Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

EXPLORING CONVOLUTIONAL NEURAL NETWORKS FOR DETECTING HAND GESTURES

Dr. B. Sravan Kumar

Dr. B. Sravan Kumar ,*Associate Professor ,CSE(AI &ML) Department, Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal(Mandal),Warangal Urban-506005(T.S,)India*

K.Raja Shekar ,*Assistant Professor ,CSE(AI &ML) Department, Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal(Mandal),Warangal Urban-506005(T.S,)India*

Gudi Sai Madhav Reddy , 20641A6669 *, UG Student, CSE(AI &ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal(Mandal),Warangal Urban-506005(T.S,)India*

Pallerla Pavan Reddy , 20641A66B3 *, UG Student, CSE(AI &ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal(Mandal),Warangal Urban-506005(T.S,)India*

Mulka Chandana , 20641A66A1 *, UG Student, CSE(AI &ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal(Mandal),Warangal Urban-506005(T.S,)India*

Kodati Sai Teja , 20641A6677 *, UG Student, CSE(AI &ML), Vaagdevi College of Engineering (Autonomous), Bollikunta, Khila Warangal(Mandal),Warangal Urban-506005(T.S,)India*

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

EXPLORING CONVOLUTIONAL NEURAL NETWORKS FOR DETECTING HAND GESTURES

ABSTRACT

Conversing to a person with hearing disability is always a major challenge. Sign language has indelibly become the ultimate panacea and is a very powerful tool for individuals with hearing and speech disability to communicate their feelings and opinions to the world. It makes the integration process between them and others smooth and less complex. However, the invention of sign language alone, is not enough. There are many strings attached to this boon. The sign gestures often get mixed and confused for someone who has never learnt it or knows it in a different language. However, this communication gap which has existed for years can now be narrowed with the introduction of various techniques to automate the detection of sign gestures. In this paper, we introduce a Sign Language recognition using American Sign Language. In this study, the user must be able to capture images of the hand gesture using web camera and the system shall predict and display the name of the captured image. We use the HSV colour algorithm to detect the hand gesture and set the background to black. The images undergo a series of processing steps which include various Computer vision techniques such as the conversion to grayscale, dilation and mask operation. And the region of interest which, in our case is the hand gesture is segmented. The features extracted are the binary pixels of the images. We make use of Convolutional Neural Network (CNN) for training and to classify the images. We are able to recognize 10 American Sign gesture alphabets with high accuracy. Our model has achieved a remarkable accuracy of above 90%.

1. INTRODUCTION

As well stipulated by Nelson Mandela [1], “Talk to a man in a language he understands, that goes to his head. Talk to him in his own language, that goes to his heart”, language is undoubtedly essential to human interaction and has existed since human civilization began. It is a medium humans use to communicate to express themselves and understand notions of the real world. Without it, no books, no cell phones and definitely not any word I am writing would have any meaning. It is so deeply embedded in our everyday routine that we often take it for granted and don’t realize its importance. Sadly, in the fast-changing society we live in, people with hearing impairment are usually forgotten and left out. They have to struggle to bring up their ideas, voice out their opinions and express themselves to people who

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

are different to them. Sign language, although being a medium of communication to deaf people, still have no meaning when conveyed to a non-sign language user. Hence, broadening the communication gap. To prevent this from happening, we are putting forward a sign language recognition system[2],[3]. It will be an ultimate tool for people with hearing disability to communicate their thoughts as well as a very good interpretation for non-sign language user to understand what the latter is saying. Many countries have their own standard and interpretation of sign gestures. For instance, an alphabet in Korean sign language will not mean the same thing as in Indian sign language. While this highlights diversity, it also pinpoints the complexity of sign languages. Deep learning must be well versed with the gestures so that we can get a decent accuracy. In our proposed system, American Sign Language is used to create our datasets. Figure 1 shows the American Sign Language (ASL) alphabets.

Identification of sign gesture is performed with either of the two methods. First is a glove based method whereby the signer wears a pair of data gloves during the capture of hand movements. Second is a vision based method, further classified into static and dynamic recognition[2]. Static deals with the 2dimensional representation of gestures while dynamic is a real time live capture of the gestures

And despite having an accuracy of over 90% [3], wearing of gloves are uncomfortable and cannot be utilized in rainy weather. They are not easily carried around since their use require computer as well. In this case, we have decided to go with the static recognition of hand gestures because it increases accuracy as compared to when including dynamic hand gestures like for the alphabets J and Z. We are proposing this research so we can improve on accuracy using Convolution Neural Network(CNN).

