

Evaluación de diferentes modelos de Redes Neuronales en la Predicción del Posicionamiento del Ventrículo Izquierdo en Ecocardiogramas

Fernando Cerriteño Magaña - A01702790, Leonardo Millán Velázquez - A01639823, and Luis Rodolfo Bojórquez Pineda - A01250513

Instituto de Estudios Superiores de Monterrey, Av. Gral Ramón Corona No 2514, Colonia Nuevo México, 45201 Zapopan, Jal. <https://tec.mx/es>

Abstract. Este documento se enfoca en el desarrollo y análisis de las predicciones realizadas con distintas técnicas en una red neuronal para poder realizar una comparación entre las técnicas.

Keywords: Modelo · U-net · Mascaras · Landmarks · Ventrículo · Ecocardiograma.

1 Introducción

El propósito de este documento es demostrar la implementación de inteligencia artificial en el sector médica, al poder realizar una segmentación correcta para la identificación el ventrículo izquierdo del corazón en un fotograma recuperado de un video desarrollado con ecocardiografía.

Para la realización del proyecto se nos proporcionó con un conjunto de videos de corta duración, los cuales constan de la dilatación y la compresión del corazón, adicionalmente se nos proporcionó con dos archivos en formato .csv, los cuales contenían la segmentación de los videos, la selección de los fotogramas a evaluar, así como un conjunto de coordenadas las cuales delimitaban la posición del ventrículo izquierdo en el fotograma especificado.

El objetivo principal de este trabajo consiste en la evaluación de diversos métodos destinados al análisis de imágenes, así como en la exploración de diversas arquitecturas de redes neuronales. Esto con el propósito de examinar y comparar los resultados obtenidos entre sí, este análisis comparativo pretende mostrar las tendencias generales y patrones observados, contribuyendo así a la comprensión más profunda de la efectividad y aplicabilidad de diversas estrategias en el ámbito del análisis de imágenes mediante redes neuronales.

2 Trasfondo

La ecocardiografía es una prueba diagnóstica la cual genera información sobre la forma, tamaño y fuerza del corazón mediante ondas sonoras de alta frecuencia. Es gracias a esta prueba con la cual se pueden generar los videos de donde

obtuvimos los datos utilizados a lo largo de este documento. El principal uso de la ecocardiografía consiste en visualizar el corazón y las válvulas cardíacas por medio de ultrasonidos a través de una sonda colocada en el esófago, siendo este tipo de estudio el más importante para estudiar y visualizar el funcionamiento de las válvulas y del circulamiento adecuado de la sangre a través del corazón. En este documento nos enfocamos más al ventrículo izquierdo, el cual se encarga del bombeo de la sangre oxigenada a través de la válvula aórtica hacia la aorta. [1]

Los datos mencionados previamente consisten de un grupo de fotogramas obtenidos a partir de un conjunto de coordenadas, el cual fue compartido por parte del socio formador, estas coordenadas representan la posición del ventrículo izquierdo en el fotograma especificado, los fotogramas cuentan con una dimensión de 112 píxeles de alto, por 112 píxeles de ancho.

3 Metodología

En esta sección se describe sobre los diferentes métodos que se utilizaron y una breve descripción del cómo funcionan. Dicha información se obtuvo de la página en internet computer vision [4] y del libro Hands-On Machine Learning with Scikit-Learn and TensorFlow [5]

3.1 UNET

UNET es una arquitectura de red neuronal utilizada para procesar imágenes y segmentar objetos en ellas, lo que significa identificar y separar áreas de la imagen que corresponden a objetos específicos.

La arquitectura UNET se asemeja a la letra "U", Fig. 1. Tiene dos partes principales:

- Codificador: En la parte superior de la "U", el codificador procesa la imagen de entrada. Utiliza capas de convolución para detectar características importantes en la imagen, como bordes y texturas. También reduce la resolución de la imagen, lo que significa que la imagen se vuelve más pequeña a medida que pasa por el codificador.
- Decodificador: En la parte inferior de la "U", el decodificador toma las características extraídas por el codificador y las utiliza para generar una máscara de segmentación. Esta máscara indica qué píxeles de la imagen pertenecen a objetos específicos. El decodificador aumenta la resolución de la máscara para que coincida con la resolución de la imagen original.

