
Past2Polygon: Un enfoque basado en Machine Learning para la digitalización de mapas históricos.

[GitHub](#)

Facultad de Matemática y Computación. Universidad de La Habana.

Ariel González Gómez lenin46ariel@gmail.com	Paula Silva Lara paula.silvalara030410@gmail.com
Leonardo Artiles Montero leo16am@gmail.com	Ricardo Cápiro Colomar rikikpiro02@gmail.com
Alex S. Bas Beovides abasbeovides@gmail.com	Daniel Machado Pérez daniel.machado.0206@gmail.com

30 de enero de 2025

Resumen

En este trabajo se presenta un proceso para la digitalización y modernización de mapas antiguos, con un enfoque en la ciudad de La Habana. El objetivo es convertir estos mapas en formatos digitales modernos para facilitar su análisis y comparación con datos actuales. El proceso incluye la eliminación de ruido y etiquetas textuales mediante técnicas de procesamiento de imágenes y aprendizaje profundo. Luego, se genera un mapa de calor para identificar calles y bloques de edificios, permitiendo una segmentación precisa. Posteriormente, los bloques se vectorizan, simplifican y georeferencian, alineándolos con mapas contemporáneos. Este enfoque permite preservar y estudiar la evolución urbana de La Habana.

Abstract

This work presents a process for the digitization and modernization of historical maps, focusing on the city of Havana. The objective is to convert these maps into modern digital formats to facilitate their analysis and comparison with current data. The process includes noise removal and text label elimination using image processing techniques and deep learning. A heatmap is then generated to identify streets and building blocks, enabling precise segmentation. Subsequently, the blocks are vectorized, simplified, and georeferenced, aligning them with contemporary maps. This approach allows for the preservation and study of Havana's urban evolution.

ÍNDICE

1 Introducción	2
2 Preprocesamiento de Imágenes	2
3 Extracción y Eliminación de Etiquetas (<i>Labels</i>)	4
3.1 Detección de Texto con PP-OCR	4
3.2 Manejo de Imágenes de Mapas Históricos de Gran Tamaño	5
3.3 Eliminación de Texto con Agrupamiento K-Means	5
4 Generación de Heatmaps de Calles	5
4.1 Creación del <i>dataset</i>	5
4.2 Arquitectura del Modelo	6
4.3 Proceso de Entrenamiento	6
4.4 Resultados	6
4.5 Generación del <i>Heatmap</i>	6
4.6 Evaluación	8
5 Segmentación de Imágenes	9
5.1 Metodología	10
5.1.1 Algoritmo de Flood Fill Modificado	10
5.1.2 Clasificación de Componentes	10
5.2 Hiperparámetros	10
5.3 Salidas	10
5.4 Evaluación	12
6 Vectorización de Polígonos	13
6.1 Vectorización del Raster	14
6.2 Simplificación de Polígonos	14
6.3 Corrección de Orientación	14
7 Identificación de Polígonos Atípicos (<i>Outliers</i>)	15
7.1 Métricas utilizadas	15
7.2 Detección de <i>Outliers</i>	15

Palabras Clave: Digitalización, Mapas Históricos, Georreferenciación, Segmentación de Imágenes, Aprendizaje Profundo, Visión por Computadora, PaddleOCR, DBSCAN, Heatmaps, Ramer-Douglas-Peucker, Optimización Bayesiana, Shapely, CNN, Índice de Jaccard

8 Matching de Outliers	16
8.1 Normalización y Preprocesamiento	16
8.2 Comparación Topológica	16
8.3 Cálculo de Centroides	16
8.4 Obtención de Coordenadas Geográficas	16
9 Georreferenciación	17
10 Conclusiones	17
10.1 Limitaciones y Trabajo Futuro	17
11 Anexos	19

1 INTRODUCCIÓN

El estudio de mapas históricos constituye una herramienta fundamental en la investigación sobre la evolución urbana, ya que permite analizar patrones de crecimiento, cambios en la infraestructura y la conservación patrimonial. En el caso de La Habana, una ciudad con una rica historia urbanística, la digitalización de mapas antiguos ofrece una oportunidad única para evaluar su transformación a lo largo del tiempo.

Los enfoques tradicionales de digitalización de mapas han recurrido a la vectorización manual o a técnicas semiautomatizadas para la detección de estructuras urbanas. Sin embargo, estos métodos presentan limitaciones en términos de escalabilidad y precisión. En este trabajo, se propone un enfoque automatizado basado en visión por computadora y aprendizaje profundo para la segmentación, extracción y georreferenciación de mapas antiguos. El proceso comienza con el preprocesamiento de los mapas, eliminando ruido y mejorando la calidad de las imágenes mediante filtros y transformaciones. La eliminación de etiquetas textuales se lleva a cabo con PaddleOCR, con optimización adicional basada en cortes y rotaciones de la imagen para mejorar la detección. Posteriormente, se emplea una red neuronal convolucional para generar un mapa de calor que distingue entre calles y bloques de edificios. Para la segmentación de los elementos urbanos, se utiliza un algoritmo de *Flood Fill*, integrando la información del mapa de calor para mejorar la identificación de bloques urbanos. Estos bloques se convierten en polígonos vectorizados, aplicando simplificación geométrica con los algoritmos de Ramer-Douglas-Peucker (RDP) y Shapely.

Una de las innovaciones clave de este trabajo es la optimización en la identificación de polígonos atípicos mediante DBSCAN, con la selección óptima de hiperparámetros a través de optimización bayesiana. Además, se introduce un algoritmo novedoso para la comparación de polígonos topológicamente similares, basado en la normalización a área unitaria, rotaciones alineadas por

pares de lados trasladados al origen y la evaluación del Índice de Jaccard o Intersection over Union (IoU). Finalmente, los mapas históricos se alinean con mapas actuales mediante georreferenciación, permitiendo su integración en plataformas SIG modernas.

Los datos fueron proporcionados por la Oficina del Historiador de la Ciudad de La Habana, y pueden ser encontrados en su [Repositorio Digital Colección de Planos](#).

19

2 PREPROCESAMIENTO DE IMÁGENES

Antes del procesamiento y análisis de los mapas, fue necesario realizar una etapa de preprocesamiento con el objetivo de mejorar la calidad de los datos y eliminar aquellos elementos que pudieran afectar el desempeño de los algoritmos de visión por computadora.

El preprocesamiento de las imágenes consistió en las siguientes tareas:

- **Recorte de bordes:** Algunos mapas presentaban daños físicos como roturas o pedazos faltantes de papel, lo que generaba ruido en la aplicación de filtros de OpenCV. Se eliminaron estas áreas afectadas para mejorar la uniformidad del dataset.
- **Selección de datos útiles:** Se identificaron mapas en mal estado, con representaciones desproporcionadas o ilegibles debido a su antigüedad. Estos fueron descartados para evitar problemas en el análisis posterior.
- **Mejora de imágenes:** Algunas imágenes fueron editadas utilizando herramientas como Photoshop para optimizar su calidad. Entre las modificaciones realizadas se encuentran ajustes de color, conversión a blanco y negro y eliminación de elementos no cartográficos como carteles o anotaciones manuscritas.

Tras el proceso de filtrado y clasificación, el conjunto de datos quedó organizado de la siguiente manera:

Datos Útiles

- **La Habana:** 30 imágenes de mapas generales de la ciudad. [1](#)
- **Barrios:** 41 imágenes de mapas detallados de distintos barrios. [2](#)
- **Azules:** 50 imágenes de planos específicos de barrios de La Habana. [3](#)

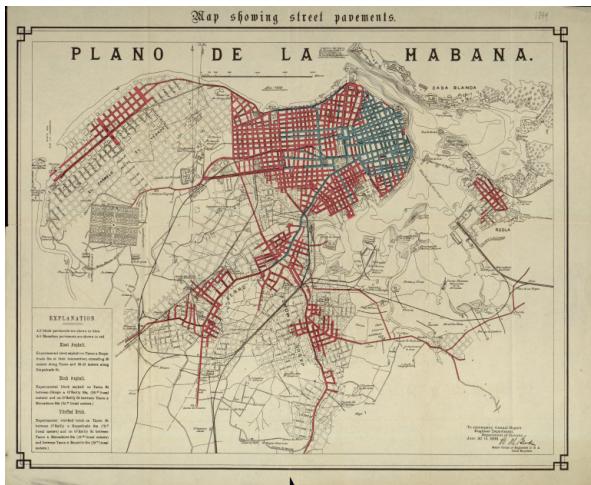


Figure 1: Mapa de La Habana



Figure 3: Plano azul de un barrio de La Habana

Datos Descartados

- **Antiguos:** 19 imágenes de mapas demasiado antiguos con poca utilidad para el análisis. 4
 - **Deficientes:** 90 imágenes en mal estado o con deficiencias graves. 5
 - **3D:** 5 imágenes con representaciones tridimensionales que no eran compatibles con el modelo. 6

