritten mail processing, in: Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on, IEEE, 2009, pp. 941–945.

[22] J. Chen, H. Cao, R. Prasad, A. Bhardwaj, P. Natarajan, Gabor features for offline arabic handwriting recognition, in: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, ACM, 2010, pp. 53–58.

[23] J. F. G. Srikantan, S. Srihari, Handprinted character/digit recognition using a multiple feature/resolution philosophy, in: Proc. Fourth Int'l Workshop Frontiers in Handwriting Recognition, 1994, pp. 57–66.

[24] A. Khémiri, A. K. Echi, A. Belaïd, M. Elloumi, Arabic handwritten words off-line recognition based on hmms and dbns, in: Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 51–55.

[25] M. Elleuch, N. Tagougui, M. Kherallah, Arabic handwritten characters recognition using deep belief neural networks, in: Systems, Signals & Devices (SSD), 2015 12th International Multi-Conference on, IEEE, 2015, pp. 1–5.

[26] M. Elleuch, N. Tagougui, M. Kherallah, Deep learning for feature extraction of arabic handwritten script, in: International Conference on Computer Analysis of Images and Patterns, Springer, 2015, pp. 371–382.

[27] M. Elleuch, N. Tagougui, M. Kherallah, Towards unsupervised learning for arabic handwritten recognition using deep architectures, in: International Conference on Neural Information Processing, Springer, 2015, pp. 363–372.

[28] M. Elleuch, M. Kherallah, An improved arabic handwritten recognition system using deep support vector machines, International Journal of Multimedia Data Engineering and Management (IJMDEM) 7 (2) (2016) 1–20.

[29] M. Elleuch, N. Tagougui, M. Kherallah, A novel architecture of cnn based on svm classifier for recognising arabic handwritten script, International Journal of Intelligent Systems Technologies and Applications 15 (4) (2016) 323–340.

[30] M. Elleuch, R. Maalej, M. Kherallah, A new design based-svm of the cnn classifier architecture with dropout for offline arabic handwritten recognition, Procedia Computer Science 80 (2016) 1712–1723.

[31] M. Elleuch, R. Mokni, M. Kherallah, Offline arabic handwritten recognition system with dropout applied in deep networks based-svms, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE, 2016, pp. 3241–3248.

[32] M. Elleuch, R. Zouari, M. Kherallah, Feature extractor based deep method to enhance online arabic handwritten recognition system, in: International Conference on Artificial Neural Networks, Springer, 2016, pp. 136–144.

[33] M. Elleuch, N. Tagougui, M. Kherallah, Optimization of dbn using regularization methods applied for recognizing arabic handwritten script, Procedia Computer Science 108 (2017) 2292–2297.

[34] R. Maalej, N. Tagougui, M. Kherallah, Online arabic handwriting recognition with dropout applied in deep recurrent neural networks, in: Document Analysis Systems (DAS), 2016 12th IAPR Workshop on, IEEE, 2016, pp. 417–421.

[35] R. Maalej, N. Tagougui, M. Kherallah, Recognition of handwritten arabic words with dropout applied in mdlstm, in: International Conference Image Analysis and Recognition, Springer, 2016, pp. 746–752.

[36] R. Maalej, M. Kherallah, Improving mdlstm for offline arabic handwriting recognition using dropout at different positions, in: International Conference on Artificial Neural Networks, Springer, 2016, pp. 431–438.

[37] N. Tagougui, M. Kherallah, Recognizing online arabic handwritten characters using a deep architecture, in: Ninth International Conference on Machine Vision, International Society for Optics and Photonics, 2017, pp. 103410L–103410L.

[38] A. Lawgali, M. Angelova, A. Bouridane, Hacdb: Handwritten arabic characters database for automatic character recognition, in: Visual Information Processing (EUVIP), 2013 4th European Workshop on, IEEE, 2013, pp. 255–259.

