Universidad de los Llanos

ORINOQUIA
Suplemento Volumen 25 - Número 1, Enero-Junio 2021

Artículo de Revisión/Review Article

Reconocimiento del abecedario de la lengua de señas colombiana con Redes Neuronales Convolucionales

Recognition of the Colombian sign language alphabet using Convolutional Neural Networks

Reconhecimento do alfabeto da língua de sinais colombiana usando redes neurais convolucionais

Recibido: 24 de noviembre de 2020. **Aceptado**: 11 de diciembre de 2020.

Néstor E. Suat-Rojas[1*],
Ing. Sist, MSc, https://orcid.org/0000-0003-2628-1173

Brayan S. Montoya-Serna[2*],
Estud. Ing. Infor. https://orcid.org/0000-0003-2405-6335

Edward M. Pinzón-Velásquez[3*],
Estud. Ing. Infor. https://orcid.org/0000-0002-0446-5310

Oscar S. Rodríguez-Galeano[4*],
Estud. Ing. Infor. https://orcid.org/0000-0001-6062-4446

[1] Docente Investigador.
Email: nestor.suat@aunarvillavicencio.edu.co
[2] Corporación Universitaria Autónoma de Nariño.
Email: lmparra89@gmail.com
[3] Desarrollador, Alcaldía de Villavicencio departamento de Sistemas. Email: Empinzon13@gmail.com
[4] Desarrollador Backend, HEWTEC SAS.
Email: oscarstep.123@gmail.com
[*] Grupo de Investigación GIAUNARVI, Corporación Universitaria Autónoma de Nariño, Programa de Ingeniería Informática, Villavicencio, Colombia.

Suplemento Orinoquia, Enero-Junio 2021; 25(1): 25-30
ISSN electrónico: 2011-2629
ISSN impreso: 0121-3709
https://doi.org/10.22579/20112629.680

Resumen

El lenguaje de señas brinda un sistema para que las personas con discapacidad oral/auditiva se comuniquen efectivamente. Sin embargo, aún falta para que el resto de la sociedad se apropie de este conocimiento. Este trabajo consiste en diseñar un método de visión artificial que identifique las señas estáticas del abecedario de la Lengua de Señas Colombiana (LSC). La metodología consiste en un algoritmo de clasificación que combina una arquitectura de Redes Neuronales Convolucionales (CNN) y técnicas de procesamiento de imágenes. Nuestro enfoque logra reconocer las señas del alfabeto sin movimiento con un 79.2% de *exactitud*. El sistema es capaz de reconocer las letras según la forma, orientación y posición de los dedos de la mano, usando un conjunto de datos desbalanceado por clase.

Palabras clave: aprendizaje automático, lengua de señas colombiana, procesamiento de imágenes, red neuronal convolucional.

Abstract

Sign language provides a system for people with speech or hearing impairments to communicate effectively. However, it is still necessary for the rest of society to appropriate this knowledge. This work consists in designing a computer vision method that recognizes the static signs of the Colombian Sign Language (LSC) alphabet. The methodology consists of a classification algorithm that combines a Convolutional Neural Network (CNN) architecture and image processing techniques. Our approach manages to recognize signs of the alphabet that don't involve any movement, with 79.2% accuracy. The system is capable of recognizing letters according to the shape, orientation and position of the fingers in each sign, using an imbalanced dataset.

Keywords: Colombian sign language, convolutional neural network, image processing, machine learning.

