Análisis comparativo de modelos de aprendizaje profundo para la segmentación semántica de nubes de puntos LiDAR aéreas en Navarra

Alex Salvatierra[1], José Antonio Sanz[1], Christian Gutiérrez[2] y Mikel Galar[1]

[1] Departamento de Estadística, Informática y Matemáticas, Institute of Smart Cities (ISC), Universidad Pública de Navarra (UPNA), Campus de Arrosadía s/n, Pamplona, 31006, Navarra, España

[2] Tracasa Instrumental, Cabárceno, 6, Sarriguren, 31621, Navarra, España

alex.salvatierra@unavarra.es

Resumen Los avances recientes en aprendizaje profundo han mejorado notablemente la segmentación semántica 3D, pero la mayoría de modelos se han desarrollado y evaluado principalmente en datos de escenas interiores o terrestres, mientras que su comportamiento en LiDAR aéreo adquirido en condiciones reales ha sido menos estudiado. Para abordar esta limitación, este trabajo evalúa varias arquitecturas del estado del arte sobre un conjunto de datos LiDAR aéreo de gran escala adquirido en condiciones reales en Navarra, España, con escenas urbanas, rurales e industriales. Se comparan KPConv, RandLA-Net, Superpoint Transformer y Point Transformer V3 sobre cinco clases habituales en cartografía aérea como suelo, vegetación baja, vegetación media y alta, edificios y vehículos. Los resultados muestran precisiones globales superiores al 93 %. KPConv obtiene la mejor mIoU (78,51 %) y el rendimiento más consistente entre clases, especialmente en categorías difíciles e infrarrepresentadas. Point Transformer V3 destaca en vehículos (75,11 % de IoU), Superpoint Transformer logra resultados competitivos con un número menor de parámetros, y RandLA-Net ofrece eficiencia computacional a cambio de menor rendimiento global.

Palabras clave: LiDAR aéreo · segmentación semántica · aprendizaje profundo · nubes de puntos · evaluación experimental

1. Introducción

El LiDAR (*Light Detection and Ranging*) [22] aéreo es una tecnología clave para adquirir información tridimensional de la superficie terrestre a gran escala. Los sistemas de escáner láser aerotransportado (*Airborne Laser Scanning*, ALS) [24] reconstruyen el terreno, vegetación y estructuras urbanas con alta precisión geométrica, dando soporte a aplicaciones de monitorización ambiental [28], gestión del territorio [30] y evaluación de infraestructuras [6]. A medida que aumentan la resolución espacial y la cobertura de estos datos, la segmentación semántica gana importancia para transformar nubes de puntos sin procesar

en productos geoespaciales interpretables, como Modelos Digitales de Superficie (DSM) y Modelos Digitales del Terreno (DTM) [15].

El aprendizaje profundo ha impulsado la segmentación semántica 3D durante la última década, principalmente a partir de experimentos interiores, terrestres y de conducción autónoma. Tras PointNet [3], arquitecturas basadas en muestreo jerárquico [16], convoluciones [20], MLP [10] y *transformers* [11,18,26,27] han alcanzado resultados sólidos en conjuntos interiores [1,5] y terrestres [2,9]. Sin embargo, estos escenarios difieren del LiDAR aéreo, que presenta densidades irregulares, gran escala territorial, fuerte desequilibrio entre clases y variaciones notables en la escala de los objetos. Por ello, no es evidente trasladar directamente a este dominio las conclusiones obtenidas en otros escenarios 3D.

La segmentación de LiDAR aéreo presenta retos propios frente a escenarios terrestres o de cartografía urbana. La perspectiva nadiral y los parámetros de vuelo generan distribuciones de puntos heterogéneas, mientras que la cobertura geográfica introduce una elevada variabilidad de escenas. Además, suelo, vegetación y edificios concentran la mayoría de puntos, y objetos pequeños como vehículos quedan infrarrepresentados. Estos factores hacen necesaria una evaluación específica en condiciones realistas de adquisición, donde el rendimiento global puede ocultar errores relevantes en clases minoritarias.

Conjuntos como DALES [21] y FRACTAL [7] han ampliado la diversidad geográfica y semántica del LiDAR aéreo. Aun así, ningún conjunto público combina, hasta donde sabemos, gran extensión espacial, alta densidad, evaluación amplia de arquitecturas y diversidad de entornos. En este contexto, resulta relevante evaluar arquitecturas recientes en datos LiDAR aéreos operativos adquiridos en condiciones reales de producción.