UNET toma una imagen, la descompone en características y luego las usa para identificar y delimitar objetos en la imagen. Esta arquitectura se utiliza en aplicaciones como la segmentación de objetos médicos en imágenes de resonancia magnética y en la detección de objetos en imágenes satelitales, entre otras. Su diseño en forma de "U" le permite ser efectiva para tareas de segmentación en imágenes.

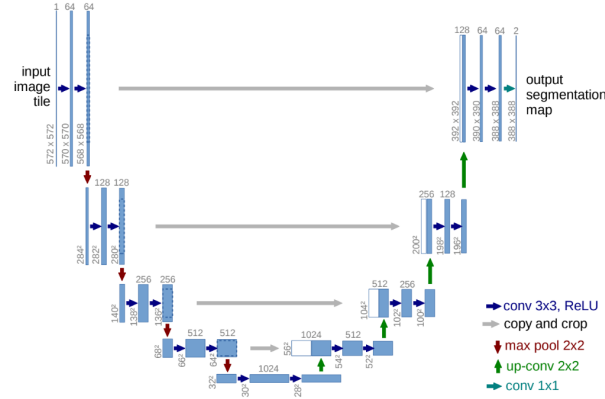


Fig. 1. Arquitectura del modelo U-net.

3.2 Método para la generación de máscaras

La generación de máscaras en la segmentación de imágenes es un proceso común en la visión por computadora. Este proceso implica la identificación y separación de regiones específicas de una imagen. Sus pasos son:

- Preprocesamiento: Las imágenes se preprocesan para mejorar la calidad de la segmentación. Esto puede incluir la eliminación de ruido, el ajuste de contraste y otros métodos para hacer que las regiones de interés sean más distinguibles.
- Segmentación: Se aplican algoritmos de segmentación para dividir la imagen en varias regiones. Los algoritmos pueden ser basados en color, textura, bordes, entre otros.
- Postprocesamiento: Las regiones segmentadas se procesan aún más para eliminar pequeñas regiones no deseadas o para combinar regiones similares.
- Generación de máscaras: Finalmente, se genera una máscara binaria para cada región de interés. En estas máscaras, los píxeles dentro de la región de interés se marcan como 1 (o blanco) y todos los demás píxeles se marcan como 0 (o negro).

3.3 Método para la generación de landmarks

La generación de landmarks en la segmentación de imágenes es un proceso que se utiliza en la visión por computadora para identificar puntos de referencia o características notables en una imagen. Sus pasos son:

- Detección de características: El primer paso en la generación de landmarks es la detección de características. Esto implica identificar puntos de interés en la imagen, como esquinas, bordes o regiones con texturas únicas.

- Extracción de características: Una vez que se han identificado los puntos de interés, se extraen las características de estos puntos. Esto puede implicar la creación de un vector de características que describe la apariencia local de la imagen alrededor de cada punto de interés.
- Matching de características: Después de extraer las características, se realiza un proceso de matching para identificar características similares en diferentes imágenes. Esto puede implicar el uso de técnicas como la correlación cruzada o el matching basado en la distancia.
- Generación de landmarks: Finalmente, los puntos de interés que han sido correctamente matched en diferentes imágenes se convierten en landmarks. Estos landmarks pueden ser utilizados para una variedad de tareas, como la alineación de imágenes, la estimación de la pose 3D, o la construcción de mapas 3D.

4 Experimentación

Todos los siguientes puntos fueron elaborados y probados en archivos ".ipynb", los cuales fueron subidos a la plataforma digital GitHub [2], cabe recalcar que todos los modelos creados fueron entrenados por medio de la GPU en vez de la CPU, esto debido a la efectividad en el tiempo resultante de aprovechar los recursos dedicados en vez de tener que utilizar la unidad de procesamiento central.