[39] M. Kherallah, N. Tagougui, A. M. Alimi, H. El Abed, V. Margner, Online arabic handwriting recognition competition, in: Document Analysis and Recognition (ICDAR), 2011 International Conference on, IEEE, 2011, pp. 1454–1458.

[40] D. Yu, M. L. Seltzer, Improved bottleneck features using pretrained deep neural networks, in: Twelfth Annual Conference of the International Speech Communication Association, 2011.

[41] T. N. Sainath, B. Kingsbury, B. Ramabhadran, Auto-encoder bottleneck features using deep belief networks, in: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, IEEE, 2012, pp. 4153–4156.

[42] H. Boubaker, A. Elbaati, N. Tagougui, H. El Abed, M. Kherallah, A. M. Alimi, Online arabic databases and applications, in: Guide to OCR for Arabic Scripts, Springer, 2012, pp. 541–557.

[43] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580.

[44] L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, R. Fergus, Regularization of neural networks using dropconnect, in: Proceedings of the 30th international conference on machine learning (ICML-13), 2013, pp. 1058–1066.

[45] X.-X. Niu, C. Y. Suen, A novel hybrid cnn–svm classifier for recognizing handwritten digits, Pattern Recognition 45 (4) (2012) 1318–1325.

[46] S. Kim, S. Kavuri, M. Lee, Deep network with support vector machines, in: International Conference on Neural Information Processing, Springer, 2013, pp. 458–465.

[47] M. Liwicki, A. Graves, H. Bunke, J. Schmidhuber, A novel approach to on-line handwriting recognition based on bidirectional long short-term

memory networks, in: Proc. 9th Int. Conf. on Document Analysis and Recognition, Vol. 1, 2007, pp. 367–371.

[48] T. Bluche, J. Louradour, M. Knibbe, B. Moysset, M. F. Benzeghiba, C. Kermorvant, The a2ia arabic handwritten text recognition system at the open hart2013 evaluation, in: Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on, IEEE, 2014, pp. 161–165.

[49] A. Tong, M. Przybocki, V. Margner, H. El Abed, Nist 2013 open handwriting recognition and translation (open hart'13) evaluation, in: Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on, IEEE, 2014, pp. 81–85.

[50] V. Pham, T. Bluche, C. Kermorvant, J. Louradour, Dropout improves recurrent neural networks for handwriting recognition, in: Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on, IEEE, 2014, pp. 285–290.

[51] I. Guyon, G. Dror, V. Lemaire, G. Taylor, D. W. Aha, Unsupervised and transfer learning challenge, in: The 2011 International Joint Conference on Neural Networks, 2011, pp. 793–800.

[52] G. Mesnil, Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D. Warde-Farley, et al., Unsupervised and transfer learning challenge: a deep learning approach, Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7 (2012) 109.

[53] A. ElAdel, R. Ejbali, M. Zaied, C. B. Amar, Dyadic multi-resolution analysis-based deep learning for arabic handwritten character classification, in: Tools with Artificial Intelligence (ICTAI), 2015 IEEE 27th International Conference on, IEEE, 2015, pp. 807–812.

[54] M. Elzobi, A. Al-Hamadi, Z. Al Aghbari, L. Dings, Iesk-ardb: a database for handwritten arabic and an optimized topological segmentation ap-

proach, International Journal on Document Analysis and Recognition (IJDAR) 16 (3) (2013) 295–308.

[55] A. El-Sawy, M. Loey, E.-B. Hazem, Arabic handwritten characters recognition using convolutional neural network, WSEAS TRANSACTIONS on COMPUTER RESEARCH 5 (2017) 11–19.

[56] S. F. Rashid, M.-P. Schambach, J. Rottland, S. von der Nüll, Low resolution arabic recognition with multidimensional recurrent neural networks, in: Proceedings of the 4th International Workshop on Multilingual OCR, ACM, 2013, p. 6.

