Detección de objetos en LiDAR aéreo: una comparativa de métodos de etiquetado de instancias

Maria Goicoechea[1], Alex Salvatierra[1][0009−0005−8732−5581], Aranzazu Jurio[1][0000−0002−4087−586X], Jose Antonio Sanz[1][0000−0002−1427−9909] y Mikel Galar[1][0000−0003−2865−6549]

Departamento de Estadística, Informática y Matemáticas
Institute of Smart Cities (ISC), Universidad Pública de Navarra (UPNA)
Campus de Arrosadía s/n, 31006 Pamplona, Navarra, España {maria.goicoechea, alex.salvatierra, aranzazu.jurio, joseantonio.sanz, mikel.galar}@unavarra.es

Resumen La anotación de nubes de puntos LiDAR aéreas con cajas delimitadoras 3D a nivel de instancia es esencial para entrenar y evaluar modelos de detección de objetos. Sin embargo, las herramientas existentes están orientadas principalmente a escenarios de conducción autónoma y no contemplan los distintos niveles de información previa disponibles en adquisiciones aéreas complejas. Para abordar esta limitación, presentamos una taxonomía y un marco de decisión para el etiquetado de instancias de objetos en nubes de puntos, estructurados en torno a tres condiciones de entrada: nubes segmentadas semánticamente, puntos semilla georreferenciados o nubes de puntos sin procesar. Para cada escenario, analizamos estrategias de anotación manuales, semiautomatizadas y automatizadas, evaluando sus ventajas y limitaciones. La validación sobre conjuntos de datos aéreos sintéticos y reales permite caracterizar los compromisos entre esfuerzo de anotación, precisión espacial y robustez frente a la complejidad del terreno. Como resultado, el marco proporciona una metodología práctica para seleccionar la estrategia de etiquetado más adecuada según el nivel de preparación de los datos disponibles.

Palabras clave: LiDAR aéreo · Anotación de nubes de puntos · Detección de objetos 3D · Segmentación de instancias · Taxonomía de etiquetado.

1. Introducción

Los sensores LiDAR (Light Detection and Ranging) emiten pulsos láser y miden el tiempo de vuelo de los pulsos reflejados para generar representaciones tridimensionales densas y precisas del entorno. Según su plataforma de adquisición, pueden clasificarse en terrestres, embarcados en vehículos o aéreos, cada uno con características propias en cuanto a cobertura, densidad de puntos y ángulo de captura. Este trabajo se centra en el LiDAR aéreo, montado sobre

Vehículos Aéreos No Tripulados (UAVs), que permite capturar escenas exteriores a gran escala con altas densidades de puntos, resultando una tecnología clave en campos como la gestión forestal y la vigilancia aérea [4].

Una tarea fundamental con cualquier tipo de sensor LiDAR es la comprensión de la escena, abordada habitualmente resolviendo tareas de segmentación semántica o de detección de objetos. Esta última se encuentra menos explorada, especialmente en el dominio aéreo. El entrenamiento y evaluación de modelos para detección de objetos requiere conjuntos de datos etiquetados (anotados) no solo a nivel de punto, como es habitual en segmentación semántica, sino también a nivel de instancia: identificando cada aparición individual de un objeto (por ejemplo, cada vehículo o persona presente en la escena) mediante una caja delimitadora 3D orientada (*Oriented Bounding Box* - OBB).

Históricamente, la anotación de nubes de puntos en exteriores se ha enfocado en la conducción autónoma, donde los sensores operan a nivel de suelo y la distribución de los puntos presenta características bien estudiadas [1]. Por contra, la investigación en LiDAR aéreo se ha limitado mayoritariamente a la segmentación semántica a nivel de punto, sin etiquetar las instancias individuales ni delimitarlas mediante cajas. Si bien existen herramientas semiautomáticas como *labelCloud* [7] o *Scale AI*[1] para la colocación manual de cajas, su utilidad es limitada cuando se dispone de información auxiliar que podría acelerar el proceso de etiquetado. Por ejemplo, si se cuenta con coordenadas GNSS de los objetos de interés o con una segmentación semántica previa de la nube de puntos.

