



Macroproceso: Formación

Proceso: Gestión Administrativa de apoyo a la Formación

Formato: Proyecto de Tesis de maestría investigación

Estudiante: Sara Geraldine Alarcón Prieto

INFORMACIÓN DEL PROYECTO

PROYECTO DE TESIS DE MAESTRÍA

TÍTULO: "Modelo de segmentación semántica basado en Deep Learning para la identificación y vectorización de edificaciones mediante imágenes aéreas y datos LiDAR"

DESCRIPCIÓN DEL PROYECTO

1. Resumen del proyecto (inglés y español)

El presente proyecto desarrollara un modelo de segmentación semántica basado en técnicas de Deep Learning para la identificación y vectorización de edificaciones que presentan interferencia de cobertura vegetal, utilizando de manera combinada imágenes aéreas y datos LiDAR. Esta investigación parte del reto que representa detectar construcciones parcialmente ocultas por vegetación, una limitante común en el análisis urbano y rural a partir de datos ópticos tradicionales. La metodología contempla, en primer lugar, la adaptación de una arquitectura de Deep Learning con reconocido desempeño en tareas de segmentación semántica. Posteriormente, se diseñará un flujo computacional que integre el preprocesamiento de imágenes aéreas y datos LiDAR, el entrenamiento del modelo con datos etiquetados y a partir de los resultados del modelo, se realizará la vectorización de las áreas segmentadas para obtener entidades espaciales georreferenciadas. Finalmente, se evaluará la exactitud comparando el desempeño con y sin la incorporación de datos LiDAR. Se espera que los resultados permitan demostrar que la integración de imágenes aéreas con datos LiDAR mejora significativamente la precisión en la segmentación semántica de edificaciones en entornos con alta interferencia vegetal. El impacto del proyecto interviene en diferentes aspectos, desde un aporte de evidencias y metodologías replicables en el campo de la visión computacional aplicada a datos geoespaciales, la actualización catastral y la planificación territorial hasta

optimizar el uso de tecnologías emergentes para la caracterización de infraestructura física, reduciendo tiempos y costos frente a métodos tradicionales de levantamiento.

2. Introducción (Revisión de Literatura y marco teórico)

2.1. Identificación y vectorización de edificaciones en entornos urbanos

La identificación automatizada de edificaciones en entornos urbanos constituye una herramienta clave para abordar los desafíos asociados con la planificación territorial, la gestión del catastro y el monitoreo del crecimiento urbano. En ciudades que enfrentan procesos acelerados de transformación espacial, contar con información detallada y actualizada sobre las estructuras construidas permite tomar decisiones informadas en temas como movilidad, uso del suelo, gestión de servicios públicos y riesgo urbano. Según Biljecki et al. (2015), el modelado preciso de edificaciones contribuye al desarrollo de ciudades inteligentes, facilita la interoperabilidad de bases de datos urbanas y fortalece la gobernanza espacial. En Colombia, entidades como el IGAC (2020) han enfatizado la necesidad de actualizar el catastro mediante metodologías automatizadas que aseguren cobertura, calidad y oportunidad de la información geoespacial.

2.2 Deep Learning en segmentación semántica de edificaciones

El surgimiento de modelos basados en Deep Learning ha revolucionado la segmentación de imágenes geoespaciales, permitiendo abordar con alta precisión tareas complejas como la identificación de edificaciones. Modelos como U-Net (Ronneberger et al., 2015) y DeepLabV3 han sido adaptados exitosamente para imágenes aéreas y satelitales, logrando buenos niveles de desempeño en la segmentación de techos urbanos. A diferencia de los enfoques tradicionales de clasificación supervisada, estas arquitecturas permiten aprender jerarquías de características espaciales y asignar etiquetas clase por píxel, una ventaja clave en contextos urbanos heterogéneos (Zhu et al., 2017; Li et al., 2020). Las métricas más comunes para evaluar su rendimiento incluyen la intersección sobre la unión (IoU), la precisión, el recall y la F1-score, utilizadas ampliamente en literatura científica y en desafíos internacionales de teledetección.