4.1 Implementación de la generación de máscaras

Para la correcta implementación de las predicciones en el modelo, se generaron los frames a partir de los archivos .csv que se nos proporcionó, se empezó por leer el archivo "FileList.csv" con el propósito de empezar a determinar dos factores, cuáles eran los frames que se tenían que extraer de cada video, y si el frame obtenido se iba a utilizar para entrenamiento, validación o segmentación. Con todo esta información se realizó un código que de forma automática obtuviera los frames de los videos y los guarde en sus carpetas correspondientes con el nombre compuesto de "frame;número del frame;_nombre del video;.jpg" un ejemplo del nombre de las imágenes sería "frame0_0X1CDE7FECA3A1754B.jpg"

Posteriormente, se realizó una lectura del archivo "VolumeTracings.csv" para poder obtener un conjunto de puntos, los cuales interpuestos en una imagen en negro generaban la forma del ventrículo, no obstante para la primera parte de este documento se necesitaba que las máscaras fueran un polígono en vez de un conjunto de puntos, por lo que haciendo uso de la librería Polygon en Python se crea un polígono como el de la Fig. 2, para mejorar la calidad de la imagen y que la red no tenga problemas en interpretar lo que tiene que predecir, se le aplicó una transformación morfológica con ayuda de la librería cv2, para este caso la transformación fue de clausura, Fig. 3, finalmente para guardar la imagen se utilizó el mismo formato que para guardar el frame excepto que en vez de utilizar la palabra "frame", se utilizó la palabra "mask", un ejemplo del nombre resultante sería "mask0_0X1CDE7FECA3A1754B.jpg", los nombres se

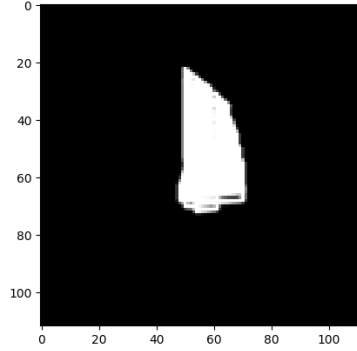


Fig. 2. Máscara sin la transformación aplicada.

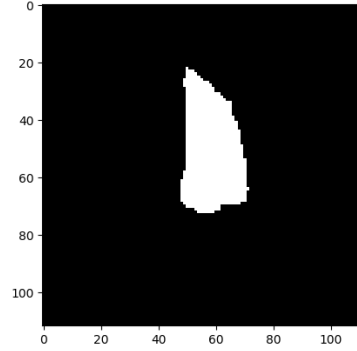


Fig. 3. Máscara con la transformación aplicada.

generaron de esta forma para poder utilizarlos e identificarlos de forma rápida y automática.

4.2 Implementación del modelo U-net

La red neuronal generada toma una imagen de entrada de tamaño 112x112 y produce un mapa de segmentación de la misma forma. La arquitectura permite la identificación de características a diferentes escalas y utiliza conexiones laterales para mejorar la precisión de la segmentación.

Para la realización de la arquitectura, se generó una estructura con 18 capas ocultas, 10 capas convolucionales descendentes y 8 capas transpuestas ascendentes, no obstante, también hay capas de Batch Normalization, ReLU, Max-Pooling, Dropout, y la capa final de salida, por lo tanto, si contamos todas estas capas, la red tiene más capas en total.

4.3 Métodos de generación de landmarks

Para poder realizar el entrenamiento de la red neuronal haciendo uso de landmarks, se tuvo que realizar primero un código que obtuviera dichos landmarks, el código funciona con base en las imágenes de las máscaras generadas previamente. Dichas máscaras son sometidas a un conjunto de operaciones especializadas.