Resumo

A linguagem de sinais fornece um sistema para pessoas com deficiência auditiva ou de fala se comunicarem de forma eficaz. Porém, ainda é necessário que o restante da sociedade se apropriе desse conhecimento. Este trabalho

Como Citar (Norma Vancouver):
Suat-Rojas NA, Montoya-Serna BS, Pinzón-Velásquez EM, Rodríguez-Galeano OS. Reconocimiento del abecedario de la lengua de señas colombiana con Redes Neuronales Convolucionales. Orinoquia, 2021;(SUPLEMENTO 1):25-30. Https://doi.org/10.22579/20112629.680

consiste no desenho de um método de visão computacional que reconhece os signos estáticos do alfabeto da Língua de Sinais Colombiana (LSC). A metodologia consiste em um algoritmo de classificação que combina uma arquitetura de Rede Neural Convolucional (CNN) e técnicas de processamento de imagens. Nossa abordagem consegue reconhecer sinais do alfabeto que não envolvem nenhum movimento, com 79,2% de precisão. O sistema é capaz de reconhecer letras de acordo com a forma, orientação e posição dos dedos em cada sinal, usando um conjunto de dados desequilibrado.

Palavras chave: aprendizado de máquina, língua de sinais colombiana, processamento de imagem, rede neural convolucional.

Introducción

El estado colombiano adoptó la lengua de señas colombiana (LSC) como lengua nativa y patrimonio cultural, priorizando su protección y divulgación a través de la Ley 982 de 2005, de esta manera convirtiéndola en la segunda lengua nativa más usada en el país.

En Colombia se estima que la población con discapacidad auditiva alcanza los 554.119 personas en el 2019, según proyecciones del DANE y el Instituto Nacional para Sordos (INSOR), esta última siendo una institución que tiene como función orientar y promover el establecimiento de entornos sociales en derecho de igualdad para personas sordas. La lengua de señas es la principal forma de comunicación de esta población, sin embargo, la falta de educación y apropiación de este lenguaje en los diferentes espacios de la sociedad aumenta la brecha de inclusión e integración social. Además, al no ser una lengua universal, por lo que cada país tiene su propia lengua, presenta un mayor reto hacia los investigadores locales, razón por la cual se ha convertido en un tema de gran interés en los distintos campos de investigación.

Muchos investigadores se han interesado en este problema creando dispositivos tecnológicos externos para la interpretación de señas, como guantes (Mehdi y Khan, 2002), brazaletes (Abreu *et al*., 2016), e incluso empleando herramientas disponibles como Kinect (Huang *et al*., 2015) y HoloLens (Fang *et al*., 2017). Sin embargo, estos dispositivos son difíciles de acceder debido a los altos costos de adquisición y mantenimiento, lo que dificulta llegar a toda la población con limitaciones auditivas.

Otros autores han explorado soluciones disponibles de bajo costo, por medio de cámaras de video convencionales, accesible desde celulares y computadoras (Lahoti *et al*., 2018), y técnicas de visión artificial (Nel *et al*., 2013). Estas propuestas realizan implementaciones basadas en procesamiento de imágenes e inteligencia artificial, empleando técnicas clásicas como HOG[1] y SVM[2] (Albino y López, 2018), sin embargo, los resultados del estado del arte son técnicas que emplean redes neuronales profundas para la extracción de características, como es el uso de las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés Convolutional Neural Networks) (Abiyev *et al*., 2020). En Colombia, la traducción de señas también ha sido un tema importante de investigación, con sistemas que van desde el uso de un perceptrón multicapa (Guerrero-Balaguera y Pérez-Holguín, 2015) hasta el uso de técnicas clásicas de clasificación como el SVM y el KNN[3] (Botina-Monsalve *et al*., 2018).

El uso de las Redes Neuronales Convolucionales ha dado buenos resultados en tareas de extracción o mapeo de características para clasificación de imágenes, además se encuentran disponibles distintos modelos pre entrenados que se pueden utilizar. En el trabajo realizado por García y Viesca (2016), construyen un sistema reentrenando el modelo Google Net con el dataset de University of Surrey y el dataset de Massey University, el sistema usa la cámara del usuario para predecir la letra, sin movimiento, más probable que pertenezca a la lengua de señas americana (ASL), logrando una *exactitud* de 98% para un dataset que incluye las 5 letras entre a-e, y un 74% para las letras a-k. En el sistema de Quirk y Kamaal (2018) diseñaron un prototipo para traducir las señas del alfabeto australiano a inglés, haciendo fine-tuning de la arquitectura Yolo-V3 y utilizando un conjunto de datos propio, de esta manera alcanzaron una *exactitud* de 86% sobre el dataset empleado.

