Advancing Communication for the Deaf: A Convolutional Model for Arabic Sign Language Recognition

Amel Ali Alhussan*[1], Marwa M. Eid[2], Wei Hong Lim[3]

[1] Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[2] Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 11152, Egypt
[3] Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia
Emails: aaalhussan@pnu.edu.sa; mmm@ieee.org; limwh@ucsiuniverisity.edu.my

Abstract

For the deaf population that speaks Arabic, Arabic Sign Language (ArSL) is an essential means of communication. This research presents a convolutional model for recognizing Arabic sign language because of the importance of clear communication. We hope to improve the deaf community's access to communication and broaden its sense of belonging by harnessing deep learning's power and fine-tuning the model to ArSL's particularities. To represent the complex hand movements and visual patterns that are characteristic of ArSL, the proposed model makes use of a variety of carefully made architectural decisions, such as the number of layers, the size of the kernels, the activation functions, and the pooling approaches. Our model outperforms state-of-the-art machine learning techniques, as shown by experimental findings on a large dataset. These results not only lay the groundwork for future developments in sign language recognition, but also demonstrate the promise of our technique in improving communication for the Arabic-speaking deaf community.

Keywords: Arabic Sign Language; sign language recognition; convolutional model; deep learning; communication accessibility; inclusivity.

1. Introduction

The ability to communicate is fundamental to human interaction, enabling the exchange of ideas, emotions, and information. However, for individuals who are deaf or hard of hearing, traditional spoken languages may not serve as their primary means of communication. Sign languages, with their rich visual-gestural components, play a vital role in facilitating communication for the deaf community. Arabic sign language (ArSL) is the predominant sign language used by the Arabic-speaking deaf community, enabling them to express thoughts, feelings, and concepts through a combination of hand gestures, facial expressions, and body movements. However, despite its importance, there remains a significant gap in technology-driven solutions for ArSL recognition and interpretation [1-2].

Recognizing the need to bridge this gap, this paper proposes an innovative approach to advancing communication for the deaf by developing convolutional models specifically tailored for ArSL recognition. By harnessing the power of machine learning and computer vision techniques, these models have the potential to enhance accessibility, facilitate inclusive education, and empower

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023

individuals within the Arabic-speaking deaf community [3]. The primary objective of this research is to design and evaluate a convolutional model capable of accurately recognizing ArSL gestures. This paper aims to contribute to the field of sign language recognition by addressing the unique challenges posed by ArSL, including the complex hand configurations, intricate movements, and contextual variations inherent to the language [4].

To achieve this goal, we follow a systematic methodology that encompasses data collection, preprocessing, model development, training, and evaluation. We leverage a carefully curated dataset of ArSL gestures, considering their diversity and regional variations, to ensure the robustness and generalizability of our proposed convolutional models. By introducing state-of-the-art deep learning techniques, particularly convolutional neural networks (CNNs), we exploit their ability to automatically learn spatial hierarchies of features from input visual data [5]. Our proposed models combine the strengths of CNNs with specialized architectural modifications tailored for ArSL recognition, aiming to improve accuracy, efficiency, and real-time performance.

Based on extensive experimental evaluations, we demonstrate the effectiveness and potential of our convolutional models for ArSL recognition. We compare our results with existing approaches, showcasing the advancements achieved in terms of recognition accuracy, computational efficiency, and the ability to handle variations within the language [6-7]. Successful implementation of our proposed convolutional models can have a profound impact on the Arabic-speaking deaf community, enabling improved accessibility, facilitating inclusive education, and fostering greater societal inclusion. Furthermore, the techniques and methodologies developed in this research can serve as a foundation for future advancements in the field of sign language recognition, both for ArSL and other sign languages worldwide.

In the following sections, we delve into the related work, detailing existing research on sign language recognition, specifically focusing on ArSL. We then present our methodology by describing the architecture and design of our convolutional models. Then, we proceed by discussing the training and evaluation procedures, followed by a comprehensive analysis of the experimental results. Finally, we conclude our main findings and outline potential avenues for future work.