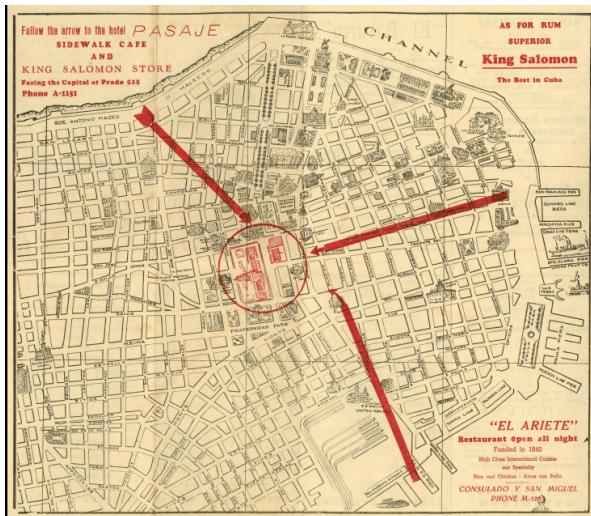


Figure 2: Mapa de un barrio de La Habana

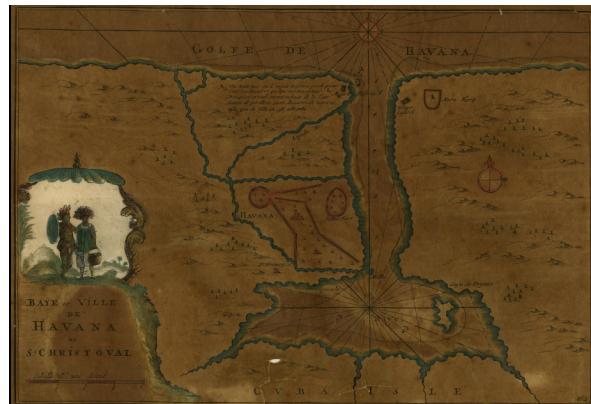


Figure 4: Mapa antiguo de La Habana



Figure 5: Mapa de la Playa Santa María, La Habana
Mapa 3D de La Habana



Figure 6: Mapa 3D de La Habana

En la Figura 7 se muestra un ejemplo de una imagen antes y después de preprocesarla



Figure 7: Mapa antes y después de preprocesar

3 EXTRACCIÓN Y ELIMINACIÓN DE ETIQUETAS (Labels)

Los mapas históricos contienen etiquetas textuales que identifican calles y lugares. Sin embargo, estas etiquetas pueden introducir ruido en los sistemas automatizados de análisis de mapas, ya que los trazos de las letras pueden ser malinterpretados como líneas que representan polígonos. La detección y eliminación de estas etiquetas es un paso crucial en el preprocesamiento para mejorar la interpretación automatizada de mapas históricos.



Figure 8: Imagen preprocesada.

3.1 DETECCIÓN DE TEXTO CON PP-OCR

PP-OCR [2], es un sistema OCR ultraligero diseñado para eficiencia y precisión. Este modelo emplea *Differentiable Binarization* (DB) [4] para una segmentación robusta del texto y está basado en MobileNetV3 [3], lo que optimiza la velocidad de procesamiento y minimiza el consumo de memoria. Su entrenamiento se ha realizado en un conjunto diverso de 97,000 imágenes, incluyendo:

- 68,000 imágenes reales obtenidas de conjuntos de datos públicos y búsquedas en Baidu.
- 29,000 imágenes sintéticas con diversos estilos y orientaciones de texto.

Los conjuntos de datos empleados incluyen:

- **LSVT**: Texto en vistas urbanas a gran escala.
- **RCTW-17**: Texto en documentos escaneados.
- **MTWI 2018**: Imágenes con texto mixto.
- **CASIA-10K**: Texto en documentos oficiales.
- **SROIE**: OCR en recibos y facturas.
- **MLT 2019**: Detección de texto en múltiples idiomas.
- **MSRA-TD500**: Señales de tráfico y carteles.
- **CCPD 2019**: Reconocimiento de matrículas vehiculares.



Figure 9: Imagen con etiquetas detectadas.

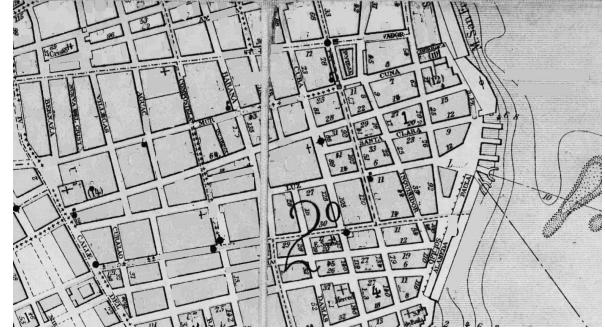


Figure 10: Imagen con etiquetas eliminadas.

3.2 MANEJO DE IMÁGENES DE MAPAS HISTÓRICOS DE GRAN TAMAÑO

Los mapas históricos suelen tener alta resolución, lo que plantea desafíos en la detección de texto con PP-OCR debido a su tamaño de entrada fijo. El redimensionamiento automático de imágenes grandes puede comprometer la resolución y afectar la detección de texto pequeño. Para abordar este problema, la imagen se segmenta en regiones más pequeñas antes del procesamiento, permitiendo detectar el texto con mayor precisión. Posteriormente, los polígonos de texto detectados se mapean a sus coordenadas originales en la imagen completa.

3.3 ELIMINACIÓN DE TEXTO CON AGRUPAMIENTO K-MEANS

Después de detectar el texto, su eliminación implica reemplazar los píxeles correspondientes con el color del fondo circundante. Métodos como la moda o el promedio de color no resultan efectivos debido a la variabilidad cromática del fondo. Por ello, se utiliza el algoritmo de agrupamiento K-Means para identificar el color predominante dentro del polígono de texto. El procedimiento es el siguiente:

- Se agrupan los colores dentro del polígono en 3 *clusters*.
- Se selecciona el *cluster* dominante como color representativo del fondo.
- Los píxeles del texto se reemplazan por este color, integrando visualmente el área editada con el resto de la imagen.

Este método ha demostrado ser efectivo sin requerir ajustes adicionales del hiperparámetro `n_clusters`.

4 GENERACIÓN DE Heatmaps DE CALLES

Se entrenó un modelo basado en redes neuronales convolucionales (CNN) para la detección de calles en mapas antiguos. El objetivo es generar un mapa de calor que permita identificar las regiones correspondientes a calles y diferenciarlas de bloques de casas. Este proceso es un paso fundamental para la posterior selección de los bloques tras la segmentación.

4.1 CREACIÓN DEL *dataset*

Para entrenar el modelo, se construyó un *dataset* a partir de recortes de mapas. Se intentó también generar un *dataset* sintético, pero este enfoque no resultó efectivo, ya que las imágenes generadas no lograban capturar la variabilidad y complejidad de los mapas reales, lo que llevó a un desempeño inferior en la detección de calles. Las imágenes fueron clasificadas en cuatro categorías 11:

- **Intersecciones de calles:** aproximadamente 1400 imágenes.
- **Calles (segmentos entre intersecciones):** 1500 imágenes.
- **Partes del mapa que no son ni calles ni intersecciones:** 1500 imágenes.
- **Recortes aleatorios de los mapas:** cantidad variable.

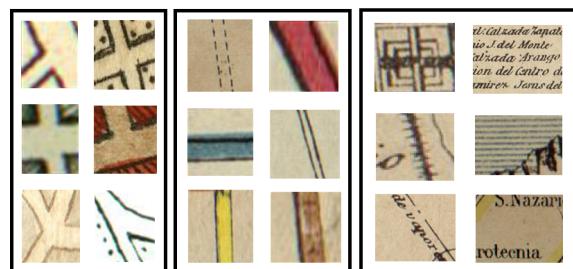


Figure 11: Ejemplos de imágenes del *dataset*.

Para mejorar la capacidad de generalización del modelo, se realizó data augmentation sobre la categoría de intersecciones de calles. En particular, cada imagen se rotó en tres ángulos diferentes (90° , 180° y 270°), cuadruplicando así el número de muestras en esta clase.

4.2 ARQUITECTURA DEL MODELO

En la Figura 12 se muestra un esquema gráfico de la arquitectura de la CNN utilizada para la detección de calles:

El modelo utilizado es una CNN con la siguiente configuración:

- **Capas convolucionales:** Cuatro capas convolucionales con 32, 64, 128 y 256 filtros respectivamente, con kernel de tamaño 3×3 y *padding* de 1.
- **Normalización y activación:** Se emplea *Batch Normalization* después de cada capa convolucional y ReLU como función de activación.
- **Regularización:** Se utiliza una capa de *dropout* con una tasa del 50% para evitar el sobreajuste.
- **Pooling:** Se aplica *max pooling* con un tamaño de ventana 2×2 y *stride* de 2, reduciendo progresivamente la dimensionalidad espacial de las características extraídas.
- **Clasificador:** Dos capas completamente conectadas, una con 512 neuronas y otra de salida con 2 neuronas para la clasificación binaria (calles vs. no calles).