[57] A. Graves, S. Fernández, J. Schmidhuber, Multi-dimensional recurrent neural networks., in: ICANN (1), 2007, pp. 549–558.

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

[59] S. Yousfi, S.-A. Berrani, C. Garcia, Arabic text detection in videos using neural and boosting-based approaches: Application to video indexing, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE, 2014, pp. 3028–3032.

[60] S. Yousfi, S.-A. Berrani, C. Garcia, Alif: A dataset for arabic embedded text recognition in tv broadcast, in: Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 1221–1225.

[61] S. Yousfi, S.-A. Berrani, C. Garcia, Deep learning and recurrent connectionist-based approaches for arabic text recognition in videos, in: Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 1026–1030.

[62] S. Yousfi, S.-A. Berrani, C. Garcia, Contribution of recurrent connectionist language models in improving lstm-based arabic text recognition in videos, Pattern Recognition 64 (2017) 245–254.

[63] M. Delakis, C. Garcia, text detection with convolutional neural networks., in: VISAPP (2), 2008, pp. 290–294.

[64] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, J. Schmidhuber, A novel connectionist system for unconstrained handwriting recognition, IEEE transactions on pattern analysis and machine intelligence 31 (5) (2009) 855–868.

[65] A. Graves, N. Jaitly, Towards end-to-end speech recognition with recurrent neural networks, in: Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014, pp. 1764–1772.

[66] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, S. Khudanpur, Recurrent neural network based language model., in: Interspeech, Vol. 2, 2010, p. 3.

[67] M. Jain, M. Mathew, C. Jawahar, Unconstrained scene text and video text recognition for arabic script, in: Arabic Script Analysis and Recognition (ASAR), 2017 1st International Workshop on, IEEE, 2017, pp. 26–30.

[68] O. Zayene, J. Hennebert, S. M. Touj, R. Ingold, N. E. B. Amara, A dataset for arabic text detection, tracking and recognition in news videos-activ, in: Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 996–1000.

[69] L. Gomez, D. Karatzas, A fine-grained approach to scene text script identification, in: Document Analysis Systems (DAS), 2016 12th IAPR Workshop on, IEEE, 2016, pp. 192–197.

[70] L. Gomez, A. Nicolaou, D. Karatzas, Improving patch-based scene text script identification with ensembles of conjoined networks, Pattern Recognition 67 (2017) 85–96.

[71] N. Sharma, R. Mandal, R. Sharma, U. Pal, M. Blumenstein, Icdar2015 competition on video script identification (cvsi 2015), in: Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, IEEE, 2015, pp. 1196–1200.

[72] B. Shi, X. Bai, C. Yao, Script identification in the wild via discriminative convolutional neural network, Pattern Recognition 52 (2016) 448–458.

[73] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazan, L. P. de las Heras, Icdar 2013 robust reading competition, in: Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, IEEE, 2013, pp. 1484–1493.

[74] J. Zdenek, H. Nakayama, Script identification using bag-of-words with entropy-weighted patches, in: The 31st Annual Conference of the Japanese Society for Artificial Intelligence, 2017.

[75] S. B. Ahmed, S. Naz, M. I. Razzak, R. Yousaf, Deep learning based isolated arabic scene character recognition, arXiv preprint arXiv:1704.06821.

[76] V. Jindal, A deep learning approach for arabic caption generation using roots-words., in: AAAI, 2017, pp. 4941–4942.

[77] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 248–255.

[78] Y. A. Alotaibi, Spoken arabic digits recognizer using recurrent neural networks, in: Signal Processing and Information Technology, 2004. Proceedings of the Fourth IEEE International Symposium on, IEEE, 2004, pp. 195–199.

[79] J. L. Elman, Finding structure in time, Cognitive science 14 (2) (1990) 179–211.

[80] B. Kingsbury, T. N. Sainath, H. Soltau, Scalable minimum bayes risk training of deep neural network acoustic models using distributed hessian-free optimization., in: Interspeech, 2012, pp. 10–13.