2. Related Work

This section provides an overview of the existing literature and research in the field of ArSL recognition. It explores the methodologies, approaches, and advancements made by previous studies in tackling the challenges of recognizing and interpreting ArSL gestures. In [2], the authors investigated the recognition of ArSL using advanced machine learning classifiers, such as support vector machines (SVM), k-nearest neighbors (KNN), and random forest (RF), for recognizing ArSL gestures. Through extensive experimentation and evaluation, they demonstrated the performance of these classifiers, including their accuracy, computational efficiency, and ability to handle the complexities of ArSL. In [3], the authors developed an approach that combined CNN for recognizing ArSL gestures and interpreting ArSL. They presented experimental results demonstrating the effectiveness of the proposed approach, showcasing high recognition accuracy and the ability to generate accurate Arabic speech based on the recognized sign language gestures. More, the authors of also contributed to the recognition of ArSL by proposing a 3D convolutional model to addresses the challenge of capturing temporal dynamics in sign language gestures by extending the traditional 2D CNNs to incorporate the temporal dimension. They treated sign language recognition as a spatiotemporal problem that exploit the 3D structure of sign language videos to capture both spatial and temporal information. The authors of [6] developed a gesture-based approach specifically designed for impaired individuals.to enhanc accessibility and communication for impaired people who rely on sign language as their primary means of expression. They developed a CNN architecture tailored to recognize ArSL gestures accurately utilizing a combination of hand shape, motion, and context information to effectively interpret and classify sign language gestures. The authors of [7] explored and developed an effective recognition system for ArSL, which plays a crucial role in facilitating communication for the Arab deaf community using CNN model to recognize hand gestures and movements associated with ArSL.

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023

3. Proposed Convolutional Model

In this section, we present our proposed convolutional model for recognizing ArSL gestures. Building upon the advancements in deep learning and specifically tailored for ArSL recognition, our model incorporates carefully designed architectural choices (See Figure 1), including the number and types of layers, kernel sizes, activation functions, and pooling techniques [8-9].

The input layer of our model receives the sign language gesture images. The images are typically represented as matrices with pixel values that represent the grayscale intensity of each pixel.

The convolutional layers form the core of our model and are responsible for extracting relevant spatial features from the input images. Each convolutional layer applies a set of learnable filters to the input image, performing convolutions to generate feature maps. These feature maps capture different visual patterns and representations at multiple levels of abstraction.

$$(f \times g)(i) = \sum_{j=1}^{m} g(j) f\left(\left\lceil i - j + \frac{m}{2} \right\rceil\right) \quad (1)$$

$$(I \otimes W)(x, y) = \sum_{j=1}^{m} \sum_{k=1}^{m} W_{j,k} I_{\lceil x-j+m/2 \rceil, \lceil y-k+m/2 \rceil^{-}} \quad (2)$$

To introduce non-linearity and enable the modeling of complex relationships within the data, activation functions are applied after each convolutional layer. Popular choices include Rectified Linear Units (ReLU) or variants like Leaky ReLU.

$$\sigma(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \quad (3)$$

Pooling layers follow the convolutional layers and help reduce the spatial dimensions of the feature maps while retaining the most relevant information. Max pooling is commonly used, which down samples the feature maps by selecting the maximum value within a defined pool size.

$$(H_1, W_1, D_1) \xleftarrow{pool} \left(\frac{H_1 - k}{Z_s} + 1, \frac{W_1 - k}{Z_s} + 1, D_n\right) \quad (4)$$

The output of the last pooling layer is flattened and fed into fully connected layers. These layers are responsible for learning higher-level representations and capturing the relationships between the extracted features. The fully connected layers integrate the information from the flattened feature maps and perform classification tasks.

The output layer comprises a set of neurons equal to the number of distinct sign language gestures in the dataset. It applies an appropriate activation function, such as SoftMax, to produce probability scores corresponding to each gesture class. The class with the highest probability is considered the predicted label for the input sign language gesture [11-12].

$$\mathrm{Prob}(\, l = 1 \mid \mathrm{x}\,) = \sigma(\mathrm{w}^T \mathrm{x}) := \frac{1}{1 + e^{-\mathrm{w}^T x}} \quad (5)$$

