USB Proceedings

The 20th International Conference on

Modeling Decisions for Artificial Intelligence

MDAI 2023, Umeå

Vicenç Torra, Yasuo Narukawa

Image from wikipedia.org

USB Proceedings

The 20th International Conference on

Modeling Decisions for Artificial Intelligence

MDAI 2023, Umeå, Sweden
19 - 22 June, 2023

Editors:

Vicenç Torra
Umeå University
Umeå, Sweden
E-mail: vtorra@ieee.org

Yasuo Narukawa
Department Management Science
Tamagawa University
Tokyo
Japan
E-mail: nrkwy@eng.tamagawa.ac.jp

ISBN: 978-91-527-7293-5

Preface

This volume contains papers that had to be presented at the 20th International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2023) celebrated in Umeå, Sweden, 19 - 22 June, 2023. The rest of papers as well as invited papers have been separately published in the Lecture Notes in Artificial Intelligence, Vol. 13890 (by Springer).

This conference followed MDAI 2004 (Barcelona), MDAI 2005 (Tsukuba), MDAI 2006 (Tarragona), MDAI 2007 (Kitakyushu), MDAI 2008 (Sabadell), MDAI 2009 (Awaji Island), MDAI 2010 (Perpinyà), MDAI 2011 (Changsha), MDAI 2012 (Girona), MDAI 2013 (Barcelona), MDAI 2014 (Tokyo), MDAI 2015 (Skövde), MDAI 2016 (Sant Julià de Lòria), MDAI 2017 (Kitakyushu), MDAI 2018 (Mallorca), MDAI 2019 (Milano), MDAI 2020, MDAI 2021 (Umeå), and MDAI 2022 (Sant Cugat).

The aim of MDAI is to provide a forum for researchers to discuss different facets of decision processes in a broad sense. This includes model building and all kinds of mathematical tools for data aggregation, information fusion, and decision-making; tools to help make decisions related to data science problems (including, e.g., statistical and machine learning algorithms as well as data visualization tools); and algorithms for data privacy and transparency-aware methods so that data processing procedures and the decisions made from them are fair, transparent, and avoid unnecessary disclosure of sensitive information.

The MDAI conference included tracks on the topics of (a) data science, (b) machine learning, (c) data privacy, (d) aggregation functions, (e) human decision-making, and (f) graphs and (social) networks.

The conference celebrates this year the 50th anniversary of graded logic, introduced by Jozo Dujmović in a paper in 1973. In such paper, he also introduced the concept of andness, a key concept to define adjustable aggregators with a variable conjunction degree.

The conference was supported by Umeå University, the European Society for Fuzzy Logic and Technology (EUSFLAT), the Catalan Association for Artificial Intelligence (ACIA), the Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT), and the UNESCO Chair in Data Privacy.

Vicenç Torra, Yasuo Narukawa
June, 2023

General Chairs

Vicenç Torra, Umeå University

Program Chairs

Vicenç Torra, Umeå University, Sweden
Yasuo Narukawa, Tamagawa University, Japan

Advisory Board

Didier Dubois, Institut de Recherche en Informatique de Toulouse, CNRS, France
Jozo Dujmović, San Francisco State University, USA
Lluis Godo, IIIA-CSIC, Spain
Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Poland
Cengiz Kahraman, Istanbul Technical University, Turkey
Sadaaki Miyamoto, University of Tsukuba, Japan
Pierangela Samarati, Università degli Studi di Milano, Italy
Sandra Sandri, Instituto Nacional de Pesquisas Espaciais, Brazil
Michio Sugeno, Tokyo Institute of Technology, Japan
Ronald R. Yager, Machine Intelligence Institute, Iona Collegue, USA

Program Committee

Kayode S. Adewole, Umeå University, Sweden
Laya Aliahmadipour, Shahid Bahonar University, Iran
Cláudia Antunes, Universidade de Lisboa, Portugal
Eva Armengol, IIIA-CSIC, Spain
Edurne Barrenechea, Universidad Pública de Navarra, Spain
Gloria Bordogna, Consiglio Nazionale delle Ricerche, Italy
Humberto Bustince, Universidad Pública de Navarra, Spain
Alina Campan, North Kentucky University, USA
Francisco Chiclana, De Montfort University, UK
Susana Díaz, Universidad de Oviedo, Spain
Josep Domingo-Ferrer, Universitat Rovira i Virgili, Spain
Yasunori Endo, University of Tsukuba, Japan
Vladimir Estivill-Castro, Griffith University, Australia
Zoe Falomir, Universitat Jaume I, Spain
Javier Fernandez, Universidad Pública de Navarra, Spain
Katsushige Fujimoto, Fukushima University, Japan
Joaquin Garcia-Alfaro, Institut Mines-Télècom and Institut Polytechnique de Paris, France