Primero, se utiliza el algoritmo Canny para la detección de bordes, con el cual se puede obtener información sobre la estructura y contornos de las áreas de interés. Posteriormente, se generan mapas de calor, los cuales representan landmarks específicos mediante la colocación de valores distintivos en coordenadas predefinidas. Estos mapas de calor se someten a un proceso de suavizado mediante un filtro gaussiano para lograr una apariencia más natural. Utilizando el contorno obtenido previamente como máscara, se ajustan los heatmaps para destacar únicamente las regiones de interés. Finalmente, las imágenes resultantes,

Fig. 4, cada una representando un landmark, son normalizadas y almacenadas en un nuevo directorio.

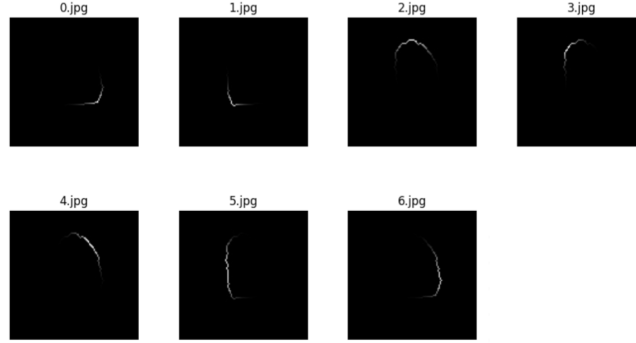


Fig. 4. Imágenes utilizadas para los landmarks

4.4 Modelo U-net con landmarks

Para desarrollar el modelo mencionado, la red neural comienza tomando una imagen de entrada con dimensiones 112x112. Se construye un modelo de segmentación, similar al enfoque de U-net con máscaras, aprovechando esta segmentación para iniciar el entrenamiento específico de la red neuronal para detectar landmarks.

En este proceso, se emplea el algoritmo Canny, al igual que en el caso del modelo entrenado con máscaras. Este algoritmo proporciona información sobre la estructura y contornos de las áreas de interés. Luego, el procedimiento sigue una metodología similar a la aplicada en el caso anterior, facilitando la integración y coherencia en el desarrollo del modelo.

5 Resultados

Para verificar los resultados se evaluó la capacidad de predicción del modelo de dos formas, la primera se realizó haciendo que el modelo prediga la posición del ventrículo en un video completo, la segunda y última forma fue presentándole un video el cual no cuenta con la misma calidad que los videos con los que se entrenó,

5.1 Modelo entrenado con máscaras

Como se había contado previamente se realizaron dos experimentos para verificar los resultados de la red, se evaluó un video entero, el cual cuenta con una duración

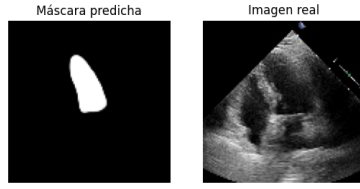


Fig. 5. Predicción con máscaras en el frame 0 del video con buena calidad.

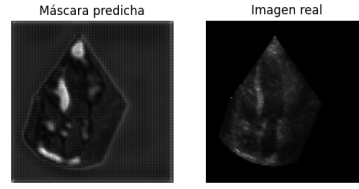


Fig. 6. Predicción con máscaras en el frame 0 del video con mala calidad.

de 2 minutos aproximadamente, para verificar la exactitud del modelo, Fig. 5, evaluando las imágenes generadas se puede determinar que el modelo generado funciona perfectamente, no obstante, cuando se le proporciona con el video el cual cuenta con menor calidad, Fig. 6, se puede observar que no se obtiene un resultado favorable.

Para intentar mejorar la predicción se decidió aplicar un Umbral medio adaptativo [3], esto con el propósito de mejorar el resultado obtenido en la predicción al mejorar un poco más a la imagen, Fig. 7, esto resultó en una mejora significativa; sin embargo, no es una predicción perfecta.



