Learning with Spaceborne LiDAR for Enhancement of Bare-Earth Digital Elevation Models from Global Data

Xiandong Cai[1,2], Matthew Wilson[1,2]

[1] Geospatial Research Institute, University of Canterbury, Christchurch, New Zealand
[2] School of Earth and Environment, University of Canterbury, Christchurch, New Zealand –
(xander.cai, matthew.wilson)@canterbury.ac.nz

Keywords: Digital elevation model, Spaceborne LiDAR, Multi-modal fusion, Depth completion, Open geospatial science.

Abstract

Bare-earth Digital Terrain Models (DTMs) are fundamental to geospatial applications, yet publicly accessible global elevation products are Digital Surface Models (DSMs) with vertical accuracies of 4–15 m RMSE at 30 m or coarser resolution. Airborne LiDAR achieves sub-metre accuracy but remains costly and spatially fragmented. ICESat-2 ATL08 spaceborne LiDAR photons offer a compelling complement: unlike optical sensors, they physically penetrate dense vegetation canopies to measure ground elevation directly, and ICESat-2 acquires new measurements daily on a 91-day repeat cycle. Incorporating ATL08 into a learning framework, however, poses two challenges: extreme measurement sparsity after quality filtering and an unstructured point geometry incompatible with the dense raster structure of imagery and DSMs. We present a deep neural network that addresses both challenges by fusing remote sensing imagery, Copernicus GLO-30 DSMs, and ATL08 photons to generate 3 m bare-earth elevation from global data. A multi-scale deformable cross-attention mechanism fuses each photon with surrounding image and DSM features in their native geometry—without rasterisation—thereby turning sparse measurements into effective guidance features. A Spatial Propagation Network (SPN) then densifies predictions guided by image structure, re-anchored at photon locations at every iteration. Evaluated on the DFC30 benchmark augmented with ATL08 measurements, our method achieves consistent improvements over optical-only baselines across all vegetation classes, improving overall RMSE by 74.5% over Copernicus GLO-30 (cubically interpolated to 3 m spatial resolution), with the largest gains in dense forest canopy (83.74% RMSE reduction) where ground elevation is most difficult to recover.

1. Introduction

Accurate bare-earth topographic data are foundational to geospatial science, driving hydrological modelling, natural hazard assessment, carbon stock estimation, and infrastructure planning. Current publicly accessible global elevation products – SRTM (Farr et al., 2007), ASTER GDEM (Tachikawa et al., 2011), and Copernicus DEM (European Space Agency et al., 2010–2014) – are Digital Surface Models (DSMs) that include vegetation canopies and built structures, with vertical accuracies of 4–15 m RMSE at 30 m or coarser resolution (Meadows et al., 2024). DEM vertical accuracy and resolution directly impact flood inundation mapping, with coarser, less accurate DEMs leading to significant over- or underestimation of inundation extent (Saksena and Merwade, 2015), illustrating the real-world consequences of this limitation.

Airborne LiDAR achieves sub-metre accuracy but is prohibitively expensive at scale—a wall-to-wall global survey has been estimated to cost £42 billion for a single acquisition (Hancock et al., 2021). Even in well-resourced countries, coverage remains incomplete: approximately 20% of New Zealand's land surface lacks airborne LiDAR coverage (Toitū Te Whenua Land Information New Zealand, 2023), disproportionately in the most remote and ecologically significant terrain where accurate elevation models are most critical.

Deep learning super-resolution methods guided by optical imagery have substantially improved global DSM accuracy. JSPSR (Joint Spatial Propagation Super-Resolution Networks) (Cai and Wilson, 2025) reduces elevation RMSE by over 70% versus Copernicus GLO-30 through spatial propagation networks. However, optical sensors cannot penetrate vegetation canopies, imposing a fundamental accuracy ceiling in forested terrain.

ICESat-2 ATL08 spaceborne LiDAR offers a direct solution: photons physically penetrate canopy to measure ground elevation, with validated terrain RMSE of 2.35 m in boreal forests (Feng et al., 2023) and continuous global coverage refreshed every 91 days (Neuenschwander and Pitts, 2019). Incorporating ATL08 into a super-resolution framework poses two structural challenges: after quality filtering, usable photons cover fewer than 0.1% of dense-forest mountain areas (Tian and Shan, 2021); and their unstructured point geometry cannot be directly handled by standard raster-based architectures.

