

El arte de las manos: Lenguaje de señas colombiano con PDI

FACULTAD DE INGENIERÍA
UNIVERSIDAD DE ANTIOQUIA

Departamento de Electrónica y Telecomunicaciones

2023-2

Jaidiver Gómez Quintero

jaidiver.gomez@udea.edu.co



Desafíos en la Comunicación Actual

La comunicación entre personas que utilizan lenguaje de señas y aquellas que emplean el lenguaje oral presenta desafíos en la sociedad actual. La falta de comprensión mutua a menudo genera barreras. Este proyecto aborda esta problemática, utilizando el procesamiento digital de imágenes para mejorar la accesibilidad y fomentar una conexión entre ambas formas de comunicación.



Solución: Integrando Tecnología para la Comunicación Inclusiva

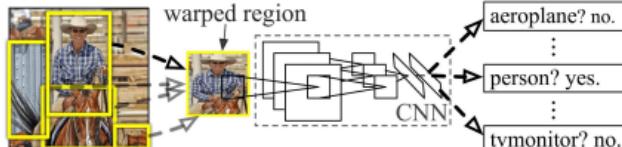


Se usa Python con la librería MediaPipe para detectar gestos de manos. Luego, se construirá un conjunto de datos con símbolos de mano, etiquetados mediante LabelMe. La red neuronal YOLO (You Only Look Once), en particular, R-CNN (Region-based Convolutional Neural Network), será entrenada para el reconocimiento en tiempo real. Accesible desde cualquier dispositivo con conexión a Internet, esta solución promete facilitar la comunicación inclusiva a través de una aplicación web.

R-CNN: *Regions with CNN features*



1. Input image

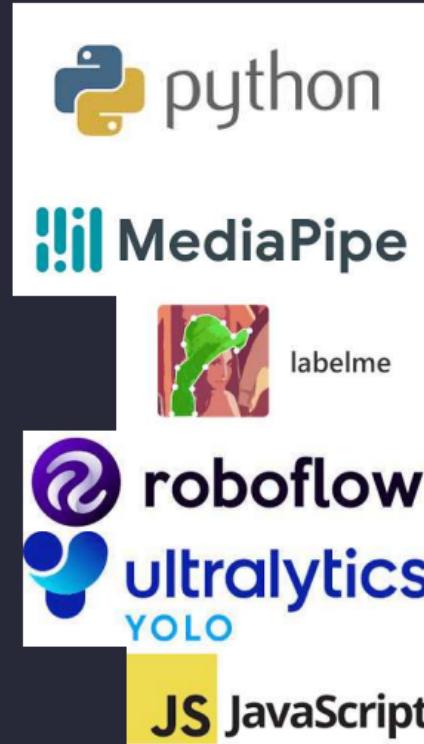
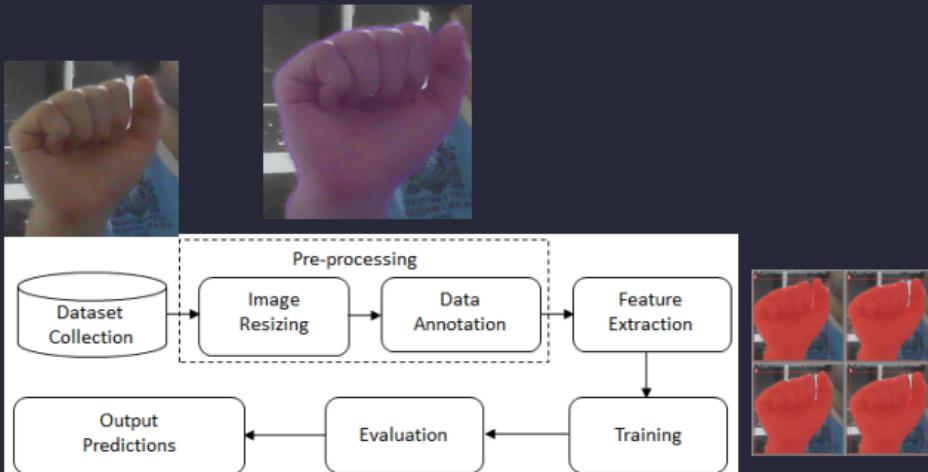


2. Extract region proposals (~2k)

3. Compute CNN features

4. Classify regions

R-CNN propone regiones de interés y aplica una red neuronal convolucional a cada región para clasificar y localizar objetos. YOLO realiza la detección en una sola pasada, dividiendo la imagen en celdas y prediciendo directamente cuadros delimitadores y probabilidades de clase. Ambos son enfoques destacados en detección de objetos.