El escenario concreto que motiva y valida este trabajo es el del proyecto CANOPER[2]: adquisiciones LiDAR aéreas sobre entornos forestales con el objetivo de detectar objetos de fabricación humana (por ejemplo vehículos) ocultos bajo el dosel arbóreo (sub-canopy) para tareas de vigilancia y salvamento. En este tipo de escenas, los objetos de interés aparecen parcial o totalmente ocultos por el follaje, dando lugar a nubes de puntos escasas e irregulares para las que las herramientas de etiquetado existentes no ofrecen flujos de trabajo adaptados. A pesar de su relevancia, el etiquetado sistemático de instancias en este contexto apenas ha sido explorado en la literatura. Para abordar esta brecha, proponemos una taxonomía y un marco de decisión para la anotación de nubes de puntos LiDAR cuyas contribuciones principales son tres:

- Definimos un flujo de trabajo que se adapta a tres escenarios de entrada comunes: nubes de puntos en bruto (*raw*), puntos semilla georreferenciados y nubes preclasificadas semánticamente.
- Introducimos y utilizamos dos nuevos conjuntos de datos desarrollados para el proyecto CANOPER: un entorno sintético de alta fidelidad y una adquisición real mediante UAV en terreno forestal complejo.
- Presentamos un análisis empírico sobre el equilibrio entre extracción automatizada y refinamiento manual de cajas, evaluando la precisión geométrica, el esfuerzo del operador y la transferencia del dominio sintético al real.

[1] `https://scale.com`

[2] `https://www.frontex.europa.eu/innovation/eu-research/research-grants/project-canoper-1-5QHhMF`

2. Trabajo Relacionado

2.1. Herramientas de etiquetado 3D existentes

La anotación de nubes de puntos 3D es una tarea notablemente más compleja que el etiquetado de imágenes 2D. Tal como indican en [2], la anotación totalmente automatizada es aún inviable, e incluso las herramientas semiautomáticas exigen un esfuerzo manual considerable. Además, mientras que una caja delimitadora 2D solo requiere definir dos esquinas, dibujar una caja 3D exige determinar la posición central, sus dimensiones y su orientación (o, en su defecto, la posición de sus 8 vértices).

Entre las soluciones de código abierto, *labelCloud* (LC) [7] ofrece una interfaz de escritorio interactiva, lo que la convierte en una de las escasas herramientas aptas para la anotación aérea *offline* sin depender de infraestructuras en la nube. Propuestas comerciales como *Scale AI*, CVAT[3] y *LiDAR360*[4] están orientadas fundamentalmente para la conducción autónoma y carecen de la flexibilidad necesaria para gestionar entornos con vegetación densa. Una limitación común de todas estas herramientas es la falta de una metodología estructurada que guíe el proceso de etiquetado según la heterogeneidad de los datos de entrada.

2.2. Segmentación de instancias en nubes de puntos

La comprensión de escenas LiDAR se aborda principalmente mediante dos tareas. La segmentación semántica, más madura, asigna una etiqueta de clase a cada punto de forma individual; métodos como KPConv [9] o RandLA-Net [5] han alcanzado resultados sólidos tanto en entornos terrestres como aéreos. La detección de objetos, en cambio, exige además localizar e identificar cada instancia individualmente, lo que la convierte en una tarea considerablemente más compleja y menos explorada. Los enfoques existentes suelen seguir dos estrategias: métodos *bottom-up*, que agrupan puntos mediante algoritmos de densidad como DBSCAN [3], o proyecciones sobre imágenes de vista de pájaro (BEV) donde modelos como *Complex-YOLO* [8] y *PointPillars* [6] aplican detectores 2D. Sin embargo, estos métodos han sido desarrollados y validados principalmente en escenarios de conducción autónoma, donde los objetos son grandes, bien separados y capturados con alta densidad de puntos; su aplicabilidad directa a LiDAR aéreo bajo cubierta forestal, donde las instancias son escasas e irregulares, permanece en gran medida sin explorar.

