CNN-based Roofing Material Segmentation using Aerial Imagery and LiDAR Data Fusion

P5 Presentation | Dimitris Mantas

31-10-2024

Agenda

01 Introduction

02 Methodology

03 Implementation

04 Results

05 Conclusions

01

Introduction

Problem Statement

Applications of Roofing Material Classification

- Hazardous material mapping
- Urban heat island studies
- Energy retrofitting studies and solar potential estimations
- Material inventorying

Current State of the Art

- Narrow application and material scope
- Reliance on MSI and HSI
- Negative disposition towards pixel-based classification
- Limited integration of heterogenous data sources

To what extent is DL–based roofing material segmentation possible using aerial imagery and LiDAR data fusion?

Research Questions

Primary Question

To what extent is **DL–based** roofing material **segmentation** possible using aerial imagery and LiDAR **data fusion**?

Supporting Questions

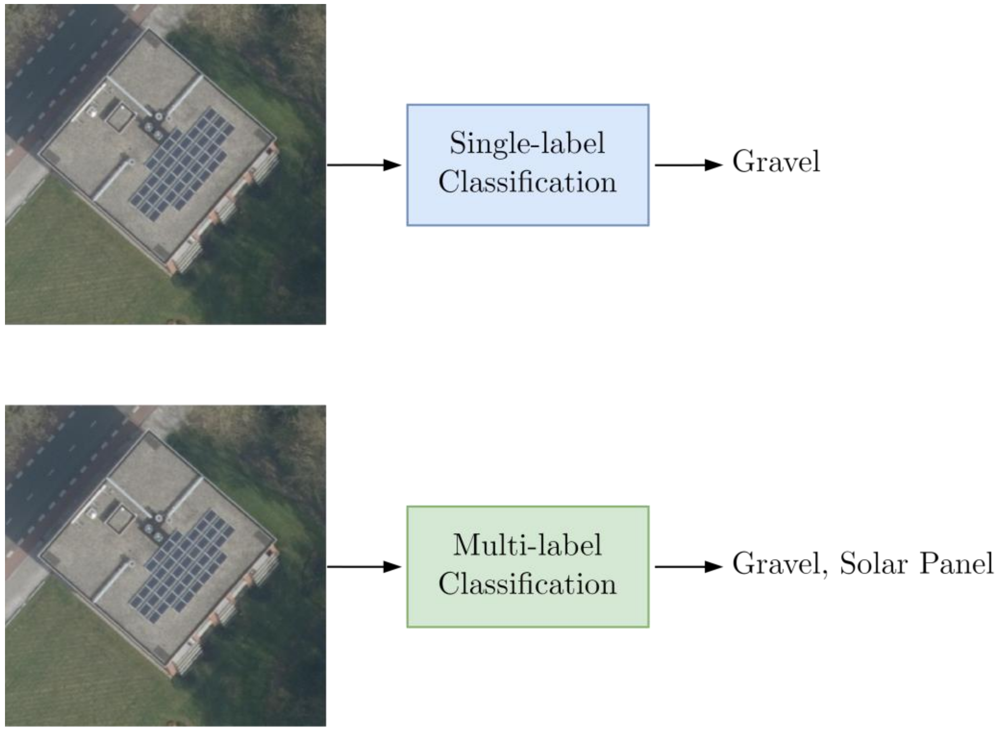
1. Which **imagery- and LiDAR-derived features** are the most contextually prevalent and effective in terms of performance?
2. Which **classification and data fusion approaches** are the most contextually prevalent and effective in terms of performance?
3. How does the **generalisation of pixel-wise material maps** to the roof segment level using each of the building LODs offered by the 3DBAG influence results?
4. How does the **availability of LiDAR-derived features** influence performance?

Image-based Classification

Assigns material labels at the image level

Key Takeaways

- Only method where DL has been applied
  - Small, custom models
- Potentially significant label localisation issues
  - Requires small or specially constructed images showing a single building or material

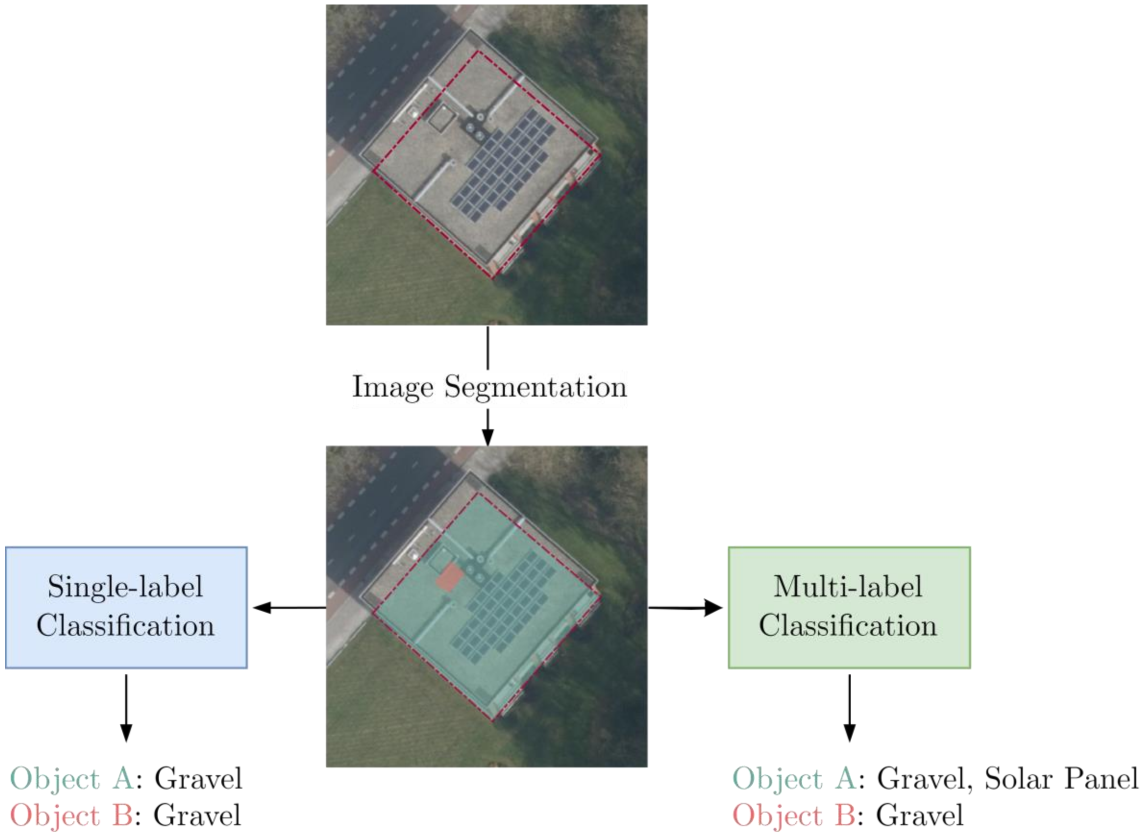


Object-based Classification

Assigns material labels at the superpixel level

Key Takeaways

- Alternative to pixel-based classification due to historically spatially inconsistent results
- Significant ambiguity in segmentation step
  - Image-specific algorithm parameters
  - “What constitutes an object”?