This paper addresses both challenges through a triple-branch multi-modal fusion network. The key contributions are:

1. A triple-branch multi-modal fusion network incorporating a multi-scale deformable cross-attention mechanism that fuses sparse ATL08 photon measurements with dense imagery and DSM features in their native geometry, addressing the dual challenges of extreme sparsity and unstructured point geometry.
2. An augmented benchmark dataset built on DFC30, enriched with spatially matched ICESat-2 ATL08 photon measurements, demonstrating that globally available, freely accessible data alone are sufficient for high-accuracy bare-earth DTM generation.
3. A fully open-source implementation of the complete pipeline, including ATL08 preprocessing tools, trained model weights, and training scripts.

2. Related Work

2.1 Global DEM Enhancement

Early deep learning approaches to DEM super-resolution directly transferred image super-resolution networks (SRCNN, SRGAN) to elevation data, but without terrain-specific design, these models struggled to reconstruct physically meaningful structures such as ridgelines and drainage networks. To address this, Ma et al. (Ma et al., 2023) introduced feature-enhanced networks that explicitly capture long-range terrain context, and JSPSR (Cai and Wilson, 2025) incorporated spatial propagation networks for depth completion that refine predictions along image-consistent paths—reducing RMSE by over 70% versus the Copernicus GLO-30 baseline. However, all these methods rely solely on optical imagery as guidance. Since optical sensors cannot penetrate vegetation canopies, they systematically overestimate ground elevation in forested terrain, leaving a fundamental accuracy ceiling that imagery-guided methods alone cannot overcome.

2.2 Spaceborne LiDAR for Terrain and Canopy Mapping

Spaceborne LiDAR directly addresses the canopy-penetration problem. ICESat-2 ATL08, validated to achieve terrain RMSE of 2.35 m in boreal forests (Feng et al., 2023) and RMSE of 0.73 m in southern Finland (Neuenschwander et al., 2020), provides globally available sub-canopy elevation measurements with full coverage at all latitudes, complementing GEDI (Dubayah et al., 2020), which is limited to 51.6°N–51.6°S. A growing number of works have exploited this by fusing ATL08 or GEDI measurements with optical imagery to produce wall-to-wall canopy height maps using deep learning (Lang et al., 2023) – demonstrating that sparse spaceborne LiDAR can serve as an effective supervision signal when combined with dense image features. However, these studies target canopy structure rather than bare-earth terrain and treat the fusion problem in a simplified way: LiDAR footprints are typically aggregated or rasterised before being fed into the network, thereby discarding the precise sub-pixel locations of individual measurements. This rasterisation bottleneck motivates a more geometry-aware fusion strategy.

2.3 Multi-modal Fusion in Remote Sensing

Transformer-based cross-attention has become the dominant paradigm for fusing heterogeneous remote sensing modalities, with architectures such as those of Roy et al. (Roy et al., 2023) showing strong performance on paired optical-SAR and optical-LiDAR raster inputs. Yet these methods still assume co-registered dense raster grids, and handling truly sparse, non-grid-aligned point data remains an open problem. Deformable DETR (Zhu et al., 2020) offered a key insight from the computer vision field: rather than attending to the full feature map, deformable attention allows each query to sample from a small set of learned offset locations, making attention tractable for sparse inputs. We build on this mechanism and adapt it to the geospatial context, enabling individual ATL08 photons to query surrounding image and DSM features in their native point geometry, directly closing the gap identified above.

3. Methodology

3.1 Network Architecture

The proposed network follows a late-fusion, triple-branch U-Net-derived architecture, as shown in Figure 1. After preprocessing, the image branch and DSM branch each use a Swin Transformer V2 (Liu et al., 2022) encoder to independently extract hierarchical feature maps at four scales (C, 2C, 4C, and 8C channels at resolutions of 1, 1/2, 1/4, and 1/8, where C is the configurable channel number). The photon branch encodes the discrete photon array using an MLP encoder with position encoding, producing photon embeddings at channel dimension C. At each encoder scale, photon features are fused with the corresponding Image and DSM features via deformable cross-attention (Section 3.2). In the decoder, features are progressively upsampled and fused with encoder skip connections at each scale via the same mechanism, recovering full resolution and producing a unified Guidance Features map. This, together with the Initial DSM and Photon Table, is passed to the SPN for final DTM prediction (Section 3.3).

