

DATA SCIENCE - PROYECTO 2 - ANÁLISIS EXPLORATORIO

RECONOCIMIENTO DE LENGUAJE DE SEÑAS

GRUPO 5: ALINA CARÍAS, MARCOS DÍAZ, IGNACIO MÉNDEZ,
ARIELA MISHAAN Y DIEGO SOTO

NUESTRA PROBLEMÁTICA

Las personas sordas y con pérdida auditiva carecen de tecnologías accesibles para ingresar texto en dispositivos móviles. Aunque el fingerspelling en ASL (American Sign Language) es más rápido que escribir en teclados virtuales, aún no existen modelos robustos que automaticen su traducción a texto.

OBJETIVOS

Desarrollar un modelo de deep learning que traduzca fingerspelling en ASL a texto escrito.

Diseñar un proceso de preprocessamiento de landmarks que reduzca el volumen de datos y preserve la información relevante.

Entrenar y optimizar un modelo exportable a TensorFlow que cumpla con restricciones de memoria y tiempo de procesamiento.





TECNOLOGÍAS Y DESAFÍOS

El dataset contiene más de tres millones de caracteres deletreados por más de 100 personas sordas, capturados mediante la cámara frontal de un teléfono inteligente en condiciones variadas de iluminación y fondo. Los datos incluyen secuencias de coordenadas espaciales extraídas de videos, que detecta landmarks en la cara, manos y postura corporal.



TÉCNICAS POSIBLES

Exploración de técnicas de aprendizaje profundo para secuencias: redes recurrentes (LSTM, GRU), transformers para series temporales, entre otras.

COMPLEJIDAD

Grabaciones en diferentes fondos, ángulos y niveles de iluminación, → realismo pero más complejidad.

MANEJO DE DATOS

Manejo de datos masivos (189 GB), lo que obliga a definir estrategias de carga parcial, muestreo o uso de servicios en la nube.

DATOS IMPORTANTES



- **Face:** 467 puntos que representan el movimiento.
- **Manos:** 21 puntos para cada una
- **Pose:** 32 puntos.

Cada mano se representa con 21 landmarks distribuidos en la palma y dedos:

- 1. Muñeca
- 2-5. Dedos pulgares
- 6-9. Dedo medio
- 14-17. Dedo anular
- 18-21. Dedo meñique

Cada landmark tiene coordenadas x, y, z, donde x e y representan la posición relativa en la imagen y z representa la profundidad o distancia desde la cámara. Estos landmarks permiten capturar con detalle la forma, orientación y movimiento de la mano durante la ejecución de una señal (Google, s.f.). Se utilizará MediaPipe de Google para detección de manos en imágenes o videos y estimación de los 21 landmarks de manera rápida y eficiente.



DESCRIPCIÓN DEL DATASET Y VARIABLES DISPONIBLES

El conjunto de datos está compuesto por 124 archivos de formatos Parquet, CSV y JSON, con un tamaño total de 189.09 GB. Esto incluye archivos de entrenamiento, metadata suplementaria y los landmarks extraídos.

Variables

- **Path**: ruta al archivo de landmarks.
- **File_id**: identificador único del archivo.
- **Participant_id**: identificador del contribuyente de datos.
- **Sequence_id**: identificador único de cada secuencia de landmarks.
- **Phrase**: etiqueta textual (frase) asociada a la secuencia.
- **Frame**: número de cuadro dentro de la secuencia.
- **x/y/z_[type]_[landmark_index]**: coordenadas espaciales para cada uno de los 543 landmarks, clasificados por tipo (face, left_hand, pose, right_hand).
- **Character_to_predictions_index.json**: mapeo entre caracteres y sus índices de predicción.

ANÁLISIS EXPLORATORIO

CARGA DE DATOS

Se leen archivos **.parquet** que contienen las coordenadas de los landmarks.
Se selecciona una secuencia específica para analizarlas

EXPLORACIÓN

Se escalan los datos para normalizarlos y se crean funciones para visualizar landmarks en distintos frames.

DEFINICIÓN DE CONEXIONES

Se crean listas que definen cómo se conectan los puntos de la mano, la cara y el cuerpo que servirán para ver y analizar los movimientos.