3. Descripción del conjunto de datos

En este trabajo se utilizan dos conjuntos de datos desarrollados en marco del proyecto CANOPER, que tiene como objetivo detectar objetos de fabricación humana (por ejemplo vehículos) ocultos bajo la vegetación. Uno de los datasets

[3] `https://www.cvat.ai/`
[4] `https://www.greenvalleyintl.com/LiDAR360`

es sintético, generado mediante simulación, y el otro real, adquirido en campo. Ambos consisten en nubes de puntos que representan entornos forestales compuestos por terreno, vegetación y cuatro categorías de objetos de interés: vehículos, tiendas de campaña, cabañas y personas. A continuación se describe el procedimiento utilizado para obtener ambos conjuntos de datos:

- **CANOPER Sintético** Este conjunto se ha generado mediante una metodología de simulación híbrida que integra datos de terreno reales con modelos 3D procedimentales utilizando HELIOS++ [10]. El dataset consta de 66 escenas que suman aproximadamente 159 millones de puntos. Cada punto cuenta con una etiqueta semántica exacta y un identificador de instancia único.
- **CANOPER Real** Este conjunto se compone de capturas LiDAR aerotransportadas (UAV) obtenidas en tres ubicaciones de Navarra (Unzué, Montejurra y Urdiain) bajo condiciones controladas. Cada escena se acompaña de coordenadas GNSS que indican la posición aproximada de cada instancia, constituyendo la información de partida sobre la que se aplica la metodología de etiquetado propuesta en este trabajo.

La Tabla 1 resume la distribución de puntos por clase y el número de instancias para ambos conjuntos de datos.

Tabla 1. Distribución de puntos e instancias en los conjuntos CANOPER sintético y real.

	CANOPER Sintético		**CANOPER Real**	
Categoría	**Millones de puntos (%)**	**Instancias**	**Millones de puntos (%)**	**Instancias**
Suelo	43.00 (27.50)	–	231.00 (20.10)	–
Vegetación	113.00 (70.50)	–	915.00 (79.00)	–
Ruido de suelo	2.33 (1.40)	–	–	–
Objetos de interés	0.78 (0.49)	1112	0.54 (0.05)	311
Vehículo	0.19 (0.12)	59	0.12 (0.03)	58
Tienda de campaña	0.20 (0.13)	89	0.35 (0.01)	48
Cabaña	0.17 (0.11)	25	–	–
Persona	0.21 (0.13)	939	0.06 (0.01)	205
Total	159.00 (100.00)	–	1146.00 (100.00)	–

4. Taxonomía de Etiquetado

El núcleo de este estudio reside en una taxonomía diseñada para el etiquetado de instancias, mediante cajas 3D orientadas, sobre nubes de puntos LiDAR en entornos de vegetación densa (ver Figura 1). Este marco organiza metodologías existentes y propuestas originales según la información previa disponible: (A) nubes con segmentación semántica, (B) nubes con puntos semilla georreferenciados, y (C) nubes de puntos en bruto (*raw*).

