PROCESAMIENTO Y SEGMENTACIÓN DE PICTOGRAMAS

UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA

INGENIERÍA EN SISTEMAS DE SISTEMAS Y COMPUTACIÓN

Hernan Carreño

2024

***Resumen*—Se desarrolló un sistema de segmentación automática de pictogramas utilizando una arquitectura de red neuronal convolucional U-Net modificada. El modelo fue entrenado para identificar y delimitar los pictogramas en imágenes digitales, generando máscaras binarias precisas. La arquitectura constó de tres partes principales: un encoder que extrae características jerárquicas, un bottleneck que integra información contextual y un decoder que reconstruye las imágenes segmentadas con precisión. Se utilizaron técnicas de normalización, regularización y augmentación de datos para mejorar el rendimiento del modelo. La evaluación del modelo se realizó utilizando el coeficiente de Dice, IoU y métricas de precisión en bordes. El sistema permite una segmentación precisa a nivel de píxel y es robusto ante variaciones de escala y rotación. Este enfoque tiene aplicaciones en la comunicación aumentativa, señalización urbana y digitalización de documentos.**

***Index Terms*—Segmentación automatizada, Pictogramas, Redes neuronales convolucionales, U-Net, Procesamiento de imágenes, Modelos de segmentación, Arquitecturas de redes neuronales.**

I. Introducción

LOS pictogramas, representaciones gráficas de relatos y símbolos creados por diversas culturas prehistóricas en superficies como rocas, cerámica y paredes de cavernas, constituyen una valiosa fuente de información sobre las primeras formas de comunicación humana. Estos símbolos, con un inmenso valor histórico y cultural, reflejan las creencias, prácticas y estructuras sociales de las civilizaciones antiguas, ofreciendo un vínculo visual hacia la comprensión de sus contextos socioculturales. Sin embargo, su estudio y conservación presentan importantes desafíos. La clasificación manual de estos motivos es limitada por factores como la subjetividad interpretativa, la degradación natural de las superficies y las variaciones estilísticas y técnicas en su ejecución a lo largo del tiempo y las regiones.

Figura 1. Fotografía de Pictograma. Fuente: Autores.

El enfoque tradicional de clasificación manual presenta limitaciones significativas debido a la variabilidad y el desgaste de los pictogramas. Para abordar estos problemas, este proyecto utiliza técnicas avanzadas de procesamiento de imágenes y redes neuronales convolucionales (CNN), que permiten no solo el análisis de imágenes bidimensionales, sino también la incorporación de características tridimensionales. Este enfoque mejora la comprensión precisa de los detalles estructurales de los pictogramas y asegura una clasificación consistente y detallada, lo que facilita la comparación entre diferentes sitios arqueológicos y culturas. [2]

La segmentación de imágenes es crucial en el campo de la visión por computadora, especialmente cuando se desea identificar y clasificar objetos o áreas de interés dentro de una imagen. La segmentación semántica, que asigna una etiqueta a cada píxel de la imagen, es particularmente útil para el análisis de pictogramas, ya que permite una clasificación detallada y precisa de los distintos elementos presentes en los mismos. Las CNN, con su capacidad para aprender representaciones jerárquicas y extraer características complejas de las imágenes, se han consolidado como una herramienta poderosa para realizar este tipo de tareas.

En este contexto, se emplea la arquitectura U-Net, un modelo ampliamente utilizado en segmentación de imágenes debido a su capacidad para preservar tanto la información espacial como los detalles finos de las imágenes. La arquitectura U-Net, compuesta por una estructura encoder-decoder, se complementa con conexiones de salto que permiten mantener las características esenciales de los pictogramas, lo que mejora la precisión de la segmentación, especialmente en los bordes y las estructuras complejas. En la Figura 2, se puede observar una representación de la arquitectura U-Net, mostrando cómo se combinan las características de alta resolución con las capas expansivas para optimizar el proceso de segmentación.

La utilización de U-Net permite no solo una segmentación precisa, sino también la mejora en la eficiencia de la clasificación y análisis de los pictogramas, facilitando así su digitalización, procesamiento y conservación.