2.3 Limitaciones de los modelos de Deep Learning en presencia de vegetación

A pesar del éxito de estos modelos, su desempeño tiende a disminuir en entornos con alta interferencia visual provocada por vegetación. En áreas tropicales, la cobertura arbórea puede ocultar parcial o totalmente las cubiertas de edificaciones, dificultando su detección. Este problema se traduce en ambigüedad espectral, ya que el color y textura del follaje pueden confundirse con los materiales constructivos, especialmente en imágenes RGB. Además, la pérdida de bordes definidos y la fragmentación estructural afectan negativamente los resultados de segmentación (Audebert et al., 2018). Estudios como el de Kampffmeyer et al. (2016) demuestran que, en estos escenarios, los modelos tradicionales presentan altos niveles de falsos positivos y negativos, lo cual compromete la confiabilidad de los mapas resultantes.

2.5 Datos LiDAR en modelos de segmentación

El uso de sensores LiDAR (Light Detection and Ranging) ha demostrado ser una alternativa efectiva para mejorar la detección de edificaciones en áreas con cobertura vegetal. Este tipo de sensor permite capturar información tridimensional precisa, útil para distinguir elementos estructurales elevados como techos de vegetación circundante, incluso cuando ambos presentan características espectrales similares. A través de la generación de modelos digitales de superficie (DSM), modelos digitales del terreno (DTM) y modelos de altura del dosel (CHM), es posible incorporar una dimensión altimétrica al análisis espacial (Lin et al., 2019; Qi et al., 2017). Según Chen et al. (2021), la inclusión de datos LiDAR mejora significativamente los resultados de segmentación semántica en entornos complejos, al aportar información estructural adicional que complementa los datos espectrales.

2.6 Modelos multicanal y multientrada: integración RGB + LiDAR

La integración de múltiples fuentes de datos en modelos de Deep Learning ha dado lugar al desarrollo de arquitecturas multicanal o multientrada, donde imágenes RGB y datos LiDAR se procesan de forma conjunta. Esta fusión puede realizarse en etapas tempranas (entrada conjunta al modelo) o en capas profundas (fusión de características), dependiendo del diseño de la red. En el caso de la segmentación de edificaciones, estos enfoques permiten que el modelo aproveche tanto la información espectral como la altimétrica para generar resultados más robustos en presencia de vegetación (Audebert et al., 2018; Chen

multibanda

et al., 2021). Investigaciones recientes en el contexto colombiano (Reyes et al., 2023) han demostrado que los modelos híbridos con entrada multicanal superan en precisión a aquellos entrenados únicamente con imágenes RGB.

2.7 Evaluación del desempeño de modelos en segmentación

La validación del desempeño de los modelos entrenados debe realizarse mediante métricas cuantitativas que permitan evaluar su ~~predicción y aplicabilidad~~ en distintos tipos de cobertura. El uso de IoU, F1-score, precisión y recall permite medir la correspondencia entre las segmentaciones generadas y las referencias de verdad de campo. Además, es fundamental diferenciar el análisis en zonas con y sin interferencia vegetal, con el fin de identificar fortalezas y limitaciones del modelo en escenarios reales de implementación (Rodríguez & Gómez, 2021; Chen et al., 2021). Esta evaluación crítica permite establecer criterios metodológicos aplicables al diseño de soluciones automatizadas en catastro, planificación urbana y monitoreo territorial en Colombia.

Arquitecturas redes neuronales convolucionales:

- U-Net: buen rendimiento con pocos datos
- DeepLab v3: captura contexto multiescala sin perder resolución
- Transformers: permite al modelo enfocarse en distintas partes de la entrada al procesarla.... Porque permiten que el modelo tenga una vista global del contexto de la escena desde las primeras capas, lo que mejora la identificación de estructuras complejas, como edificaciones parcialmente ocultas por vegetación. Modelos como SegFormer y Swin-UNet combinan ventajas de los Transformers y las CNN para mejorar resultados en segmentación precisa.

