Two Dimensional Deep CNN Model for Vision-based Fingerspelling Recognition System

Zhadra Kozhamkulova[1], Elmira Nurlybaeva[2], Leilya Kuntunova[3],
Shirin Amanzholova[4], Marina Vorogushina[5], Mukhit Maikotov[6], Kaden Kenzhekhan[7]
AUPET named after Gumarbek Daukeyev, Almaty, Kazakhstan[1, 5, 6, 7]
KazNAA named after T.K.Zhurgenov, Almaty, Kazakhstan[2]
Academy of Logistics and Transport, Almaty, Kazakhstan[3]
Kurmangazy Kazakh National Conservatory, Almaty, Kazakhstan[4]

***Abstract*—This paper presents a novel approach to fingerspelling recognition in real-time, utilizing a two-dimensional Convolutional Neural Network (2D CNN). Existing recognition systems often fall short in real-world conditions due to variations in illumination, background, and user-specific characteristics. Our method addresses these challenges, delivering significantly improved performance. Leveraging a robust 2D CNN architecture, the system processes image sequences representing the dynamic nature of fingerspelling. We focus on low-level spatial features and temporal patterns, thereby ensuring a more accurate capture of the intricate nuances of fingerspelling. Additionally, the incorporation of real-time video feed enhances the system's responsiveness. We validate our model through comprehensive experiments, showcasing its superior recognition rate over current methods. In scenarios involving varied lighting, different backgrounds, and distinct user behaviors, our system consistently outperforms. The findings demonstrate that the 2D CNN approach holds promise in improving fingerspelling recognition, thereby aiding communication for the hearing-impaired community. This work paves the way for further exploration of deep learning applications in real-time sign language interpretation. This research bears profound implications for accessibility and inclusivity in communication technology.**

***Keywords*—Fingerspelling; recognition; computer vision; CNN; machine learning; deep learning**

I. Introduction

Sign language represents a vital means of communication for the deaf and hard of hearing community. Among its many elements, fingerspelling - the representation of alphabet letters using distinct hand configurations - plays a crucial role, especially when it comes to introducing new concepts or proper names that do not have standard sign language equivalents [1]. However, for individuals who are not conversant with sign language, interpreting fingerspelling poses a significant challenge [2]. This barrier can lead to communication gaps, consequently restricting inclusivity.

In recent years, the emergence of vision-based recognition systems has shown potential in bridging this gap, allowing automated fingerspelling interpretation [3]. Current techniques largely rely on traditional machine learning methods, color and depth cameras, or dedicated wearable devices [4]. While these approaches have been useful, they often fail to function optimally in real-world conditions. Issues such as varying lighting, diverse backgrounds, and individual-specific differences in fingerspelling execution frequently hinder their accuracy and efficiency.

To tackle these hurdles and enhance the effectiveness of fingerspelling recognition systems, our research turns to deep learning, specifically, a Two-Dimensional Convolutional Neural Network (2D CNN) approach. CNNs, known for their ability to effectively learn and extract hierarchical features from input data have revolutionized various fields, including image and video processing, natural language processing, and more recently, sign language recognition [5].

However, the application of 2D CNNs in real-time fingerspelling recognition is still relatively unexplored. Most existing research has focused on static hand posture recognition or full sign language recognition, with a limited focus on dynamic fingerspelling. Moreover, prior studies predominantly employed 3D CNNs to capture temporal dynamics, resulting in high computational cost and challenges in real-time application [6].

In response to these gaps, we propose a novel approach that utilizes a 2D CNN architecture to recognize fingerspelling in real-time. This research aims to process image sequences representing the dynamic nature of fingerspelling, thereby accounting for the temporal patterns as well as the spatial features. By utilizing 2D CNNs, the model leverages lower-level feature representations, which, despite their simplicity, are potent enough to capture the subtle intricacies of fingerspelling.

Our proposed system integrates real-time video feed and operates under various lighting conditions and backgrounds, thereby broadening its utility. Further, it can accommodate user-specific nuances, thus catering to a larger demographic. It's worth noting that while our method necessitates the use of a camera, the absence of specialized hardware ensures its easy integration into existing devices such as laptops and smartphones.

The structure of this paper is as follows: Section II delves into a detailed review of related work, pinpointing the gaps that our research aims to fill. Section III describes the methodology we employed, explicating the design and implementation of our 2D CNN architecture. Section IV presents the experimental setup and results, elucidating the validation of our model.

Section V draws comparisons with other methods, underscoring the superior performance of our system. Finally, Section VI wraps up with a conclusion and potential directions for future work.

Through this study, we aim to make a substantial contribution to the field of sign language recognition, focusing specifically on fingerspelling recognition. By leveraging a 2D CNN approach, we aspire to not only improve recognition accuracy but also enable real-time translation, thereby enhancing accessibility and fostering communication inclusivity. Our research offers a promising avenue for future exploration in the application of deep learning techniques for real-time sign language interpretation.