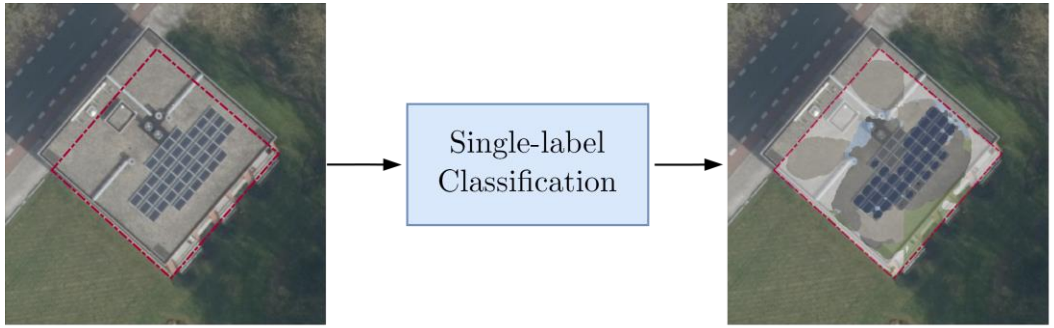


Pixel-based Classification

Assigns material labels at the pixel level

Key Takeaways

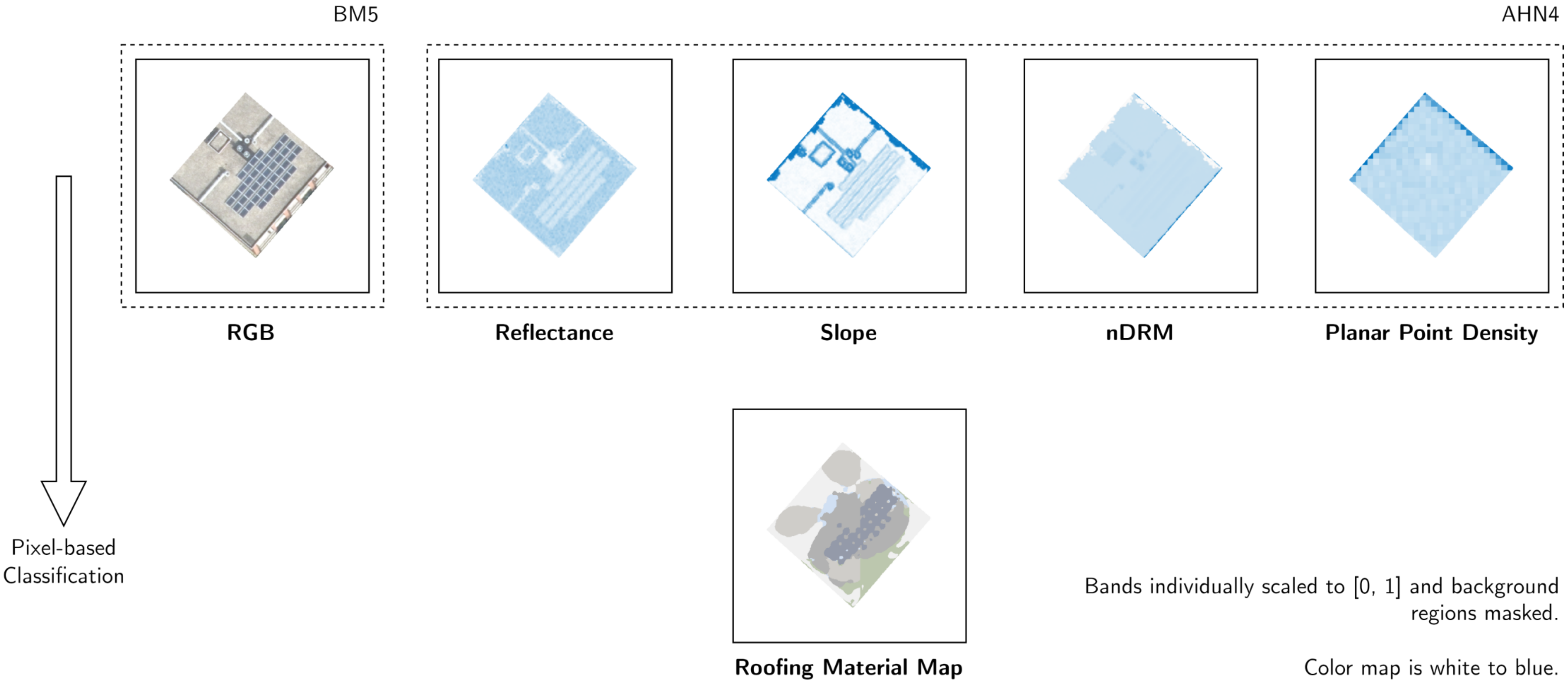
- No explicit input or preprocessing requirements
- "Infinite" output label resolution
    - Label merging becomes a matter of post-processing
- Potentially significant mixed pixel issues with models which cannot infer spatial context



02

Methodology

Overview



Bands individually scaled to [0, 1] and background regions masked.

Color map is white to blue.

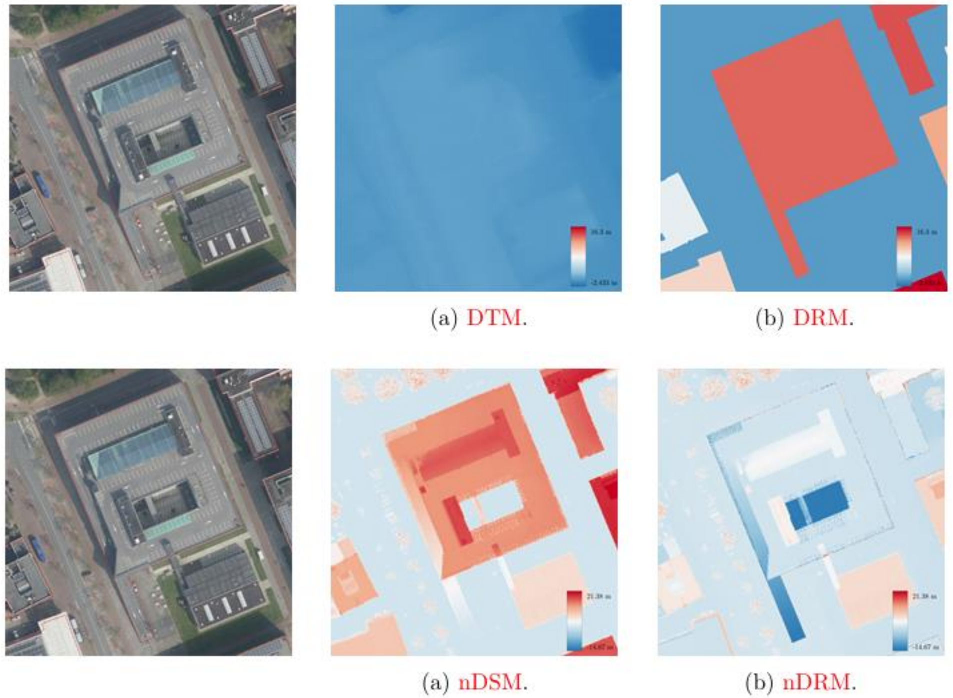
Normalised Digital Roof Model (nDRM)

Digital Roof Model

- Median roof elevation (MRE) where buildings are present and zero elsewhere

nDRM

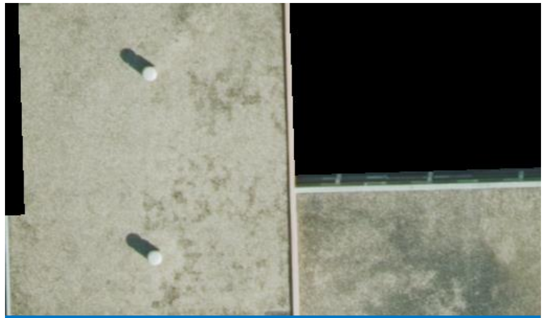
- Signed elevation difference from MRE where buildings are present and DSM elsewhere
- Complements slope by assigning it a rough direction



Material Classes

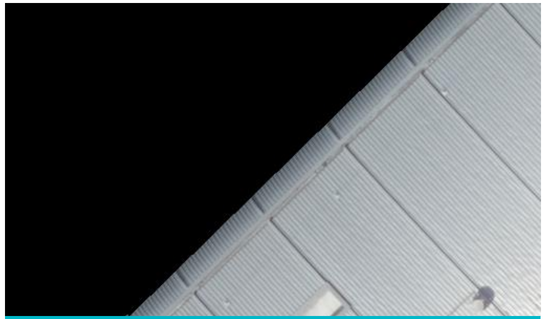
Dark-coloured Mem.

Ceramic Tiles

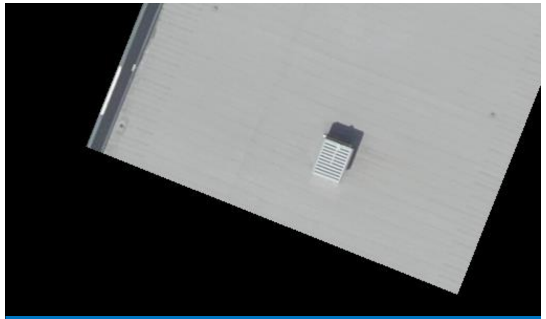


Gravel

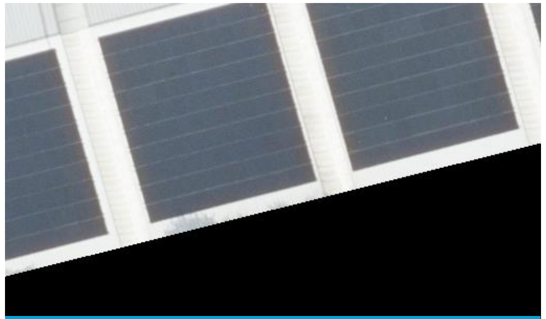
Light-permitting Surf.



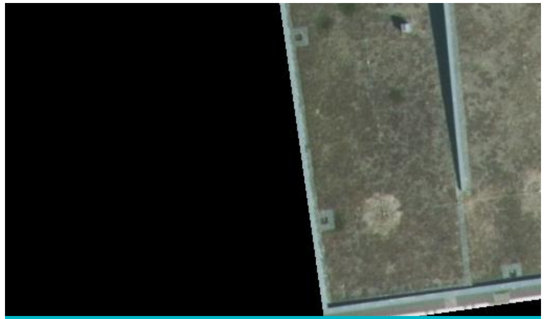
Metal



Light-coloured Mem.



Solar Panels



Vegetation

03

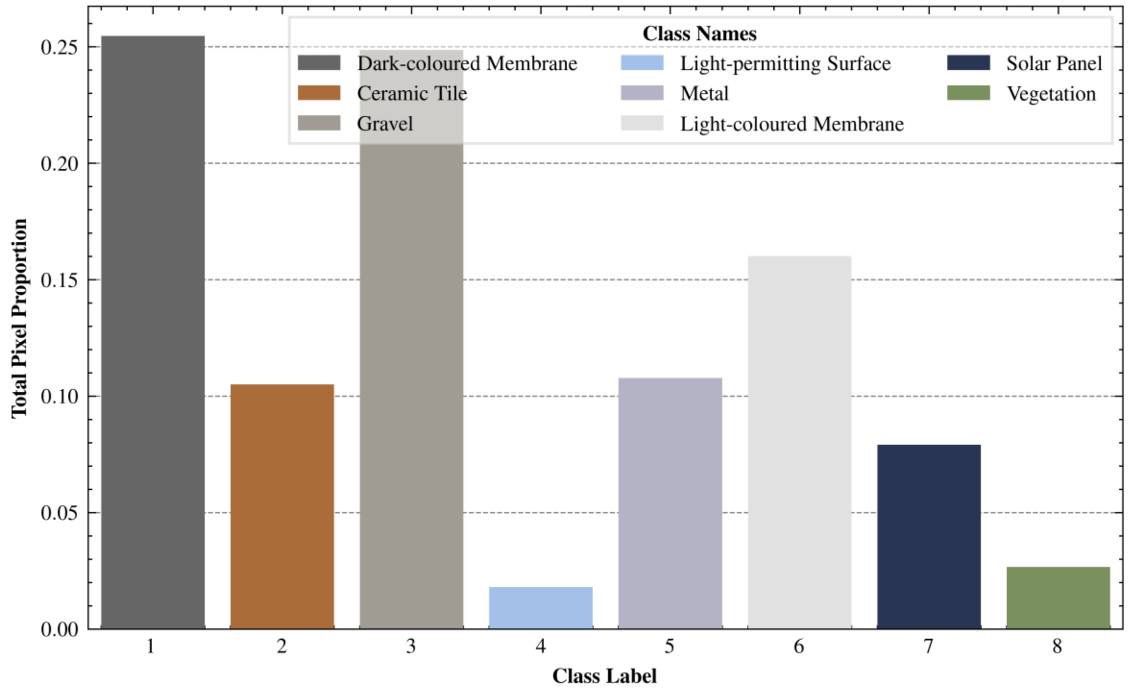
Implementation

Reference Dataset

Randomly sampled 512 x 512 px. chips from tiles within 15 km from one of 30 cities with at least 100K population