ANÁLISIS EXPLORATORIO



VALORES DE LAS COORDENADAS

- **Valores de x:** entre -1.2848 y 2.4284
 - **Valores de y:** entre -0.0286 y 3.2895
 - **Valores de z:** entre -6.6786 y 6.2112
- Será necesaria la normalización de los datos para el entrenamiento del modelo.



VALORES NULOS

Se analizó un parquet:

- La mayoría de valores nulos está en la coordenada z.
- Hay alrededor de 16 millones de NaNs en total (normal porque Mediapipe no reconoce todos los landmarks).
- Hay muchos más valores en la mano derecha que en la mano izquierda (esta es solamente auxiliar).

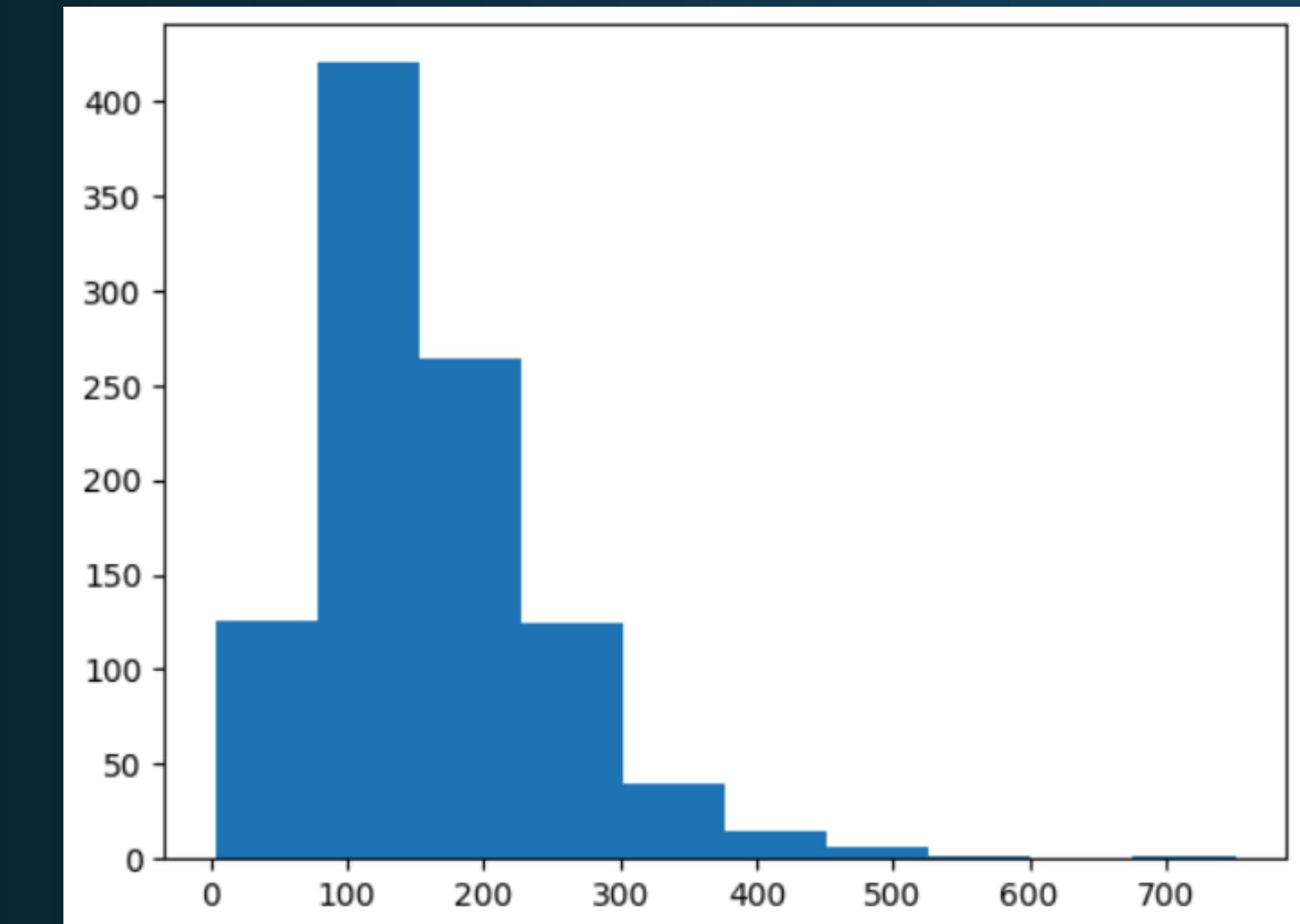
ANÁLISIS EXPLORATORIO



CANTIDAD DE FRAMES DE LAS SECUENCIAS

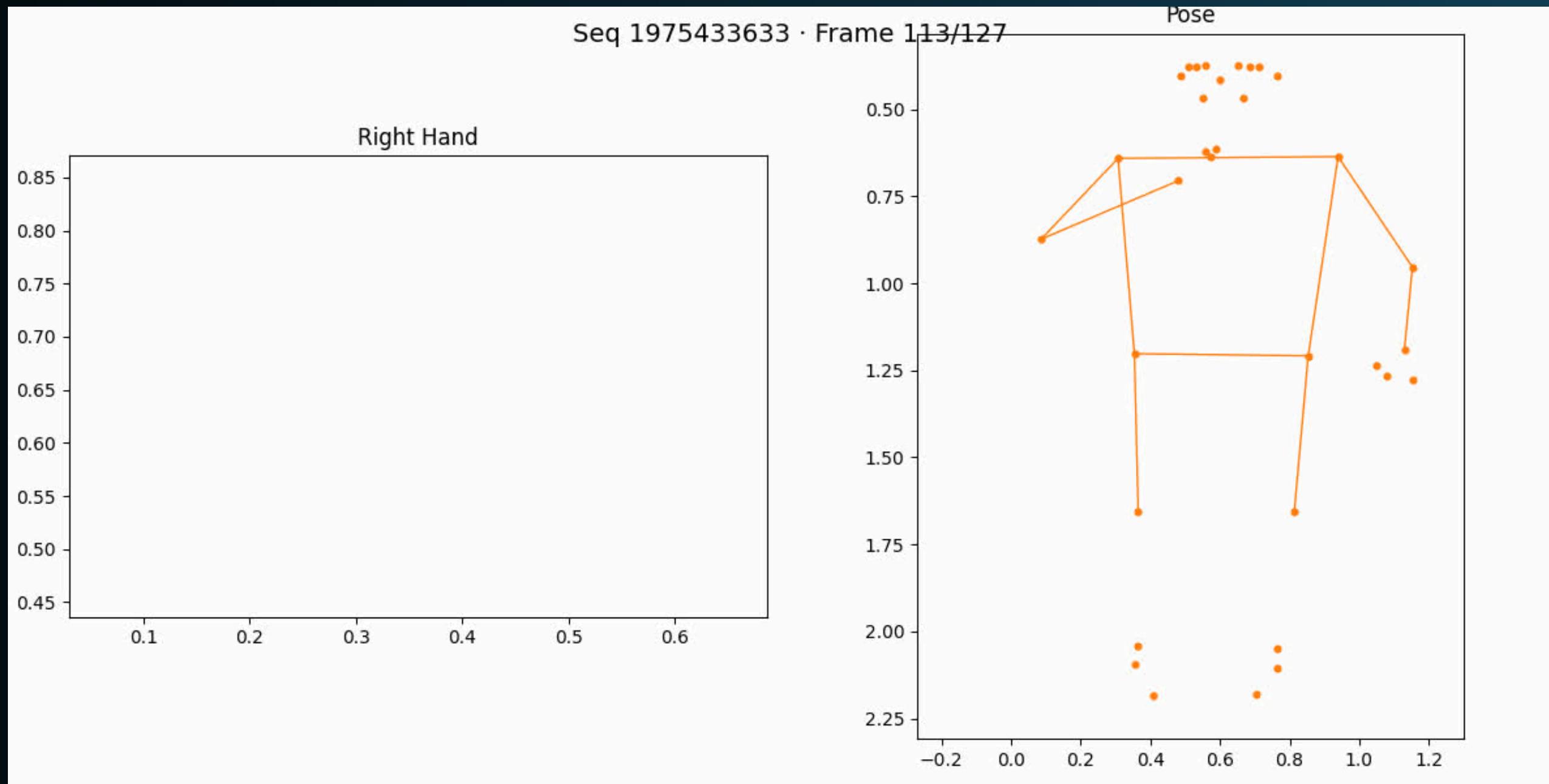
Toma los landmarks en coordenadas (x, y, z) de cada frame y calcula la distribución de valores para cada coordenada.