Nuestro enfoque se basa en un método de clasificación de imágenes de señas sin movimiento del abecedario LSC; se combina técnicas de procesamiento de imágenes y redes neuronales; se construye un modelo de clasificación con un conjunto de datos conforma-

[1] Histogram of oriented gradients (conocido en español como Histograma de Gradientes Orientados)

[2] Support Vector Machine (conocido en español como Máquinas de Vectores de Soporte)

[3] K-nearest neighbors algorithm (conocido en español como el método K-vecinos-más-cercanos)

do por 2364 imágenes. El resto del documento está estructurado de la siguiente manera. En la sección *Materiales y Métodos* se continúa con la explicación de la metodología del enfoque propuesto para la clasificación de las señas. En la sección *Resultados* se presenta los resultados obtenidos de la experimentación. En la sección *Discusión se* continúa con la discusión y análisis de los resultados. Se finaliza con la última sección *Conclusiones* donde se resalta las conclusiones principales del sistema propuesto, en relación con el reconocimiento y clasificación de las señas estáticas pertenecientes al abecedario de la Lengua de Señas Colombiana (LSC).

Materiales y métodos

En esta sección, presentamos el enfoque utilizado para la recolección de las imágenes, el procesamiento, y clasificación de señas fijas del abecedario de la lengua de señas colombiana. Primero, para la adquisición de las imágenes se usaron las manos de 12 voluntarios y videos disponibles en la red social YouTube. Segundo, se eliminó el fondo; se realizó un proceso de *data augmentation*, la cual es una técnica usada para para aumentar el número de datos modificando las imágenes originales; y se redimensionó la imagen a 32x32 pixeles. Finalmente, una arquitectura CNN es implementada para la clasificación de las letras del abecedario de señas que no involucran movimiento.

Adquisición de datos

En este trabajo se ignoró las señas que involucran movimiento, solo se tuvo en cuenta 21 letras seleccionadas que corresponden a señas estáticas [a, b, c, d, e, f, i, k, l, m, n, o, p, q, r, t, u, v, w, x, y]. La recolección de las imágenes se realizó en dos fases: A) Se tomaron fotografías de 12 personas voluntarias haciendo las señas. B) y se recolectaron imágenes extraídas de 7 videos de YouTube donde daban a conocer el abecedario de LSC.

Preprocesamiento de la imagen

Las imágenes son capturadas en ambientes no controlados y con cámaras disponibles en dispositivos móviles, conteniendo ruido y fondos que distorsionan el proceso de aprendizaje del modelo computacional. Los siguientes pasos de preprocesamiento son aplicados:

A. ***Eliminar fondo:*** El espacio de color RGB se transforma a YCrCB; se aplica un suavizado utilizando desenfoque Gaussiano; se remueve el fondo de la imagen detectando el contorno de la mano por medio del tono de piel.

B. ***Redimensión de tamaño:*** Se redimensiona la imagen a 32x32 pixeles.

C. ***Data Augmentation:*** Para aumentar el tamaño del conjunto de datos se genera nuevas imágenes a partir de las existentes, variando parámetros como la ubicación, giro horizontal, rotación entre (+/-) 15°, teniendo especial cuidado con las señas que al cambiar de rotación pueden significar otra letra.

En la Figura 1, se muestra un resumen del proceso de preprocesamiento y data augmentation.

