

La segmentación de instancias, no la semántica, limita la detección de árboles individuales en nubes de puntos ULS densas

Arturo Cantú Olivares, Diego Sebastian Cruz Cervantes*, Luis Fernando Maldonado, Jesús Gabriel Gudiño Lara, Daniel Cantón Enriquez

Resumen—La segmentación automatizada de árboles individuales (ITS) en nubes de puntos de escaneo láser de UAV (ULS) de alta densidad es fundamental para los inventarios forestales escalables, pero la contribución relativa del preprocesamiento semántico y la segmentación de instancias al error de ITS de extremo a extremo rara vez se aísla experimentalmente. Presentamos una comparación factorial controlada (diseño de 3×2) de tres estrategias de preprocesamiento (sin preprocesamiento, Random Forest (RF) en características geométricas y agrupamiento multiescala (MSG) de PointNet++) combinadas con dos variantes de un segmentador volumétrico interpretable Watershed 3D (siembra CHM clásica y una nueva siembra basada en densidad), calibradas independientemente por pipeline y evaluadas en el benchmark FOR-instance V1 en cinco tipos de bosques. A pesar de una calidad semántica casi perfecta (tree IoU > 0,97 para ambos clasificadores), la instancia F1 alcanza un pico de 0,209, lo que revela una fuerte disociación entre la calidad semántica y el rendimiento a nivel de instancia: la principal fuente de error surge de la separación de instancias en lugar del preprocesamiento semántico. Dentro de las configuraciones evaluadas, RF con características geométricas supera a PointNet++ MSG en todas las combinaciones de pipeline sin requerir una GPU, y la siembra basada en densidad mejora consistentemente sobre CHM (+47 % para RF, +74 % para PointNet++), particularmente en dosel denso de altura uniforme donde CHM es casi plano. Estos hallazgos son específicos de condiciones ULS densas, donde la densidad extrema de puntos y la continuidad del dosel amplifican la complejidad geométrica de la separación de árboles individuales.

Palabras clave—Segmentación de instancias, semántica, detección de árboles individuales, nubes de puntos ULS densas.

Instance Segmentation, Not Semantics, Limits Individual Tree Detection in Dense ULS Point Clouds

Abstract—Automated individual tree segmentation (ITS) in high-density UAV laser scanning (ULS) point clouds is critical for scalable forest inventories, yet the relative contribution of semantic preprocessing and instance segmentation to end-to-end ITS error is rarely isolated experimentally. We present a controlled factorial comparison (3×2 design) of three preprocessing strategies —no preprocessing, Random Forest (RF)

on geometric features, and PointNet++ multi-scale grouping (MSG)— combined with two variants of an interpretable volumetric Watershed 3D segmenter (classic CHM seeding and a novel density-based seeding), calibrated independently per pipeline and evaluated on the FOR-instance V1 benchmark across five forest types. Despite near-perfect semantic quality (tree IoU > 0.97 for both classifiers), instance F1 peaks at 0.209, revealing a strong dissociation between semantic quality and instance-level performance: the dominant source of error emerges from instance separation rather than semantic preprocessing. Within the evaluated configurations, RF with geometric features outperforms PointNet++ MSG across all pipeline combinations without requiring a GPU, and density-based seeding consistently improves over CHM (+47% for RF, +74% for PointNet++), particularly in dense uniform-height canopies where the CHM is nearly flat. These findings are specific to dense ULS conditions, where extreme point density and canopy continuity amplify the geometric complexity of individual tree separation.

Index Terms—Individual tree segmentation, UAV laser scanning (ULS), semantic segmentation, watershed 3D, density seeding, Random Forest, PointNet++, FOR-instance.

I. INTRODUCCIÓN

A. Motivación

Los inventarios forestales automatizados a nivel de árbol individual son esenciales para la estimación de carbono almacenado, el monitoreo de la biodiversidad y la silvicultura de precisión. El escaneo láser desde drones (ULS) proporciona densidades de puntos entre uno y dos órdenes de magnitud superiores a las del escaneo láser aerotransportado convencional (ALS) ($\sim 7,000$ pts/m² vs. ~ 5 pts/m²), lo que permite caracterizar árboles individuales, incluidas sus copas, troncos, ramas y sotobosque. Sin embargo, la segmentación automática de árboles individuales (ITS) en datos ULS sigue siendo un problema abierto debido al ruido en los modelos de altura, la ambigüedad entre copa y sotobosque, y la variabilidad estructural entre tipos de bosque, particularmente en bosques densos donde la alta densidad de puntos y la continuidad vertical hacen que múltiples árboles compartan regiones espaciales continuas.

Los pipelines de ITS suelen operar en dos etapas encadenadas: un clasificador semántico que aísla los puntos de árbol respecto al suelo, sotobosque y ruido, seguido de un segmentador de instancias que asigna a cada punto de árbol un identificador único.

Los métodos de aprendizaje profundo, como PointNet++ [2], han generado altas expectativas respecto

Manuscript received on 12/02/2026, accepted for publication on 11/04/2026. Corresponding author is Diego Sebastian Cruz Cervantes (dcruz54@alumnos.uaq.mx).

The authors are with Universidad Autónoma de Querétaro (UAQ), Querétaro, México.

a la calidad semántica; sin embargo, su ventaja práctica frente a clasificadores clásicos como Random Forest en el contexto específico del preprocesado para ITS aún no está clara. Al mismo tiempo, los métodos volumétricos de Watershed 3D han evolucionado más allá del enfoque clásico 2D basado en el modelo de altura del dosel (canopy height model, CHM) [9], permitiendo explotar directamente la densidad vertical de puntos. A pesar de estos avances, el impacto relativo de cada etapa sobre el rendimiento extremo a extremo rara vez se ha aislado experimentalmente [2], [5], [1], la mayoría de los trabajos recientes en ITS se enfocan principalmente en incrementar el desempeño global mediante arquitecturas end-to-end cada vez más complejas.

Sin embargo, al acoplar estrechamente la clasificación semántica y la segmentación de instancias dentro de modelos tipo black-box, resulta difícil identificar cuál de las etapas domina realmente el error final del pipeline.

Esto motiva una pregunta fundamental: en ambientes forestales ULS densos, ¿el principal límite del ITS moderno proviene de la clasificación semántica o de la separación de instancias?

