UNC Universidad Nacional de Córdoba

FCEFyN
FACULTAD DE CIENCIAS EXACTAS, FÍSICAS y NATURALES

LIDeSIA

Problemas Emergentes

Intérprete Inteligente

LIDeSIA

Objetivos del Proyecto

- Detectar y diferenciar señas, para su posterior traducción/emisión.
- Evitar la “robotización” de la lectura. Emitiendo un mensaje continuo y coherente.

Lengua de Señas

La lengua de señas es un sistema de comunicación visual que posee una estructura, la misma contempla:

- Gramática y estructura lingüística propia
- Variaciones regionales
- Expresiones faciales y corporales incluidas

La Glosa

En el lenguaje de señas, la glosa es el sistema de representación de las palabras habladas mediante gestos, movimientos de manos, expresiones faciales y otros elementos propios de la lengua de señas para transmitir la información de manera clara y comprensible.

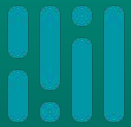
- La glosa se utiliza para interpretar el mensaje hablado en tiempo real.

- Describe la estructura gramatical de la lengua de señas. Esto puede incluir la representación de la configuración de mano, el movimiento, la ubicación en el espacio y otros aspectos gramaticales específicos.

- Incluye descripciones de las expresiones faciales y corporales para transmitir el significado completo de una frase.

Arquitecturas de Estudio

Intérprete Inteligente



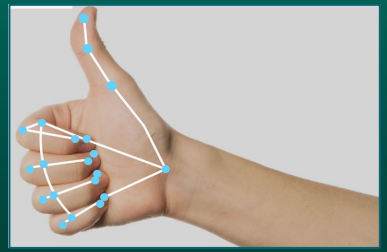
MediaPipe

- Implementación

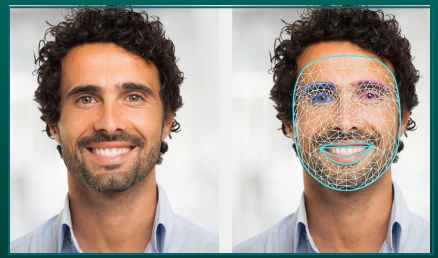
MediaPipe

MediaPipe es una biblioteca y conjunto de herramientas desarrolladas por Google que se utiliza principalmente para el procesamiento y análisis de datos visuales en tiempo real.

Utiliza una arquitectura de redes neuronales para llevar a cabo tareas específicas de procesamiento de medios y visión por computadora.



MediaPipe Hands

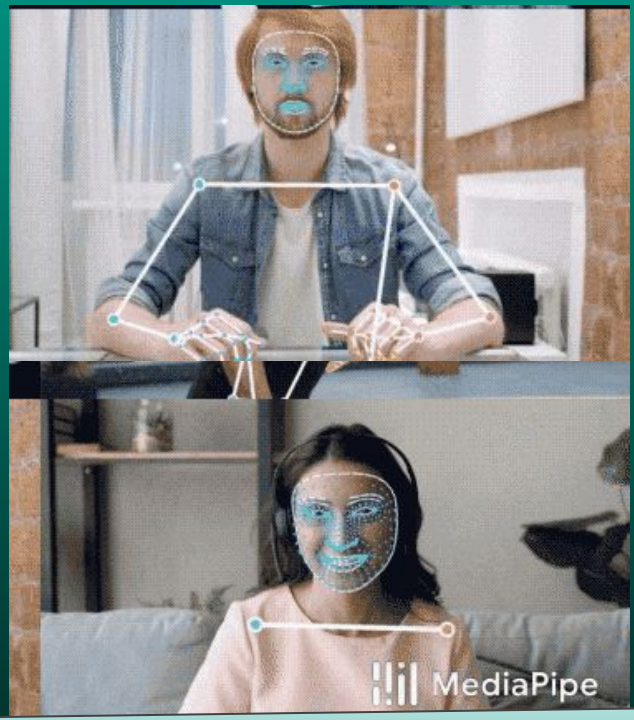


MediaPipe Face Detection

MediaPipe Pose

Holistic Landmarker

Permite combinar componentes de los puntos de referencia de postura , cara y mano para crear un punto de referencia completo para el cuerpo humano. Se puede utilizar esta herramienta para analizar gestos, posturas y acciones de todo el cuerpo. Genera un total de 543 puntos de referencia (33 puntos de referencia de pose, 468 puntos de referencia de cara y 21 puntos de referencia de mano por mano) en tiempo real.



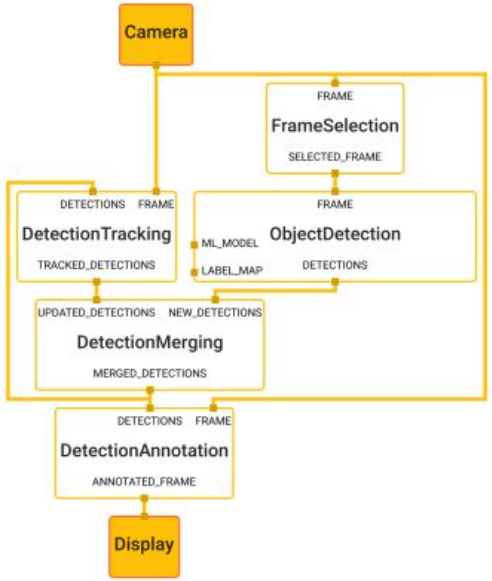


Figure 1: Object detection using MediaPipe. The transparent boxes represent computation nodes (calculators) in a MediaPipe graph, solid boxes represent external input/output to the graph, and the lines entering the top and exiting the bottom of the nodes represent the input and output streams respectively. The ports on the left of some nodes denote the input side packets. See Section 6.1 for details.

Primer implementación y obtención de datos exploratorios de Media Pipe se utilizó Hands landmarks detection.

- Trabaja mediante el uso de un modelo de red neuronal identificando las coordenadas de manos y dedos.
- Puede recibir datos estáticos o un flujo continuo.
- Genera puntos de referencia de las manos en coordenadas de imagen, puntos de referencia de las manos en coordenadas globales y la lateralidad de múltiples manos.

- **Input :**

  El input en MediaPipe Hand Landmarks es un flujo de imágenes en tiempo real que representa las manos o las regiones de las manos que se desean rastrear y analizar. Este consiste en una secuencia de cuadros de video (frames) capturados por una cámara o una fuente de video en tiempo real. Cada cuadro de video es una imagen en la que se buscarán los landmarks (puntos de referencia) de la mano.