Clasificación

La última etapa de la metodología es realizar la clasificación de las imágenes pre procesadas prediciendo la letra del abecedario que corresponde. Nuestro clasificador se basa en una implementación de arquitectura profunda CNN como se muestra en la figura 2, los parámetros de entrenamiento son: batch size de 32,

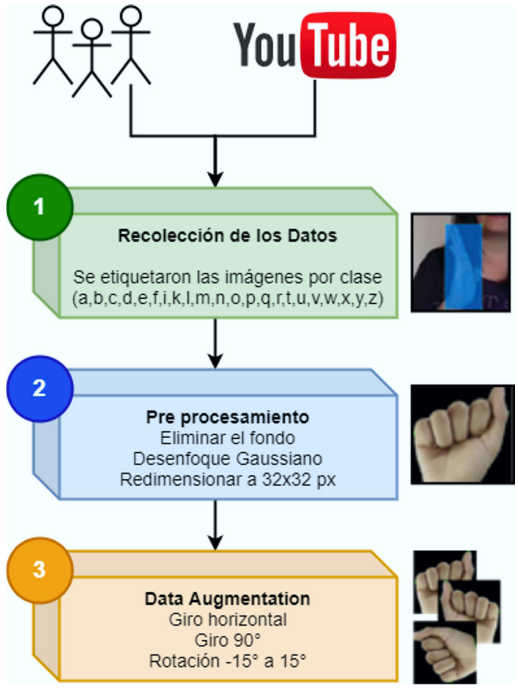


Figura 1. Ejemplo de preprocesamiento de una imagen.

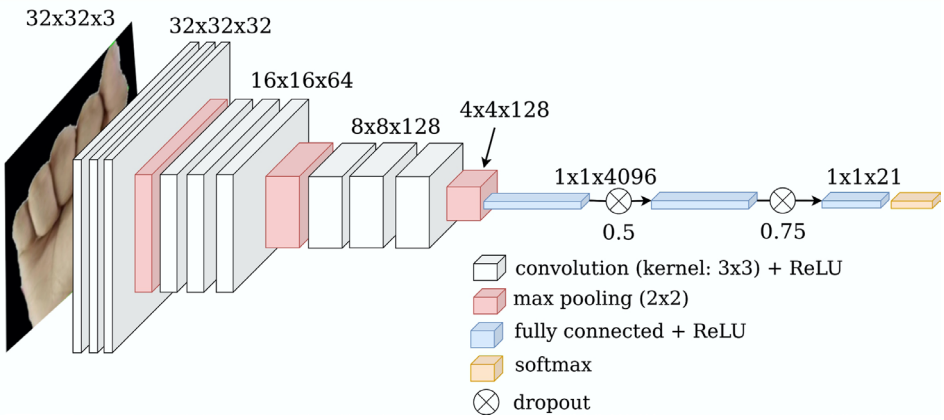


Figura 2. Arquitectura CNN implementada.

distribución uniforme Xavier (Glorot y Bengio, 2010) para la inicialización de los pesos de la red y el Descenso de Gradiente Estocástico de Nesterov con una taza de aprendizaje 0.001 y caída de la taza de 1e-6 y *momentum* 0.9.

Resultados

Recolectamos 2364 imágenes que corresponden a las 21 señas del alfabeto que no involucran movimiento (a, b, c, d, e, f, i, k, l, m, n, o, p, q, r, t, u, v, w, x, y). La figura 3 contiene algunos ejemplos de las imágenes. El conjunto de datos se dividió en 1875 imágenes que se utilizan para entrenar el clasificador y 489 imágenes para evaluar el desempeño del clasificador. En la figura 4 se muestra la distribución de imágenes por seña, no todas las letras tienen la misma cantidad.