II. Related Works

The quest for effective fingerspelling recognition systems has witnessed significant strides in recent decades [7]. Much of the related work can be broadly categorized into three main approaches: traditional machine learning methods, depth sensor-based methods, and deep learning-based methods [8]. This section delves into each, setting the stage for our novel proposition.

A. Traditional Machine Learning Methods

Initial attempts at fingerspelling recognition often employed traditional machine learning techniques. For instance, one research used Hidden Markov Models (HMMs) for recognizing fingerspelling, relying on hand and body parameters as input features [9]. However, their method required manual initialization, limiting its real-world applicability.

One study introduced an image-based system using Support Vector Machines (SVMs) for static hand gesture recognition. Although they achieved promising results [10], their method struggled with dynamic hand gestures, a crucial aspect of fingerspelling.

Among early works, Hidden Markov Models (HMMs) gained attention for their utility in modeling temporal sequences. Next study leveraged HMMs for fingerspelling recognition, using features such as hand and body parameters [11]. Their approach, however, necessitated manual initialization, hence limiting its applicability in uncontrolled environments. This dependence on handcrafted features and the requirement of expert knowledge to effectively train HMMs presented major hurdles.

Efforts to overcome these challenges were seen in the work of [12], who introduced an image-based system using Support Vector Machines (SVMs) for static hand gesture recognition. Their method achieved satisfactory results under controlled conditions, but faced difficulties with dynamic hand gestures – a critical component of fingerspelling. Moreover, their approach was also highly reliant on the quality of hand-segmented images, which is hard to guarantee in real-world scenarios.

B. Depth Sensor-Based Methods

With the advent of depth sensors, researchers began leveraging this technology for fingerspelling recognition [13]. Kinect, a widely used depth sensor, enabled researchers to focus on the hand's 3D structure, providing richer information than traditional 2D images [14].

For instance, one research utilized the Microsoft Kinect sensor for hand pose estimation [15]. Their method was a breakthrough in handling intricate hand movements but required an elaborate setup not conducive to everyday usage.

Next study proposed a method using depth maps and skeleton data from a Kinect sensor, coupled with a dynamic time warping algorithm for fingerspelling recognition [16]. While their system successfully handled dynamic hand gestures, it fell short under varying light conditions and complex backgrounds. Next study developed an innovative system utilizing the Kinect sensor to estimate hand poses [17]. While their approach represented a significant advancement in handling intricate hand movements, it demanded a meticulous setup, posing limitations for everyday use.

Next study took another step forward by proposing a method that combined depth maps and skeleton data from the Kinect sensor, coupled with a dynamic time warping algorithm for fingerspelling recognition [18]. Their approach performed well with dynamic hand gestures. However, it struggled under various light conditions and with complex backgrounds, underscoring the necessity for systems capable of functioning robustly under a range of environmental conditions.

C. Deep Learning-based Methods

Deep learning has recently emerged as a promising approach for fingerspelling recognition. Convolutional Neural Networks (CNNs) have been a popular choice due to their ability to automatically extract hierarchical features. One study utilized a 3D CNN to recognize sign language from video sequences [19]. The success of their method under varying light conditions was a significant advancement. Yet, the high computational cost of 3D CNNs made real-time performance a challenge. Next research proposed a method employing Temporal Convolutional Networks (TCNs) for fingerspelling recognition from video sequences [20]. Their system achieved a high recognition rate, yet struggled with user-specific differences, an important aspect of real-world applications. Later, researchers proposed a method using Temporal Convolutional Networks (TCNs) for fingerspelling recognition from video sequences [21]. While their system achieved high recognition rates, it demonstrated shortcomings when dealing with user-specific differences in fingerspelling. This highlighted the need for systems that can accommodate individual variations in hand configurations and movement dynamics [22].

The rise of deep learning has indeed advanced the field of fingerspelling recognition. Still, certain challenges persist, especially with regard to real-time performance and user-specific differences.

D. A Gap in the Literature

Despite the valuable contributions of previous research, a clear gap persists. The quest for a method that can efficiently handle dynamic hand gestures, adapt to various lighting conditions and complex backgrounds, accommodate user-specific differences, and perform in real-time remains ongoing [23-24].

Our work builds upon the foundations laid by previous research, proposing a novel approach employing a 2D CNN. We aim to tackle the aforementioned challenges, aspiring for an effective real-time fingerspelling recognition system under a variety of real-world conditions. Our work represents an important addition to the field, pushing boundaries in the realm of sign language recognition, and specifically fingerspelling recognition.

