ASLR: Arabic Sign Language Recognition Using Convolutional Neural Networks

Asma Althagafi[1†], Ghofran Althobaiti[2†] Tahani Alsubait[3†*], Tahani Alqurashi[4††]

[†]College of Computer and Information systems, Umm Al-Qura University, Saudi Arabia
[††]Common First Year Deanship, Computer Science Department, Umm Al-Qura University, Makkah, Saudi Arabia

Abstract
Sign language is the native language of deaf and hearing-impaired people which they use on their daily life. Few interpreters are available to facilitate communication between deaf and vocal people. However, this is neither practical nor possible for all situations. Advances in information technology encouraged the development of systems that can facilitate the automatic translation between sign language and spoken language, and thus removing barriers facing the integration of deaf people in society.
Objective: The main objective of this paper is to present an Arabic Sign Language recognition system that automatically recognizes 28 letters using a CNN model taking RGB images as input.
Methods: In this work, we propose a new framework based on convolutional neural networks, fed with a real dataset, which will automatically recognize letters of Arabic Sign Language. In order to validate our scheme, we have performed a comparative analysis that demonstrates the efficacy and robustness of our proposed method compared to conventional methods.
Results: we tested the CNN model on (10810 images), we measured the accuracy of the model as well as the error rate, the accuracy increases, the error decreases during the training and testing phases, and we achieved 92.9% of recognition accuracy.
Key words:
Arabic Sign Language Convolutional Neural Networks CNN

1. Introduction

Deaf and hard hearing people make up a significant number in the Arab world and 70 million around the world. They use the sign language as the main contact language. Sign language varies from country to country and even within the same region. There are several sign languages in Arabic countries, such as Egyptian, Jordanian, Saudi, and Yemeni [1]. The Arab States League (LAS) in 1999 and the Arab League Educational, Cultural and Science Organization (ALECSO) has standardized the Arabic Sign Language (ArSL) and has written a dictionary of about 3200 words [2]. ArSL language is used in almost all Arabic countries and mainly in countries of the Arab Gulf such as Qatar and the United Arab Emirates. But instead of using this language to replace some countries national language, they combine it with their local sign language. Most TV channels in Arabic countries currently use ArSL to translate their news and programs.

The sign language consists of a set of manual or nongestural used for communication between deaf people in particular. Most sign language gestures are manual gestures, while non-manual gestures such as head movements, body movements and facial expressions also play an important role in sign communications. Unfortunately, the use of sign language is usually confined to the deaf community, resulting in limited communication with the rest of the world. As a result, they are usually excluded from society and deprived of their right to equal educational and career opportunities.

Arabic Sign Language Recognition (ASLR) systems are designed to translate this language into speech or text in order to address this issue. Developing efficient SLR systems can make it easier for deaf people to communicate in society and remove barriers for them. As in the oral language, sign language is not universal; it varies by country or even by region. Sign language has recently been recognized and documented in the Arab world. A great deal of effort has been made to establish the sign language used in individual countries, including Jordan, Egypt and the Gulf States; In an attempt to standardize the language and spread it among the members of the deaf community and those concerned. Such efforts have produced many sign languages, almost as many as Arabic-speaking countries, but with the same sign alphabets. Gestures are used in Arabic Sign Language Alphabets. Arabic sign language (ASL) is a spoken, visual and non-verbal language. Moreover, both the sign language and the spoken language have the same features [3].

SLs are used by deaf, hard-hearing, dumb people. This helps them to communicate with others through signs and gestures. In addition, ASL is not a new innovation; it occurs in the same way as verbal dialects. Sign languages are naturally formed in each area, as other verbal dialects, as independent communication systems. They cannot be traced to anyone. As a result, the growing deaf population is making it important to set up automated systems.

Manual and non-manual signs are the key components that make up sign languages. Manual signs are hand location, orientation, form, and direction. Non-manual stands for

Manuscript received July 7, 2020
Manuscript revised July 20, 2020

body motion and facial expressions [4]. However, some researches such as [3] and [4] concentrate on manual signs as they provide the key information, which helps non-manual signs to explain and highlight the significance of manual signs.