4.3 PROCESO DE ENTRENAMIENTO

Se empleó la función de pérdida de entropía cruzada y el optimizador AdamW con una tasa de aprendizaje ajustable mediante un *scheduler* que reducía la tasa si la pérdida de validación dejaba de mejorar.

La base de datos se dividió en 80% para entrenamiento y 20% para validación. Se utilizó un tamaño de batch de 32 y se entrenó por 10 épocas.

4.4 RESULTADOS

Durante el entrenamiento, se observó un comportamiento inusual en las primeras épocas: la pérdida en el conjunto de validación era menor que la pérdida en el conjunto de entrenamiento, y el *accuracy* de validación también era mayor que el de entrenamiento. Este fenómeno puede deberse a varios factores:

- **Regularización efectiva:** Es posible que el modelo generalizara bien desde el principio debido al uso de *Batch Normalization* y *Dropout*, lo que impide un sobreajuste rápido a los datos de entrenamiento.

- **Efecto del data augmentation:** Como el conjunto de entrenamiento incluye imágenes con transformaciones adicionales, esto puede haber aumentado su complejidad inicial en comparación con las imágenes del conjunto de validación, que no sufrieron las mismas alteraciones.

A medida que avanzaron las épocas, esta diferencia se redujo y el modelo terminó mostrando una tendencia más esperada, con la pérdida de entrenamiento disminuyendo más rápidamente 13 y el *accuracy* de entrenamiento alcanzando valores superiores a los de validación. 14

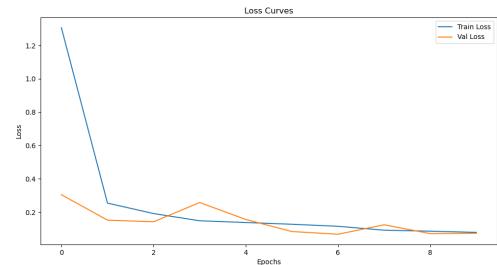


Figure 13: Gráfica de la función de pérdida.

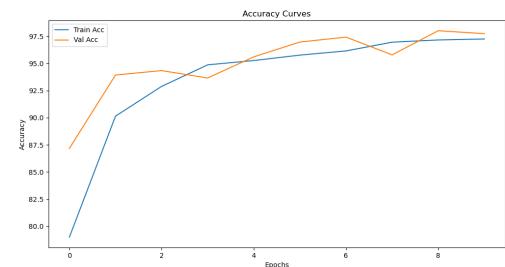


Figure 14: Gráfica de la función de accuracy.

Se probaron diferentes enfoques para el entrenamiento del modelo. El mejor desempeño se obtuvo cuando se entrenó el modelo exclusivamente con la categoría de intersecciones de calles, aplicando data augmentation mediante rotaciones. Este enfoque permitió mejorar significativamente la capacidad del modelo para reconocer patrones en mapas antiguos y distinguir calles de bloques de casas.

4.5 GENERACIÓN DEL Heatmap

Para generar un mapa de calor que permita identificar las zonas correspondientes a calles en los ma-

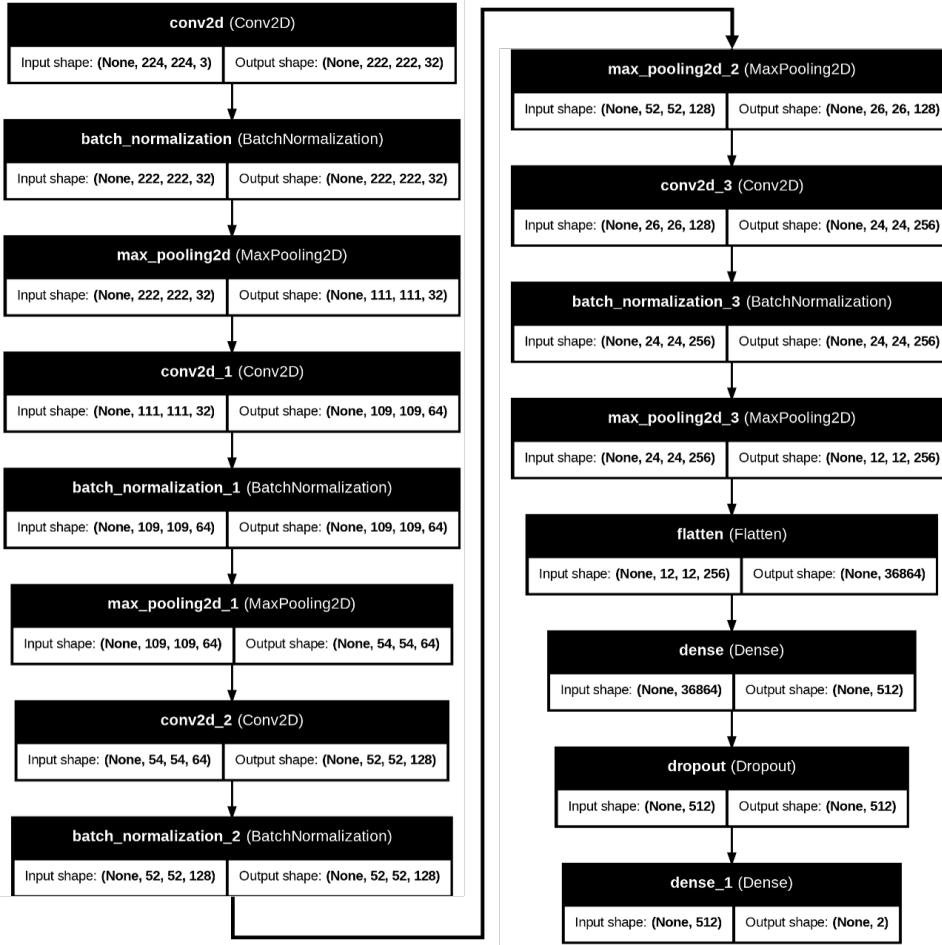


Figure 12: Esquema de la arquitectura de la CNN utilizada.

pas antiguos, se utilizó un enfoque basado en una ventana deslizante sobre la imagen. Se aplicó el modelo entrenado a pequeñas secciones de la imagen original, obteniendo una probabilidad de presencia de calles en cada región. Los resultados se combinaron en un mapa de calor normalizado para representar gráficamente las zonas con mayor probabilidad de ser calles.



Figure 15: Ejemplo de heatmap aprendido.

Además del enfoque basado en la ventana deslizante, se experimentó con técnicas como Grad-CAM y cuadriculado para visualizar la relevancia de las activaciones del modelo. Sin embargo, estos métodos no ofrecieron resultados tan precisos como el enfoque de ventana deslizante, por lo que se optó por este último en la generación final del heatmap.

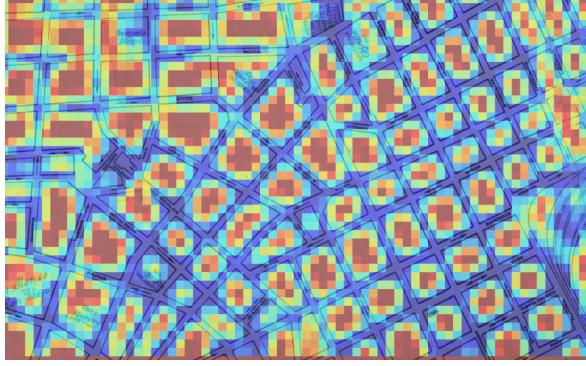


Figure 16: Ejemplo de *heatmap* aprendido.

4.6 EVALUACIÓN

La evaluación de esta parte se llevó a cabo mediante un proceso comparativo que involucró la generación de un conjunto de datos de referencia (*ground truth*, Figura 17) a partir de anotaciones manuales y la evaluación de tres modelos entrenados con diferentes estrategias.



Figure 17: Segmentación anotada manualmente.

Se evaluaron los siguientes tres modelos para la generación del *heatmap*:

- Modelo Manual:** Entrenado con datos etiquetados manualmente, en el que se definieron dos clases: *intersecciones+calles* y *no intersecciones ni calles*.
- Modelo con Intersecciones y Recortes Aleatorios:** Se empleó un conjunto de datos que contenía instancias de intersecciones junto con recortes aleatorios, con el fin de incrementar la diversidad del entrenamiento.
- Modelo Sintético:** Entrenado con datos generados totalmente de forma sintética, que simulaban las características de intersecciones y calles.

Cada uno de estos modelos fue evaluado en seis imágenes seleccionadas, y, adicionalmente, se probó la versión *labelless* (sin etiquetas) para analizar el

comportamiento en ausencia de anotaciones. La comparación se efectuó mediante tres métricas de evaluación no binarias, las cuales se describen a continuación:

- **AUC (Area Under the ROC Curve):** [12] Mide la capacidad del modelo para discriminar entre las dos clases. Un valor de AUC cercano a 1 indica una excelente discriminación, mientras que valores cercanos a 0.5 reflejan un desempeño aleatorio.
- **Precisión Promedio:** [14] Calcula la media de la precisión obtenida en diferentes niveles de *recall*, ofreciendo una medida global del equilibrio entre la exactitud y la capacidad de recuperar instancias relevantes.
- **Brier Score:** [13] Evalúa la exactitud de las probabilidades pronosticadas. Se define como el promedio del error cuadrático entre la probabilidad predicha y la etiqueta real, donde valores menores indican una mejor calibración del modelo.