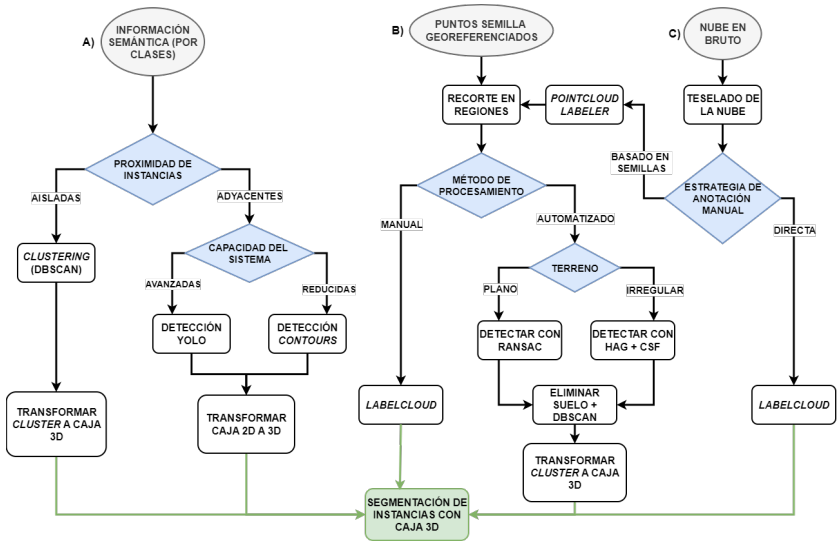


Figura 1. Diagrama resumen de la taxonomía propuesta

4.1. Escenario A: Nube de puntos con segmentación semántica

Este escenario asume que la nube de puntos ya dispone de una clasificación semántica densa a nivel de punto (obtenida, por ejemplo, mediante una red de segmentación semántica), pero carece de distinción entre instancias individuales de un mismo objeto. El objetivo es aislar y parametrizar mediante una caja cada instancia independiente. Se definen dos estrategias según lo cercanas que estén las instancias:

DBSCAN. Si existe separación visible entre las instancias, aplicamos el algoritmo de clustering DBSCAN, sobre cada uno de los tipos de objetos identificados, para identificarlas. Para evitar problemas con la inclinación del terreno, normalizamos las alturas a HAG (*Height Above Ground*). Dada la geometría de algunos objetos de interés (como coches o tiendas de campaña) la mayor separación entre las instancias se encuentra a partir de cierta altura. Por ello, aplicamos el clustering a los puntos cuya altura supere un umbral (HAG $> 1{,}2\,\mathrm{m}$). Posteriormente, los puntos inferiores a dicho umbral se asocian a su núcleo más cercano usando árboles k-d (hasta una distancia $d_{\mathrm{máx}}$); el resto se descarta como ruido.

La OBB 3D se parametriza asumiendo *pitch* y *roll* nulos: la orientación en planta se estima mediante PCA (Análisis de Componentes Principales) sobre la proyección horizontal de los puntos, y la altura se obtiene de la extensión en el eje Z. Las instancias con un número insuficiente de puntos se descartan.

Proyección BEV. Cuando las instancias están muy próximas, el agrupamiento por densidad, como el realizado por DBSCAN, tiende a fusionar instancias

independientes. Para mitigarlo, se proyecta la nube filtrada semánticamente sobre un plano de Vista de Pájaro (*Bird's Eye View*, BEV) para realizar una segmentación en el dominio de la imagen.

Para ello se define una extensión espacial mínima en el plano XY y se rasteriza la nube preservando la relación de aspecto. Si la nube carece de color RGB, los valores HAG se mapean a falso color para resaltar el contraste. A continuación se aplica un cierre morfológico para consolidar contornos, seguido de una erosión binaria para separar instancias adyacentes. Según la capacidad computacional del sistema, existen dos opciones para la detección en la imagen 2D respectivamente:

- **Opción I: YOLO.** Se emplea el modelo **YOLO11x-OBB** (variante *extra-large*) que incorpora una cabeza de predicción orientada. Dado que el modelo opera sobre imágenes BEV de LiDAR fuera de su dominio de entrenamiento, no se aprovecha la predicción de clase sino únicamente la localización de instancias: la inferencia se ejecuta en modo agnóstico a la clase.
- **Opción II: Contornos.** Como alternativa computacionalmente ligera que no requiere entrenamiento previo, se extraen los contornos vectoriales de los componentes conectados de la imagen. Para cada polígono que supere un área mínima, se calcula su rectángulo orientado de área mínima (*Minimum Area Bounding Box*).

Independientemente del método de detección, las coordenadas 2D de las esquinas del rectángulo orientado se proyectan de vuelta al sistema de coordenadas global. Una vez delimitada la base, se determina la altura de la caja como la mínima necesaria para englobar a todos los puntos de esa clase que están sobre esa base.

4.2. Escenario B: Puntos semilla georreferenciados para instancias

Esta vía se utiliza cuando la nube de puntos está vinculada a un archivo de coordenadas GNSS (X_s, Y_s, Z_s), que identifica la posición (mediante un único punto) y clase de cada instancia, pero carece de información volumétrica. El sistema determina un recorte de la nube de puntos centrado en cada punto semilla lo suficientemente grande para asegurar que la instancia está dentro (tamaño parametrizado por clase). Sobre estos recortes se ofrecen dos opciones:

Anotación manual. El operador procesa cada recorte manualmente mediante la herramienta *labelCloud*. Al aislar al anotador de la escena global y suprimir millones de puntos ajenos, el ajuste de la OBB 3D se agiliza considerablemente y se hace más preciso.