B. Contribuciones

Las contribuciones de este trabajo son:

- (1) un diseño factorial controlado 3×2 capaz de aislar experimentalmente la contribución relativa del preprocesado semántico y de la segmentación de instancias dentro de pipelines ITS para ambientes ULS densos;
- (2) evidencia experimental de una fuerte disociación entre calidad semántica e instancia, donde valores tree IoU superiores a 0.97 no se traducen en mejoras proporcionales de segmentación individual;
- (3) un análisis comparativo entre Random Forest y PointNet++ MSG bajo estrategias CHM y Density dentro de un mismo segmentador Watershed 3D interpretable; y
- (4) evidencia de que el principal cuello de botella del pipeline ITS evaluado emerge de la separación de instancias y no del preprocesado semántico.

II. TRABAJO RELACIONADO

A. Detección de árboles individuales y segmentación de instancias

En ALS convencional (<10 pts/m²), el enfoque dominante para ITS detecta máximos locales en el CHM y delimita copas mediante watershed 2D [8], [7], [9]. Estos métodos se degradan en copas densas y estratificadas, donde las copas se superponen. Los datos ULS ($\sim 10^3$ – 10^4 pts/m²) habilitan enfoques volumétricos 3D que separan troncos adyacentes y sotobosque mediante llenado volumétrico, en lugar de depender de una proyección horizontal [9]; sin embargo, pocos trabajos aíslan la contribución del preprocesado semántico respecto del segmentador de instancias. Esta brecha motiva nuestro diseño factorial controlado.

B. Segmentación semántica de nubes de puntos forestales

Los métodos clásicos emplean características geométricas diseñadas manualmente junto con Random Forest [5] o SVM [11], [12], mientras que los métodos de aprendizaje profundo (PointNet [3], PointNet++ [2]) aprenden representaciones de extremo a extremo. PointNet++ ha mostrado resultados prometedores en ULS con características normalizadas por altura, pero la clasificación semántica se evalúa predominantemente como tarea final; su papel como preprocesado para ITS sigue poco estudiado.

C. Benchmarks

FOR-instance V1 [1] re'une parcelas ULS de cinco tipos de bosque distribuidos en cuatro continentes, con anotaciones semánticas y de instancia a nivel de punto. FOR-instanceV2 [10] amplía este recurso; en este trabajo evaluamos sobre V1 para mantener la comparabilidad. NeonTreeEvaluation [6] ofrece datos multimodales, pero está limitado a ecosistemas de Norteamérica.

Métodos end-to-end recientes como TreeLearn [13], SegmentAnyTree [14] y ForestFormer3D [10] integran representación semántica y agrupamiento de instancias en un mismo modelo, alcanzando mejores métricas globales pero dificultando aislar qué componente domina el error final.

III. DATASET

A. FOR-instance

FOR-instance V1 [1] es un benchmark ULS con anotaciones semánticas e de instancias a nivel de punto. Utilizamos cinco de las seis colecciones (Tabla I), excluyendo NIBIO2 por sobre-representar el bosque boreal noruego, lo que sesgaría las métricas agregadas. Las densidades varían entre 454 y 10156 pts/m² (mediana por colección: RMIT 473, TUWIEN 1406, CULS 2510, SCION 3064, NIBIO 7246). La diversidad geográfica y estructural es clave: NIBIO representa el 50% del test (bosque boreal denso), mientras que RMIT y TUWIEN presentan estructuras de copa radicalmente diferentes.

B. Splits y reproducibilidad

Se utiliza el split oficial (70% dev / 30% test) con subdivisión interna train/val (80/20%) a nivel de plot con semilla fija, siguiendo [13], [14].

IV. METODOLOGÍA

A. Visión general del pipeline

En lugar de proponer una nueva arquitectura ITS end-to-end, este trabajo adopta un diseño factorial interpretable que prioriza el aislamiento experimental del efecto del preprocesado semántico y la estrategia de separación de instancias sobre la maximización directa del desempeño absoluto. Watershed 3D fue seleccionado como segmentador

base por su dependencia explícita de la generación geométrica de semillas, permitiendo observar cómo los errores semánticos se propagan hacia la segmentación de instancias bajo condiciones controladas.

Los flujos comparten una arquitectura de dos etapas: (1) preprocesado semántico y (2) segmentador Watershed 3D, variadas factorialmente en cinco flujos (Fig. 1). Los parámetros del segmentador se calibran independientemente por flujo mediante grid search sobre val (F1 de instancia, IoU 3D $\geq$ 0.5).

B. Flujo A — Sin preprocesado (baseline)

La nube completa (excluyendo clases $\{0,3\}$) se pasa directamente al segmentador. Todos los flujos comparten el cálculo de la altura sobre el suelo (HAG) a partir de un modelo digital de terreno (DTM) interpolado sobre puntos de suelo, truncada a $[0, 50]$ m.

C. Flujo B — Random Forest sobre features geométricas

Cada punto se representa mediante 27 features geométricas: 8 descriptores eigen-basados y 5 descriptores locales (HAG, verticalidad, rugosidad, rango vertical, intensidad normalizada), calculados a dos vecindades $k \in \{20, 50\}$, más densidad volumétrica local ($r = 0,5$ m) [11], [12]. Se impone un radio máximo de búsqueda de 5,0 m. El clasificador es un Random Forest ($n = 200$ árboles, balanceo por clase) entrenado con submuestreo de 12 000 pts/clase/plot.

D. Flujo C — PointNet++ MSG

Se emplea PointNet++ con agrupamiento multiescala (MSG) [2] con cuatro módulos Set Abstraction y cuatro Feature Propagation [4]. La entrada son 5 dimensiones por punto ($XYZ + HAG_norm + intensidad$), garantizando equivalencia de información con el Flujo B. La red se entrena durante 100 épocas con submuestreos de 8 192 puntos. En inferencia se realizan múltiples pasadas con promediado de probabilidades; los puntos residuales se etiquetan por interpolación KNN ($k = 3$).

E. Segmentador de instancias — Watershed 3D volumétrico

El segmentador opera sobre los puntos clasificados como árbol en HAG y sigue cuatro pasos: (1) voxelización 3D en (X, Y, HAG) ; (2) suavizado gaussiano 3D de la grilla de densidad; (3) detección de semillas; y (4) watershed volumétrico sobre la densidad suavizada invertida con máscara de vóxeles ocupados [9]. A diferencia del watershed 2D sobre CHM, el fill volumétrico separa correctamente troncos adyacentes y puntos de sotobosque.

E1. Variante CHM (Flujos A, B, C): Las semillas se detectan sobre la envolvente superior del volumen (equivalente a un CHM derivado del propio grid 3D) mediante `peak_local_max`.