In this paper, we are going to concentrate on the manual part in our job through developing an Arabic sign language recognition system that automatically recognizes 28 letters using a CNN model fed with RGB images. The rest of this paper is organized as follows: Section II introduces the related works with a comparison on related works in the area of Arabic sing language. Section III introduces our methods and materials used in this research. Section IV exposes our results and discussions. Finally, Section V presents the conclusion and future work.

2. Related Work and Comparison of Related Work

2.1 Related Work

In order to help vocal people who do not know sign language, communicate with deaf and hearing-impaired people, a lot of research has been done on developing systems for different sign languages from around the word. Mohandes et al. [5] presented a review of the developed system for the recognition of Arabic alphabet signs using the newly introduced leap motion controller (LMC). Ten samples of each of the 28 Arabic alphabet signs were collected from a single signatory. Ten frames were taken from each sample letter sign to provide a total of 2,800 data frames. Twelve features were selected from the 23 values provided by the LMC for the representation of each frame in the LMC coverage area. For classification, they compared the performance of the Nave Bayes Classifier (NBC) with the Multilayer Perceptron (MLP) of the back-propagation algorithm. The overall accuracy of the sign recognition using the NBC was approximately 98.3% while the accuracy of the MLP was approximately 99.1%.

Aly and Mohammed [6] presented another contribution in the field of the sign language made by Basma Hisham and Alaa Hamouda from Al-Azhar University, Egypt. They propose a model to recognize static gestures like numbers, letters, ...etc and dynamic gestures such as movement and motion in performing the signs. Moreover, they propose a segmentation method in order to segment a sequence of continuous signs in real time based on tracking the palm velocity, which is useful in translating pre-segmented signs and continuous sentences. They used Leap Motion controller device, which detects and tracks the hands' and fingers' motion and position in an accurate manner. The proposed model applies several machine learning algorithms as Support Vector Machine (SVM), K- Nearest Neighbour (KNN), Artificial Neural Network (ANN) and Dynamic Time Wrapping (DTW) depending on two different features sets. However, researchers assume that this study will increase the chance for the Arabic hearing-impaired and deaf persons to communicate easily using Arabic Sign language (ArSLR).

The proposed model works as an interface between hearing-impaired and normal persons to overcome the gap between them.

Learning Arabic Sign Language (ARSL) has been identified as a difficult process. This is what inspired Hisham and Hamouda [7] to propose a smart tutoring system for the Arabic Sign Language (ArSL Tutor). The aim is to provide a learning platform that supports learners of ArSL by allowing them to practice, get instant feedback and self-assess their learning. Sign recognition based on a series of features was implemented using the KNN algorithm. System architecture and design were provided in this study together with an overview of the accepted methodology for sign recognition. The system was assessed in terms of user acceptance based on the Technology Acceptance Model. The results showed a satisfactory level of acceptance by users and a willingness to use the system.

Hisham and Hamouda [7] proposed a system for automatic translation of Arabic Sign Language to Arabic Text (ATASAT) System. This system acts as a translator between deaf and dumb from one side and the normal people from the other side to enhance their communication. The proposed System consists of five main stages Video and Images capture, Video and images processing, Hand Signs Construction, Classification, and Text transformation and interpretation. This system depends on building two datasets for Arabic sign language gestures alphabets from two resources: Arabic Sign Language dictionary and gestures from different signer's human, also using gesture recognition techniques, which allows the user to interact with the outside world. This system offers a novel technique of hand detection, which detect and extract hand gestures of Arabic Sign from Image or video.

The researcher used a set of appropriate features in setting up hand sign construction and classification based on different classification algorithms such as KNN, MLP, C4.5, VFI and SMO.

Satori et al. [8] investigated the speech recognition problem regarding Arabic Language. They proposed a new approach to build an Arabic Automated Speech Recognition System (ASR). This system is based on an open source CMU Sphinx-4, from the Carnegie Mellon University. CMU Sphinx is a large-vocabulary; speaker-independent, continuous speech recognition system based on discrete Hidden Markov Models (HMMs). They built a model using utilities from the Open Source CMU Sphinx.

They demonstrated the possible adaptability of this system to Arabic voice recognition.