3.2 Photon Feature Fusion via Deformable Cross-Attention

Fusing sparse photon measurements with dense raster features is the central challenge: standard attention across a full feature map is computationally prohibitive and spatially imprecise. We address this with a multi-scale deformable cross-attention mechanism, as shown in Figure 2, applied at all four scales in both encoder and decoder. Photon features serve as queries, predicting learned sampling offsets into the surrounding image and DSM features rather than attending to the entire map. This allows each photon to adaptively sample only spatially relevant context. The enriched photon representations are then injected back into the dense feature maps via a learned gate, so that sub-canopy elevation information actively shapes raster features throughout the network.

3.3 SPN Refinement with Photon Re-injection

The SPN (Figure 3) takes three inputs: Guidance Features, the Initial DSM, and the Photon Table. From the fused Guidance Features and Initial DSM, the SPN predicts per-pixel spatial sampling offsets and aggregation weights, which are applied via a modulated deformable convolutional layer to produce an elevation residual, propagated in a non-local, content-adaptive manner along image-consistent structures such as ridgelines and vegetation boundaries. ATL08 photon measurements are then re-injected through a learnable gating mechanism, and the final DTM is reconstructed by combining the predicted residual, the Initial DSM, and the photon-injected correction.

4. Experiments and Results

4.1 Dataset

We evaluate our framework on the 3 m-resolution subset of the DFC30 dataset (Cai and Wilson, 2025), a benchmark built on the GRSS Data Fusion Contest 2022 (DFC2022) dataset, which provides paired high-resolution remote sensing imagery and airborne LiDAR ground truth across diverse land-cover types in France. DFC30 further supplements DFC2022 by using Copernicus COP30 as the low-resolution DSM input, along with FABDEM and FathomDEM as additional reference elevation surfaces. Each sample has a spatial extent of 334×334 pixels, from which four 256×256 patches are extracted via tiled cropping, yielding an effective training set of 12,728 and a test set of 3,196 patches. All elevation values, including DSM patches and ATL08 photon measurements, are normalised prior to training using relative elevation log-scale normalisation following JSPSR.