Sampling Requirements

- Chips accepted based on outcome of Bernoulli trial
- At most 80% background content
- No more than three accepted chips in a row



Summary Statistics

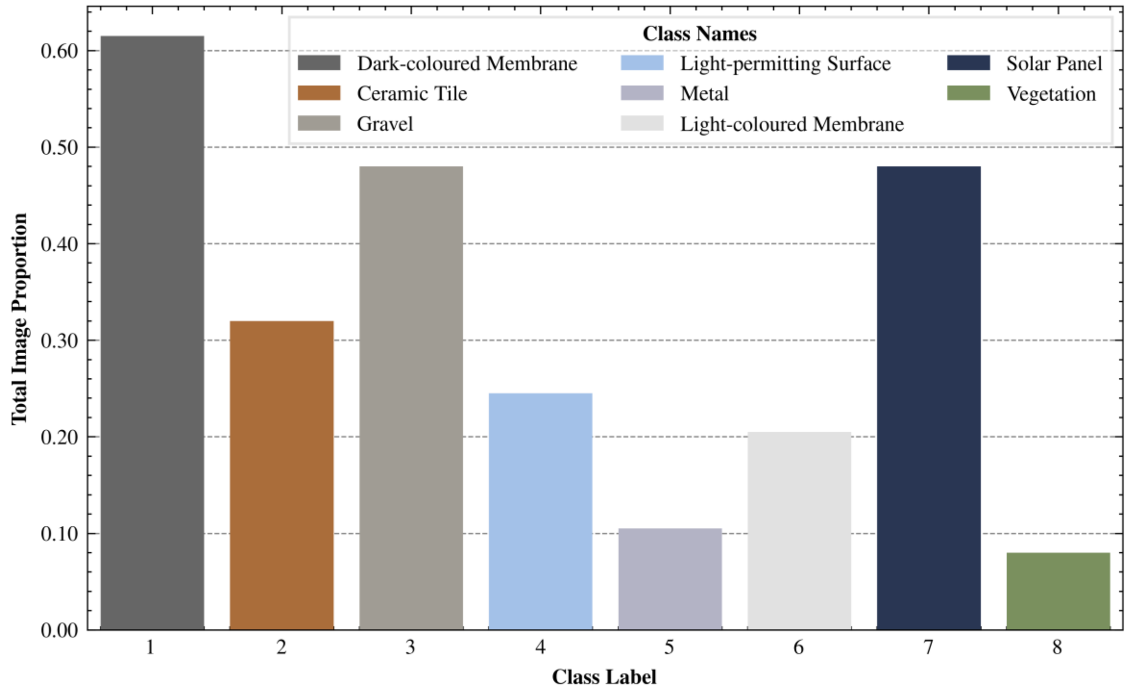
- 200 chips
- 4 tiles
  1x Delft, 2x Dordrecht, 1x Enschede
- 15M+ px.
  100+ ha @ 8 cm GSD

Reference Dataset

Randomly sampled 512 x 512 px. chips from tiles within 15 km from one of 30 cities with at least 100K population

Sampling Requirements

- Chips accepted based on outcome of Bernoulli trial
- At most 80% background content
- No more than three accepted chips in a row

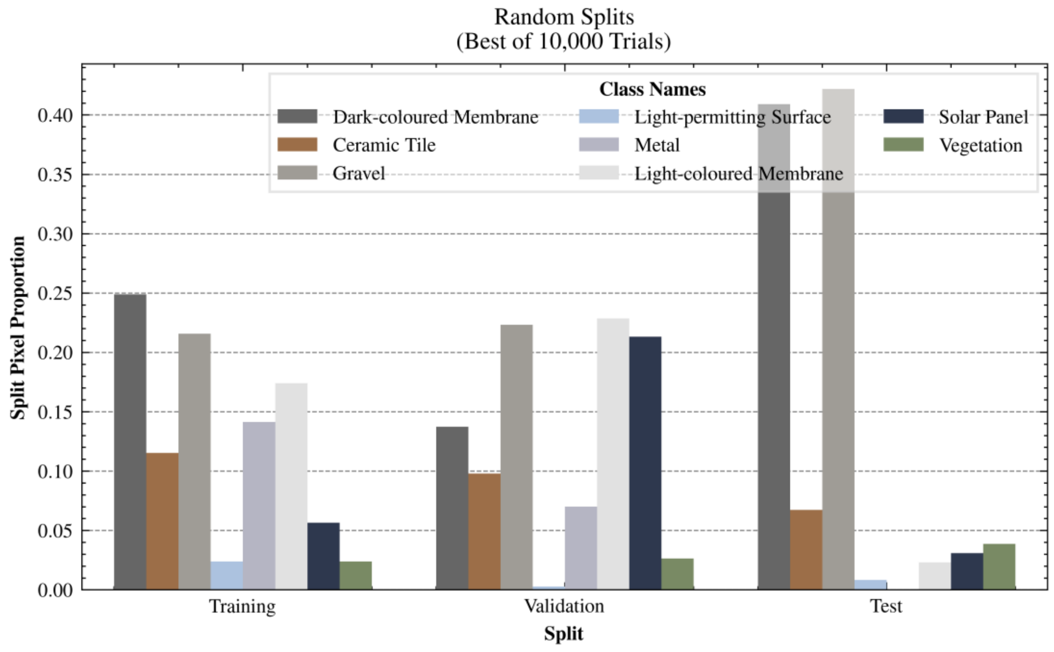


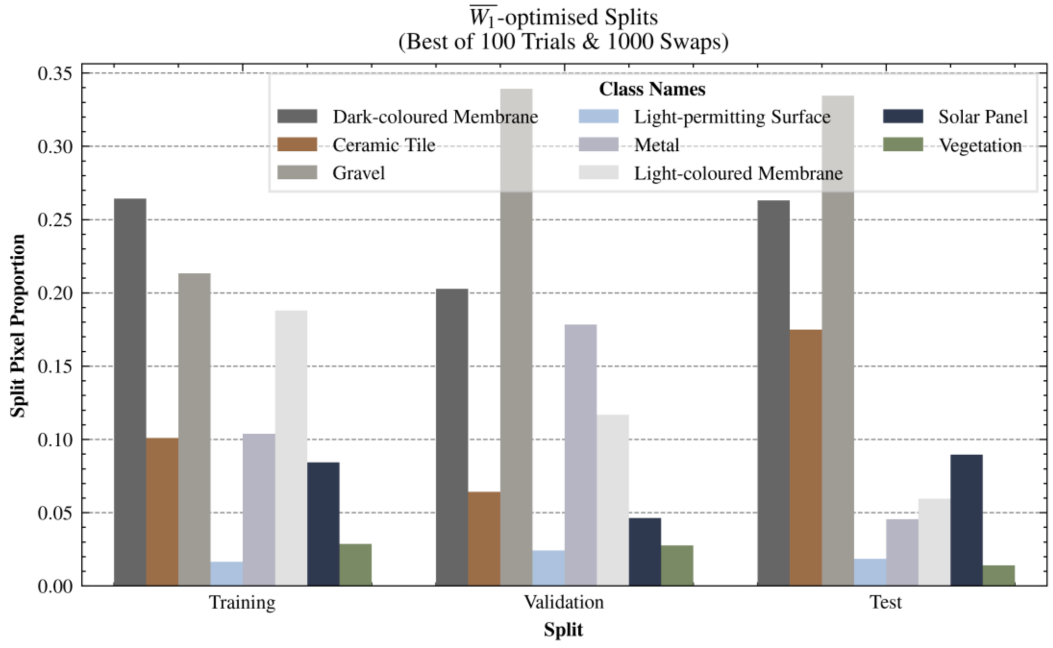
Summary Statistics

- 200 chips
- 4 tiles
  1x Delft, 2x Dordrecht, 1x Enschede
- 15M+ px.
  100+ ha @ 8 cm GSD

