RoofSense: A Multimodal Semantic Segmentation Dataset for Roofing Material Classification

Dimitris Mantas[1], Weixiao Gao[1], Hugo Ledoux[1]

[1] Delft University of Technology, The Netherlands - dimitris.mantas@outlook.com, (w.gao-1, h.ledoux)@tudelft.nl

Keywords: Aerial Imagery, Lidar, Data Fusion, Roofing Materials, Roofing Material Classification, Semantic Segmentation.

Abstract

Roofing material classification is critical for urban sustainability, energy efficiency, public health, environmental protection, and regulatory compliance. Despite the need for scalable solutions, existing approaches are hindered by reliance on oftentimes expensive and rare multi- or hyper-spectral satellite imagery, application-specific assumptions and biases, and oversight of deep learning and multimodal data fusion. This paper addresses these gaps by introducing RoofSense, a multimodal semantic segmentation dataset for roofing material classification in diverse urban contexts, leveraging 8 cm aerial true-color imagery and airborne laser scanning data. Representing eight classes and encompassing over 138 ha and 480 buildings across five Dutch cities, RoofSense is the largest publicly available dataset of its kind. By combining spectral and geometric information at the pixel level and adopting a novel weighting scheme to address class imbalance, RoofSense can be used to achieve competitive classification and segmentation performance in downstream tasks. This was demonstrated in a comprehensive purpose-designed benchmarking experiment with an off-the-shelf model based on ResNet-18-D and DeepLabv3+. Although lidar-derived features improved performance in difficult classes and materials commonly used on pitched roofs, results were sensitive to material and building context, clutter, and modality alignment, indicating that the theoretical benefits of data fusion are not straightforward. The implementation is publicly accessible at https://github.com/DimitrisMantas/RoofSense.

1. Introduction

Roofing material classification is vital for urban sustainability and building management, addressing energy efficiency, public health, and environmental protection (Abbasi et al., 2022). Material choices influence the urban heat island effect (Ilehag et al., 2018), wind turbulence (Zheng et al., 2021), and regulatory mandates (Abbasi et al., 2022) necessitate mapping for retrofitting (Gibril et al., 2017). Although essential, comprehensive material inventories are scarce, as in situ audits are expensive and limited in application and scale (Abriha et al., 2018).

Existing datasets are constrained by two-dimensional (2D) true-color (RGB) imagery and application-specific assumptions and biases, oftentimes representing pseudo-labelled materials (e.g., "Material 1"; Tommasini et al. 2019) or small study areas (Samsudin et al., 2016). Furthermore, the performance potential of incorporating lidar-derived features data to capture three-dimensional (3D) structural features (e.g., elevation, slope, etc.) remains largely underexplored (Hamedianfar et al., 2014b; Norman et al., 2020). Image classification methods typically assign a single label to each input scene, making material delineation difficult and unsuitable for detailed mapping. Object-based image analysis (OBIA) addresses this issue, but requires data- and study area-specific tuning which can lead to improper segmentation if suboptimal (Hamedianfar et al., 2014a). Although semantic segmentation does not inherently have such issues, it is often ignored, as most relevant works favour classical machine learning models, which are inefficient in dense prediction tasks (Feng and Fan, 2021).

This paper addresses these research gaps by introducing RoofSense, a multimodal semantic segmentation dataset for roofing material classification in diverse urban contexts, leveraging 8 cm aerial RGB imagery and airborne laser scanning (ALS) data. Representing eight classes and encompassing over 138 ha and 480 buildings across five Dutch cities, RoofSense is the largest publicly available dataset of its kind. By combining spectral and geometric information at the pixel level and adopting a novel weighting scheme to address class imbalance, RoofSense was used to achieve competitive classification and segmentation performance in a purpose-designed, comprehensive benchmarking experiment with an off-the-shelf model based on ResNet-18-D (He et al., 2019) and DeepLabv3+ (Chen et al., 2018). While lidar-derived features improved performance in difficult classes and materials commonly used on pitched roofs, results were sensitive to material and building context, clutter, and modality alignment, indicating that the theoretical benefits of data fusion are not straightforward.

2. Related Work

2.1 Roofing Material Classification Datasets

Existing roofing material classification datasets primarily use aerial RGB imagery or multi- or hyper-spectral satellite products, with spatial resolutions for the former datasets typically ranging between 5 cm and 25 cm (Abbasi et al., 2022). Although satellite imagery provides superior spectral resolution, its limited availability and high cost have popularised more common modalities, such as near-infrared (Ilehag et al., 2018). Recent studies indicate that RGB imagery can achieve performance comparable to hyperspectral products when combined with deep learning (DL) (Krówczyńska et al., 2020). Furthermore, incorporating lidar-derived features, such as digital surface models, intensity, and slope, with optical imagery shows considerable promise. Specifically, pixel-level fusion demonstrates accuracy improvements ranging between 8% and 25% in difficult classes (Hamedianfar et al., 2014b; Norman et al., 2020). However, existing datasets are limited in material and spatial coverage, use inconsistent class definition and annotation protocols, and inadequately address inter- and intra-class variability. This paper addresses these research gaps by

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W6-2025-153-2025 | © Author(s) 2025. CC BY 4.0 License.

introducing RoofSense, a large-scale, multimodal semantic segmentation dataset providing broad material coverage across diverse urban contexts. It overcomes the shortcomings of relevant works by adopting comprehensive class definition and annotation schemes and thoroughly investigating the fusion of high-resolution RGB imagery with ALS data.