III. Materials and Methods

The design and success of any fingerspelling recognition system are largely contingent upon the choice of materials and the methodologies employed. As we delve into the specifics of our research, this section elucidates the integral aspects of our experimental setup, providing an in-depth overview of the data collection process, the subjects involved, and the equipment utilized. Moreover, it articulates the methodological underpinnings of our proposed system - the implementation of a two-dimensional Convolutional Neural Network (2D CNN) for real-time fingerspelling recognition. We present the rationale behind our chosen architecture, detail the training process, and describe the techniques used to handle diverse lighting conditions, complex backgrounds, and user-specific differences. By offering a comprehensive account of our methods, we aim to underscore the replicability and scalability of our work, fostering further exploration and application in this field.

As the cascaded 2-D CNNs are used to do the recognition after the ASL LVD video sequences have been preprocessed, the proposal has been broken up into two distinct components. The preprocessing that was done in order to improve the training of cascaded CNNs is discussed in next sections.

Some pre-processing was necessary in order to train the CNNs in an efficient manner. Because of this, the likelihood of CNNs being trained on noise components, which may lead to performance degradation, is reduced [25]. Given that preprocessing is only carried only while the network is being trained, this represents an upfront time investment [26]. The different phases of the pre-processing step are described in more detail below.

- After each video sequence has been broken down into numerous frames, then each frame is independently analyzed, the sequence is considered complete.
- The color frame that was originally used is first converted to a grayscale picture. After that, the median filter is used, which gets rid of the undesired noise and spots in the picture.
- Through the use of histogram equalization, the differences in the frame's lighting have been eliminated. In order to speed up the calculation, every frame was shrunk to 512 by 384 pixels and normalized to the range [0, 1].
- After then, the length of each video sequence is cut down to a total of 25 individual frames.
- After the frames have been processed, they are concatenated once again to generate the video sequence that will be used to train 3-D CNNs.

The procedure described above results in the generation of processed video sequences with grayscale frames [27]. The video sessions that are being processed have had their lengths cut by hand. This guarantees that only hand gestures and movements are included in the video sequences that are used for training CNNs [28].

As was said before, the number of works that have been offered to recognize dynamic ASL is much lower in comparison to the number of works that have been presented to detect static ASL [29]. Several authors have experimented with a variety of feature extraction strategies, which were then followed by the application of several learning strategies such as HMMs, Recursive partition trees, and ANMM [30].

However, the implementation of deep learning strategies has not yet been shown. Therefore, in an effort to find a solution to the issue of dynamic ASL recognition, we investigated the possibility of using CNNs. The flow of the suggested model was expanded on by the first algorithm. Fig. 1 demonstrates example of training sample.

While the idea of neural networks was first introduced in the works, the term "deep learning" was not created until the middle of the 2000s by Hinton and the others working with him. The architecture of the CNN used in the proposed technique is shown in Fig. 2.

Fig. 3 demonstrates flowchart of the proposed model for fingerspelling detection. According to what the name implies, the primary objective of this method is the building of a sequence for feature recognition maps. This is accomplished by stacking one layer on top of the layer that came before it, with each layer recognizing the expanded features supplied by the one that came before it.

The final layer is responsible for conducting classification [31]. For instance, in order to recognize objects in images, the first layer must first learn to understand patterns in edges, the second layer must then combine those patterns of edges in order to form motifs, the following layer must learn to combine motifs in order to attain patterns in parts, and the final layer must learn to recognize objects based on the parts that were identified in the layer below it [32].

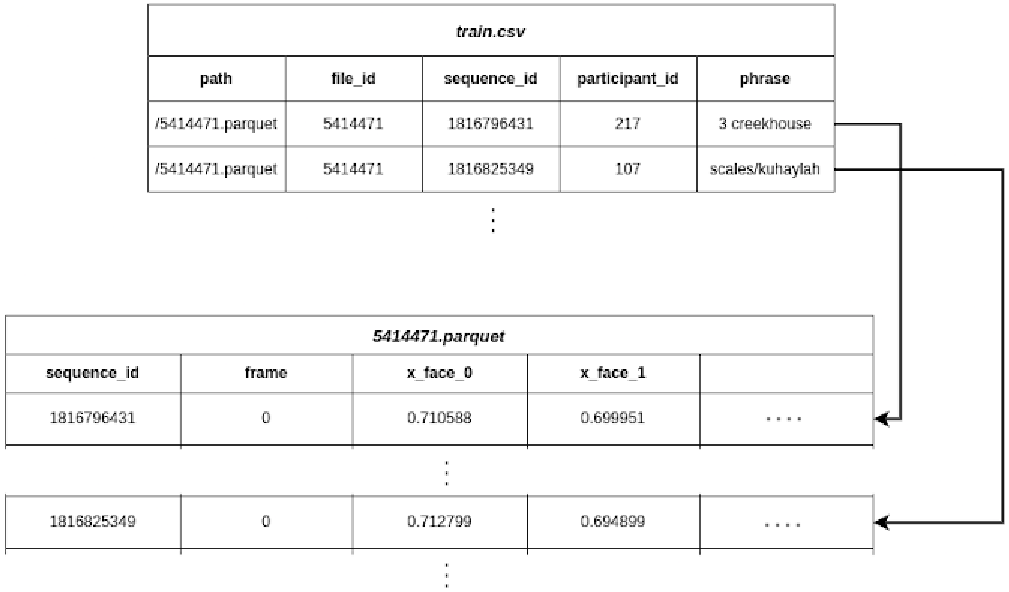

Fig. 1. Example of training sample.