Michel Grabisch, Université Paris I Panthéon-Sorbonne, France
Yukihiro Hamasuna, Kindai University, Japan
Tove Helldin, University of Skövde, Sweden
Enrique Herrera-Viedma, Universidad de Granada, Spain
Aoi Honda, Kyushu Institute of Technology, Japan
Van-Nam Huynh, JAIST, Japan
Masahiro Inuiguchi, Osaka University, Japan
Simon James, Deakin University, Australia
Aránzazu Jurío, Universidad Pública de Navarra, Spain
Yuchi Kanzawa, Shibaura Institute of Technology, Japan
Ali Karaşan, Yildiz Technical University, Turkey
Hiroaki Kikuchi, Meiji University, Japan
Petr Krajča, Palacky University Olomouc, Czech Republic
Marie-Jeanne Lesot, Université Pierre et Marie Curie (Paris VI), France
Giovanni Livraga, Università degli Studi di Milano, Italy
Jun Long, National University of Defense Technology, China
Beatriz López, University of Girona, Catalonia, Spain
Jean-Luc Marichal, University of Luxembourg, Luxembourg
Radko Mesiar, Slovak University of Technology, Slovakia
Andrea Mesiarová-Zemánková, Slovak Academy of Sciences, Slovakia
Anna Monreale, University of Pisa, Italy
Pranab K. Muhuri, South Asian University, India
Toshiaki Murofushi, Tokyo Institute of Technology, Japan
Guillermo Navarro-Arribas, Universitat Autònoma de Barcelona, Spain
Shekhar Negi, Umeå University, Sweden
Jordi Nin, Esade, Universitat Ramon Llull, Spain
Miguel Nunez-del-Prado, Universidad del Pacífico, Peru
Anna Oganyan, National Institute of Statistical Sciences (NISS), USA
Gabriella Pasi, Università di Milano Bicocca, Italy
Oriol Pujol, University of Barcelona, Catalonia, Spain
Maria Riveiro, Jönköping University, Sweden
Julian Salas, Universitat Oberta de Catalunya, Catalonia, Spain
Robyn Schimmer, Umeå University, Sweden
H. Joe Steinhauer, University of Skövde, Sweden
László Szilágyi, Sapientia-Hungarian Science University of Transylvania, Hungary
Aida Valls, Universitat Rovira i Virgili, Spain
Paolo Viappiani, Université Paris Dauphine, France
Zeshui Xu, Southeast University, China

Local Organizing Committee Chair

Vicenç Torra, Umeå University, Sweden

Additional Referees

Sergio Martinez Lluis, Najeeb Moharram Salim Jebreel, Rami Haffar

Supporting Institutions

Umeå University
The European Society for Fuzzy Logic and Technology (EUSFLAT)
The Catalan Association for Artificial Intelligence (ACIA)
The Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
The UNESCO Chair in Data Privacy

Table of Contents

Regular Papers

Hand Pose Recognition through MediaPipe Landmarks

Manuel Gil-Martín[1[0000-0002-4285-6224]], Rubén San-Segundo[1[0000-0001-9659-5464]] and Ricardo de Córdoba[1[0000-0002-7136-9636]]

[1] Speech Technology and Machine Learning Group (T.H.A.U. Group),
Information Processing and Telecommunications Center, E.T.S.I. de Telecomunicación,
Universidad Politécnica de Madrid, 28040, Madrid, Spain