Satori et al. [9] presented one more study, which was conducted at the Department of Electrical Engineering, King Fahd University of Petroleum & Minerals, Saudi Arabia. In this paper, the researchers proposed an image-based system for Arabic Sign Language (ArSL) recognition. The algorithm starts by detecting the face of the signer using a model of Gaussian skin color. The centroid of the detected face is then used as a reference point for tracking the hands' movements. The hands regions are segmented using a region-growing algorithm assuming the signer wears a yellow and an orange-colored gloves. From the segmented hands regions, an optimal set of features is extracted. To represent the time varying feature patterns, a Hidden Markov Model (HMM) is then used. Before using HMM in testing, the number of states and the number of Gaussian mixtures are optimized. The proposed system was implemented for both signer dependent and signer independent conditions. The experimental results show that an accuracy of more than 95% can be achieved with a large database of 300 signs. The results outperform previous work on ArSL mainly restricted to small vocabulary size.

Kurdyumov and Ng [10] have built a program to provide instant feedback for those studying the American Sign Language, users only need to use their computer webcams to practice the American Sign Language, and the program tells them how well they perform the gesture and how they can improve it. For feature extraction, they normalized and scaled their gesture images to 20 x 20px and used pixels as their features, used K-NN and SVM for classification, and found that SVM had more classification accuracy than K-NN, while SVM had an accuracy of about 93%, which surpassed K-NN by 10%.

Hayani et al. [3] presented a new system based on convolutional neural networks, fed with a real dataset, which automatically recognize the numbers and letters of the Arabic sign language. In order to validate the framework, a comparative study has been conducted which shows the effectiveness and robustness of the proposed method compared to conventional approaches based on K-nearest neighbors (KNN) and support vector machines (SVM).

Elsayed and Fathy [11] presented a solution to this problem by offering an Arabic sign language translation program that uses ontology and deep learning techniques. This is to translate the sign of the user to various meanings. This paper has introduced a sign language domain ontology to address some of the sign language problems. Deep Convolution Neural Network (CNN) architecture has been trained and tested on the pre-made Arabic sign language dataset and the dataset collected in this paper to obtain better recognition accuracy. The accuracy of the training set (80% of the dataset) was 98.06% and the accuracy of the test set (20% of the dataset) was 88.87%.

2.2 Comparison of Related Work

Overall, we reviewed several recent studies highly related to ours. Some of these studies as shown in Table I applied Arabic Sign Language Recognition on images to extract the features by using many methods:

Table 1: Comparison of Related Work

Studies	***Main Method***	**Accuracy %**
H. Satori[9] study	Hidden Markov Model (HMM)	95%
Ruslan Kurdyumov[10]study	support vector machines (SVM)	93%
Hayani,[3]study	support vector machines (SVM)	90.02%
Elsayed,[11]study	Convolution Neural Network (CNN)	88.87%

3. Methods and Materials

3.1 Dataset:

We used an image dataset called ArSL2018 [12] to validate our approach. The ArSL2018 is a modern, extensive, completely labeled dataset of Arabic Sign Language Images launched at Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia, to be made available to Machine Learning and Deep Learning researchers. It is useful for the advancement of technologies and tools in the area of assistive technology for the support of deaf and hard-to hearing individuals.

The ArSL2018 dataset consists of 54,049 grayscale images with a 6464 dimension. Variations of photographs have been implemented with specific lighting and context. Figure 1 displays a selection of the Arabic sign language and alphabets in the dataset.

Fig. 1 ArSL2018 image dataset

3.2 Methods

- Preprocessing Steps:

Digital image processing is the use of computer algorithms for digital image processing. As a sub-field of digital signal processing, digital image processing has many advantages over analog image processing. This allows a much broader variety of algorithms to be used for input data. The purpose of image processing is to enhance image data (features), filter for noise removal, reduce ray disruption, and so on.

- Image Processing Steps:

1) **Read images:** We stored the path of our image dataset in a variable, and then created a function to load folders containing images into arrays by importing the libraries that we will use.
2) **Remove noise:** We used noise removal to smooth our image to remove possible noise.
3) **Converting:** All pictures have been converted to a grayscale before being fed into a script.
4) Resizing: Each image was resized to a common resolution of 64 64 pixels.
5) **Normalization**: is a process that changes the range of pixel intensity values, so we normalised the range of pixel from 0-255 to 0-1 by dividing on 255.