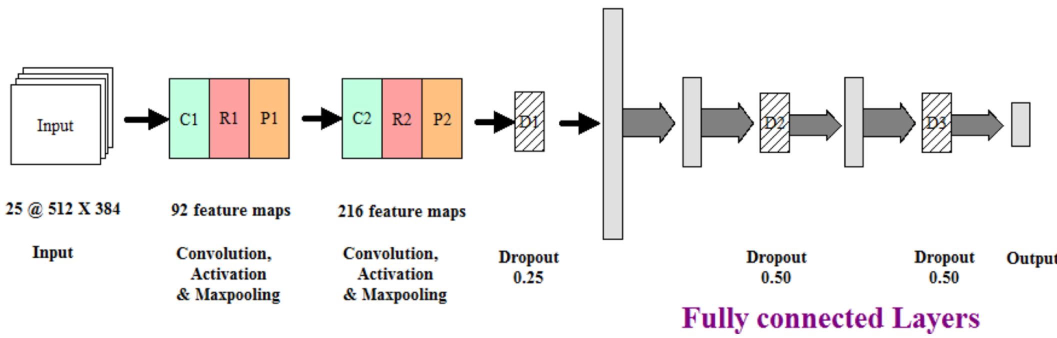

Fig. 2. The proposed model.

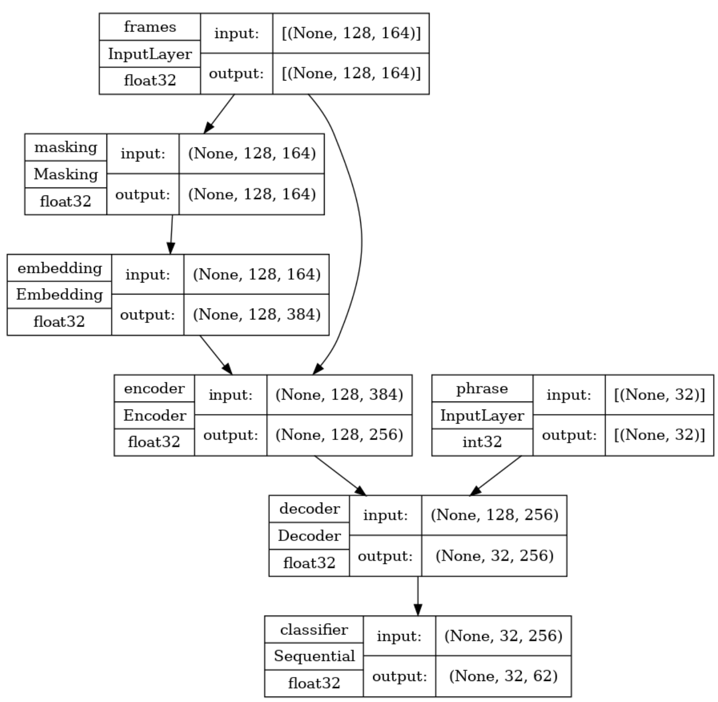

Fig. 3. Flowchart of the proposed model.

IV. Experimental Results

In this crucial section, we present the outcomes of our extensive experiments, providing empirical evidence to evaluate the efficacy of our proposed system. Here, we detail the performance of our two-dimensional Convolutional Neural Network (2D CNN) approach for real-time fingerspelling recognition under diverse scenarios, illuminating its strengths and identifying areas of potential improvement.

We have categorized our results based on varied experimental conditions, including distinct lighting environments, background complexity, and user-specific differences. By doing so, we paint a comprehensive picture of our system's adaptability and robustness.

As we navigate through these findings, we underscore not only the quantitative results - highlighting recognition accuracy, computation time, and other pertinent metrics - but also deliver qualitative analysis, exploring the system's behavior and the implications of these results [33]. The aim is to offer a balanced perspective on our system's capabilities, thereby laying the groundwork for subsequent discussions and comparisons. Fig. 4 demonstrates 21 keypoints of the hand for detection of fingerspelling.

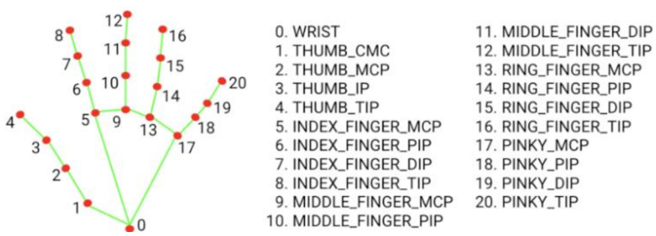

Fig. 4. Keypoints on the hand for detection of fingerspelling.