Para evaluar el desempeño del clasificador propuesto, y compararlos con otros trabajos, empleamos la métrica de *Exactitud* (o Accuracy en inglés) que indica las veces que las señas son clasificadas correctamente, y se calcula dividiendo el número de predicciones correctas sobre el número total de predicciones. El algoritmo de CNN es estocástico por lo que se realizaron

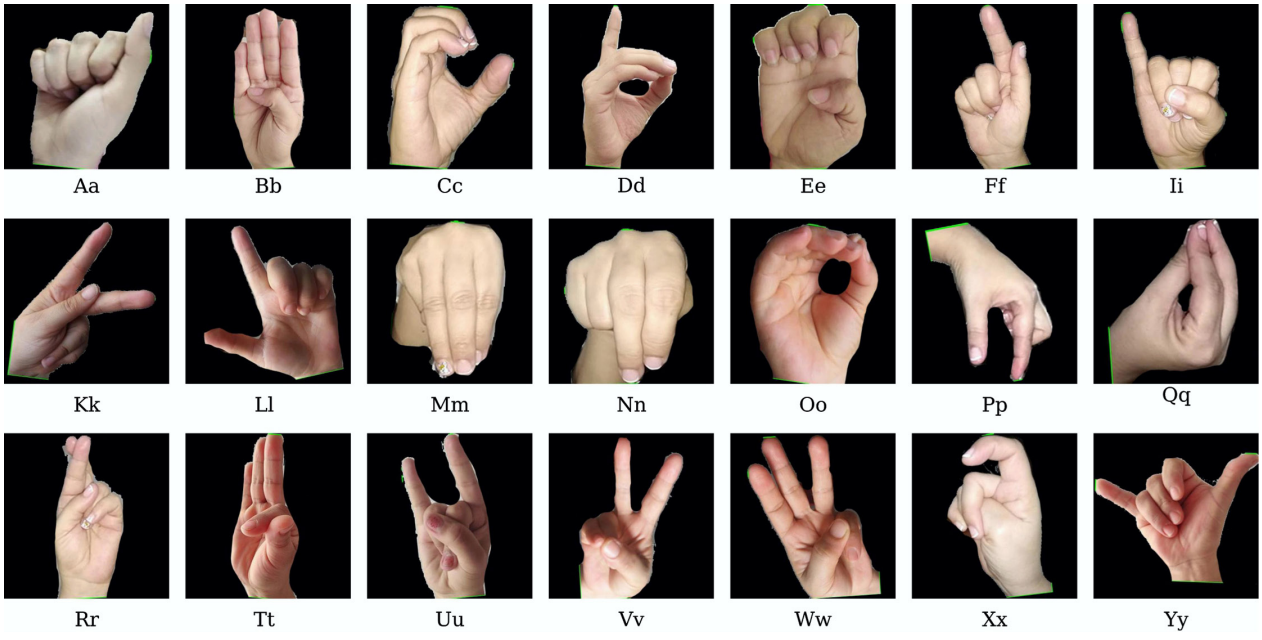
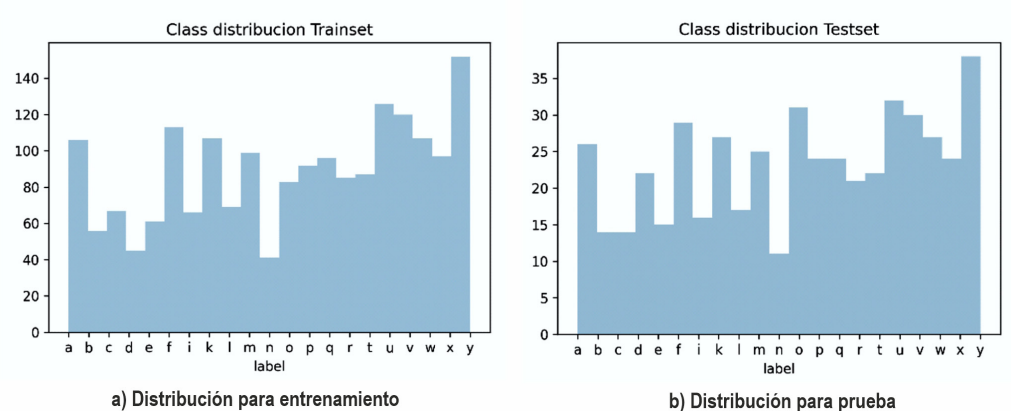


Figura 3. Ejemplos de las imágenes recolectadas por letra.



a) Distribución para entrenamiento

b) Distribución para prueba

Figura 4. Distribución de la cantidad de imágenes por cada letra, a) conjunto de entrenamiento y b) conjunto para evaluación.

