



UANL

Autor

Marco Antonio Obregón Flores

Universidad

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas



FCFM

Mejora de Voz con Autoencoders: Enfoque en Comandos Direccionales

1 Introducción

Este estudio se dirige a un desafío muy concreto dentro del reconocimiento de voz: la interpretación precisa de comandos de voz para controles direccionales. La finalidad es adaptar esta tecnología para ser especialmente útil para personas con discapacidades motoras, quienes podrían beneficiarse de una interacción más efectiva con dispositivos electrónicos mediante comandos de voz.

Mi teoría es que a través de la combinación de técnicas avanzadas de procesamiento de señales y modelos de aprendizaje profundo, se puede desarrollar un sistema que interprete comandos direccionales con buena precisión. Este avance sería una contribución a la calidad de vida de las personas con discapacidades motoras.

2 Objetivos

Principal: Desarrollar un sistema de reconocimiento de voz de alta precisión para interpretar comandos direccionales, destinado a mejorar la interacción de personas con discapacidades motoras con dispositivos electrónicos.

Secundarios:

- Investigar y aplicar técnicas de preprocesamiento y extracción de características avanzadas para optimizar la precisión del sistema.
- Aplicar y validar modelos de aprendizaje profundo para interpretar comandos de voz de manera más efectiva que los enfoques tradicionales.
- Diseñar e implementar estrategias para mitigar el impacto del ruido externo en la precisión del sistema, permitiendo una mayor robustez en entornos no ideales.

3 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"

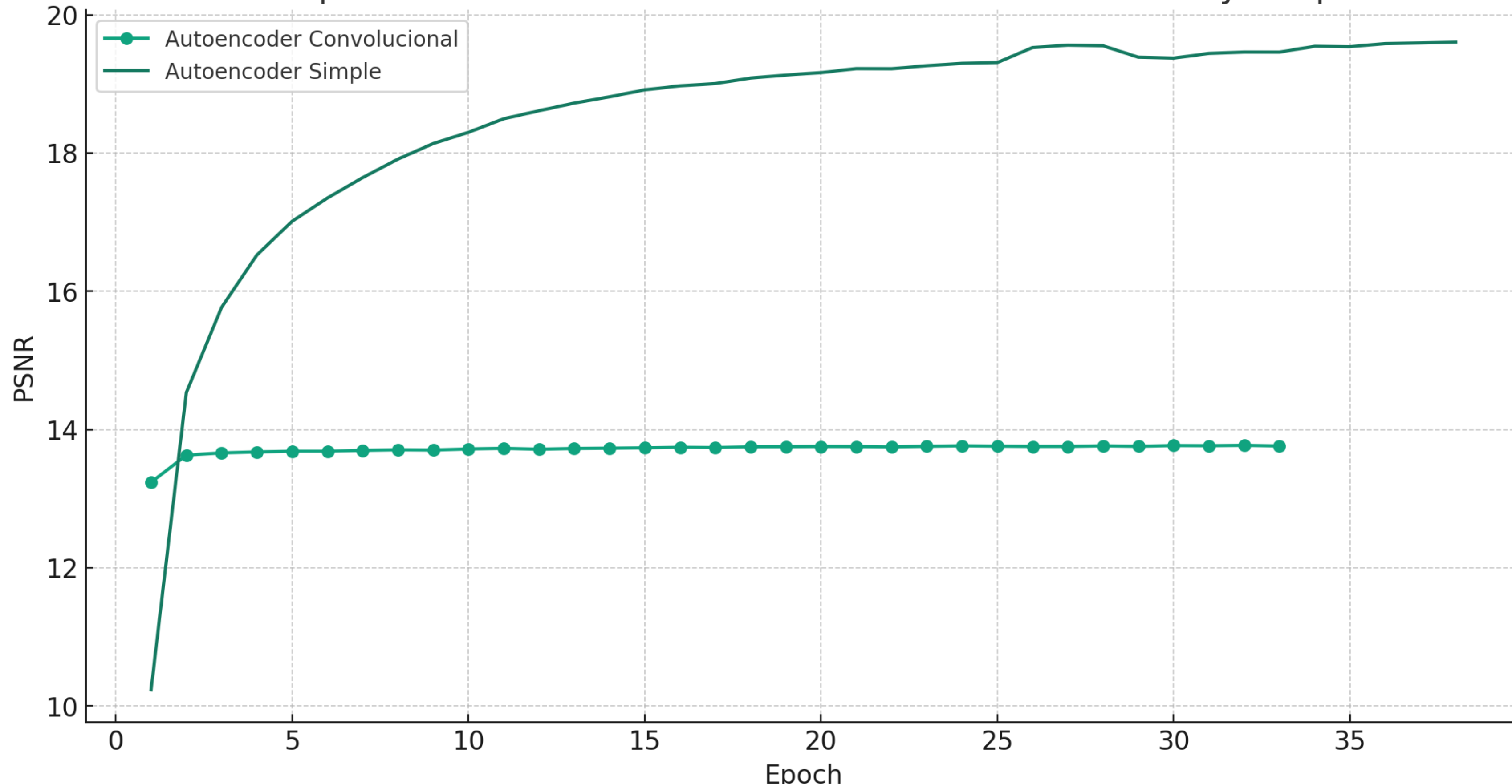
4 Resultados

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 Gráfica 1, el autoencoder simple supera al convolucional con un PSNR consistentemente más alto, reflejando una mejor calidad en la reconstrucción de audio y demostrando su eficacia en la reducción de ruido para el reconocimiento de comandos de voz.

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.

Gráfica 1. Comparación de PSNR entre Autoencoder Convolucional y Simple



5 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.

6 Referencias

Consultar referencias en el código QR