Para abordar esta limitación, este trabajo evalúa KPConv [20], RandLA-Net [10], Superpoint Transformer [18] y Point Transformer V3 [26] sobre datos LiDAR aéreos operativos adquiridos en Navarra, España. La comparación cubre arquitecturas convolucionales, basadas en MLP y basadas en *transformers*, y analiza su robustez en escenas urbanas, rurales e industriales, sin filtrado adicional y con el desequilibrio natural de las clases. El objetivo no es solo identificar el mejor modelo en términos globales, sino estudiar cómo cambian sus fortalezas y debilidades por clase bajo condiciones de adquisición real.

El resto del artículo se organiza como sigue. La Sección 2 revisa el trabajo relacionado. La Sección 3 describe la configuración experimental, incluyendo datos, modelos, entrenamiento y métricas. La Sección 4 presenta los resultados cuantitativos y cualitativos, y la Sección 5 recoge las conclusiones.

2. Trabajo relacionado

La segmentación semántica de nubes de puntos 3D ha avanzado rápidamente con el desarrollo de arquitecturas de aprendizaje profundo. PointNet [3] estableció el procesamiento directo de conjuntos de puntos no ordenados, y PointNet++ [16] añadió agregación jerárquica para capturar estructura local a distintas escalas. Después se han explorado enfoques convolucionales, como KP-

Conv [20], PointCNN [13] y PointConv [25]; modelos basados en MLP, como RandLA-Net [10]; métodos basados en grafos [12,23]; y, más recientemente, *transformers* [8,18,26,27,34]. Estos paradigmas introducen sesgos inductivos distintos, por lo que pueden comportarse de manera diferente ante clases pequeñas, fronteras de clases o regiones con densidad irregular.

La mayoría de estas arquitecturas se han desarrollado y evaluado en datos interiores o terrestres, como S3DIS [1], ScanNet [5], SemanticKITTI [2] y Semantic3D [9]. Estos conjuntos difieren del LiDAR aéreo en perspectiva de adquisición, densidad, extensión espacial y distribución de clases. En consecuencia, una arquitectura que obtiene buenos resultados en interiores o conducción autónoma no tiene por qué mantener el mismo comportamiento en escenas aéreas extensas, adquiridas desde una perspectiva nadiral y clases muy desequilibradas.

En LiDAR aéreo, el aprendizaje profundo ha recibido atención más recientemente gracias a conjuntos de datos específicos de ALS. ISPRS Vaihingen [19] estandarizó el etiquetado semántico con nueve clases, aunque su cobertura (0,2 km^2) y densidad (6 pts/m^2) son limitadas para entrenar modelos modernos a gran escala. DFC2019 [29] y DublinCity [35] ampliaron cobertura y densidad, respectivamente, pero siguieron limitados en diversidad de escenarios o extensión geográfica.

DALES [21] supuso un hito, con 10 km^2, 500 millones de puntos, 50 pts/m^2 y ocho clases, evaluando arquitecturas como KPConv, PointNet++ y Superpoint Graphs en escenas urbanas. Otros trabajos se centraron en aspectos concretos: LASDU [31] en entornos urbanos densos, OpenGF [17] en filtrado de suelo a gran escala (47 km^2), YUTO Semantic [32] en varias misiones urbanas y CENAGIS-ALS [33] en densidades muy altas (275 pts/m^2).

Más recientemente, FRACTAL [7] presentó 250 km^2 de paisajes heterogéneos, con 9260 millones de puntos etiquetados a 37 pts/m^2, aunque evaluó únicamente RandLA-Net. ECLAIR [14] aportó 10 km^2, 50 pts/m^2 y once clases, pero se centró en arquitecturas basadas en Minkowski [4].

A pesar de estos avances, los conjuntos ALS públicos siguen limitados en extensión, densidad, diversidad ambiental o variedad de arquitecturas evaluadas. Muchos trabajos se centran en escenas urbanas o en protocolos muy concretos, y el desarrollo de modelos basados en *transformers* ha avanzado más rápido que su evaluación en LiDAR aéreo. Como resultado, arquitecturas como Superpoint Transformer [18] y Point Transformer V3 [26] todavía están poco estudiadas en este dominio, especialmente bajo condiciones operativas y con desequilibrio de clases.

Este estudio responde a esa necesidad evaluando KPConv, RandLA-Net, Superpoint Transformer y Point Transformer V3 sobre LiDAR aéreo operativo y escenarios heterogéneos, con especial atención al rendimiento por clase.