E1a. Variante Density (Flujos D, E): Se reemplaza el CHM por una superficie de densidad acumulada en una banda de `top_band_m` metros bajo el vóxel más alto de cada columna. Esta superficie refleja concentraciones locales de puntos incluso cuando el CHM es casi plano, como en bosques boreales densos de altura uniforme.

E1b. Calibración: Parámetros calibrados por grid search independiente por flujo (Tabla II, Fig. 2).

F. Detalles computacionales y reproducibilidad

Implementación en Python (Open3D, Scikit-learn, PyTorch). RF se entrena en CPU (~ 16 min/ejecución); PointNet++ en GPU NVIDIA RTX 3060 12 GB (~ 2.5 h, 100 épocas). Código, scripts y configuraciones: <https://github.com/Cantuuu/ITS-Pipeline-Comparison>.

V. PROTOCOLO DE EVALUACIÓN

Semántica. Se reportan tree IoU ($TP/(TP + FP + FN)$), seg mIoU (promedio de IoU de ambas clases) y seg F1 binario (árbol/no-árbol) a nivel de punto.

Instancias. Las instancias predichas $\{P_i\}$ se emparejan con las GT $\{G_j\}$ mediante matching *greedy* por IoU 3D punto a punto con umbral $\tau = 0,5$, siguiendo [13], [14]:

$$IoU_{3D}(P_i, G_j) = \frac{|P_i \cap G_j|}{|P_i \cup G_j|} \quad (1)$$

A partir del IoU 3D definido en la Ec. (1), se computan precisión, recall, F1 y *coverage* (fracción de árboles GT con al menos una predicción con $IoU_{3D} \geq \tau$, sin matching único): la brecha Coverage — Recall indica fragmentación de copas. Reportar ambas tareas permite aislar en cuál etapa se concentra el error.

VI. RESULTADOS

A. Resultados agregados

Las Tablas III y IV presenta los resultados sobre el conjunto de test. El Flujo D (RF + Density) alcanza el mejor desempeño agregado de segmentación de instancias (F1 = 0.209), mientras que el seeding por densidad mejora consistentemente ambos clasificadores respecto al CHM tradicional.

Ambos clasificadores logran tree IoU > 0.97 , pero esta calidad no se traduce en instancia (F1 máximo 0.209 ± 0.161), evidenciando al segmentador como cuello de botella. La alta dispersión entre plots (SD comparable o superior a la media en todos los flujos) refleja la heterogeneidad estructural entre los cinco tipos de bosque evaluados. Notablemente, el density seeding reduce la variabilidad de RF (SD: $0.226 \rightarrow 0.161$), sugiriendo mayor robustez entre sitios.

TABLE I
COLECCIONES DEL DATASET FOR-INSTANCE V1 UTILIZADAS EN ESTE TRABAJO. GT: GROUND TRUTH (ANOTACIONES DE REFERENCIA).

Código	País	Tipo de bosque	Plots test	Árboles GT
NIBIO	Noruega	Boreal conífero (<i>P. abies</i>)	6	161
CULS	Rep. Checa	Templado conífero (<i>P. sylvestris</i>)	1	20
TUWIEN	Austria	Mixto deciduo aluvial	1	35
RMIT	Australia	Esclerófilo seco (eucaliptos)	1	64
SCION	Nueva Zelanda	Plantación conífera (<i>P. radiata</i>)	2	43
Total			11	323

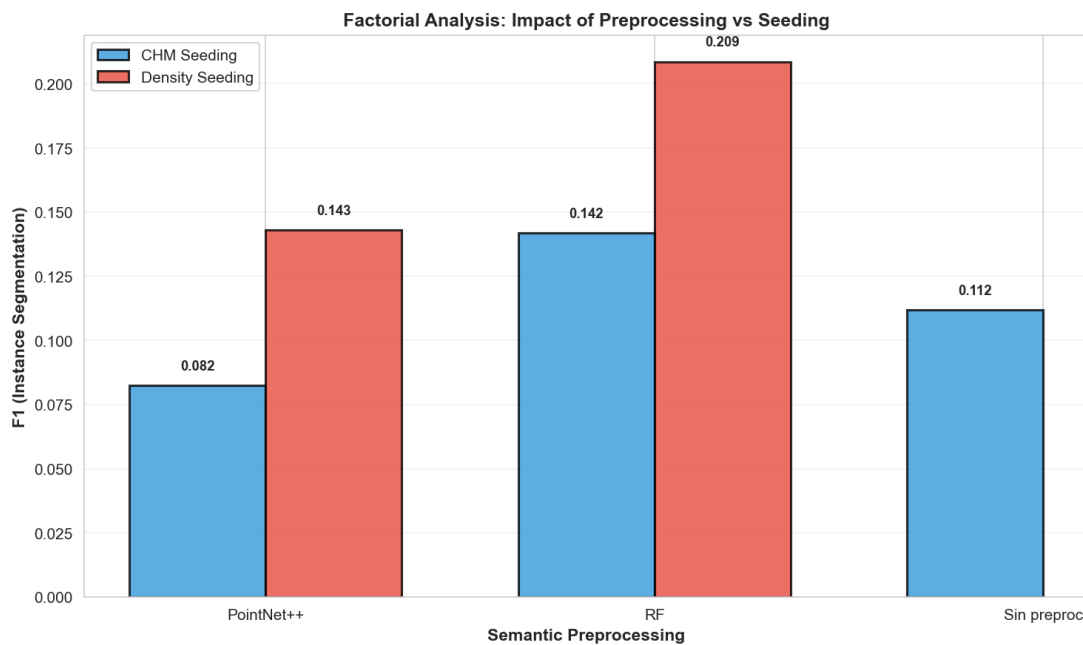


Fig. 1. Análisis factorial: el seeding por densidad (rojo) supera al CHM (azul) en todos los métodos (+47–74%), con mejora mayor que el cambio de preprocesado (+27%).

TABLE II
PARÁMETROS ÓPTIMOS DEL WATERSHED 3D POR FLUJO (GRID SEARCH SOBRE VAL, F1 IoU 3D ≥ 0.5).

Parámetro	Espacio	A	B (RF)	C (PN++)	D (RF+Dens.)	E (PN+++Dens.)
voxel_size (m)	{0.1, 0.2, 0.3}	0.2	0.1	0.1	0.2	0.3
gaussian_sigma	{0.3, 0.5, 1.0}	0.3	0.3	1.0	0.5	0.3
min_crown_radius_m	{0.5, 1.0, 1.5, 2.0}	0.5	1.0	1.5	2.0	1.5
top_band_m	{2.0, 3.0, 5.0}	—	—	—	2.0	5.0
F1 val (medio)		0.272	0.371	0.377	0.467	0.450

TABLE III
CALIDAD SEMÁNTICA POR FLUJO (TEST, NIVEL DE PUNTO).