2.2 Roofing Material Classification Methods

Existing roofing material classification methods are image-(Raczko et al., 2022), object-(Trevisiol et al., 2022), or pixel-based (Abriha et al., 2018). Image-based methods typically assign a single label to each input scene, requiring minimal labelling effort. However, capturing material variations within a single building or roof segment requires special measures, making this method unsuitable for granular mapping. Even then, the corresponding material may be difficult to delineate. OBIA addresses this issue by first segmenting each input scene into superpixels which it then operates on as described above, but it requires data- and study area-specific tuning which can lead to improper segmentation if suboptimal. Furthermore, the superpixels continue to suffer from label localisation issues. Pixel-based, more commonly known as semantic segmentation, methods, effectively solved these issues by providing dense predictions, which can later be aggregated, as required. Although relevant DL models, such as convolutional neural networks, have addressed past difficulties in image processing and feature extraction, most relevant works favour classical machine learning models, which are inefficient in dense prediction tasks. This paper addresses these research gaps by using RoofSense to achieve competitive classification and segmentation performance in a comprehensive purpose-designed benchmarking experiment, demonstrating the contemporary relevance of the underlying method.

3. The RoofSense Dataset

This paper introduces RoofSense (Figure 1), a multimodal semantic segmentation dataset for roofing material classification in diverse urban contexts. RoofSense comprises 300 images and corresponding ground truth masks spanning Den Hoor, Dordrecht, Enschede, Hoofddorp, and Papendrecht, the Netherlands. The images were extracted from five parent rasters, corresponding to randomly sampled 3DBAG (Peters et al., 2022) tiles (Section 3.4). The total annotated area is approximately 138.58 ha, spanning 488 buildings, making RoofSense the largest publicly available dataset of its kind. Each image in RoofSense is a seven-band $512 \times 512\,\text{px}^2$ raster. The first three bands form its RGB component (Section 3.1), while the latter four derive from ALS data (Section 3.2). These components were fused at the pixel level with respect to the RGB constituent (Section 3.3). Cells on the exterior of the corresponding building footprints, referred to as the background, have been masked. Finally, each mask provides an integer mapping for each pixel to one of eight roofing materials (Section 3.5) or the background. The semi-automated annotation process (Section 3.6) required circa 80 hours.

3.1 True-colour Component Construction

The RGB component of each parent image was constructed using BM5, a dataset containing aerial RGB imagery of the Netherlands at a ground sampling distance (GSD) of 8 cm (Het Waterschapshuis, 2023b). Relevant images were cropped to the extent of the corresponding LoD2.2 (Level of Detail; Biljecki et al. 2016) building footprints, extracted from 3DBAG, and the resulting images were then collated at their original GSD.

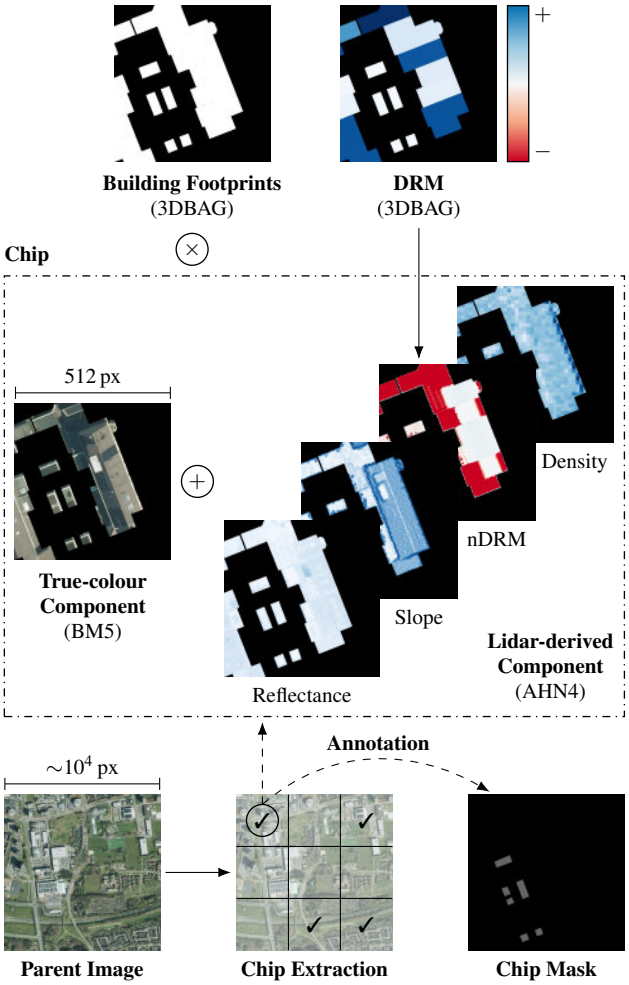


Figure 1. Construction of a single image-mask pair. The corresponding parent image is first constructed and split into chips. The RGB and lidar-derived components of the image are fused at the pixel level. The nDRM band relies on auxiliary per-building attributes (DRM), extracted from 3DBAG. The building footprints used to mask the background are also contained in this dataset. Finally, the RGB component of the chip is annotated.

3.2 Lidar-derived Component Construction

To address the limitations of 2D RGB imagery and improve the discrimination of visually similar materials, lidar-derived features were incorporated. The lidar-derived component of each parent image was constructed using the AHN4 point cloud (Het Waterschapshuis, 2023a) and relevant 3DBAG attributes, capturing 3D structural features (e.g., elevation, slope, etc.) and complementing visual cues with reflectance information.

3.2.1 Point Cloud Preprocessing: To minimise computational overhead and ensure correct density calculations, the relevant point cloud tiles were cropped, merged, and duplicate points were removed (Figure 2a). In the context of this paper, points with identical X and Y data records were considered duplicates. Duplicate sets were resolved by preserving the point with the largest z-coordinate.

https://doi.org/10.5194/isprs-annals-X-4-W6-2025-153-2025

3.2.2 Point Cloud Rasterisation: Once the point cloud was processed, its elevation and reflectance (RIEGL, 2017) fields were rasterised using inverse distance weighting (IDW) interpolation at the centre of each target cell with a power of two. Only points whose 2D projection intersected the cell were considered. The target GSD was set to 24 cm. This ensured spatial alignment between the components while maximising lidar-derived feature exploitation. To eliminate NaN values and avoid relevant processing issues, empty cells were filled with a single targeted IDW pass using GDAL.