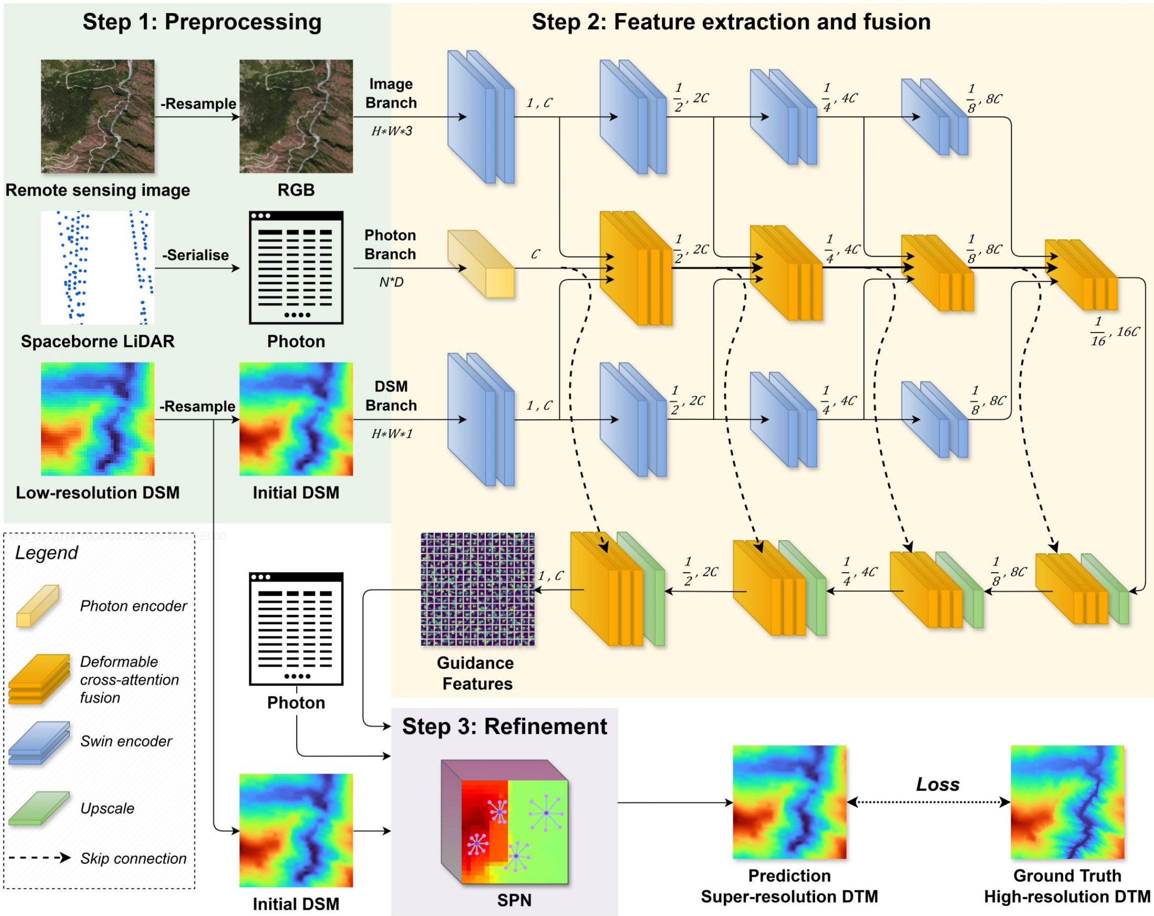

Figure 1. The proposed network workflow: (1) resample each sample to target resolution, including low-resolution DEM, high-resolution DEM, remote sensing image and photon data; (2) fuse input multi-modal data and extract guidance features in the backbone; (3) utilising fused guidance features, the spatial propagation parameters in the Spatial Propagation Network (SPN) (Liu et al., 2017) optimise their weights during training and refine initial DSMs as prediction during inferring.

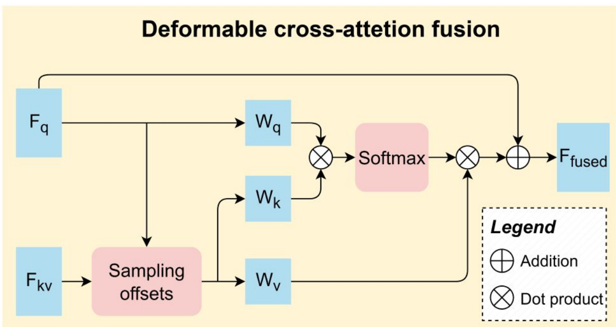

Figure 2. Deformable cross-attention fusion module takes encoded features from different modalities at various scales and fuses them using a cross-attention mechanism. In the encoding (downscaling) phase, the F_q are photon features to query rich information from imagery and DSM features. In the decoding (upscaling) phase, the F_q are DSM features to retrieve high-accuracy elevation information from enhanced photon features.

We augment DFC30 with spatially matching ICESat-2 ATL08 measurements retrieved and quality-filtered for each tile. To avoid excessive sparsity, the 20 m segment resolution of ATL08 is selected over the 100 m default, and measurements are subjected to a two-stage filtering process: a coarse filter based on the ATL08 quality indicators `te_quality_score` (terrain quality score) and `n_te_photons` (number of terrain photons) to remove low-confidence measurements, followed by dynamic outlier rejection using Median Absolute Deviation (MAD) against FathomDEM as a reference surface to remove large elevation errors. This strategy retains substantially more photons than standard quality filtering alone while maintaining measurement reliability, yielding 869,800 valid photon measurements from 1,065,516 candidates across the full dataset, with an average of 218.5 photons per sample (approximately 0.196% pixel coverage). Note that FathomDEM's role in the filtering pipeline precludes its use as an independent baseline for comparison of results.

4.2 Implementation Details

All models are trained on an NVIDIA RTX 5090 GPU using the AdamW optimiser with a step learning-rate schedule (initial lr = 1×10^{-4}, weight decay = 0.05) for 100 epochs. Swin transformer V2 encoders are initialised from ImageNet-22K pre-trained weights. Data augmentation includes random horizontal and vertical flipping, and random rotation at 90° increments. The full framework and training scripts are implemented in PyTorch and will be released under the Apache 2.0 licence. Performance is evaluated using the Root Mean Square Error (RMSE) between predicted elevation and airborne LiDAR ground truth, categorised by land use class.