Flujo	Seeding	sem mIoU	tree IoU	sem F1
A — Sin preproc.	CHM	—	—	—
B — Random Forest	CHM	0.985	0.990	0.996
C — PointNet++ MSG	CHM	0.974	0.982	0.994
D — Random Forest	Density	0.985	0.990	0.996
E — PointNet++ MSG	Density	0.974	0.982	0.994

B. Resultados por sitio

La Fig. 3 muestra la clasificación semántica sobre CULS 2. RF concentra errores principalmente en bordes de copa, mientras que PointNet++ produce falsos positivos aislados que posteriormente generan semillas espurias durante la segmentación.

La Fig. 4 muestra el F1 por sitio. CULS presenta el mejor desempeño debido a la separación clara entre copas, mientras que NIBIO representa el escenario más desafiante por su continuidad de dosel y alta densidad estructural.

El seeding por densidad mejora consistentemente todos los sitios, especialmente en RMIT y TUWIEN, donde el CHM pierde capacidad discriminativa bajo estructuras de copa complejas.

C. Visualización de segmentación de instancias

Las Figs. 5 y 6 muestran vistas superiores de predicciones: CULS exhibe buena separación de copas mientras que NIBIO muestra sobre-fusión persistente.

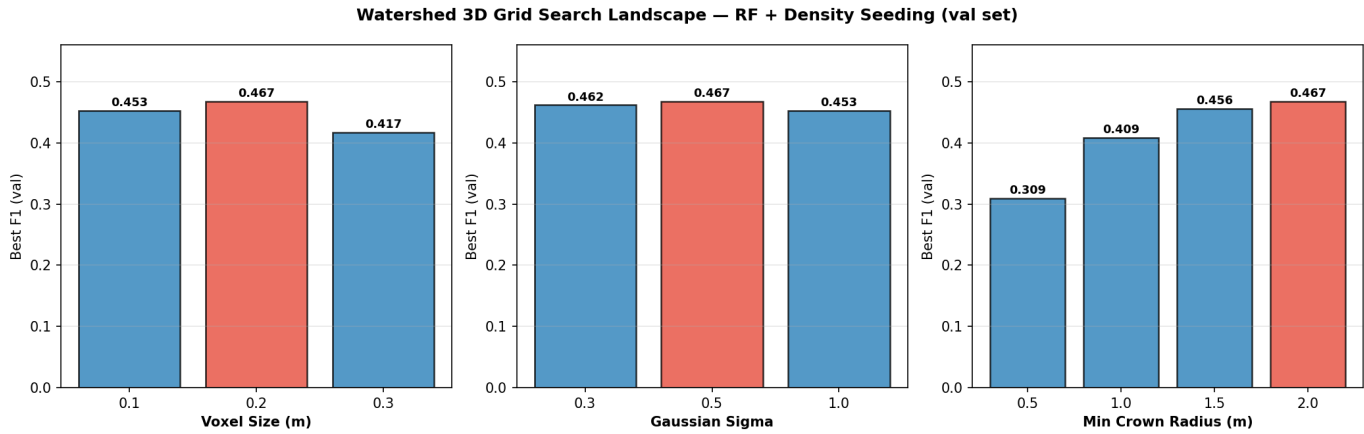


Fig. 2. Paisaje del grid search para el Flujo D (RF + Density). Barras rojas: valor óptimo. El radio mínimo de copa (2.0 m) es el parámetro más determinante.

TABLE IV
RESULTADOS DE SEGMENTACIÓN DE INSTANCIAS (TEST, IOU 3D ≥ 0.5 , MEDIA $\pm$ SD SOBRE $n = 11$ PLOTS).

Flujo	Seeding	inst Prec	inst Recall	inst F1	Coverage
A — Sin preproc.	CHM	0.138 $\pm$ 0.149	0.133 $\pm$ 0.143	0.112 $\pm$ 0.116	0.133 $\pm$ 0.143
B — Random Forest	CHM	0.394 $\pm$ 0.402	0.110 $\pm$ 0.118	0.142 $\pm$ 0.226	0.110 $\pm$ 0.118
C — PointNet++ MSG	CHM	0.172 $\pm$ 0.259	0.060 $\pm$ 0.079	0.082 $\pm$ 0.162	0.060 $\pm$ 0.079
D — Random Forest	Density	0.418 $\pm$ 0.291	0.150 $\pm$ 0.097	0.209 $\pm$ 0.161	0.150 $\pm$ 0.097
E — PointNet++ MSG	Density	0.209 $\pm$ 0.148	0.118 $\pm$ 0.076	0.143 $\pm$ 0.109	0.118 $\pm$ 0.076

[†]SD > media en algunos flujos refleja distribuciones de F1 fuertemente asimétricas con acumulación en cero; véase mediana en resultados complementarios.

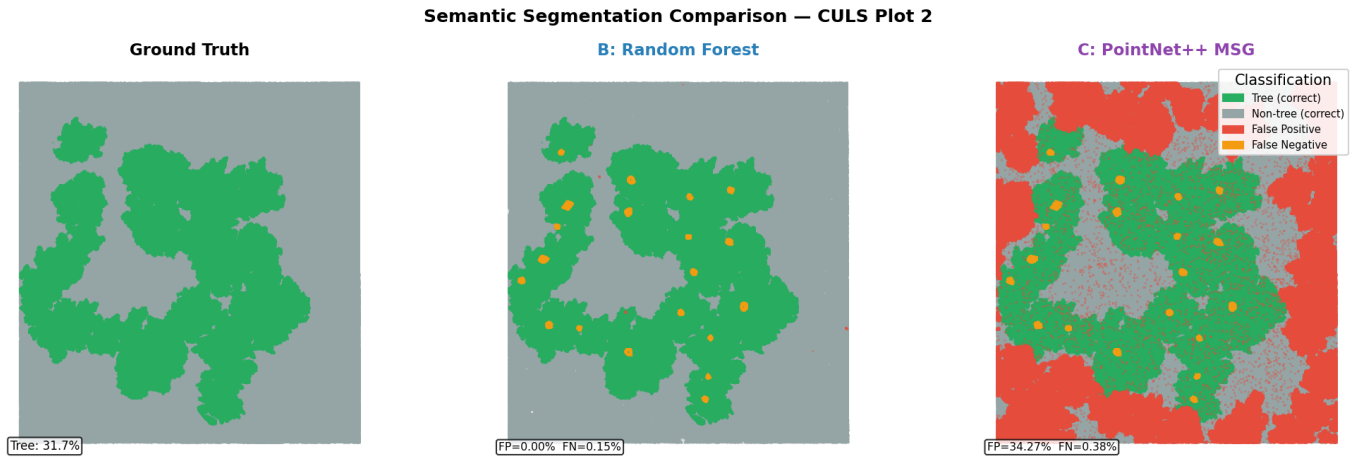


Fig. 3. Clasificación semántica sobre CULS 2. Verde: árbol correcto; rojo: FP; naranja: FN. RF concentra errores en bordes de copa; PointNet++ produce FP aislados en suelo.