Abstract. This paper proposes a framework to recognize hand poses using a limited number of landmarks from images. This Hand Pose Recognition (HPR) system is composed of a signal processing module that extracts and processes the coordinates of specific points of the hand called landmarks, and a deep neural network module that models and classifies the hand poses. These specific points or landmarks are extracted automatically through MediaPipe software. Detecting hand poses from these points has two main advantages compared to traditional computer vision approaches: the information sent to the recognition module is smaller (points' coordinates vs. a full image) and the classification is not affected by additional information included in the images (like the background). The experiments were carried out over two different datasets using the experimental setups of previous works. The proposed framework was able to obtain better performance than the best results reported in previous works. For example, in case of using the Tiny Hand Gesture Recognition Dataset, we obtained classification accuracies of 98.74 ± 0.08 % and 98.22 ± 0.06 % with simple or complex backgrounds, while the best reported accuracies in previous works (using the whole image) were 97.10 % and 85.30 % respectively. The proposed solution is able to provide high recognition performance independently of the background where the image is taken.

Keywords: Hand Pose Recognition, MediaPipe, Hand landmarks, Deep learning, Convolutional Neural Networks.

1 Introduction

Hand Pose Recognition consists in detecting the posture or pose that people perform using their hands. This technology could be useful to develop human computer interaction systems and could improve the user experience across a wide variety of different domains. For example, it could be seen as the basis for sign language understanding and hand gesture control applications. For instance, a person could ask for taking a picture using the front camera of a smartphone by opening and closing the hand palm. In these applications, it is crucial to accurately recognize the hand pose or gesture to perform specific actions with smart devices, a computer or an automatic transmission machine. In this context, computer vision based approaches have been applied reaching promising results.

However, computer vision based approaches are usually based on feeding the systems by raw images that include sensitive information like the face or the background that people would like not to share. In addition, these images could have large sizes and cause strain on bandwidth in real applications. This way, it could be great to study solutions that extract the strictly necessary information from the images in order to develop lighter systems that could respect the individual's privacy.

This paper aims to propose a framework to detect hand poses using a limited number of landmarks from images. The main contributions of the paper are:

- The proposal of a framework to detect hand poses from specific points of images that is not affected by the background of the images nor the people who perform the pose.
- The evaluation of the proposal using two datasets and a comparison to previous works that used the whole images as input using the same experimental setups.

This paper is organized as follows. Section 2 reviews the related work on hand pose recognition. Section 3 describes the material and methods used, including the datasets, the system architecture including the signal processing and deep learning approaches and the evaluation details. Section 4 discusses the experiments and the obtained results. Finally, Section 6 summarizes the main conclusions of the paper.

2 Related work

Multiple previous works have been focused on Human Activity Recognition in order to optimize the physical activity classification using wearables or cameras [1-4] that could be traditionally applied to sports monitoring purposes [5-7] such as fitness tracking, personal incentivizing, or rehabilitation [8]. However, there exist a lower number of works focused on detecting hand poses or gestures. Most of these works use images as inputs of their systems and follow a hand localization step as first stage. Afterwards, they extracted handcrafted features or descriptors [9] from the hand and feed an inference algorithm that classifies the different hand poses. As mentioned in the introduction, most hand detection systems are based on computer vision approaches [10-12] which often use raw images.

For example, Wang et al. [13] developed a hand pose recognition system where they first obtained a segmented hand map using Kinect software development kit, then they extracted a volumetric shape descriptor using the line between the center of hand and wrist as polar axis of polar coordinates and finally they used a Support Vector Machines classifier to perform hand pose recognition.

Another previous work [14] proposed a Convolutional Neural Network based system to model ten different hand poses from ten different people. They pre-filtered the images using a Gabor filter and used skin color distribution as descriptor of the hand

to feed the deep learning architecture. They obtained an accuracy of 97% in the person-independent test.

In the same way, a previous work [15] segmented the image into the hand in different regions and obtained the Histogram of Oriented Gradients and a Local Binary Pattern from each region. Afterwards, they combined k-means and Support Vector Machines in order to classify the hand poses, obtaining an F1 score near 96% using data from 25 subjects and 16 different hand poses.

Similarly a previous work [16] feed a deep Convolutional Neural Network to directly classify hand poses in images without any previous segmentation. They classified the hand pose with average accuracy of 97.1% in the images with simple backgrounds and 85.3% in the images with complex backgrounds. They used a dataset with data from 40 subjects and seven different hand poses.

