Slate detection in orthophotos of building roof panels

Jiajun Li[1], Frédéric Bosché[1], Chris Xiaoxuan Lu[2], Lyn Wilson[3] and Boan Tao[1]

[1]School of Engineering, University of Edinburgh, UK
[2]Department of Computer Science, University College London, UK
[3]Historic Environment Scotland, UK

Jiajun.Li@ed.ac.uk, f.bosche@ed.ac.uk, xiaoxuan.lu@ucl.ac.uk, lyn.wilson@hes.scot, boan.tao@ed.ac.uk

Abstract -

Traditional roof inspection methods rely heavily on manual labour, which poses challenges in efficiency and safety. In this paper, we propose an overall approach and 4 specific methods for slate detection in orthophotos of building roof panels, using edge detection by Canny detector (Method 1), SAM (Segment Anything Model) (Method 2), SAM with bounding box prompt (Method 3), and edge detection on original SAM masks (Method 4). The methods are developed using a dataset composed of 4 sample roof segments and 3 complete orthophotos from real roof panels. Method 4 reaches the best results (overall results for all slates: mIoU = 0.67, Recall = 0.81) with Method 1 also performing well (mIoU = 0.57, Recall = 0.74). Such performance demonstrates the potential of the proposed approach and these methods to provide a robust foundation for slate detection and subsequently classification tasks.

Keywords -

Slate; Edge detection; Segmentation; Roof orthophoto

1 Introduction

Roof systems are a critical building fabric element serving to protect interior spaces and inhabitants against the elements (in particular rain). Slated roof systems are common type of roofs in both historic and modern buildings, requiring monitoring and timely maintenance to prevent damages, including structural damages. This is being made increasing challenging by climate change. As global weather patterns become increasingly erratic, slated roofs are subject to more frequent and severe fluctuations in temperature, moisture, and wind intensity. These environmental stressors can exacerbate the natural vulnerabilities of slates, a metamorphic rock that, while durable, has specific material properties—such as its anisotropic nature and inherent brittleness—that make it susceptible to damage under variable climate conditions [1, 2]. Besides, the nails used to secure slates can rust and fail. As a result, slated roofs often experience a range of deterioration issues, including broken, displaces, or detached (missing) slates.

The repair and maintenance of slated roof necessitate a targeted approach due to the localised nature of damage. Given that damage typically affects individual slates rather than large sections of the roof, restoration efforts often focus on a slate-by-slate basis. This time-saving and cost-effective approach is essential, not only to maintain the structural integrity and aesthetic quality of roofs but also to extend the lifespan of overall roofing systems. Therefore, a meticulous analysis of each individual slate is warranted to assess the type (if any) and extent of deterioration.

Current research on roof defect detection methods has advanced from manual inspection towards automation. For flat roofs, visual inspection by skilled experienced professional, with auxiliary means by sensors such as electric and ehermohygrometric methods, is the dominant method for non-destructive testing [3]. But, to capture data from buildings without the need for human access provision, Unmanned Aerial Vehicle (UAV)-based close-range sensing is increasingly used [4]. With the development of Artificial Intelligence (AI), deep learning models such as U-Net and HRNet are being developed to detect and categorise flat roof defects from aerial images, such as hollows, swellings, etc. These defects show continuous and large-area features [5]. In studies targeting at generic roof types, Mostafa et al. [6] apply Convolutional Neural Networks (CNNs) to classify roof damage types, achieving high accuracy in prioritising repairs.

However, during the evaluation and repair process of pitched roofs, workers must locate, assess, and restore individual slates at elevated heights, posing safety risks and inefficiencies. This further underscores the importance of precisely identifying the position and defect status of each slate to enable targeted physical intervention. As a result, pitched roofs, especially traditional pitched roofs, present unique challenges for automated and remote inspection: 1. Different from flat roofs, defects (mostly structural damages) of slated roofs are linked to single slate units. Therefore, defect detection should ideally start with slate mapping. 2. The complexity of slate arrangement and the variability in appearance, alongside variations in lighting and occlusions in roof images, further complicates automated analysis. 3. Defect categories of flat roofs differ

from those of pitched roofs, making existing defect detection model inapplicable. Addressing these challenges necessitates an advanced methodology capable of detecting individual slates and evaluating their condition in a robust and scalable manner.