3. Configuración experimental

3.1. Datos y preprocesamiento

Los experimentos se realizaron sobre un conjunto de datos LiDAR aéreo adquirido en condiciones operativas de vuelo en el área de Pamplona en Navarra, España. El conjunto de datos cubre aproximadamente 4 km^2 con entornos urbanos, industriales, rurales y semirrurales (Fig. 1). El área se dividió antes del teselado en zonas espacialmente disjuntas de entrenamiento (65 %) y prueba (35 %), con una densidad media aproximada de 50 pts/m^2. Cada punto está descrito por sus coordenadas (x, y, z), intensidad, RGB, número de retorno, número total de retornos y el índice de vegetación de diferencia normalizada (NDVI).

El conjunto contiene cinco clases habituales en cartografía aérea: *suelo*, *vegetación baja*, *vegetación media/alta*, *edificio* y *vehículo*. La Tabla 1 muestra el fuerte desequilibrio entre clases: *vehículo* y *vegetación baja* representan juntas alrededor del 2 % de los puntos etiquetados.

Para el entrenamiento y la evaluación, las nubes originales se dividieron en teselas solapadas de 50×50 m y se normalizaron coordenadas y atributos. En inferencia, las predicciones solapadas se fusionaron promediando probabilidades por punto. Esta estrategia mantiene contexto espacial adicional en los bordes de tesela y actúa como una forma implícita de aumento durante la evaluación.

3.2. Modelos

Se seleccionaron cuatro arquitecturas representativas: KPConv, RandLA-Net, Superpoint Transformer (SPT) y Point Transformer V3. En conjunto, cubren paradigmas complementarios en el procesamiento de nubes de puntos 3D y han mostrado un rendimiento sólido en conjuntos públicos, lo que permite comparar modelos con sesgos geométricos y costes computacionales distintos.

KPConv [20] define convoluciones continuas sobre vecindarios de puntos mediante núcleos de puntos. Su diseño permite preservar detalle espacial fino y modelar superficies locales de forma robusta.

RandLA-Net [10] es un modelo ligero y escalable para nubes de gran escala que combina muestreo aleatorio y agregación local, reduciendo el consumo de memoria. Esta eficiencia lo hace especialmente atractivo para escenas extensas, aunque puede implicar pérdida de detalle local.

Superpoint Transformer (SPT) [18] opera sobre superpuntos geométricamente coherentes y combina agrupamiento espacial con atención para capturar contexto de largo alcance sin procesar todos los puntos de forma independiente.

Point Transformer V3 (PTv3) [26] serializa las nubes de puntos en secuencias y aplica atención por parches, evitando vecindades KNN costosas y ampliando el campo receptivo mediante una formulación más eficiente.

Todos los modelos se implementaron a partir de código público[3] y se entrenaron bajo un protocolo común para garantizar comparabilidad. Las configuracio-

[3] KPConv: `https://github.com/torch-points3d/torch-points3d`; RandLA-Net: `https://github.com/IGNF/myria3d`; SPT: `https://github.com/drprojects/superpoint_transformer`; PTv3: `https://github.com/Pointcept/Pointcept`.

Tabla 1. Puntos etiquetados por clase en millones (M) y porcentaje (%).

Clase	Puntos (M)	%
Suelo	87,99	62,50
Vegetación baja	1,98	1,41
Vegetación media/alta	22,77	16,17
Edificio	27,09	19,24
Vehículo	0,96	0,68
Total	**140,80**	**100**

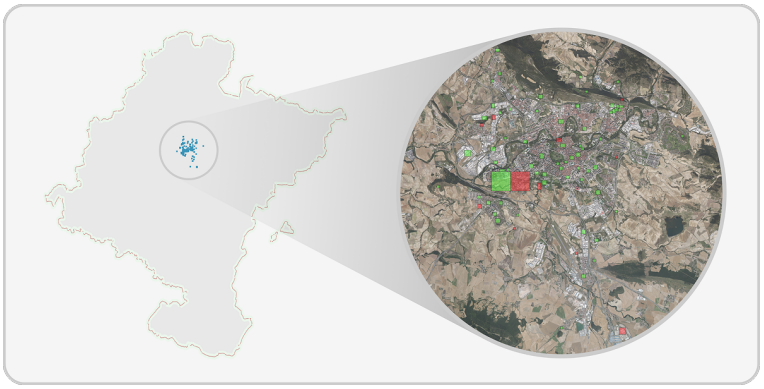


Figura 1. Distribución geográfica de las escenas LiDAR evaluadas alrededor de Pamplona, Navarra, España. Las áreas en verde indican zonas de entrenamiento y las áreas en rojo representan el conjunto de prueba.

nes siguieron las publicaciones originales, con pequeños ajustes para adaptarlas a la escala espacial y a la densidad del LiDAR aéreo.