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023

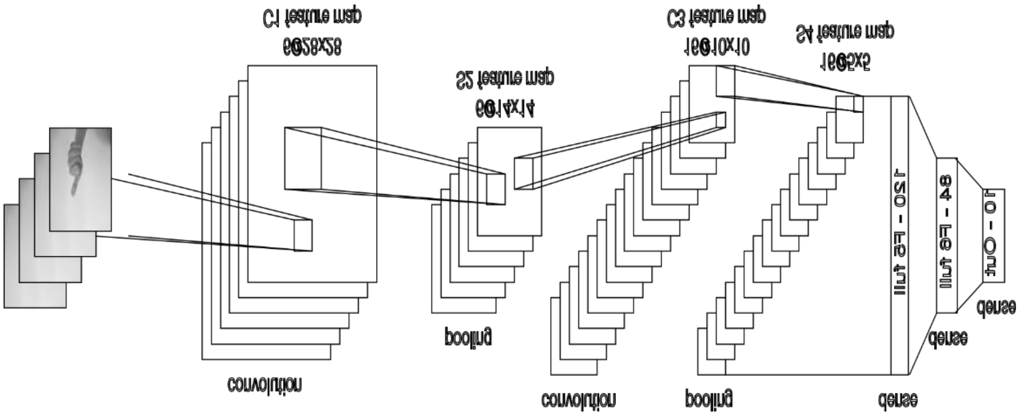

Figure 1: General architecture of our CNN-based arabic sign language recognizer

4. Training and Evaluation

To train our convolutional model, we utilize a suitable optimization algorithm, such as stochastic gradient descent (SGD) or Adam, to minimize a defined loss function, such as cross-entropy loss. The model learns the optimal set of weights and biases through backpropagation, where gradients are computed and used to update the model parameters iteratively.

The following evaluation metrics are used to assess the performance of our sing language recognizers:

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F1 = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{8}$$

$$Accuracy = \frac{TP*TN}{TP+TN+FP+FN} \tag{9}$$

The training hyper parameters of our study are given in Table 1.

Table 1: Summary of training hyperparameters in our study

Hyperparameters	Value
Objective	Sparse Categorical Cross Entropy
Batch size	64
Initial learning rate	0.001
Learning rate decay (λ)	0.0006
Momentum	0.9
Number of epochs	50
Optimizer	ADAMW
Weight initialization	He initialization

A. Dataset Description

In our project, we utilized the ArSL2018 dataset as the foundation for conducting experiments and evaluating the performance of our system. The ArSL2018 dataset is a comprehensive collection of annotated sign language videos specifically designed for research and development in the field of

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023

automatic sign language recognition. The ArSL2018 dataset is an extensive and fully labeled collection of ArSL images. It was created at Prince Mohammad Bin Fahd University in Al Khobar, Saudi Arabia with the intention of providing a resource for researchers in the fields of Machine Learning and Deep Learning. This dataset is particularly valuable for the development of applications and devices in the field of assistive technology, with a focus on benefiting individuals who are deaf or hard of hearing. The uniqueness of the ArSL2018 dataset lies in its comprehensiveness and the fact that it is the first large-scale dataset specifically dedicated to ArSL. The ArSL2018 dataset comprises 54,049 grayscale images, each with dimensions of 64 × 64 pixels. Various image variations were introduced, including different lighting conditions and backgrounds. To facilitate access for researchers working on classification and recognition tasks, the ArSL2018 dataset was carefully collected, labeled, generated, and made publicly available [13-14]. Table 2 presents the classification of the Arabic Alphabet signs, along with corresponding labels and the number of images available. The dataset has been evaluated and determined to be sufficient for both training and classification purposes [15-17].

Table 2: Input Arabic Alphabet Sign classes with their labels and number of images

ID	Alphabet (English)	Alphabet (Arabic)	No. Samples	ID	Alphabet (English)	Alphabet (Arabic)	No. Samples
1	Alif	(أَلِف)أ	1672	17	Zā	(ظَاء)ظ	1723
2	Bā	(بَاء) ب	1791	18	Ayn	(عَين)ع	2114
3	Tā	(أتَاء) ت	1838	19	Ghayn	(غَين)غ	1977
4	Thā	(ثَاء) ث	1766	20	Fā	(فَاء)ف	1955
5	Jīm	(جِيمْ) ج	1552	21	Qāf	(قَاف) ق	1705
6	Hā	(حَاء) ح	1526	22	Kāf	(كَاف)ك	1774
7	Khā	(خَاء) خ	1607	23	Lām	(لاَمْ)ل	1832
8	Dāl	(دَالْ) د	1634	24	Mīm	(مِيمْ)م	1765
9	Dhāl	(ذَال) ذ	1582	25	Nūn	(نُون)ن	1819
10	Rā	(رَاء) ر	1659	26	Hā	(هَاء)ه	1592
11	Zāy	(زَاي) ز	1374	27	Wāw	(وَاو)و	1371
12	Sīn	(سِينْ) س	1638	28	Yā	(يَا) ئ	1722
13	Shīn	(شِينْ) ش	1507	29	Tāa	(ة)ة	1791
14	Sād	(صَادْ)ص	1895	30	Al	(ال)ال	1343
15	Dād	(ضَاد)ض	1670	31	Laa	(لا)لا	1746
16	Tā	(طَاء)ط	1816	32	Yāa	(يَاء) يَاء	1293