TU Delft

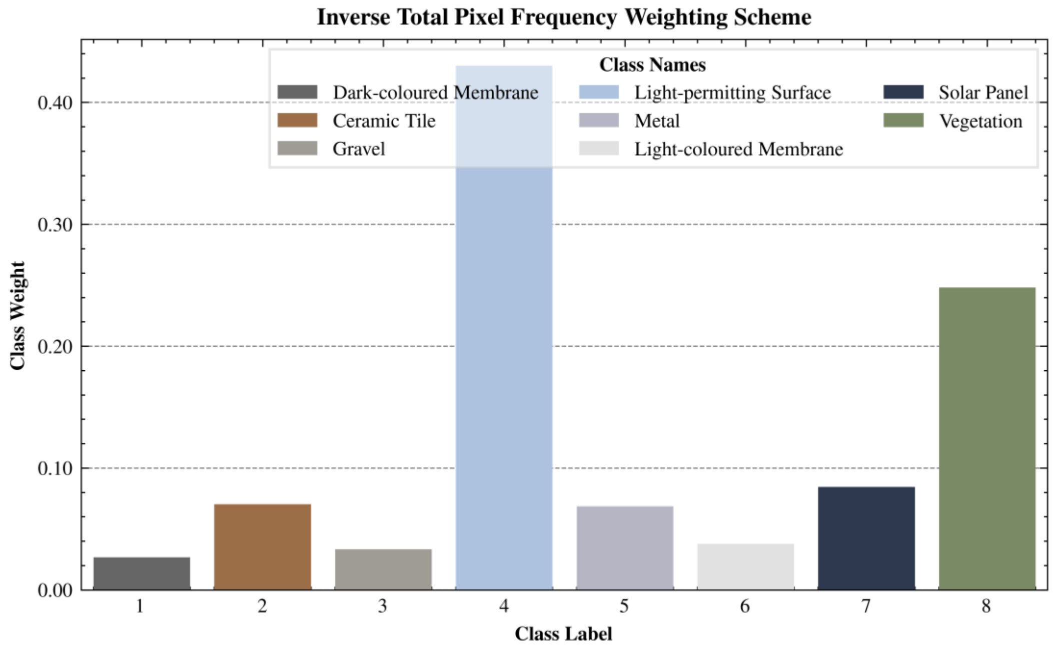
Reference Dataset Splitting

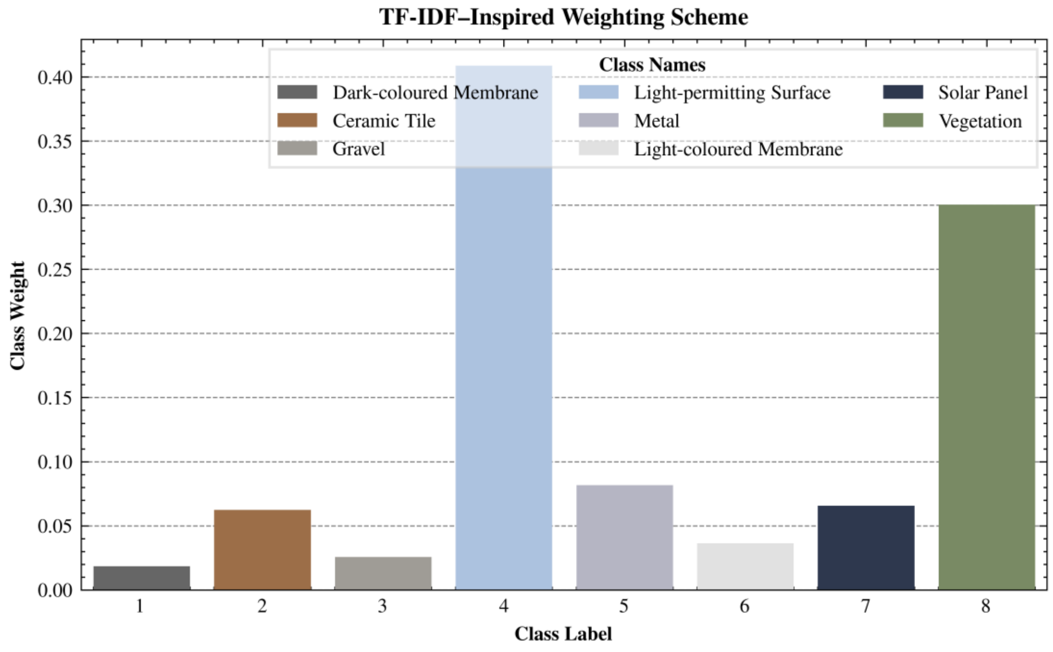




A custom stratification algorithm was designed to preserve inherent class imbalance across splits

Training Subset Class Weighting





A custom weighting scheme was applied to the CE component of the loss function to account to mitigate image-level class imbalance in the training set

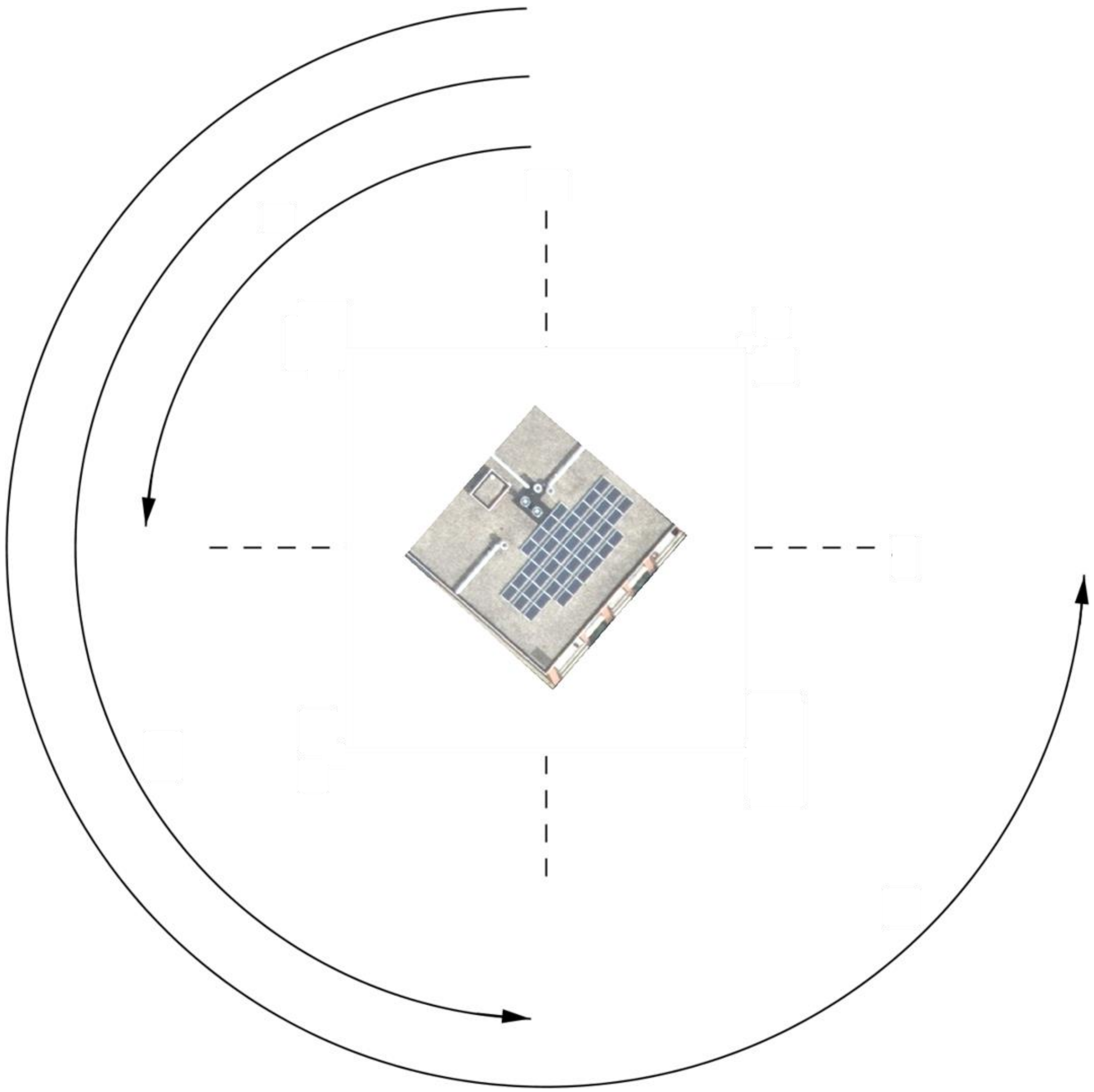
Data Augmentation

Training

- Individual band scaling to [0, 1] based on training subset statistics to prevent data leakage
- D4 dihedral group symmetries except the (off)-diagonal reflections applied sequentially with a probability of 50%

Inference

- Band scaling to [0, 1]