3.3. Protocolo de entrenamiento

Todos los modelos usaron los mismos atributos de entrada y se entrenaron y evaluaron bajo un protocolo común. El entrenamiento e inferencia se realizaron en un servidor con GPU NVIDIA RTX 6000 Ada de 48 GB, procesador AMD EPYC 9454 y 64 GB de RAM. Cada modelo se entrenó tres veces con distintas semillas aleatorias, y se reporta el rendimiento medio para reducir el efecto de variaciones en el entrenamiento.

Todos los modelos utilizaron el mismo submuestreo tras la normalización. El conjunto de prueba se reservó exclusivamente para la evaluación final, sin emplearse para ajustar hiperparámetros. Los hiperparámetros principales se fijaron para cada arquitectura a partir de sus configuraciones originales y se resumen en la Tabla 2.

Tabla 2. Hiperparámetros de entrenamiento para los modelos evaluados.

Modelo	Tamaño del lote	Tasa de aprendizaje	Épocas
KPConv	10	0,001	20
RandLA-Net	24	0,004	80
SPT	4	0,01	100
PTv3	10	0,002	100

3.4. Métricas de evaluación

El rendimiento se evaluó con métricas estándar de segmentación semántica: precisión global (*Overall Accuracy*, OA), intersección sobre unión por clase (IoU) e IoU medio (mIoU). Estas métricas combinan precisión global y consistencia por clase, incluyendo categorías dominantes y minoritarias.

El IoU por clase (Ec. 1), el IoU medio (Ec. 2) y la precisión global (Ec. 3) se definen como:

$$\mathrm{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c} \tag{1}$$

$$\mathrm{mIoU} = \frac{1}{N_c} \sum_{c=1}^{N_c} \mathrm{IoU}_c \tag{2}$$

$$\mathrm{OA} = \frac{\sum_{c=1}^{N_c} TP_c}{\sum_{c=1}^{N_c} (TP_c + FN_c)} \tag{3}$$

donde TP_c, FP_c y FN_c denotan el número de puntos verdaderos positivos, falsos positivos y falsos negativos para la clase c, respectivamente, y N_c es el número total de clases.

OA resume el acierto global, mientras que mIoU pondera todas las clases por igual y resulta más informativa con datos desequilibrados. Por ello se reportan tanto métricas globales como métricas por clase, especialmente relevantes para categorías infrarrepresentadas.

4. Resultados

Esta sección presenta los resultados cuantitativos y cualitativos de los experimentos. El análisis considera métricas globales, rendimiento por clase y eficiencia computacional, con especial atención a las diferencias entre categorías dominantes y minoritarias.

4.1. Rendimiento cuantitativo

La Tabla 3 muestra la precisión global, el IoU medio y la IoU por clase de los modelos evaluados, mientras que la Tabla 4 resume su coste computacional en términos de número de parámetros, tiempo de entrenamiento y tiempo de

Tabla 3. precisión global, IoU medio e IoU por clase en el conjunto de prueba. Los mejores valores de cada columna aparecen en **negrita**.

Método	OA	mIoU	Suelo	Veg. baja	Veg. med./alta	Edificio	Vehículo
KPConv	**96,16**	**78,51**	**95,11**	**33,61**	**95,12**	**93,97**	74,76
RandLA-Net	93,39	71,98	91,23	23,61	93,57	82,49	69,02
SPT	95,37	74,27	94,42	20,61	91,20	92,23	72,87
PTv3	95,79	73,56	94,68	11,23	93,10	93,69	**75,11**

Tabla 4. Eficiencia computacional. Los parámetros se reportan en millones (M) y los tiempos en minutos (min). El entrenamiento corresponde al protocolo completo indicado en la Tabla 2, y la inferencia al conjunto de prueba completo.

Modelo	Parámetros (M)	Entrenamiento (min)	Inferencia (min)
KPConv	23,26	611	40
RandLA-Net	1,11	307	17
SPT	0,21	513	19
PTv3	46,18	531	14

inferencia. En conjunto, ambas tablas permiten analizar no solo la calidad de la segmentación, sino también el compromiso entre rendimiento y eficiencia.