Segmentación automatizada. El primer paso consiste en la eliminación del suelo del recorte. Se proponen dos alternativas según la complejidad del relieve:

- **Opción I: RANSAC:** Se estima el plano dominante del recorte mediante RANSAC. Los puntos cercanos al plano (según un umbral adaptativo) se clasifican como terreno y se eliminan. Este enfoque es eficiente pero muy sensible en relieves abruptos.
- **Opción II: HAG vía Filtro CSF:** Para terrenos complejos, se emplea el *Cloth Simulation Filter* (CSF), que modela una malla elástica virtual depositada sobre la nube invertida. Esto permite calcular el HAG global de manera robusta. Se elimina como suelo todos aquellos puntos cuya HAG sea menor a un umbral (10 cm.).

Tras la eliminación del suelo, se define una semilla flotante en el centro del recorte, asumiendo que estará contenida en la instancia. Finalmente, se ejecuta DBSCAN para obtener el clúster que contenga la semilla, y la OBB 3D se calcula igual que en 4.1.

4.3. Escenario C: Nube de puntos en bruto (*Raw*).

Representa el escenario de máxima complejidad, al contar únicamente con la nube de puntos, sin anotaciones o referencias previas. Para poder trabajar con ella, la nube se divide en una cuadrícula de teselas regulares. Para cada una, se normalizan las alturas (HAG) a través del modelado del terreno (CSF) A continuación, se eliminan los puntos cuya altura sea mayor que un umbral, para evitar las copas de los árboles y que emerjan los objetos de interés localizados debajo. Una vez filtradas las nubes de puntos teseladas, un operador debe realizar la segmentación manualmente. Se plantean dos alternativas:

- **Opción I: Anotación directa de cajas manuales.** El operador determina las cajas de las instancias detectadas en (*labelCloud*). Para las instancias localizadas entre dos recortes, la caja se anota en el recorte que contenga la mayor proporción de puntos, determinando sus dimensiones estimadas completas. Las cajas resultantes se proyectan posteriormente sobre la nube completa, por lo que el teselado no afecta a la cobertura final del etiquetado.
- **Opción II: Anotación de puntos semilla.** Como alternativa semiautomática, el operador marca un único punto semilla en el centro aproximado de cada instancia a ras de suelo, anotando su clase. Esto transfiere automáticamente el flujo al pipeline automatizado del escenario B, evitando el ajuste manual de cajas.
  Para dar soporte a este etiquetado de semillas se desarrolló la herramienta *Point Cloud Labeler* [5], que permite colocar semillas sobre la nube con distintos modos de visualización y lanza automáticamente el pipeline del Escenario B tras su colocación.

5. Metodología de Evaluación

La taxonomía presentada encapsula cada instancia en una caja OBB. Sin embargo, también podemos obtener un etiquetado a nivel de punto, proyectando

[5] `https://github.com/arin-upna/pointCloudLabeler`

dichas cajas sobre la nube. Por ello, en la evaluación de resultados vamos a utilizar dos familias de métricas: una a nivel de caja (centrada en la localización y alineación de las OBB) y otra a nivel de punto (enfocada en la precisión del etiquetado denso).

5.1. Métricas de detección a nivel de caja

Se comparan las OBB predichas (B) frente a las de referencia o *ground truth* (B^{gt}) bajo dos criterios complementarios: la precisión geométrica ($\mathrm{IoU_v}$) y la recuperación de instancias (F1-score).

La IoU volumétrica se define como:

$$\mathrm{IoU_v}(B,\, B^{\mathrm{gt}}) = \frac{\mathrm{Vol}(B \cap B^{\mathrm{gt}})}{\mathrm{Vol}(B \cup B^{\mathrm{gt}})} \tag{1}$$

El volumen de intersección se calcula mediante el algoritmo Quickhull sobre la envolvente convexa de la región solapada, lo que hace a esta métrica sensible a errores de traslación, rotación y escala.

