



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

¿Cómo?

¿Cómo?

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

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 precisión 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

Precisión

recall

en los métodos

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).

Prob.
model

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

→ GPU
(Unidad, procesa-
mico-digital)

completamente las redes convolucionales esta 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.

Las imágenes aéreas y datos LiDAR tienen la misma central práctica que la separación de datos satelitales mejora significativamente la precisión de los modelos. La clave es la reducción de interferencia vegetal.

5. Pregunta de Investigación

exactitud

¿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?

6. Objetivos

General:

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 el 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.

Lin, Y., Hyppä, J., & Jaakkola, A. (2019). Mini-UAV-borne LiDAR for fine-scale mapping. *Remote Sensing*, 11(13), 1569.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652–660.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234–241). Springer.
https://doi.org/10.1007/978-3-319-24574-4_28

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>

Altamirano P, L. (2021). Deep Learning aplicado a la Segmentación Semántica de Imágenes áereas. Universidad de Valparaíso, Chile.

Reporte de originalidad (reporte creado por una herramienta de evaluación de la originalidad en la producción y en el manejo documental, a la cual se haya sometido el documento de Proyecto de Tesis. Este reporte debe contar con el aval del director propuesto.

Kim (2020) Semantic Segmentation of Urban Buildings
Propone una red neuronal de alta resolución, mejorada para la segmentación semántica con img de alta resolución.

Transforme: Tipo de modelo de red neuronal que (evolucionó el campo del procesamiento del lenguaje natural. (mecanismos de atención)) Permite que el modelo enfocue su atención en diferentes partes ~~para~~ de la secuencia de entrada al procesar una frase (palabra...)

Ventajas:

- Paralelismo: las T pueden procesar toda la secuencia de una vez acelerando el entrenamiento.
- Escalabilidad: Maneja grandes volúmenes de datos

Segformer. Mod Segmentación semántica basado en T

- Segformer: Simple and Efficient design for S.S. with Transformers.

→ Qué hace

Combina la eficiencia de las CNN con la capacidad del contexto global de los T.

→ S.S. rápida sin necesidad de mecanismos complejos.

Mezcla de escalas: fusiona representaciones de distintas resoluciones

- Bajo costo computacional y precisión