- **Output :**

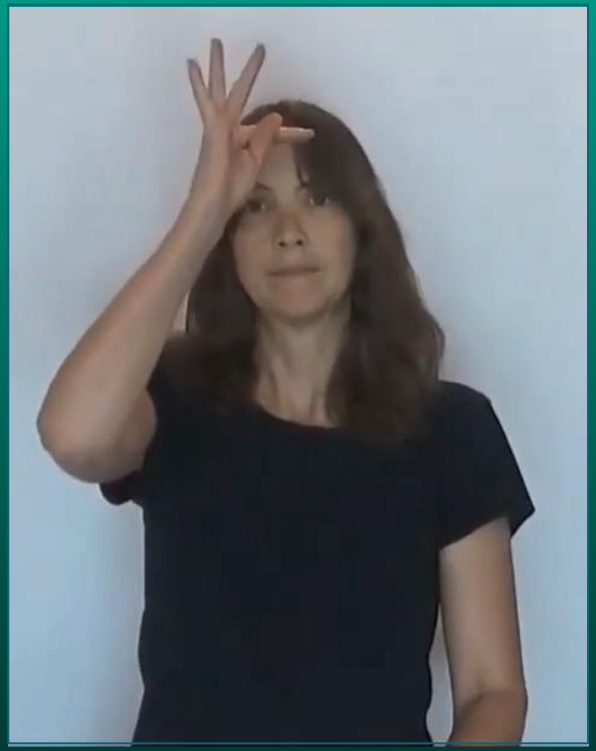
  Muestra los puntos de referencia de las manos detectados y las líneas que los conectan. Los puntos de referencia de las manos detectados por MediaPipe incluyen la posición de la muñeca, la punta de los dedos, la base de los dedos, puntos anatómicos clave de las manos. También se puede acceder a la información adicional. Por ejemplo, la posición y la orientación de las manos, la dirección de los dedos y otros detalles anatómicos.

El input puede variar en resolución y formato, pero típicamente, se utiliza una cámara RGB estándar para capturar los cuadros de video. Los algoritmos de MediaPipe Hand Landmarks pueden funcionar con diversas resoluciones de video, lo que permite su uso en una variedad de aplicaciones, desde dispositivos móviles hasta cámaras web de computadoras.

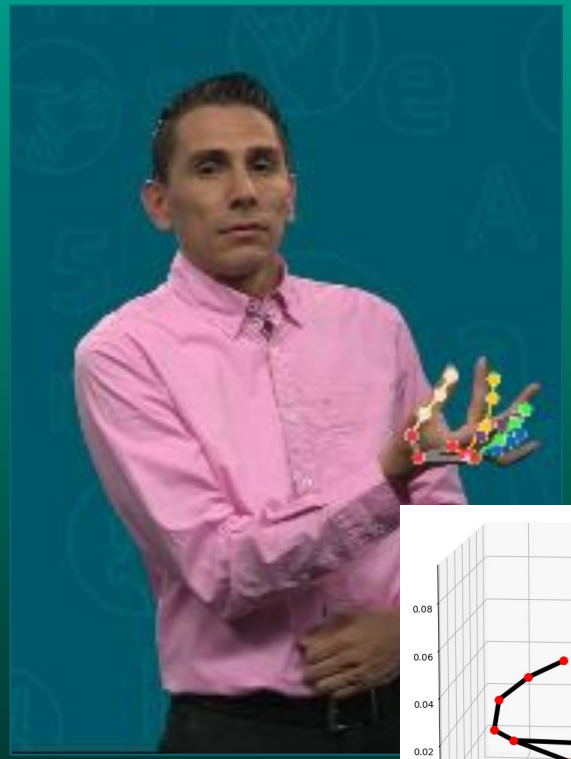
La velocidad de cuadros por segundo (FPS) requerida para el seguimiento de landmarks de manos con MediaPipe Hand Landmarks dependerá de la aplicación específica y de los requisitos de rendimiento de esa aplicación. En general, una velocidad de cuadros más alta proporcionará una experiencia de seguimiento más suave y precisa, pero también requerirá más recursos computacionales.

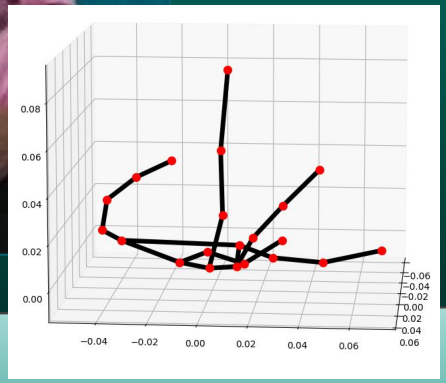
- Un mínimo de 30 FPS.

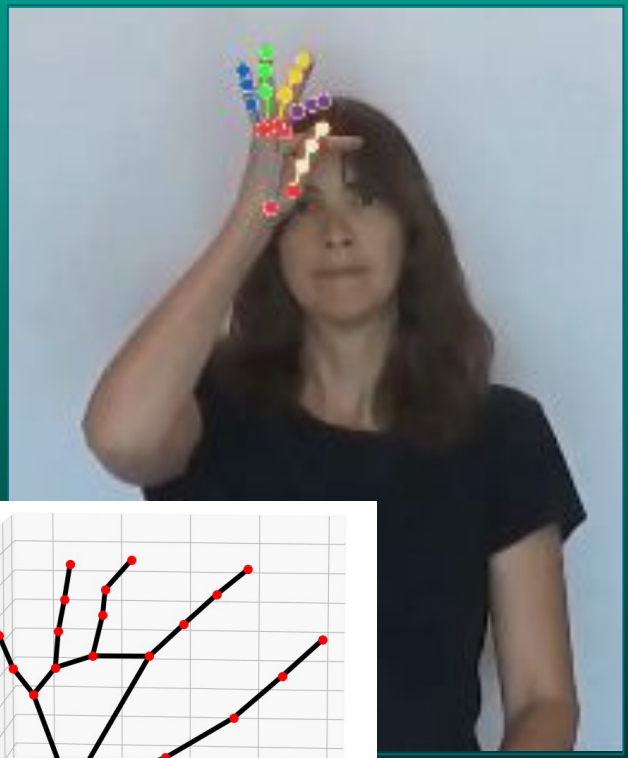
Implementación Media Pipe



Implementación Media Pipe







Convolutional Neural Networks

- Arquitectura
- Antecedentes

Recurrent Neural Networks

- Arquitectura
- Antecedentes

Transfer Learning

- YOLOV8

YOLO se enfoca en la detección de objetos en imágenes y no tiene la capacidad de comprender y traducir el lenguaje de señas como el intérprete que buscamos, que es una tarea mucho más compleja que involucra la interpretación de gestos, movimientos y expresiones faciales. Pero fue capaz de diferenciar señas en movimiento en una prueba simple de 2 palabras.