To summarize, existing methods combined a hand segmentation step with a handcrafted features extraction step to obtain a descriptive hand pose representation. Afterwards, they fed a machine learning module to model and classify the hand poses. The aim of this work is to use a powerful and effective library to directly extract representative relevant points from the hand images (landmarks) to model and classify hand poses through a deep learning solution. In this sense, it is hypothesized that hand pose recognition task could be performed by the combination of landmarks extraction and a deep learning architecture offering higher performance without handcrafted methods to process the input images.

3 Material and methods

This section includes information about the datasets used in this work, the proposed system architecture including the signal processing and the deep neural network and the evaluation of the system.

3.1 Datasets

For this work, we have used two publicly available hand pose datasets: Multi-modal Leap Motion dataset for Hand Gesture Recognition [15] and Tiny Hand Gesture Recognition Dataset [16].

Multi-modal Leap Motion dataset for Hand Gesture Recognition includes data from 25 subjects (8 women and 17 men) that performed 16 different hand poses. Each subject was placed in front of a computer with the Leap Motion located on a table between the subject and the computer for image collection. Each subject was free to move the right hand over the device inside the Leap Motion field of view. The hand poses included in this dataset are: L, fist moved, index, ok, C, heavy, hang, two, three, four, five, palm, down, palm moved, palm up, and up. This dataset contains a total number of frames of 65,156 related to hand poses.

Tiny Hand Gesture Recognition Dataset contains data from 40 subjects (14 women and 26 men) that performed seven different hand poses. Half of the subjects were recorded with gray simple background and the rest with complex background. The considered complex backgrounds are highly cluttered and the illumination undergoes large variations. The hand poses included in this dataset are: fist, L, ok, palm, pointer, thumb down and thumb up. This dataset contains a total number of frames of 260,796.

3.2 System architecture

Fig 1 shows a diagram module of the system: a data acquisition step where the images are collected, a signal processing module where landmarks are extracted from the images and processed, and a deep learning network to model and classify the hand poses.

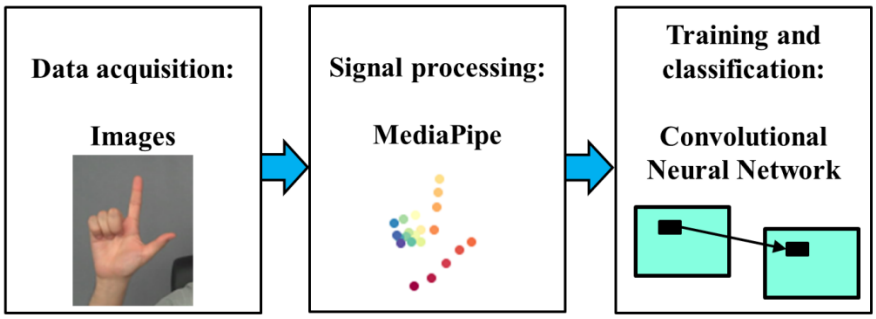

Fig 1. System architecture for hand poses recognition using MediaPipe.

Signal processing module

MediaPipe [17] is a library with the capacity to track hands from input frames or video streams. This framework offers a wide variety of solutions, such as face detection, face mesh, hair segmentation, object detection or pose and hands tracking. In particular, we used the MediaPipe Hands software to extract x and y coordinates of 21 landmarks from the hand. These coordinates are normalized to [0.0, 1.0] interval by the image width and height respectively. The 21 landmarks correspond to different location of the hand area: wrist and four points along the five fingers. Fig 2 shows the landmarks of different hand poses in the datasets used in this work.

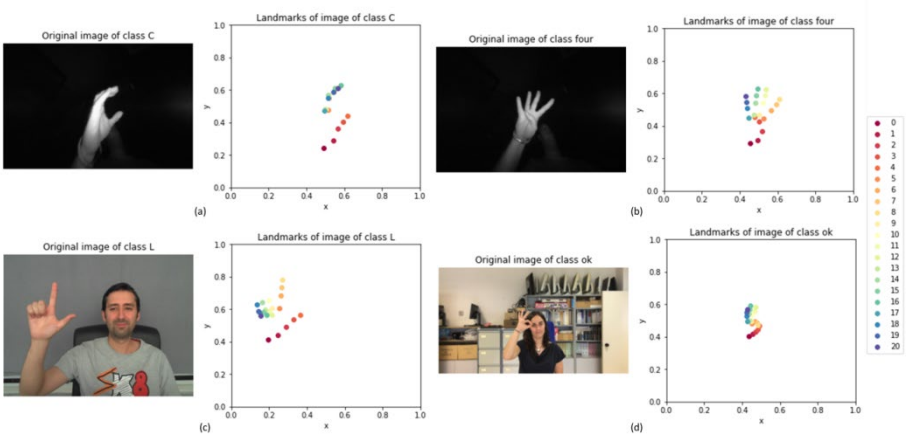

Fig 2. Original images and landmarks of different examples of the datasets used in this work (a) and (b) from Multi-modal Leap Motion Dataset and (c) and (d) from Tiny Hand Gesture Recognition Dataset.