D. Análisis factorial: preprocesado $\times$ seeding

El análisis factorial (Tabla V, Fig. 7) confirma que el cambio de seeding (CHM $\rightarrow$ Density) produce una mejora mayor (+47–74%) que el cambio de preprocesado (baseline $\rightarrow$ RF, +27%), identificando al segmentador como cuello de botella dominante. La interacción entre ambos factores es relevante: la semántica más limpia de RF produce menos semillas espúreas, resultando en menor sobre-segmentación (Precisión RF+Density: 0.418 vs. PN++Density: 0.209).

D0a. Análisis de significancia estadística.: Se emplea un test de permutación exacto pareado (bilateral, 2^{11}

permutaciones, $n = 11$ plots) con intervalos bootstrap BCa al 95% (10 000 remuestreos); se prefiere sobre Wilcoxon o t -Student por preservar la magnitud de las diferencias y no requerir supuestos distribucionales sobre la distribución F1, fuertemente asimétrica con acumulación en cero.

La Tabla VI resume los resultados: RF supera a PointNet++ con significancia plena bajo CHM ($p = 0.016$) y Density ($p = 0.020$), y RF+Density supera al baseline ($p = 0.016$). El density seeding alcanza significancia marginal ($p = 0.096$ para RF, $p = 0.066$ para PointNet++) limitada por $n = 11$; el efecto es consistente en 8/11 plots, los IC BCa para

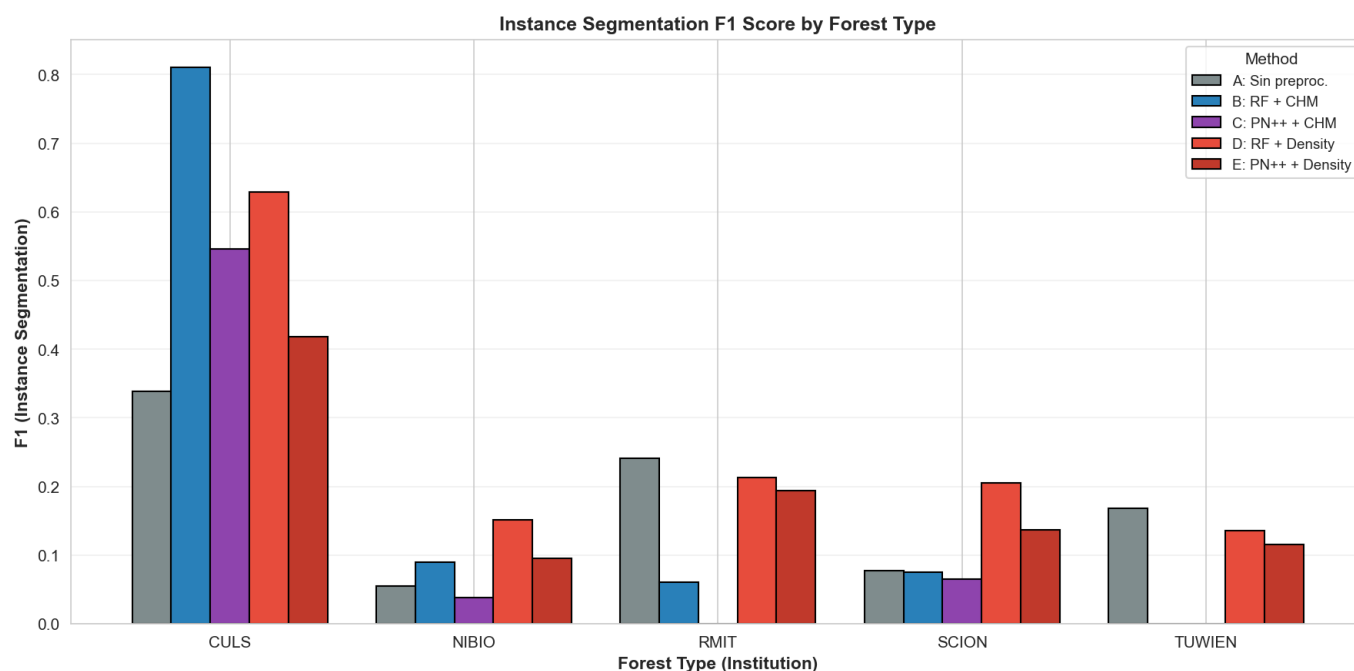


Fig. 4. F1 de instancia por tipo de bosque. El Flujo D (RF + Density, rojo) domina en todos los sitios excepto CULS (F1=0.811 con Flujo B).

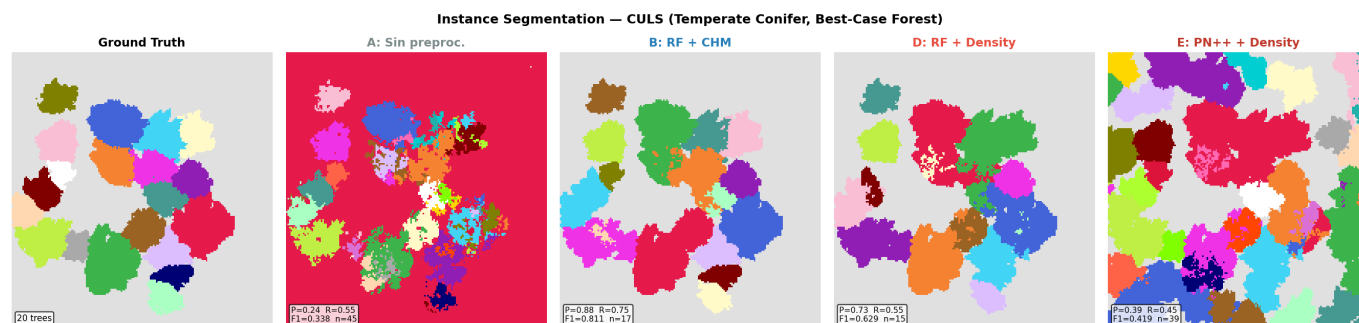


Fig. 5. Vista superior en CULS 2. El Flujo D logra la mejor separación (F1=0.629); el CHM tiende a fusionar copas adyacentes.