2. LITERATURESURVEY

This research focuses on developing a real-time sign language recognition system that utilizes Convolutional Neural Networks (CNN) [4]-[8]to facilitate seamless communication between individuals with hearing disabilities and the hearing world. Sign language is a powerful tool for expressing emotions and opinions, but the communication gap persists due to the complexity of sign gestures and variations in different sign languages. Our proposed system aims to bridge this gap by enabling users to capture hand gestures using a web camera. The captured images are then processed using Computer Vision techniques, including HSV color algorithms and segmentation to isolate the hand gesture region. The binary pixel features are extracted and fed into a CNN for training and classification. Specifically, we concentrate on recognizing the American Sign Language (ASL) alphabet gestures. Through rigorous experimentation, we have achieved exceptional accuracy,

exceeding 90%, for recognizing [10] ASL alphabet signs. This innovative system can significantly enhance communication and inclusivity for individuals with hearing and speech disabilities.

This project proposes a multi-modal approach for sign language recognition and translation, combining Convolutional Neural Networks (CNN) [9]with Natural Language Processing (NLP) techniques. Our system aims to recognize hand gestures captured through a web camera using CNN. Once the gestures are identified, they are translated into text using NLP algorithms. The system allows users to communicate with the hearing world by converting their sign language gestures into understandable text. To enhance the accuracy and robustness of the model, we employ data augmentation techniques and recurrent neural networks to handle temporal dependencies in sign language gestures. The resulting model is capable of recognizing and translating complex sign language sentences with high accuracy, making communication easier for individuals with hearing disabilities.

In this study, we present an improved sign language recognition system that utilizes transfer learning and data augmentation in combination with Convolutional Neural Networks (CNN). By leveraging pre-trained CNN models, we can accelerate the training process and finetune the model for sign language recognition. Furthermore, data augmentation techniques are applied to artificially increase the diversity of the training dataset, making the model more robust and capable of handling variations in hand gestures. The proposed system is tested on a dataset of American Sign Language gestures [11], achieving remarkable accuracy in recognizing a wide range of sign symbols. This approach contributes to the advancement of assistive technology for individuals with hearing impairments, enabling them to communicate effortlessly and effectively with the broader community.

This research introduces a real-time mobile-based sign language recognition system that employs Convolutional Neural Networks (CNN) and Edge Computing [12] to provide on-device recognition capabilities. By utilizing Edge Computing, the processing and inference of sign language gestures occur directly on the user's mobile device, eliminating the need for continuous internet connectivity and ensuring privacy and low-latency response. The system allows users to interact seamlessly by capturing and recognizing hand gestures in real-time, making communication efficient and practical. The CNN model is optimized for mobile devices[13]-[15], providing a balance between accuracy and computational efficiency. Through extensive testing, our system demonstrates reliable and rapid sign language recognition, empowering individuals with hearing impairments to communicate effortlessly

in various situations.

This project proposes an innovative approach to sign language recognition and synthesis using a combination of Convolutional Neural Networks (CNN) [16]and Generative Adversarial Networks (GAN). We focus on recognizing hand gestures captured through a web camera using CNN while simultaneously employing GAN to synthesize sign language animations. The synthesis of sign language animations enhances the visual expressiveness of communication and enables clearer understanding for both hearing-impaired individuals and their hearing counterparts. Our system provides a comprehensive solution for bridging the communication gap by recognizing sign gestures and generating corresponding animated signs. This holistic approach contributes to more inclusive communication and a richer user experience for individuals with hearing and speech disabilities.

3. PROBLEM STATEMENT

Sign language, as one of the most widely used communication means for hearing-impaired people, is expressed by variations of hand-shapes, body movement, and even facial expression. Since it is difficult to collaboratively exploit the information from hand-shapes and body movement trajectory, sign language recognition is still a very challenging task. This paper proposes [17] an effective recognition model to translate sign language into text or speech in order to help the hearing impaired communicate with normal people through sign language.

Technically speaking, the main challenge of sign language recognition lies in developing descriptors to express hand-shapes and motion trajectory. In particular, hand-shape description involves tracking hand regions in video stream, segmenting hand-shape images from complex background in each frame and gestures recognition problems. Motion trajectory is also related to tracking of the key points and curve matching. Although lots of research works have been conducted on these two issues for now, it is still hard to obtain satisfying result for SLR due to the variation and occlusion of hands and body joints. Besides, it is a nontrivial issue to integrate the hand-shape features and trajectory features together. To address these difficulties, we develop a CNNs to naturally integrate hand-shapes, trajectory of action and facial expression. Instead of using commonly used color images as input to networks like [1, 2], we take color images, depth images and body skeleton images simultaneously as input which are all provided by [18] Microsoft Kinect.

Kinect is a motion sensor which can provide color stream and depth stream. With the public Windows SDK, the body joint locations can be obtained in real-time as shown in Fig.1. Therefore, we choose

Kinect as capture device to record sign words dataset. The change of color and depth in pixel level are useful information to discriminate different sign actions. And the variation of body joints in time dimension can depict the trajectory of sign actions. Using multiple types of visual sources as input leads CNNs paying attention to the change not only in color, but also in depth and trajectory. It is worth mentioning that we can avoid the difficulty of tracking hands, segmenting hands from background and designing descriptors for hands because CNNs [19] have the capability to learn features automatically from raw data without any prior knowledge.