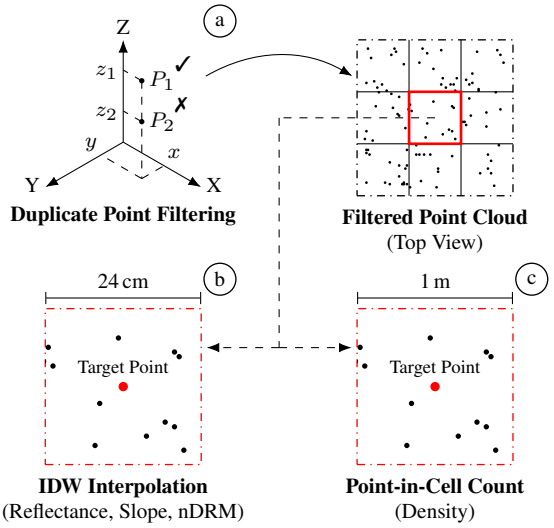


Figure 2. Point cloud preprocessing and rasterisation. First, duplicate point sets are resolved by preserving the point with the largest z-coordinate. Subsequently, the preprocessed point cloud is rasterised. The reflectance, slope, and nDRM bands are constructed using IDW interpolation considering the points whose 2D projection intersects each target cell (b). For the density band, the relevant points are counted, and their sum is assigned to the target point (c).

3.2.3 Reflectance, Slope, and nDRM Band Construction: Following rasterisation, the first three component bands were constructed: (1) The **reflectance** band, serving as a range-normalised proxy of the titular material property, was produced by converting the source raster to a linear scale and clipping values corresponding to non-Lambertian reflectors (i.e., ¿1) to one. This modification ensured correct upsampling to the GSD of the RGB component and guaranteed that such values would not be present in the resampled band; (2) The **slope** band was derived from the elevation raster using Zevenbergen and Thorne (1987); (3) The normalized digital roof model (**nDRM**), a quasi-normalised elevation model relative to the median roof level of each corresponding building, was also constructed using the elevation raster as well as the relevant rasterised attribute (DRM), extracted from 3DBAG. This signed metric distinguishes low from high roof surfaces, enabling material differentiation by position (e.g., membranes at a lower height relative to solar panels installed above them).

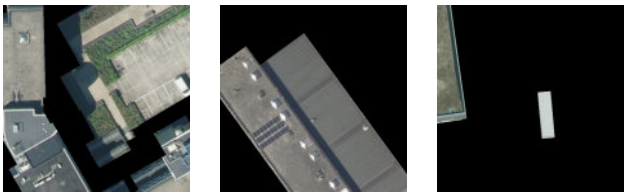
3.2.4 Density Band Construction: Finally, the density band, indicating surface completeness, opacity, and texture, was constructed at a coarser GSD of 1 m to better capture point distribution patterns. Each target cell was assigned the count of points whose 2D projection intersected it (Figure 2c). To preserve its total sum, representing the point cloud population, the resulting raster was not post-processed.

3.3 Component Fusion

To create a unified image and leverage complementary information, the RGB and lidar-derived components were fused at the pixel level. The lidar-derived component of each parent image was upsampled to the GSD of the RGB component for spatial alignment. To preserve their range and physical interpretation, the continuous reflectance, slope, and normalized digital roof model (nDRM) bands were resampled using bilinear interpolation. The nDRM band was clipped $\left(\left[2^{\text{nd}}, 98^{\text{th}}\right]\right.$ percentiles$\left.\right)$ to reduce temporal misalignment artefacts between AHN4 and 3DBAG. The discrete density band was resampled using nearest-neighbour interpolation. The resulting raster was scaled by its pre- and post-resampling total sum ratio, preserving the point cloud population. Each assembled image contained seven bands along its spectral axis, inserted by construction order.

3.4 Chip Extraction

To generate model-ready inputs and facilitate granular quality control, parent images were split into non-overlapping $512 \times 512\,\text{px}^2$ chips. Each chip was added to RoofSense only if its background content was at most 80% (Figure 3), ensuring sufficient roof coverage. This measure was computed by masking the chip with the corresponding building footprints and counting the masked cells. To avoid confusion, the background cells of included chips were filled with zero (i.e., invalidated), ensuring they would not contribute to feature maps. To increase spatial coverage, each chip was initially selected for masking using a Bernoulli trial, meaning that only approximately half of all possible chips were considered. Consequently, more images were required to construct RoofSense. Finally, to prevent redundant sampling and maximise spatial exploitation, no more than three sequentially incorporated chips per parent image were permitted.

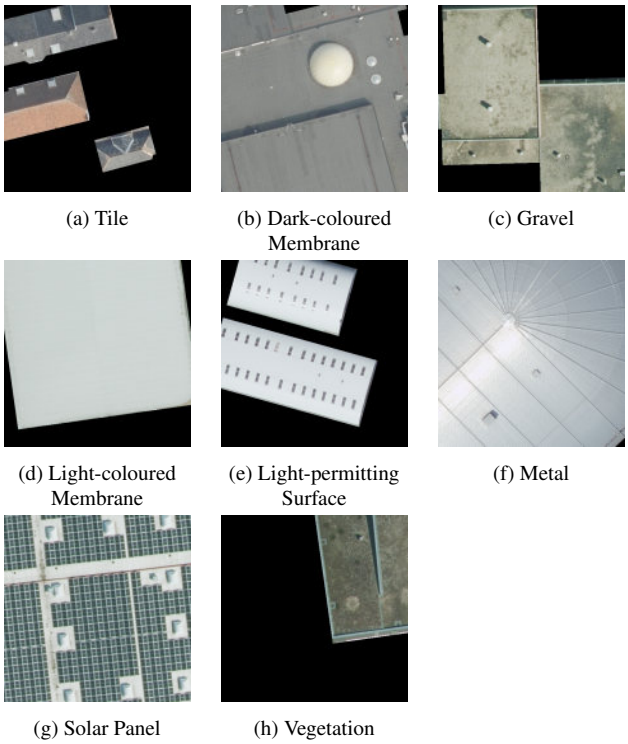


(a) $(30 \pm 5)\%$ (b) $(60 \pm 5)\%$ (c) $(90 \pm 5)\%$

Figure 3. Chips with increasing background content. Each percentage indicates the ratio of background cells relative to the corresponding pixel count (i.e., 512^2).