- La mayoría de los datos se concentra entre 90 y 150 frames. Luego.
- **Promedio de frames por secuencia:** 161.78
- **Máximo de frames:** 751
- **Mínimo de frames:** 3

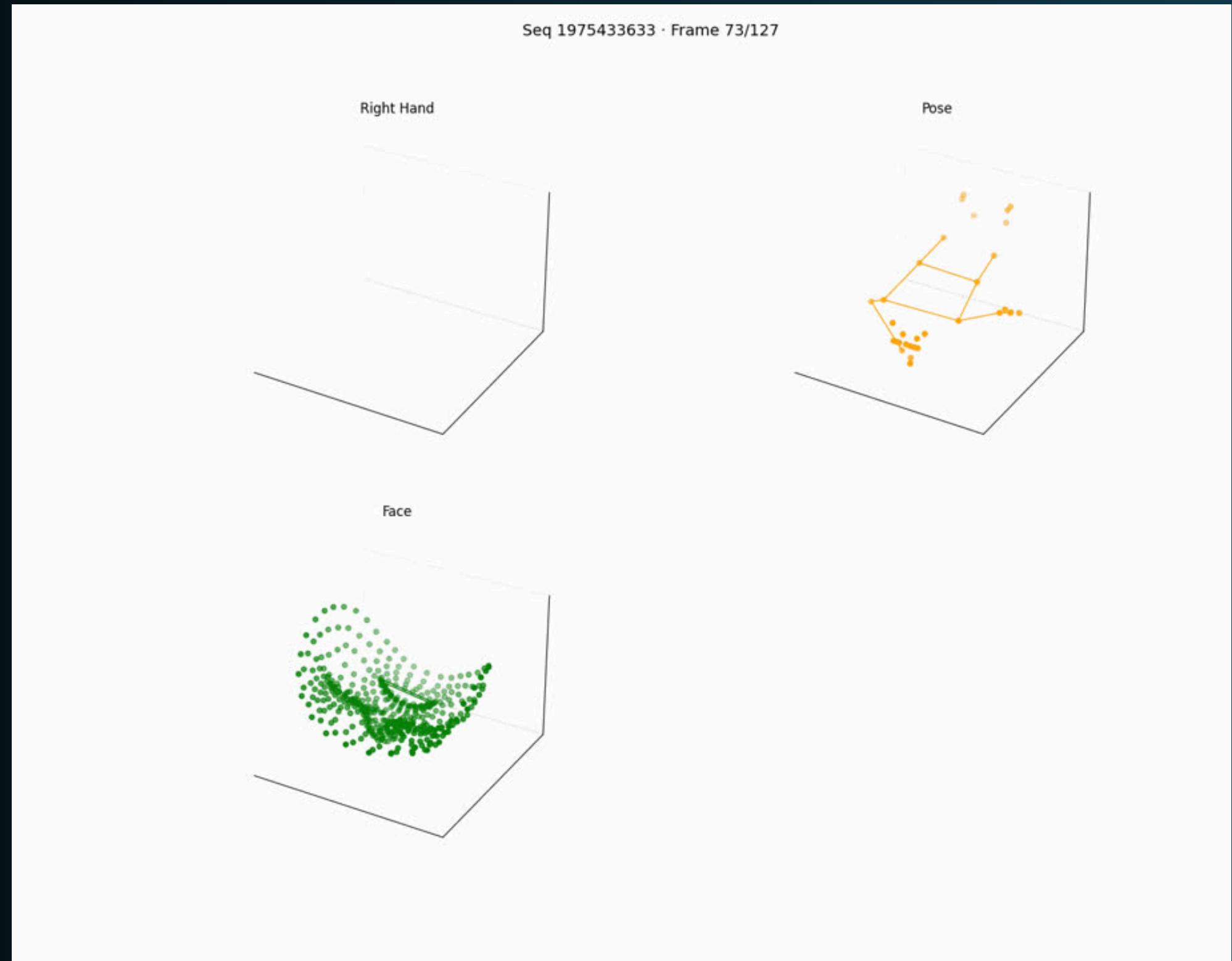


GIFS QUE MUESTRAN LOS MOVIMIENTOS

- Notar que la mano izquierda no tiene tanta importancia.
- Algunos frames tienen valores nulos en la mano porque el movimiento no es continuo.



GIFS QUE MUESTRAN LOS MOVIMIENTOS



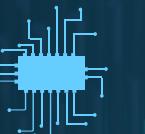
NUESTRAS CONCLUSIONES

VIABILIDAD DEL RECONOCIMIENTO



Las visualizaciones demuestran que los landmarks de manos, rostro y postura capturan adecuadamente las trayectorias del fingerspelling. Esto respalda la viabilidad de un modelo de traducción de ASL a texto.

SELECCIÓN DE MANO



El criterio automático para elegir la mano más visible mejora la consistencia de las secuencias usadas en las animaciones y futuros experimentos.

DIAGNÓSTICO DE CALIDAD DE DATOS



El conteo de NaN y la inspección por secuencia evidencian la presencia de frames incompletos y variabilidad entre participantes.

ESCALABILIDAD Y RETOS



El tamaño del dataset obliga a planificar lectura incremental y preprocesamiento eficiente para no exceder memoria en etapas anteriores.

SUGERENCIAS PARA SIGUIENTES FASES

Utilizar batch processing

Limpieza y normalización

Reducir la dimensionalidad

Métricas y evaluación

Implementar lectura por lotes de .parquet y construcción de tensores por sequence_id con padding/truncamiento, evitando cargar todo en memoria.

Formalizar reglas (mínimo de frames, descarte de filas con Nan críticos) y aplicar normalización/estandarización consistente (por participante o global).

Mantener principalmente landmarks de manos y un pequeño set de referencia (cara/pose) para reducir el número de features.