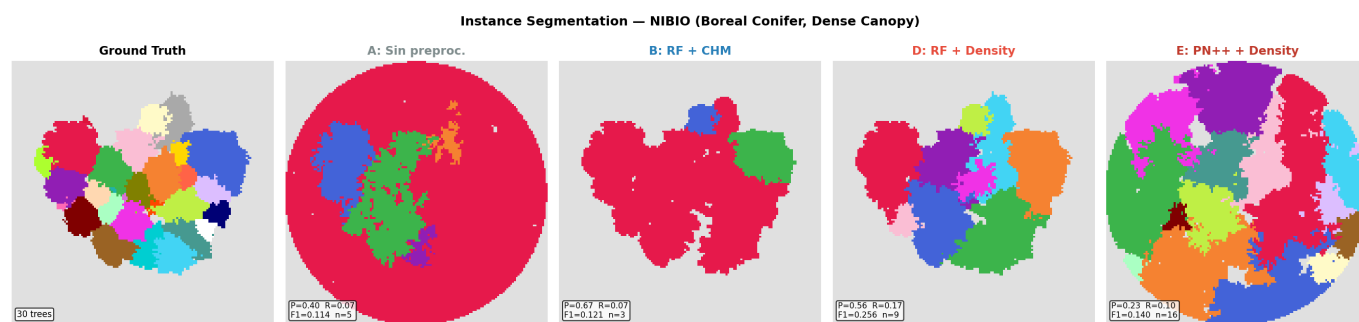


Fig. 6. Vista superior en NIBIO 17 (boreal denso). Under-segmentation severo en todos los métodos; F1 máximo 0.256 con Flujo D.

TABLE V
F1 DE INSTANCIA (TEST, MEDIA $\pm$ SD) EN EL DISEÑO FACTORIAL. Δ : MEJORA RELATIVA DEL SEEDING DENSITY SOBRE CHM.

Preprocesado	CHM	Density	Δ
Sin preproc. (baseline)	0.112 $\pm$ 0.116	—	—
Random Forest	0.142 $\pm$ 0.226	0.209 $\pm$ 0.161	+47%
PointNet++ MSG	0.082 $\pm$ 0.162	0.143 $\pm$ 0.109	+74%

PointNet++ excluyen el cero, y la reducción de SD de F1 ($0.226 \rightarrow 0.161$ para RF) indica mayor robustez entre tipos de bosque. Validar significancia estricta del seeding requerirá FOR-instanceV2 [10], con mayor número de plots y diversidad.

E. Importancia de features — Random Forest

La HAG domina la clasificación RF con el 61.5% de la importancia Gini acumulada en sus dos escalas ($k=50$: 0.325; $k=20$: 0.290), confirmando que la altura sobre el suelo es el predictor primario de la separación árbol/no-árbol. La verticalidad (0.087) y la rugosidad (0.072) a escala gruesa aportan información geométrica complementaria.

VII. DISCUSIÓN

A. El segmentador como cuello de botella

La Fig. 8 ilustra cómo el Flujo D asigna correctamente puntos de tronco y ramas en CULS 2, mientras que la proyección 2D del CHM confundiría estos puntos. El fill volumétrico del Watershed 3D es la diferencia estructural decisiva. Cabe notar que la calibración independiente por flujo (Tabla II) produce parámetros óptimos diferentes: el density seeding prefiere vóxeles más gruesos (0.2–0.3 m) para acumular suficiente masa de puntos en la banda superior, mientras que el CHM requiere alta resolución (0.1 m) para preservar los picos locales de la envolvente. Esta divergencia no es un sesgo experimental sino una restricción geométrica de cada estrategia de seeding; la calibración independiente permite que cada pipeline alcance su techo de rendimiento.

El hallazgo central (Fig. 7) es que la semántica casi perfecta ($\text{tree IoU} > 0.97$) no se traduce en calidad de instancia proporcional (F1 máximo 0.209). El principal factor limitante observado en el pipeline evaluado es la capacidad de separación de instancias bajo estructuras de copa altamente continuas y superpuestas. El análisis factorial confirma que la inversión más rentable en un pipeline ITS de dos etapas no está en mejorar la semántica — que RF ya supera ampliamente — sino en mejorar el segmentador de instancias. El efecto es dependiente del sitio: en CULS (copas separadas), RF+CHM alcanza $F1=0.811$ y el Density seeding no aporta mejora sustancial; en NIBIO, RMIT y TUWIEN, donde las copas se superponen, el CHM falla sistemáticamente y el Density seeding es necesario para obtener rendimiento no trivial.

La dificultad observada refleja un problema geométrico más general en ULS denso: cuando las copas se superponen y la altura es uniforme, múltiples árboles comparten una misma masa volumétrica de puntos, reduciendo la separabilidad incluso antes de aplicar un algoritmo específico. Cualquier método basado principalmente en señales geométricas locales enfrentará dificultades similares; Watershed 3D funciona aquí como instrumento interpretable para exponer esta limitación estructural más que como una afirmación de optimalidad algorítmica. En adquisiciones ALS de menor densidad, donde

las copas son más dispersas, la relación entre error semántico e instancia podría diferir.

Aunque métodos recientes como TreeLearn [13], SegmentAnyTree [14] y ForestFormer3D [10] probablemente alcancen mejores métricas absolutas, el objetivo de este trabajo no es competir contra arquitecturas estado-del-arte, sino analizar la relación entre calidad semántica y separación de instancias dentro de un pipeline interpretable.

B. RF vs. PointNet++: viabilidad sin GPU y generalización

RF obtiene mejores resultados que PointNet++ en las configuraciones evaluadas, particularmente bajo escenarios con alta heterogeneidad estructural entre sitios. Esta ventaja no se explica por semántica superior (brecha en $\text{tree IoU} < 1\text{ pp}$), sino por mejor interacción con el segmentador: los errores de RF son falsos positivos de borde de copa, mientras que PointNet++ genera falsos positivos aislados en suelo que producen semillas espúreas. Además, RF se entrena e infiere en CPU (~ 16 min) sin requerir GPU.

PointNet++ muestra caída severa en RMIT y TUWIEN ($F1=0.000$ con CHM), debido a que su submuestreo global de 8 192 puntos no captura adecuadamente sitios con densidades y morfologías muy diferentes a los datos de entrenamiento (dominados por NIBIO boreal). RF, al operar sobre features locales, generaliza mejor entre tipos de bosque diversos.