DISADVANTAGES

- CNNs typically require a large amount of labeled data for effective training. Acquiring labeled data for counterfeit image identification can be challenging and time-consuming, especially for niche or specialized domains.
- CNNs are susceptible to adversarial attacks were small carefully crafted perturbations to the input can cause the model to misclassify the image. This vulnerability can be exploited in security-sensitive applications like counterfeit image identification.

4. PROPOSEDSYSTEM

We developed a CNN model for sign language recognition. Our model learns and extracts both spatial and temporal features by performing 2D [12] convolutions. The developed deep architecture extracts multiple types of information from adjacent input frames and then performs convolution and sub-sampling separately. The final feature representation combines information from all channels. We use multilayer perception classifier to classify these feature representations. For comparison[20], we evaluate both CNN on the same dataset. The experimental results demonstrate the effectiveness of the proposed method.

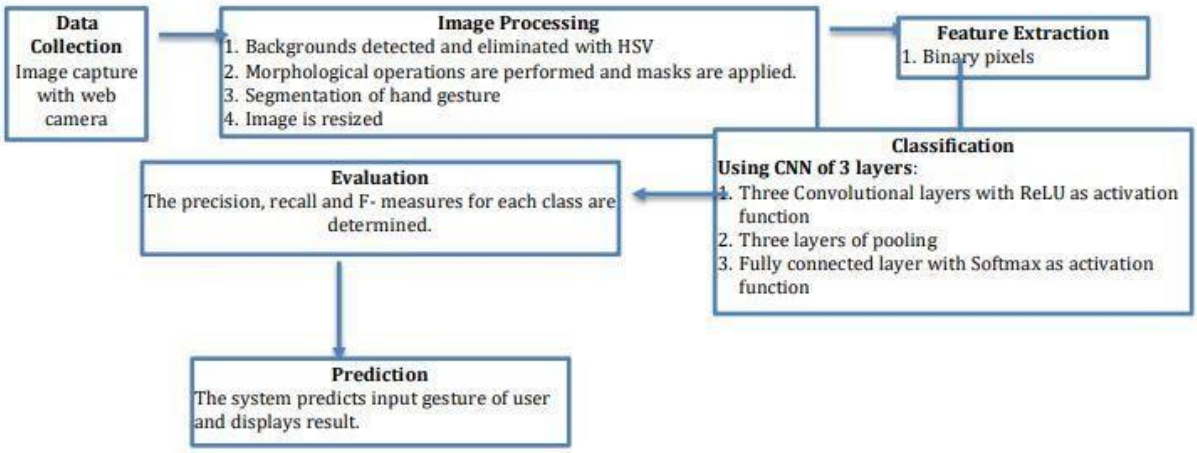
ADVANTAGES

- CNNs can automatically learn hierarchical features from images. This ability allows them to capture both low-level features like edges and textures and high-level features like shapes and patterns, which are crucial for identifying counterfeit images.
- Pre-trained CNN models, such as those trained on ImageNet, are available. These pre-trained models can be fine-tuned on specific counterfeit image identification tasks, reducing the need for extensive training from scratch.
- CNNs can scale well with computational resources. They can be trained on powerful hardware like GPUs and TPUs to handle large datasets and complex model architectures efficiency

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

5. SYSTEM ARCHITECTURE



6. IMPLEMENTATION

1. DATA COLLECTION

Data collection is indelibly an essential part in this research as our result highly depends on it. We have therefore created our own dataset of ASL having 2000 images of 10 static alphabet signs. We have 10 classes of static alphabets which are A,B,C,D,K,N,O,T and Y. Two datasets have been made by 2 different signers. Each of them has performed one alphabetical gesture 200 times in alternate lighting conditions. The dataset folder of alphabetic sign gestures is further split into 2 more folders, one for training and the other for testing. Out of the 2000 images captured, 1600 images are used for training and the rest for testing. To get higher consistency, we have captured the photos in the same background with a webcam each time a command is given. The images obtained are saved in the png format .It is to be pinpointed that there is no loss in quality whenever an image in png format is opened ,closed and stored again.PNG is also good in handling high contrast and detailed image. The webcam will capture the images in the RGB colourspace.

2.DATA PROCESSING

Since the images obtained are in RGB colourspaces, it becomes more difficult to segment the hand gesture based on the skin colour only. We therefore transform the images in HSV colourspace. It is a model which splits the colour of an image into 3 separate parts namely: Hue,Saturation and value. HSV is a powerful tool to improve stability of the images by setting apart brightness from the

Chromaticity.