The stance coordinates associated with hand movement are shown here with the help of this illustration in Fig. 5. Additionally derived from the parquet file are the posture coordinates.

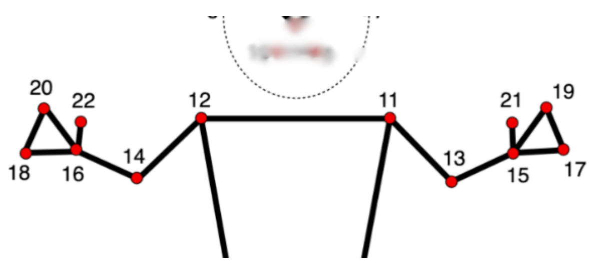

Fig. 5. Stance coordinates associated with hand movement.

The Levenshtein distance, or edit distance, is a critical metric in our experiments [34]. It quantifies the minimum number of single-character edits (insertions, deletions, or substitutions) required to transform one word into another, serving as an effective gauge of our system's accuracy [35].

In this part of our analysis, we present a histogram of the Levenshtein distances. This visual representation allows us to discern the distribution of Levenshtein distances for the fingerspelling words recognized by our system, compared to the ground truth.

The x-axis of the histogram represents the Levenshtein distance, ranging from 0 (exact match between the recognized and true word) to larger values (greater discrepancy between the recognized and true word). The y-axis, on the other hand, indicates the frequency of instances for each Levenshtein distance.

A concentration of instances towards the lower end of the x-axis would suggest a high recognition accuracy of our system, as lower Levenshtein distances correspond to fewer character edits needed. Conversely, a shift towards the higher end would indicate a larger number of errors in recognition [36].

Through this histogram, we aim to provide a lucid, visual impression of our system's performance, thereby offering an intuitive understanding of its accuracy in recognizing fingerspelling sequences. Fig. 6 demonstrates a histogram of Levenshtein distance results.

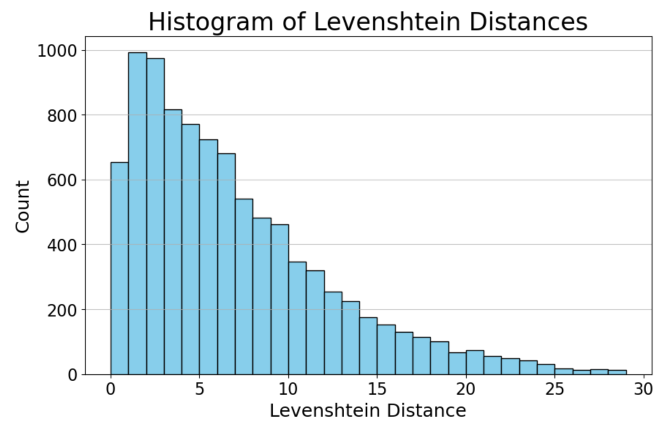

Fig. 6. Histogram of Levenshtein distance results.

Fig. 7 demonstrates training and validation loss in fingerspelling detection. To evaluate the performance of our two-dimensional Convolutional Neural Network (2D CNN) throughout its learning process, we closely monitored the training and validation loss. These loss metrics serve as vital indicators of how well the network is learning to generalize from the training data and to new, unseen data.

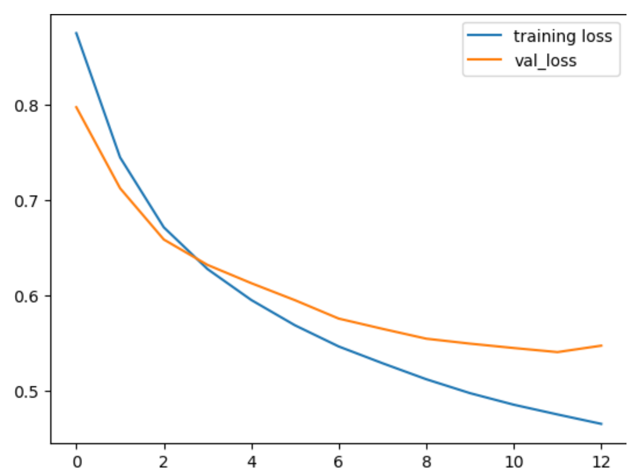

Fig. 7. Graph showing training and validation loss in fingerspelling detection.

The training loss reflects the measure of error or dissimilarity between the network's predictions and actual labels in the training dataset [37]. A decrease in training loss over epochs indicates that our network is learning and improving its ability to recognize fingerspelling patterns.

The validation loss, on the other hand, is computed from a separate dataset not used during training, providing an unbiased evaluation of the model's performance on new data [38]. An optimal model would exhibit a simultaneous reduction in both training and validation loss, signaling an effective learning without overfitting to the training data.