3.5 Material Classes

To ensure comprehensive and relevant material coverage, RoofSense represents eight classes (Figure 4), informed by relevant research (Wyard et al., 2023) and adapted to the typical Dutch roof morphology, source datasets, and for the absence of prior knowledge of the spatial distribution of certain materials (e.g., asbestos). Therefore, RoofSense balances generalisability with regional specificity. The intended distinction between light- and dark-coloured materials refers to their intrinsic colour under neutral viewing conditions, unaffected by texture. As this property is difficult to capture in aerial imagery, the actual distinction was subject to expert opinion. Finally, to accommodate its inherent diversity, the tile class is intentionally broad, encompassing asphalt shingles, ceramic, concrete, metal tiles, etc., addressing morphological and data limitations which prevented further subclassing (i.e., lack of prior knowledge).



(a) Tile (b) Dark-coloured Membrane (c) Gravel

(d) Light-coloured Membrane (e) Light-permitting Surface (f) Metal

(g) Solar Panel (h) Vegetation

Figure 4. Material classes represented in RoofSense.

3.6 Annotation Process

Each ground truth mask was constructed by annotating the RGB component of the corresponding chip in Roboflow Annotate (Roboflow, 2025). Besides traditional tools, a semi-automatic utility was used to annotate multiple nearby objects of the same material (e.g., solar panel arrays). All other annotations were manual. Generally, objects larger than 1–10 m^2 were annotated, depending on class. A special label was used to denote ambiguous, irrelevant, or unknown objects (e.g., chimneys, electromechanical equipment, flashing, gutters, ridge caps, vents, etc.) within valid annotation regions. Isolated, irrelevant regions (e.g., façade segments[1], poorly illuminated surfaces, roofs under construction or featuring severe clutter or unknown materials, etc.) were not annotated, and recreational areas (e.g., atria, balconies, decks, patios, etc.) were ignored. Finally, the background was remasked to correct annotation errors along building edges, and all invalid and unlabelled cells were mapped to it. The resulting material class encoding is presented in Table 1.

Label	Colour	Name	Label	Colour	Name
1		Tile	2		Dark-coloured Membrane
3		Gravel	4		Light-coloured Membrane
5		Light-permitting Surface	6		Metal
7		Solar Panel	8		Vegetation

Table 1. Material class encoding in RoofSense.

[1] BM5 is orthorectified using a digital terrain model, and hence elevated objects suffer from lens distortion effects.

4. Benchmark Design

4.1 Dataset Splitting Scheme

RoofSense is inherently imbalanced due to factors influencing local roof morphology (e.g., architecture, climate, zoning laws, etc.). To preserve this imbalance, an iterative stratification algorithm, inspired by Xiao et al. (2018), was developed. First, RoofSense was randomly split into training (70%), validation (15%), and test (15%) sets, ensuring non-zero pixel support per class and split. Subsequently, the splits were greedily optimised in a pairwise fashion by swapping randomly selected elements to minimise the mean Jensen-Shannon distance (mJSD) of the corresponding area-normalised histograms. This process continued for 1000 swaps or until convergence ($\Delta\text{mJSD} \leq 10^{-6}$). The entire routine was repeated 100 times, selecting the split with the lowest overall mJSD.

4.2 Class Weighting Scheme

Since RoofSense represents objects of varying typical prominence and size, certain classes with relatively large pixel support appear in few images, and vice versa (Figure 5).

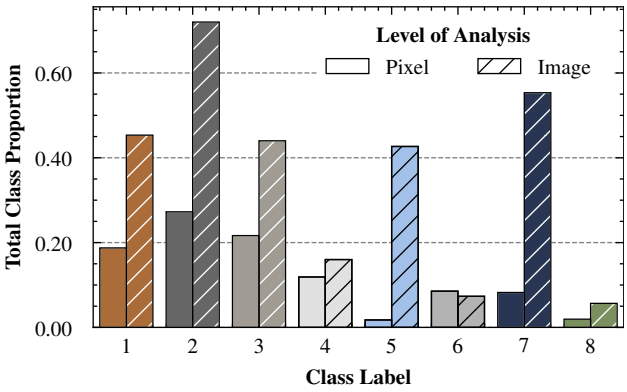


Figure 5. Pixel- and image-level class proportions in RoofSense.

To address this imbalance in the training set, a macroscopic weighting scheme, inspired by term frequency-inverse document frequency (Manning et al., 2008), was developed. Let $\mathcal{M} \coloneqq \left\{M_i \in \{0, 1, \ldots, C\}^{H \times W}, i = 1, 2, \ldots, N\right\}$ denote the set of $N \coloneqq 300$ ground truth masks, where each mask M_i is an $H \times W$ integer matrix, encoding the class of each pixel in the corresponding image: 0 for the background and $1, \ldots, C$ the $C \coloneqq 8$ material classes in RoofSense. Then, assuming that $C \geq 2$, the weight w_c of a particular class c is defined as:

$$w_c \coloneqq \frac{N}{\sum\limits_{M \in \mathcal{M}} \text{count}\,(M; c)} \cdot \left[\log_C \left(\frac{N+1}{|\mathcal{M}^c| + 1}\right) + 1\right] \quad (1)$$

where $\text{count}\,(M; c) \coloneqq \sum\limits_{i=1}^{H} \sum\limits_{j=1}^{W} \delta_{c m_{ij}}$

$$\mathcal{M}^c \coloneqq \{M \in \mathcal{M} | \,\text{count}\,(M; c) > 0\}$$

Here, δ denotes the Kronecker delta and m_{ij} represents the entry of M at its i^{th} row and j^{th} column. The selected logarithmic base scales the information content of c based on its overall prominence in RoofSense, relative to the remaining classes. To prevent division by zero, $|\mathcal{M}^c|$ is incremented by one, effectively assuming an additional mask containing c. Furthermore, a unit factor is added to the logarithmic component

of Equation 1 to account for it being zero if c is present in all masks. The background was assigned zero weight.

