Hand Gesture Recognition with ConvNets for School-Aged Children to Learn Basic Arithmetic Operations

Alvaro Teran-Quezada[1]
[1]*Facultad de Ingeniería de Sistemas Computacionales*
Universidad Tecnológica de Panamá
Panamá, Panamá
alvaro.teran@utp.ac.pa

Victor Lopez-Cabrera[1]
[1]*Facultad de Ingeniería de Sistemas Computacionales*
Universidad Tecnológica de Panamá
Panamá, Panamá
victor.lopez@utp.ac.pa

Jose Carlos Rangel[1,2]
[1]*Facultad de Ingeniería de Sistemas Computacionales*
Universidad Tecnológica de Panamá
[2]*Sistema Nacional de Investigación*
Panamá, Panamá
jose.rangel@utp.ac.pa

Javier E. Sanchez-Galan[1,2,*]
[1]*Facultad de Ingeniería de Sistemas Computacionales*
Universidad Tecnológica de Panamá
[2]*Sistema Nacional de Investigación*
Panamá, Panamá
javier.sanchezgalan@utp.ac.pa

***Abstract*—Hand Gesture and Sign Language Recognition are recognized as non-trivial problems in the computer vision community. Deep Learning (DL) has provided both, a novel analytical framework and solutions. Moreover, Convolutional Neural Networks (CNN) have shown there are still goals to be met and solutions to be given regarding Hand Posture Recognition with the architectures and technologies we have today. This work proposes building a Hand Gesture Recognizer able to identify 13 classes: numbers from 0 to 9, and the signs corresponding to the arithmetic operations of addition, subtraction, and multiplication. The core model aims to translate digits and math symbols signed by students and teachers in a school environment to both make the teaching-learning process more engaging and to promote sign language learning in the student community. Based on Object Detection and leveraging a pretrained model using Transfer Learning, this DL model was retrained with a data set of 1,365 images (105 per class/sign) of Panamanian Sign Language hand-shapes. An accuracy of 88% was reached on the validation data, with proved usefulness on American Sign Language (ASL) for these hand-shapes, and being easily adaptable to other sign systems, including two-handed finger spelling ones like International Sign System (IS).**

***Keywords*—Hand Gesture Recognition, Sign Language Recognition, Convolutional Neural Networks, Object Detection, Transfer Learning, Educational AI.**

I. Introduction

There are more than 300 different sign languages in the world, used daily by tens of millions of people [1], [2]. In Panama, the last estimate of the number of people with hearing disabilities was above 15,000 [3]. It is expected that this number has increased in the last decade [2] and, considering it, can serve as a reference to have an idea of the number of sign language users in the country.

[*]**Corresponding Author**: javier.sanchezgalan@utp.ac.pa

Panamanian Sign Language (PSL) users have seen, like their counterparts in most of the world, that technology has gone from making itself a space in education, to considerably influence this purposeful activity, to finally become the core of the development of educational systems worldwide [4]. However, there has been no such use of technology for certain communities, including the deaf community of Panama, with which there are many debts [5]. This community has limited implementations within the classroom; beyond the use of, mainly, electronic devices, or the virtual classroom as the learning environment that served as a solution in response to the quarantine due to the COVID-19 pandemic [6].

In any case, the efforts in terms of information technology applied to learning sign language exist and should be highlighted [7]. Also, different types of developments around this subject matter leave space to continue working on various topics around it, contributing to the community [8]–[10]. These lines of research have acquired, so far, the form of a mobile application with documentation and vocabulary on PSL [11], a web platform with a series of translations of concepts in Spanish to their equivalent in PSL signs [12], [13] and, more recently, a model capable of translating vowel signs according to the PSL manual alphabet into Spanish [14], [15].

The objective of this work is to build a Hand Gesture Recognizer able to identify 13 classes of gestures, including: numbers (0 to 9), and arithmetic operations (addition, subtraction, and multiplication). Also, to evaluate if the model can be extended and used with other sign languages that share similar signs. The final intention for this model is that it can translate digits and math symbols signed by students and teachers in a school setting. Finally, enhancing the overall teaching-learning process.