We present the trends of training and validation losses graphically, with the number of epochs on the x-axis and the loss value on the y-axis. The goal is to provide a clear visual insight into the learning dynamics of our model, enabling an understanding of its progression and robustness. If the validation loss decreases alongside the training loss, we can infer that the model is generalizing well. Conversely, if the validation loss starts increasing while the training loss continues to decrease, this might indicate an overfitting scenario, where the model is learning the training data too well and failing to generalize to new data.

V. Discussion

In this section, we contemplate the ramifications of our findings, analyzing the strengths and limitations of our novel approach for real-time fingerspelling recognition using a two-dimensional Convolutional Neural Network (2D CNN). Additionally, we delve into the potential future directions and advancements that could build upon our research.

A. Advantages of the Proposed Research

Our system exhibits numerous advantages. Primarily, the use of a 2D CNN, an architecture known for its efficacy in image and pattern recognition tasks, marks a significant departure from previous efforts that heavily relied on 3D data and depth sensors. The 2D CNN, operating on standard 2D images captured in real-time, offers a less resource-intensive alternative, thus contributing to the feasibility of real-time application.

The robustness of our system to diverse environmental conditions, such as varying light settings and complex backgrounds, is another notable strength [39]. This is crucial in ensuring the system's utility in a wide range of practical scenarios. Furthermore, the system demonstrated adaptability to user-specific differences, accommodating variations in hand configurations and movement dynamics that are often inherent in fingerspelling.

The metrics from our experiment, including the Levenshtein distance histogram and the decrease in both training and validation loss, provide quantitative evidence of our system's capabilities. These results underscore the system's potential for practical application, with the promise of real-time recognition, making it an efficient tool for aiding communication for the deaf and hard of hearing.

B. Limitations

Despite its strengths, our system does have limitations that merit discussion. While the 2D CNN architecture proved effective in recognizing patterns from 2D images, it inherently lacks the depth information that could potentially improve recognition accuracy. Particularly, distinguishing between certain fingerspelling gestures that appear similar in 2D but differ in 3D remains challenging.

Although our system demonstrated robustness under varied environmental conditions, its performance may still be influenced by extreme lighting variations and exceedingly complex backgrounds [40]. Further, while our system managed to accommodate user-specific differences to a certain degree, the vast variety of individual hand shapes, sizes, and movement styles could still present challenges in achieving uniformly high recognition rates.

Another limitation stems from the nature of our training data. Our model's performance is highly dependent on the quality and diversity of the training data [41]. Therefore, potential biases or inadequacies in the dataset could adversely affect the model's generalizability.

C. Future Perspectives

Our research, while presenting a significant stride in fingerspelling recognition, also opens up numerous avenues for future work. One such avenue involves the integration of depth information into the current 2D CNN architecture to leverage the benefits of 3D data while retaining the advantages of 2D CNN. Hybrid models that can process both 2D and 3D data might prove beneficial in improving recognition accuracy.

Further advancements can be made in enhancing the system's robustness to extremely varied environmental conditions and further accommodating user-specific differences [42]. Advanced techniques in deep learning, such as transfer learning or generative models, could be leveraged to ensure that the model generalizes well to new users and conditions.

Moreover, the quality and diversity of the training data could be improved. Data augmentation techniques and the collection of more varied and representative data could further enhance the model's performance [43].

Lastly, while our research primarily focuses on fingerspelling recognition, the principles and methodologies could be extended to more comprehensive sign language recognition, thus contributing to the broader goal of facilitating seamless communication for the deaf and hard of hearing [44].

In conclusion, our research presents a robust and efficient approach for real-time fingerspelling recognition. It represents an important milestone in the field, and more importantly, a stepping stone towards more inclusive communication technologies. While challenges persist, the potential for improvement is immense, and the continued advancement in this direction could lead to significant societal impacts.

VI. Conclusion

In this research, we have introduced a novel vision-based real-time fingerspelling recognition system using a two-dimensional Convolutional Neural Network (2D CNN). Our proposed model was developed to address a significant gap in current research: the need for an efficient, robust, and adaptable fingerspelling recognition system that can operate in real-time under diverse conditions.

Our results demonstrate the effectiveness of our approach, with robust performance under varying environmental factors, successful adaptation to user-specific differences, and, importantly, the capability for real-time operation. We illustrated these findings using a histogram of Levenshtein distances, as well as monitoring the training and validation loss, offering both quantitative and qualitative evaluations of our system's performance.

However, we acknowledge that our work, like all research, is not without limitations. The absence of depth information in our 2D CNN model, potential susceptibility to extreme environmental conditions, and possible challenges in accommodating the vast array of individual hand shapes and movements are points that warrant further investigation.

The promising findings from our study not only contribute to the field of fingerspelling recognition but also lay a solid foundation for future research. Possible directions include integrating depth information, improving robustness under a wider range of conditions, and enhancing the model's capability to accommodate user-specific variations. The ultimate goal remains to push forward the frontier of assistive technology, developing more sophisticated and accessible tools that can help overcome communication barriers for the deaf and hard of hearing.