Las Figuras 18, 19 y 20 presentan la comparación de estas métricas entre los distintos modelos, considerando tanto la versión con extracción de texto como la sin ella. Adicionalmente, las Figuras 21, 22 y 23 muestran la distribución de cada métrica por modelo, permitiendo evaluar la variabilidad y consistencia del desempeño.

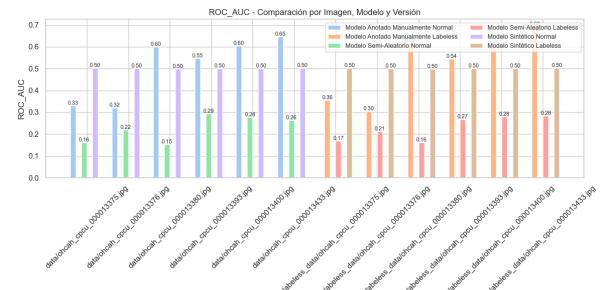


Figure 18: Comparación del *AUC* entre los distintos modelos, sin y con extracción de texto.

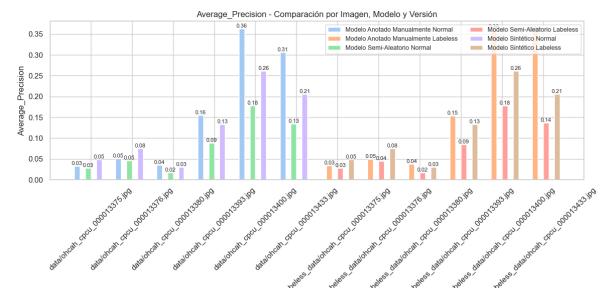


Figure 19: Comparación del *promedio de la precisión* entre los distintos modelos, sin y con extracción de texto.

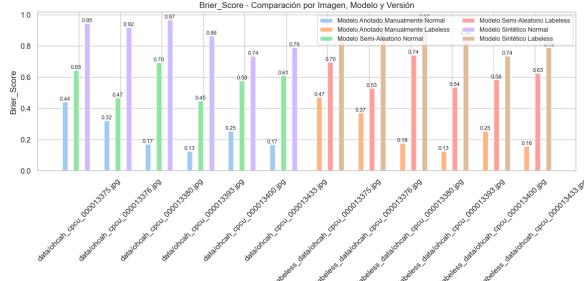


Figure 20: Comparación del *Brier Score* entre los distintos modelos, sin y con extracción de texto.

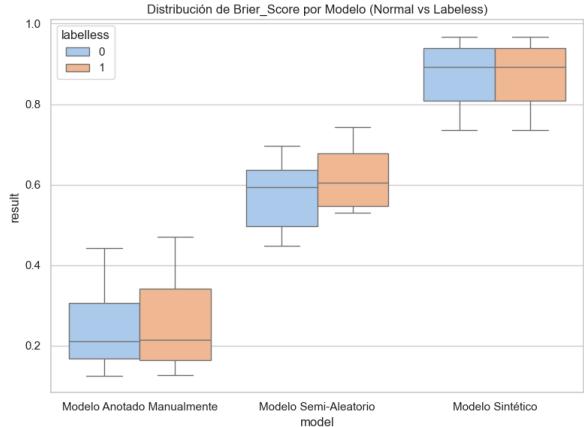


Figure 23: Distribución del *Brier Score* por modelo.

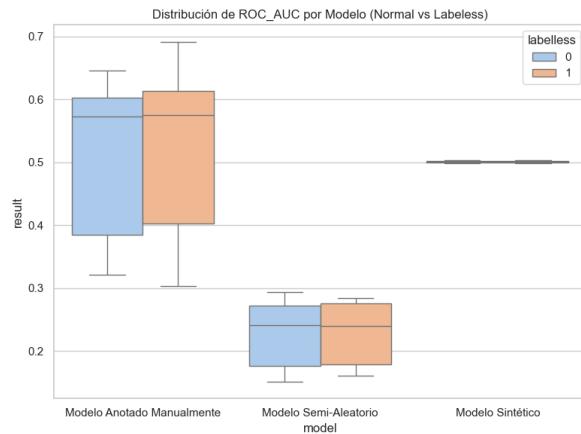


Figure 21: Distribución del *AUC* por modelo.

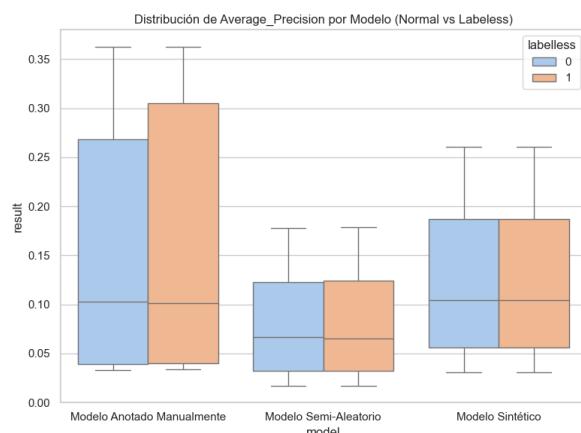


Figure 22: Distribución del *promedio de la precisión* por modelo.

Los resultados indicaron que el modelo entrenado con datos etiquetados manualmente alcanzó un desempeño superior, reflejado en un AUC mayor, un *Brier Score* menor y una precisión promedio más alta, en comparación con el modelo sintético y el modelo basado en intersecciones con recortes aleatorios. La versión labelless, aunque mostró ciertos desafíos en la generación del *heatmap*, proporcionó resultados comparables en términos de consistencia espacial. La matriz de confusión de dicho modelo se muestra en la Figura 24.

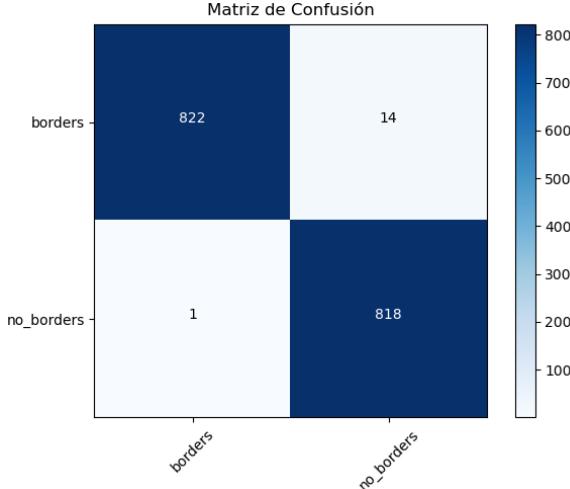


Figure 24: Matriz de Confusión de la CNN entrenada con datos etiquetados manualmente.

Las tablas de los resultados de las métricas aplicadas pueden ser consultadas en los Anexos.

5 SEGMENTACIÓN DE IMÁGENES

En esta fase del proyecto, se aborda la segmentación de las imágenes con el objetivo de identificar y aislar componentes conexas cromáticamente similares.

Este proceso es crucial para la posterior vectorización de bloques de edificios y generación de puntos de control. La segmentación se realiza sobre las imágenes obtenidas de la fase de eliminación de etiquetas, generando como resultado un conjunto de máscaras binarias que representan los contornos de las componentes identificadas, además de información asociada a cada componente.

5.1 METODOLOGÍA

La segmentación se lleva a cabo mediante un algoritmo de *Flood Fill* [7] modificado, similar al funcionamiento de la herramienta "cubo de pintura" en programas de edición de imágenes. Este algoritmo se aplica de forma iterativa a cada píxel no visitado, explorando sus vecinos y agrupando aquellos que comparten características de color similares.

5.1.1 ALGORITMO DE FLOOD FILL MODIFICADO

El algoritmo de *Flood Fill* se adapta para la segmentación considerando la similitud cromática entre píxeles. El proceso se describe como sigue:

1. Se inicia con un píxel no visitado como semilla.
2. Se expande a los píxeles vecinos, incluyendo los píxeles adyacentes por los lados, y opcionalmente por las diagonales, verificando si la diferencia de color con respecto al píxel inicial (o al último píxel visitado) es menor que un umbral (k).
3. Si la diferencia de color es menor que k , el píxel se agrega a la componente actual y se marca como visitado.
4. El proceso se repite iterativamente hasta que no queden píxeles conectados que cumplan con el umbral de diferencia de color.
5. Se obtienen los píxeles que forman la componente segmentada.