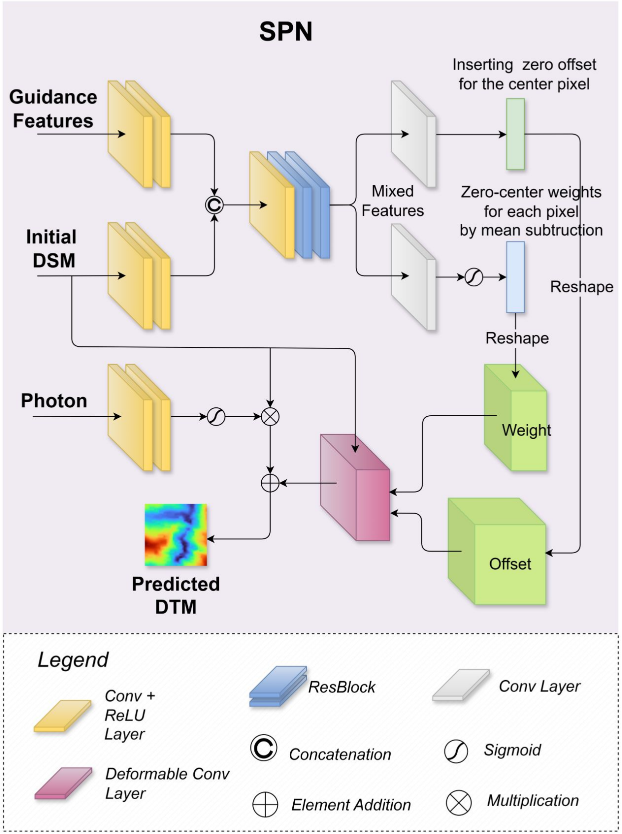

Figure 3. The refinement module structure. It first generates offset and weight predictions for each pixel in the initial DSM using mixed features. Then, it utilises a modulated deformable convolutional layer to sample neighbouring pixels using offsets and weights, thereby predicting the SR DEM in a non-local manner. Finally, photon data is re-injected into the initial DSM, incorporating the output of the deformable convolutional layer to generate the DTM prediction.

The network is trained with a three-component weighted loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sparse}} + \lambda_{\text{aux}} \cdot \mathcal{L}_{\text{aux}} + \lambda_{\text{tv}} \cdot \mathcal{L}_{\text{tv}} \tag{1}$$

$\mathcal{L}_{\text{sparse}}$ is computed only at ATL08 photon locations via bilinear sampling, constraining the network only where the ground truth is reliable and preventing erroneous gradients in photon-free regions. Each photon's loss is weighted by its quality score $q \in [0, 1]$:

$$\mathcal{L}_{\text{sparse}} = \frac{1}{M} \sum_{i=1}^{M} (q_i + 0.5) \cdot |\hat{h}_i - h_i| \tag{2}$$

$\mathcal{L}_{\text{aux}}$ is the multi-scale deep supervision loss (Lee et al., 2015) computed at each decoder stage k against the downsampled ground truth, with deeper stages assigned lower weights:

$$\mathcal{L}_{\text{aux}} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{k} \cdot \left\| \hat{D}_k - D_k^{\downarrow} \right\|_1 \tag{3}$$

$\mathcal{L}_{\text{tv}}$ is an L1 total variation regularisation term (Rudin et al., 1992) that penalises high-frequency artefacts in photon-free regions:

$$\mathcal{L}_{\text{tv}} = \frac{1}{HW} \sum_{i,j} \left(|\hat{D}_{i+1,j} - \hat{D}_{i,j}| + |\hat{D}_{i,j+1} - \hat{D}_{i,j}| \right) \tag{4}$$

The weights are set to $\lambda_{\text{aux}} = 0.1$ and $\lambda_{\text{tv}} = 0.01$.

4.3 Quantitative Results

Table 1 compares our method against JSPSR and two cubic-interpolated 3 m baseline DEMs: BaseCOP (Copernicus COP30, which also serves as the DSM input to our network) and BaseFAB (FABDEM). Our method achieves consistent improvements across all land use types, with the largest gains in dense forest, where optical-only methods are most severely limited. DEMs are classified pixel-wise into categories using land-use data from DFC30. Our method reduces the overall RMSE from 3.536 m (BaseCOP) to 1.085 m, a 74.5% improvement, and from 9.36 m to 1.522 m in the forests, an 83.74% improvement.

Land use class	RMSE			
	BaseCOP	BaseFAB	JSPSR	Ours
No information	3.844	1.352	1.336	**1.283**
Urban fabric	1.37	0.755	**0.671**	0.688
Industrial, etc., Units	1.741	0.873	0.902	**0.857**
Construction, etc., sites	2.884	**2.822**	2.9	2.876
Artifical vegetated areas	4.25	2.404	1.547	**1.2**
Arable land (annual crops)	0.936	0.925	0.563	**0.521**
Permanent crops	0.778	0.767	**0.526**	0.548
Pastures	1.502	0.782	0.618	**0.625**
Forests	9.36	3.417	2.11	**1.522**
Herbaceous vegetation	2.207	1.136	1.05	**0.924**
Open spaces	2.539	2.62	1.981	**1.69**
Wetlands	1.082	0.665	**0.529**	0.532
Water	2.21	2.089	1.88	**1.712**
All classes	3.536	1.682	1.154	**1.085**

Metric[Unit]: RMSE[m]

Table 1. Elevation RMSE at 3 m spatial resolution across classes. Best results in **bold**.