II. Materials and Methods

A. Construction of the hand sign data set

Images were taken using and integrated RGB HD Camera of 0.92 megapixels in still image mode. These images have dimensions of 1280×720, and a 96-dpi value for both horizontal and vertical resolution. The distance between the camera and the subject was approximately of 0.6m (around 2ft). As state before, each image corresponds to an individual making the signs of the decimal system digits and the mathematical symbols, according to Panamanian Sign Language exclusively.

A total of 1,560 pictures were taken, 120 per class/sign. For this purpose, OpenCV [16] was used. The images were stored locally, dividing them into sets for training (87.5%; 1,365: 105 per class) and for validation (12.5%; 195: 15 per class)

B. Image Preprocessing

Manual labeling of the images was done with the *LabelImg* tool [17]. The labeling process includes determining the Region of Interest (ROI) within each training data set image. Labeling also includes generating a *.xml* file for each image. This linked .xml have the information of the object name (the sign label), the image size, and the bounding box coordinate point values. Further, preprocessing operations as image resizing in order to fit the model structure were made in TensorFlow directly [18].

C. Usage of Transfer Learning

As a computer vision problem, the task of building a sign language recognizer can be based on an object detector. That is, a detector that considers hand postures as objects or elements of interest to be detected and discard the rest of elements on the image as background. Finally, identify and learn the patterns involved in the hand gesture, to build a model that generalizes their recognition.

Figure 1 shows a methodological diagram, representing the overall architecture and subsystems needed to build an object detection model [19]. In particular, a model that focuses on sign classes, reassigning its task to the specific use case of developing a static sign translator to be further optimized.

Since the proposed solution is based on object detection, Transfer Learning (TL) technique [20], [21] was selected for this task, as it has been previously used for hand gesture recognition [22]. In general terms, the main idea of TL is that having a model that has been pretrained on a great number of images, a significantly well-trained model is achieved without a large sample for retraining purposes. See Figure 2 for a conceptual representation.

The proposed solution model was pretrained using the *MobileNet V2 / FPNLite 640×640* model. This model is a Single Shot Multibox Detector with a Feature Pyramid Network (FPN) feature extractor trained with the *COCO*

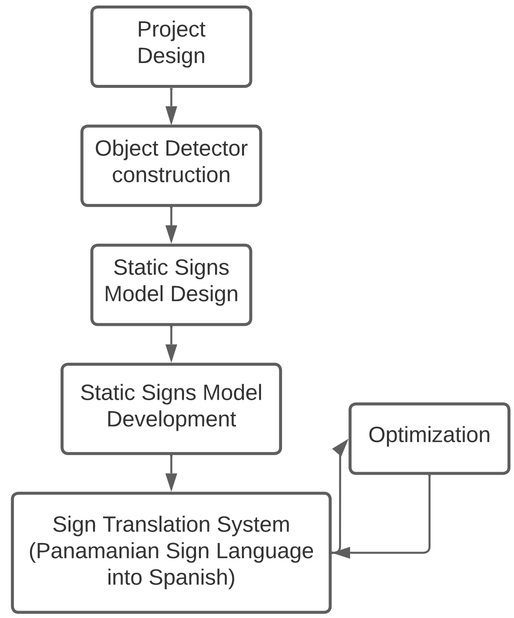

Fig. 1. Methodology Diagram based on Object Detection for Sign Language Recognition.

2017 data set [23]. The score prediction speed is 39 ms and it has a Mean Average Precision (mAP) of 28.2. This data set belongs to the TensorFlow 2 Detection Model Zoo, described in detail in [24].

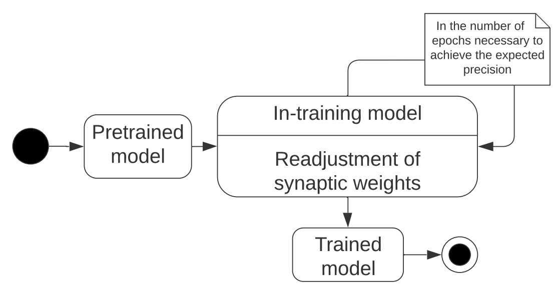

Fig. 2. Model states in the training process.

D. Model Training

The ready-to-use model would typically receive, as input, images where signs are made by the people in them, being these signs the objects to detect in real time. An overall view of the functioning of the system is depicted in Figure 3.