The Hue element is unaffected by any kind of illumination, shadows and shadings and can thus be considered for background removal. A track-bar having H ranging from 0 to 179, S ranging from 0-255 and V ranging from 0 to 255 is used to detect the hand gesture and set the background to black. The region of the hand gesture undergoes dilation and erosion operations with elliptical kernel.

3.SEGMENTATION

The first image is then transformed to grayscale. As much as this process will result in the loss of colour in the region of the skin gesture, it will also enhance the robustness of our system to changes in lighting or illumination. Non-black pixels in the transformed image are binarized while the others remain unchanged, therefore black. The hand gesture is segmented firstly by taking out all the joined components in the image and secondly by letting only the part which is immensely connected, in our case is the hand gesture. The frame is resized to a size of 64 by 64 pixel. At the end of the segmentation process, binary

images of size 64 by 64 are obtained where the area in white represents the hand gesture, and the black coloured area is the rest.

4.FEATURE EXTRACTION

One of the most crucial part in image processing is to select and extract important features from an image. Images when captured and stored as a dataset usually take up a whole lot of space as they are comprised of a huge amount of data. Feature extraction helps us solve this problem by reducing the data after having extracted the important features automatically. It also contributes in maintaining the accuracy of the classifier and simplifies its complexity. In our case, the features found to be crucial are the binary pixels of the images. Scaling the images to 64 pixels has led us to get sufficient features to effectively classify the American Sign Language gestures . In total, we have 4096 number of features, obtained after multiplying 64 by 64 pixels.

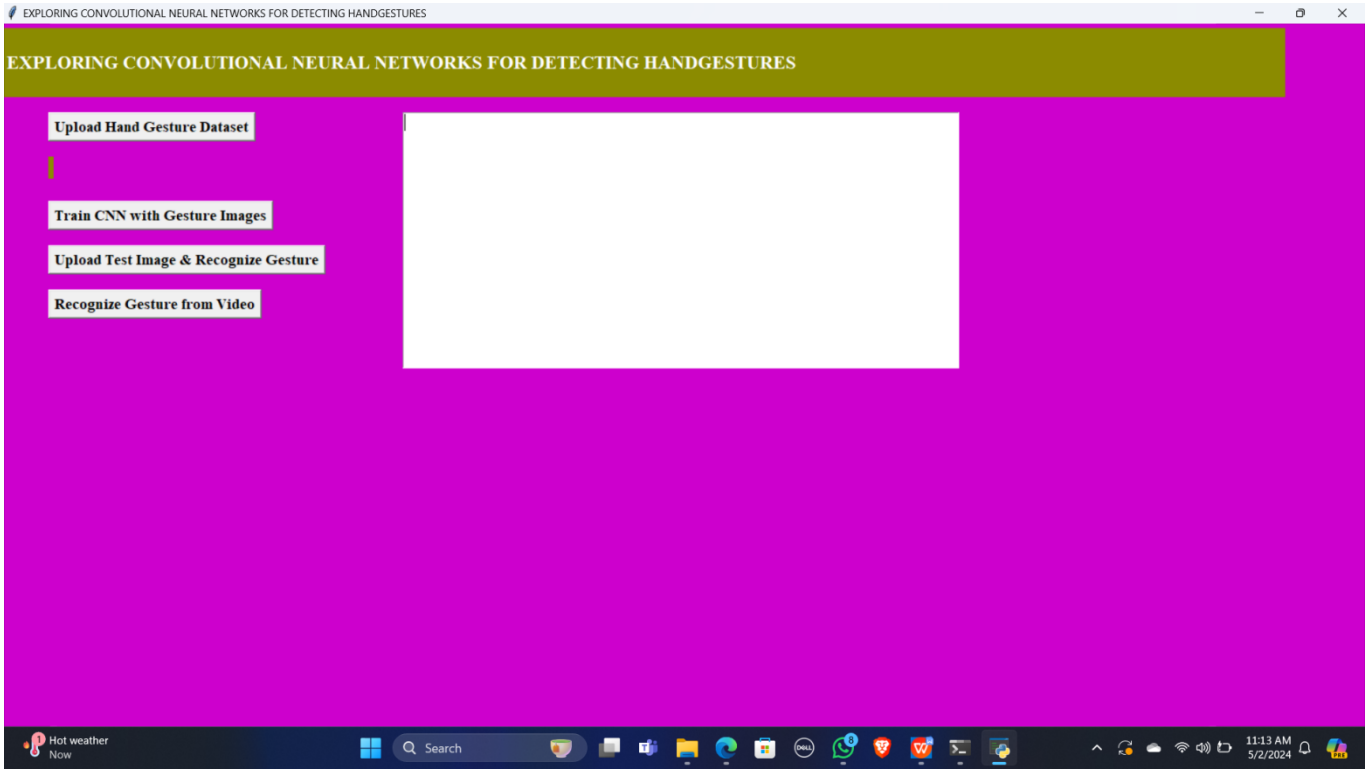
Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