Todos los modelos alcanzan una precisión global elevada, entre el 93,39 % de RandLA-Net y el 96,16 % de KPConv, lo que confirma la capacidad general de las arquitecturas modernas para la segmentación de LiDAR aéreo. Sin embargo, el IoU medio muestra diferencias más marcadas, de 71,98 % a 78,51 %, lo que refleja diferencias en la consistencia entre clases.

Esta diferencia entre OA y mIoU es importante en datos operativos. Cuando una o dos clases concentran la mayor parte de los puntos, pequeñas mejoras en clases dominantes pueden ocultar fallos persistentes en categorías minoritarias. Por ello, el análisis por clase no es un complemento secundario, sino una parte necesaria para entender la utilidad real de cada arquitectura en productos geoespaciales.

KPConv obtiene el mejor rendimiento global, con 96,16 % de OA y 78,51 % de mIoU. Lidera cuatro de las cinco clases: *suelo*, *vegetación baja*, *vegetación media/alta* y *edificio*. Este comportamiento sugiere una buena capacidad para preservar detalle geométrico y fronteras locales. No obstante, también implica el mayor coste: 611 minutos de entrenamiento, 40 de inferencia y 23,26 M parámetros.

RandLA-Net alcanza el menor mIoU (71,98 %), aunque mantiene resultados competitivos en clases dominantes como *suelo* y *vegetación media/alta*. Su muestreo aleatorio reduce el coste, pero parece perder detalle geométrico en regiones minoritarias y fronteras, como muestran *vegetación baja* y *edificio*. Con 1,11 M parámetros, presenta el menor tiempo de entrenamiento y el segundo menor tiempo de inferencia, por lo que sigue siendo una alternativa interesante cuando prima la eficiencia.

Superpoint Transformer ocupa la segunda posición en mIoU (74,27 %) y ofrece un buen compromiso entre calidad y eficiencia paramétrica. Su principal limitación aparece en *vegetación baja* (20,61 %), donde el agrupamiento en superpuntos puede fusionar vegetación dispersa con el terreno. Con 0,21 M parámetros, es la arquitectura más ligera, aunque su rendimiento depende de que los superpuntos respeten adecuadamente las fronteras semánticas.

Point Transformer V3 logra el mejor resultado en *vehículo* (75,11 % de IoU), superando ligeramente a KPConv, lo que confirma el interés de la atención para objetos pequeños y aislados. En cambio, obtiene el peor valor en *vegetación baja* (11,23 %), posiblemente por la dificultad de la serialización ante patrones difusos y poco densos. Aunque tiene más parámetros (46,18 M), consigue la inferencia más rápida gracias a su organización por secuencias y parches.

La diferencia más clara aparece entre clases mayoritarias y minoritarias. Todas las arquitecturas superan el 91 % de IoU en *suelo* y *vegetación media/alta*, pero caen en *vegetación baja*, con valores entre 11,23 % y 33,61 %. Esta clase representa solo el 1,41 % de los puntos y además tiene una geometría difusa, próxima al suelo y parcialmente solapada con otras categorías. *Vehículo*, aunque es aún menos frecuente (0,68 %), alcanza IoU superiores (69,02 %–75,11 %) por su estructura más compacta y distintiva.

4.2. Análisis cualitativo

La Figura 2 muestra una escena de prueba segmentada por las cuatro arquitecturas, junto con la imagen RGB y la referencia. Todos los modelos identifican las estructuras principales, pero difieren en fronteras, clases minoritarias y regiones ambiguas. Estas diferencias permiten interpretar mejor los resultados por clase de la Tabla 3.

Entre las clases, *vegetación baja* es la más difícil. Su geometría se solapa con *suelo* y *vegetación media/alta*, formando capas finas y discontinuas. Esto la hace más ambigua que objetos compactos y geométricamente distintivos como *vehículos*, aunque estos también estén muy infrarrepresentados.

KPConv produce la segmentación más coherente: delimita bien cubiertas y fronteras de vegetación, presenta poca confusión entre clases y clasifica correctamente más puntos de *vegetación baja* que el resto de modelos. Esta observación es coherente con su mejor rendimiento medio y con su ventaja en la clase minoritaria más problemática.