III. Results and Discussion

Plots showing aspects of the training process per epoch (for 2,500 epochs) can be seen in Figure 4. The total loss per epoch Figure 4-(A) shows a tendency for the loss to decrease, while the learning rate Figure 4-(B), shows close

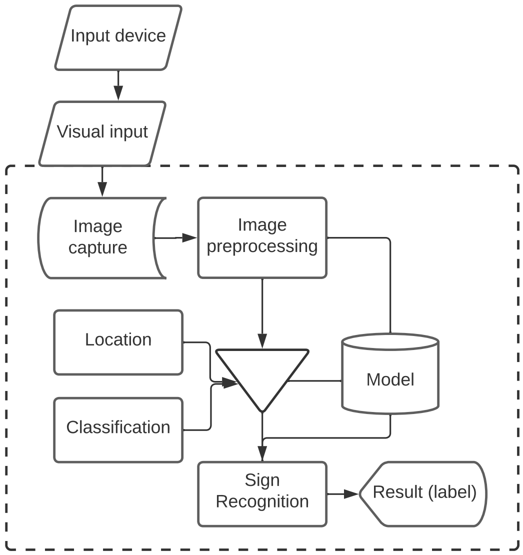

Fig. 3. System's Diagram.

to no change since the middle of the process. Finally, the steps per second, Figure 4-(C), have a similar behaviour, decreasing from a relatively early part of the training.

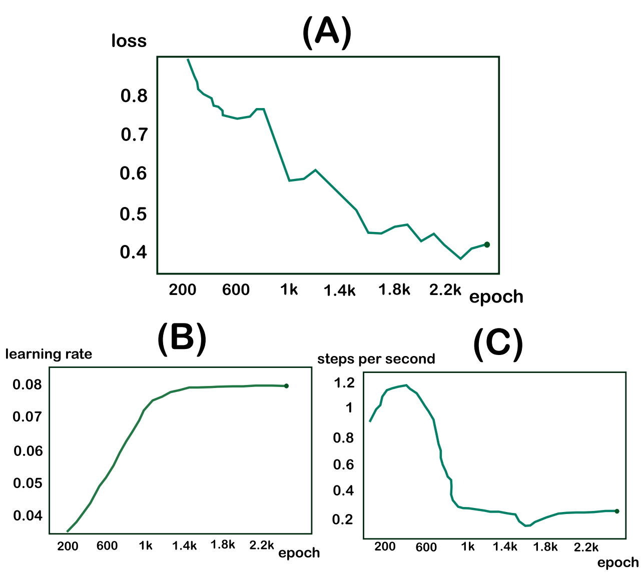

Fig. 4. Total loss per epoch (A), learning rate of the model per epoch (B), and steps per second per epoch on the training process (C).

A. Model Evaluation

The rest of the images (195 images in total, about 12.5% of the total data set) was considered for the validation of the retrained model. These images have the same characteristics as those used for training, however data leakage was avoided.

The validation process resulted in an accuracy of 88.1%, as well as a total loss of 0.49. Other metrics of interest are shown in the Table I.

TABLE I
PERFORMANCE VALIDATION OF THE MODEL

Metric	Value
Average Precision	0.881
Average Recall	0.749
Classification Loss	0.250
Localization Loss	0.089
Regularization Loss	0.149
Total Loss	0.489
Epochs	2,500

B. Analysis of Results by Sign Subgroup

As the overall resulting model provided comparably good results, one further was to understand and evaluate how good of a performance the model exhibited to recognize sign classes. For this task, it is fair to consider three subgroups among the sign classes, as:

- *math symbol signs - Subgroup 1*: this include addition, subtraction, and multiplication, plus the zero sign, the reason being that they're not easily confused with each other nor with the rest of the classes considered for this model.
- *numbers from zero to five - Subgroup 2*: this classification responds to the fact of these signs being effortlessly executable because of the natural hand postures that have to be done by the person who is signing them.
- *numbers from six to nine - Subgroup 3*: for which the signing of them is reasonably harder.

Each class subgroup is defined by their own characteristics. Examples of each subgroup can be seen in Figure 5.