Métricas cuantitativas:

- IoU
- F1-score
- Precision
- recall

3. Planteamiento del problema

La identificación automatizada de edificaciones a partir de imágenes se ha consolidado como una herramienta fundamental para la planificación urbana, el catastro, y el monitoreo del desarrollo territorial. En este contexto, los métodos basados en Deep Learning especialmente los enfoques de segmentación semántica y detección de objetos han demostrado un desempeño sobresaliente en entornos urbanos densos y homogéneos (Zhu et al., 2017; Li et al., 2020). Sin embargo, su efectividad disminuye significativamente en escenarios donde existe interferencia visual generada por vegetación, como ocurre en ciudades tropicales con cobertura arbórea moderada. La presencia de dicha vegetación en zonas urbanas pueden provocar obstrucción parcial de las cubiertas, ambigüedad espectral y pérdida de bordes claros, lo que afecta la precisión de la captura (Audebert et al., 2018; Kampffmeyer et al., 2016).

En Países como Colombia, donde muchas ciudades intermedias combinan crecimiento urbano irregular con vegetación natural, este problema se vuelve particularmente crítico. Pese a ello, existe una escasez de estudios que evalúen de manera sistemática cómo la interferencia vegetal afecta la exactitud de los modelos de Deep Learning en contextos urbanos tropicales. Esta brecha limita tanto la aplicabilidad práctica de dichos modelos como el diseño de estrategias metodológicas adaptadas a entornos con alta heterogeneidad visual.

Por tanto, se plantea el entrenamiento especializado de un modelo de Deep Learning para la identificación y vectorización de edificaciones en ciudades tropicales colombianas, incorporando como entrada tanto fotografías aéreas como datos LiDAR. Esta integración busca enriquecer la representación espacial y estructural de las edificaciones, especialmente en entornos urbanos con alta interferencia visual causada por cobertura vegetal.

4. Justificación

La creciente necesidad de información espacial precisa ha llevado a adoptar métodos automatizados para la detección y vectorización de edificaciones en entornos urbanos con aplicaciones en diferentes campos (Biljecki et al., 2015). Los modelos de Deep Learning, especialmente aquellos basados en segmentación semántica como U-Net (Ronneberger et al., 2015), FCN (Long et al., 2015), DeepLabv3+ (Chen et al., 2018) y por ultimo la idea de reemplazar

completamente las redes convolucionales es el transformer (Vaswani et al., 2017) han demostrado un desempeño notable en ciudades con estructuras compactas, bordes definidos y baja interferencia visual (Zhu et al., 2017; Li et al., 2020).

Sin embargo, su eficacia se reduce considerablemente en entornos urbanos tropicales donde la vegetación interfiere visualmente con las cubiertas edificadas. Este fenómeno genera ambigüedad espectral, pérdida de continuidad estructural y falsos positivos o negativos en la segmentación (Audebert et al., 2018; Kampffmeyer et al., 2016). En estos escenarios, los árboles pueden ocultar techos parcial o totalmente, dificultando la identificación precisa incluso en imágenes de alta resolución (Chen et al., 2021). En Colombia, esta situación es crítica debido a la presencia de ciudades intermedias con abundante vegetación natural integrada al entorno urbano (DANE, 2022; IDEAM, 2020).

A pesar de la disponibilidad de datos satelitales y modelos preentrenados, la literatura evidencia una escasez de estudios que evalúen de manera sistemática el efecto de la cobertura vegetal sobre la precisión de modelos de segmentación y detección en contextos urbanos tropicales (Reyes et al., 2023; Rodríguez & Gómez, 2021). Esta brecha limita la aplicabilidad de estas metodologías en proyectos de catastro multipropósito o monitoreo territorial en zonas con alta heterogeneidad visual.