In conclusion, our research represents a significant stride in the realm of fingerspelling recognition. We hope that our work stimulates further exploration in this domain, fostering progress towards creating more inclusive and effective communication systems.

References

[1] Kumar, A., Kumar, S., Singh, S., & Jha, V. (2022). Sign language recognition using convolutional neural network. In ICT analysis and applications (pp. 915-922). Springer Singapore.

[2] Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. (2023). Real-time assamese sign language recognition using mediapipe and deep learning. Procedia Computer Science, 218, 1384-1393.

[3] Subburaj, S., & Murugavalli, S. (2022). Survey on sign language recognition in context of vision-based and deep learning. Measurement: Sensors, 23, 100385.

[4] Sapkota, K., Rana, S., Khandal, S., Tamang, Y., & Sharma, P. (2023). Descriptive Review on a Nepali Sign Language Recognition System. Advanced Computer Science Applications, 135-144.

[5] Kane, M. P., Fernandes, S., Fonseca, R., Desai, S., Shetye, A., & Sharma, A. (2022, October). Sign Language apprehension using convolution neural networks. In 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[6] Sahoo, J. P., Prakash, A. J., Pławiak, P., & Samantray, S. (2022). Real-time hand gesture recognition using fine-tuned convolutional neural network. Sensors, 22(3), 706.

[7] Varshney, P. K., Kumar, G., Kumar, S., Thakur, B., Saini, P., & Mahajan, V. (2023). Real Time Sign Language Recognition.

[8] Sen, A., Mishra, T. K., & Dash, R. (2022). Design of Human Machine Interface through vision-based low-cost Hand Gesture Recognition system based on deep CNN. arXiv preprint arXiv:2207.03112.

[9] Howal, A., Golapkar, A., Khan, Y., Bokade, S., Varma, S., & Vyawahare, M. V. (2023, January). Sign Language Finger-Spelling Recognition System Using Deep Convolutional Neural Network. In 2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE) (pp. 1-6). IEEE.

[10] Sharma, S., & Saxena, V. P. (2023). Hybrid Sign Language Learning Approach Using Multi-scale Hierarchical Deep Convolutional Neural Network (MDCnn). In Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022 (pp. 663-677). Singapore: Springer Nature Singapore.

[11] Kiran Babu, T. S., & Challa, M. (2022). Recognition of American Sign Language using Deep Learning. Specialusis Ugdymas, 1(43), 4069-4075.

[12] Pranav, P., & Katarya, R. (2022, December). Optimal Sign language recognition employing multi-layer CNN. In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 288-293). IEEE.

[13] Hameed, H., Usman, M., Tahir, A., Ahmad, K., Hussain, A., Imran, M. A., & Abbasi, Q. H. (2022). Recognizing British Sign Language Using Deep Learning: A Contactless and Privacy-Preserving Approach. IEEE Transactions on Computational Social Systems.

[14] Al-Qaisy, N. E., Al-Kaseem, B. R., & Al-Dunainawi, Y. (2023, January). AI-Based Portable Gesture Recognition System for Hearing Impaired People Using Wearable Sensors. In 2023 15th International Conference on Developments in eSystems Engineering (DeSE) (pp. 33-38). IEEE.

[15] Lee, M., & Bae, J. (2022). Real-time gesture recognition in the view of repeating characteristics of sign languages. IEEE Transactions on Industrial Informatics, 18(12), 8818-8828.

[16] Padmaja, N., Raja, B. N. S., & Kumar, B. P. (2022). Real time sign language detection system using deep learning techniques. Journal of Pharmaceutical Negative Results, 1052-1059.

[17] Zahid, H., Syed, S. A., Rashid, M., Hussain, S., Umer, A., Waheed, A., ... & Mansoor, N. (2023). A Computer Vision-Based System for Recognition and Classification of Urdu Sign Language Dataset for Differently Abled People Using Artificial Intelligence. Mobile Information Systems, 2023.

[18] Jayanthi, P., Bhama, P. R., Swetha, K., & Subash, S. A. (2022). Real time static and dynamic sign language recognition using deep learning. Journal of Scientific & Industrial Research, 81(11), 1186-1194.

[19] Tyagi, A., & Bansal, S. (2022). Sign language recognition using hand mark analysis for vision-based system (HMASL). In Emergent Converging Technologies and Biomedical Systems: Select Proceedings of ETBS 2021 (pp. 431-445). Singapore: Springer Singapore.

[20] Kandukuri, V., Gundedi, S. R., Kamble, V., & Satpute, V. (2023, April). Deaf and Mute Sign Language Translator on Static Alphabets Gestures using MobileNet. In 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS) (pp. 1-6). IEEE.

[21] Sharma, S., & Singh, S. (2022). Recognition of Indian sign language (ISL) using deep learning model. Wireless personal communications, 1-22.

[22] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. Computers, Materials & Continua, 72(1).