It was observed that in execution the model performs at its best with the first subgroup. While, the performance is substantially reduced for recognizing signs of the third subgroup.

As part of the optimization process of the model, a threshold of 0.5 was set. This made the model performed better for both when it had one or more signs as an input. Figure 6 shows recognition on random static images as the visual input. One can see that it provides recognition with 85% or more for samples Figure 6-(A), Figure 6-(B), and Figure 6-(C). Moreover, a low recognition confidence level for Figure 6-(D), and a moderate to high level for Figure 6-(E) which is two-handed like (B) is.

C. Applicability to American Sign Language (ASL)

To validate further the usage of the system, the possibility of using the model with signs executed in a different way, that is, from another sign language or in an alternative manner, was explored.

The idea was to evaluate if the system could have an acceptable performance for signs from the American Sign

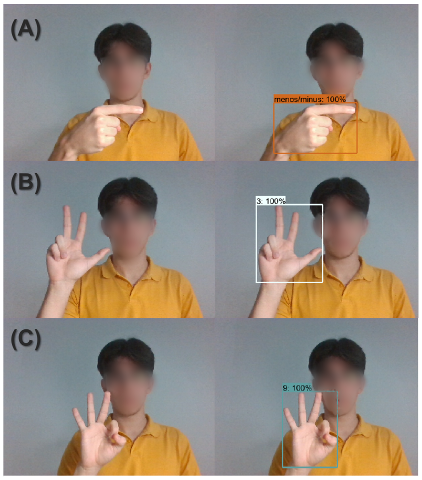

Fig. 5. Examples of signs for each subgroup. Math Symbol sings (A), numbers from one to five (B), and numbers from six to nine (C).

Fig. 6. Model performance for static images.

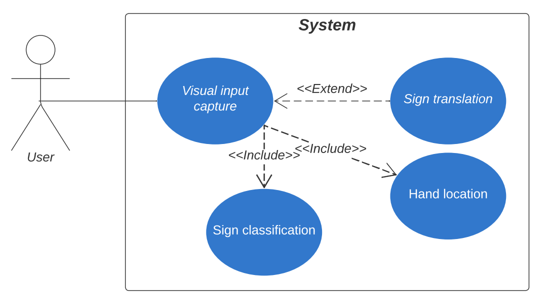

Fig. 7. Use Case Diagram of the Model.

Language (ASL) [25]. For this, an additional set of 65 images, was created. The same concepts of mathematical symbols and the decimal system numerals were considered with the difference being the sign language.

ASL signs for the 13 classes were used. The hypothesis was that, since there is some similarity in the way this ideas (each sign) are signed in both sign languages, there could be a good performance for the model that was exclusively trained with PSL signs. It should be noted that some of these signs have similar, equivalent or identical versions (either the most common or the alternative ones were used) for these sign languages.

An important factor is to notice, that the proposed system will only work if there are commonalities between the sign languages. For instance, in PSL the signs for numbers are usually made with the palm facing the person with whom communication is taking place [4]. In ASL something similar happens. Thus, the additional set of images described before considers those signs with the palm facing the person who is signing them. If the languages differ drastically the system, will not work efficiently.

The recognition results for the ASL signs were fairly good (60%-70%). In terms of model performance, as it can bee seen in Figure 8. In general, returning the right class label, but having a low prediction confidence.

Figure 8, also summarizes three examples of the model response to the tests. Figure 8-(A), shows a positive inference, which was expected for the similarity of the signing, (palm facing opposite directions). Figure 8-(B), shows a negative inference, which was also expected because there is the hand posture for the output given and the hand that is in front barely covers the one behind (hence, the identified sign). Finally, Figure 8-(C), that had a positive inference as well, which was not expected because of how different the signs are. Yet the model could make sense because of the way the data is processed.

Apart from the hits and misses describe above, due to the architecture of the neural network, there is a great opportunity of applying this model to other sign languages.