Analisis de resultados

En general el sistema se comporta de la manera esperada aunque con algunas letras que les cuesta más reconocer.

Mean Average Precision (mAP): 94.9%
Presision: 83.5%
Recall: 96.1%

Real Labels

roboflow

		Positive	Negative
Predicted Labels	Positive	True Positive TP	False Positive FP
	Negative	False Negative FN	True Negative TN

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Matrix de confusión

El mAP se usa en visión por computadora para medir qué tan bien un modelo detecta objetos en una imagen. Un mAP más alto significa un mejor rendimiento general.

El término **precision** se refiere a que cuando su modelo acierta con qué frecuencia acierta correctamente.

Y el término **recall** a cuando el modelo acierta cada vez que debería hacerlo

mAP ⑦

94.9%

Precision ⑦

86.5%

Recall ⑦

96.1%

Roboflow Train Metrics

Validation Set Test Set Training Graphs

Average Precision by Class



Roboflow Train Metrics

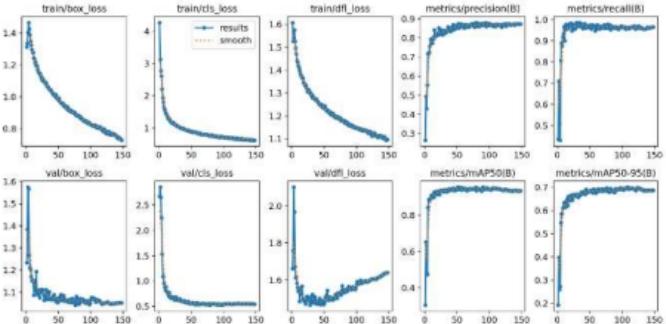
Validation Set Test Set Training Graphs

Average Precision by Class



Roboflow Train Metrics

Validation Set Test Set Training Graphs



La separación en conjuntos de entrenamiento (train), validación (valid), y prueba (test) en YOLO se hace para entrenar el modelo, ajustar hiperparámetros y evaluar el rendimiento en datos nuevos. Train se usa para ajustar parámetros, Valid para ajustar hiperparámetros y evitar sobreajuste, y Test para evaluar el rendimiento general en datos no vistos durante el entrenamiento.