The proposed framework extracts the landmarks from the images using the MediaPipe library. After obtaining the landmarks, a specific normalization of the coordinates is applied in order to help the neural network to model the hand poses. This normalization consists in using the lower landmark of the palm (wrist) as reference and subtracting their coordinates to the rest of landmarks. **Fig 3** shows the original and normalized landmarks of an example of class four, where it is possible to observe that the reference of the different poses becomes the coordinate origin instead of the wrist landmark.

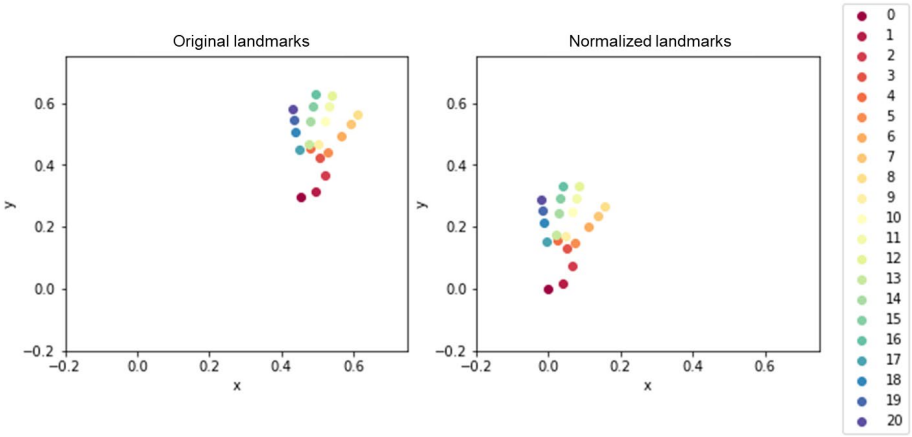

Fig 3. Original and normalized landmarks of an example.

Deep learning approach

The deep learning architecture used in this work was composed of two main parts: a feature learning subnet and a classification subnet. The first subnet learnt features from the x and y coordinates of the different landmarks, using two convolutional layers. The second subnet used fully connected layers to classify the learned features as a predicted hand pose. The architecture included dropout layers (0.3) after convolutional and and fully connected layers to avoid overfitting during training. The last layer used a softmax activation function to offer the predictions of each class for every analysis frame, while intermediate layers used ReLU for reducing the impact of gradient vanishing effect. We used categorical cross-entropy as loss metric and the root-mean-square propagation method as optimizer. We adjusted the epochs and batch size of the deep learning structure for each dataset: 300 and 500 for the Multi-modal Leap Motion dataset for Hand Gesture Recognition and 5 and 500 for the Tiny Hand Gesture Recognition Dataset. The difference between the numbers of epochs in each configuration is related of the number of examples to train the network, which is higher in the second dataset. Fig 4 represents the architecture used in this work to model and classify the hand poses of the datasets, where C indicates the number of recognized hand poses.

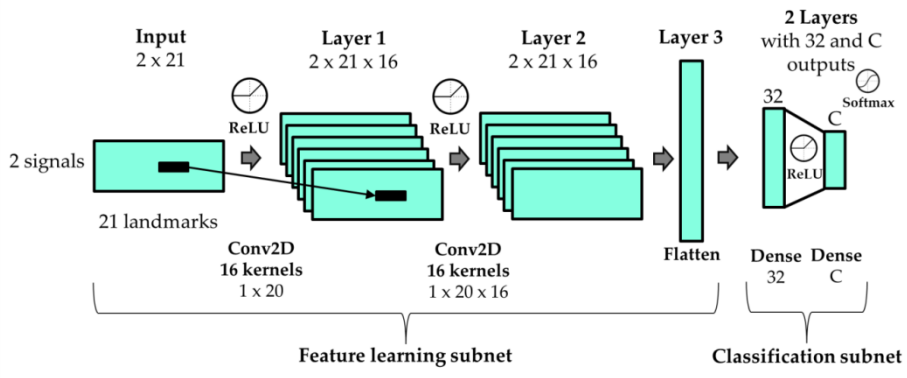

Fig 4. Convolutional Neural Network Architecture used in this work for all the datasets.