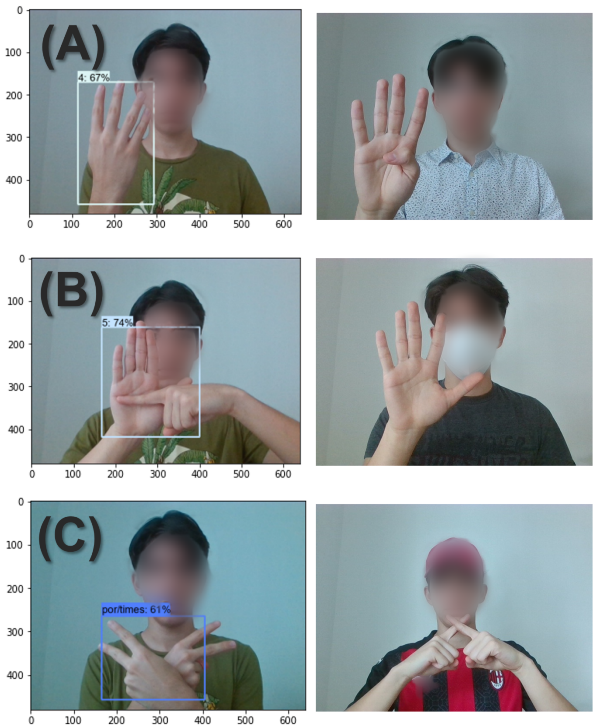

Fig. 8. Inferences for signs made differently (left), and how they were signed for the training data set (right).

IV. Conclusions

In light of the results above presented. There is, once more, demonstration that Object Detection-based Sign Language Recognition models are feasible to a number of applications. In particular, those relating numbers and symbols, as an interesting group of signs that for many sign languages are mostly static ones (i.e., just hand postures without any movement).

This study suggests that, for Panamanian Sign Language, the developed model can be leveraged in applications such as a math learning-evaluation system for kids, and that it could also be applicable to other sign languages, including American Sign Language, without necessarily creating a whole new set of data for its signs as new classes since they're in their majority quite similar.

With training on different data for any other sign language, the model could serve as part of a solution that involves static signs. It befits sign languages no matter their characteristics as if it is either one-hand signed or two-handed. Also, the referred solutions could have real time execution as a fundamental feature given that the inference is made shot by shot in just a few milliseconds.

As part of the limitations of the model, there are the ones associated with the dynamic signs (i.e., signs that need some movement or ideas that need more than one sign to be expressed) for which it cannot be trained properly. The signs for numbers are, in fact, statically made just by digit. So, in solutions involving numbers, these should be presented digit by digit for figures higher than 10 instead of the actual signs made in PSL.

For kids, while learning PSL, that is the way of signing numbers for up to three (sometimes four) digit amounts, making it possible and reasonable to use the model as a solution for the mentioned system; being already usable for this purpose in this stage, to some extent. This model does not currently translate other PSL static signs, namely, most of the letters in the PSL manual alphabet.

This project could represent visualization to the deaf community in Panama and could be utilized as an entertaining way of promoting PSL learning in the early educational stages. Another important element is being able to see the latent need of applying tech to communities as an achievable goal. In the same order of ideas, this one is yet another project that highlights the relevance of tech applied to education, also giving a chance to make part of the learning process a more engaging one.

Acknowledgment

J.C.R. and J.E.S.-G. are members of the Sistema Nacional de Investigación (SNI) of the Secretaría Nacional de Ciencia, Tecnología e Innovación de Panamá (SENACYT). The authors acknowledge administrative support provided by Universidad Tecnológica de Panamá.

References

[1] United Nations, U.N., "International day of sign language." https://www.un.org/en/observances/sign-languages-day, 2021.

[2] World Health Organization, W.H.O, "Deafness and hearing loss." https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss, 2021. Online available; accessed at 16/06/2022.

[3] A. Oviedo, "Panamá, atlas sordo – cultura sorda." https://cultura-sorda.org/panama-atlas-sordo/, 2015. Online available; accessed at 13/04/2022.

[4] D. Pimentel, R. Walker, and M. Fajardo, *Lengua de señas panameñas*. Panama, Panama: Editora Panamá América, 2018.

[5] Á. Arjona, "Desafíos de la educación a personas sordas en panamá." https://www.laestrella.com.pa/nacional/politica/171005/sordas-panama-desafios-personas-educacion, 2017. Online available; accessed at 24/02/2022.