Al no disponer de puntuaciones de confianza, el emparejamiento entre cajas se realiza mediante una búsqueda voraz basada en la superposición espacial. Una predicción se considera Verdadero Positivo (TP) si su $\mathrm{IoU_v} > \tau$; en caso contrario, se computa como Falso Positivo (FP). Las cajas de referencia sin emparejar se contabilizan como Falsos Negativos (FN).

El **F1-score** penaliza tanto la sobresegmentación como las omisiones. En este trabajo se prioriza la evaluación de la recuperación de instancias fijando $\tau = 0$.

5.2. Métricas de clasificación a nivel de punto

Siguiendo la práctica habitual en segmentación semántica de nubes de puntos, se evalúa la precisión del etiquetado denso resultante de proyectar las OBB sobre la escena, tratando el conjunto de prueba como una única adquisición agregada. Para cada clase $c \in \mathcal{C}$ se acumulan los recuentos globales de TP_c, FP_c y FN_c, y se reportan dos variantes de IoU:

- **Macro IoU**: media aritmética de la IoU por clase, que asigna el mismo peso a cada categoría independientemente de su frecuencia. Es la métrica idónea para evaluar el rendimientoen presencia de clases minoritarias o desequilibrio de datos:
- **Micro IoU**: agrega los recuentos de todas las clases antes del cálculo del ratio. Esta variante pondera implícitamente cada clase por su abundancia de puntos, ofreciendo una visión de la predicción general de la nube de puntos etiquetada:

6. Resultados y Discusión

En esta sección se evalúa la calidad de las cajas delimitadoras obtenidas mediante la taxonomía propuesta, estructurando el análisis en dos bloques: datos

sintéticos y reales. En el conjunto sintético, al disponer de la localización exacta de las instancias, se evalúan los escenarios A (partiendo de segmentación semántica) y C (nube en bruto). En el conjunto real, dado que no existe *ground-truth* automático, se generó un etiquetado de referencia punto a punto por dos operadores humanos de forma independiente a la taxonomía propuesta. Debido al gran volumen de las adquisiciones reales, el escenario C se descarta por su elevado coste manual; la evaluación se centra por tanto en el escenario B y en el escenario A, este último posible gracias a las etiquetas semánticas del etiquetado de referencia.

6.1. Resultados en el conjunto de datos sintético

Tal y como muestra la Tabla 2, las estrategias del escenario A (partiendo de segmentación semántica) rozan la perfección en el entorno simulado. YOLO alcanza el F1@0 más alto (1.000), aunque el margen sobre DBSCAN (0.994) es insignificante, sugiriendo que el enfoque más simple basado en densidad es suficiente para datos limpios. La detección por contornos sufre una ligera caída (F1@0 = 0.894) impulsada casi en su totalidad por la clase *Persona* (0.868 en la Tabla 3), dado que su huella en vista de pájaro (BEV) es pequeña e irregular.

Tabla 2. Resumen de estrategias evaluadas en el conjunto CANOPER Sintético.

Ruta	**Estrategia**	**F1@0**	**IoU$_v$**	**Micro IoU**	**Macro IoU**
Escenario A	DBSCAN	0.994	0.997	0.510	0.752
	YOLO	1.000	0.984	0.514	0.758
	Contornos	0.894	0.920	0.502	0.716
Escenario C	LabelCloud cajas	0.657	0.740	0.527	0.651
	Semilla+LabelCloud	0.719	0.791	0.397	0.490
	Semilla+HAG	0.719	0.657	0.407	0.525
	Semilla+RANSAC	0.719	0.520	0.383	0.487

En el escenario C, sin información previa, el rendimiento global (F1@0 $\approx$ 0,7) es significativamente inferior. La Tabla 3 muestra que la clase *Persona* sigue siendo la más problemática. La similitud geométrica entre las personas, los troncos de los árboles y ciertos elementos de vegetación baja provocan confusiones durante el proceso de etiquetado, de modo que algunas personas presentes en las escenas no fueron anotadas. A nivel de calidad volumétrica, la anotación manual por recortes georreferenciados (*Semilla+LabelCloud*) produce las cajas más ajustadas (IoU$_v$ = 0.791). Entre los métodos automáticos equivalentes, HAG supera a RANSAC en calidad geométrica (0.657 vs. 0.520) porque el filtro CSF de HAG se adapta mucho mejor a la curvatura del terreno.