- Create CNN Architecture:

Convolutional neural networks (CNN), first demonstrated by Fukushima [13], which have wide-ranging applications for behavior recognition, sentence classification, text recognition, face recognition, object detection and image characterization, etc. A Convolutional Neural Network (CNN) is a Neural network algorithm designed to process visual inputs and perform tasks such as image recognition, segmentation and object detection that can be useful for autonomous vehicles. CNNs is the most popular type of neural networks, especially in high-dimensional data, such as images and videos. It is a multi-layer neural network (NN) architecture, induced by visual cortex neurobiology, which includes a convolutional layer(s) accompanied by a fully connected (FC) layer(s).
The biggest benefit of CNN over its predecessors is that it automatically identifies essential features without any human supervision. CNN is also computationally efficient and achieves high accuracy [14].

We created CNN architecture with several layers:
1) Sequential: it's a linear stack of layers, used to construct deep learning models like the Sequential Model case, and layers will be added to it piecewise.
2) Conv2D: This 2D convolution layer basically creates a filter for the kernel that is convolved with the input layer (at the initial level with the image) and then creates a new array.
3) MaxPooling2D: This layer conducts downsampling on the input of this layer, which means that it compresses the input image with more information into less detailed images.
4) Flatten: This flattens layer, which flattens the inputs, will be added after a series of convolutional and pooling layers, followed by a dense, fully connected layer.
5) Dropout: Dropout is applied to the data. This is a technique used to prevent the model from being overfitting by making the weights of certain redundant neurons in a particular layer equal to zero.
We trained the CNN architecture model on 43239 images, which is about the 80% of the dataset. The CNN architecture was built with 300 epochs and two layers of Dropout (0.1) after the Activation function ('ReLU'), and one more Dense (128) layer before the Dense(32) layer. The accuracy of the CNN model on this training set was 0.96%.

3.3 Experimental Environment

Our CNN model is trained on OS: Linux 4.15.0-96-generic
97-Ubuntu , 8GB RAM computer and CPU:

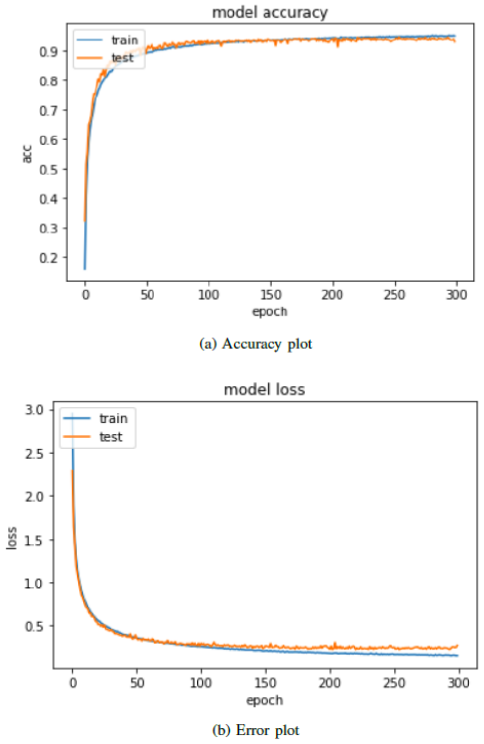
- Architecture:x86 64.
- CPU(s): 4.
- Model name: Intel Core i7-4510U CPU @ 2.00GHz.

The model is trained with a total of 43239 images after feature extraction, it takes an average of 5.96 minutes to complete training. And the model testing with a total of

10810 images takes an average of 4.37 seconds to complete testing.

4. Results and Discussions

We tested the CNN model on (10810 images), which is about the 20% of the dataset and we measured the accuracy of the model as well as the error.
Figure 2 illustrates the accuracy and error of the training and the test datasets. As we can see, the accuracy increases and the error decreases during the training and testing phases. Precisely, we achieved 92.9% of recognition accuracy. This research is inspired by the work of Sign Language Semantic Translation System using Ontology and Deep Learning [11] in which the researchers presented the results of a case study on Arabic Sign Language Recognition by used ArSL2018 dataset. They used 42960 images, which is about 80% of ArSL2018 dataset, to training the model of CNN architecture. The accuracy of this training dataset was 98.06%. Then they tested their CNN architecture on 11089 images, which is about 20% of the ArSL2018 dataset. And They achieved an accuracy of 88.87%. In contrast, we have achieved a higher accuracy, due to that our model was built with 300 training epochs and two layers of Dropout(0.1) and used the Activation function ('ReLU'), and one more Dense(128) layers before the Dense(32) layer.