La organización del trabajo se estructura de la siguiente manera: en la Sección II se presentan los objetivos del estudio; en la Sección III se expone el marco teórico que sustenta el proyecto; la Sección IV describe la metodología CRISP-DM aplicada al proyecto; en la Sección V se detalla la implementación de dicha metodología. A continuación, la Sección VI presenta los resultados obtenidos. Finalmente, las conclusiones se exponen en la Sección VII.

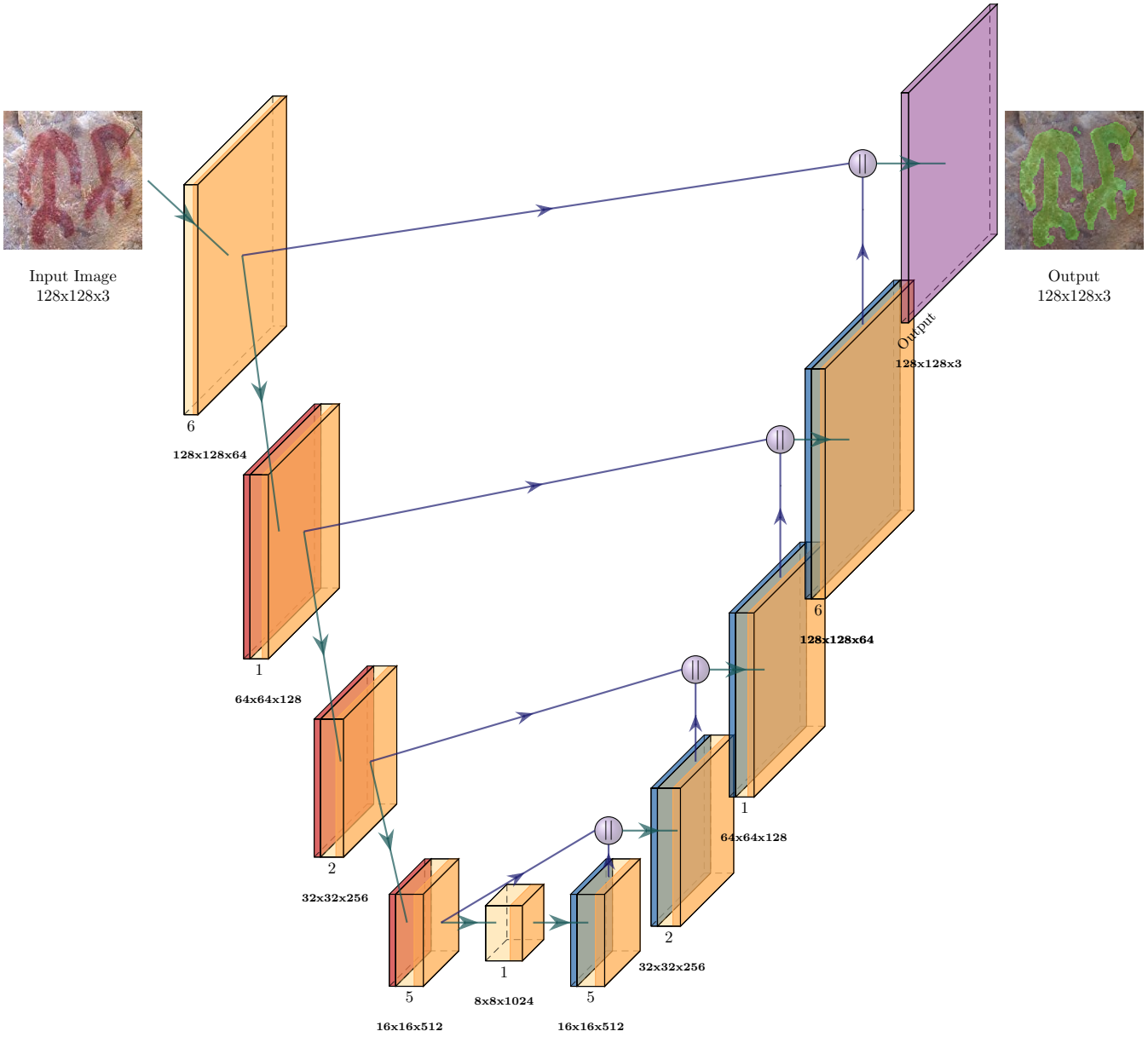

Figura 2. Arquitectura U-Net, adaptado de [1]

II. Marco Teórico

II-A. Conceptos y Tecnologías Relevantes

II-A1. Segmentación de Imágenes: La segmentación de imágenes es una técnica fundamental en la visión por computadora, que permite dividir una imagen en regiones homogéneas. Esta división facilita el análisis de los objetos dentro de una imagen. En la segmentación semántica, se asigna una etiqueta a cada píxel de la imagen, lo que permite identificar objetos específicos, como los pictogramas, dentro de un conjunto visual complejo.