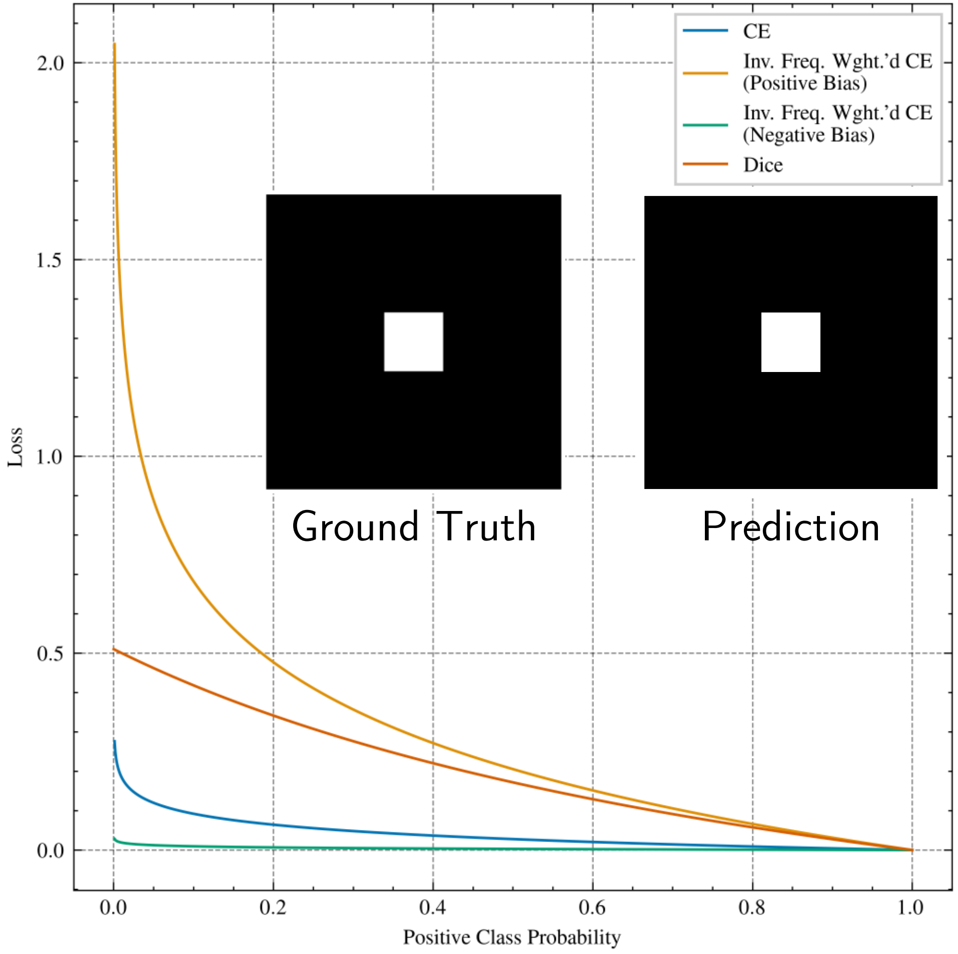
Loss Function

Sum of weighted CE and Dice loss

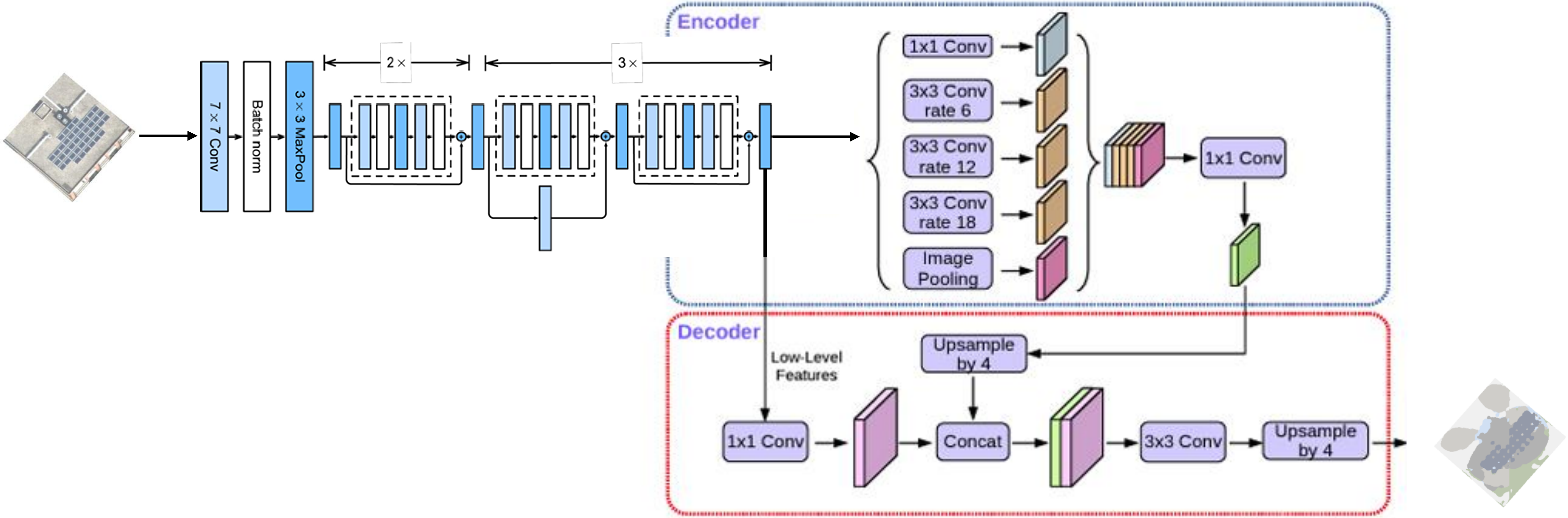
Key Advantages

- Metric of interest is directly maximised while maintaining favourable loss landscape properties
- Handling of inappropriate weighting scenarios

$$\ell(X, Y) = -w \ln \frac{\exp(X)}{\sum_{c \in C} \exp(X_c)} - \left(1 - \frac{2X \odot Y}{X^2 + Y^2}\right)$$



Model Design



04

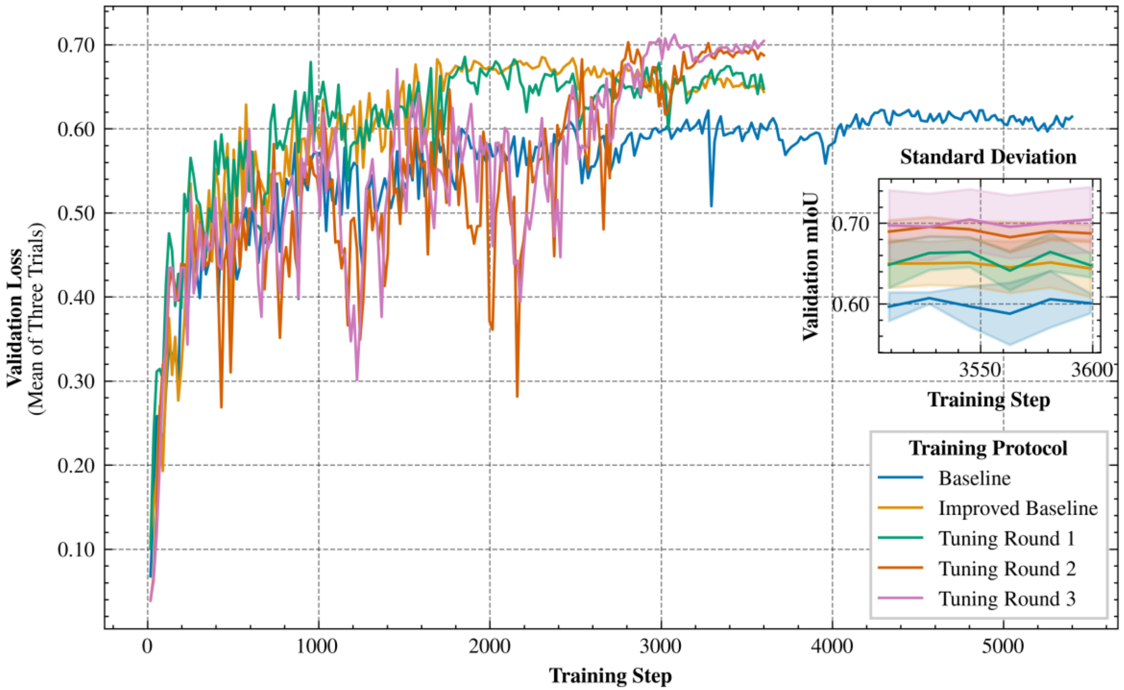
Results

Hyperparameter Optimisation

Exploration to gain insights into the problem and model dynamics, followed by greedy exploitation of obtained knowledge

Process Outline

- Manual Experimentation
- Tuning Round 1
  Input Data & Model
- Tuning Round 2
  Optimiser & LR Scheduler Config.
- Tuning Round 3
  Continuous Parameters



Key Changes

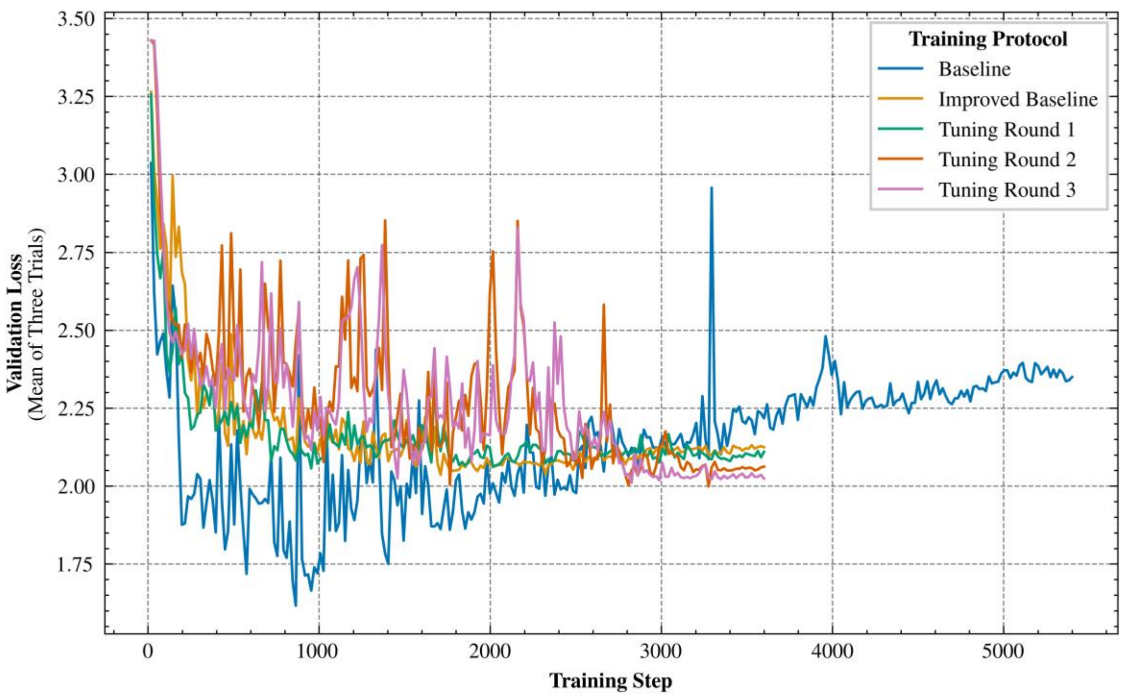
- Swapped ResNet-18 for ResNet-18-D and added ECA to each block
- Added 10% label smoothing in CE
- Increased max. LR by 4.5x and added linear warmup for 67.5% of training duration

Hyperparameter Optimisation

Exploration to gain insights into the problem and model dynamics, followed by greedy exploitation of obtained knowledge