Using UAV-based photogrammetric reconstruction, Digital Twin -based data processing, and Artificial Intelligence have been shown to be useful to automated slated roof monitoring Li et al. [7], from orthophoto generation for each roof panel [8] to roof sub-component segmentation [9]. These processes realise the gradual deconstruction of roof in a more measurable and analysable way. To further detect slate defects, one approach can be to first detect each individual slate in the roof orthophotos, and then classify each detected slate according to its condition, supporting high-precision diagnostics that can automate and streamline the repair process. This study proposes and compares a few methods of detecting individual slates, such as using traditional image processing in edge detection, the innovative Segment Anything Model (SAM).

2 Related work

2.1 Object detection using construction images

Image processing has emerged as a cornerstone of construction informatics, facilitating automated analyses for tasks like defect detection, progress monitoring, and material classification. Among them, object detection is the most common application. Simple yet effective techniques such as thresholding (e.g., Otsu's method) and filtering, are commonly applied in crack detection [10]. Canny edge-detector and watershed transformation are used in building component (column) detection [11]

With the advent of deep learning, modern models are built for object detection, like convolutional neural networks (CNNs), transformer-based vision architectures or more advanced zero-shpt SAM. For example, CNN refine input data for edge detection, making it easier for the network to learn spatial features. Such integrations enable applications like real-time construction site monitoring, where drones capture data that is processed for safety analysis [12], equipment detection [13]. While SAM-related research enables provide more accurate and comprehensive results [14]. Image processing in construction not only enhances operational efficiency but also addresses challenges like labour shortages and the need for faster project delivery, making it indispensable for modern construction workflows.

Based on object detection, segmentation helps further clarify the boundary of each object. To realise precise segmentation, edge detection-based methods and direct segmentation methods can be considered.

2.2 Edge detection -based methods

Traditional edge detection -based methods focus on explicitly identifying intensity discontinuities in images, often marking the boundaries of objects. These methods typically involve gradient-based techniques, such as:

1. *Canny Edge Detection* combines Gaussian smoothing, gradient calculation, non-maximum suppression, and hysteresis thresholding to produce clean and precise edges [15].
2. *Sobel and Prewitt Operators* compute gradients in horizontal and vertical directions to highlight regions with high intensity variation, making them ideal for detecting straight lines or simple contours [16, 17].
3. *Laplacian of Gaussian (LoG)* identifies edges by detecting zero-crossings in the second derivative of the image, providing robust results for blob-like structures [18].

Recent advancements in deep learning have significantly enhanced traditional edge detection methods, shifting from purely gradient-based approaches to data-driven, learning-based frameworks. Modern edge detection methods leverage convolutional neural networks (CNNs) and other architectures to achieve robust and adaptive performance in diverse scenarios. For instance, models like Holistically-Nested Edge Detection (HED) use deep CNNs to learn hierarchical edge features, allowing for precise edge maps with context-aware refinement [19]. Transformer-based approaches, such as DETR (DEtection TRansformer), implicitly model edge-like information by attending to boundary regions during object localisation and segmentation tasks [20].

2.3 Segment Anything Model (SAM)

Segment Anything Model (SAM) is a groundbreaking zero-shot learning model designed for universal image segmentation. It is trained using segmentation dataset with over 1 billion masks on 11M images. Unlike task-specific models, SAM employs a flexible prompting mechanism, enabling it to generalise across diverse image types and tasks without fine-tuning. It accepts various prompts such as points, bounding boxes, and even textual descriptions, making it adaptable to many domains. For example, SAM can segment complex natural scene images with minimal human intervention, showcasing its versatility [21].

Traditional edge detection methods offer efficient, lightweight solutions for extracting structural details, especially in scenarios where computational resources are limited, while deep learning approaches excel in capturing complex patterns and contextual information through data-driven feature learning. SAM adds a high-level semantic understanding to segmentation tasks, enabling precise object delineation across diverse and challenging vi-

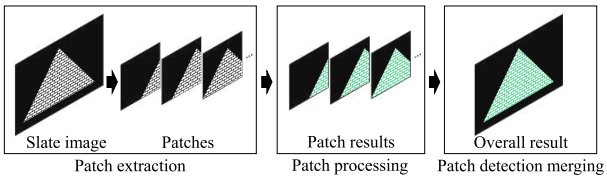


Figure 1. Approach overview

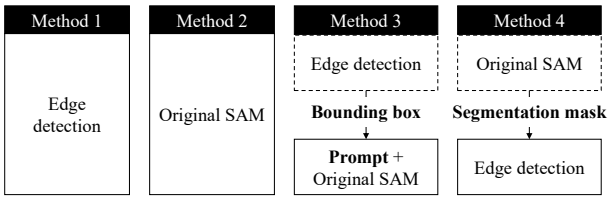


Figure 2. Method development of slate detection

sual contexts.