3.3 Evaluation setup

In this work, we considered the data distributions of the previous works: specific train and test subsets for Multi-modal Leap Motion dataset for Hand Gesture Recognition and a cross-validation strategy for Tiny Hand Gesture Recognition Dataset.

In case of training and testing subset, data from the same subject were included in both subsets. This methodology provided an optimistic scenario where the system was evaluated with recordings from subjects who were processed during the training step. This methodology was used for the Multi-modal Leap Motion dataset for Hand Gesture Recognition to follow the same experimental setup of a previous work [15] using

this dataset, where the training subset contained 48,436 frames and the testing subset contained 16,720 frames.

In case of the cross-validation experimental setup, 25 people were used for training, 5 subjects for validation, and 10 people for testing. In these experiments, it was assured that all the recordings from the same subject are included only in a subset. Once the system model is fitted on the training subset, the validation subset was used for optimizing the model hyperparameters. Finally, the system was evaluated with the testing subset. This process was repeated several times leaving different subjects for testing in each iteration. The results were averaged along all trials. This methodology simulated a difficult scenario because the system was evaluated with recordings from subjects different to those used for training. This methodology was used for the Tiny Hand Gesture Recognition Dataset to follow the same experimental setup of a previous work [16] using this dataset.

As evaluation metrics, we used accuracy, which is defined as the ratio between the number of correctly classified samples and the number of total samples. Considering a classification problem with N testing samples and C classes, accuracy is defined in Equation (1).

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{C} P_{ii} \tag{1}$$

Considering R_i as the sum of all examples in a column of the confusion matrix, and S_i as the sum of all examples in a row, precision, recall and F1 score metrics are defined as follows:

$$\text{precision} = \frac{1}{C}\sum_{i=1}^{C} \frac{P_{ii}}{R_i} \tag{2}$$

$$\text{recall} = \frac{1}{C}\sum_{i=1}^{C} \frac{P_{ii}}{S_i} \tag{3}$$

$$\text{F1 score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

Confidence intervals are used to show statistical significance values and provide confidence about the results reliability. These intervals include plausible values for a specific metric. We will assure that there exists a significant difference between results of two experiments when their confidence intervals do not overlap. Equation (5) represents the computation of confidence intervals attached to a specific metric value and N samples when the confidence level is 95%.

$$\mathrm{CI}(95\%) = \pm 1.96 \sqrt{\frac{\text{metric} \cdot (100 - \text{metric})}{\mathrm{N}}} \tag{5}$$

4 Experiments and Discussion

We firstly analyzed the effect of normalizing the coordinates using the lower landmark of the palm (wrist) as reference and subtracting their coordinates to the rest of landmarks. In this sense, the reference of the different poses becomes the coordinate origin. We observed that we could increase the recognition accuracy from 96.45 ± 0.28 % to 97.25 ± 0.25 % for Multi-modal Leap Motion dataset for Hand Gesture Recognition and from 98.52 ± 0.09 % to 98.74 ± 0.08 % for simple backgrounds of Tiny Hand Gesture Recognition Dataset. This normalization offers a slight increment of performance. However, it is fair to say that this improvement of performance is significant even in the difficult situation when performance is high. One of the reasons of this improvement is that thanks to this normalization, the representation of examples of the same pose become similar independently of the location of the hand in the image. For example, the representation of a hand pose consisting in pointing a screen with one finger could differ when the person performs the pose at right of left side of the image. However, thanks to normalizing using the wrist landmark, both representations become similar since both use the coordinate origin as reference. Regarding computational cost, the normalization does not heavily increase the processing time thanks to simple operations that are used.