[81] A. E.-D. Mousa, H.-K. J. Kuo, L. Mangu, H. Soltau, Morpheme-based feature-rich language models using deep neural networks for lvcsr of egyptian arabic, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 8435–8439.

[82] T. AlHanai, W.-N. Hsu, J. Glass, Development of the mit asr system for the 2016 arabic multi-genre broadcast challenge, in: Spoken Language Technology Workshop (SLT), 2016 IEEE, IEEE, 2016, pp. 299–304.

[83] P. Cardinal, A. Ali, N. Dehak, Y. Zhang, T. A. Hanai, Y. Zhang, J. R. Glass, S. Vogel, Recent advances in asr applied to an arabic transcription system for al-jazeera, in: Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[84] A. Ali, Y. Zhang, P. Cardinal, N. Dahak, S. Vogel, J. Glass, A complete kaldi recipe for building arabic speech recognition systems, in: Spoken Language Technology Workshop (SLT), 2014 IEEE, IEEE, 2014, pp. 525–529.

[85] S. Thomas, G. Saon, H.-K. J. Kuo, L. Mangu, The ibm bolt speech transcription system, in: Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[86] S. Khurana, A. Ali, Qcri advanced transcription system (qats) for the arabic multi-dialect broadcast media recognition: Mgb-2 challenge, in: Spoken Language Technology Workshop (SLT), 2016 IEEE, IEEE, 2016, pp. 292–298.

[87] N. Tomashenko, K. Vythelingum, A. Rousseau, Y. Estève, Lium asr systems for the 2016 multi-genre broadcast arabic challenge, in: Spoken Language Technology Workshop (SLT), 2016 IEEE, IEEE, 2016, pp. 285–291.

[88] Y. Kim, Y. Jernite, D. Sontag, A. M. Rush, Character-aware neural language models., in: AAAI, 2016, pp. 2741–2749.

[89] S. Harrat, K. Meftouh, K. Smaïli, Machine translation for arabic dialects (survey), Information Processing & Management.

[90] T. Deselaers, S. Hasan, O. Bender, H. Ney, A deep learning approach to machine transliteration, in: Proceedings of the Fourth Workshop on Statistical Machine Translation, Association for Computational Linguistics, 2009, pp. 233–241.

[91] F. Guzmán, H. Bouamor, R. Baly, N. Habash, Machine translation evaluation for arabic using morphologically-enriched embeddings, in: COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 1398–1408.

[92] S. Malmasi, M. Zampieri, N. Ljubešic, P. Nakov, A. Ali, J. Tiedemann, Discriminating between similar languages and arabic dialect identification: A report on the third dsl shared task, VarDial 3 (2016) 1.

[93] Y. Belinkov, J. Glass, A character-level convolutional neural network for distinguishing similar languages and dialects, arXiv preprint arXiv:1609.07568.

[94] C. Guggilla, Discrimination between similar languages, varieties and dialects using cnn-and lstm-based deep neural networks, VarDial 3 (2016) 185.

[95] Y. Samih, M. Attia, M. Eldesouki, A. Abdelali, H. Mubarak, L. Kallmeyer, K. Darwish, A neural architecture for dialectal arabic segmentation, in: Proceedings of the Third Arabic Natural Language Processing Workshop, 2017, pp. 46–54.

[96] Y. Yao, Z. Huang, Bi-directional lstm recurrent neural network for chinese word segmentation, in: International Conference on Neural Information Processing, Springer, 2016, pp. 345–353.

[97] Y. Samih, M. Eldesouki, M. Attia, K. Darwish, A. Abdelali, H. Mubarak, L. Kallmeyer, Learning from relatives: Unified dialectal arabic segmentation, in: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), 2017, pp. 432–441.

[98] V. Jindal, A personalized markov clustering and deep learning approach for arabic text categorization, ACL 2016 (2016) 145.