3 Approach overview

Roof panels, and consequently roof panel orthophotos come in various sizes. Since the performance slate detection algorithms is dependent on the size of the slates in the processed image, it is necessary to process each roof panel orthophoto in patches representing roofs at similar scale. As a result, the overall pipeline for slate detection in roof panel orthophotos includes the following three steps (fig. 1):

1. **Patch extraction**: the image is split into an overlapping set of patches;
2. **Patch processing**: slates are detected in each patch; and
3. **Patch detection merging**: all detections are merged in the original image.

Sections 4 to 6 detail those three steps.

4 Patch extraction

Small fixed-size patches (200×200 pixels) are extracted from the complete slate images using a sliding window approach. Square patches are the best option for inputting in SAM model to prevent distortion. The step size is 25% of the window size (i.e. 50 pixels), creating overlapping patches. This ensures that each slate will appear fully in at least one patch. The mask of the slated area in the orthophoto is also a known input from the orthophoto generation theory [8].

5 Slate detection

Four slate detection methods are investigated in this study (see fig. 2). They are developed according to the following logic: **Method 1** is developed based on traditional image processing using edge detection. Thanks to the well-developed AI model by meta, SAM, **Method 2** is developed by using SAM as is. However, as will be shown, **Method 2** shows lower performance than **Method 1**. As a result, Methods 3 and 4 are investigated. **Method 3** uses the bounding boxes obtained using **Method 1** as prompt to SAM. **Method 4** employs the same detection strategy of **Method 1** but applied to the SAM segmentation map. The 4 methods are described in more detail in the following sub-sections.

5.1 Method 1: Edge detection by Canny detector

The process of the application of **Method 1** to each patch is as follows:

1. *Noise removal*: Gaussian blur is applied to the patch (or the overall image) to reduce noise.
2. *Edge detection*: The Canny edge detector is applied to the patch, followed by oriented horizontal edge detection using Sobel.
3. *Slate row detection:* Probabilistic Hough Transform-based line detection is applied to the edge map to identify horizontal slate rows in the image. Only lines where the absolute difference between the y-coordinates of the two endpoints is less than 10 are retained. The most common distance between neighbouring lines — which should correspond to the roof panel *exposure*, also called *margin* — is calculated, and any line whose distances to neighbouring lines are smaller than the exposure is discarded. Similarly, lines are added between pairs of lines whose distance is twice (or three times) the exposure. The final set of lines is used to split the image into strips, with each strip expected to contain one slate row.
4. *Slate detection*: For each slate row, the average vertical grayscale value is computed across the strip, and valleys in this resulting 1D signal are assumed to correspond to the locations of vertical boundaries between slates in the strip, and the strip is segmented correspondingly into boxes defining initial slate boundaries, The slates boxes detected across all strips are gathered in the patch.
5. *Slate detection filtering*: Boxes that contain more than 90% black pixels (i.e. background pixels in the orthophoto that are not part of the roof panel) are discarded as noise.
6. *Slate detection adjustment*: Because slates may not be laid in a strict horizontal direction, each slate box is slightly resized by locally adjusting their top and bottom edges to better match the local minimum average grayscale along the edges. The resizing is restricted

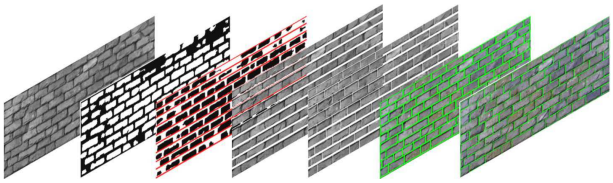


Figure 3. Method 4 pipeline, left to right: grayscale image - SAM segmentation mask - line extraction based on Hough Transform - strip extraction - strip valley calculation - box visualisation - box shifting.

within a distance set to 30% of the slate exposure.