Process Outline

- Manual Experimentation
- Tuning Round 1
  Input Data & Model
- Tuning Round 2
  Optimiser & LR Scheduler Config.
- Tuning Round 3
  Continuous Parameters



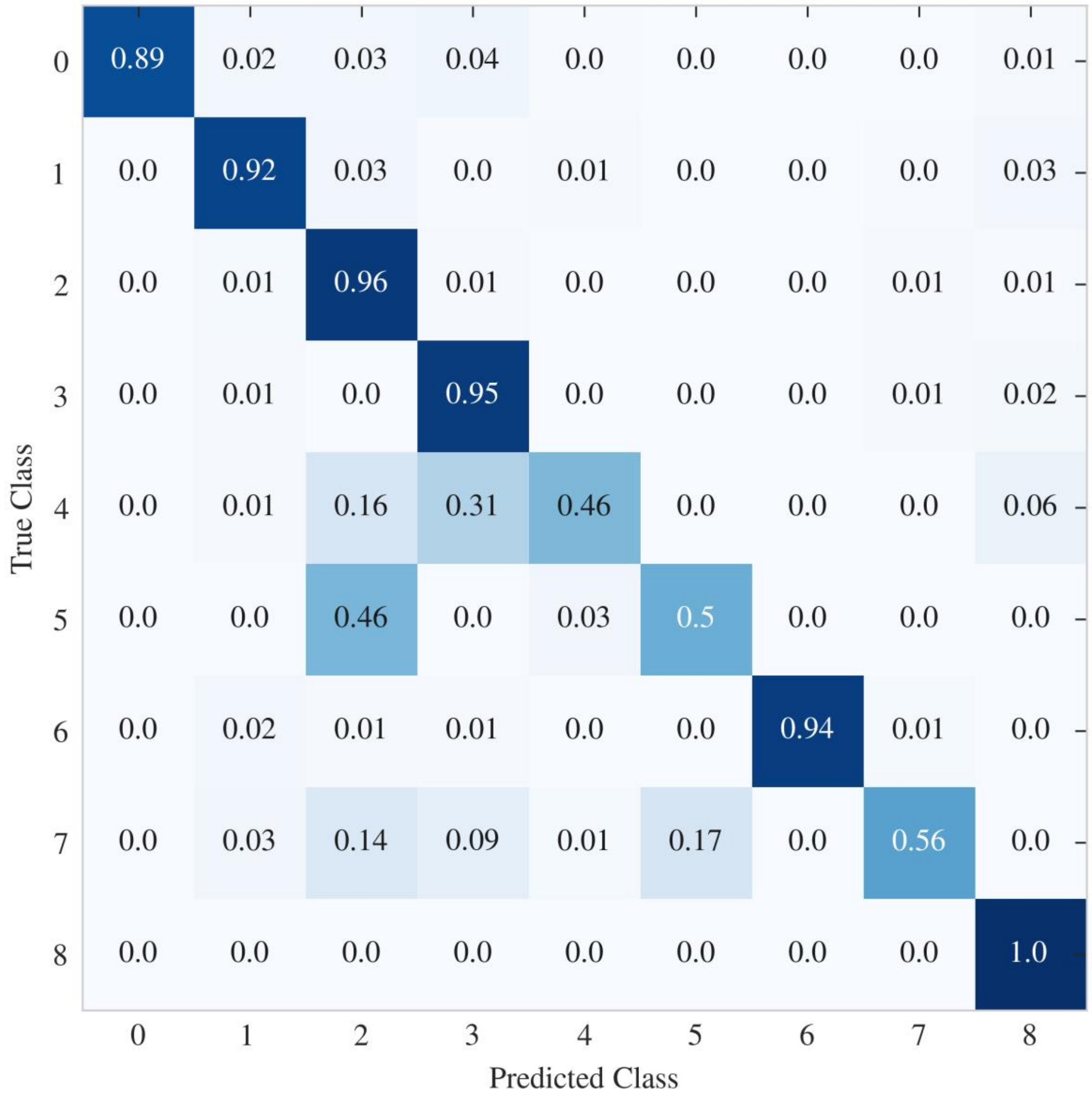
Key Changes

- Swapped ResNet-18 for ResNet-18-D and added ECA to each block
- Added 10% label smoothing in CE
- Increased max. LR by 4.5x and added linear warmup for 67.5% of training duration

Test Performance

Key Findings

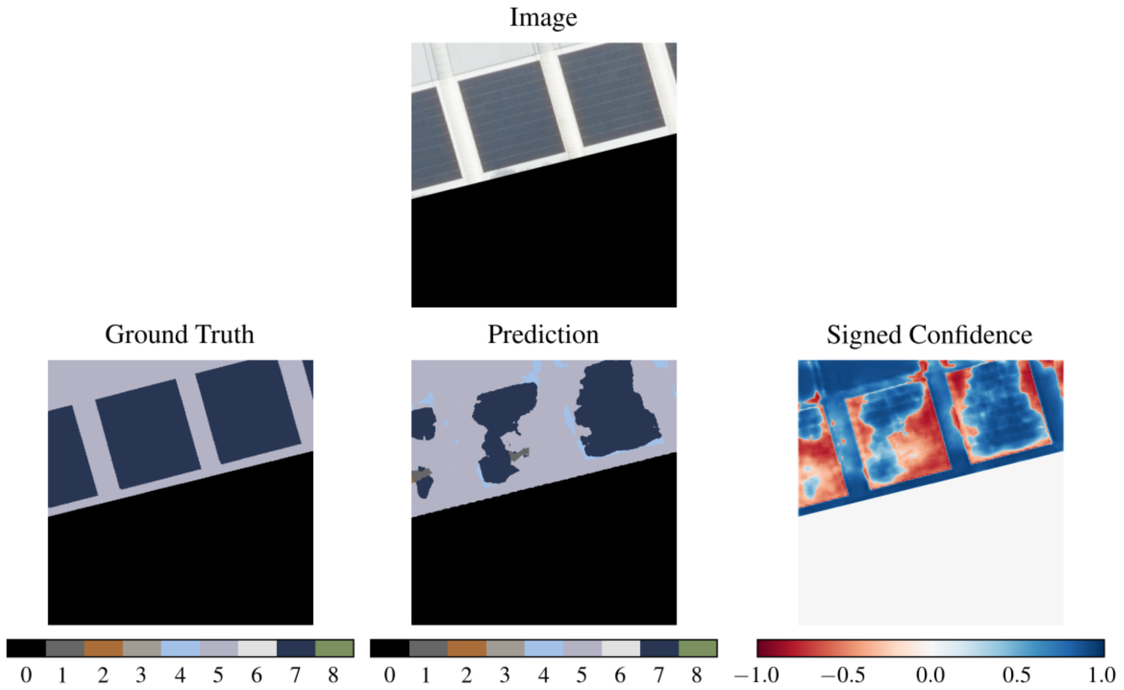
- OA: 85.15%; Avg. F_1: 78.09%
- mIoU: 64.68%
- Class division into three groups based on general performance:
  1. Dark-coloured Membrane
     Gravel
     Light-coloured Membrane
  2. Ceramic Tile
     Solar Panel
     Vegetation
  3. Light-permitting Surface (LPS)
     Metal



Test Performance

Key Findings

- OA: 85.15%; Avg. F_1: 78.09%
- mIoU: 64.68%
- Class division into three groups based on general performance:
  1. Dark-coloured Membrane
     Gravel
     Light-coloured Membrane
  2. Ceramic Tile
     Solar Panel
     Vegetation
  3. Light-permitting Surface (LPS)
     Metal



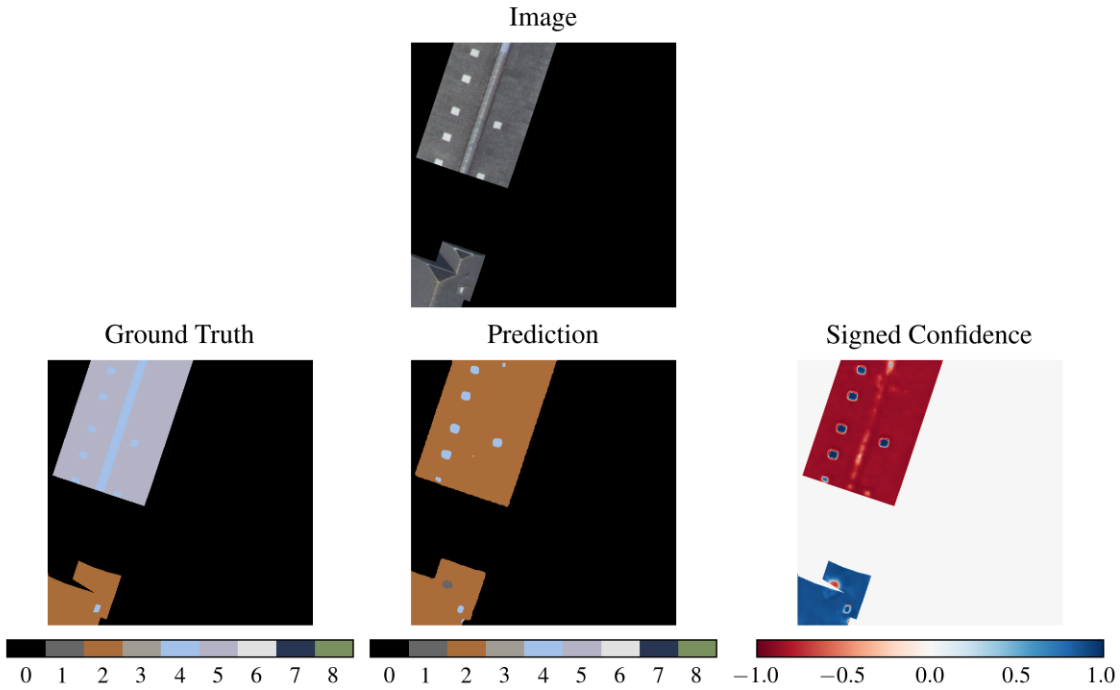
Key Issues

- **Incomplete solar panel detections**

Test Performance

Key Findings

- OA: 85.15%; Avg. F_1: 78.09%
- mIoU: 64.68%
- Class division into three groups based on general performance:
  1. Dark-coloured Membrane
     Gravel
     Light-coloured Membrane
  2. Ceramic Tile
     Solar Panel
     Vegetation
  3. Light-permitting Surface (LPS)
     Metal