5.CLASSIFICATION

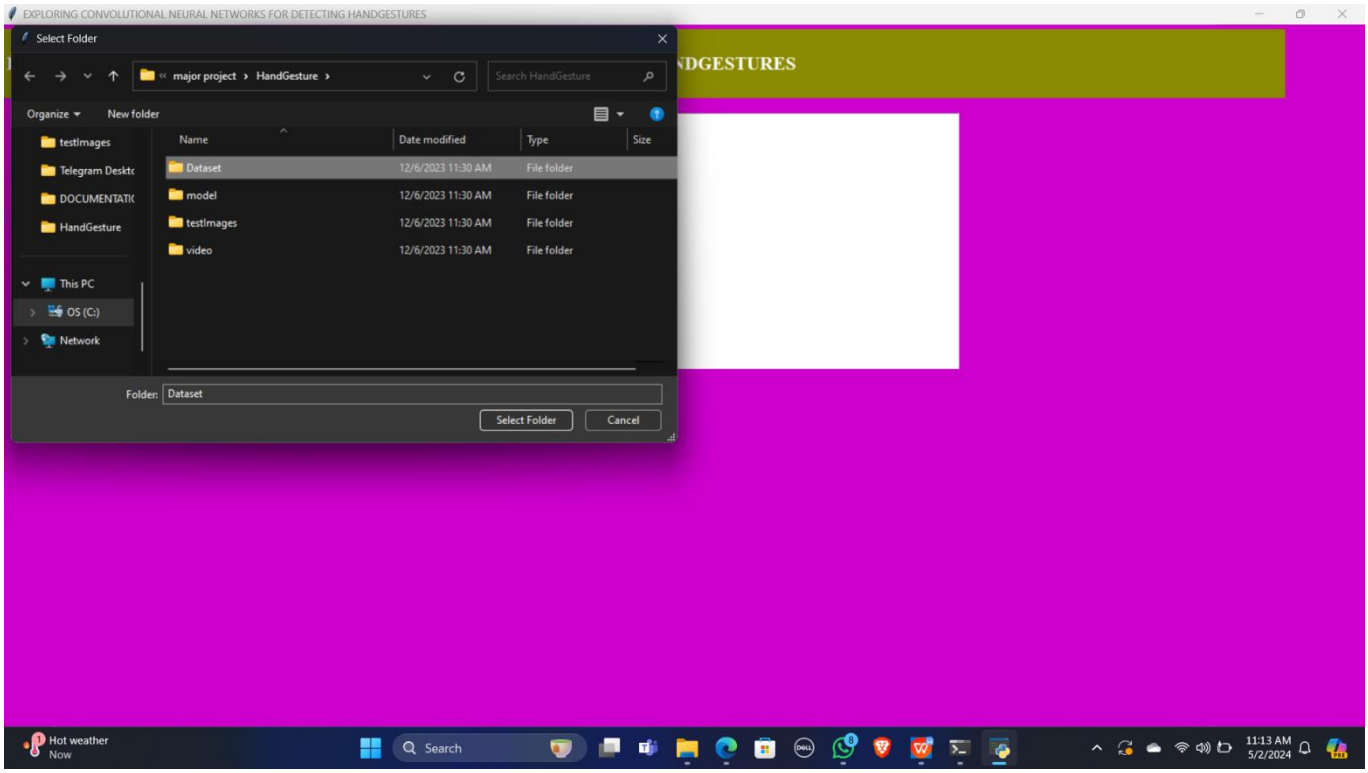
In our proposed system, we apply a 2D CNN model with a tensor flow library. The convolution layers scan the images with a filter of size 3 by 3. The dot product between the frame pixel and the weights of the filter are calculated. This particular step extracts important features from
the input image to pass on further. The pooling layers are then applied after each convolution layer. One pooling layer decrements the activation map of the previous layer. It merges all the features that were learned in the previous layers' activation maps. This helps to reduce overfitting of the training data and generalises the features represented by the network. In our case, the input layer of the convolutional neural network has 32 feature maps of size 3 by 3, and the activation function is a Rectified Linear Unit. The max pool layer has a size of 2×2. The dropout is set to 50 percent and the layer is flattened. The last layer of the network is a fully connected output layer with ten units, and the activation function is Softmax. Then we compile the model by using category cross-entropy as the loss function and Adam as the optimiser.