5.2 Method 2: Original SAM

In **Method 2**, the original SAM is used without any other input to generate the probability map of segments, followed by a segmentation mask by setting the threshold as the median value of the entire probability map. Bounding boxes are then generated by extracting the largest connected components after removing noise (small objects, pixel count < 25) and smoothing edges. Due to the transition area from slate to slate gap, segments by SAM tend to be smaller than the actual slates. Therefore, a two pixel extension is added to the four boundary sides to make the boxes closer to reality and touch each other. Theoretically, this approach may be able to better generalise than **Method 1**, since SAM offers strong generalisation capabilities and high-level semantic understanding.

5.3 Method 3: SAM with bounding box from edge detection as prompt

In **Method 3**, the original SAM is used with the bounding box information from **Method 1** used as prompt, with the aim of increasing the probability of a slate segment inside the box while retaining the theoretical capability of SAM to detect slates that may not have a rectangle appearance. The detection box refinement steps of **Method 1** are then applied the same way.

5.4 Method 4: Edge detection by original SAM mask

Because we noticed that the segmentation masks produced by SAM have clear slate boundaries, possibly clearer than in the original grayscale image, in **Method 4** SAM is applied to the grayscale image, outputting a probability map indicating pixel-level segmentation confidence. By setting the threshold as the median probability value, a segmentation mask is generated to which the Canny edge detector of **Method 1** is applied. The detection box refinement steps of **Method 1** are then applied the same way.

6 Patch slate detection results merging

All slate detections from all patches are gathered in the overall roof panel orthophoto. The resulting set of detections contains redundant detections from overlapping detections as well as boxes with abnormal sizes (noise). To merge and filter those results, the following steps are applied:

1. *Boundary boxes*: Boxes located on the boundary of overlapping patches are discarded.
2. *Small boxes*: Boxes whose height or width is $< 0.5 \times exposure$ are discarded, but only if $> 90\%$ of the area of the orthophoto covered by the box is already covered by another box.
3. *Redundant boxes*: Non-Maximum Suppression (NMS) is used to discard redundant (overlapping) boxes. When boxes are overlapping, NMS keeps the most significant one based on a predefined Intersection-over-Union (IoU) threshold (0.3). However, like above, a box is discarded only if $> 90\%$ of the area of the orthophoto that it covers is already covered by another box.

7 Experimental results

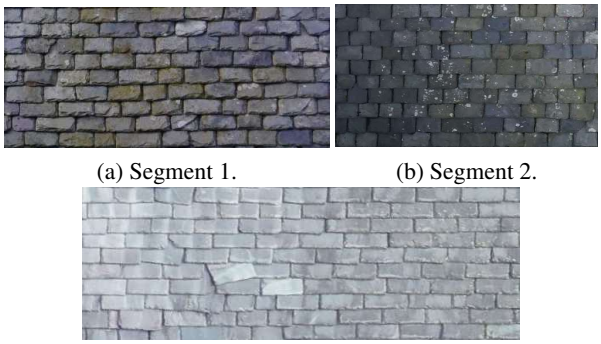
7.1 Dataset

The dataset used in the study is a set of orthophotos obtained from real slated roof panels using the method in [7]. The orthophotos are of roof panels with different shapes and deterioration conditions (see fig. 7) . Panel A and B are expected to be easier due to their smaller size, while Panel C is much larger and has the highest deterioration level. The three panels collectively contain 5,386 slates. Ground Truth (GT) labelling for each slate is done manually with axis-aligned bounding box (even though some slates are not rectangular). The results are presented for sample images first and then for the overall roof panels.

7.2 Performance evaluation metrics

Recall and Intersection over Union (IoU) are widely used metrics for evaluating object detection and segmentation tasks. Recall helps evaluate how well a model identifies all objects in an image, while IoU provides a robust measure of spatial overlap between predicted and ground truth (GT) regions[22]. Slate detection performance is thus evaluated here using both Recall and IoU. Note that each predicted box is matched to the GT box with highest IoU, and classified according to the IoU value as summarised below:

- *TP (True Positive)*, if $IoU >= 0.5$.
- *FP (False Positive)*, if $0.1 <= IoU < 0.5$. It is further classified into:



(a) Segment 1. (b) Segment 2.

(c) Segment 3.

Figure 4. Three segments extracted from the three roof panels.

Method	mIoU	Recall
1	0.86	0.97
2	0.62	0.65
3	0.76	0.89
4	0.86	0.97

Table 1. Performance (mIoU and Recall) of the four slate detection methods applied to the three sample orthophoto segments.