The largest improvements occur precisely where airborne LiDAR data gaps are most consequential: in forests with dense understory and complex terrain, where ATL08 photons, fused via deformable cross-attention, provide direct sub-canopy elevation constraints that optical-only methods fundamentally cannot recover.

Figure 4 visualises the elevation error distribution and RMSE. Compared to BaseFAB and JSPSR, our method has a narrower error distribution and obtains better performance.

4.4 Qualitative Results

Figure 5 shows elevation profiles extracted along transects through forested areas, comparing BaseCOP, BaseFAB, our method, and airborne LiDAR ground truth. To maintain visual clarity and avoid overlapping curves that would reduce discriminability, JSPSR is excluded from the profile plots. The profiles illustrate the characteristic behaviour of each method. In open terrain, all methods produce visually similar results, consistent with the quantitative findings. The differences become pronounced in vegetated areas. BaseCOP and BaseFAB capture broad terrain structure but produce systematically elevated surfaces in dense forest, following the canopy rather than the ground. Our method closely tracks the LiDAR ground truth

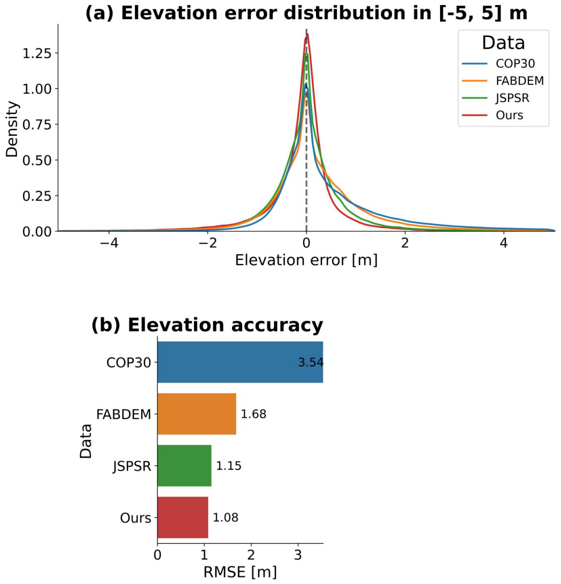

Figure 4. Comparison between baselines and the proposed methods. (**a**) Elevation error distribution between −5 m and 5 m. (**b**) Elevation vertical accuracy.

profile where photons appear, guiding the predicted elevations towards the true ground level beneath the dense canopy. The profile also highlights the role of SPN refinement, in which corrections initiated at sparse photon locations propagate laterally along terrain-consistent paths, thereby recovering accurate ground elevation even in areas between photon tracks where no direct measurements exist.

5. Discussion

Our results demonstrate that spaceborne LiDAR, despite its extreme sparsity in challenging terrain, provides information that optical imagery fundamentally cannot supply: direct sub-canopy elevation measurements. The deformable cross-attention mechanism is critical to extracting value from this sparse signal—standard rasterisation of ATL08 photons degrades accuracy by introducing interpolation artefacts that the network must subsequently unlearn.

This work is not intended to replace commercial high-resolution elevation products where they exist. Rather, it targets the many regions where such data are unavailable, unaffordable, or insufficiently up to date, precisely the contexts where global open datasets such as Copernicus DEM, Sentinel-2, and ICESat-2 represent the only viable source of terrain information. The operational implications for New Zealand are significant: the trained model enables 3 m DTM generation for approximately 268,000 km^2 of unmapped national land surface, supporting applications including landslide susceptibility mapping, hydrological network delineation, carbon stock assessment in native forest, and rural infrastructure planning.

A current limitation is that the ATL08 track density is uneven across the Southern Alps. Some high-relief catchments contain fewer than 10 photon measurements per 1 km^2 tile after quality filtering, which constrains the benefit of the photon branch in these areas. More broadly, ATL08 photon quality and quantity degrade in high-slope terrain, where geometric distortion and signal noise reduce the reliability of terrain retrievals—a known limitation of the ATL08 algorithm that affects precisely the rugged mountain areas where accurate bare-earth elevation is most needed.