(a) Accuracy plot

(b) Error plot

Fig. 2 Accuracy and Error for the training and test datasets.

5. Conclusion and Future Work

Error detection and correction are required to achieve better recognition accuracy. The method of fusion recognition is important to incorporate a system that is closer to real life.
In order to make contact between the deaf and the hearing simpler, it is also important to use the opposite way, to translate speech or text into the Arabic sign language. The biggest challenge is how to boost recognition accuracy with adequate processing time for real-life applications. This paper presented a develop an Arabic sign language recognition system that automatically recognizes 28 letters using a CNN model feed with RGB images.

References

[1] K. Al-Fityani and C. Padden, "Sign language geography in the arab world," Sign languages: A Cambridge survey, vol. 20, 2010.
[2] H. Luqman, S. A. Mahmoud, et al., "Transform-based arabic sign language recognition," Procedia Computer Science, vol. 117, pp. 2–9, 2017.
[3] S. Hayani, M. Benaddy, O. El Meslouhi, and M. Kardouchi, "Arab sign language recognition with convolutional neural networks," in 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), pp. 1–4, IEEE, 2019.
[4] H. Cooper, B. Holt, and R. Bowden, "Sign language recognition," in Visual Analysis of Humans, pp. 539–562, Springer, 2011.
[5] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," in 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE), pp. 960–965, IEEE, 2014.
[6] S. Aly and S. Mohammed, "Arabic sign language recognition using spatio-temporal local binary patterns and support vector machine," in International Conference on Advanced Machine Learning Technologies and Applications, pp. 36–45, Springer, 2014.
[7] B. Hisham and A. Hamouda, "Arabic static and dynamic gestures recognition using leap motion.," JCS, vol. 13, no. 8, pp. 337–354, 2017.
[8] H. Satori, H. Hiyassat, M. Haiti, and N. Chenfour, "Investigation arabic speech recognition using cmu sphinx system.," International Arab Journal of Information Technology (IAJIT), vol. 6, no. 2, 2009.
[9] H. Satori, M. Harti, and N. Chenfour, "Introduction to arabic speech recognition using cmusphinx system," arXiv preprint arXiv:0704.2083, 2007.
[10] R. Kurdyumov, P. Ho, and J. K. Ng, "Sign language classification using webcam images," 2011.
[11] E. K. Elsayed and D. R. Fathy, "Sign language semantic translation system using ontology and deep learning," Sign, vol. 11, no. 1, 2020.
[12] G. Latif, N. Mohammad, J. Alghazo, R. AlKhalaf, and R. AlKhalaf, "Arasl: Arabic alphabets sign language dataset," Data in brief, vol. 23, p. 103777, 2019.

[13] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," Biological cybernetics, vol. 36, no. 4, pp. 193–202, 1980.
[14] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," Progress in Artificial Intelligence, pp. 1–28, 2019.

Asma Althagafi received her Bachellors degree in Computer Sciences from Al-Taif University in 2017. She is currently a Masters of Artificial Intelligence student in Umm Al-Qura University, working on a wide-range of AI related research.

Ghofran Althobaiti received her bachellors degree in computer sciences form Al-Taif University in 2014. She is currently a Masters of Artificial Intelligence student in Umm Al-Qura University**.**

Tahani Alsubait is a faculty member of College of Computer and Information Systems at Umm Al-Qura University. She earned her PhD in AI and instruction from the University of Manchester. She hold a Bachelor's in Computer Science from King Saud University and a Master's from King Abdulaziz University. Her research interests include AI, knowledge representation and reasoning, data analytics and HCI.

Tahani Alqurashi is a faculty member at Common First Year Deanship, Computer Science Department, Umm Al-Qura University. She obtained her master degree in Knowledge Discovery and Data mining and a Phd in Machine Learning from University of East Anglia in UK. Her Current research interests include Unsupervised Machine Learning, Ensemble Methodology, Data Science applications to industrial and Web data, text mining and Natural Language Processing.