En este contexto, la integración de datos altimétricos provenientes de sensores LiDAR representa una oportunidad significativa para enriquecer la entrada de los modelos, al proporcionar información tridimensional que permite diferenciar estructuras verticales como edificaciones y vegetación (Lin et al., 2019; Qi et al., 2017). Al combinar imágenes aéreas ortorrectificadas con datos LiDAR procesados como el modelo digital de elevación y el de terreno, es posible entrenar modelos de segmentación más robustos y adaptados al contexto colombiano. Este proyecto propone el entrenamiento de un modelo basado en Deep Learning con entrada multicanal (Imágenes aéreas + LiDAR), con el objetivo de mejorar la identificación y vectorización de edificaciones bajo condiciones de interferencia vegetal.

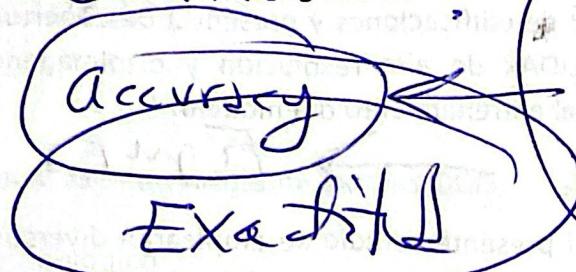
5. Pregunta de Investigación

¿En qué medida la combinación de datos LiDAR y fotografías aéreas incrementa la precisión de un modelo de Deep Learning para la segmentación semántica de edificaciones parcialmente cubiertas por vegetación?

exactitud

6. Objetivos

General:



- Precision

- Recall

- F1-Score

Desarrollar un modelo de segmentación semántica basado en Deep Learning para la identificación y vectorización de edificaciones con interferencia vegetal mediante el uso combinado de imágenes aéreas y datos LiDAR.

Específicos:

01. Adaptar una arquitectura Deep Learning con buen desempeño en segmentación semántica para el análisis de imágenes aéreas y datos LiDAR.
02. Desarrollar un flujo computacional para el procesamiento de datos y desarrollo de un modelo de segmentación semántica de edificaciones.
03. Vectorizar las edificaciones segmentadas a partir de los resultados del modelo y generar entidades georreferenciadas.
04. Evaluar el desempeño de la segmentación semántica generada por el modelo.

7. Materiales y Métodos

7.1. Enfoque metodológico

Esta investigación adopta un enfoque cuantitativo aplicado, de tipo experimental y exploratorio-comparativo, orientado a evaluar la eficacia de modelos de segmentación semántica basados en Deep Learning (DL) para la identificación de edificaciones parcialmente cubiertas por vegetación, utilizando datos combinados de imágenes aéreas y datos LiDAR. La hipótesis central plantea que la integración de datos LiDAR mejora significativamente la precisión de los modelos DL frente a escenarios de interferencia vegetal.

7.2 Área de estudio

La zona de estudio corresponde a la ciudad de Pereira y Valledupar, seleccionada por su alta densidad de edificaciones y presencia de cobertura vegetal. Esta zona cuenta con datos LiDAR de alta resolución y ortoimágenes aéreas recientes, indispensables para el entrenamiento del modelo.

7.3 Materiales y datos

Para el desarrollo del presente estudio se emplearan diversos materiales y datos esenciales para la implementación del modelo de segmentación semántica. Los datos utilizados incluyen ortoimágenes las cuales permiten identificar con alta precisión los detalles constructivos del área de estudio. Asimismo, se utilizaran nubes de puntos LiDAR, las cuales proporcionan información altimétrica detallada y tridimensional, útil para diferenciar edificaciones del entorno natural. Como base para la validación, se integraran capas catastrales o planos de referencia existentes, que permitirán evaluar la exactitud de la segmentación generada. El procesamiento y análisis se llevaron a cabo utilizando software especializado, como QGIS para el manejo SIG, Global Mapper para la manipulación de datos LiDAR, y entornos de desarrollo como Python 3.11 con bibliotecas de aprendizaje profundo (TensorFlow o PyTorch). Las actividades de programación y experimentación se realizaran en Jupyter Notebooks y Visual Studio Code. En cuanto al hardware, se contara con una estación de trabajo personal equipada con GPU dedicada para acelerar el entrenamiento de redes neuronales, así como servicios de almacenamiento en la nube para la gestión y respaldo de los datos. Adicionalmente, se realizara un proceso de etiquetado manual asistido para generar las máscaras de entrenamiento, utilizando plugins específicos de QGIS.