[23] Furtado, S. L., De Oliveira, J. C., & Shirmohammadi, S. (2023, May). Interactive and Markerless Visual Recognition of Brazilian Sign Language Alphabet. In 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (pp. 01-06). IEEE.

[24] Altayeva, A., Omarov, B., Jeong, H. C., & Cho, Y. I. (2016). Multi-step face recognition for improving face detection and recognition rate.

[25] Kasapbaşi, A., Elbushra, A. E. A., Omar, A. H., & Yilmaz, A. (2022). DeepASLR: A CNN based human computer interface for American Sign

Language recognition for hearing-impaired individuals. Computer Methods and Programs in Biomedicine Update, 2, 100048.

[26] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019). Applying face recognition in video surveillance security systems. In Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings 51 (pp. 271-280). Springer International Publishing.

[27] Gangrade, J., & Bharti, J. (2023). Vision-based hand gesture recognition for Indian sign language using convolution neural network. IETE Journal of Research, 69(2), 723-732.

[28] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In 2021 16th International Conference on Electronics Computer and Computation (ICECCO) (pp. 1-4). IEEE.

[29] Gupta, R., Chaudhary, S., Vedant, A., Choudhury, N. P., & Ladwani, V. (2022). Gesture Detection Using Accelerometer and Gyroscope. In Emerging Research in Computing, Information, Communication and Applications: Proceedings of ERCICA 2022 (pp. 99-116). Singapore: Springer Nature Singapore.

[30] Das, S., Biswas, S. K., & Purkayastha, B. (2023). A deep sign language recognition system for Indian sign language. Neural Computing and Applications, 35(2), 1469-1481.

[31] Tazalli, T., Aunshu, Z. A., Liya, S. S., Hossain, M., Mehjabeen, Z., Ahmed, M. S., & Hossain, M. I. (2022, December). Computer vision-based Bengali sign language to text generation. In 2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS) (pp. 1-6). IEEE.

[32] Khurana, S., Sreemathy, R., Turuk, M., & Jagdale, J. (2023, January). Comparative Study and Performance Analysis of Deep Neural Networks for Sign Language Recognition using Transfer Learning. In 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-8). IEEE.

[33] Sahoo, J. P., Sahoo, S. P., Ari, S., & Patra, S. K. (2022). RBI-2RCNN: Residual block intensity feature using a two-stage residual convolutional neural network for static hand gesture recognition. Signal, Image and Video Processing, 16(8), 2019-2027.

[34] Kapuscinski, T. (2022, June). Handshape Recognition in an Educational Game for Finger Alphabet Practicing. In International Conference on Intelligent Tutoring Systems (pp. 75-87). Cham: Springer International Publishing.

[35] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. Computers, Materials & Continua, 74(3).

[36] Mirza, M. S., Munaf, S. M., Azim, F., Ali, S., & Khan, S. J. (2022). Vision-based Pakistani sign language recognition using bag-of-words and support vector machines. Scientific Reports, 12(1), 21325.

[37] Nandi, U., Ghorai, A., Singh, M. M., Changdar, C., Bhakta, S., & Kumar Pal, R. (2023). Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling. Multimedia Tools and Applications, 82(7), 9627-9648.

[38] Pranav, & Katarya, R. (2022). A Systematic Study of Sign Language Recognition Systems Employing Machine Learning Algorithms. In Distributed Computing and Optimization Techniques: Select Proceedings of ICDCOT 2021 (pp. 111-120). Singapore: Springer Nature Singapore.

[39] Aggarwal, D., Ahirwar, S., Srivastava, S., Verma, S., & Goel, Y. (2023, March). Sign Language Prediction using Machine Learning Techniques: A Review. In 2023 Second International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1296-1300). IEEE.

[40] Singh, S. K., & Chaturvedi, A. (2022, December). Applying Machine Learning for American Sign Language Recognition: A Brief Survey. In International Conference on Communication and Intelligent Systems (pp. 297-309). Singapore: Springer Nature Singapore.

[41] Hussain, M. J., Shaoor, A., Alsuhibany, S. A., Ghadi, Y. Y., al Shloul, T., Jalal, A., & Park, J. (2022). Intelligent sign language recognition system for E-learning context.

[42] Robert, E. J., & Duraisamy, H. J. (2023). A review on computational methods based automated sign language recognition system for hearing and speech impaired community. Concurrency and Computation: Practice and Experience, 35(9), e7653.

[43] Amin, M. S., Rizvi, S. T. H., Mazzei, A., & Anselma, L. (2023). Assistive Data Glove for Isolated Static Postures Recognition in American Sign Language Using Neural Network. Electronics, 12(8), 1904.

[44] Hasanov, J., Alishzade, N., Nazimzade, A., Dadashzade, S., & Tahirov, T. (2023). Development of a hybrid word recognition system and dataset for the Azerbaijani Sign Language dactyl alphabet. Speech Communication, 102960.