RandLA-Net muestra los errores más evidentes: algunas cubiertas planas y estructuras ambiguas se clasifican erróneamente como *suelo*, y aparecen pequeñas predicciones de *vehículo* sobre edificios o superficies planas. Este comportamiento refleja la pérdida de detalle local que puede introducir el muestreo aleatorio en zonas de frontera.

Superpoint Transformer genera máscaras de *edificio* limpias y, en esta escena, segmenta bien algunas zonas de *vegetación baja*. No obstante, en el conjunto completo su rendimiento en esta clase sigue siendo bajo (20,61 % de IoU), lo que indica que este comportamiento no se mantiene de forma homogénea. También identifica correctamente una estructura de edificio compleja que resulta difícil

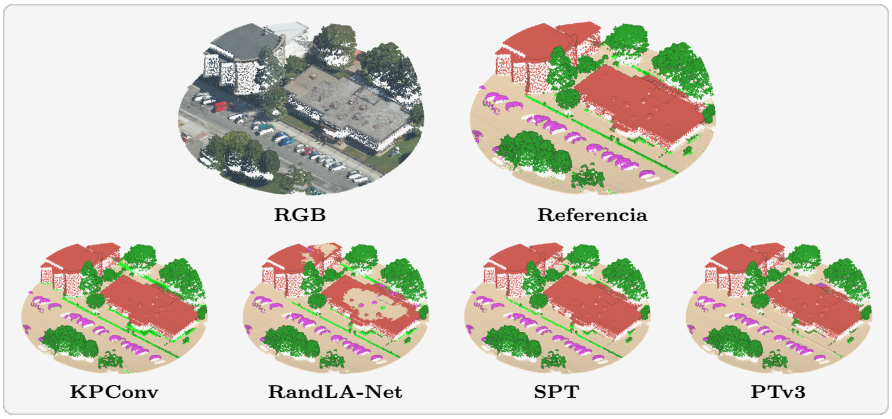


Figura 2. Ejemplo de escena segmentada por cuatro modelos de aprendizaje profundo. Arriba se muestran la escena RGB de entrada y las etiquetas semánticas de referencia. Abajo aparece la misma escena segmentada por *KPConv*, *RandLA-Net*, *Superpoint Transformer* (SPT) y *Point Transformer V3* (PTv3). Colores: suelo (beis), vegetación baja (verde intenso), vegetación media/alta (verde helecho), edificio (bermellón) y vehículo (magenta).

para otras arquitecturas, mostrando la utilidad de los superpuntos en geometrías estructuradas.

Point Transformer V3 detecta mejor los *vehículos*, beneficiándose de su amplio campo receptivo y el procesamiento de características mediante atención. Sin embargo, muestra cierta confusión entre *vegetación* y *edificio* en zonas entrelazadas y rara vez predice *vegetación baja*, que tiende a confundirse con *suelo*.

En conjunto, el análisis cualitativo confirma las diferencias observadas en precisión en fronteras de clases, predicción de clases minoritarias y preservación del detalle geométrico.

5. Conclusiones

Este trabajo evalúa KPConv, RandLA-Net, Superpoint Transformer y Point Transformer V3 sobre nubes de puntos LiDAR aéreas capturadas en condiciones operativas en Navarra, España, con entornos heterogéneos y fuerte desequilibrio entre clases. Los experimentos reflejan un escenario práctico de cartografía aérea, donde las clases no aparecen balanceadas y la densidad de puntos varía de forma natural.

Todas las arquitecturas alcanzan una precisión global elevada (>93 %), pero el IoU medio muestra diferencias más claras (71,98 %–78,51 %). Esto ocurre porque la precisión global está dominada por clases mayoritarias, mientras que mIoU pondera todas las categorías por igual. KPConv obtiene el mejor rendimiento global y destaca en *vegetación baja* (33,61 % de IoU), mientras que Point Transformer V3 funciona especialmente bien en *vehículo* (75,11 %). Superpoint

Transformer ofrece resultados competitivos con alta eficiencia en el número de parámetros, y RandLA-Net reduce el coste computacional a cambio de menor rendimiento.

Las dificultades con *vegetación baja*, con IoU entre 11,23 % y 33,61 %, reflejan la ambigüedad geométrica de las regiones próximas al suelo y el impacto del desequilibrio entre clases en vuelos operativos. Este resultado apunta a la necesidad de estrategias específicas para clases minoritarias y geométricamente ambiguas.