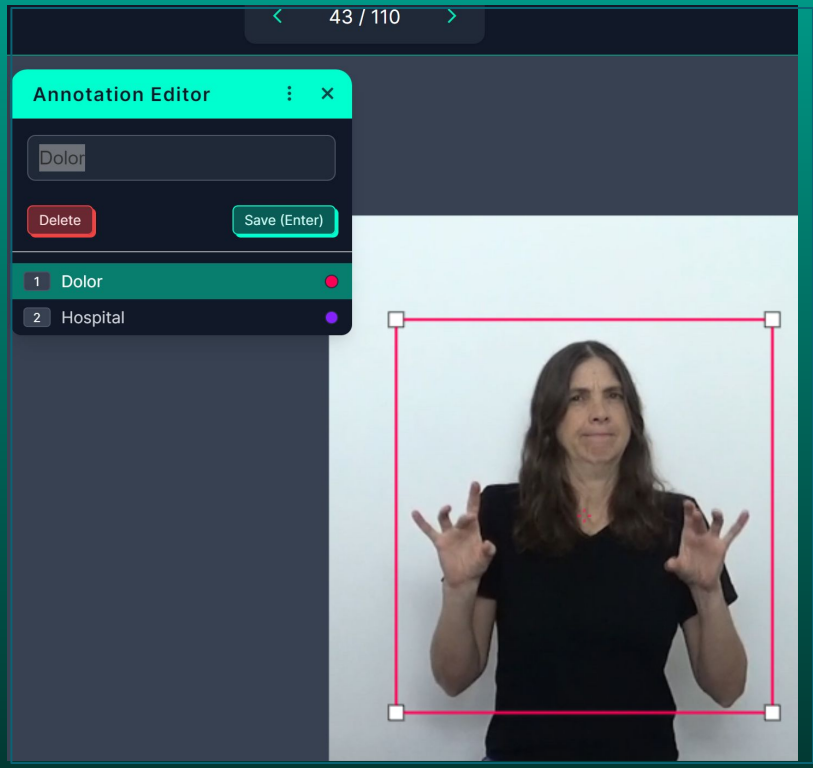
Implementación YoloV8

Se seleccionaron 2 palabras de datos públicos de INSOR:

- Hospital
- Dolor

Pre procesamiento:

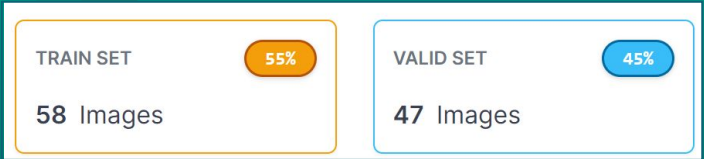
- 5 videos por palabra
- 4 fps por video



Hospital

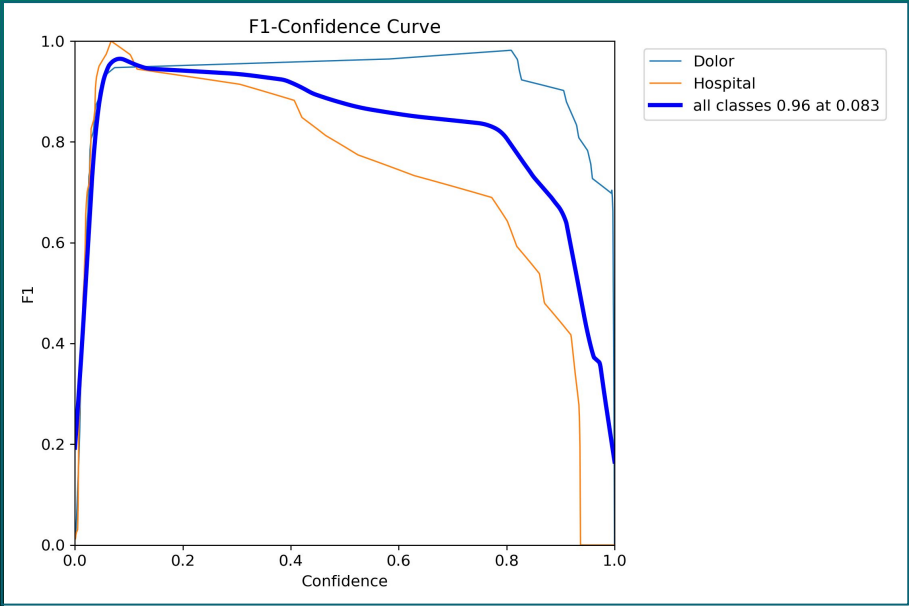
Dolor

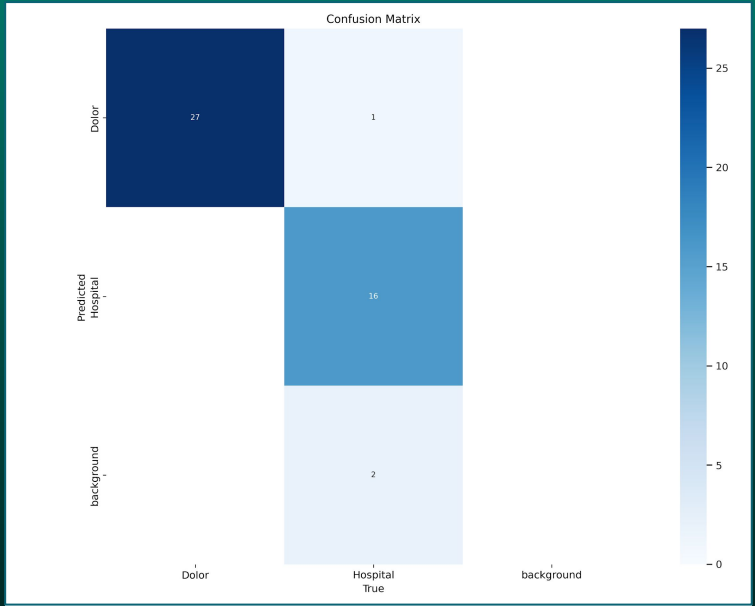
Entrenamiento



- 105 imágenes
- 25 épocas

Métricas





Detección YoloV8

Transformers

- Bert

El modelo BERT (Bidirectional Encoder Representations from Transformers) es una poderosa arquitectura de procesamiento de lenguaje natural pre-entrenada que ha demostrado ser efectiva en una amplia gama de tareas de procesamiento de texto. Es a partir de esto, que consideramos que para realizar un intérprete automático de lenguaje de señas basado en inteligencia artificial, debemos tener en cuenta una prueba con un modelo como BERT.

Capacidad de Aprendizaje por Transferencia: BERT es conocido por su capacidad de aprendizaje por transferencia. Esto significa que podemos aprovechar modelos BERT pre-entrenados en grandes cantidades de texto en lenguaje natural, el objetivo sería afinar el modelo específicamente para el lenguaje de señas.

Contexto Bidireccional: BERT tiene la capacidad de entender el contexto de las palabras en dos direcciones, lo que lo haría efectivo para comprender las relaciones y matices en el lenguaje de señas. Esta característica se vuelve esencial, ya que el lenguaje de señas depende en muchas circunstancias fuertemente del contexto y la expresión facial.