[6] J. R. Loaiza, K. Rao, G. A. Eskildsen, E. Ortega-Barria, M. J. Miller, and R. A. Gittens, "Covid-19 pandemic in panama: lessons of the unique risks and research opportunities for latin america," *Revista Panamericana de Salud Pública*, vol. 44, p. e86, 2020.

[7] C. Hernández, J. L. Pulido, and J. E. Arias, "Las tecnologías de la información en el aprendizaje de la lengua de señas," *Revista de Salud Pública*, vol. 17, pp. 61–73, 2015.

[8] M. I. Pimentel Araúz, *Sitio web para el aprendizaje de Lengua de Señas Panameñas*. PhD thesis, Universidad de Panamá. Vicerrectoría de Investigación y Postgrado, 2018.

[9] L. Alain and R. Vejarano, "Alternativas tecnológicas para mejorar la comunicación de personas con discapacidad auditiva en la educación superior panameña," *Revista de Educación de la Universidad de Granada*, vol. 23, pp. 219–235, 2016.

[10] A. Rodríguez-Fuentes, L. Alaín-Botaccio, and F. García-García, "Presentation and evaluation of a digital tool for sign language (presentación y evaluación de una herramienta digital para la lengua de signos)," *Culture and Education*, pp. 1–31, 2022.

[11] Gestión Empresarial 3000, “Iphe inclusivo,” 2018. Mobile app.

[12] L. Alaín Botacio, *Desarrollo y validación de una aplicación web y cd educativa inclusiva para el aprendizaje de la lengua de señas panameña*. PhD thesis, Universidad de Granada, 2019.

[13] A. R. Fuentes, L. Alain, and F. G. García, “Ensenias: Technological tool to learn, teach, improve and use panamanian sign language,” *Íkala*, vol. 25, no. 3, pp. 663–678, 2020.

[14] R. Bodmer, L. Liu, W. Liu, and J. C. Rangel, “Sign language recognition with machine learning for elementary school children,” *Revista de Iniciación Científica, Edición Especial*, vol. 6, 2020.

[15] A. Flores, E. González, J. Z. Pan, V. Villarreal, and L. Muñoz, “Sistema de aprendizaje de lengua de señas panameña (lsp) a través de un brazo robótico articulado con reconocimiento de gestos,” in *Memorias de Congresos UTP*, pp. 168–173, 2019.

[16] DataFlair, “Sign language recognition using python and opencv.” https://data-flair.training/blogs/sign-language-recognition-python-ml-opencv/, 2022. Online available; accessed at 23/02/2022.

[17] T. Lin, “heartexlabs/labelimg.” https://github.com/heartexlabs/labelImg, 2022. Online available; accessed at 25/02/2022.

[18] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. California, USA: O’Reilly Media, 2019.

[19] G. Jocher, “Yolov5 documentation.” https://docs.ultralytics.com/, 2020. Online available; accessed at 07/01/2022.

[20] H. Davidson, “Introducing transfer learning as your next engine to drive future innovations.” https://medium.datadriveninvestor.com/introducing-transfer-learning-as-your-next-engine-to-drive-future-innovations-5e81a15bb567, 2020. Online available; accessed at 15/02/2022.

[21] D. Martinez, “Is transfer learning the final step for enabling ai in aviation? - datascience.aero.” https://datascience.aero/transfer-learning-aviation/, 2020. Online available; accessed at 19/05/2022.

[22] T. Ozcan and A. Basturk, “Transfer learning-based convolutional neural networks with heuristic optimization for hand gesture recognition,” *Neural Computing and Applications*, vol. 31, no. 12, pp. 8955–8970, 2019.

[23] A. Nikishaev, “Feature extraction and similar image search with opencv for newbies.” https://medium.com/machine-learning-world/feature-extraction-and-similar-image-search-with-opencv-for-newbies-3c59796bf774, 2018. Online available; accessed at 25/04/2022.

[24] V. Birodkar, “Tensorflow 2 detection model zoo.” https://tinyurl.com/t2detzoo, 2020. Online available; accessed at 25/05/2022.

[25] A. Duarte, S. Palaskar, L. Ventura, D. Ghadiyaram, K. DeHaan, F. Metze, J. Torres, and X. Giro-i Nieto, “How2sign: A large-scale multimodal dataset for continuous american sign language,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2735–2744, 2021.