En conjunto, esta evaluación aporta evidencia cuantitativa y cualitativa sobre las capacidades y limitaciones actuales de estos modelos en condiciones de uso reales. También muestra que la elección de arquitectura no depende solo de métricas globales como OA o mIoU, sino del equilibrio entre rendimiento por clase y coste computacional. Como trabajo futuro, se plantea evaluar mejor la consistencia espacial, analizar patrones de error, definir particiones de validación y explorar estrategias de entrenamiento o arquitectura adaptadas al LiDAR aéreo.

Agradecimientos

Alex Salvatierra es beneficiario de un contrato predoctoral financiado por la Cátedra Tracasa de Ciencias de la Computación e Inteligencia Artificial de la Universidad Pública de Navarra. Este trabajo ha sido parcialmente financiado por el proyecto 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. Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S.: 3D Semantic Parsing of Large-Scale Indoor Spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1534–1543 (2016)
2. Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J.: SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. 2019 IEEECVF Int. Conf. Comput. Vis. ICCV pp. 9296–9306 (2019). https://doi.org/10.1109/ICCV.2019.00939
3. Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: 2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR. pp. 77–85 (2017). https://doi.org/10.1109/CVPR.2017.16
4. Choy, C., Gwak, J., Savarese, S.: 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. In: 2019 IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR. pp. 3070–3079 (2019). https://doi.org/10.1109/CVPR.2019.00319
5. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Niessner, M.: ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5828–5839 (2017)
6. Feroz, S., Dabous, S.A.: Uav-based remote sensing applications for bridge condition assessment. Remote Sens. **13**(9) (2021). https://doi.org/10.3390/rs13091809

7. Gaydon, C., Daab, M., Roche, F.: FRACTAL: An Ultra-Large-Scale Aerial Lidar Dataset for 3D Semantic Segmentation of Diverse Landscapes (2024). https://doi.org/10.48550/arXiv.2405.04634
8. Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M.: PCT: Point cloud transformer. Comp. Visual Media **7**(2), 187–199 (2021). https://doi.org/10.1007/s41095-021-0229-5
9. Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M.: SEMANTIC3D.NET: A NEW LARGE-SCALE POINT CLOUD CLASSIFICATION BENCHMARK. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **IV-1-W1**, 91–98 (2017). https://doi.org/10.5194/isprs-annals-IV-1-W1-91-2017
10. 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. In: 2020 IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR. pp. 11105–11114 (2020). https://doi.org/10.1109/CVPR42600.2020.01112
11. Lai, X., Liu, J., Jiang, L., Wang, L., Zhao, H., Liu, S., Qi, X., Jia, J.: Stratified Transformer for 3D Point Cloud Segmentation. In: 2022 IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR. pp. 8490–8499 (2022). https://doi.org/10.1109/CVPR52688.2022.00831
12. Landrieu, L., Simonovsky, M.: Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs. In: 2018 IEEECVF Conf. Comput. Vis. Pattern Recognit. pp. 4558–4567 (2018). https://doi.org/10.1109/CVPR.2018.00479
13. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B.: PointCNN: Convolution On X-Transformed Points. In: Adv. Neural Inf. Process. Syst. vol. 31. Curran Associates, Inc. (2018). https://doi.org/10.48550/arXiv.1801.07791
14. Melekhov, I., Umashankar, A., Kim, H.J., Serkov, V., Argyle, D.: ECLAIR: A High-Fidelity Aerial LiDAR Dataset for Semantic Segmentation (2024). https://doi.org/10.48550/arXiv.2404.10699
15. Moore, I., Grayson, R., Ladson, A.: Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrol. Processes **5**(1), 3–30 (1991). https://doi.org/10.1002/hyp.3360050103
16. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In: Adv. Neural Inf. Process. Syst. vol. 30. Curran Associates, Inc. (2017). https://doi.org/10.48550/arXiv.1706.02413
17. Qin, N., Tan, W., Ma, L., Zhang, D., Li, J.: OpenGF: An Ultra-Large-Scale Ground Filtering Dataset Built Upon Open ALS Point Clouds Around the World (2021). https://doi.org/10.48550/arXiv.2101.09641
18. Robert, D., Raguet, H., Landrieu, L.: Efficient 3D Semantic Segmentation with Superpoint Transformer. In: 2023 IEEECVF Int. Conf. Comput. Vis. ICCV. pp. 17149–17158 (2023). https://doi.org/10.1109/ICCV51070.2023.01577
19. Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U.: THE ISPRS BENCHMARK ON URBAN OBJECT CLASSIFICATION AND 3D BUILDING RECONSTRUCTION. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **I-3**, 293–298 (2012). https://doi.org/10.5194/isprsannals-I-3-293-2012
20. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.: KPConv: Flexible and Deformable Convolution for Point Clouds. In: 2019 IEEECVF Int. Conf. Comput. Vis. ICCV. pp. 6410–6419 (2019). https://doi.org/10.1109/ICCV.2019.00651
21. Varney, N., Asari, V.K., Graehling, Q.: DALES: A Large-scale Aerial LiDAR Data Set for Semantic Segmentation (2020). https://doi.org/10.48550/arXiv.2004.11985