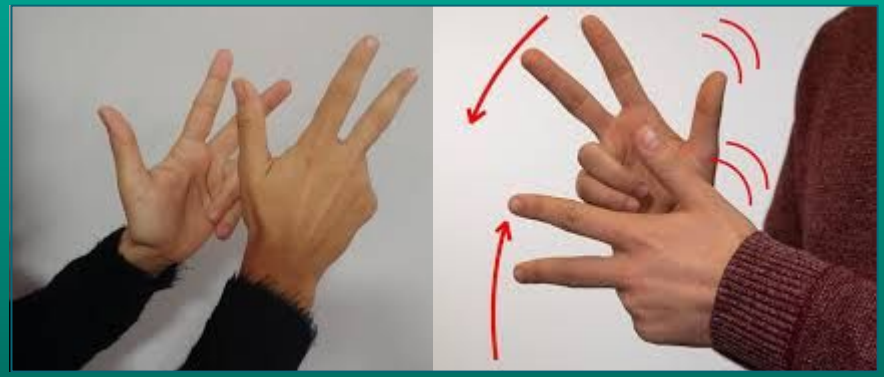
Atención Multi Atención (Self-Attention): BERT utiliza la atención para asignar pesos a las partes relevantes del texto de entrada. Esto sería beneficioso para traducir el lenguaje de señas, ya que puede ayudar a identificar las relaciones entre las señas y las expresiones emocionales.

Base de Datos

Intérprete Inteligente

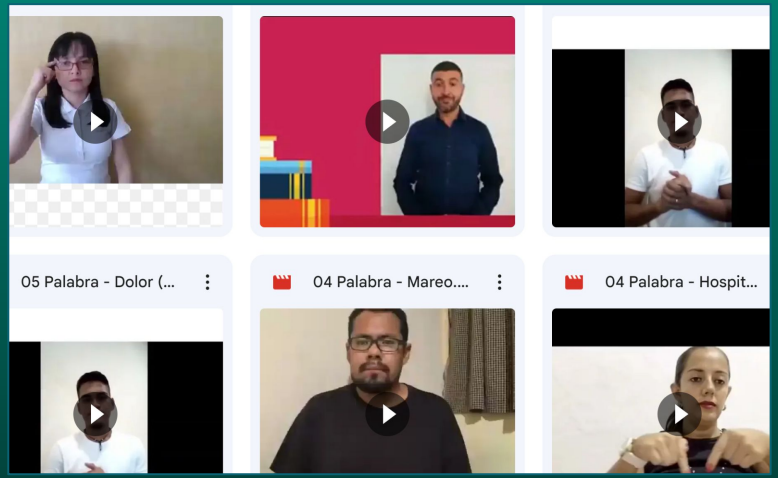
Base de Datos

La implementación de parte de los modelos mencionados requiere la recopilación de una base de datos de videos de lenguaje de señas, que posteriormente debe ser sometida al pre-procesamiento correspondiente a cada uno, con el propósito de ser empleada en el entrenamiento del intérprete automático.



Reunión con equipo de estudiantes IEEE y LIDeSIA

Intercambio de metodología de trabajo sobre la construcción de la base de datos de **Colombia** la cual se comenzó a partir de videos propuestos por el instituto nacional para sordos INSOR y otros recopilados de la web.

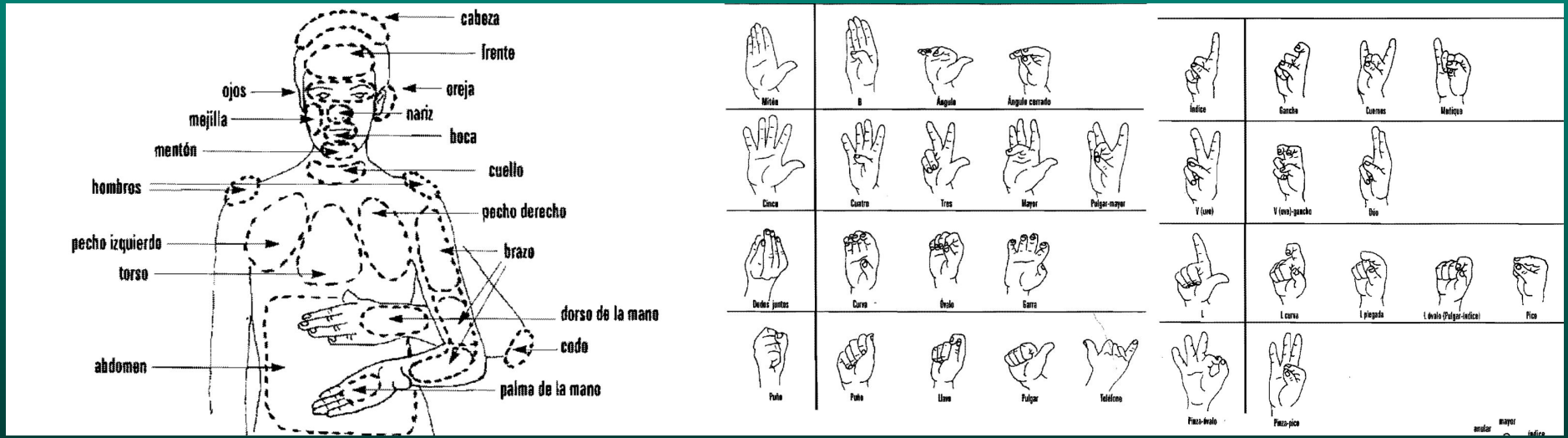


Base de datos de Argentina LSA64

Presenta un conjunto de datos de 64 signos de la Lengua de Señas Argentina (LSA). Contiene 3200 videos de 64 signos LSA diferentes grabados por 10 sujetos, y es un primer paso hacia la construcción de un conjunto de datos integral a nivel de investigación de signos argentinos, específicamente diseñado para el reconocimiento de lenguaje de signos u otras tareas de aprendizaje automático.

Los sujetos que realizaron las señales usaron guantes de colores para facilitar los pasos de seguimiento y segmentación de las manos, lo que permitió que los experimentos en el conjunto de datos se centrarán específicamente en el reconocimiento de señales.

Diccionario de Lengua de Señas Argentina



https://biblioteca.malvinasargentinas.ar/gobierno/GOB_senas-introduccion-diccionario.pdf

Avances de investigación exploratoria LIDeSIA

- Avances sobre la propuesta de arquitectura basada en inteligencia artificial.
- Recopilación de antecedentes.

Antecedentes Colombia

2021

Diseño y Construcción de Prototipo de Software para Reconocer Lenguaje de Señas de Personas con Discapacidad Auditiva

Contexto: ¿Puede implementarse un software que sea capaz de reconocer algunos gestos de LSC de manera automática y que sea capaz de detectarlas con el menor número de desaciertos?. El objetivo es diseñar e implementar un prototipo de software que traduce automáticamente en texto el lenguaje de señas empleado por la población sorda de Colombia usando técnicas de aprendizaje de máquina.

Método: CNN, modificando el modelo base se obtuvo diferentes modelos modificando su arquitectura agregando más capas convolucionales o Dropout e incluso usando los pesos de las capas iniciales y entrenando las capas adicionales usando Transfer Learning y Fine Tuning.

Antecedentes Colombia

2021

Resultados: Este modelo logró una eficiencia del 68% en la clasificación de 15 de las 22 clases disponibles. Comparado con un modelo que simplemente asignaría clases al azar (con una eficiencia del 4.5%), el modelo es significativamente mejor. Sin embargo, el autor destaca, que la eficiencia del modelo no alcanza el umbral aceptable del 80%.