Future work will investigate incorporating ICESat-2 ATL06 (land ice) and ATL03 (raw photon cloud) products to augment coverage, and will explore uncertainty quantification to identify regions where predictions should be treated with greater caution.

We note that the results presented here reflect an early iteration of the framework. Ablation studies and more comprehensive evaluations are ongoing and will be reported in a forthcoming full journal publication.

6. Conclusion

This paper presented an open-source deep learning framework for generating 3 m bare-earth elevation by fusing Copernicus GLO-30 DSMs, remote sensing imagery, and ICESat-2 ATL08 spaceborne LiDAR photons. A multi-scale deformable cross-attention mechanism addresses the geometric mismatch between sparse photon measurements and dense raster data, while a sparse-to-dense progressive supervision strategy enables effective learning from extremely sparse photon constraints (0.196% pixel coverage and 218.5 photons per sample on average). Evaluated on the DFC30 benchmark in 3 m spatial resolution, our method outperforms all baselines across the majority of land use classes, with the most significant gains in forested areas where RMSE is reduced by 83.74%, demonstrating that ICESat-2 ATL08 photons provide sub-canopy elevation information that optical-only methods fundamentally cannot recover. All code, trained weights, and preprocessing tools are released under an open-source licence to support reproducibility and community adaptation.

https://doi.org/10.5194/isprs-archives-L-4-W1-2026-27-2026 | © Author(s) 2026. CC BY 4.0 License.