[99] M. Elrazzaz, S. Elbassuoni, K. Shaban, C. Helwe, Methodical evaluation of arabic word embeddings, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vol. 2, 2017, pp. 454–458.

[100] A. Dahou, S. Xiong, J. Zhou, M. H. Haddoud, P. Duan, Word embeddings and convolutional neural network for arabic sentiment classification, in: COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 2418–2427.

[101] M. A. Aly, A. F. Atiya, Labr: A large scale arabic book reviews dataset., in: ACL (2), 2013, pp. 494–498.

[102] M. Nabil, M. A. Aly, A. F. Atiya, Astd: Arabic sentiment tweets dataset., in: EMNLP, 2015, pp. 2515–2519.

[103] E. Refaee, V. Rieser, An arabic twitter corpus for subjectivity and sentiment analysis., in: LREC, 2014, pp. 2268–2273.

[104] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, M. Al-Ayyoub, Arabic sentiment analysis: Lexicon-based and corpus-based, in: Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on, IEEE, 2013, pp. 1–6.

[105] H. ElSahar, S. R. El-Beltagy, Building large arabic multi-domain resources for sentiment analysis., in: CICLing (2), 2015, pp. 23–34.

[106] A. A. Altowayan, L. Tao, Word embeddings for arabic sentiment analysis, in: Big Data (Big Data), 2016 IEEE International Conference on, IEEE, 2016, pp. 3820–3825.

[107] T. Mikolov, W.-t. Yih, G. Zweig, Linguistic regularities in continuous space word representations., in: hlt-Naacl, Vol. 13, 2013, pp. 746–751.

[108] A. A. Al Sallab, R. , G. Badaro, H. Hajj, W. El Hajj, K. B. Shaban, Deep learning models for sentiment analysis in arabic, in: ANLP Workshop 2015, 2015, p. 9.

[109] G. Badaro, R. Baly, H. Hajj, N. Habash, W. El-Hajj, A large scale arabic sentiment lexicon for arabic opinion mining, ANLP 2014 165.

[110] A. Al-Sallab, R. Baly, H. Hajj, K. B. Shaban, W. El-Hajj, G. Badaro, Aroma: A recursive deep learning model for opinion mining in arabic as a low resource language, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 16 (4) (2017) 25.

[111] R. Baly, H. Hajj, N. Habash, K. B. Shaban, W. El-Hajj, A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in arabic, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 16 (4) (2017) 23.

[112] A. Pasha, M. Al-Badrashiny, M. T. Diab, A. El Kholy, R. Eskander, N. Habash, M. Pooleery, O. Rambow, R. Roth, Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic., in: LREC, Vol. 14, 2014, pp. 1094–1101.

[113] R. Baly, G. Badaro, G. El-Khoury, R. Moukalled, R. Aoun, H. Hajj, W. El-Hajj, N. Habash, K. B. Shaban, A characterization study of arabic twitter data with a benchmarking for state-of-the-art opinion mining models, WANLP 2017 (co-located with EACL 2017) (2017) 110.

[114] A. M. Alayba, V. Palade, M. England, R. Iqbal, Arabic language sentiment analysis on health services, arXiv preprint arXiv:1702.03197.

[115] S. Al-Azani, E.-S. M. El-Alfy, Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short arabic text, Procedia Computer Science 109 (2017) 359–366.

[116] S. Al-Azani, E.-S. M. El-Alfy, Hybrid deep learning for sentiment polarity determination of arabic microblogs, in: International Conference on Neural Information Processing, Springer, 2017, pp. 491–500.

[117] M. AL-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, B. Gupta, Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels reviews, Journal of Computational ScienceTo appear.