II-A2. Redes Neuronales Convolucionales (CNN): Las redes neuronales convolucionales (CNN) son una clase de redes neuronales profundas que han mostrado gran eficacia en tareas de visión por computadora. Estas redes extraen características jerárquicas de las imágenes mediante capas convolucionales, seguidas de capas de pooling que mejoran la robustez del modelo ante transformaciones como rotación, escala y desplazamientos. [4]

II-A3. U-Net: U-Net es una arquitectura de red neuronal convolucional diseñada especialmente para la segmentación de imágenes. Con una estructura encoder-decoder, U-Net es capaz de aprender representaciones detalladas de las imágenes, y gracias a sus *skip connections*, puede reconstruir imágenes segmentadas con alta precisión, preservando detalles importantes que de otra forma se perderían. [2]

II-A4. Augmentación de Datos: La augmentación de datos es una técnica que permite aumentar la variabilidad del conjunto de datos de entrenamiento mediante la aplicación de transformaciones como rotación, escalado, cambios de brillo y recortes. Esta técnica es especialmente útil cuando se dispone

de un conjunto de datos limitado, como es común en proyectos con imágenes especializadas como los pictogramas.

III. Metodología

En la figura 3 se muestra la metodología usada para el desarrollo del proyecto, pasando desde la comprensión del negocio, la recolección de los datos, el entrenamiento del modelo y su evaluación.

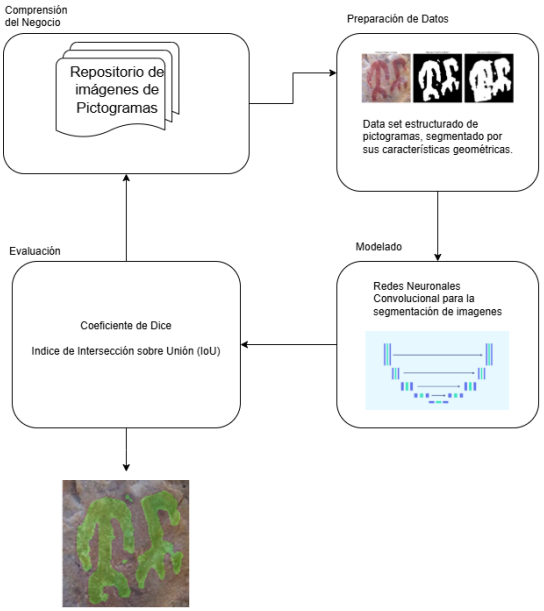

Figura 3. Metodología Crisp-DM aplicada, Fuente: autores.

la metodología empleada en este estudio para la segmentación automatizada de pictogramas. El proceso inicia con la creación de un repositorio de imágenes construido por los autores mediante la captura fotográfica individual de cada pictograma bajo condiciones controladas. Estos datos estructurados, organizados según características geométricas fundamentales, alimentan un sistema basado en redes neuronales convolucionales especializadas en segmentación semántica. El modelo se evalúa cuantitativamente mediante métricas robustas como el coeficiente de Dice y el índice de Intersección sobre Unión (IoU), que permiten validar la precisión en la delimitación de los elementos gráficos.

IV. Desarrollo

El desarrollo del sistema de segmentación de pictogramas se llevó a cabo en varias etapas, que incluyen la preparación del conjunto de datos, la implementación del modelo de red neuronal convolucional (U-Net), la aplicación de técnicas de augmentación de datos y la evaluación del modelo.

IV-A. Preparación del Conjunto de Datos

El primer paso en el desarrollo del sistema consistió en la recopilación y preparación de un conjunto de datos adecuado de pictogramas. Las imágenes fueron obtenidas o recolectadas a partir de los mismos integrantes del equipo en el municipio de Sáchica. Las imágenes fueron acompañadas de sus respectivas máscaras de segmentación, las cuales fueron generadas manualmente para crear las etiquetas de píxel que el modelo usaría para aprender a segmentar los pictogramas.

