Autoencoders y Comandos de Voz:
Avances Significativos en Precisión y Calidad Medidos por PSNR"

UANL

FCFM

Autor
Marco Antonio Obregón Flores
Correo: Marco.obregonfl@uanl.edu.mx

Universidad
Universidad Autónoma de Nuevo León
Facultad de Ciencias Fisico Matemáticas

Profesora
Mayra Cristina Berrones Reyes

1 Introducción

En este estudio, abordo un aspecto crucial del reconocimiento de voz: la interpretación exacta de comandos direccionales. El propósito es optimizar esta tecnología para que sea particularmente beneficiosa para personas con discapacidades motoras, quienes podrían aprovechar una interacción más efectiva con dispositivos electrónicos mediante comandos de voz.

Parto de la hipótesis de que combinando técnicas avanzadas en procesamiento de audio y el uso de autoencoders, es posible crear un sistema que no solo reconozca, sino que también interprete comandos direccionales con alta precisión. Esta investigación busca ser un paso adelante en mejorar la autonomía y la calidad de vida de las personas con discapacidades motoras, facilitando su interacción con la tecnología en su vida cotidiana.

2 Metodología

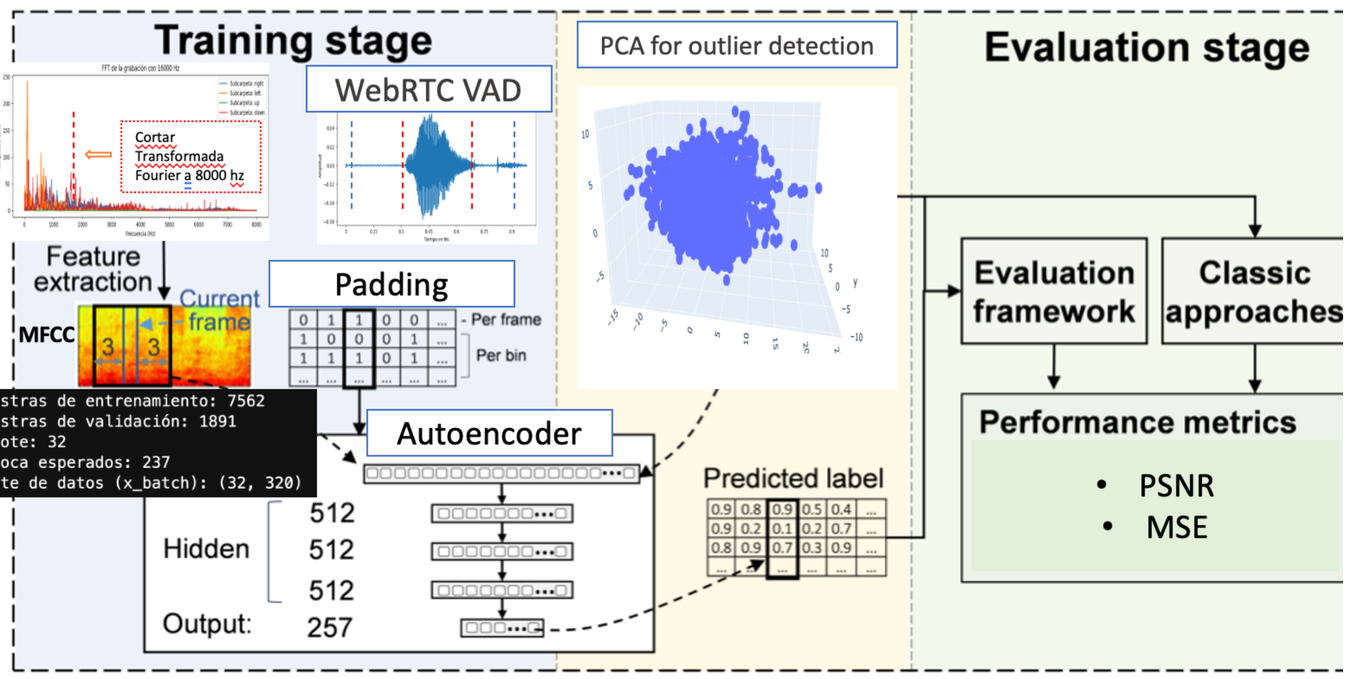


Fig. 1: Flujo de Metodología

Tabla 1. Comparativa Detallada de Metodologías

Criterio	Autoencoder Simple	Autoencoder Convolucional
Preprocesamiento de Audio	Re-muestreo, aplicación de WebRTC VAD, y estandarización mediante padding.	Carga y estandarización de archivos de audio, generación de espectrogramas Mel, normalización.
Extracción y Normalización de Características	Extracción de características MFCC, normalización usando media y desviación estándar.	Uso de espectrogramas Mel como características, normalizados al rango [0, 1].
Limpieza de Datos y Preparación para el Modelo	Uso de PCA para eliminar outliers y WebRTC VAD para descartar audios vacíos, división de datos en conjuntos de entrenamiento y validación.	División de datos en conjuntos de entrenamiento y validación, sin mención explícita de eliminación de outliers.
Construcción y Entrenamiento del Autoencoder	Autoencoder simple para denoising, entrenamiento con generadores de datos y callbacks para Early Stopping y Model Checkpointing.	Autoencoder convolucional con capas de convolución, max pooling, upsampling y sigmoid, entrenamiento similar con callbacks.