[118] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, G. Eryiğit, Semeval-2016 task 5 : aspect based sentiment analysis, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, 2016, pp. 19–30.
URL http://www.aclweb.org/anthology/S16-1002

[119] M. AL-Smadi, O. Qwasmeh, B. Talafha, M. Al-Ayyoub, Y. Jararweh, E. Benkhelifa, An enhanced framework for aspect-based sentiment analysis of hotels' reviews: Arabic reviews case study, in: Internet Technology and Secured Transactions (ICITST), 2016 11th International Conference for, IEEE, 2016, pp. 98–103.

[120] S. R. El-Beltagy, M. E. Kalamawy, A. B. Soliman, Niletmrg at semeval-2017 task 4: Arabic sentiment analysis, arXiv preprint arXiv:1710.08458.

[121] S. Rosenthal, N. Farra, P. Nakov, Semeval-2017 task 4: Sentiment analysis in twitter, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 502–518.

[122] S. R. El-Beltagy, T. Khalil, A. Halaby, M. Hammad, Combining lexical features and a supervised learning approach for arabic sentiment analysis, arXiv preprint arXiv:1710.08451.

[123] J.-A. González, F. Pla, L.-F. Hurtado, Elirf-upv at semeval-2017 task 4: Sentiment analysis using deep learning, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 723–727.

[124] R. Baly, G. Badaro, A. Hamdi, R. Moukalled, R. Aoun, G. El-Khoury, A. Al Sallab, H. Hajj, N. Habash, K. Shaban, et al., Omam at semeval-2017 task 4: Evaluation of english state-of-the-art sentiment analysis models for arabic and a new topic-based model, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 603–610.

[125] S. Romeo, G. Da San Martino, Y. Belinkov, A. Barrón-Cedeño, M. Eldesouki, K. Darwish, H. Mubarak, J. Glass, A. Moschitti, Language processing and learning models for community question answering in arabic, Information Processing & Management.

[126] P. Nakov, D. Hoogeveen, L. Màrquez, A. Moschitti, H. Mubarak, T. Baldwin, K. Verspoor, Semeval-2017 task 3: Community question answering, in: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 27–48.

[127] G. A. Abandah, A. Graves, B. Al-Shagoor, A. Arabiyat, F. Jamour, M. Al-Taee, Automatic diacritization of arabic text using recurrent neural networks, International Journal on Document Analysis and Recognition (IJDAR) 18 (2) (2015) 183–197.

[128] Y. Belinkov, J. Glass, Arabic diacritization with recurrent neural networks., in: EMNLP, 2015, pp. 2281–2285.

[129] I. Zitouni, R. Sarikaya, Arabic diacritic restoration approach based on maximum entropy models, Computer Speech & Language 23 (3) (2009) 257–276.

[130] M. A. Rashwan, A. A. Al Sallab, H. M. Raafat, A. Rafea, Automatic arabic diacritics restoration based on deep nets, in: ANLP 2014, 2014, p. 65.

[131] M. A. Rashwan, A. A. Al Sallab, H. M. Raafat, A. Rafea, Deep learning framework with confused sub-set resolution architecture for automatic arabic diacritization, IEEE Transactions on Audio, Speech, and Language Processing 23 (3) (2015) 505–516.

[132] R. Al-Rfou, B. Perozzi, S. Skiena, Polyglot: Distributed word representations for multilingual nlp, arXiv preprint arXiv:1307.1662.

[133] R. Al-Rfou, V. Kulkarni, B. Perozzi, S. Skiena, Polyglot-ner: Massive multilingual named entity recognition, in: Proceedings of the 2015 SIAM International Conference on Data Mining, SIAM, 2015, pp. 586–594.

[134] M. A. Zahran, A. Magooda, A. Y. Mahgoub, H. M. Raafat, M. Rashwan, A. Atyia, Word representations in vector space and their applications for arabic., in: CICLing (1), 2015, pp. 430–443.

[135] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781.

[136] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in neural information processing systems, 2013, pp. 3111–3119.

[137] J. Pennington, R. Socher, C. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.