Tabla 3. Rendimiento por clase (recuperación de instancias F1@0 y precisión volumétrica IoU_v) en el conjunto CANOPER Sintético.

Ruta	Estrategia	Vehículo		Tienda		Cabaña		Persona	
		F1@0	IoU_v	F1@0	IoU_v	F1@0	IoU_v	F1@0	IoU_v
Escenario A	DBSCAN	1.000	1.000	1.000	0.985	1.000	1.000	0.993	1.000
	YOLO	1.000	0.980	1.000	0.974	1.000	1.000	1.000	0.978
	Contornos	1.000	0.980	1.000	0.960	1.000	1.000	0.868	0.948
Escenario C	LC cajas	1.000	0.777	0.941	0.787	1.000	0.748	0.558	0.649
	Semilla+LC	1.000	0.771	1.000	0.965	1.000	0.803	0.630	0.621
	Semilla+HAG	1.000	0.729	1.000	0.610	1.000	0.676	0.636	0.613
	Semilla+RANSAC	1.000	0.491	1.000	0.420	1.000	0.477	0.636	0.693

6.2. Resultados en el conjunto de datos real y transferencia

El rendimiento disminuye en datos reales (Tabla 4) debido a la variabilidad de densidad, oclusiones, capturas incompletas y la mayor presencia de vegetación baja, poco representada en los datos sintéticos.

Tabla 4. Rendimiento de las estrategias de segmentación (clase a instancia y georreferenciada) en CANOPER Real, incluyendo métricas de precisión y coste temporal.

Ruta	**Estrategia**	**F1@0**	**IoU_v**	**Micro IoU**	**Macro IoU**	**Tiempo Medio (min)**
Escenario A	DBSCAN	0.825	0.318	0.816	0.661	2.2
	YOLO	0.644	0.314	0.617	0.416	3.0
	Contornos	0.616	0.258	0.686	0.495	2.5
Escenario B	Manual	0.988	0.313	0.971	0.952	12,4
	HAG	0.870	0.266	0.433	0.144	2.3
	RANSAC	0.910	0.248	0.385	0.076	1.9

El enfoque manual mediante recortes (*Manual*) demuestra ser altamente resiliente, logrando una recuperación de instancias casi perfecta (F1@0 = 0.988) y un excelente Micro IoU a nivel de punto de 0.971. Por el contrario, los métodos automatizados de eliminación de suelo colapsan en métricas semánticas a nivel de punto: el Macro IoU se desploma a niveles residuales (0.144 y 0.076 para HAG y RANSAC respectivamente). La causa principal radica en el estrecho contacto espacial de la densa vegetación arbórea, la cual sobrevive al filtrado del suelo y es absorbida por el algoritmo de agrupamiento, inflando artificialmente el tamaño de las cajas delimitadoras resultantes y penalizando el IoU_v geométrico.

A nivel de clases, la *Tienda de campaña* se vio severamente afectada por hiperparámetros calibrados en simulación que no funcionaron bien en formas irregulares reales. Finalmente, la clase *Persona* sufre el mayor problema en la

Tabla 5. Rendimiento por clase en el conjunto Real (Vehículos, Tiendas y Personas).

Ruta	Estrategia	Vehículo		Tienda		Persona	
		F1@0	IoU_v	F1@0	IoU_v	F1@0	IoU_v
Escenario A	DBSCAN	0.992	0.411	0.658	0.282	0.799	0.260
	YOLO	0.667	0.430	0.515	0.293	0.669	0.219
	Contornos	0.737	0.336	0.833	0.253	0.487	0.185
Escenario B	Manual	1.000	0.356	1.000	0.361	0.980	0.223
	HAG	0.975	0.364	0.804	0.232	0.852	0.202
	RANSAC	0.992	0.209	0.990	0.350	0.863	0.184

brecha del dominio (*sim-to-real gap*): mientras que en simulación su silueta es estática y limpia, en el mundo real los micromovimientos de los sujetos durante el escaneo causan un "difuminado temporal"de los puntos, convirtiéndolos en nubes alargadas irreconocibles que se confunden fácilmente con maleza o ruidos verticales. Estas diferencias se ilustran en la Figura 2.