La diferencia de color entre dos píxeles se calcula mediante una distancia, la cual puede ser la diferencia absoluta de los valores RGB, o la distancia Euclídea en el espacio de color RGB. Sean $c_1 = (r_1, g_1, b_1)$ y $c_2 = (r_2, g_2, b_2)$ los colores de dos píxeles, la diferencia de color se calcula como:

- Distancia Absoluta:

$$\text{diff}(c_1, c_2) = |r_1 - r_2| + |g_1 - g_2| + |b_1 - b_2|$$

- Distancia Euclídea:

$$\text{diff}(c_1, c_2) = \sqrt{(r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2}$$

Una vez completado el proceso de segmentación de una componente, se calcula el *bounding box* que la engloba, almacenando su ubicación (píxel de la esquina superior izquierda) y dimensiones (ancho y alto).

5.1.2 CLASIFICACIÓN DE COMPONENTES

Tras segmentar las imágenes en componentes, se realiza una clasificación para distinguir entre componentes que representan bloques de edificios y aquellas que no. Esta clasificación se basa en el análisis de un mapa de calor previamente generado en la fase anterior. Para cada componente, se calcula el promedio de las probabilidades de ser parte de un bloque de edificios de todos sus pixeles.

Si el promedio de probabilidades supera un umbral determinado, la componente se considera un bloque de edificios; de lo contrario, se clasifica como un componente no relacionado. Además, se descartan componentes de tamaño inferior a un umbral mínimo, ya que se consideran ruido o texto en la imagen.

5.2 HIPERPARÁMETROS

La precisión de la segmentación depende de la configuración de varios hiperparámetros, los cuales se ajustan de forma empírica:

- k : Umbral de similitud de color. Determina la tolerancia en la diferencia de color para considerar dos píxeles como parte de la misma componente.
- *use8Way*: Booleano que activa o desactiva el uso de los 8 vecinos en la búsqueda de componentes conexas. Si es falso, solo los 4 vecinos ortogonales.
- *euclidif*: Booleano que activa o desactiva el uso de la distancia euclídea en vez de la absoluta para calcular la diferencia de color.
- *adj*: Booleano que indica si la comparación se hace con respecto al ultimo pixel visitado o el inicial.
- *minComponentSize*: Tamaño mínimo (en número de píxeles) para que una componente sea considerada válida.
- *buildingBlockTreshold*: Umbral de probabilidad promedio para clasificar una componente como un bloque de edificios.

5.3 SALIDAS

Como resultado de esta fase, se obtienen los siguientes archivos por cada imagen procesada:

- Máscaras binarias: Imágenes en formato JPG que contienen las máscaras binarias de cada componente identificada, separadas en dos carpetas: una para componentes clasificadas como bloques de edificios y otra para las no relacionadas.

- Imagen de segmentación: Una imagen en formato JPG que visualiza la segmentación, donde cada componente se muestra con un color aleatorio.

- Imagen de componentes de edificios: Una imagen en formato JPG que visualiza las componentes clasificadas como bloques de edificios.

- Archivo de información: Un archivo en formato JSON con información detallada sobre cada componente extraída, incluyendo su bounding box (coordenadas de la esquina superior izquierda, ancho y alto) y la probabilidad de que sea un bloque de edificios.

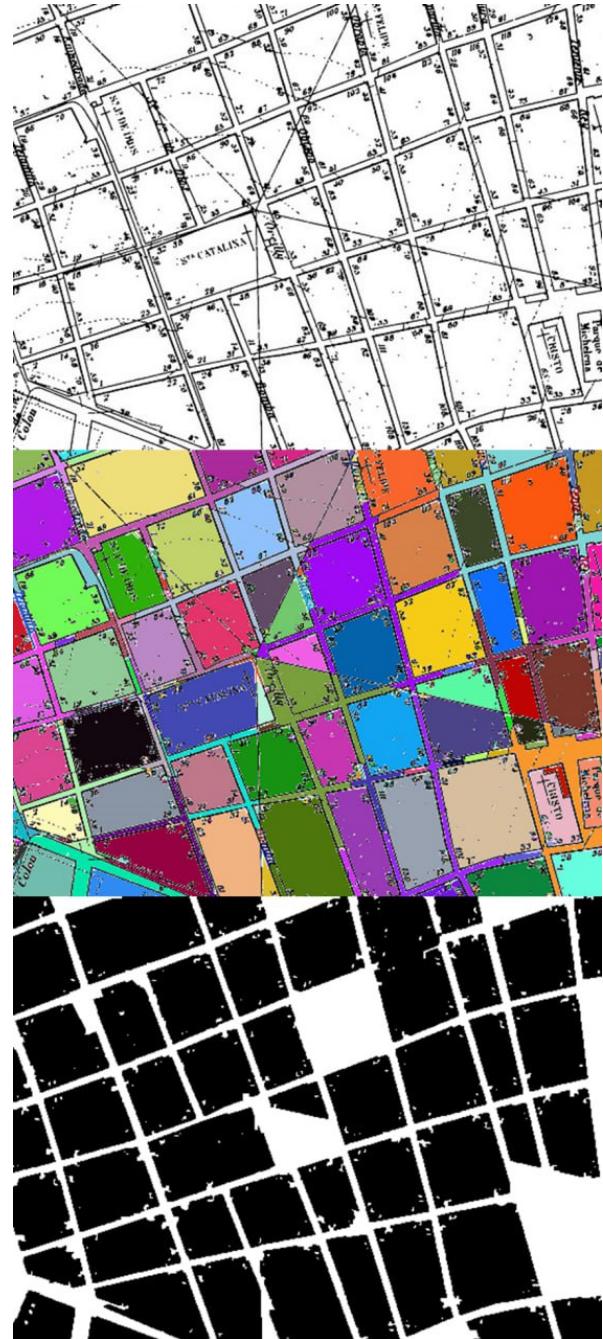


Figure 25: Mapa preprocessado (arriba), Segmentación coloreada (medio), Máscara de bloques de edificios (abajo).

Estos resultados constituyen la entrada para las siguientes etapas del proyecto, en particular, la vectorización de bloques de edificios.



Figure 26: Mapa preprocessado (arriba), Segmentación coloreada (medio), Máscara de bloques de edificios (abajo).

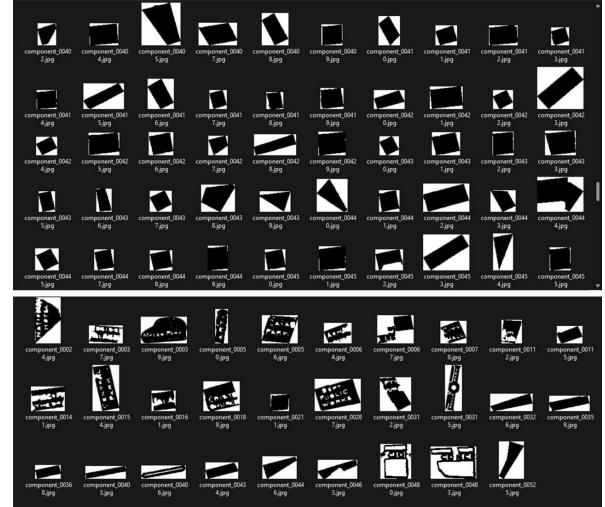


Figure 27: Componentes clasificadas como bloques de edificios (arriba), Resto de componentes (abajo).

5.4 EVALUACIÓN

Como la etapa de segmentación produce un mapa binario que distingue las áreas de interés (por ejemplo, manzanas o bloques) del fondo, se emplearon métricas de clasificación binaria para evaluar su calidad. En concreto, se utilizaron:

- **Dice Score:** [15] Mide la superposición entre el mapa predicho y el mapa de referencia, con valores cercanos a 1 indicando una mayor similitud.
- **IoU (Intersection over Union):** [16] Calcula la proporción de la intersección entre la región segmentada y la región real frente a su unión total. Valores más altos reflejan mayor precisión en la delimitación de la región.
- **Precisión:** Mide cuántos de los píxeles clasificados como positivos (manzanas) son realmente positivos, ofreciendo una perspectiva sobre la exactitud de la detección.

En esta fase se compararon dos configuraciones: una en la que se mantenían las etiquetas o texto dentro de la imagen (*Normal*), y otra en la que se habían eliminado dichos elementos (*Labeless*). Los resultados, ilustrados en las Figuras 28, 29 y 30, muestran que la versión *Labeless* presenta valores superiores de *Dice*, *IoU* y *Precisión*, lo cual indica una mejor segmentación. Esto sugiere que la eliminación de etiquetas textuales contribuyó a reducir el ruido en la imagen y, en consecuencia, a mejorar la detección de las áreas de interés.

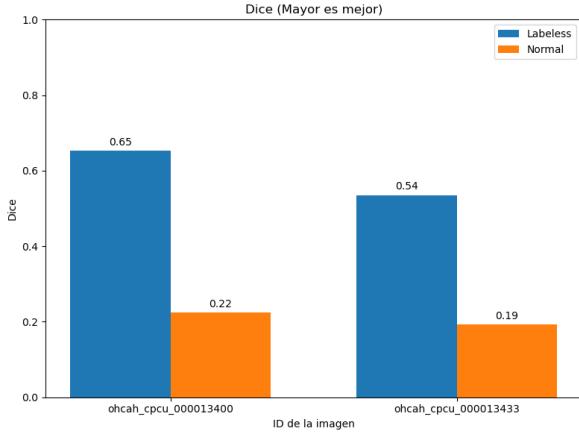


Figure 28: Comparación del *Dice Score* entre imágenes con y sin texto.