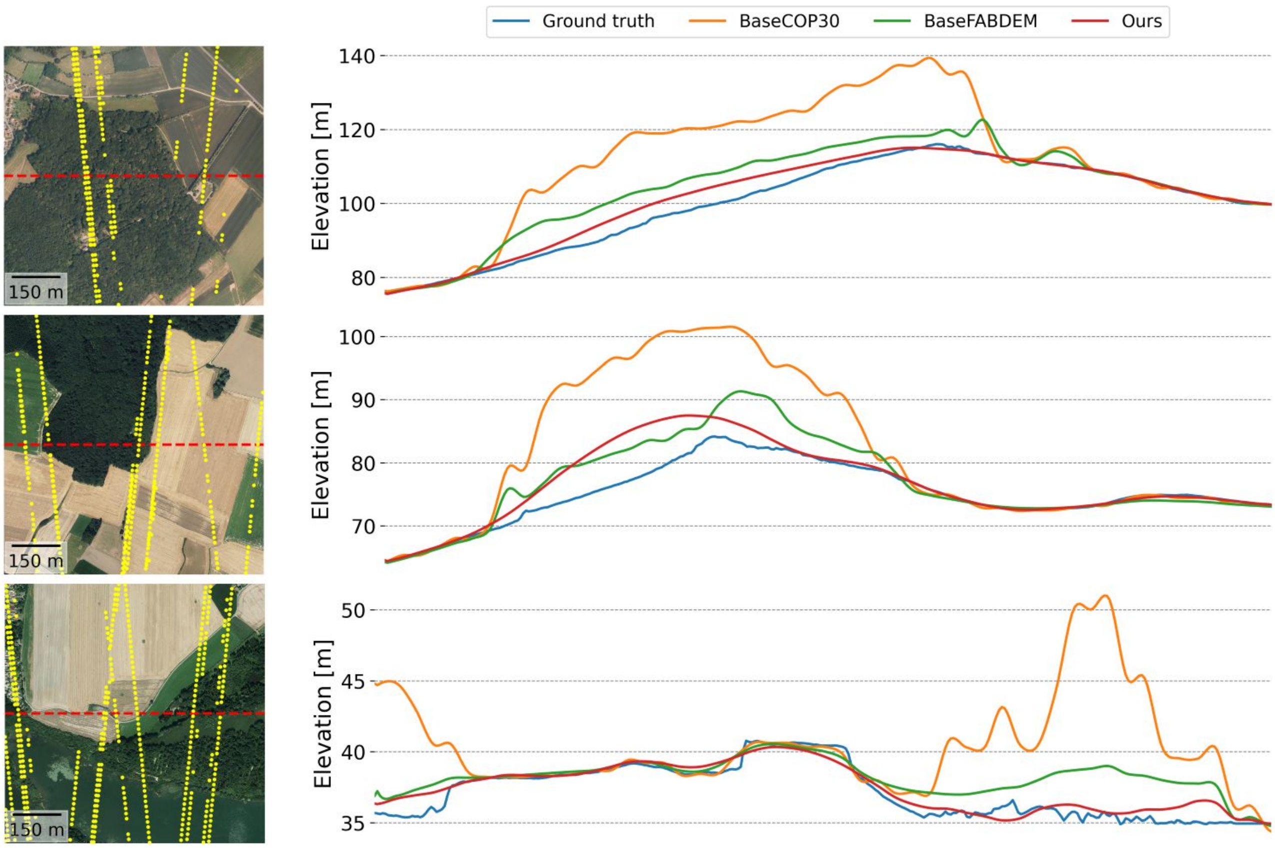

Figure 5. Elevation profiles examples. The red-dashed lines are elevation cross-section projections on the remote sensing images. The yellow points are the unfiltered ATL08 photons.

References

Cai, X., Wilson, M. D., 2025. JSPSR: Joint Spatial Propagation Super-Resolution Networks for Enhancement of Bare-Earth Digital Elevation Models from Global Data. *Remote Sensing*, 17(21), 3591.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S. et al., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth’s forests and topography. *Science of remote sensing*, 1, 100002.

European Space Agency, Airbus Defence and Space GmbH, DLR e.V., 2010–2014. Copernicus digital elevation model (dem) - global and european digital elevation model. ESA/Copernicus Data Space Ecosystem. doi.org/10.5270/ESA-c5d3d65.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L. et al., 2007. The shuttle radar topography mission. *Reviews of geophysics*, 45(2).

Feng, T., Duncanson, L., Montesano, P., Hancock, S., Minor, D., Guenther, E., Neuenschwander, A., 2023. A systematic evaluation of multi-resolution ICESat-2 ATL08 terrain and canopy heights in boreal forests. *Remote sensing of environment*, 291, 113570.

Hancock, S., McGrath, C., Lowe, C., Davenport, I., Woodhouse, I., 2021. Requirements for a global lidar system: spaceborne lidar with wall-to-wall coverage. *Royal Society open science*, 8(12), 211166.

Lang, N., Jetz, W., Schindler, K., Wegner, J. D., 2023. A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*, 7(11), 1778–1789.

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z., 2015. Deeply-supervised nets. *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 562–570.

Liu, S., De Mello, S., Gu, J., Zhong, G., Yang, M.-H., Kautz, J., 2017. Learning affinity via spatial propagation networks. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (eds), *Advances in Neural Information Processing Systems*, 30, Curran Associates, Inc.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L. et al., 2022. Swin transformer v2: Scaling up capacity and resolution. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12009–12019.

Ma, X., Li, H., Chen, Z., 2023. Feature-enhanced deep learning network for digital elevation model super-resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 5670–5685.

Meadows, M., Jones, S., Reinke, K., 2024. Vertical accuracy assessment of freely available global DEMs (FABDEM, Copernicus DEM, NASADEM, AW3D30 and SRTM) in flood-prone environments. *International Journal of Digital Earth*, 17(1), 2308734.

Neuenschwander, A., Guenther, E., White, J. C., Duncanson, L., Montesano, P., 2020. Validation of ICESat-2 terrain and canopy heights in boreal forests. *Remote Sensing of Environment*, 251, 112110. doi.org/10.1016/j.rse.2020.112110.

Neuenschwander, A., Pitts, K., 2019. The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sensing of Environment*, 221, 247–259. doi.org/10.1016/j.rse.2018.11.005.

Roy, S. K., Deria, A., Hong, D., Rasti, B., Plaza, A., Chanussot, J., 2023. Multimodal fusion transformer for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–20.

Rudin, L. I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4), 259–268. doi.org/10.1016/0167-2789(92)90242-F.

Saksena, S., Merwade, V., 2015. Incorporating the effect of DEM resolution and accuracy for improved flood inundation mapping. *Journal of Hydrology*, 530, 180–194.

Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D. B., Oimoen, M. J., Zhang, Z., Danielson, J. J., Krieger, T., Curtis, B., Haase, J. et al., 2011. ASTER global digital elevation model version 2-summary of validation results. *USGS Report*, 37.

Tian, X., Shan, J., 2021. Comprehensive evaluation of the ICESat-2 ATL08 terrain product. *IEEE Transactions on Geoscience and Remote Sensing*, 59(10), 8195–8209.

Toitū Te Whenua Land Information New Zealand, 2023. Elevation data. www.linz.govt.nz/products-services/data/types-linz-data/elevation-data (1 December 2025).

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J., 2020. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*.