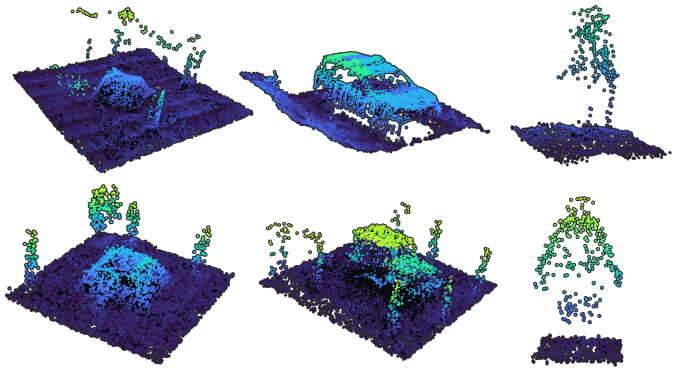


Figura 2. Ejemplos de objetos de interés en datos reales (arriba) y sintéticos (abajo): tienda, vehículo y persona (de izquierda a derecha).

7. Conclusiones

En este trabajo se ha presentado una taxonomía estructurada de estrategias para el etiquetado de instancias en nubes de puntos LiDAR bajo el dosel arbóreo. Mediante la evaluación de tres escenarios de entrada, se ha establecido un marco de decisión que permite equilibrar la precisión geométrica con el esfuerzo manual.

El hallazgo más relevante es que el nivel de información previa disponible es el factor determinante en el rendimiento. Mientras que las estrategias de clase a instancia alcanzan una precisión casi óptima en datos sintéticos, los métodos automatizados demuestran fragilidad frente a la complejidad del terreno real y la vegetación densa.

Los resultados subrayan que el salto del entorno simulado al real se manifiesta tanto en errores de detección como en distorsiones geométricas. Por ello, cualquier despliegue de estos flujos de trabajo en un nuevo dominio requeriría una recalibración específica.

Agradecimientos Este trabajo ha sido parcialmente financiado por el proyecto 2025/279 CANOPER-1, financiado por FRONTEX; por PID2022-136627NB-I00, financiado por MCIN/AEI/10.13039/501100011033 y por FEDER, UE; y por el proyecto PID2025-174319OB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

Referencias

1. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J.: Semantickitti: A dataset for semantic scene understanding of lidar sequences. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9297–9307 (2019)
2. Dong, S., Liu, F., Lin, G.: Leveraging Large-Scale Pretrained Vision Foundation Models for Label-Efficient 3D Point Cloud Segmentation (Nov 2023). https://doi.org/10.48550/arXiv.2311.01989
3. Ester, M., Kriegel, H.P., Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise
4. Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M.: Deep Learning for 3D Point Clouds: A Survey (Jun 2020). https://doi.org/10.48550/arXiv.1912.12033
5. Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A.: RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds (May 2020). https://doi.org/10.48550/arXiv.1911.11236
6. Lang, A.H., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O.: PointPillars: Fast Encoders for Object Detection from Point Clouds (May 2019). https://doi.org/10.48550/arXiv.1812.05784
7. Sager, C., Zschech, P., Kuhl, N.: labelCloud: A Lightweight Labeling Tool for Domain-Agnostic 3D Object Detection in Point Clouds. Computer-Aided Design and Applications **19**(6), 1191–1206 (Mar 2022). https://doi.org/10.14733/cadaps.2022.1191-1206
8. Simon, M., Milz, S., Amende, K., Gross, H.M.: Complex-YOLO: Real-time 3D Object Detection on Point Clouds (Sep 2018). https://doi.org/10.48550/arXiv.1803.06199
9. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: KPConv: Flexible and Deformable Convolution for Point Clouds (Aug 2019). https://doi.org/10.48550/arXiv.1904.08889
10. Winiwarter, L., et al.: Virtual laser scanning with helios++: A novel take on ray tracing-based simulation of topographic full-waveform 3d laser scanning. Remote Sensing of Environment **269**, 112772 (2022). https://doi.org/10.1016/j.rse.2021.112772