Conclusiones: la construcción de modelos CNN es compleja debido a la gestión de múltiples hiperparámetros. Aumentar la conectividad o el número de capas no garantiza un mejor aprendizaje; a menudo, esto empeora el rendimiento. Las técnicas como Transfer Learning y Fine Tuning pueden acelerar el entrenamiento, pero no siempre mejoran la precisión. En el caso del LSC, la similitud de patrones entre señas y las variaciones ambientales representan desafíos significativos para la detección de patrones. Se requiere un ambiente de captura más controlado para obtener mejoras sustanciales.

Antecedentes Colombia

2021

https://revistas.sena.edu.co/index.php/conciencia/article/view/3926/4602

Contexto: El presente trabajo hace parte de una investigación en desarrollo para el reconocimiento de la Lengua de Señas Colombiana (LSC) en donde se usará la arquitectura de Red Neuronal propuesta en esta fase para el reconocimiento de los símbolos usados para formar las letras del abecedario. Se realiza ajuste de hiperparámetros de una Red Neuronal Convolucional (CNN) para el reconocimiento de manos.

Método: Se propone una arquitectura preliminar y se diseña un experimento para evaluar el porcentaje de clasificación en el reconocimiento de mano derecha e izquierda usando diferentes tamaños de filtros en cada una de las capas de Convolución.

Antecedentes Colombia

2021

Resultados: Los resultados obtenidos han permitido ajustar la red para lograr una clasificación del 100% en pocas épocas de entrenamiento.

Conclusiones: Se realizó un estudio sistemático de los hiperparámetros que influyen en el porcentaje de clasificación, principalmente centrados en el taman;o de los filtros de convolución en una arquitectura CNN.La arquitectura propuesta y el ajuste han permitido un aprendizaje en pocas épocas de entrenamiento, logrando el 100% de clasificación tanto en datos de entrenamiento como en los de validación. El tamaño de los kernel en las primeras dos capas convolucionales son determinantes en la capacidad de aprendizaje de la CNN. Se puede observar que se obtienen mejores resultados cuando los kernel disminuyen su tamaño en capas posteriores.

Antecedentes Colombia 2022

Reconocimiento de lengua de señas colombiana mediante redes neuronales convolucionales y captura de movimiento

Contexto: Este artículo presenta el diseño de un modelo predictivo computacional que facilita el reconocimiento de la lengua de señas colombiana (LSC) en un entorno hotelero y turístico.

Método: Se aplicaron técnicas de inteligencia artificial y redes neuronales profundas en el aprendizaje y la predicción de gestos en tiempo real, los cuales permitieron construir una herramienta para disminuir la brecha y fortalecer la comunicación. Se implementaron algoritmos de redes neuronales convolucionales sobre captura de datos en tiempo real. Se capturó movimiento mediante cámaras de video de dispositivos móviles; así, se obtuvieron las imágenes que forman el conjunto de datos. Las imágenes se utilizaron como datos de entrenamiento para un modelo computacional óptimo que puede predecir el significado de una imagen recién presentada.

Antecedentes Colombia

2022

Resultados: Se evaluó el rendimiento del modelo usando medidas categóricas y comparando diferentes configuraciones para la red neuronal. Adicional a esto, todo está soportado con el uso de herramientas como Tensorflow, OpenCV y MediaPipe.

Conclusiones: Se obtuvo un modelo capaz de identificar y traducir 39 señas diferentes entre palabras, números y frases básicas enfocadas al sector hotelero, donde se logró una tasa de éxito del 97,6 % en un ambiente de uso controlado.

Antecedentes Argentina

2021

Reconocimiento de Lengua de Señas con Redes Neuronales Recurrentes Iván Mindlin

Contexto: El reconocimiento automático de lenguas de señas apunta a convertir señas capturadas por video a texto, en un lenguaje escrito dado. Típicamente consiste en un pipeline de tareas que comienzan con el reconocimiento de la forma de las manos de quien señala, su movimiento y posición, la forma de los labios y expresiones faciales, además del trasfondo, en cada frame del video. Luego, las señas en el video deben ser clasificadas y traducidas al lenguaje escrito, como español o inglés.

Antecedentes Argentina

2021

Método: Se desarrollan modelos de reconocimiento de lenguas de señas basados en técnicas de aprendizaje automático profundo (Deep Learning) para clasificar las señas de la base de datos LSA64. Para los experimentos de reconocimiento se crearon tres modelos base los cuales fueron llamados, ConvLSTM, MobileNet y Conv3D debido a las arquitecturas utilizadas para crearlos. Aunque todos tienen capas convolucionales, varían principalmente en su procesamiento de la codificación temporal de los videos.

Antecedentes Argentina 2021

Reconocimiento de Lengua de Señas con Redes Neuronales Recurrentes Iván Mindlin

Resultados: En la tabla se muestra los 10 mejores resultados obtenidos en una serie de experimentos previos. En estos experimentos, se evaluó el rendimiento de diferentes técnicas en el conjunto de datos LSA64, que implica el procesamiento de videos. Consideran que el mejor resultado se obtuvo utilizando la arquitectura Conv3D y videos muestreados a 32 frames, con segmentación de manos. Aunque este resultado es ligeramente inferior al obtenido en la LSTM con procesamiento complejo de los videos, la diferencia es pequeña, lo que es notable ya que las técnicas utilizadas son estándar en el campo de Visión por Computadora.

Antecedentes Argentina 2021

Cuadro 4.2: Comparación de resultados de distintos modelos entrenados para clasificar LSA64. El modelos basado en Conv3D con segmentación de las manos alcanza el estado del arte (marcado en negrita).

Modelo	Features	accuracy promedio (%)	σ
ConvLSTM	RGB-64x64-32 frames	97.65	0.0102
ConvLSTM	RGB-64x64-16 frames	98.18	0.083
ConvLSTM	RGB-64x64-16 frames-segmented hands	98.37	0.0143
MobileNet	RGB-128x128-32 frames	98.43	0.0071
MobileNet	RGB-128x128-16 frames	98.90	0.0038
MobileNet	RGB-128x128-16frames-segmented hands	98.98	0.0030
Conv3D	RGB-128x128-32 frames	97.65	0.0128
Conv3D	RGB-128x128-16 frames	97.87	0.0027
Conv3D	RGB-128x128-16 frames-segmented hands	**99.4**	0.0007
LSTM + DSC [37]	Skeletal + RGB + Optical Flow	99.84	-

Conclusiones: El modelo con mayor de velocidad de procesamiento fue Conv3D así como el que alcanzó mayor accuracy. La mayor cantidad de errores en clasificación se dio por señas con movimientos similares y distinta configuración de mano. La base de datos LSA64 es fácil de utilizar debido a su buena organización, pero no representa un desafío para el estado del arte en reconocimiento de lenguas de señas. Con métodos estándar basados en Deep Learning para la visión por computador y reconocimiento de acciones se obtuvieron resultados cercanos a los mejores registrados hasta el momento.