100 experimentos y se calculó el promedio, en cada experimento se usó la estrategia de parada temprana o *early stopping* (alrededor de 35 epochs). También utilizamos la matriz de confusión para analizar el desempeño del clasificador en cada una de las letras, donde cada columna representa el número de predicciones de cada clase y cada fila representa a las instancias en la clase real (ver Figura 5). En general la *exactitud* obtenida con el conjunto de prueba del enfoque propuesto es de 79.2% para las 21 clases.

Discusión

El desempeño por clase se observa en la matriz de confusión en la figura 5. Hay letras que involucran una figura exclusiva de la mano que no se repiten con las demás, esto hace que el clasificador las reconozca fácilmente, como son los casos de las letras A, B, C, O, R, U, W y Y que estuvieron cerca al 100% de reconocimiento. Incluso con un conjunto desbalanceado, así estas letras posean una menor cantidad de imágenes, su forma y figura única de la mano es una caracte-

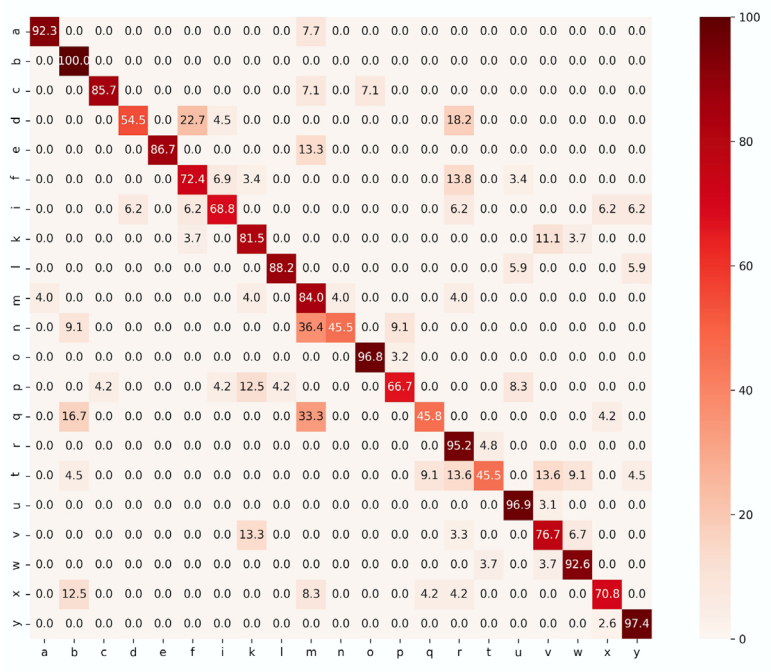


Figura 5. Matriz de confusión del desempeño de las 21 letras.

rística local fuerte para que la red CNN sea más fácil encontrar diferencias, como es el caso de las letras B y E en la figura 4.

Por otro lado, el desempeño del clasificador desciende para las señas que tienen características comunes de la figura de la mano. Por ejemplo, las letras D, F, I obtienen un menor reconocimiento de las imágenes, estas 3 letras involucran levantar uno de los dedos de la mano mientras los demás se mantienen abajo, este efecto se puede ver en la figura 3. De esta misma forma hay grupos de letras que son muy similares, este es el caso de las letras M/N, R/T y K/V/W, involucran una pequeña variación como cambiar la posición de un dedo de la mano. En la matriz de confusión de la figura 5 se ve cómo esta similitud afecta el reconocimiento con las leras N, T y K.