• Datos:

- Ortoimágenes RGB con resolución ≤ 10 cm.
- Nubes de puntos LiDAR.
- Capas catastrales o planos de referencia para validación.

Software:

- QGIS, LAStools, CloudCompare
- Python 3.11, TensorFlow o PyTorch
- Entornos Jupyter y VSCode

- **Hardware:**

- Estación de trabajo personal con GPU dedicada.
- Almacenamiento en nube.

- **Otros Insumos:**

- Etiquetado manual asistido mediante plugins QGIS.

7.4 Procedimiento metodológico

La metodología se organiza en cuatro fases, alineadas con los objetivos específicos:

Objetivo 1:

Fase 1: Adaptación de la arquitectura Deep Learning

- Revisión bibliográfica sobre arquitecturas DL para segmentación semántica.
- Selección inicial de una arquitectura robusta entre U-Net, DeepLabV3, transformer (SegFormer)
- Modificación para entrada multicanal: fusión de bandas RGB y datos derivados de LiDAR.
- Ajustes de hiperparámetros iniciales: tasa de aprendizaje, batch size, optimizador.

Objetivo 2:

Fase 2: Desarrollo del flujo computacional

seguimiento sucesivo

- Preprocesamiento de datos:
 - Registro y alineación entre LiDAR y ortoimágenes.
 - Generación de capas derivadas.
 - Etiquetado manual o semiautomático de edificaciones para entrenar el modelo.
- División del dataset: Entrenamiento (70%), validación (15%), test (15%).
- Entrenamiento del modelo:
 - Uso de augmentación de datos (rotación, ruido, escalado).
 - Entrenamiento en múltiples iteraciones (épocas).
 - Registro de métricas de pérdida y precisión por época.

Fase 3: Vectorización de edificaciones

- Postprocesamiento de resultados de segmentación.

- Conversión de raster a vectores mediante contornos y simplificación topológica.
- Filtrado por tamaño y forma para eliminar falsos positivos.
- Asignación de atributos espaciales desde LiDAR

Objetivo 4

Fase 4: Evaluación de la ~~precision~~ exactitud

exactitud

- Comparación con datos de referencia (cartografía básica)
- Cálculo de métricas de evaluación: las describo si o no
- Evaluación comparativa:
 - Modelo entrenado solo con ortoimágenes.
 - Modelo entrenado con ortoimágenes + LiDAR.
- Análisis estadístico de resultados para validar hipótesis.

8. Resultados Esperados

Este proyecto tiene como principal resultado el desarrollo de un modelo de segmentación semántica que combine imágenes aéreas y datos LiDAR para identificar edificaciones que estén parcialmente cubiertas por vegetación. Se espera que este modelo sirva de insumo por empresas dedicadas a la cartografía, sistemas de información geográfica (SIG) y por entidades encargadas de la planeación urbana.

Además, se espera construir un flujo de trabajo automatizado que permita procesar los datos y entrenar el modelo de manera reproducible. Este flujo estará documentado y podrá ser reutilizado por investigadores, estudiantes y desarrolladores que trabajen con inteligencia artificial y análisis espacial.

También se obtendrán capas vectoriales, es decir, mapas digitales donde estarán representadas las edificaciones identificadas por el modelo, incluyendo información como el tamaño, la forma y la altura de cada una. Estas capas podrán ser utilizadas por entidades públicas como el IGAC, las alcaldías y las oficinas de catastro.