Antecedentes Argentina

2022

Inteligencia artificial que interpreta lenguaje de señas | UNICEN

Contexto: Surgió a partir de querer comunicarse con una persona sordomuda de la forma más fluida posible y la motivación fue usar inteligencia artificial.

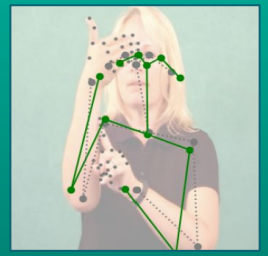
"El objetivo es poder conformar un equipo con programadores, organizaciones y personas vinculadas a la lengua de señas para poder seguir desarrollando la inteligencia artificial ampliando sus términos y haciendo que funcione mejor"

Antecedentes Argentina

2022

Observaciones: En los compromisos establecidos, el autor destaca que, a pesar de los avances en la legislación, el acceso a la comunicación de las personas con discapacidad, y en particular de las personas sordas, sigue enfrentando desafíos significativos. Un aspecto crítico es la falta de reconocimiento de la Lengua de Señas Argentina (LSA) como un idioma y un valioso patrimonio lingüístico y cultural de la comunidad sorda. Esta carencia tiene un impacto sustancial en la creación de obstáculos que limitan el acceso a sus derechos en igualdad de condiciones, según se advierte.

Antecedentes Transformers



Sign Psedose-Ba Transformer for Word-Level Sign Language Recognition

Contexto: En este artículo se presenta un sistema para el reconocimiento de lenguaje de señas a nivel de palabras basado en el modelo de transformer. El objetivo es una solución con un bajo coste computacional, ya que vemos un gran potencial en el uso de dicho sistema de reconocimiento de dispositivos portátiles. Basamos el reconocimiento en la estimación de la pose del cuerpo humano en forma de puntos de referencia en 2D. Introducimos un esquema robusto de normalización de poses que toma en consideración el espacio de signos y procesa las poses de la mano en un sistema de coordenadas local separado, independiente de la postura del cuerpo.

Antecedentes Transformers

Describen que lograron resultados de primer nivel con tecnologías de punta en los conjuntos de datos WLASL y LSA64. Para WLASL podemos reconocer con éxito el 63,18% de las grabaciones de señas en el subconjunto 100-gloss, lo que supone una mejora relativa del 5% con respecto de la técnica anterior. Para el subconjunto de 300-gloss, logramos una tasa de reconocimiento del 48,78% lo que representa una mejora relativa del 3,8%. Con el conjunto LSA64 informamos un precisión de reconocimiento de prueba del 100%.

Método: SPOTER - Sign POsebased TransformER refleja el hecho de que manejamos la pose del cuerpo de acuerdo con el espacio de signos detectado, y usaron un transformer para clasificar la secuencia de poses en una glosa de signos.

Antecedentes

Sign Pose-Based Transformer for Word-Level Sign Language Recognition

Conclusión: Proponemos un enfoque novedoso de utilizar Transformer para esta tarea. El modelo opera sobre representaciones de secuencias de poses corporales, aplicamos el conocimiento de la lingüística SL para crear una técnica de normalización robusta, así como nuevas técnicas de aumento de datos específicas para el SL.

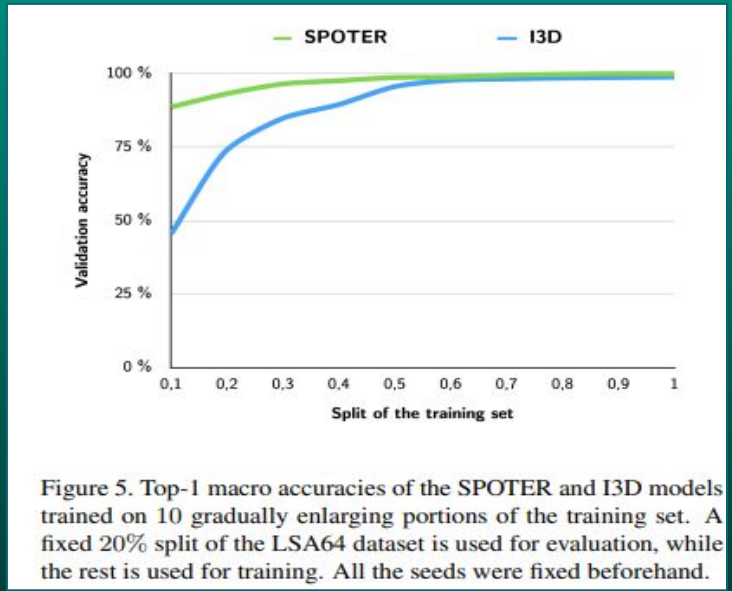


Figure 5. Top-1 macro accuracies of the SPOTER and I3D models trained on 10 gradually enlarging portions of the training set. A fixed 20% split of the LSA64 dataset is used for evaluation, while the rest is used for training. All the seeds were fixed beforehand.

Antecedentes

Validamos nuestro enfoque en dos conjuntos de datos para aislados. Logramos resultados generales de vanguardia para el LSA64 y resultados de última generación establecidos en la categoría de modelo basado en poses para WLASL. También hemos realizado un estudio de rendimiento comparando nuestro modelo con el Línea de base I3D, que demostró que la arquitectura recientemente propuesta es sustancialmente menos exigente y se generaliza bien incluso en conjuntos de entrenamiento muy pequeños.

Antecedentes

Joint End-to-End Sign Language Recognition and Translation

Contexto: Trabajos anteriores sobre traducción de lengua de señas han mostrado que tener una glosa de señas de nivel medio mejora drásticamente el rendimiento de la traducción. De hecho, el estado actual de la traducción requiere tokenización de nivel de glosa para funcionar. Presentamos una novedosa arquitectura basada en transformer que aprende conjuntamente el reconocimiento y la traducción continua de lenguaje de seña y al mismo tiempo se puede entrenar de un extremo a otro. Eso se logra mediante una perdida de CTC (Clasificación Temporal Conexionista) para unir problemas de reconocimiento y traducción en una única arquitectura unificada.

No requiere ninguna información de tiempo real, resuelve simultáneamente dos problemas de aprendizaje de secuencias codependientes y conduce a mejoras significativas en el rendimiento.

Evaluamos los datos de reconocimiento y traducción de nuestros enfoques en el conjunto de datos RWTHPHOENIX-Weather-2014T (PHOENIX14T).

Antecedentes

Conclusión: Nuestras redes de traducción superan a los modelos de traducción de videos de señas a lenguaje hablado y de glosa a lenguaje hablado y en algunos casos duplica su rendimiento. Nuestro primer conjunto de datos ha demostrado que el uso de características que fueron previamente entrenadas en datos de signos supero el uso de representaciones genéricas espaciales basadas en ImageNet. Además el aprendizaje conjunto de reconocimiento y traducción mejoro el rendimiento de ambas tareas. Hemos superado los resultados de traducción texto a texto que se estableció como límite superior virtual. A futuro se quiere ampliar el enfoque para modelar múltiples articuladores de signos es decir rostros, manos y cuerpo y alentar a nuestras a redes a aprender la relación lingüística entre ellos.