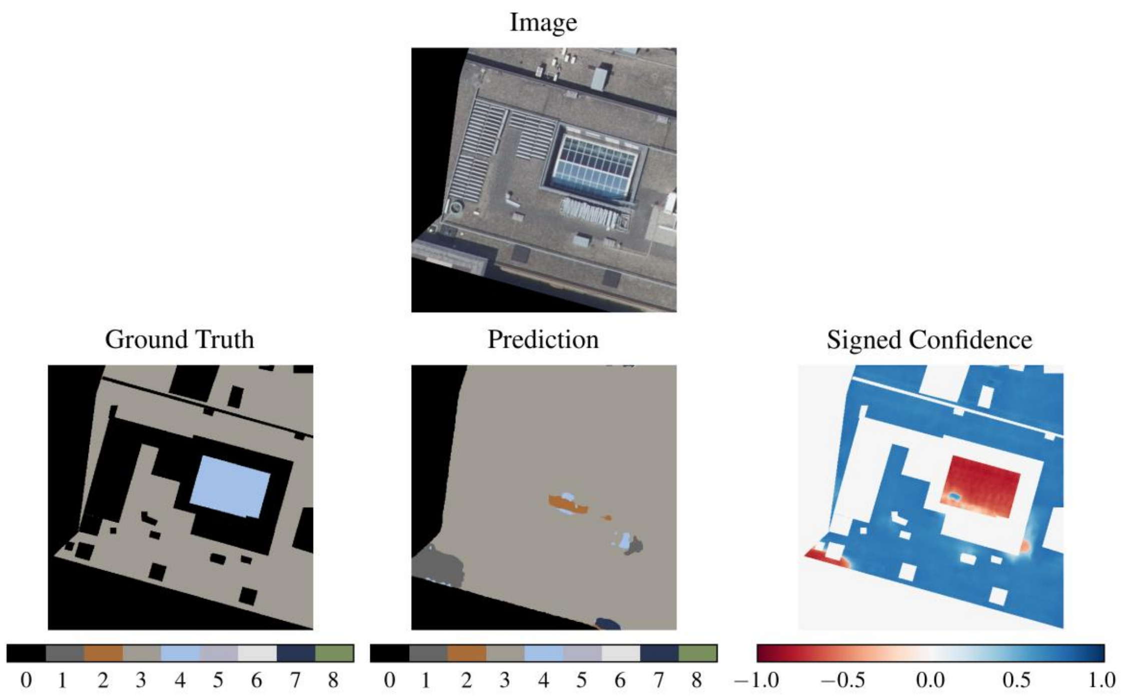
Key Issues

- Incomplete solar panel detections
- **Confusion between dark metal and ceramic tiles**

Test Performance

Key Findings

- OA: 85.15%; Avg. F_1: 78.09%
- mIoU: 64.68%
- Class division into three groups based on general performance:
    1. Dark-coloured Membrane
       Gravel
       Light-coloured Membrane
    2. Ceramic Tile
       Solar Panel
       Vegetation
    3. Light-permitting Surface (LPS)
       Metal



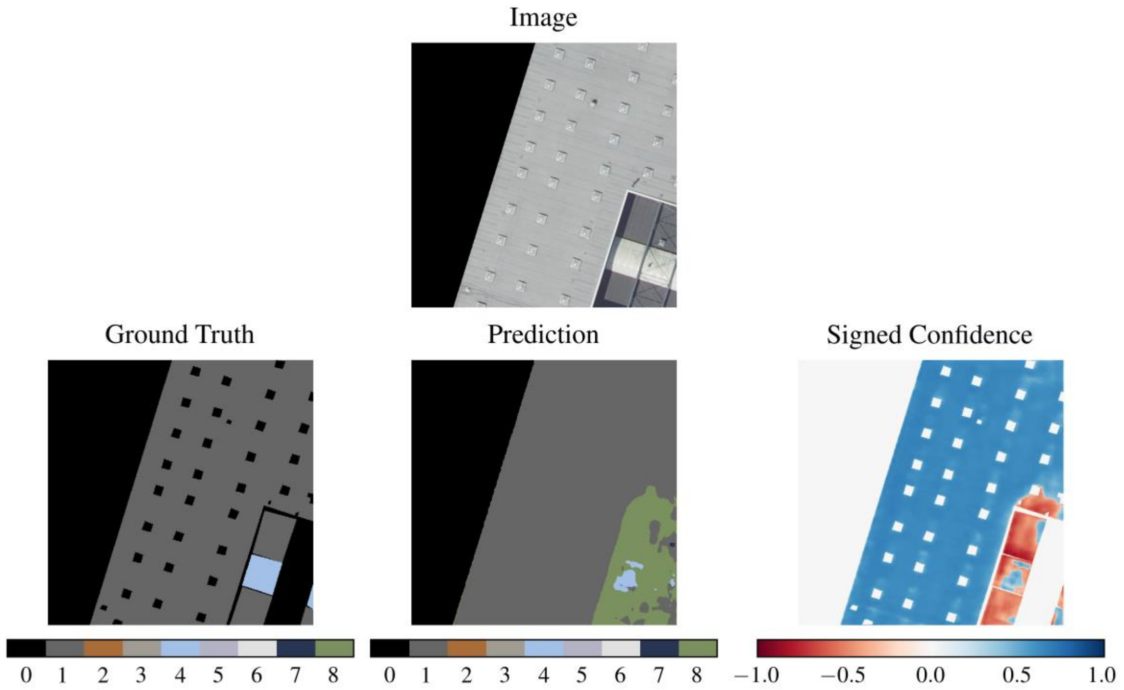
Key Issues

- Incomplete solar panel detections
- Confusion between dark metal and ceramic tiles
- **Failure to detect small LPSs**

Test Performance

Key Findings

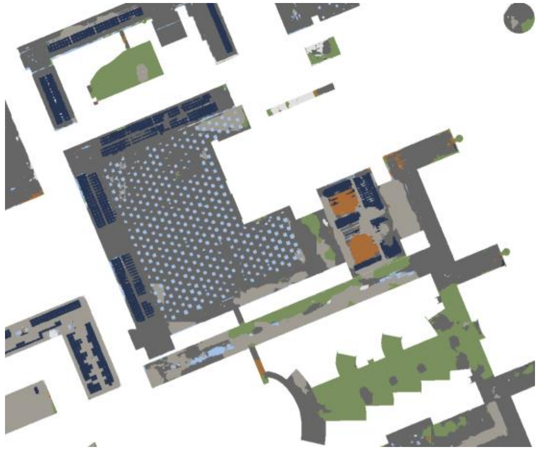
- OA: 85.15%; Avg. F_1: 78.09%
- mIoU: 64.68%
- Class division into three groups based on general performance:
  1. Dark-coloured Membrane
     Gravel
     Light-coloured Membrane
  2. Ceramic Tile
     Solar Panel
     Vegetation
  3. Light-permitting Surface (LPS)
     Metal



Key Issues

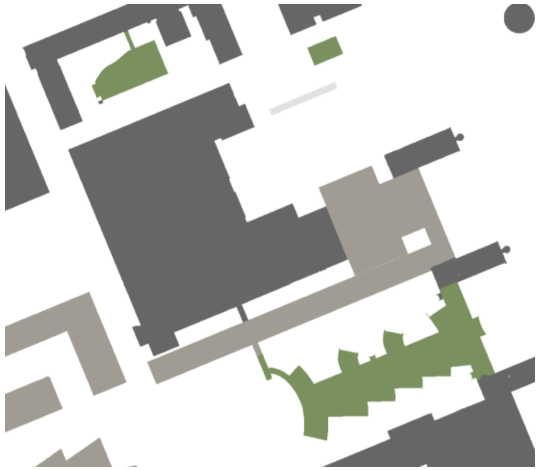
- Incomplete solar panel detections
- Confusion between dark metal and ceramic tiles
- Failure to detect small LPSs
- **Vegetation hallucinations**

Pixel-wise Material Map Generalisation



Pixel-wise Map

LoD1.2 Map



LoD1.3 Map

LoD2.2 Map

Generalisation can be applied as a post-processing step to cater to applications which do not require the full resolution (and errors) of pixel-wise maps

Generalised Test Performance

Key Findings

- Light-permitting surfaces absent from all LoDs
- Improved performance across the board except in metal and solar panels
- Large reduction in vegetation hallucinations
- Largest improvement observed in Group 2

LoD	Avg. Prec. (%)	Avg. Recall (%)	mIoU (%)
N/A	82.73	95.40	78.69
1.2	93.44	97.60	91.21
1.3	86.56	97.00	83.61
2.2	86.56	94.40	84.22

(*) Scores do not include metal and solar panel classes

Ablation Study

Key Findings

- Reflectance does not behave as NIR band as originally assumed
- Slope and nDRM most impactful
- Performance degradation due to density ablation not significant
- Confusion amongst certain classes increased or dropped similarly regardless of ablated band
- No LiDAR not too bad

Ablated Band	Avg. Prec. (%)	Avg. Recall (%)	mIoU (%)
N/A	77.55	78.63	64.68
Refl.	74.63	76.88	60.84
Slope	69.51	71.75	54.37
nDRM	71.47	74.25	56.07
Density	75.77	78.88	63.27
All	76.90	73.63	60.68