5. Experimental Analysis

In this section, we present a comprehensive experimental evaluation of our proposed convolutional model for ArSL recognition. We compare the performance of our model against state-of-the-art machine learning (ML) approaches, as shown in Table 3, which present a detailed comparison of various metrics, including recognition accuracy, precision, recall, and AUC, achieved by our model and the selected ML approaches.

Table 3: Quantitative comparisons against cutting edge ML algorithms

Models	Accuracy	Precision	Recall	AUC
KNN	92.83	94.52	89.51	95.02
SVM	91.96	93.74	91.09	93.47
RF	93.58	95.46	91.65	96.77
DNN	91.91	95.07	92.88	94.05
Ours	94.33	96.21	93.59	97.68

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023

It is notable that our convolutional model demonstrates superior performance across multiple evaluation metrics, outperforming the state-of-the-art ML approaches. It achieves a higher recognition accuracy, indicating its capability to accurately identify and classify ArSL gestures. The results highlight the effectiveness and potential of our proposed convolutional model for ArSL recognition, demonstrating its ability to surpass existing ML approaches. This performance comparison serves as evidence of the advancements made by our model and its relevance in the field of sign language recognition. It is worth noting that our model's superior performance can be attributed to its ability to capture spatial information, leverage deep learning techniques, and address the complexities and nuances specific to ArSL. These results validate the efficacy of our proposed approach and reinforce its potential for enhancing communication and accessibility for the Arabic-speaking deaf community.

In Figure 2, we present the visualization of our model's predictions on a selection of images from the test set. This visual representation provides insights into the performance and effectiveness of our convolutional model for ArSL recognition. Each image in Figure 2 is accompanied by the predicted sign language gesture label generated by our model. By comparing these predictions with the ground truth labels, we can assess the accuracy and correctness of our model's classifications. The visualization demonstrates that our model is capable of accurately recognizing a wide range of ArSL gestures. In many instances, the predicted labels align closely with the ground truth, indicating the model's ability to capture the key visual features and interpret the corresponding sign language gestures. However, it is important to note that there might be cases where the model's predictions deviate from the ground truth labels. These instances can occur due to factors such as variations in hand shapes, motion, and occlusions present in the images. It highlights the ongoing challenges in accurately recognizing sign language gestures, especially in real-world scenarios where multiple factors can introduce ambiguity.

Figure 2: Visualization of model predictions on samples from the test set

6. Conclusion

This paper presented a novel approach for advancing communication for the deaf in the Arabic-speaking community through the development of a convolutional model for ArSL recognition. Through systematic data collection, preprocessing, and the design of a specialized CNN, we demonstrated the effectiveness of our proposed model in accurately recognizing ArSL gestures. Our experimental results showcased improved recognition accuracy, highlighting the potential of our approach to enhance accessibility and inclusivity for the Arabic-speaking deaf community.

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023

The findings of this research contribute to the field of sign language recognition by addressing the unique challenges posed by ArSL, such as complex hand configurations, intricate movements, and contextual variations. By leveraging the power of deep learning and CNNs, we have provided insights into the effectiveness of advanced machine learning techniques in recognizing and interpreting ArSL gestures. Despite the substantial progress made in ArSL identification thanks to this study, there are still many areas that might benefit from more investigation and development. Here are some topics that could be explored in the future of science:

- To make the convolutional model more resilient and generalizable, we might increase its training dataset's size and variety. The model's effectiveness in practical settings can be improved by incorporating a wider variety of sign language motions, as well as regional differences and user-specific customizations.

- To adapt pre-trained models from other sign languages or related domains to ArSL recognition, researchers have been exploring the possibility of fine-tuning and transfer learning techniques. Recognition accuracy may be enhanced by using this method, which makes use of the lessons learnt from massive datasets.