22. Wang, R., Peethambaran, J., Chen, D.: LiDAR Point Clouds to 3-D Urban Models: A Review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **11**(2), 606–627 (2018). https://doi.org/10.1109/JSTARS.2017.2781132
23. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic Graph CNN for Learning on Point Clouds. ACM Trans. Graph. **38**(5), 146:1–146:12 (2019). https://doi.org/10.1145/3326362
24. Wehr, A., Lohr, U.: Airborne laser scanning—an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing **54**(2), 68–82 (1999). https://doi.org/10.1016/S0924-2716(99)00011-8
25. Wu, W., Qi, Z., Fuxin, L.: PointConv: Deep Convolutional Networks on 3D Point Clouds. In: 2019 IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR. pp. 9613–9622 (2019). https://doi.org/10.1109/CVPR.2019.00985
26. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point Transformer V3: Simpler, Faster, Stronger. In: 2024 IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR. pp. 4840–4851 (2024). https://doi.org/10.1109/CVPR52733.2024.00463
27. Wu, X., Lao, Y., Jiang, L., Liu, X., Zhao, H.: Point Transformer V2: Grouped Vector Attention and Partition-based Pooling. Adv. Neural Inf. Process. Syst. **35**, 33330–33342 (2022). https://doi.org/10.48550/arXiv.2210.05666
28. Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T.: Lidar sampling for large-area forest characterization: A review. Remote Sensing of Environment **121**, 196–209 (2012). https://doi.org/10.1016/j.rse.2012.02.001
29. Xu, Y., Du, B., Zhang, L., Cerra, D., Pato, M., Carmona, E., Prasad, S., Yokoya, N., Hänsch, R., Le Saux, B.: Advanced Multi-Sensor Optical Remote Sensing for Urban Land Use and Land Cover Classification: Outcome of the 2018 IEEE GRSS Data Fusion Contest. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **12**(6), 1709–1724 (2019). https://doi.org/10.1109/JSTARS.2019.2911113
30. Yan, W.Y., Shaker, A., El-Ashmawy, N.: Urban land cover classification using airborne LiDAR data: A review. Remote Sensing of Environment **158**, 295–310 (2015). https://doi.org/10.1016/j.rse.2014.11.001
31. Ye, Z., Xu, Y., Huang, R., Tong, X., Li, X., Liu, X., Luan, K., Hoegner, L., Stilla, U.: LASDU: A Large-Scale Aerial LiDAR Dataset for Semantic Labeling in Dense Urban Areas. ISPRS Int. J. Geo-Inf. **9**(7), 450 (2020). https://doi.org/10.3390/ijgi9070450
32. Yoo, S., Ko, C., Sohn, G., Lee, H.: YUTO SEMANTIC: A LARGE SCALE AERIAL LIDAR DATASET FOR SEMANTIC SEGMENTATION. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **XLVIII-1-W2-2023**, 209–215 (2023). https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-209-2023
33. Zachar, P., Bakuła, K., Ostrowski, W.: CENAGIS-ALS BENCHMARK - NEW PROPOSAL FOR DENSE ALS BENCHMARK BASED ON THE REVIEW OF DATASETS AND BENCHMARKS FOR 3D POINT CLOUD SEGMENTATION. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **XLVIII-1-W3-2023**, 227–234 (2023). https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-227-2023
34. Zhao, H., Jiang, L., Jia, J., Torr, P., Koltun, V.: Point Transformer. In: 2021 IEEECVF Int. Conf. Comput. Vis. ICCV. pp. 16239–16248 (2021). https://doi.org/10.1109/ICCV48922.2021.01595
35. Zolanvari, S.M.I., Ruano, S., Rana, A., Cummins, A., da Silva, R.E., Rahbar, M., Smolic, A.: DublinCity: Annotated LiDAR Point Cloud and its Applications (2019). https://doi.org/10.48550/arXiv.1909.03613