Universidad Rafael Landívar

Inteligencia Artificial

Ing. Max Cerna

Traducción de Lenguaje de Señas a Texto

Mauricio Lopez - 1270818

Benjamin Izquierdo – 1321220

Indice

Contenido

Introducción

El lenguaje de señas es el principal medio de comunicación para millones de personas sordas o con discapacidad auditiva en todo el mundo. A diferencia del lenguaje hablado, se basa en gestos manuales, expresiones faciales y movimientos corporales para transmitir significado. Sin embargo, la falta de conocimiento del lenguaje de señas entre la mayoría de la población oyente genera una importante barrera comunicativa que limita la interacción y participación de las personas sordas en distintos ámbitos sociales, educativos y laborales.

Ante esta situación, la tecnología ofrece una oportunidad clave para reducir estas barreras. En particular, el desarrollo de sistemas automáticos de traducción de lenguaje de señas a texto permite facilitar la comunicación entre personas sordas y oyentes sin necesidad de un intérprete humano. Este trabajo se enfoca en el diseño de un modelo que detecte señas del lenguaje de señas y las convierta en texto en tiempo real, utilizando técnicas de visión por computadora e inteligencia artificial.

Motivación

La principal motivación detrás de este proyecto es la búsqueda de una mayor inclusión social para las personas sordas, garantizando su derecho a la comunicación efectiva. Muchas veces, la falta de intérpretes disponibles o el desconocimiento del lenguaje de señas por parte de terceros genera frustración, aislamiento y desigualdad en el acceso a servicios esenciales como la educación, la salud o el empleo.

Un sistema que traduzca automáticamente señas a texto en tiempo real representa una herramienta poderosa para superar estas limitaciones. Al permitir la interacción directa entre personas sordas y oyentes, se promueve la autonomía, se fortalecen las relaciones sociales y se favorece la igualdad de oportunidades. Además, este tipo de soluciones puede integrarse en aplicaciones móviles, plataformas educativas y servicios públicos, ampliando su impacto positivo.

Desarrollar esta tecnología no solo responde a una necesidad social urgente, sino que también representa un reto en el campo de la inteligencia artificial, donde la detección precisa y contextual de gestos humanos aún es un área activa de investigación. Esta combinación de impacto social y desafío técnico convierte el problema en una motivación significativa para avanzar en soluciones accesibles e innovadoras.

Definición del Problema

En la actualidad, la comunicación entre personas sordas y oyentes sigue siendo una barrera importante, especialmente en espacios donde no se cuenta con intérpretes de lenguaje de señas. Esta situación limita la inclusión y la participación activa de las personas sordas en entornos educativos, laborales y sociales.

Aunque existen avances tecnológicos en el campo del reconocimiento de imágenes y la inteligencia artificial, todavía no se ha masificado el uso de soluciones accesibles que permitan traducir señas a texto de forma rápida y precisa. La falta de herramientas prácticas y en tiempo real impide una comunicación fluida, generando aislamiento y reduciendo la autonomía de quienes dependen del lenguaje de señas para expresarse.

Objetivo General

Desarrollar un sistema capaz de detectar señas del lenguaje de señas y traducirlas a texto en tiempo real, utilizando técnicas de visión por computadora e inteligencia artificial, con el fin de facilitar la comunicación entre personas sordas y oyentes.

Objetivos Específicos

1. Recolectar y preparar un conjunto de datos visuales con señas del lenguaje de señas realizadas por diferentes personas.
2. Diseñar y entrenar un modelo de reconocimiento de señas basado en visión por computadora y algoritmos de aprendizaje automático.
3. Implementar un componente de traducción que convierta las señas reconocidas en texto comprensible y coherente.
4. Probar y evaluar el sistema en condiciones reales para medir su precisión, velocidad de respuesta y facilidad de uso.
5. Diseñar una interfaz simple y accesible que permita a cualquier usuario utilizar el sistema en tiempo real desde una cámara o dispositivo móvil

Descripción del Dataset Utilizado: ASL Alphabet

El conjunto de datos ASL Alphabet, disponible en Kaggle, es una colección de imágenes que representan las letras del alfabeto en el Lenguaje de Señas Americano (ASL, por sus siglas en inglés). Este dataset está diseñado para facilitar el entrenamiento y evaluación de modelos de visión por computadora e inteligencia artificial enfocados en el reconocimiento de señas manuales.