Second, we compared our solution to previous works using the same datasets and their data distributions. A previous work [15] using the Multi-modal Leap Motion dataset for Hand Gesture Recognition segmented the image into the hand in different regions and obtained the Histogram of Oriented Gradients and a Local Binary Pattern from each region. Afterwards, the system combined k-means and Support Vector Machines in order to classify the hand poses, obtaining an F1 score near 96% using specific training and testing subsets. Another previous work [16] used the Tiny Hand Gesture Dataset and increased the number of samples by performing synthetic translations over the whole images, reaching a total number of 500,000 hand gesture samples for training. This system used a deep convolutional neural network (composed by 9 convolutional layers, 4 pooling layers, 3 fully connected layers, interlaced with ReLU and dropout layers) to directly classify hand poses in images without any previous segmentation. This previous work classified the hand pose with average accuracy of 97.1% in the images with simple backgrounds and 85.3% in the images with complex backgrounds following a cross-validation experimental setup, where 25 people were used for training, 5 subjects for validation, and 10 people for testing. Table 1 includes the comparison of these previous works and our work using the mentioned normalization of landmarks.

Table 1. Results considering baseline experimental setups for the datasets.

Dataset	Work	Accuracy (%)	F1 score (%)
Multi-modal Leap Motion Dataset for Hand Gesture Recognition	[15]	-	96.00
	This work	97.25 ± 0.25	97.23 ± 0.25
Tiny Hand Gesture Recognition Dataset – Simple background	[16]	97.10	-
	This work	98.74 ± 0.08	98.74 ± 0.08
Tiny Hand Gesture Recognition Dataset – Complex background	[16]	85.30	-
	This work	98.22 ± 0.06	98.23 ± 0.06

As observed, we obtained better performance than these previous works. One of the interesting aspects of using our approach is that the system does not suffer a decrement of performance when dealing with complex backgrounds. One of the reasons is that once the landmarks are extracted, the process is the same independently of the context and background where the image was taken. Additionally, the system can process images of different dimensions: in previous works, the deep neural architecture that directly processes the images requires a specific input image dimensions and uses specific convolutional kernel sizes to learn relevant features from them. Nevertheless, as the extraction of landmarks does not depend on the neural architecture in the proposed approach, it is not restricted to specific image dimensions.

However, the proposed approach could have some limitations that should be address in future works: managing blurred images or images without complete hands. In these situations, the MediaPipe tool can have problems extracting the landmarks.

5 Conclusions

This paper proposes an alternative framework to detect hand poses using a limited number of landmarks from images. This approach for Hand Pose Recognition automatically extracts 21 MediaPipe landmarks (x and y coordinates of specific points) from the hand and feeds a deep neural architecture to model and recognize different hand poses. This solution obtained better results than previous works using the same datasets. For example, in case of using the Tiny Hand Gesture Recognition Dataset, classification accuracies of 98.74 ± 0.08 % and 98.22 ± 0.06 % with simple or complex backgrounds, respectively. Moreover, detecting hand poses from these points or landmarks has two main advantages compared to traditional computer vision approaches: the information sent to the recognition module is smaller (coordinates from the points vs. a full image) and the classification is not affected by additional information included in the images (like the background). In this sense, the proposed system could offer a high performance without handcrafted methods to process the input images.

As future work, it would be interesting to recognize gestures as a sequence of frames using landmarks, study the effect of changing the brightness and contrast of

the recorded images before extracting the landmarks, study the effect of including a non-gesture class to distinguish when the system is not able to extract landmarks, detect which hand is processed to avoid errors when both hands appear in the image, perform the hand pose recognition using a Leave One Subject Out Cross Validation methodology and automatically compute the remaining landmarks when only a part of the hand appears in the original image. In addition, it would be interesting to apply this framework for other datasets with wider variety of backgrounds and/or related to sign language recognition with a higher number of classes.