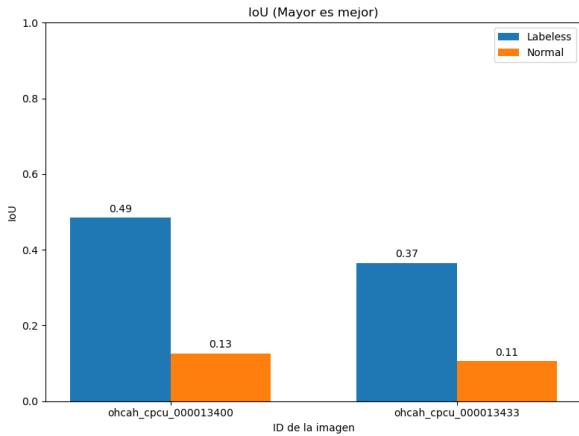


Figure 29: Comparación del *IoU* entre imágenes con y sin texto.

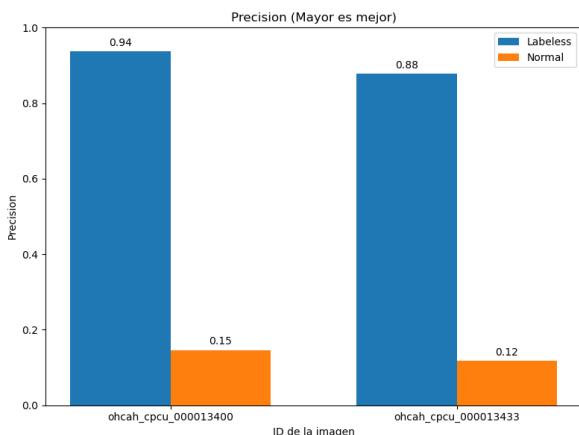


Figure 30: Comparación de la *precision* entre imágenes con y sin texto.

La notable diferencia de desempeño entre ambas configuraciones refuerza la hipótesis de que la presencia de etiquetas o texto en las imágenes puede

interferir en la etapa de segmentación, generando falsos positivos o confundiendo la red. En cambio, al retirar estos elementos, el modelo se centró en la morfología de las manzanas y logró una segmentación más fiel a la estructura urbana real.

6 VECTORIZACIÓN DE POLÍGONOS

La vectorización constituyó un paso esencial en el procesamiento de los datos, permitiendo la conversión de estructuras raster a representaciones geométricas vectoriales. Este proceso comenzó con la recepción de imágenes que contenían los polígonos de interés, junto con la información complementaria almacenada en el archivo `components_info.json`. La combinación de estas fuentes de datos permitió garantizar una correspondencia precisa entre los elementos geoespaciales y sus atributos asociados.

Inicialmente, cada imagen fue procesada para su preparación antes de la vectorización. El procedimiento incluyó la conversión a escala de grises y la binarización mediante la aplicación de un umbral fijo.



Figure 31: Ejemplo de imágenes a vectorizar.

Posteriormente, la imagen binarizada fue almacenada en formato GeoTIFF con la proyección espacial EPSG:4326. Este formato permitió conservar las referencias espaciales y facilitó la integración con herramientas SIG, como QGIS [8], para su visualización y análisis posterior.

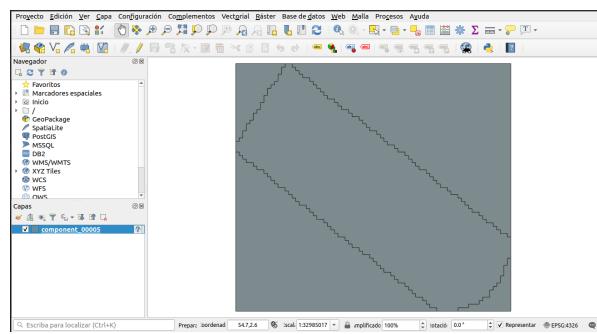


Figure 32: Ejemplo de uso de QGIS.

6.1 VECTORIZACIÓN DEL RASTER

Una vez obtenida la imagen binarizada, se empleó el módulo `gdal_polygonize.py` para la vectorización. Este proceso generó un conjunto de polígonos almacenados en formato Shapefile. En esta etapa, se aplicó un filtrado para seleccionar el polígono principal basado en su *área* como criterio de prioridad. Este filtrado permitió eliminar elementos espurios y garantizar que la información extraída fuera relevante para el análisis.

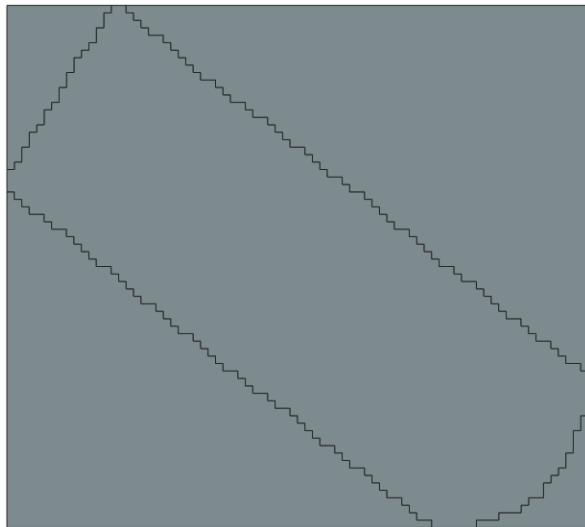


Figure 33: Resultado de vectorizar el polígono.

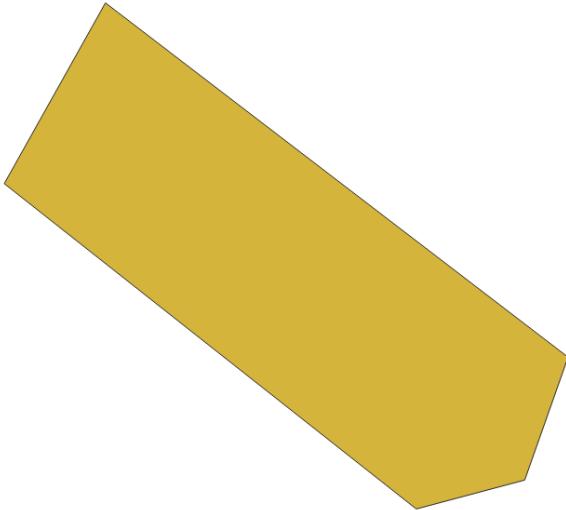


Figure 35: Resultado de simplificar la geometría del polígono.

Se realizaron comparaciones entre ambos enfoques para evaluar su efectividad en la preservación de la estructura original del polígono mientras se minimizaba la cantidad de vértices. Las principales diferencias en cuanto a los resultados se sustentan en lo siguiente:

- **rdp:** Implementa únicamente Douglas-Peucker, lo que lo hace específico para simplificación basada en la distancia.
- **shapely:** Es más general, basado en un algoritmo de topología-preservación y puede ser más preciso en algunos casos.

Una diferencia evidente se puede observar en la Figura 36

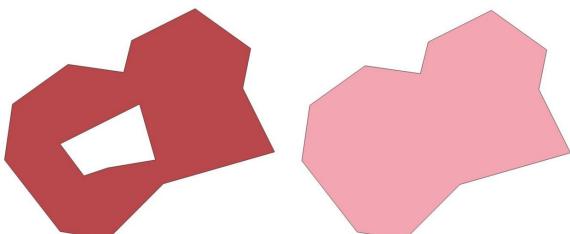


Figure 36: Polígono vectorizado con rdp (izquierda) y con shapely (derecha).

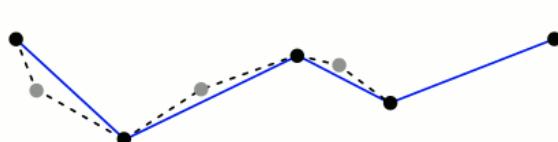


Figure 34: Ejemplo ilustrativo del proceso de simplificación de polígonos con RDP.

6.3 CORRECCIÓN DE ORIENTACIÓN

Dado que el proceso de vectorización podía generar una inversión en la orientación de los polígonos, se aplicó una transformación geométrica (*flip*) para corregir esta situación. Esto aseguró la correcta orientación de los datos conforme a la configuración esperada en el sistema de referencia.