4.3 Data Augmentation Policy

To standardise its magnitude while preserving its physical interpretation, each input band was individually scaled to $[0, 1]$. Images were also concatenated with the scaled CIELAB representation of their RGB component. Each image-mask pair underwent sequential horizontal and vertical reflection, followed by up to three 90° rotations, each with a probability of 50%.

4.4 Loss Function

The benchmark model was trained using a loss function based on cross entropy (CE), weighted using the proposed scheme. A label smoothing (Szegedy et al., 2016) factor of 10% was applied to the target labels. To prevent rare classes (e.g., vegetation) from dominating more prominent ones (e.g., gravel, membranes, solar panels, etc.), due to class weighting, CE was combined with an unmodified Dice component (Milletari et al., 2016), each with equal contribution. The background was ignored.

4.5 Model Configuration & Training Protocol

The benchmark model was based on DeepLabv3+ (Chen et al., 2018) and used a ResNet-18-D (He et al., 2019) encoder. Encoder blocks were augmented with anti-aliasing (Zhang, 2019) and efficient channel attention (Wang et al., 2020) modules. The dilation rates in the atrous spatial pyramid pooling block were set to $(20, 15, 6)$. To mitigate labelling errors and improve predictions in small regions, the decoder output stride was set to 16. The encoder was initialised with pre-trained ImageNet-1K weights (Wightman et al., 2021); the decoder was randomly initialised. The resulting training protocol is in the implementation repository.

4.6 Tiled Inference Strategy

Although the benchmark model was fully convolutional, computational constraints necessitated a tiled inference strategy (Section 5.2). Each input was masked and split into $512 \times 512\,\text{px}^2$ patches with a 256 px overlap in a sliding window fashion. Patch segments extending beyond the input bounds were filled with zero. To minimise artefacts, patches which did not border the input edges were cropped by 16 px. Finally, overlapping probabilities were averaged.

5. Evaluation

5.1 Quantitative Evaluation

Test results using model checkpoint which achieved the highest validation mean intersection over union (mIoU) are presented in Table 2. Generally, the macroscopic accuracy (84.99%) and F_1 score (84.20%) were similar to relevant works, although differing test datasets prevented direct comparison. Similarly, the achieved mIoU of 74.74% was considered competitive.

The model best detected gravel, membranes, metal, and tiles, where the corresponding F_1 score and mIoU were in the order of 90% and 80%, respectively. This was expected as these classes were not only the most prominent at the pixel-level but also represent typically large, uniform objects on flat roofs, which are associated with relatively higher performance due to their morphology (Fiumi et al., 2014). Conversely, performance degraded in light-permitting surfaces, solar panels, and vegetation. In the case of solar panels, errors primarily involved confusion with tiles, with the low corresponding IoU suggesting incomplete boundaries[2] or omissions. This was likely linked to their relatively small inherent size and large decoder output stride, particularly concerning black panels on dark or poorly illuminated tiles (Figure 6a). The most significant issues related to light-permitting surfaces involved confusion with discoloured skylights, light-coloured membranes, and metal of similar hue (Figure 6b). Furthermore, confusion between vegetation and gravel was concentrated in a single test case featuring a gravel-ballasted green roof with little healthy vegetation, exposing both materials (Figure 6c).

Class Label	Precision (%)	Recall (%)	IoU (%)
1	90.49	94.83	86.24
2	96.22	88.26	85.30
3	93.64	98.36	92.22
4	93.95	92.10	86.63
5	57.57	54.59	38.93
6	88.82	97.96	87.20
7	82.56	75.74	65.29
8	66.63	78.09	56.14
Overall	**Avg. Acc.** (%)	**Avg. F_1** (%)	**mIoU** (%)
	84.99	84.20	74.74

Table 2. Test performance of the benchmark model.

Finally, the low precision and IoU in light-permitting surfaces and vegetation suggested oversegmentation or hallucinations. This was expected in the case of light-permitting surfaces (Figure 6d) due to their size and visual similarity with small, (i.e., $<$1–2 m^2) reflective, metal objects, as well as the decoder output stride. For vegetation, misdetections in dark regions (Figure 6e), albeit with low margin[3], potentially stemmed from annotation quality issues, adjacent tall vegetation, and modality misalignment.