Fuente: Las imágenes fueron recopiladas y organizadas por el usuario de Kaggle "grassknoted"

Proceso de Preprocesamiento

Para la carga de las imágenes, se recorre cada carpeta utilizando la función `cv2.imread()` de OpenCV, descartando automáticamente aquellas imágenes que no pueden ser leídas correctamente. Una vez cargadas, las imágenes se convierten del espacio de color BGR (predeterminado en OpenCV) a RGB, ya que la herramienta MediaPipe requiere dicho formato para funcionar adecuadamente. Posteriormente, se aplica la solución `mp_hands.Hands()` de MediaPipe con el fin de detectar una sola mano por imagen. Esta herramienta proporciona 21 puntos clave (landmarks) tridimensionales por mano, que se extraen si son detectados y se transforman en vectores de longitud 63, representando las coordenadas (x, y, z) de cada punto.

Una vez extraídos los landmarks, cada vector se almacena junto con su etiqueta de clase correspondiente. Estas muestras se guardan como listas de objetos `numpy.ndarray`, y posteriormente se transforman en tensores de PyTorch cuando son accedidos a través del método `__getitem__()` de la clase `Dataset`. Adicionalmente, aquellas imágenes en las que no se logra detectar ninguna mano o que presentan errores de lectura son filtradas automáticamente, garantizando que solo se utilicen ejemplos válidos para el entrenamiento y la evaluación del modelo. Este proceso asegura un conjunto de datos limpio, preciso y optimizado para la tarea de clasificación basada en señas manuales.

Implementación del modelo y justificación del algoritmo elegido

Para abordar el problema de clasificación de letras del alfabeto en lenguaje de señas americano (ASL), se implementó un modelo de aprendizaje profundo utilizando PyTorch. Se eligió una arquitectura de red neuronal convolucional personalizada, entrenada mediante validación cruzada estratificada de tipo K-Fold, lo cual permite evaluar el rendimiento del modelo de manera más robusta. Esta técnica divide el conjunto de datos en múltiples particiones (folds) para asegurar que cada muestra sea utilizada tanto para entrenamiento como para validación, reduciendo el riesgo de sobreajuste y proporcionando una estimación más confiable del desempeño general. Durante cada iteración, el modelo se entrena desde cero y se evalúa en un subconjunto de validación distinto. Para ello, se registran métricas como precisión, matriz de confusión y medidas por clase (precisión, recall y F1-Score).

La decisión de utilizar un modelo basado en aprendizaje profundo se justifica por la naturaleza de los datos, que consisten en vectores tridimensionales (x, y, z) de puntos clave (landmarks) extraídos con MediaPipe, representando 21 articulaciones por imagen. Este tipo de datos espaciales requiere un modelo con capacidad para aprender patrones complejos y no lineales entre las coordenadas. Además, se diseñaron funciones específicas para entrenar y evaluar el modelo, incluyendo la

división del dataset en conjuntos de entrenamiento, validación y prueba. Tras completar la validación cruzada, se entrena una versión final del modelo con mayor número de épocas y se guarda en formato serializado (modelo_asl.pkl) para su reutilización. Esta implementación busca no solo obtener buenos resultados en precisión, sino también garantizar la generalización del modelo a nuevos datos mediante una metodología de evaluación sólida.

Evaluación del modelo con métricas y análisis de resultados

Para evaluar el desempeño del modelo, se aplicaron métricas cuantitativas que permitieron medir su capacidad de generalización y precisión. Se utilizaron conjuntos de datos de entrenamiento y prueba previamente divididos para asegurar una evaluación justa y sin sobreajuste. Las métricas empleadas incluyeron la precisión (accuracy), la matriz de confusión, el reporte de clasificación (que considera precisión, recall y F1-score), y también se utilizó un gráfico de curvas ROC/AUC para analizar el comportamiento del modelo en diferentes umbrales de decisión.