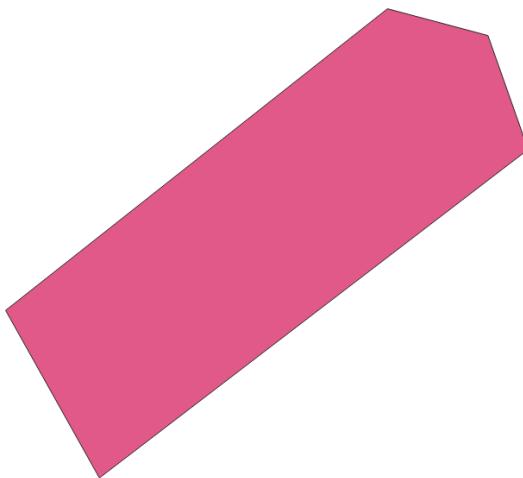


Figure 37: Resultado de invertir verticalmente el polígono.

7 IDENTIFICACIÓN DE POLÍGONOS ATÍPICOS (*Outliers*)

Para detectar polígonos con características significativamente diferentes al resto, se utilizó el algoritmo DBSCAN. Primero, a partir de las coordenadas de los polígonos identificados en las fases anteriores, se calcularon diversas métricas relevantes para el análisis.

7.1 MÉTRICAS UTILIZADAS

- **Número de lados:** Los polígonos con mayor cantidad de lados tienden a ser más atípicos.
- **Área:** Es una característica fundamental para otras métricas.

$$A = \frac{1}{2} \left| \sum_{i=1}^n (x_i y_{i+1} - x_{i+1} y_i) \right|$$

[1]

- **Perímetro:** También es una característica importante para otras métricas.

$$P = \sum_{i=1}^n \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

- **Compacidad:** Evalúa qué tan "compacto" es un polígono. Para un polígono dado, un valor más alto indica que tiene un área grande en relación con su perímetro.

$$\text{Compacidad} = \frac{\text{Area}}{\text{Perimeter}^2}$$

- **Circularidad:** La circularidad compara un polígono con un círculo perfecto, que tiene el valor máximo posible de circularidad (1). Los polígonos más cercanos a un círculo tendrán valores altos, mientras que formas más irregulares (estrechas, alargadas o dentadas) tendrán valores bajos.

$$\text{Circularidad} = \frac{4\pi \text{Area}}{\text{Perimeter}^2}$$

- **Convexidad:** Esta métrica mide qué fracción del área del *convex hull* está ocupada por el polígono. Un valor de convexidad muy cercano a 1, significa que el polígono es muy cercano a ser convexo.
- **Coeficiente de variación:** Esta métrica permite identificar cuánto varían las longitudes de los lados en relación con su media.

$$CV = \frac{\sigma}{\mu}$$

Donde σ representa la desviación estándar y μ la media de los datos.

7.2 DETECCIÓN DE *Outliers*

Después de calcular las métricas seleccionadas como características para cada polígono, se procede a normalizar o estandarizar los datos. Si los valores siguen una distribución normal, se realiza una estandarización (restando la media y dividiendo por la desviación estándar); en caso contrario, se aplica normalización (ajustando los valores a un rango definido, como [0,1]).

Luego, se utiliza la optimización bayesiana con el objetivo de encontrar los mejores hiperparámetros para próximamente aplicar el algoritmo DBSCAN. La optimización bayesiana es un método para encontrar los valores óptimos de hiperparámetros en funciones costosas de evaluar. Se basa en la construcción de un modelo probabilístico (usualmente un proceso gaussiano) para estimar la función objetivo y seleccionar los próximos puntos a evaluar de manera eficiente.

Utiliza una función de adquisición para equilibrar la exploración y explotación, reduciendo la cantidad de evaluaciones necesarias en comparación con métodos de búsqueda exhaustivos. Con esta técnica son hallados los hiperparámetros **eps** y **min_samples** de DBSCAN, maximizando la detección de outliers en los polígonos analizados.

Función objetivo:

$$\arg \min_{\epsilon \in [0.1, 1.0], \text{ min_samples} \in \{2, \dots, 20\}} -\text{Outliers}(\epsilon, \text{min_samples})$$

Posteriormente, se aplica el algoritmo DBSCAN para detectar los *outliers*, los cuales corresponden a

aquellos polígonos considerados atípicos en relación con el resto de la muestra.

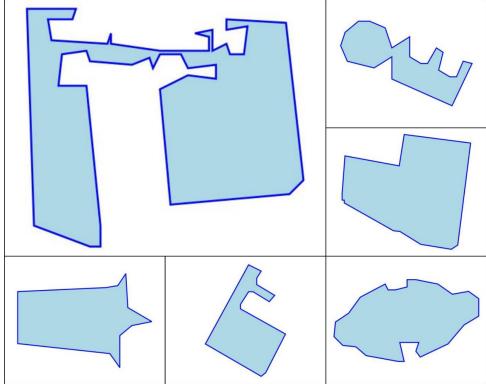


Figure 38: Polígonos *Outliers*.

los polígonos, considerando sus longitudes relativas y ángulos.

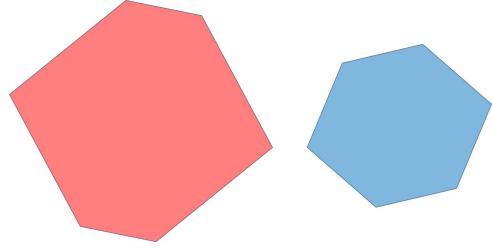


Figure 39: Dos polígonos topológicamente similares a comparar.

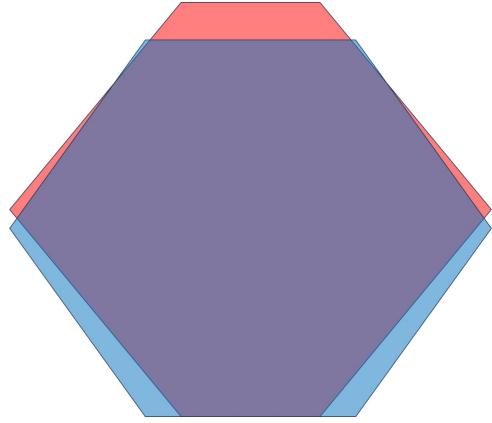


Figure 40: Paso intermedio del algoritmo, luego de normalizar a área unitaria, rotar y trasladar al origen, se muestra el área de intersección.

8 MATCHING DE OUTLIERS

El proceso de matching de outliers constituye una etapa fundamental en la georreferenciación de mapas históricos, permitiendo establecer correspondencias entre estructuras urbanas atípicas identificadas tanto en mapas antiguos como actuales. Este procedimiento se desarrolla mediante la comparación topológica de polígonos previamente clasificados como outliers en ambos conjuntos de datos.

8.1 NORMALIZACIÓN Y PREPROCESAMIENTO

Para garantizar una comparación efectiva entre polígonos de diferentes escalas y orientaciones, se implementa un proceso de normalización. Inicialmente, cada polígono se escala a un área unitaria, eliminando así las diferencias dimensionales. Posteriormente, los polígonos se trasladan al origen de coordenadas, facilitando la comparación de sus características geométricas fundamentales.

8.2 COMPARACIÓN TOPOLOGICA

La determinación de similitud entre polígonos se realiza mediante un análisis exhaustivo que considera múltiples orientaciones. Para cada par de polígonos candidatos, se ejecuta un proceso iterativo de rotación que evalúa la similitud geométrica en diferentes ángulos. La métrica principal de similitud se basa en dos indicadores complementarios:

- Índice de Jaccard (IoU): [11] Cuantifica la superposición entre los polígonos mediante la relación entre el área de intersección y el área de unión.
- Análisis de pares de lados: Evalúa la correspondencia entre los segmentos que conforman

8.3 CÁLCULO DE CENTROIDES

Una vez establecida la correspondencia entre polígonos, se procede al cálculo de sus centroides utilizando el algoritmo del "shoelace" [9] (fórmula del área de Gauss). Para un polígono definido por n vértices (x_i, y_i) , el centroide se calcula mediante las siguientes expresiones:

$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

donde A representa el área del polígono:

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$$

8.4 OBTENCIÓN DE COORDENADAS GEGRÁFICAS

El proceso culmina con la extracción de las coordenadas geográficas correspondientes a los centroides

de los polígonos matching en el mapa actual. Estas coordenadas sirven como puntos de control para el proceso posterior de georreferenciación, permitiendo establecer una transformación espacial precisa entre el mapa histórico y su contraparte moderna.

9 GEORREFERENCIACIÓN

La georreferenciación se basa en el uso de puntos de control previamente identificados para calcular una matriz de transformación. En este caso, se emplea una matriz de homografía, la cual permite convertir coordenadas de píxeles de una imagen a coordenadas geográficas (latitud y longitud). A partir de esta transformación, se determinan los límites geográficos de la imagen, definiendo así su extensión en el mapa de *Open Street Map* [10]. Finalmente, las imágenes georreferenciadas se superponen como capas sobre un mapa interactivo, permitiendo su visualización y análisis.



Figure 41: Mapa superpuesto.

10 CONCLUSIONES

En este trabajo se ha presentado un método automatizado para la digitalización y modernización de mapas históricos, con un enfoque particular en la ciudad de La Habana. Se han integrado diversas técnicas de visión por computadora y aprendizaje profundo para abordar cada fase del procesamiento de mapas, desde la eliminación de etiquetas textuales hasta la georreferenciación final.