LEY 27.710 de LENGUA DE SEÑAS ARGENTINA

Aprobada el 13 de Abril del 2023

- Artículo 1° - La presente ley tiene por objeto reconocer a la Lengua de Señas Argentina (LSA) como una **lengua natural y originaria** que conforma un legado histórico inmaterial como parte de la **identidad lingüística** y la **herencia cultural** de las personas sordas en todo el territorio de la Nación Argentina, y que **garantiza su participación e inclusión plena**.

- Artículo 2° - LSA como aquella que se transmite en la **modalidad visoespacial**. Posee una **estructura gramatical completa, compleja y distinta del castellano**. Al ser visual, la LSA es completamente accesible desde el punto de vista **perceptual** para las personas sordas, como así también para todas las personas que elijan utilizar para:

Comunicarse, transmitir sus deseos e intereses, informarse, defender sus derechos y construir una identidad lingüística y cultural positiva que les permita participar y trascender plenamente en todos los aspectos de la vida social.

LEY 27710 de LENGUA DE SEÑAS ARGENTINA

- Artículo 4° - **Promoción de la Lengua de Señas Argentina (LSA)**.

Accesibilidad efectiva y plena a la vida social - Eliminar barreras comunicacionales y actitudinales facilitando el acceso a la comunicación e información por parte de las personas que se comunican en la LSA en su interacción con el entorno - Equiparar oportunidades tendientes a impulsar y fortalecer su independencia, y autonomía personal y toma de decisiones - Diseñar y ejecutar estrategias que aseguren la accesibilidad comunicacional en todas las políticas públicas dirigidas a la sociedad.

ELdeS - Uruguay

“Primera plataforma del mundo en permitir la enseñanza de Lengua de Señas de forma autoadministrada e interactiva mediante el uso de Inteligencia Artificial y detección de movimientos."
Agencia Nacional de Investigación e Innovación (ANII)

La lengua de señas es propia de cada país y por eso coordinan con los referentes locales para poder brindarte un contenido de excelencia.

Los respaldan las instituciones públicas y privadas referentes en educación, tecnología y Lengua de Señas del país donde se aplique la plataforma.

Creada para personas, colegios, empresas. https://www.somoseldes.com/

EldeS Plataforma Argentina de IA y detección de movimientos

Desarrollo del Centro Universitario de Idiomas de la UBA con la coordinación de la profesora Romina Aza, que dirige el instituto Señas de Comunicación junto a Enseñanza de Lengua de Señas, con el aval de la UTN.

Es la **primera plataforma de LSA en operar con inteligencia artificial** y ello representa un avance significativo en la enseñanza y en la preservación de la lengua de señas argentina.

La plataforma detecta los movimientos de las manos y el rostro para hacer una devolución inmediata sobre la práctica.
https://www.somoseldes.com/

LIDeSIA


ELdeS • LSA

Competencias Google

Reconocimiento de Deletreo de Dedos en American Sign Language (ASL)

El objetivo de esta competencia fue detectar y traducir el lenguaje de señas americano (ASL) de deletreo de dedos en texto. El objetivo fue crear un modelo entrenado con el conjunto de datos más grande de su tipo, lanzado específicamente para esta competencia. Los datos incluyen más de tres millones de caracteres deletreados con los dedos producidos por más de 100 personas sordas, capturados a través de la cámara frontal de un teléfono inteligente en una variedad de fondos y condiciones de iluminación.

https://www.kaggle.com/competitions/asl-fingerspelling

Propuesta Ganadora

La solución describe una arquitectura de red neuronal que consta de dos partes principales: un codificador y un decodificador.

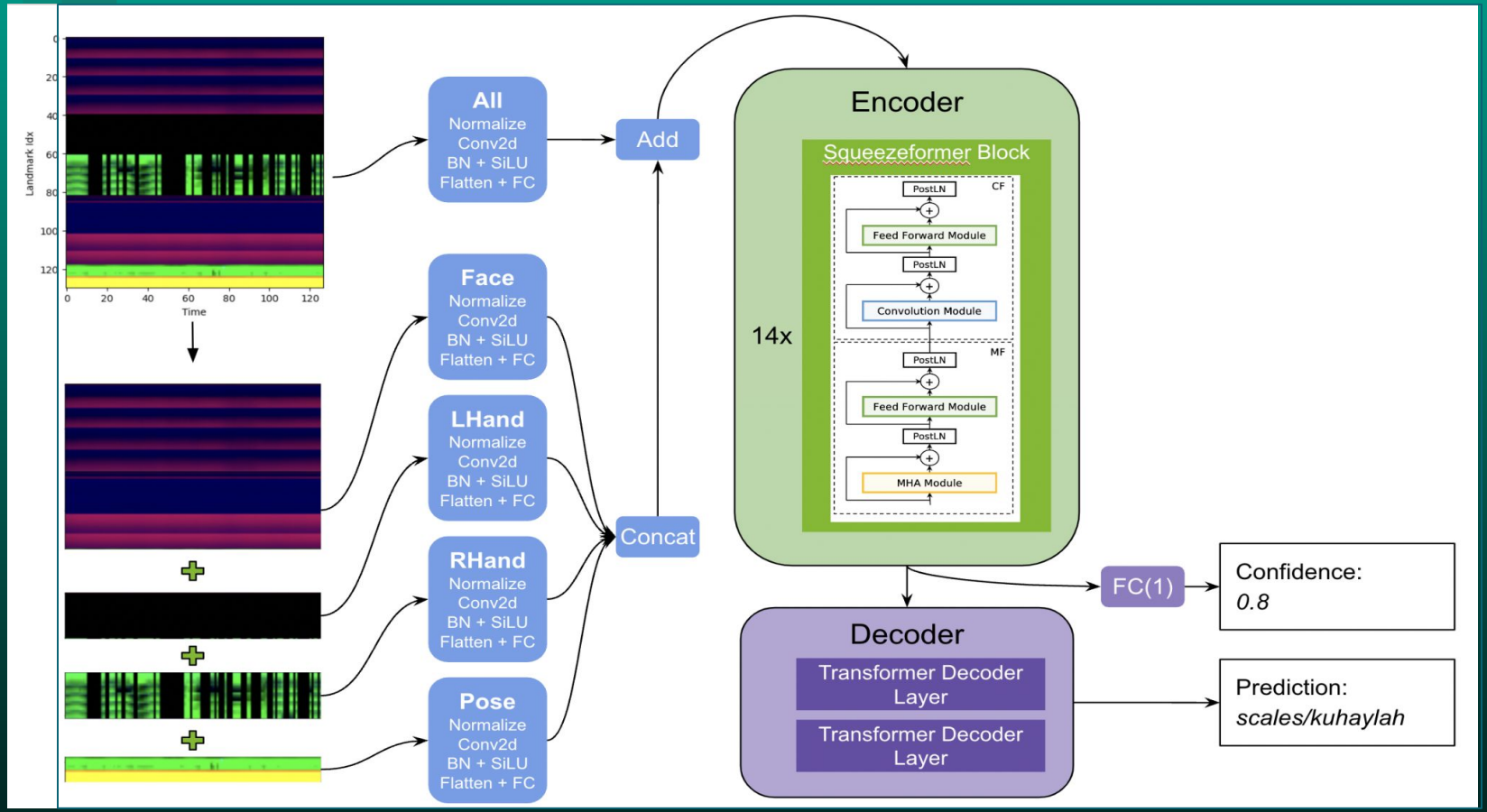
Codificador Mejorado: Mencionan que han mejorado significativamente un modelo llamado "Squeezeformer" para que pueda manejar puntos de referencia generados por Mediapipe en lugar de señales de voz.