Se utilizaron técnicas de preprocesamiento para asegurar que las imágenes estuvieran en un formato adecuado para ser utilizadas en el modelo. Estas técnicas incluyeron la normalización de los valores de los píxeles y la adaptación del tamaño de las imágenes para que coincidiera con la entrada esperada por el modelo de U-Net.

IV-B. Implementación del Modelo U-Net

Para la implementación del modelo de segmentación, se utilizó una red neuronal convolucional de arquitectura U-Net. Esta arquitectura es adecuada para tareas de segmentación debido a su estructura de encoder-decoder, que permite aprender representaciones jerárquicas de las imágenes mientras preserva los detalles finos en las etapas de reconstrucción.

El modelo fue implementado utilizando la librería `segmentation-models-pytorch`, que ofrece una implementación de U-Net y otros modelos de segmentación preentrenados. Se cargaron las imágenes de entrenamiento y sus correspondientes máscaras, y el modelo fue configurado para realizar tareas de segmentación semántica.

- **Encoder**: La parte del encoder del modelo fue responsable de extraer las características relevantes de las imágenes de entrada. Se utilizaron capas convolucionales y de pooling para reducir las dimensiones de las imágenes y aprender representaciones jerárquicas de alto nivel.
- **Decoder**: El decoder reconstruyó las imágenes segmentadas utilizando técnicas de upsampling y convoluciones, lo que permitió recuperar la resolución original de las imágenes y generar máscaras precisas.
- **Skip Connections**: Se implementaron conexiones de salto (*skip connections*) entre el encoder y el decoder para preservar los detalles espaciales a medida que la imagen se procesaba y reconstruía.

Estas técnicas ayudaron a mejorar la capacidad de generalización del modelo, permitiéndole manejar diferentes configuraciones de pictogramas en las imágenes.

IV-C. Entrenamiento del Modelo

El modelo U-Net fue entrenado utilizando el conjunto de datos preparado, y se utilizaron las imágenes y sus máscaras de segmentación como entrada y salida, respectivamente. Se emplearon las siguientes configuraciones para el entrenamiento:

- **Métricas de evaluación**: Durante el entrenamiento, se monitorearon las métricas de precisión en los bordes y se calculó el coeficiente de Dice y el índice de intersección sobre unión (IoU) para evaluar la calidad de la segmentación generada por el modelo.

IV-D. Evaluación del Modelo

Para evaluar el rendimiento del modelo, se utilizaron las siguientes métricas de segmentación:

- **Coeficiente de Dice**: El coeficiente de Dice es una métrica que mide la superposición entre las máscaras predicha y real. Un valor cercano a 1 indica una buena segmentación.
- **Índice de Intersección sobre Unión (IoU)**: Mide la relación entre la intersección de las máscaras predicha y real y la unión de ambas. Un valor alto de IoU también indica una segmentación precisa.

Se utilizó un conjunto de datos de prueba independiente para medir estas métricas, y los resultados mostraron que el modelo fue capaz de segmentar los pictogramas con alta precisión a nivel de píxel, incluso con variaciones en la escala y la rotación.

V. Resultados

Para visualizar el rendimiento del modelo, se implementó una función de visualización que mostraba las imágenes originales, sus máscaras reales y las predicciones generadas por el modelo. La comparación visual entre las imágenes y sus segmentaciones permitió una evaluación rápida y efectiva de la calidad de los resultados.