Otro resultado importante será la evaluación de la ~~precision~~ exactitud del modelo, para conocer qué tan bien funciona en zonas con vegetación. Esto se hará mediante métricas y comparaciones con información de referencia, y será útil para la

comunidad académica y los responsables de tomar decisiones en proyectos urbanos o ambientales.

público

Finalmente, todo el trabajo quedará documentado a través de un repositorio que incluirá el código, una guía técnica y una memoria metodológica para posterior uso.

9. Cronograma de Actividades

Actividad	Tiempo en Meses											
	1	2	3	4	5	6	7	8	9	10	11	12
Revisión bibliográfica, definición del área y recolección de datos	x	x										
Preprocesamiento de imágenes y nubes de puntos			x	x								
Diseño y adaptación de la arquitectura DL					x	x						
Entrenamiento y validación del modelo							x	x	x			
Vectorización de resultados y generación de capas SIG										x		
Evaluación de precisión y comparación entre modelos										x	x	
Redacción de informe final, entrega de resultados y socialización												x

10 Presupuesto

Tabla 1. Presupuesto global de la propuesta por fuentes de financiamiento (en miles de \$)

Rubros	Fuentes		Total
	Contrapartida UN	Cofinanciación	
Personal	Tiempo director		\$20.000.000
Servicios Técnicos	Etiquetado manual		\$1.500.000
Equipos	Almacenamiento		\$600.000

	nube, alquiler		
Materiales y Suministros	Disco externo, licencias		\$2.500.000
Viáticos y Gastos de Viaje	NA		
Impresos y Publicaciones	Entregables finales		\$1.000.000
Patentes	NA		
TOTAL			\$25.000.000

ORGANIZACIÓN
FICHEROS

11. Bibliografía citada

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need*. In Advances in Neural Information Processing Systems (Vol. 30). <https://arxiv.org/abs/1706.03762>

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). *Encoder–Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. En Proceedings of the European Conference on Computer Vision (ECCV). Recuperado de <https://arxiv.org/abs/1802.02611>

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>

Reyes Quintana, N. M. (2024). Modelo de segmentación semántica de imágenes satelitales basado en redes neuronales convolucionales para la clasificación de cobertura de la tierra en páramos [Tesis de maestría, Universidad Nacional de Colombia]. Repositorio Institucional UNAL. <https://repositorio.unal.edu.co/handle/unal/86362>

Vali, A., Comai, S., & Matteucci, M. (2020). Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15), 2495. <https://doi.org/10.3390/rs12152495>

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2019). Joint deep learning for land cover and land use classification. *Remote Sensing of Environment*, 221, 173–187. <https://doi.org/10.1016/j.rse.2018.11.014>

~~Buscar artículos~~ } DL → edificación
Chen, L.-C., et al. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. ECCV.

He, K., et al. (2017). Mask R-CNN. IEEE ICCV.

Jensen, J. R. (2007). Remote Sensing of the Environment: An Earth Resource Perspective.

Liu, Y., et al. (2020). Building footprint extraction from high-resolution satellite images using deep learning. Remote Sensing.

Martha, T. R., et al. (2011). Object-oriented approach for landslide detection using remote sensing and GIS. IEEE Transactions.

Wang, X., et al. (2020). Automatic building extraction in aerial scenes using CNN-based instance segmentation. ISPRS Journal.

Zhou, T., et al. (2019). Occlusion-aware deep learning framework for building segmentation. Remote Sensing.

Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. ISPRS Journal of Photogrammetry and Remote Sensing, 140, 20-32.
<https://doi.org/10.1016/j.isprsjprs.2017.04.009>

Kampffmeyer, M., Salberg, A. B., & Jenssen, R. (2016). Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. CVPR Workshops, 1-9.
<https://doi.org/10.1109/CVPRW.2016.77>

Li, W., Fu, H., Yu, L., & Cracknell, A. (2020). Deep learning based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1-15.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8-36.
<https://doi.org/10.1109/MGRS.2017.2762307>