05

Conclusions

Research Questions

Review–based Questions

1. Which **imagery- and LiDAR-derived features** are the most contextually prevalent and effective in terms of performance?
   - Images: Spectral indices; PCA; no apparent performance gap between MSI/HSI and RGB/CIR
   - LiDAR: DTM, DSM, nDSM, intensity, slope
2. Which **classification and data fusion approaches** are the most contextually prevalent and effective in terms of performance?
   - Classification: Image-based methods competitive but suffer from label localisation issues; OBIA most widely used; Pixel-based techniques have no inherent technical limitations but still underexploited
   - Data Fusion: Band concatenation

Research Questions

Experiment–based Questions

3. How does the subsequent **generalisation of roofing material maps** to the roof segment level using each of the building LODs offered by the 3DBAG influence results?
    - No explicit performance improvement
        - Largely reduced gross errors
        - Small and incomplete predictions may be missing

    - LoD1.1 cannot model multi-material roofs with varying height
    - LoD1.3 solves this issue
    - LoD2.2 contains many gross errors but only option for dormer and skylight visualisation

Research Questions

Experiment–based Questions

4. How does the **availability of LiDAR-derived features** influence performance?
   - Slope and nDRM most impactful
   - Performance degradation due to density ablation not significant

   - LiDAR bands semantically coupled
     - Better to use them as a single dataset or not at all if not all bands are available

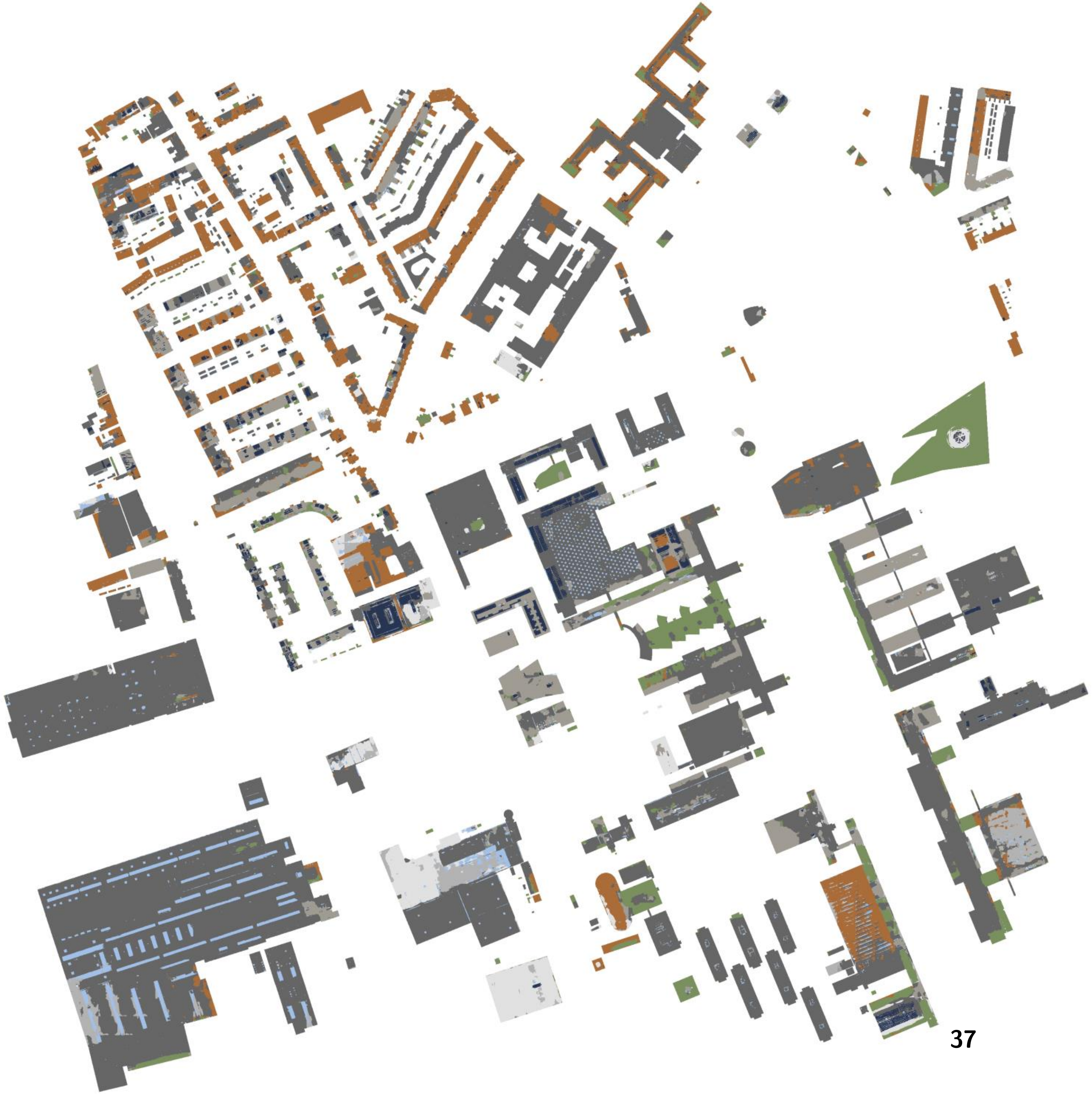
Contributions & Limitations

Scientific Contributions

- Dataset
- Dataset splitting and class weighting schemes
- Optimised model parameters
- Software implementation

Limitations

- Optimal performance in large, isolated buildings with flat roofs of uniform material
- Dataset size and quality



Future Work

High-priority

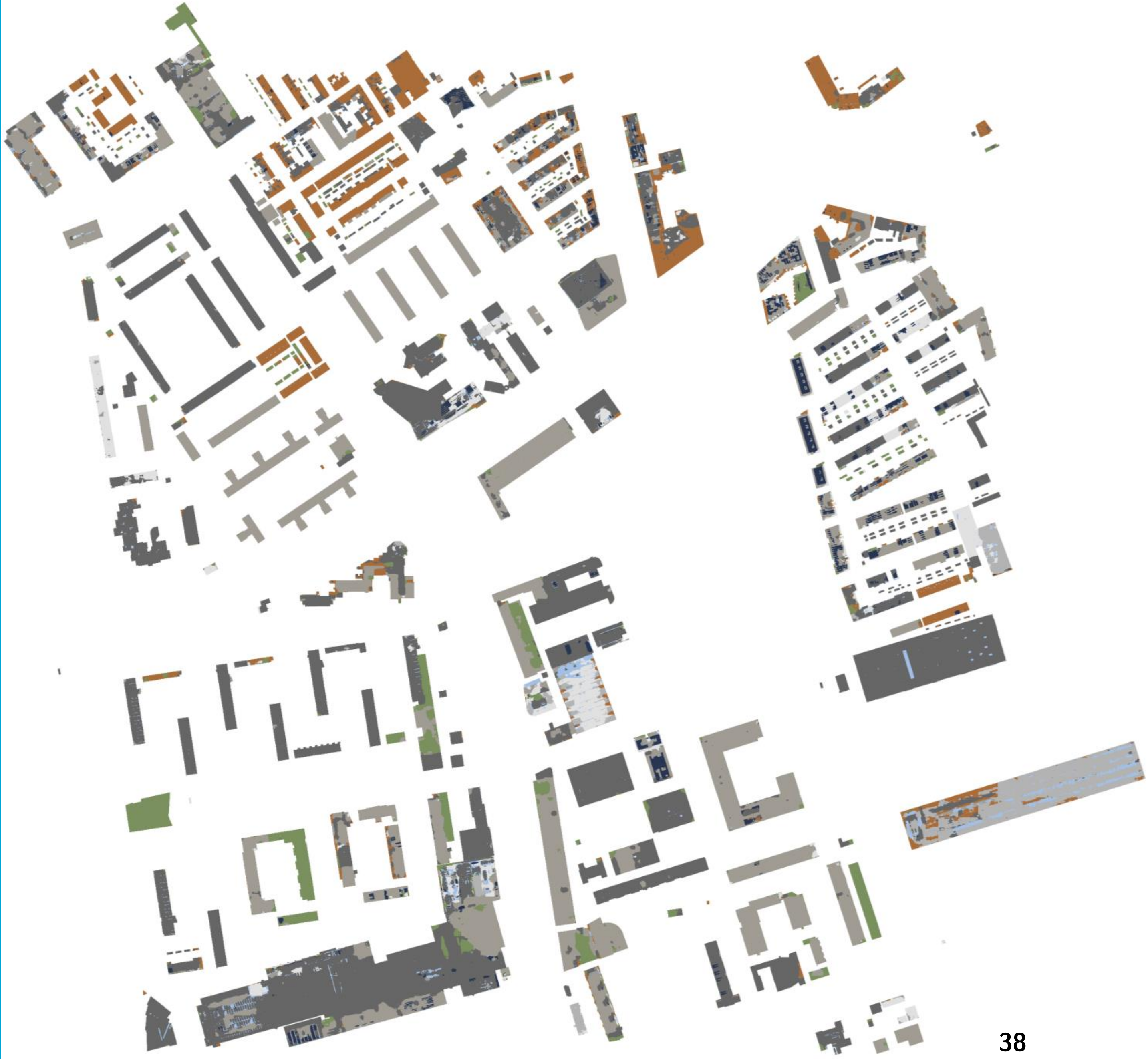
- Increasing the dataset size

Medium-priority

- Exploring (semi-) automated annotation methods and implementing quality control
- Exploring self-supervised pretraining methods

Low-priority

- Better datasets, more efficient pre-processing, extra features for small object detection



Thank you for your attention!

P5 Presentation | Dimitris Mantas

31-10-2024