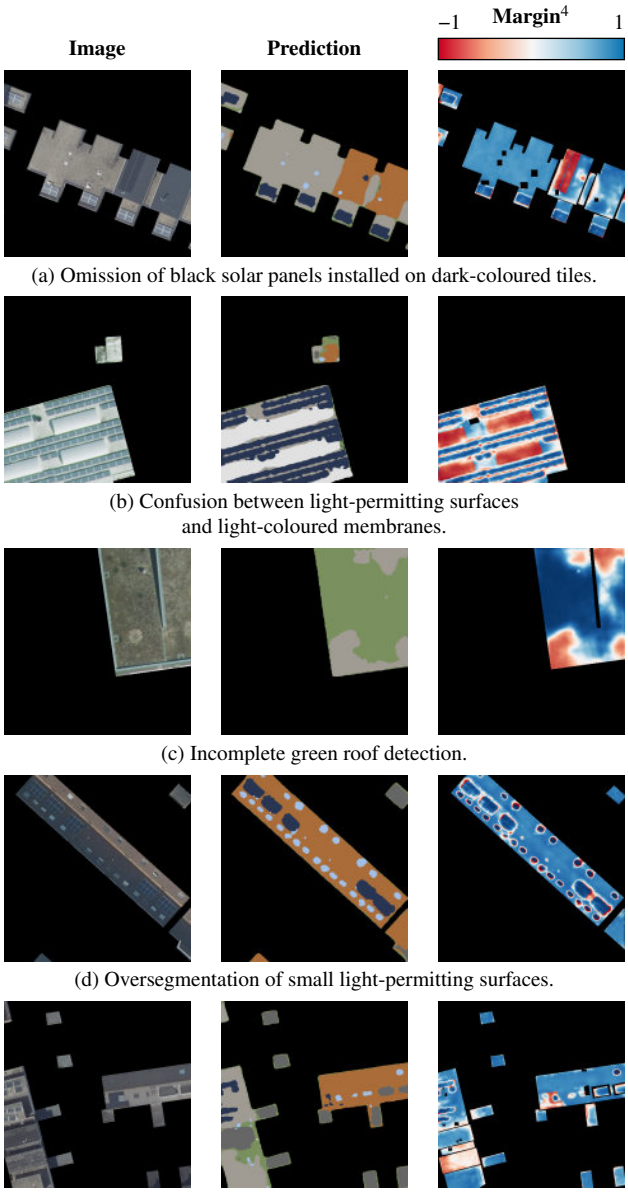
5.2 Qualitative Evaluation

Because the size of the test set (i.e., 45 images) limited the impact of quantitative conclusions, additional qualitative evaluation was performed on a separate 3DBAG tile containing buildings of varying use, including two with unrepresented materials. Tiled inference was performed according to Section 4.6.

In general, performance was best in isolated commercial and industrial buildings with uniform flat roofs (Figures 9a and 9b), consistent with test performance in the corresponding materials. Conversely, the model struggled with adjacent residential buildings featuring clutter and relatively small objects represented by materials featuring with and high inter- and intra-class variability (Figure 7c). Suggesting the effectiveness of the test set, and hence the proposed splitting scheme, this result was also aligned with the corresponding quantitative results. Furthermore, unknown materials were handled appropriately. For instance, asphalt was mainly labelled as dark-coloured membranes (Figure 7a-Top Left), the most visually similar class in RoofSense.

[2] Relatively high precision implied that positive predictions were generally smaller than the corresponding ground truth.

[3] Defined as the difference between the top two corresponding probabilities (Scheffer et al., 2001).



(a) Omission of black solar panels installed on dark-coloured tiles.

(b) Confusion between light-permitting surfaces and light-coloured membranes.

(c) Incomplete green roof detection.

(d) Oversegmentation of small light-permitting surfaces.

(e) Vegetation hallucinations.

Figure 6. Characteristic prediction errors of the benchmark model in the test set.

In summary, performance depended heavily on material and building context (e.g., use, isolation, complexity, etc.), clutter, and modality alignment (Figure 8).

5.3 Performance Impact of Lidar Data

To assess the performance impact of the lidar-derived component, an ablation study was conducted (Table 3). Each experiment was repeated three times using different seeds. The last model checkpoint was used in the name of fairness.

In general, performance was only marginally affected by the ablation of the lidar-derived component, indicating potential interference amongst the corresponding bands, or with the RGB component. Specifically, the reflectance and slope bands appeared to offer little overall added value, with slope ablation even resulting in negligible mean segmentation improvement. Furthermore, the limited effectiveness of the reflectance band

[4] Incorrect prediction scores are negated and invalid regions ignored.

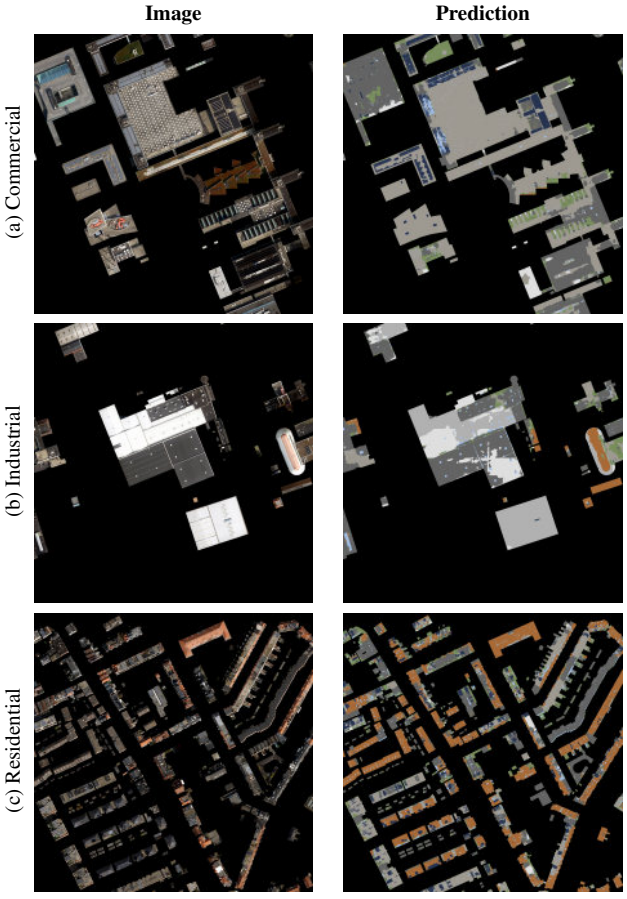
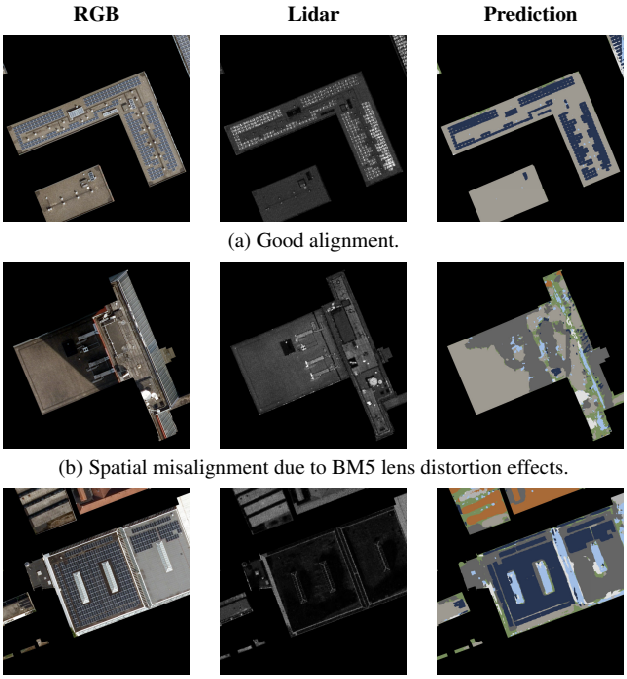


Figure 7. Tiled inference in various neighbourhoods using the benchmark model.