Los principales logros de esta investigación incluyen:

- **Preprocesamiento Eficiente:** Se implementaron técnicas avanzadas para mejorar la calidad de las imágenes históricas, reduciendo ruido y eliminando etiquetas textuales mediante el uso de PaddleOCR y agrupamiento con K-Means.

- **Generación de Mapas de Calor con Redes Neuronales Convolucionales:** Se desarrolló un modelo basado en CNNs para detectar calles y bloques urbanos en los mapas históricos, proporcionando información clave para la segmentación de imágenes.
- **Segmentación Precisa con Flood Fill:** Se empleó un algoritmo modificado de Flood Fill para identificar componentes conexas en los mapas, permitiendo una mejor extracción de bloques urbanos mediante el uso del heatmap generado.
- **Vectorización y Simplificación de Polígonos:** Se aplicaron técnicas de vectorización utilizando GDAL y simplificación mediante los algoritmos de Ramer-Douglas-Peucker y Shapely, optimizando la representación geométrica de los bloques urbanos.
- **Identificación de Polígonos Atípicos mediante DBSCAN:** Se implementó un modelo de detección de outliers utilizando DBSCAN, con hiperparámetros optimizados mediante optimización bayesiana, logrando una mejor detección de estructuras urbanas inusuales.
- **Matching de Outliers para Georreferenciación:** Se estableció un procedimiento basado en normalización, análisis topológico y comparación con el índice de Jaccard para asociar estructuras en mapas históricos con sus equivalentes en mapas modernos.
- **Superposición Geoespacial de Mapas:** Finalmente, se realizó la georreferenciación de los mapas históricos mediante homografía y superposición en sistemas de información geográfica, permitiendo su integración con herramientas modernas de análisis geoespacial.

10.1 LIMITACIONES Y TRABAJO FUTURO

A pesar de los avances logrados, el método propuesto presenta algunas limitaciones, entre ellas:

- La calidad de los resultados depende en gran medida del estado de los mapas históricos originales. Mapas con degradación severa o detalles borrosos pueden afectar la precisión del proceso.
- El modelo de clasificación basado en CNN podría beneficiarse de un dataset más extenso y variado para mejorar su capacidad de generalización.
- La georreferenciación se basa en la identificación de outliers urbanos, lo que puede no ser suficiente en algunos casos con mapas muy homogéneos.

Como trabajo futuro, se propone mejorar el modelo de clasificación de calles incorporando técnicas de aprendizaje semi-supervisado y aumentar la precisión de la georreferenciación mediante la integración con datos complementarios, como imágenes satelitales y registros cartográficos adicionales.

REFERENCIAS

- [1] Arteaga Moreno, F. J. (2009). Cálculo del área de un polígono simple: una demostración personal. Lección inaugural, septiembre de 2009.
- [2] Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., ... & Wang, H. (2020). Pp-ocr: A practical ultra lightweight ocr system. arXiv preprint arXiv:2009.09941.
- [3] Koonce, B., & Koonce, B. (2021). MobileNetV3. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization, 125-144.
- [4] Liao, M., Zou, Z., Wan, Z., Yao, C., & Bai, X. (2022). Real-time scene text detection with differentiable binarization and adaptive scale fusion. IEEE transactions on pattern analysis and machine intelligence, 45(1), 919-931.
- [5] Ehret, U., & Neuper, M. (2014, May). Applying the Ramer-Douglas-Peucker algorithm to compress and characterize time-series and spatial fields of precipitation. In EGU General Assembly Conference Abstracts (p. 13537).
- [6] Gillies, S. (2013). The shapely user manual. URL <https://pypi.org/project/Shapely>.
- [7] Shuaeb, S. M., Kamruzzaman, M., & Ali, M. H. (2021). Extracting a bounded region from a map using flood fill algorithm. Asian Journal of Research in Computer Science, 7(1), 14-20.
- [8] Moyroud, N., & Portet, F. (2018). Introduction to QGIS. QGIS and generic tools, 1, 1-17.
- [9] Lee, Y., & Lim, W. (2017). Shoelace formula: Connecting the area of a polygon and the vector cross product. The Mathematics Teacher, 110(8), 631-636.
- [10] Map, O. S. (2017). Open street map. Acessado em, 12.
- [11] Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index. Journal of the American Society for Information Science and Technology, 59(1), 77-85.
- [12] Marzban, C. (2004). The ROC curve and the area under it as performance measures. Weather and Forecasting, 19(6), 1106-1114.
- [13] Rufibach, K. (2010). Use of Brier score to assess binary predictions. Journal of clinical epidemiology, 63(8), 938-939.
- [14] Arora, M., Kanjilal, U., & Varshney, D. (2016). Evaluation of information retrieval: precision and recall. International Journal of Indian Culture and Business Management, 12(2), 224-236.
- [15] Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., & Blaschko, M. B. (2019). Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22 (pp. 92-100). Springer International Publishing.
- [16] Cheng, B., Girshick, R., Dollár, P., Berg, A. C., & Kirillov, A. (2021). Boundary IoU: Improving object-centric image segmentation evaluation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 15334-15342).

11 ANEXOS

Table 1: ROC_AUC (Normal)

model	Modelo Anotado Manualmente	Modelo Semi-Aleatorio	Modelo Sintético
ohcah_cpcu_000013375.jpg	0.3309	0.1623	0.5031
ohcah_cpcu_000013376.jpg	0.3208	0.2189	0.5000
ohcah_cpcu_000013380.jpg	0.5990	0.1516	0.4983
ohcah_cpcu_000013393.jpg	0.5473	0.2940	0.5006
ohcah_cpcu_000013400.jpg	0.6038	0.2762	0.5017
ohcah_cpcu_000013433.jpg	0.6459	0.2624	0.5027

Table 2: ROC_AUC (Labelless)

model	Modelo Anotado Manualmente	Modelo Semi-Aleatorio	Modelo Sintético
ohcah_cpcu_000013375.jpg	0.3559	0.1678	0.5031
ohcah_cpcu_000013376.jpg	0.3038	0.2129	0.4999
ohcah_cpcu_000013380.jpg	0.6156	0.1610	0.4983
ohcah_cpcu_000013393.jpg	0.5434	0.2671	0.5006
ohcah_cpcu_000013400.jpg	0.6072	0.2788	0.5017
ohcah_cpcu_000013433.jpg	0.6914	0.2837	0.5027

Table 3: Average_Precision (Normal)

model	Modelo Anotado Manualmente	Modelo Semi-Aleatorio	Modelo Sintético
ohcah_cpcu_000013375.jpg	0.0330	0.0276	0.0490
ohcah_cpcu_000013376.jpg	0.0506	0.0455	0.0755
ohcah_cpcu_000013380.jpg	0.0355	0.0171	0.0307
ohcah_cpcu_000013393.jpg	0.1552	0.0879	0.1333
ohcah_cpcu_000013400.jpg	0.3625	0.1774	0.2608
ohcah_cpcu_000013433.jpg	0.3056	0.1341	0.2052

Table 4: Average_Precision (Labelless)

model	Modelo Anotado Manualmente	Modelo Semi-Aleatorio	Modelo Sintético
ohcah_cpcu_000013375.jpg	0.0341	0.0276	0.0490
ohcah_cpcu_000013376.jpg	0.0494	0.0449	0.0755
ohcah_cpcu_000013380.jpg	0.0372	0.0172	0.0307
ohcah_cpcu_000013393.jpg	0.1537	0.0852	0.1333
ohcah_cpcu_000013400.jpg	0.3622	0.1784	0.2608
ohcah_cpcu_000013433.jpg	0.3557	0.1367	0.2052

Table 5: Brier_Score (Normal)

model	Modelo Anotado Manualmente	Modelo Semi-Aleatorio	Modelo Sintético
ohcah_cpcu_000013375.jpg	0.4433	0.6465	0.9453
ohcah_cpcu_000013376.jpg	0.3243	0.4682	0.9202
ohcah_cpcu_000013380.jpg	0.1696	0.6956	0.9672
ohcah_cpcu_000013393.jpg	0.1256	0.4480	0.8647
ohcah_cpcu_000013400.jpg	0.2538	0.5782	0.7360
ohcah_cpcu_000013433.jpg	0.1671	0.6098	0.7909

Table 6: Brier_Score (Labelless)

model	Modelo Anotado Manualmente	Modelo Semi-Aleatorio	Modelo Sintético
ohcah_cpcu_000013375.jpg	0.4713	0.6959	0.9453
ohcah_cpcu_000013376.jpg	0.3713	0.5305	0.9202
ohcah_cpcu_000013380.jpg	0.1764	0.7434	0.9672
ohcah_cpcu_000013393.jpg	0.1269	0.5358	0.8647
ohcah_cpcu_000013400.jpg	0.2544	0.5836	0.7360
ohcah_cpcu_000013433.jpg	0.1595	0.6262	0.7909