  - *FP - Over segmented*, if GT box size > Prediction box size.
  - *FP - Under segmented*, if GT box size < Prediction box size.
- *FN (False Negative)*, if $IoU < 0.1$.

Overall performance is evaluated using:

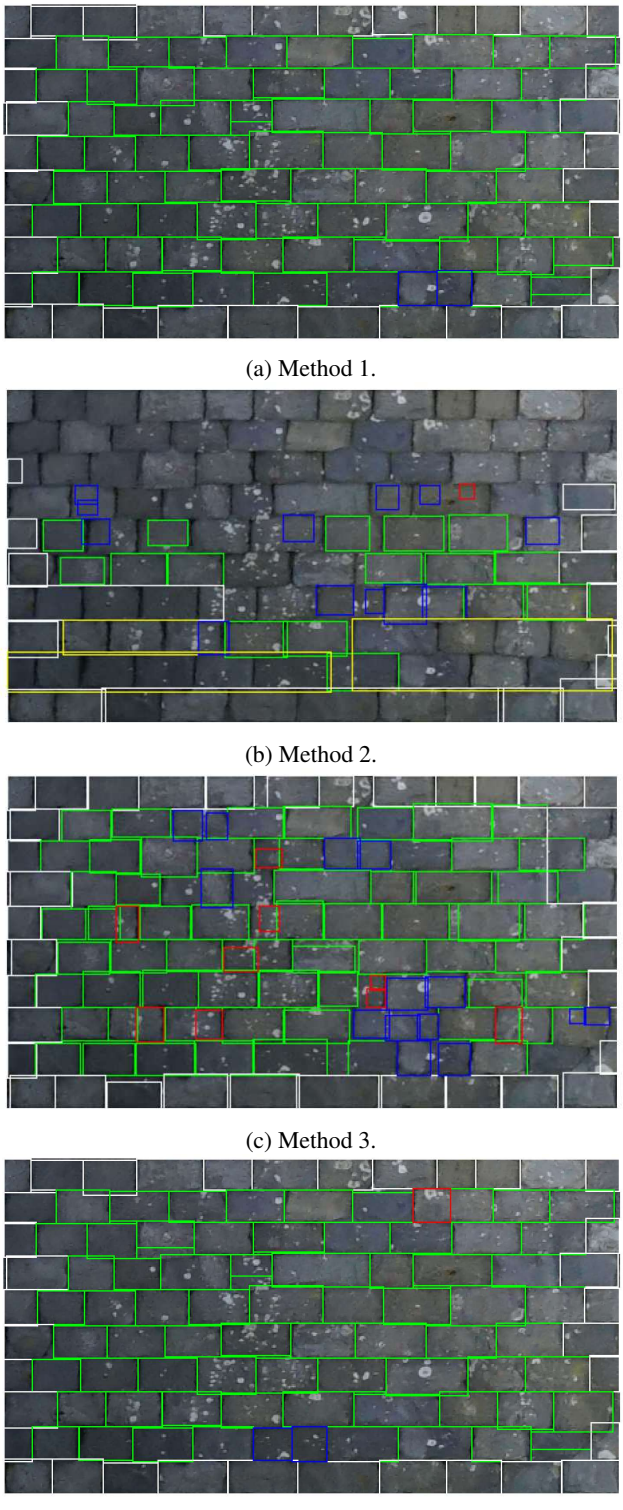
- *mIoU*, the mean IoU values;
- *Recall*, the proportion of slates correctly detected.

7.3 Detection results on sample roof segment

Results using the 4 different methods were first obtained for three roof segments extracted from the three panels, shown in fig. 4. The results for the three segments are summarised in table 1, and results can be visualised for one segment in fig. 5.

Note that the slates located at the boundaries of the segments are typically incomplete, and should thus be discarded from the performance analysis. The three segments collectively contain 202 non-boundary slates, and 114 discarded boundary slates, and the results in table 1 are obtained by considering only the non-boundary slates.

The first observation is that **Method 1** clearly outperforms **Method 2**. **Method 2** performs poorly because some its segmentation map sometimes includes small connections between the detecting areas of neighbouring slates leading to the detection of one large (that is subsequently filtered out) instead of two slates.