(a) Good alignment.

(b) Spatial misalignment due to BM5 lens distortion effects.

(c) Temporal misalignment due to AHN4 precedence.

Figure 8. Characteristic modality (mis)alignemnt cases in the test set. The lidar-derived component is visualised using the reflectance band.

Ablated Band	ΔPrecision (%)	ΔRecall (%)	ΔmIoU (%)
Reflectance	0.58 ± 0.14	1.01 ± 0.16	0.07 ± 0.20
Slope	0.26 ± 0.26	0.58 ± 0.21	0.11 ± 0.34
nDRM	0.85 ± 0.29	1.24 ± 0.29	1.75 ± 0.18
Density	1.76 ± 1.05	2.46 ± 0.63	2.85 ± 1.12
All	0.08 ± 0.40	0.33 ± 0.23	0.18 ± 0.48

Table 3. Macroscopic test performance of the benchmark model for each ablated band of the lidar-derived component of the corresponding images. The mean of three trials and corresponding standard error are reported relative to the baseline. Negative changes are denoted in red.

was attributed to uncalibrated atmospheric, instrument, target, and operating parameters (RIEGL, 2024). Although slope information may appear redundant at a macroscopic level given the context provided by the nDRM band, its ablation actually degraded performance in tiles, suggesting the importance of lidar in providing relevant 3D cues for detecting materials on pitched roofs.

Similarly, nDRM ablation resulted in relatively noticeable performance degradation, supporting the contextual added value of the corresponding band. This was reflected at the class level, with a performance decrease observed in tiles, light-permitting surfaces, and solar panels. The density band also appeared to be impactful overall, and its removal disproportionately affected segmentation performance in light-permitting surfaces and solar panels, consistent with the implicit author assumption that light-permitting surfaces were associated with low density.

The analysis also revealed more complex interactions between the RGB and lidar-derived components. Specifically, the incorporation of lidar-derived features resulted in marginal performance degradation in certain classes. For instance, classification performance in metal improved slightly when all lidar-derived bands were removed. This outcome suggests that, for materials with a highly distinct visual appearance due to colour, reflectance or texture (i.e., light-permitting surfaces, metal, solar panels, etc.), lidar can introduce confounding signals rather than complementary information, particularly due to modality misalignment. Similarly, performance in solar panels was mixed. While slope, reflectance, and nDRM ablation generally decreased performance, the removal of the density band particularly impacted segmentation more than classification. In addition, the ablation of the lidar-derived component resulted in a slight increase of the corresponding F_1 score. This finding reflects scenarios where certain panels were easily identifiable in the visible light spectrum, while lidar introduced conflicts in other cases due to modality misalignment.

In summary, the performance effect of the lidar-derived component, albeit potentially positive, was not immediately apparent. This discrepancy with relevant works may stem from differences in the particular study area and source datasets, the size of RoofSense increasing uncertainty and limiting statistical significance of such experiments, and the dominance of the RGB component due to the annotation process in combination with sensitivity to modality alignment. These results suggest that the relatively simple pixel-level fusion strategy employed in the context this paper may not be sufficient to optimally leverage the complementary information provided by lidar, and that a more refined approach to feature engineering (e.g., radiometric reflectance calibration; Wu et al. 2021) and data fusion could have yielded better results.

6. Conclusion

This paper addresses pressing research gaps in the field of roofing material classification by introducing RoofSense, a multimodal semantic segmentation dataset designed for use in diverse urban contexts, leveraging 8 cm aerial RGB imagery and ALS data. Representing eight diverse classes and spanning more than 138 ha and 480 buildings across five Dutch cities, RoofSense is the largest publicly available dataset of its kind. Using an off-the-shelf model based on ResNet-18-D and DeepLabv3+, RoofSense achieved competitive classification and segmentation performance in a purpose-designed, comprehensive benchmarking experiment, thus demonstrating the contemporary relevance of the underlying method, particularly a novel weighting scheme to address class imbalance.

Despite this outcome, experimental results were ultimately sensitive to material and building context, clutter, and modality alignment, hindering the otherwise positive effect of the incorporated lidar-derived features. A relevant ablation study revealed that the performance impact of the lidar component was marginal and, in certain cases, counterintuitive. This observation indicates that, despite the theoretical benefits of data fusion, its implementation remains critical. Hence, the incorporated features and corresponding preprocessing and fusion strategies should be refined to better leverage the complementary information provided by lidar while addressing clutter and modality alignment issues. Furthermore, future work should focus on extending RoofSense and enriching its annotation protocol with semi-automated tools, facilitating the eventual inclusion of diverse geographic regions outside of the Netherlands to increase its robustness and applicability. These efforts will further solidify the role of semantic segmentation for roofing material classification in enabling data-driven urban sustainability and infrastructure management, particularly for applications such as energy studies and material inventorying.

References

Abbasi, M., Mostafa, S., Vieira, A. S., Patorniti, N., Stewart, R. A., 2022. Mapping Roofing with Asbestos-Containing Material by Using Remote Sensing Imagery and Machine Learning-Based Image Classification: A State-of-the-Art Review. *Sustainability*, 14(13).

Abriha, D., Kovács, Z., Ninsawat, S., Bertalan, L., Balázs, B., Szabó, S., 2018. Identification of roofing materials with Discriminant Function Analysis and Random Forest classifiers on pan-sharpened WorldView-2 imagery – a comparison. *Hungarian Geographical Bulletin*, 67(4), 375–392.