Acknowledgements

The work leading to these results was supported by the Spanish Ministry of Science and Innovation through the project AMIC-PoC, and BeWord (PDC2021-120846-C42 and PID2021-126061OB-C43, funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR"). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

References

1. M. Gil-Martin, R. San-Segundo, F. Fernandez-Martinez, and J. Ferreiros-Lopez, "Time Analysis in Human Activity Recognition," *Neural Processing Letters,* 2021, doi: 10.1007/s11063-021-10611-w.
2. M. Gil-Martín, R. San-Segundo, F. Fernández-Martínez, and R. de Córdoba, "Human activity recognition adapted to the type of movement," *Computers & Electrical Engineering,* vol. 88, p. 106822, 2020/12/01/ 2020, doi: https://doi.org/10.1016/j.compeleceng.2020.106822.
3. M. Gil-Martin, R. San-Segundo, F. Fernandez-Martinez, and J. Ferreiros-Lopez, "Improving physical activity recognition using a new deep learning architecture and post-processing techniques," *Engineering Applications of Artificial Intelligence,* vol. 92, Jun 2020, Art no. 103679, doi: 10.1016/j.engappai.2020.103679.
4. S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, "A Review on Human Activity Recognition Using Vision-Based Method," *Journal of Healthcare Engineering,* vol. 2017, 2017 2017, Art no. 3090343, doi: 10.1155/2017/3090343.
5. Y. Hsu, S. Yang, H. Chang, and H. Lai, "Human Daily and Sport Activity Recognition Using a Wearable Inertial Sensor Network," *IEEE Access,* vol. 6, pp. 31715-31728, 2018, doi: 10.1109/ACCESS.2018.2839766.
6. Z. Zhuang and Y. Xue, "Sport-Related Human Activity Detection and Recognition Using a Smartwatch," *Sensors,* vol. 19, no. 22, Nov 2019, Art no. 5001, doi: 10.3390/s19225001.
7. D. Hendry, K. Chai, A. Campbell, L. Hopper, P. O'Sullivan, and L. Straker, "Development of a Human Activity Recognition System for Ballet Tasks," *Sports Medicine - Open,* vol. 6, no. 1, p. 10, 2020/02/07 2020, doi: 10.1186/s40798-020-0237-5.

8. M. Gil-Martin, W. Johnston, R. San-Segundo, and B. Caulfield, "Scoring Performance on the Y-Balance Test Using a Deep Learning Approach," *Sensors,* vol. 21, no. 21, pp. 7110-7110, Nov 2021.
9. P. Trindade, J. Lobo, and J. P. Barreto, "Hand gesture recognition using color and depth images enhanced with hand angular pose data," in *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 13-15 Sept. 2012 2012, pp. 71-76, doi: 10.1109/MFI.2012.6343032.
10. M. Oudah, A. Al-Naji, and J. Chahl, "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques," *Journal of Imaging,* vol. 6, no. 8, p. 73, 2020.
11. G. Zhang, L. Wang, L. Wang, and Z. Chen, "Hand-raising gesture detection in classroom with spatial context augmentation and dilated convolution," *Computers & Graphics,* vol. 110, pp. 151-161, 2023/02/01/ 2023, doi: https://doi.org/10.1016/j.cag.2022.11.009.
12. P. Trigueiros, F. Ribeiro, and L. P. Reis, "Hand Gesture Recognition System Based in Computer Vision and Machine Learning," in *Developments in Medical Image Processing and Computational Vision*, J. M. R. S. Tavares and R. Natal Jorge Eds. Cham: Springer International Publishing, 2015, pp. 355-377.
13. Y. Wang, R. Yang, and Ieee, "Real-Time Hand Posture Recognition based on Hand Dominant Line using Kinect," in *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, San Jose, CA, 2013, Jul 15-19 2013, in IEEE International Conference on Multimedia and Expo Workshops, 2013. [Online]. Available: <Go to ISI>://WOS:000335245800022
14. D. Núñez Fernández and B. Kwolek, "Hand Posture Recognition Using Convolutional Neural Network," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Cham, M. Mendoza and S. Velastín, Eds., 2018// 2018: Springer International Publishing, pp. 441-449.
15. T. Mantecon, C. R. del-Blanco, F. Jaureguizar, and N. Garcia, "A real-time gesture recognition system using near-infrared imagery," *Plos One,* vol. 14, no. 10, Oct 3 2019, Art no. e0223320, doi: 10.1371/journal.pone.0223320.
16. P. Bao, A. I. Maqueda, C. R. del-Blanco, and N. Garcia, "Tiny Hand Gesture Recognition without Localization via a Deep Convolutional Network," *Ieee Transactions on Consumer Electronics,* vol. 63, no. 3, pp. 251-257, Aug 2017, doi: 10.1109/tce.2017.014971.
17. C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," *ArXiv,* vol. abs/1906.08172, 2019.