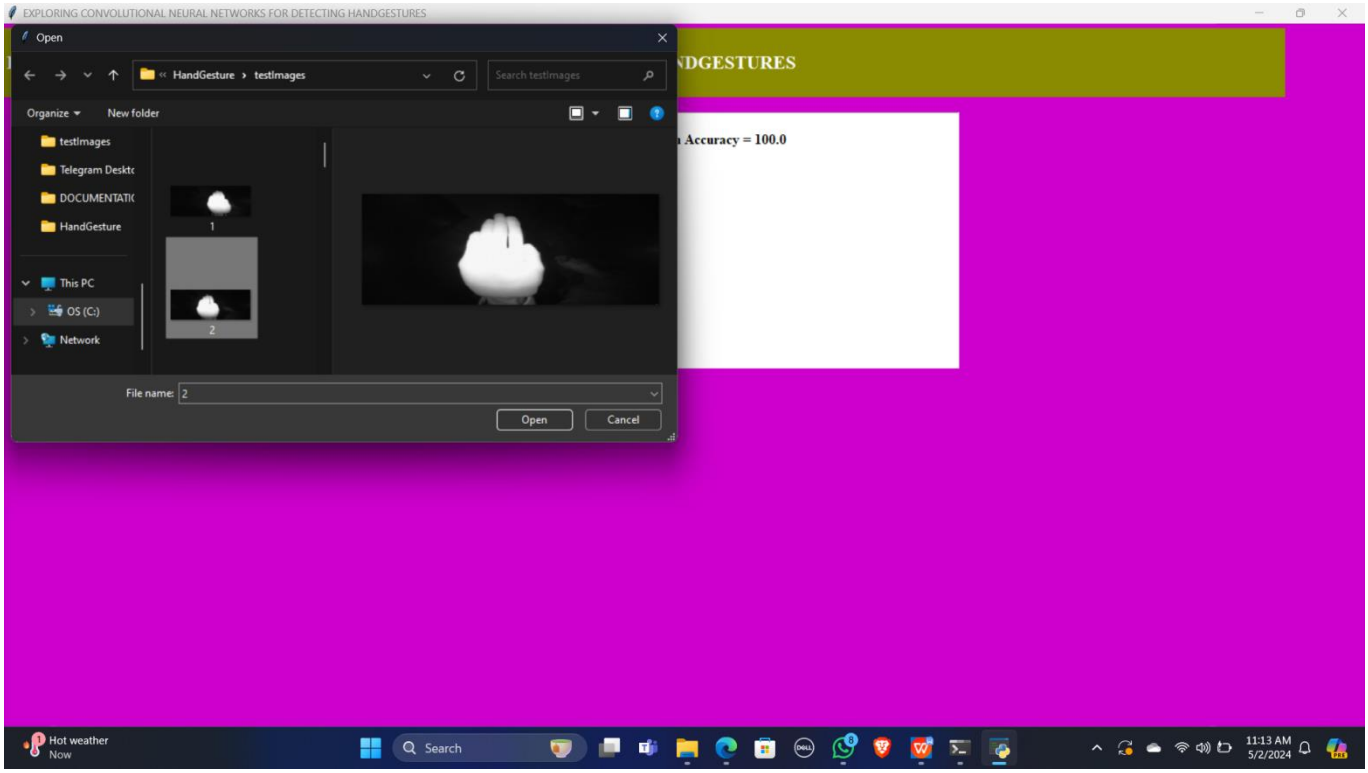
7. OUTCOMES



Industrial Engineering Journal
ISSN: 0970-2555

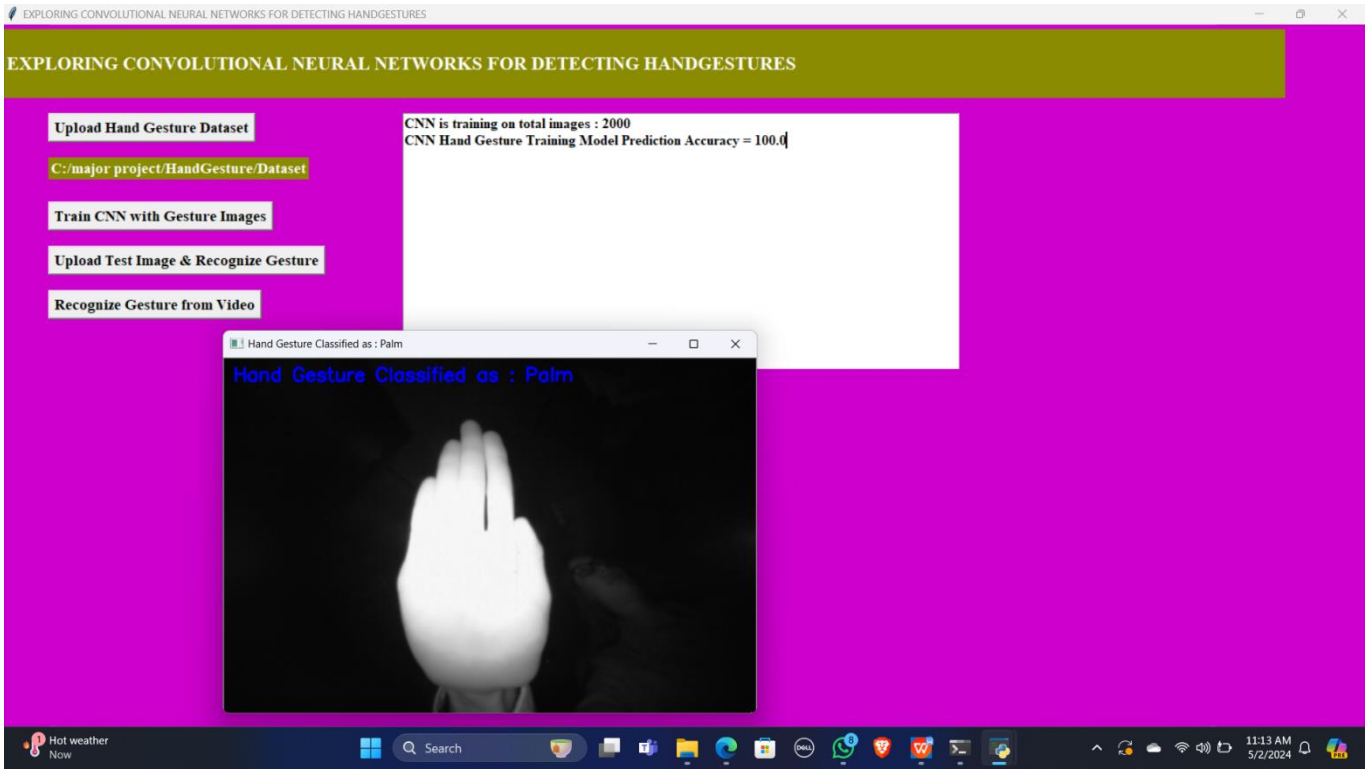
Volume : 53, Issue 5, May : 2024

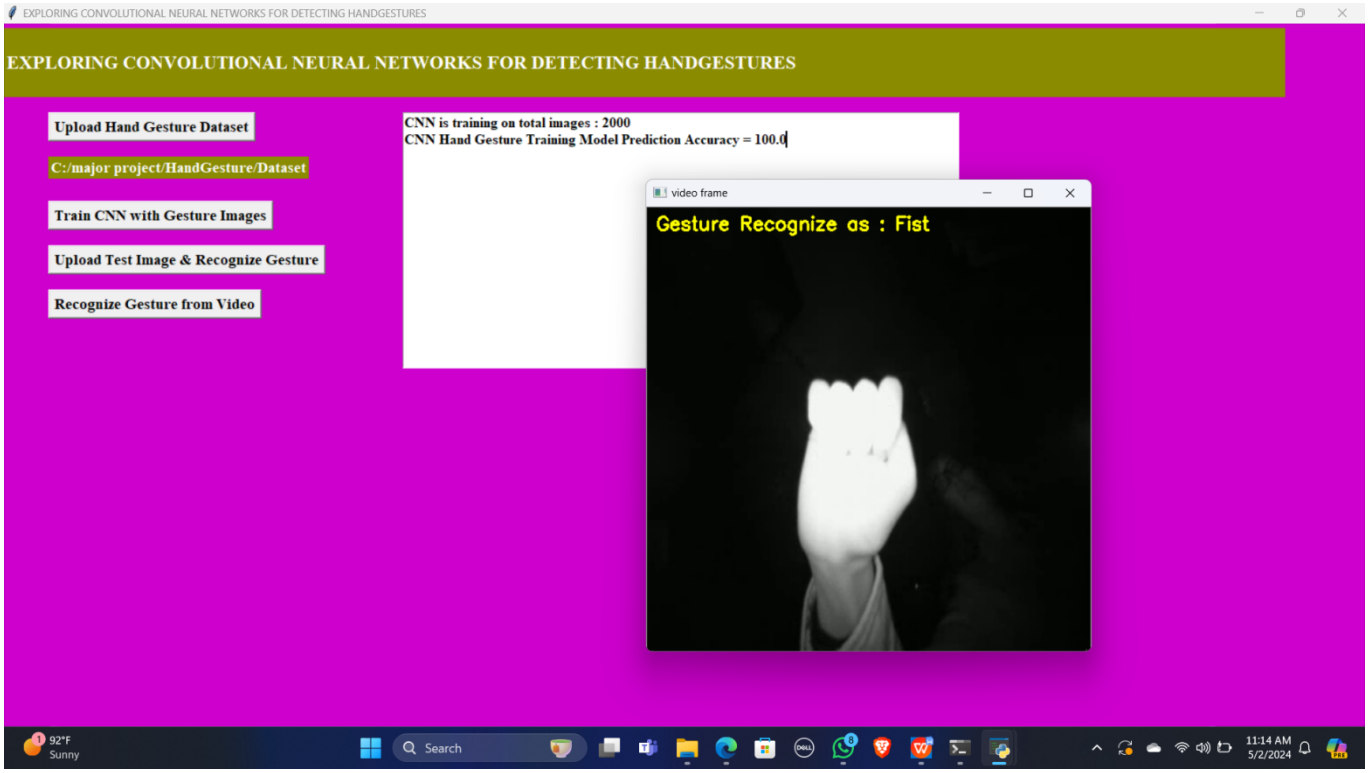




Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024





Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

8.CONCLUSION

Our study on sign language recognition using Convolutional Neural Networks (CNN) highlighted the immense diversity and complexity of sign languages, which differ across countries in terms of gestures, body language, and sentence structures. Capturing precise hand movements and creating a comprehensive dataset posed challenges, as some gestures proved difficult to reproduce accurately. Consistent hand positions during data collection were critical to maintaining dataset quality. Furthermore, understanding the unique grammatical rules and contextual nuances of each sign language was essential to develop a robust recognition system. Despite the challenges, our research underscored the significance of recognizing and preserving the richness and expressiveness of sign languages, and we remain committed to advancing assistive technologies for improved communication and inclusivity in the future.

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

8.1 FUTURESCOPE

Certainly! Future scopes for exploring CNN for detecting hand gestures:

1. Enhanced Architectures:

Develop streamlined CNN architectures tailored for hand gesture recognition, focusing on simplicity, efficiency, and accuracy.

2. Edge Computing Optimization:

Optimize CNN models for real-time inference on edge devices like smartphones and IoT gadgets, ensuring minimal computational resources while maintaining performance.

3. Transfer Learning:

Explore transfer learning techniques to leverage pre-trained models and adapt them to new hand gesture datasets, facilitating faster training and improved accuracy.

4. Privacy Preservation:

Investigate privacy-preserving methods to ensure user data confidentiality in hand gesture recognition systems, promoting trust and security.

5. Multi-modal Fusion: Integrate multiple sensory modalities, like depth sensors or infrared cameras, with visual data to enhance gesture recognition robustness and accuracy.

6. Gesture Understanding:

Expand gesture recognition systems beyond classification tasks, focusing on gesture understanding and interaction for applications like human-computer interaction and sign language recognition.

7. Accessibility and Inclusivity:

Incorporate human-centric design principles to make gesture-based interfaces accessible and usable for diverse user populations, including those with disabilities.

By pursuing these avenues, researchers and developers can advance hand gesture recognition

technology, making it more efficient, accurate, and inclusive for a wide range of applications

9. REFERENCESAND BOOKS

1. https://peda.net/id/08f8c4a8511
2. K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.
3. CABRERA, MARIA & BOGADO, JUAN & Fermín, Leonardo & Acuña, Raul & RALEV,
4. DIMITAR. (2012). GLOVE-BASED GESTURE RECOGNITION SYSTEM.
5. 10.1142/9789814415958_0095.
6. He, Siming. (2019). Research of a Sign Language Translation System Based on Deep Learning. 392-396. 10.1109/AIAM48774.2019.00083.
7. International Conference on Trendz in Information Sciences and Computing (TISC). : 30-35, 2012.
8. Herath, H.C.M. & W.A.L.V.Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi. (2013). IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE