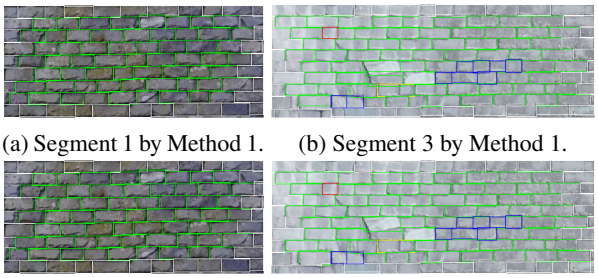
(a) Method 1.

(b) Method 2.

(c) Method 3.

(d) Method 4.

Figure 5. Slate detection results obtained by the four methods on Segment 2 (White: Boundary; Green: TP; Blue: FP - Over Segmented; Yellow: FP - Under segmented; Red: FN).



(a) Segment 1 by Method 1. (b) Segment 3 by Method 1.

(c) Segment 1 by Method 4. (d) Segment 3 by Method 4.

Figure 6. Visualisation of results obtained with Method 4 for Segments 1 and 3.

Method	Panel	Non-boundary		All slates	
		mIoU	Recall	mIoU	Recall
1	A	0.72	0.89	0.68	0.85
	B	0.57	0.78	0.54	0.73
	C	0.57	0.75	0.56	0.73
	Overall	0.58	0.77	0.57	0.74
4	A	0.78	0.93	0.76	0.91
	B	0.63	0.83	0.59	0.78
	C	0.69	0.83	0.67	0.81
	Overall	0.69	0.84	0.67	0.81

Table 2. Comparison between different segmentation methods in raw slate images.

The results of **Method 3** demonstrate an obvious improvement compared to **Method 2**. The bounding box prompt helps increase the probability of segments inside each prompt bounding box, making **Method 3** distinguish individual slates more robustly than **Method 2**.

Method 4 performs even better, to a similar level to **Method 1**. However, because of the mechanism behind the two methods (Canny detector and SAM segmentation mask), they differ in their sensitivity to edges, resulting in both methods seemingly failing in different locations.

fig. 6 show the results obtained with **Method 4** on the remaining two segments.

7.4 Detection results on complete slate images

The results obtained for the three orthophoto segments in section 7.3 clearly show the superiority of **Method 1** and **Method 4**. Therefore the performance on the complete orthophotos is now evaluated for only those two methods. The results are summarised in table 2 and those obtained with **Method 4** can be visually explored in fig. 7. Note that the analysis distinguishes the performance with and without *boundary* slates, that are the slates whose GT box intersects the slate mask, i.e. intersect the orthophoto background of black pixels. The reason for this distinction is that boundary slates typically present a greater range of shape (e.g. truncated) and appearance (e.g. presence of leadwork pixels), which may challenge the detectors.

Overall, **Method 4** performs best, with $mIoU = 0.67$ and $Recall = 0.81$. The better seems because the SAM segmentation map demonstrates clearer slate boundaries than the Canny edge detector. Such advantage is more obvious in the complete orthophotos that include a broader range of complex shape and conditions (e.g. boundary slates, holes, etc.).

However, **Method 4** still makes a number of errors. For example, there are a lot of noisy bounding boxes generated in areas containing biological growth (see the upper and right part of Panel C). The uneven brightness in the bottom part of Panel A also affects the vertical slate edge detection, leaving some non-boundary slates undetected. The unclear slate boundary in the top part of Panel B results in more slates detected in FP (Under segmentation) status.

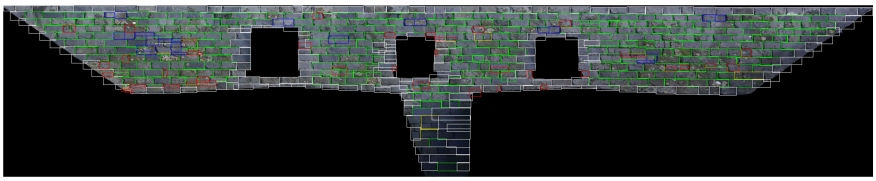
The comparison of the results for all slates and non-boundary slates shows a drop of 2 percentage points in the performance for all slates. As noted above, a drop was expected due to the fact that boundary slates tend to have a greater range of shapes and appearance. However, the drop appears to be only minor.