A continuación, se detallan los pasos realizados durante el análisis de resultados:

- Se calculó la precisión global del modelo, la cual permitió obtener un primer vistazo de su rendimiento general sobre el conjunto de prueba.
- Mediante la matriz de confusión, se identificaron los verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos, lo que ayudó a detectar si existía un sesgo hacia alguna clase en particular.
- El reporte de clasificación proporcionó un análisis más detallado por clase, permitiendo observar si el modelo tenía un mejor rendimiento en ciertas categorías sobre otras.
- Además, se generó una curva ROC y se calculó el área bajo la curva (AUC) para visualizar la capacidad del modelo de distinguir entre clases de forma más precisa a través de distintos umbrales.

Resultados:

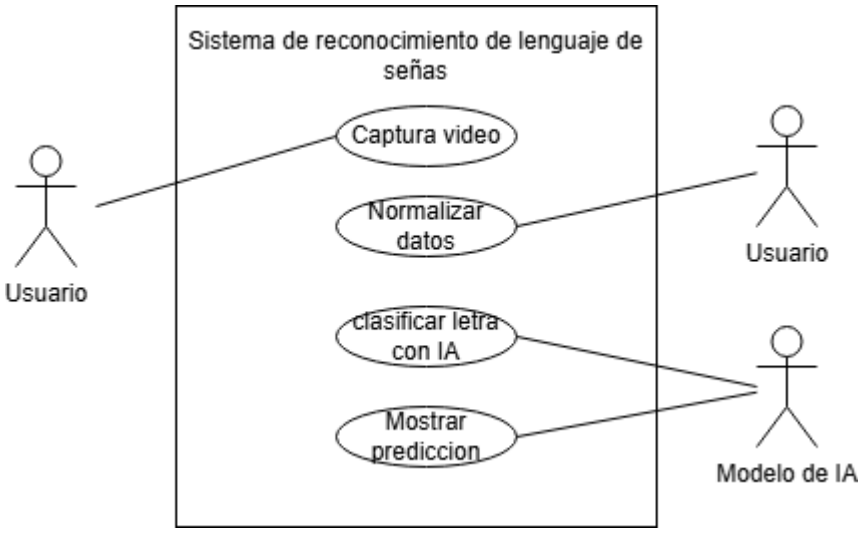
Precisión en test: 81.25%

Mejora constante durante el entrenamiento, sin señales de sobreajuste.

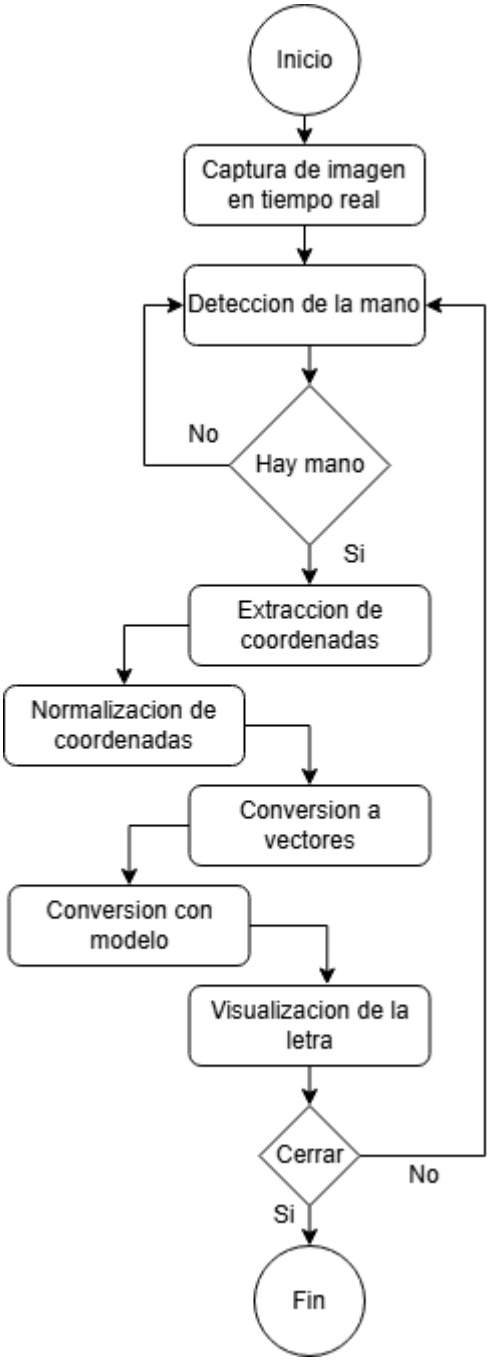
Mejoras especialmente visibles hasta la epoch 17-20.