Decodificador simple de 2 capas: Describen que el decodificador es una red neuronal simple con solo 2 capas.

Propuesta Ganadora

Puntuación de Confianza: Además de la tarea principal, también mencionan que han incorporado un componente en la red que predice una puntuación de confianza. Lo que consideran puede ser útil en el procesamiento posterior de los resultados.
Además para mejorar el rendimiento y la generalización del modelo, introdujeron técnicas de aumento de datos. Ayudando a que el modelo sea más robusto frente a diferentes tipos de datos.

Squeezeformer + TransformerDecoder + aumentos inteligentes mejorados



Reconocimiento de Lenguaje de Señas Aislado para mejorar los juegos educativos de PopSign para aprender ASL.

El objetivo de esta competencia es clasificar signos aislados del Lenguaje de Señas Americano (ASL). Crear un modelo TensorFlow Lite entrenado con datos de referencia etiquetados extraídos utilizando la solución MediaPipe Holistic.

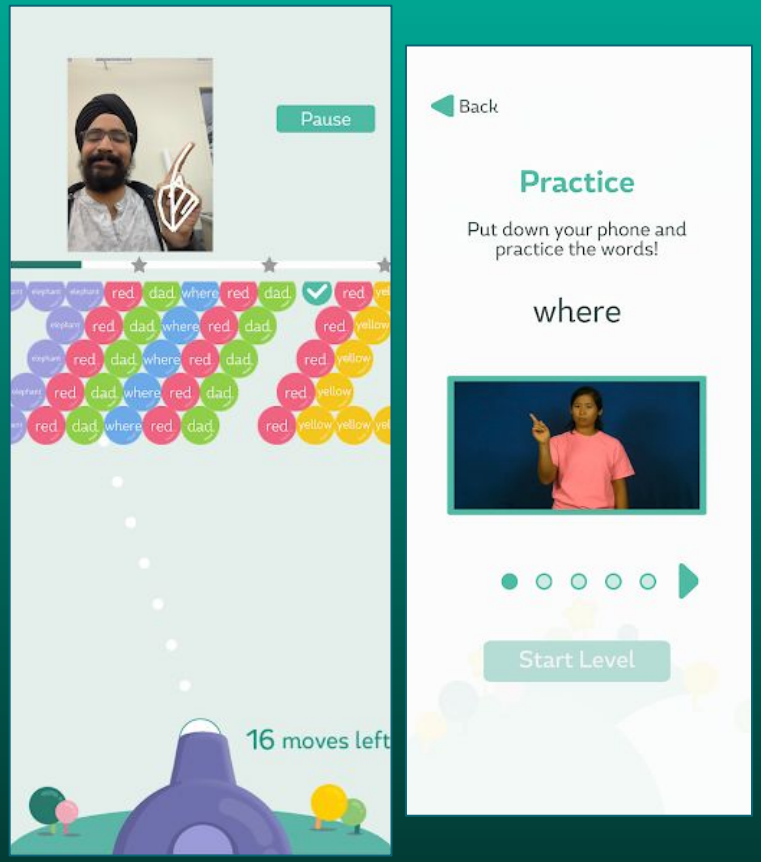
Reconocimiento de Lenguaje de Señas Aislado para mejorar los juegos educativos de PopSign para aprender ASL.

PopSign es una aplicación de juego para teléfonos inteligentes que hace que aprender el Lenguaje de Señas Americano sea divertido, interactivo y accesible. Los jugadores combinan videos de signos en ASL con burbujas que contienen palabras escritas en inglés para explotarlas. PopSign está diseñado para ayudar a los padres de niños sordos a aprender ASL, pero está abierto a cualquier persona que quiera aprender el vocabulario del lenguaje de señas. Al agregar un reconocedor de lenguaje de señas de esta competencia, los jugadores de PopSign podrán firmar el tipo de burbuja que desean disparar, lo que les brindará la oportunidad de practicar el signo ellos mismos en lugar de simplemente ver videos de otras personas haciendo señas. https://www.kaggle.com/competitions/asl-signs

Popsign

PopsignAI combina la jugabilidad de Popsign con el reconocimiento del lenguaje de señas, y se desarrolló a partir de más de 220.000 ejemplos totales de 250 señas recopiladas por Deaf Professional Arts Network de 47 firmantes sordos para quienes la ASL es su primer idioma.

DPAN.TV



Propuesta Ganadora

1D CNN combinado con Transformer

En la solución se menciona la combinación de una Convolutional Neural Network (CNN) 1D y un Transformer para abordar el problema. Se entrenó el modelo desde cero utilizando todos los datos de entrenamiento disponibles en la competencia. Para entrenar se utilizó TensorFlow y una Unidad de Procesamiento Tensorial (TPU) en Google Colab para garantizar la compatibilidad con TensorFlow Lite.

En su prueba el ganador comenta descubrir que la CNN 1D pura superó fácilmente al Transformer y, como resultado, logró una puntuación alta en una métrica pública (LB public score) de 0.80 utilizando solo la CNN 1D al final.
A pesar de que la CNN 1D fue efectiva, encontró la utilidad del Transformer, usándolo en conjunto con la CNN 1D. Visualizó a la CNN 1D como una especie de "tokenizador entrenable" que se combina con el Transformer para mejorar aún más el rendimiento del modelo.

Enlaces de Interés

- *CAS – Confederación Argentina de Sordos*. (2016). Cas.org.ar. https://cas.org.ar/

- *ASAM*. (2022). Asamutual.org.ar. https://www.asamutual.org.ar/

- *Asociaciones | FUNDASOR*. (2023). Fundasor.org.ar. https://www.fundasor.org.ar/senando-en-familia/asociaciones/

- admin. (2023, April 19). *INSOR | Instituto Nacional para Sordos – Trabajando por la Población Sorda Colombiana*. Insor.gov.co. https://www.insor.gov.co/home/

Equipo de Trabajo

- *Directora: Dra. Laura Cecilia Diaz Davila (LIDeSIA - FCEFyN)*
- *Coordinadora: Aybar Lourdes (LIDeSIA - FCEFyN)*
- *Benitez Josefina Victoria (LIDeSIA - FCEFyN)*

LIDeSIA