Definir métricas a nivel de carácter o frase (accuracy) y análisis por participante/condición para detectar sesgos.

REFERENCIAS

Redacción. (2023, 24 septiembre). Sordera: 4 datos que quizás no conocías. BBC News Mundo.

<https://www.bbc.com/mundo/articles/cglelv87zgmo#:~:text=%C2%BFCu%C3%A1ntos%20sordos%20hay%20en%20el,cifra%20superar%C3%A1%20los%20700%20millones>

Google. (s. f.). American Sign Language Fingerspelling Recognition. Kaggle. <https://www.kaggle.com/competitions/asl-fingerspelling>

World Health Organization (WHO). (2025, 26 febrero). Deafness and hearing loss. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>

Population | Documentation. (s. f.). <https://sign.mt/docs/docs/facts/population.html>

Google Research. (2023). ASL Fingerspelling Recognition Competition. Kaggle. <https://www.kaggle.com/competitions/asl-fingerspelling>

Koller, O., Zargaran, S., Ney, H., & Bowden, R. (2019). Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers. *Computer Vision and Image Understanding*, 141, 108–125. <https://doi.org/10.1016/j.cviu.2015.09.013>

Hand landmarks detection guide. (s. f.). Google AI For Developers. https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

MediaPipe Solutions guide. (s. f.). Google AI For Developers. <https://ai.google.dev/edge/mediapipe/solutions/guide>

Zhang, Y., Kim, J., & Shi, B. (2023). Landmark-based sign language recognition with deep temporal models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6), 7254–7268. <https://doi.org/10.1109/TPAMI.2023.3245678>

Jones, T. (2023, 8 mayo). 5 Parameters of ASL: Sign Language 101. Learn Bright. <https://learnbright.org/5-parameters-of-asl-sign-language-101/>

Strickland, J. (2021, 15 abril). How sign language works. HowStuffWorks. <https://people.howstuffworks.com/sign-language.htm>