Diagramas

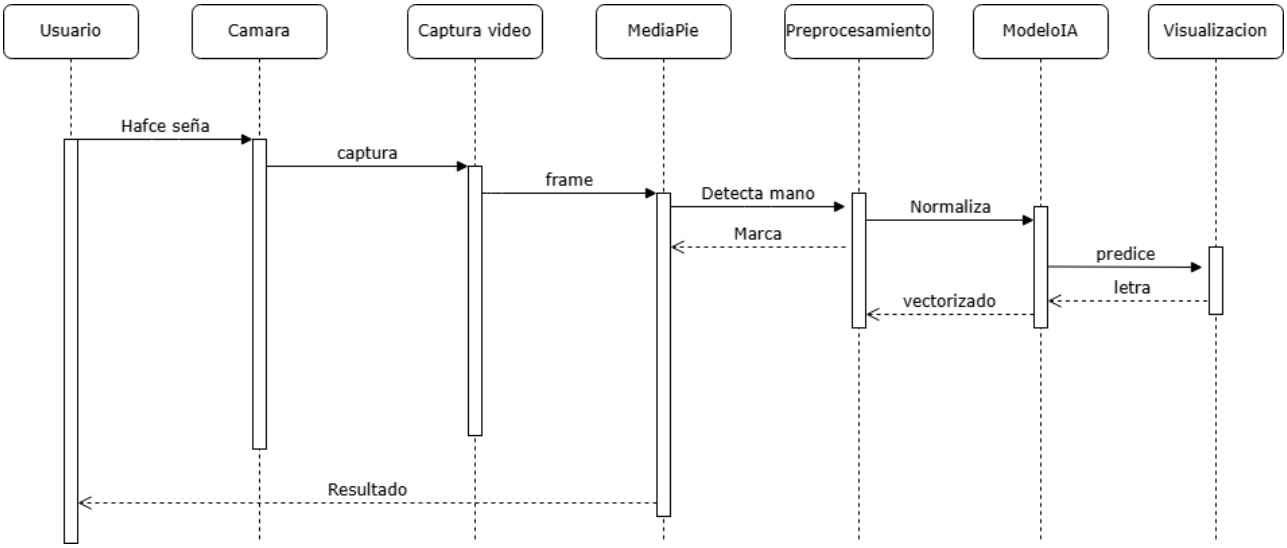
Casos de uso



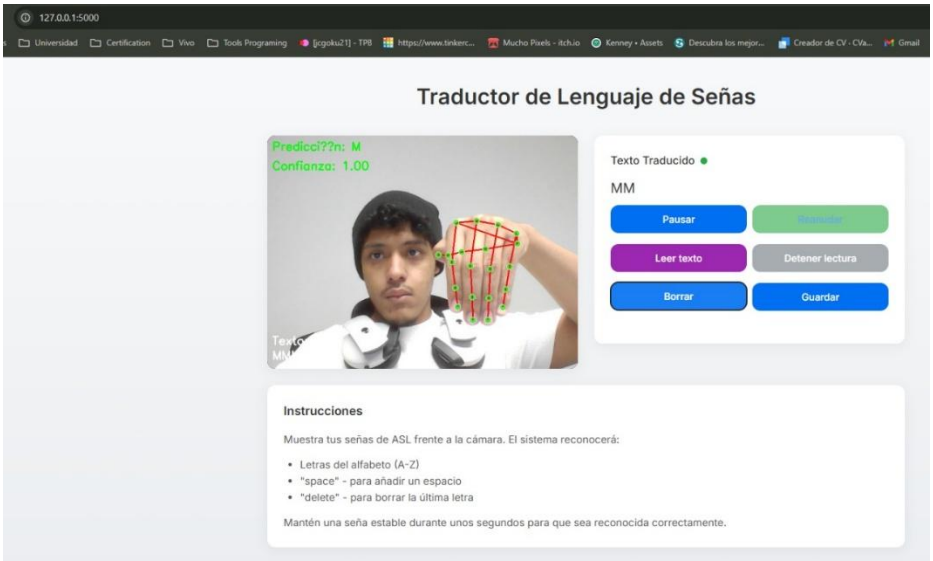
Flujo general del sistema

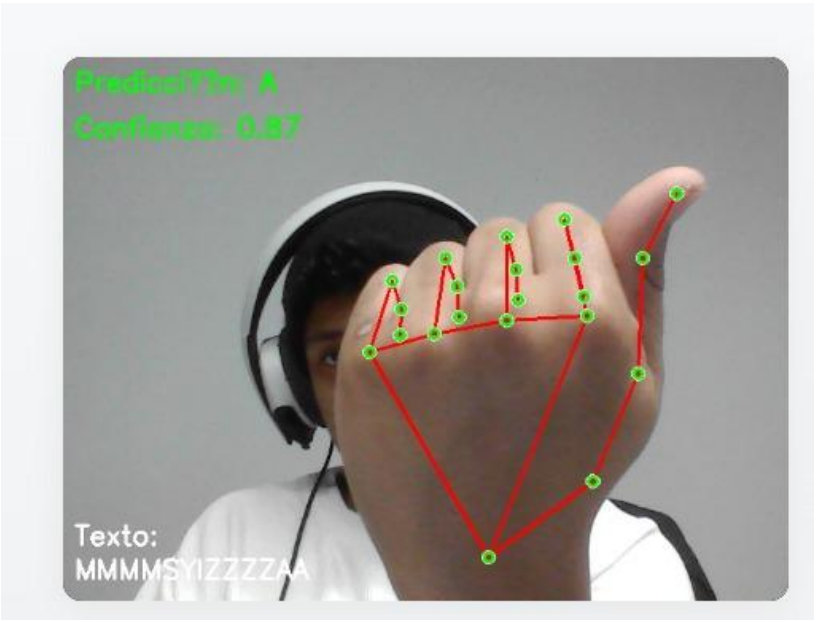


Componentes y secuencia de interacción.



Evidencias del funcionamiento





Traductor de Len

Github: https://github.com/BenIzq/Proyecto-IA-2

Concusiones y aprendizaje

Conclusiones

- Viabilidad del reconocimiento con IA:
  Este proyecto demostró que es posible implementar un sistema funcional de reconocimiento de lenguaje de señas utilizando técnicas de visión por computadora e inteligencia artificial, permitiendo traducir señas a texto en tiempo real.

- Uso eficiente de herramientas modernas:
  Herramientas como OpenCV y MediaPipe facilitaron la captura y procesamiento de imágenes, reduciendo considerablemente la complejidad del desarrollo y permitiendo centrarse en la lógica de reconocimiento y predicción.

- Impacto social del proyecto:
  El sistema desarrollado representa un paso importante hacia la inclusión, al proporcionar una herramienta que puede ayudar a mejorar la comunicación con personas con discapacidad auditiva o del habla.

- Modularidad y escalabilidad:
  La arquitectura modular del sistema permite una fácil mejora y mantenimiento, así como la integración de nuevas señas, mejoras al modelo de IA o cambios en la interfaz de usuario.

Aprendizajes

- Integración de visión por computadora con IA:
  Se profundizó en el uso combinado de bibliotecas de visión artificial y modelos de aprendizaje automático para resolver problemas del mundo real.

- Preprocesamiento y normalización de datos:
  Se adquirió experiencia en la preparación de datos para su uso en modelos de IA, desde la extracción de características hasta su normalización y vectorización.

- Diseño de arquitecturas interactivas:
  El desarrollo del flujo del sistema permitió comprender cómo estructurar una aplicación interactiva que responda en tiempo real a entradas del usuario.

- Desarrollo responsable con impacto social:
  El trabajo dejó en claro la responsabilidad ética que implica crear soluciones tecnológicas con impacto directo en la calidad de vida de los usuarios finales.