9. M. Geetha and U. C. Manjusha, , “A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation”, International Journal on Computer Science and Engineering (IJCSE), vol. 4, no. 3, pp. 406-415. 2012.
10. Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham. https://doi.org/10.1007/978-3-319-16178-5_40
11. Escalera, S., Baró, X., Gonzàlez, J., Bautista, M., Madadi, M., Reyes, M., . . . Guyon, I. (2014). ChaLearn Looking at People Challenge 2014: Dataset and Results. Workshop at the European Conference on Computer Vision (pp. 459-473). Springer, . Cham.
12. Huang, J., Zhou, W., & Li, H. (2015). Sign Language Recognition using 3D convolutional neural networks. IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-

6). Turin: IEEE.

13. Jaoa Carriera, A. Z. (2018). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on (pp. 47244733). IEEE. Honolulu.
14. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A LargeScale Hierarchical Image Database. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255). IEEE. Miami, FL, USA .
15. Soomro, K., Zamir , A. R., & Shah, M. (2012). UCF101: A Dataset of 101 Human Actions
16. Classes From Videos in The Wild. Computer Vision and Pattern Recognition, arXiv:1212.0402v1, 1-7.
17. Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011). HMDB: a large video database for human motion recognition. Computer Vision (ICCV), 2011 IEEE International Conference on (pp. 2556-2563). IEEE
18. Zhao, Ming & Bu, Jiajun & Chen, C.. (2002). Robust background subtraction in HSV color space. Proceedings of SPIE MSAV, vol. 1. 4861. 10.1117/12.456333.
19. [11][16] Chowdhury, A., Sang-jin Cho, & Ui-Pil Chong. (2011). A background subtraction method using color information in the frame averaging process. Proceedings of 2011 6th
20. International Forum on Strategic Technology. doi:10.1109/i .