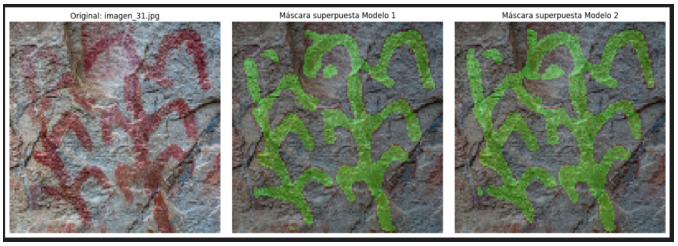

Figura 4. Ejemplo de segmentación de pictogramas

En el cuadro 1 se pueden visualizar las métricas obtenidas en el proceso de entrenamiento. La red muestra un patrón típico de aprendizaje saludable. Tanto la pérdida (Loss) como el coeficiente de Dice mejoran consistentemente a lo largo de las 10 épocas. La mejora más dramática se observa en IoU score, pasando de valores bajos (0.26) a moderados (0.55). Esto sugiere que inicialmente había poco solapamiento entre las predicciones y las máscaras verdaderas, pero la red gradualmente aprendió a localizar mejor las regiones de interés. Finalmente, con un valor final de 0.707, Dice Coefficient indica una segmentación razonablemente buena. Valores superiores a 0.7 generalmente se consideran aceptables para muchas aplicaciones de segmentación médica o de imágenes.

Epoch	Loss	IoU Score	Dice Coefficient	Accuracy
0	0.437961996	0.255975932	0.406542391	0.794532955
1	0.326114148	0.433585525	0.603049219	0.859549463
2	0.304004788	0.46730122	0.634744227	0.869850516
3	0.287798703	0.488970101	0.654318094	0.877284706
4	0.281001389	0.498717517	0.662687182	0.88033551
5	0.273806363	0.50828588	0.671436071	0.883311987
6	0.265291125	0.521205723	0.682442665	0.887355149
7	0.258565247	0.529395998	0.689608753	0.889805019
8	0.250418007	0.54052645	0.698823512	0.892688811
9	0.244434297	0.55045104	0.707037568	0.895370126

Cuadro I
Métricas en cada época del entrenamiento

Así mismo, la figura 5 muestra muestra los resultados de las métricas obtenidas en el proceso de entrenamiento.

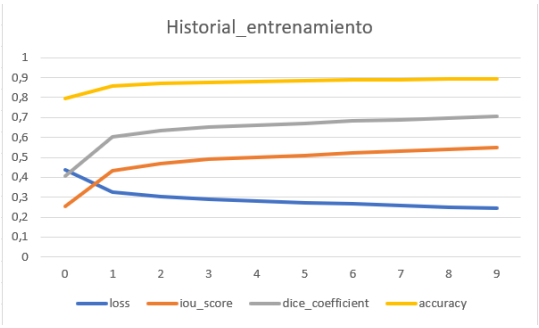

Figura 5. Gráfica Historial de rendimiento.

En la figura 6 se muestra el resultado obtenido del programa desarrollado para usar de una forma simple el modelo.

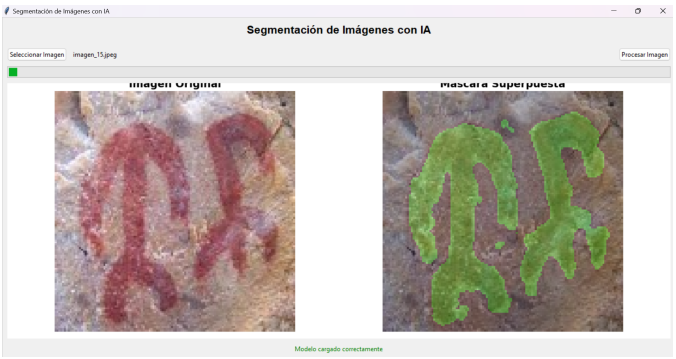

Figura 6. Funcionamiento Interfaz Gráfica segmentación de Pictogramas.

VI. Conclusiones

El sistema desarrollado logró segmentar con precisión los pictogramas en imágenes, superando los desafíos derivados de la variabilidad en las imágenes. La utilización de la arquitectura U-Net y la argumentación de datos fueron claves para mejorar la precisión de la segmentación. Este enfoque puede ser adaptado a otros tipos de símbolos visuales o implementado en aplicaciones prácticas como la digitalización de documentos o sistemas de asistencia.

Referencias

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv preprint arXiv:1505.04597*, 2015. [Online]. Available: http://lmb.informatik.uni-freiburg.de/

[2] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323. [Online]. Available: https://proceedings.mlr.press/v15/glorot11a.html

[3] A. Bhatnagar, M. Sharma, and R. S. Patil, "Pictogram recognition using deep learning methods," *International Journal of Computer Applications*, vol. 178, no. 13, pp. 34–39, 2019. [Online]. Available: https://www.ijcaonline.org/

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9. [Online]. Available: https://ieeexplore.ieee.org/document/7298594