Fig. 7. Predicción con máscaras con el umbral medio adaptativo

Para intentar mejorar la predicción, aunque sea un poco, se creó un nuevo código con el que se entrenó un modelo U-net, pero con la adición de que a algunas de las imágenes de entrenamiento se les aplique un umbral medio adaptativo, esto con el propósito de que la red se acostumbre a predecir imágenes con esas condiciones, la razón por la cual no se entrenó el modelo solamente con el filtro es debido a que cuando se trataba de probar con imágenes normales, el modelo no predecía bien, por lo tanto, se optó por este entrenamiento híbrido entre imágenes con filtro, Fig. 8, e imágenes sin filtro, Fig. 9. Con este nuevo modelo se volvió a realizar la predicción y se obtuvieron los resultados mostrados en la Fig. 10.

Cabe aclarar que este nuevo modelo, no se toma en consideración para la comparación entre el modelo de landmarks y el modelo de máscaras, este nuevo

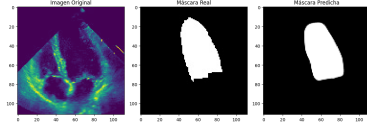


Fig. 8. Predicción sin filtro.

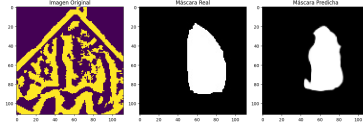


Fig. 9. Predicción con filtro.

modelo solo se realizó para ver si se puede mejorar para situaciones en donde la imagen no cuenta con buena calidad.



Fig. 10. Predicción con el nuevo modelo de máscaras

5.2 Modelo entrenado con Landmarks

Al igual que en los resultados con máscaras, para el modelo entrenado con landmarks se realizaron dos experimentos clave. El primer experimento, siendo que genere las predicciones del mismo video que se le dio al modelo de máscaras, Fig. 11, analizando las predicciones a simple vista, se puede determinar que el modelo de máscaras es mejor, esto es debido a que la predicción es más limpia y se ve más clara con máscaras que con landmarks.

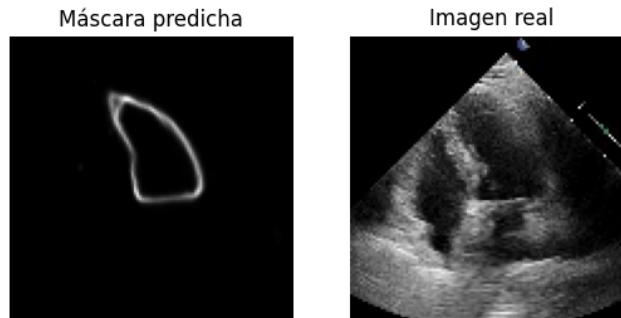


Fig. 11. Predicción con landmarks en el frame 0 del video con buena calidad

No obstante, cuando se realiza la predicción con landmarks en el video que no cuenta con la misma calidad, Fig. 12, podemos apreciar una mejora significativa en comparación a la predicción hecha con máscaras, Fig. 6.

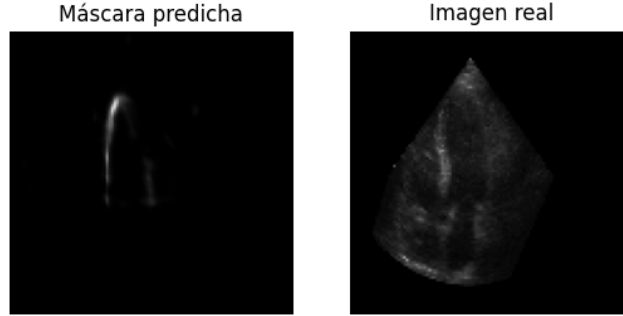


Fig. 12. Predicción con landmarks en el frame 0 del video con mala calidad

Mientras que no tenemos una respuesta clara del porqué es mejor el modelo con landmarks bajo estas condiciones, creemos que se debe a que en el modelo de landmarks se enfoca más en aspectos más relevantes, como lo puede ser el tamaño de la figura, la forma general, entre otros, mientras que, el modelo de máscaras se enfoca más en figuras más completas.