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, 59, 25-37.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Computer Vision – ECCV 2018*, 833–851.

Feng, S., Fan, F., 2021. Analyzing the Effect of the Spectral Interference of Mixed Pixels Using Hyperspectral Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 1434–1446.

Fiumi, L., Congedo, L., Meoni, C., 2014. Developing expeditious methodology for mapping asbestos-cement roof coverings over the territory of Lazio Region. *Applied Geomatics*, 6(1), 37–48.

Gibril, M. B. A., Shafri, H. Z. M., Hamedianfar, A., 2017. New semi-automated mapping of asbestos cement roofs using rule-based object-based image analysis and Taguchi optimization technique from WorldView-2 images. *International Journal of Remote Sensing*, 38(2), 467–491.

Hamedianfar, A., Shafri, H. Z. M., Mansor, S., Ahmad, N., 2014a. Combining data mining algorithm and object-based image analysis for detailed urban mapping of hyperspectral images. *Journal of Applied Remote Sensing*, 8(1), 085091.

Hamedianfar, A., Shafri, H. Z. M., Mansor, S., Ahmad, N., 2014b. Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data. *International Journal of Remote Sensing*, 35(5), 1876–1899.

He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M., 2019. Bag of Tricks for Image Classification with Convolutional Neural Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 558–567.

Het Waterschapshuis, 2023a. Actuel Hoogtebestand Nederland (AHN). GeoTiles. https://geotiles.citg.tudelft.nl (1 July 2025).

Het Waterschapshuis, 2023b. Beeldmateriaal Nederland (BM). https://www.beeldmateriaal.nl (1 July 2025).

Ilehag, R., Bulatov, D., Helmholz, P., Belton, D., 2018. Classification and Representation of Commonly used Roofing Material using Multisensorial Aerial Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1, 217–224.

Krówczyńska, M., Raczko, E., Staniszewska, N., Wilk, E., 2020. Asbestos—Cement Roofing Identification Using Remote Sensing and Convolutional Neural Networks (CNNs). *Remote Sensing*, 12(3).

Manning, C. D., Raghavan, P., Schütze, H., 2008. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*, 565–571.

Norman, M., Shafri, H. Z. M., Mansor, S., Yusuf, B., Radzali, N. A. W. M., 2020. Fusion of multispectral imagery and LiDAR data for roofing materials and roofing surface conditions assessment. *International Journal of Remote Sensing*, 41(18), 7090–7111.

Peters, R., Dukai, B., Vitalis, S., van Liempt, J., Stoter, J., 2022. Automated 3D Reconstruction of LoD2 and LoD1 Models for All 10 Million Buildings of the Netherlands. *Photogrammetric Engineering & Remote Sensing*, 88(3), 165–170.

Raczko, E., Krówczyńska, M., Wilk, E., 2022. Asbestos roofing recognition by use of convolutional neural networks and high-resolution aerial imagery. Testing different scenarios. *Building and Environment*, 217, 109092.

RIEGL, 2017. LAS Extrabytes Implementation in RIEGL Software. 4TU. https://data.4tu.nl/file/1aac46fb-7900-4d4c-a099-d2ce354811d2/7ade80c4-aa45-4e87-b887-ee8478c96181 (1 July 2025).

RIEGL, 2024. RIEGL VQ(R)-1560 II-S Data Sheet. http://www.riegl.com/uploads/tx_pxpriegldownloads/RIEGL_-VQ-1560II-S_Datasheet_2024-03-22.pdf (1 July 2025).

Roboflow, 2025. Roboflow Annotate. https://roboflow.com/annotate (1 July 2025).

Samsudin, S. H., Shafri, H. Z. M., Hamedianfar, A., 2016. Development of spectral indices for roofing material condition status detection using field spectroscopy and WorldView-3 data. *Journal of Applied Remote Sensing*, 10(2), 025021.

Scheffer, T., Decomain, C., Wrobel, S., 2001. Active Hidden Markov Models for Information Extraction. *Advances in Intelligent Data Analysis*, 309–318.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826.

Tommasini, M., Bacciottini, A., Gherardelli, M., 2019. A QGIS Tool for Automatically Identifying Asbestos Roofing. *ISPRS International Journal of Geo-Information*, 8(3).

Trevisiol, F., Lambertini, A., Franci, F., Mandanici, E., 2022. An Object-Oriented Approach to the Classification of Roofing Materials Using Very High-Resolution Satellite Stereo-Pairs. *Remote Sensing*, 14(4).

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11531–11539.

Wightman, R., Touvron, H., Jegou, H., 2021. ResNet strikes back: An improved training procedure in timm. *NeurIPS 2021 Workshop on ImageNet: Past, Present, and Future*.

Wu, Q., Zhong, R., Dong, P., Mo, Y., Jin, Y., 2021. Airborne LiDAR Intensity Correction Based on a New Method for Incidence Angle Correction for Improving Land-Cover Classification. *Remote Sensing*, 13(3).

Wyard, C., Fauvel, H., Palmaerts, B., Beaumont, B., Hallot, E., 2023. From DL approach conception to operational product design : identifying roof materials for policy makers. *2023 Joint Urban Remote Sensing Event (JURSE)*, 1–4.

Xiao, T., Liu, Y., Zhou, B., Jiang, Y., Sun, J., 2018. Unified Perceptual Parsing for Scene Understanding. *Computer Vision – ECCV 2018*, 432–448.

Zevenbergen, L. W., Thorne, C. R., 1987. Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms*, 12(1), 47–56.

Zhang, R., 2019. Making Convolutional Networks Shift-Invariant Again. *Proceedings of the 36th International Conference on Machine Learning*, 97, 7324–7334.

Zheng, X., Montazeri, H., Blocken, B., 2021. CFD analysis of the impact of geometrical characteristics of building balconies on near-façade wind flow and surface pressure. *Building and Environment*, 200, 107904.