Esta ventaja se observa dentro de las configuraciones evaluadas y no debe leerse como superioridad general de clasificadores clásicos sobre deep learning (cf. §7.3); aún así, refuerza la hipótesis central de desacoplamiento, ya que mejoras semánticas no se traducen en mejoras proporcionales a nivel de instancia.

C. Limitaciones

C1. Tamaño y balance del conjunto de evaluación.: El conjunto de test comprende 11 plots y 323 árboles GT, con una concentración significativa en un solo tipo de bosque: NIBIO (boreal conífero) aporta 6 de 11 plots ($\sim 50\%$) y 161 de 323 árboles ($\sim 50\%$), mientras que CULS, TUWIEN y RMIT contribuyen un único plot cada uno. Esta distribución introduce dos efectos cuantificables: (i) alta varianza en las estimaciones por sitio — los intervalos de confianza bootstrap al 95% de F1 para el mejor flujo (D) abarcan $[0.128, 0.314]$, reflejando que sitios individuales como CULS ($F1=0.629$) o NIBIO_5 ($F1=0.000$) dominan la dispersión; y (ii) posible sesgo en las métricas agregadas hacia el comportamiento en bosque boreal denso, que es el escenario más desafiante para todos los métodos.

C2. Análisis de errores de segmentación.: Aunque cuantificamos F1, precisión, recall y coverage, no descomponemos sistemáticamente los errores en sobre/infra-segmentación; un análisis formal por tipo de bosque queda como trabajo futuro.

TABLE VI
TEST DE PERMUTACIÓN EXACTO PAREADO (BILATERAL, 2¹¹ PERMUTACIONES, $n = 11$ PLOTS) E INTERVALOS DE CONFIANZA BOOTSTRAP BCa AL 95% PARA DIFERENCIAS EN F1 DE INSTANCIA.

Comparación	Δ F1	SD(Δ)	p_{perm}	IC 95% BCa (Δ)	Plots con mejora
RF: CHM $\rightarrow$ Density (B vs. D)	+0.067	0.119	0.096*	[−0.004, +0.129]	8/11
PN++: CHM $\rightarrow$ Density (C vs. E)	+0.061	0.094	0.066*	[+0.003, +0.110]	8/11
Baseline $\rightarrow$ RF+Density (A vs. D)	+0.097	0.103	0.016**	[+0.027, +0.161]	9/11
Baseline $\rightarrow$ RF+CHM (A vs. B)	+0.030	0.181	0.665	—	7/11
RF vs. PN++ con CHM (B vs. C)	−0.060	0.082	0.016**	—	—
RF vs. PN++ con Density (D vs. E)	−0.066	0.077	0.020**	—	—

**Significativo a $\alpha = 0.05$. *Marginalmente significativo a $\alpha = 0.10$. IC BCa: 10 000 remuestreos, semilla 42.

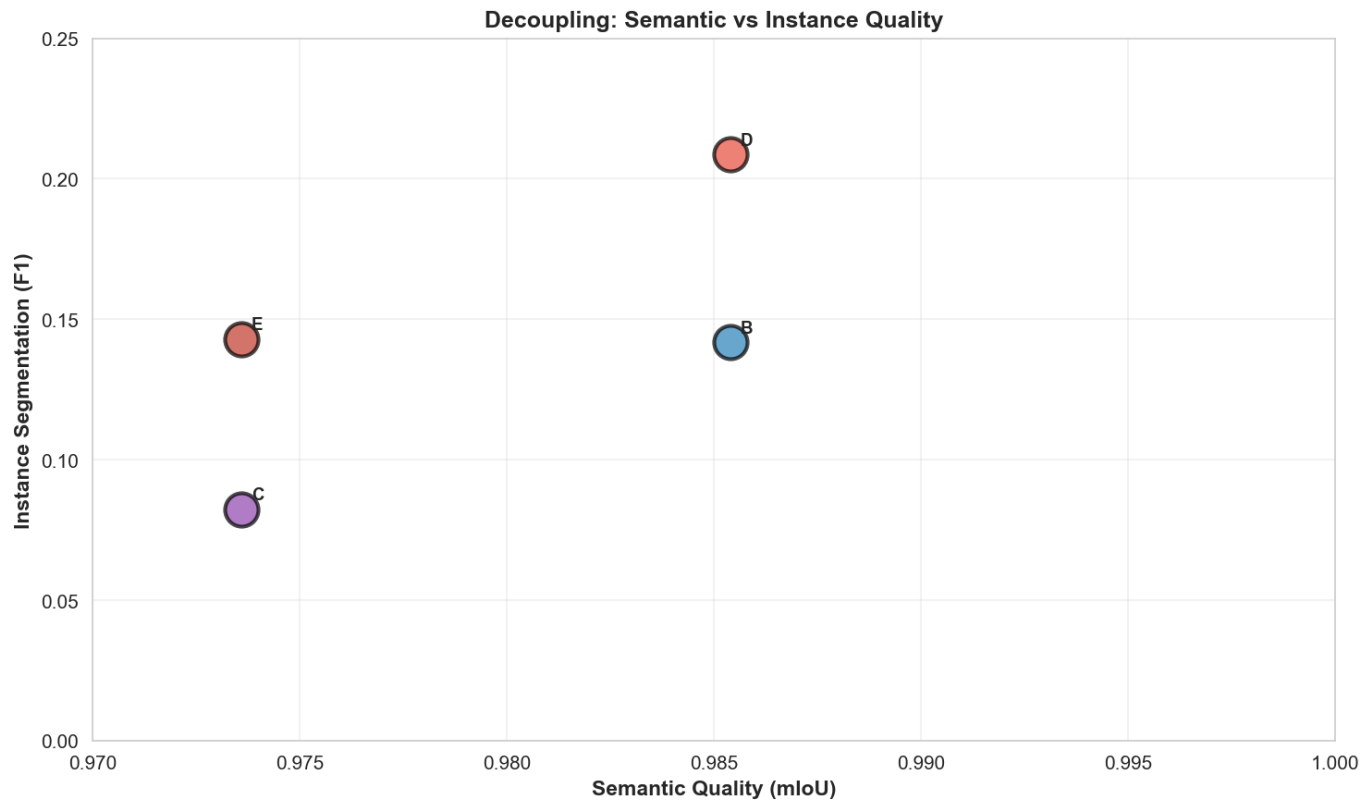


Fig. 7. Desacoplamiento semántico-instancia. Tree IoU > 0.97 para ambos clasificadores, pero F1 inst. difiere 1.7× (RF vs. PN++ con CHM). El segmentador es el factor dominante.