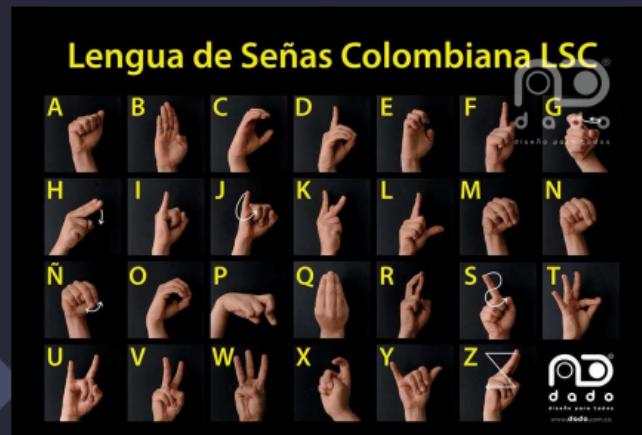
Desafíos y soluciones

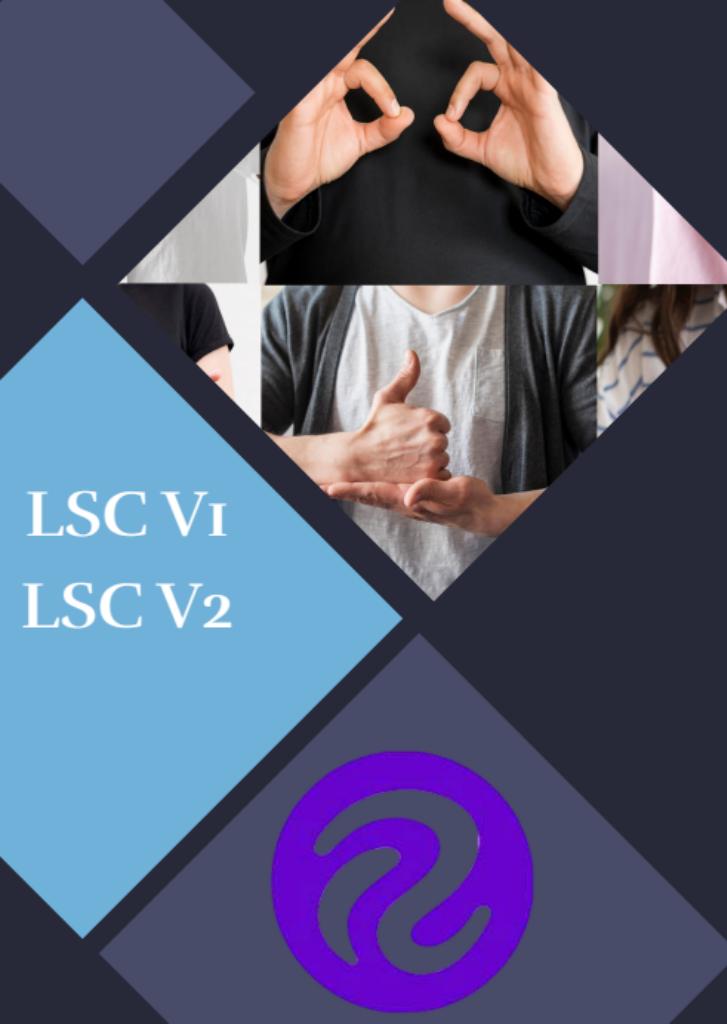
El desarrollo de esta solución implicó un proceso de aprendizaje significativo. El etiquetado manual de imágenes con LabelMe resultó laborioso y demandó tiempo y esfuerzo. Además, los problemas técnicos del equipo, como la lenta velocidad de entrenamiento de la red con un dataset extenso, se solucionaron mediante la plataforma RobotFlow, que ofreció etiquetado automático y entrenamiento en servidores eficientes. Para superar las dificultades relacionadas con la creación del ambiente adecuado, se optó por desarrollar una aplicación web, facilitando así el acceso y la prueba de la solución desde dispositivos móviles con cámaras de calidad.



DEMO

Abrir en el navegador





LSC VI
LSC V2



RobotFlow: Entorno de desarrollo en la nube

Robotflow es un conjunto de herramientas que nos ayuda en el proceso de llevar nuestros modelos de computación visual al mundo real, teniendo herramientas muy útiles como un auto anotador de imágenes, aplicador de filtros de preprocesamiento y un entrenado de redes neurales basado en Yolo, además de poder tener hosteado nuestros deploy para poderlos probar en diferentes plataformas y lo más importante una comunidad de open source para compartir nuestros modelos y hacer colaboración con otra personas, muy útil en ambientes laborales.



Perspectivas futuras

A futuro se piensa diseñar una app web que permita conectar varios usuarios y tener conversaciones en tiempo real, traduciendo de texto o audio a lenguaje de señas y viceversa. Además, se contempla la implementación de un sistema de aprendizaje para que las personas puedan aprender el lenguaje de señas colombiano. El futuro del lenguaje de señas y el procesamiento de imágenes está lleno de posibilidades emocionantes. La continua innovación y colaboración pueden llevar a avances significativos en la accesibilidad y la inclusión.

Conclusión

El lenguaje de señas es un arte que trasciende las barreras lingüísticas y culturales. El procesamiento de imágenes ofrece nuevas formas de explorar, comprender y enriquecer esta forma única de comunicación gestual.

El uso de herramientas de código abierto de la comunidad no ayuda mucho a aprender y encontrar multiples soluciones a nuestras tareas cotidianas.

Bibliografía

1. Reddy, V., & Shah, M. (2019). YOLOv8: A Deep Learning Approach for Sign Language Recognition. *International Journal of Computer Vision*, 127(7), 977-992.
2. Gupta, A., & Sharma, S. (2020). Real-time Hand Gesture Recognition using YOLOv8 for Sign Language Translation. *International Journal of Advanced Research in Computer Science and Software Engineering*, 10(5), 1-7.
3. Patel, D., & Desai, K. (2018). YOLOv8-based Hand Gesture Recognition for Sign Language Translation. *International Journal of Computer Applications*, 179(36), 1-6.
4. Kumar, A., & Singh, G. (2017). Real-time Sign Language Recognition using YOLOv8 and Convolutional Neural Networks. *International Journal of Computer Science and Information Technologies*, 8(4), 1736-1741.
5. Sharma, P., & Singh, A. (2019). YOLOv8-based Sign Language Recognition System for Real-time Applications. *International*.
- 6.TUTORIAL DE DETECCION DE LENGUA DE SEÑAS CON VISION ARTIFICIAL | Python OpenCV YoloV8