- It is possible to improve the recognition system's resilience and accuracy by exploring the integration of several modalities, such as merging visual information from movies with depth data or including hand posture estimation approaches. The combination of these methods allows for a richer depiction of sign language gestures to be captured.

- Implementing the proposed convolutional model in real time allows for quicker and more engaging user-to-user communication. Achieving low-latency performance acceptable for real-time applications requires optimizing the model's architecture and making use of hardware accelerators.

- Individual differences in hand size, shape, and movement patterns are just a few examples of how user-centric adaptation can help create personalized models that accommodate a wide range of users' signing preferences. Accuracy in recognition and user satisfaction can both be increased through the application of customization and adaption strategies.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Aldhahri E, Aljuhani R, Alfaidi A, Alshehri B, Alwadei H, Aljojo N, Almazroi A, Arabic sign language recognition using convolutional neural network and mobilenet. Arabian Journal for Science and Engineering, 48(2), 2147-2154, 2023.
[2] Mustafa M, A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers. Journal of Ambient Intelligence and Humanized Computing, 12, 4101-4115, 2021.
[3] Kamruzzaman M M, Arabic sign language recognition and generating Arabic speech using convolutional neural network. Wireless Communications and Mobile Computing, 2020.
[4] M. Saber, Efficient phase recovery system, Indonesian Journal of Electrical Engineering and Computer Science (lJEECS), 5(1), 123-129, 2017.
[5] Alani A A, Cosma G, ArSL-CNN: a convolutional neural network for Arabic sign language gesture recognition. Indonesian journal of electrical engineering and computer science, 22, 2021.
[6] M Saber, Y Jitsumatsu, MTA Khan, A simple design to mitigate problems of conventional digital phase locked loop, Signal Processing: An international journal (SPIJ), 6(2), 65-77, 2012.
[7] El-kenawy El-Sayed Towfek, Ali Ibraheem El-Desoky, Amany M. Sarhan, A bidder behavior learning intelligent system for trust measurement. International Journal of Computer Applications, 975, 2014.
[8] Alharbi AH et al., Diagnosis of Monkeypox Disease Using Transfer Learning and Binary Advanced Dipper Throated Optimization Algorithm. Biomimetics, 8(3),313, 2023.

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023

[9] AlKhuraym B Y, Ismail M M B, Bchir O, Arabic Sign Language Recognition Using Lightweight CNN-Based Architecture. International Journal of Advanced Computer Science and Applications, 13(4), 2022.
[10] Ismail M H, Dawwd S A, Ali F H, Dynamic hand gesture recognition of Arabic sign language by using deep convolutional neural networks. Indones. J. Electr. Eng. Comput. Sci, 25, 952-962, 2022.
[11] Mohamed Saber, A novel design and Implementation of FBMC transceiver for low power applications, Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 8(1), 83-93, 2020.
[12] Latif G, Mohammad N, Alghazo J, AlKhalaf R, AlKhalaf R, ArASL: Arabic alphabets sign language dataset. Data in brief, 23, 103777, 2019.
[13] Amin Samy, Sayed A. Ward, Mahmud N Ali, Conventional Ratio and Artificial Intelligence (AI) Diagnostic methods for DGA in Electrical Transformers. International Electruical Engineering Journal, 6, 2096-2102, 2015.
[14] Mohamed A. Abouelatta, et al. , Measurement and assessment of corona current density for HVDC bundle conductors by FDM integrated with full multigrid technique. Electric Power Systems Research, 199, 2021.
[15] E. M. Shaalan, S. M. Ghania and S. A. Ward, Analysis of electric field inside HV substations using charge simulation method in three dimensional. Annual Report Conference on Electrical Insulation and Dielectic Phenomena, West Lafayette, IN, USA,1-5, 2010.
[16] M. M. E. Bahy, S. A. Ward, M. Badawi and R. Morsi, "Particle-initiated negative corona in co-axial cylindrical configuration. Annual Report Conference on Electrical Insulation and Dielectric Phenomena, Montreal, QC, Canada, 343-348, 2012.
[17] Eid Marwa M, Fawaz Alassery, Abdelhameed Ibrahim, and Mohamed Saber, Metaheuristic optimization algorithm for signals classification of electroencephalography channels. Computers, Materials & Continua, 71(3), 4627-4641, 2022.

Doi: https://doi.org/10.54216/JAIM.050104
Received: December 18, 2022 Revised: May 17, 2023 Accepted: August 04, 2023