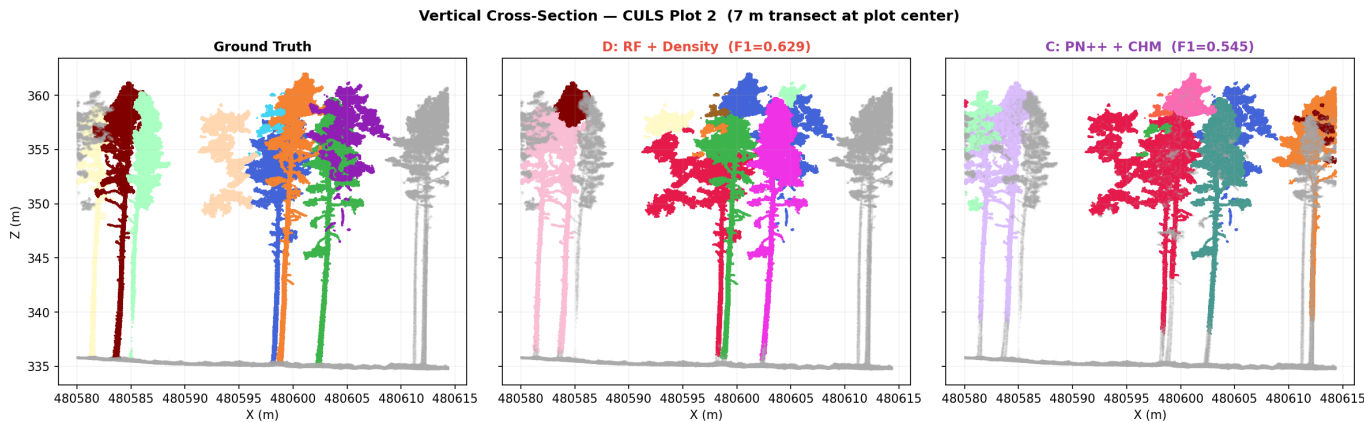


Fig. 8. Perfil vertical (7 m) en CULS Plot 2. El Flujo D (centro) separa correctamente los árboles usando fill volumétrico; PointNet++ + CHM (derecha) confunde troncos compartidos.

C3. *Otras limitaciones.*: PointNet++ se evalúa con una única configuración de hiperparámetros; por ello, la comparación RF vs. PointNet++ debe leerse como observación interna del pipeline evaluado y no como veredicto entre paradigmas. Las conclusiones sobre el cuello de botella son específicas a Watershed 3D volumétrico y a datos ULS (454–10 156 pts/m²); otros segmentadores (TreeLearn [13], SegmentAnyTree [14]) o adquisiciones ALS podrían mostrar una relación distinta entre error semántico e instancia.

VIII. CONCLUSIONES

Mediante un diseño factorial 3×2 sobre FOR-instance V1, demostramos que: (1) el segmentador de instancias, no el preprocesado semántico, es el cuello de botella principal del pipeline ITS; (2) Random Forest con features geométricas supera significativamente a PointNet++ MSG en todas las combinaciones evaluadas ($p = 0.016$, test de permutación exacto), sin requerir GPU; y (3) el seeding por densidad de copa mejora consistentemente al seeding CHM (+47% para RF, +74% para PointNet++), con significancia marginal ($p = 0.096$ y $p = 0.066$, respectivamente) limitada por el tamaño muestral del benchmark ($n = 11$ plots), siendo especialmente efectivo en bosques densos de altura uniforme.

La combinación RF + Watershed 3D Density alcanza el mejor F1 de instancia (0.209 en test). La brecha entre calidad semántica y de instancia sugiere que los avances más impactantes en ITS vendrán de segmentadores capaces de separar copas adyacentes en canopías densas mediante priors geométricos, arquitecturas end-to-end o seeding adaptativo. Trabajo futuro: comparación con TreeLearn [13] y SegmentAnyTree [14], extensión a FOR-instanceV2 [10], y análisis formal de sobre/infra-segmentación.

REFERENCES

- [1] S. Puliti *et al.*, “FOR-instance: a UAV laser scanning benchmark dataset for semantic and instance segmentation of individual trees,” arXiv:2309.01279, 2023.
- [2] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “PointNet++: Deep hierarchical feature learning on point sets in a metric space,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5099–5108.
- [3] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “PointNet: Deep learning on point sets for 3D classification and segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 652–660.
- [4] X. Yan, “Pointnet_Pointnet2_pytorch: PyTorch implementation of PointNet and PointNet++,” 2019. [Online]. Available: https://github.com/yanx27/Pointnet_Pointnet2_pytorch
- [5] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [6] B. G. Weinstein *et al.*, “A benchmark dataset for canopy crown detection in airborne RGB, LiDAR and hyperspectral imagery (NEON),” *PLOS Comput. Biol.*, vol. 17, no. 7, p. e1009180, 2021.
- [7] C. A. Silva *et al.*, “Imputation of individual longleaf pine tree attributes from field and LiDAR data,” *Can. J. Remote Sens.*, vol. 42, no. 5, pp. 554–573, 2016.
- [8] S. C. Popescu *et al.*, “Estimating plot-level tree heights with lidar via canopy-height-based variable window filtering,” *Comput. Electron. Agric.*, vol. 37, no. 1–3, pp. 71–95, 2002.
- [9] M. Dalponte and D. A. Coomes, “Tree-centric mapping of forest carbon density from ALS and hyperspectral data,” *Methods Ecol. Evol.*, vol. 7, no. 10, pp. 1236–1245, 2016.

- [10] B. Xiang *et al.*, “ForestFormer3D: A unified framework for end-to-end segmentation of forest LiDAR 3D point clouds,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2025, arXiv:2506.16991.
- [11] M. Weinmann *et al.*, “Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers,” *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 286–304, 2015.
- [12] M. Weinmann *et al.*, “Geometric features and their relevance for 3D point cloud classification,” *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. IV-1/W1, pp. 157–164, 2017.
- [13] J. Henrich *et al.*, “TreeLearn: deep learning for individual tree segmentation in ground-based LiDAR,” *Ecol. Inform.*, vol. 84, p. 102888, 2024.
- [14] M. Wielgosz *et al.*, “SegmentAnyTree: Sensor- and platform-agnostic deep learning for tree segmentation,” *Remote Sens. Environ.*, vol. 313, p. 114367, 2024.