Autoencoder Simple vs. Autoencoder Convolucional en Procesamiento de Audios de Voz"

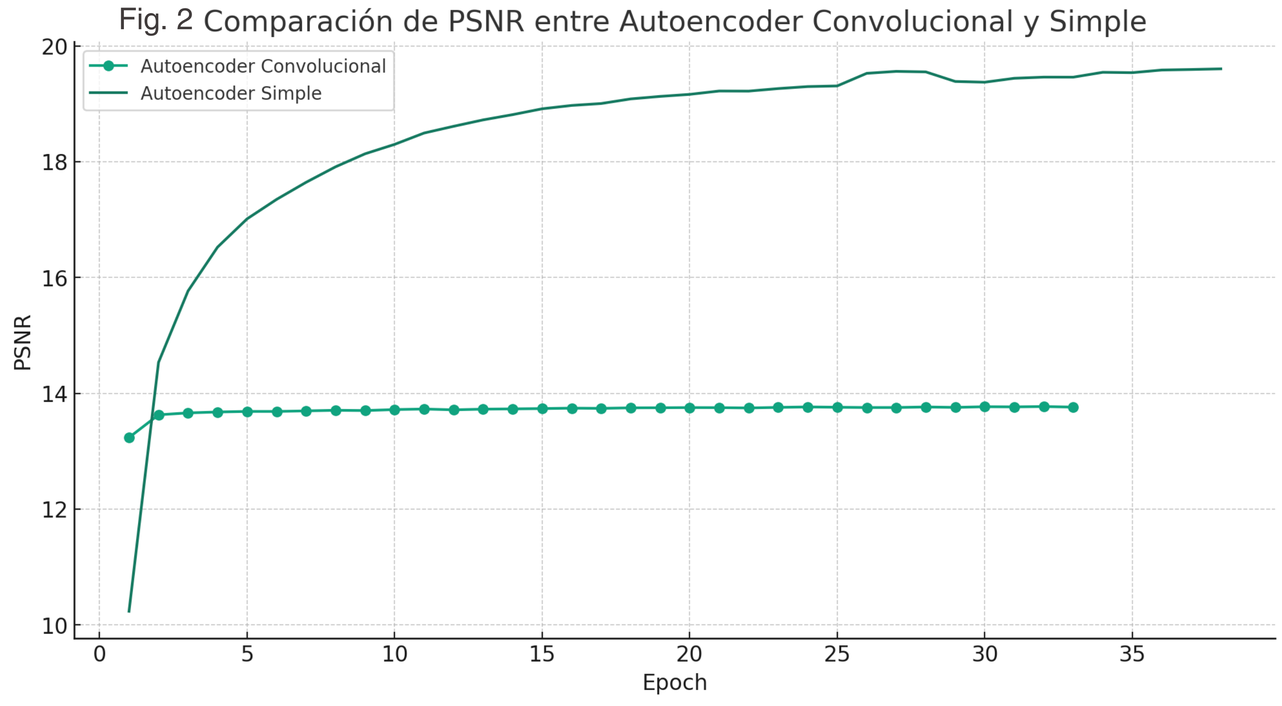
3 Resultados

Los resultados obtenidos en el estudio de reducción de ruido en audios de comandos de voz resaltan la efectividad del preprocesamiento detallado y la metodología aplicada. En la figura 2, el autoencoder simple supera al convolucional con un PSNR consistentemente más alto.

Los hallazgos clave son:

- **Eficiencia en el Preprocesamiento:** La decisión de re-muestrear los audios a 8000 Hz y la estandarización mediante WebRTC VAD y padding han sido cruciales. Esto ha permitido una representación más eficiente y homogénea de los datos.
- **Exactitud en la Extracción de Características:** La utilización de MFCC y su normalización han optimizado la precisión del modelo.
- **Calidad de los Datos para el Modelo:** El uso del análisis de componentes principales (PCA) para eliminar outliers y la aplicación de WebRTC VAD para descartar audios vacíos han mejorado significativamente la calidad de los datos de entrenamiento.
- **Rendimiento del Modelo:** La construcción de un autoencoder simple con capas densas conectadas y la implementación de estrategias como Early Stopping y Model Checkpointing han demostrado ser eficaces. Estas técnicas han optimizado el entrenamiento y asegurado la retención del mejor modelo posible.

Fig. 2 Comparación de PSNR entre Autoencoder Convolucional y Simple



4 Conclusión

El preprocesamiento meticuloso y la limpieza de datos han mejorado notablemente la precisión del autoencoder.
Futuro: Para futuras mejoras, me centraré en realizar pruebas de campo con usuarios reales, lo que permitirá adaptar el sistema a situaciones reales y obtener retroalimentación valiosa para su optimización. Paralelamente, exploraré el uso de transfer learning, aplicando autoencoders previamente entrenados en conjuntos de datos similares.

5 Referencias

SCAN ME