6 Conclusiones

A lo largo del documento, se abordó la implementación y evaluación de dos modelos de redes neuronales para predecir el posicionamiento del ventrículo izquierdo en ecocardiogramas. Se desarrollaron y compararon modelos basados en máscaras y landmarks.

Ambos modelos enfrentaron desafíos al lidiar con videos de menor calidad, indicando la necesidad de investigar estrategias adicionales para mejorar la robustez de los modelos en condiciones adversas. Se sugirió la exploración de técnicas avanzadas de preprocesamiento como la aplicación de aumento de datos.

En cuanto a las conclusiones con base en los resultados, se podría argumentar que el modelo de landmarks es mejor debido a que predice mejor, el video de mala calidad, en comparación con el de máscaras, se tiene que tomar en cuenta que a pesar de que ambos modelos fueron entrenados con la misma arquitectura, las predicciones son más claras cuando se usan máscaras, por lo que creemos que el mejor modelo dependería del uso que se le vaya a dar, debido a que implementar un modelo de landmarks como el realizado, puede resultar muy agotador, y si el uso que se le va a dar es simple, puede resultar mejor implementar un modelo de máscaras.

6.1 Leonardo

En este reto se puso en desafío el conocimiento que se obtuvo, por lo que pude reforzar los conocimientos adquiridos en este bloque, también pude sentir más fluido en trabajar en equipo.

También pude aprender las distintas técnicas hay y los modelos que hay para entrenar modelos.

6.2 Luis Rodolfo

En este desafío, he tenido la oportunidad de sumergirme en un proyecto fascinante que me ha permitido aplicar y fortalecer mis conocimientos en el campo de la inteligencia artificial y el procesamiento de imágenes médicas. Trabajar en equipo en la implementación y evaluación de modelos de redes neuronales para predecir el posicionamiento del ventrículo izquierdo en ecocardiogramas ha sido un proceso enriquecedor.

Durante el desarrollo del proyecto, he aprendido a abordar desafíos prácticos, explorando diversas técnicas para mejorar la calidad de las predicciones. La colaboración con mis compañeros ha sido clave para enfrentar las complejidades y encontrar soluciones efectivas. Además, he tenido la oportunidad de experimentar con diferentes arquitecturas de redes neuronales, lo que ha ampliado mi comprensión en este campo.

Este proyecto no solo ha contribuido a mi desarrollo profesional y académico, sino que también ha proporcionado lecciones valiosas sobre la aplicación práctica de la inteligencia artificial en el ámbito médico. Estoy agradecido por la oportunidad de participar en este proyecto y por el aprendizaje significativo que he obtenido a lo largo de este desafío.

6.3 Fernando

Creo que este reto ha sido de los más interesantes que he tenido que realizar, esto es debido a las diversas actividades que tuvimos que realizar y después ver el que tanto se dificultaron otros equipos o el que tanto nos dificultamos nosotros en algunas partes del proyecto, debido a que contábamos con la libertad de hacer el reto de la forma que nosotros queramos.

En cuanto a los conocimientos que logre desarrollar, creo que puedo destacar el aprendizaje de diferentes técnicas para analizar, mejorar y generar imágenes, así como la importancia de las capas en la arquitectura de la red neuronal.

También puedo destacar que me pareció bastante entretenido y retador el ver los diferentes métodos que se podían hacer por aparte para poder analizar si el modelo de máscaras podía superar al modelo de landmarks.

References

1. Ecocardiografía <https://www.cun.es/>. Last accessed 21/11/2023

2. Github con los códigos elaborados <https://github.com/FernandoCerriteno/Deeplearning-for-ventricle-prediction>. Last accessed 27/11/2023
3. OpenCV Adaptive Mean Thresholding docs.opencv.org. Last accessed 27/11/2023
4. U-Net: Convolutional Networks for Biomedical Image Segmentation <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/>. Last accessed 27/11/2023
5. Aurélien, G: Hands-on machine learning with Scikit-learn and TensorFlow. 2nd edn. O'Reilly, Location (2019)