Conclusiones

Este artículo demuestra un potencial de reconocer las señas de la lengua de señas colombiana por medio de detección automática de imágenes. Implementamos una metodología que recolecta, pre procesa y reconoce las imágenes de las señas utilizando técnicas de visión artificial. El sistema propuesto con arquitectura CNN logra una *exactitud* del 79.2% usando el dataset de prueba de 21 letras.

El sistema es capaz de reconocer señas fijas según las características de forma y figura de la mano, usando un conjunto de datos desbalanceado por clase. Removemos el fondo de las imágenes para que el modelo se concentre en las características propias de la mano, como son la posición de los dedos.

En trabajos futuros se pretende implementar este enfoque en un sistema de tiempo real para traducir las señas a texto, incluyendo las señas que implican movimiento y aplicando técnicas de modelamiento de lenguaje para sugerir la formación rápida de palabras.

Agradecimientos

Se agradece a la Corporación Universitaria Autónoma de Nariño – Extensión Villavicencio por incentivarnos y apoyarnos en el desarrollo de este proyecto, y por prestarnos espacios para realizar la experimentación de nuestro sistema propuesto.

Referencias

Abiyev RH, Arsla M, Idoko JB. Sign Language Translation Using Deep Convolutional Neural Networks. *KSII Transactions on Internet & Information Systems*, 2020;14(2): 631-653. DOI: 10.3837/tiis.2020.02.009.

Abreu JG, Teixeira JM, Figueiredo LS, Teichrieb V. 2016. Evaluating Sign Language Recognition Using the Myo Armband. *2016 XVIII Symposium on Virtual and Augmented Reality (SVR).*

Albino-Huertas EA, López-Olivos LT. 2018. Visión computacional para la traducción en tiempo real del lenguaje de señas a texto en idioma español.

Botina-Monsalve DJ, Domínguez-Vásquez MA, Madrigal-González CA, Castro-Ospina AE. 2018. Clasificación automática de las vocales en el lenguaje de señas colombiano.

Fang B, Co J, Zhang M. 2017. Deepasl: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation. *In Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems.* pp.1-13.

Garcia B, Viesca SA. Real-time American sign language recognition with convolutional neural networks. *Convolutional Neural Networks for Visual Recognition.* 2016;2:225-232.

Glorot X, Bengio Y. 2010. Understanding the difficulty of training deep feedforward neural networks. *In Proceedings of the thirteenth international conference on artificial intelligence and statistics.* pp.249-256.

Guerrero-Balaguera JD, Pérez-Holguín WJ. Sistema traductor de la lengua de señas colombiana a texto basado en FPGA. *Dyna*. 2015;82(189):172-181.

Huang J, Zhou W, Li H, Li W. 2015. Sign language recognition using 3d convolutional neural networks. *In 2015 IEEE international conference on multimedia and expo* (ICME). pp.1-6. IEEE.

Lahoti S, Kayal S, Kumbhare S, Suradkar I, Pawar V. 2018. Android based american sign language recognition system with skin segmentation and svm. *In 2018 9th International Conference on Computing, Communication and Networking Technologies* (ICCCNT). pp.1-6. IEEE.

Ley 982 de 2005. Por la cual se establecen normas tendientes a la equiparación de oportunidades para las personas sordas y sordociegas y se dictan otras disposiciones. 09 de agosto de 2005. D.O. No. 45995.

Mehdi SA, Khan YN. 2002. Sign language recognition using sensor gloves. *Proceedings of the 9th International Conference on Neural Information Processing, 2002*. ICONIP '02.

Nel W, Ghaziasgar M, Connan J. 2013. An integrated sign language recognition system. *In Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference.* pp.179-185.

Quirk TTL, Kamaal K. 2018. How we used AI to translate sign language in real time. [en línea]. [Citado el 02 julio, 2020]. Disponible en internet: <https://medium.com/@coviu/how-we-used-ai-to-translate-sign-language-in-real-time-782238ed6bf>.