8 Discussion

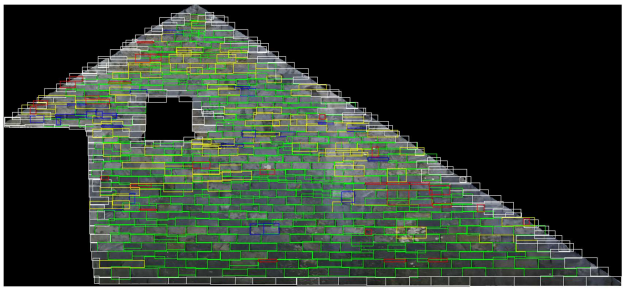
The results of **Method 1** and **Method 4** already demonstrate promising results, with mIoU=0.67 and recall=0.81 for **Method 4** and mIoU=0.57 and recall=0.74 for **Method 1**, which can form the basis for defect assessment of individual slates, e.g. through classification. However, the orthophoto quality noticeably influences the accuracy, especially in Panel B where the slates at the bottom of the panel are not as clear. This motivates further work to improve data collection and orthophoto generation. Besides, FPs are noticed to be common in areas where there is the presence of biological growth. Therefore, methods should be explored to better discriminate slate boundaries from biological growth in the image input to the Hough Transform-based line detection (from Canny edge detector or SAM segmentation map).

We note that the proposed slate segmentation approach is designed for typical traditional slated roofs (laid in in horizontal rows) and does not require any roof-specific parameters. The approach is thus applicable to most slated buildings. However, slated roofs may at times have different patterns, with slates cut and laid in other ways (e.g. fish scale patterns). For those roofs, **Method 1** and **Method 4** would not be applicable, but **Method 3** could prove suitable.

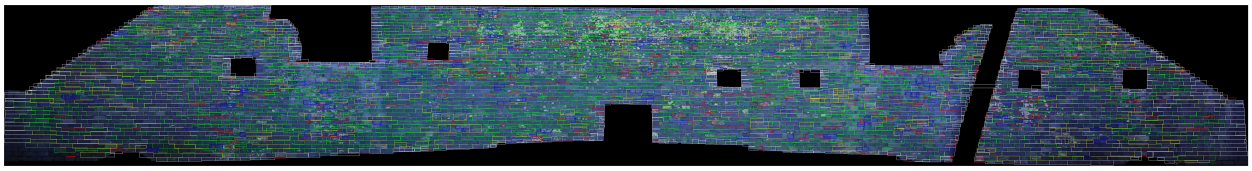
Finally, while SAM is shown to be able to highlight edge features, some alternative deep learning model for image segmentation (and slate detection) could be explored.



(a) Panel A.



(b) Panel B.



(c) Panel C.

Figure 7. Visualisation of complete slate image results obtained by Method 4.

9 Conclusion

This study proposed an overall approach and 4 specific methods for slate detection in orthophotos of building roof panels, comparing: edge detection by Canny detector (**Method 1**), SAM (**Method 2**), SAM with bounding box prompt from **Method 1** (**Method 3**), and edge detection on the original SAM mask (**Method 4**). All four methods were first evaluated on three cropped orthophoto segments, and **Method 1** and **Method 4** stood out with similar and high mIoU (0.86) and Recall (0.97). The two methods were then further evaluated on complete roof panel orthophotos, on which **Method 4** showed superior performance. The results highlight the encouraging levels of performance of **Method 1** and **Method 4**, but also their difficulty in areas containing significant biological growth and their sensitivity to roofs acquired in poorly lit conditions. Their distinct performance in different areas also suggest that future research could — alongside evaluating the methods on more extensive datasets — possibly explore integrating their results.

10 Acknowledgements

This paper was made possible thanks to research funding from Historic Environment Scotland (HES) and the Engineering and Physical Sciences Research Council (EPSRC) [grant reference EP/W524384/1]. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of HES and EPSRC. The authors would also like to acknowledge the HES Digital Documentation and Innovation team, Stirling City Heritage Trust for providing us with the data used in the experiments reported in this paper.

References

[1] J. Wichert. *Mining of Slate and Production*, pages 159–201. Springer International Publishing, Cham, 2020. doi:10.1007/978-3-030-35667-5_6.

[2] V. Cárdenes, J. P. Cnudde, J. Wichert, D. Large, A. López-Mungira, and V. Cnudde. Roofing slate standards: A critical review. *Construction and Building Materials*, 115:93–104, 2016. doi:10.1016/j.conbuildmat.2016.04.042.

[3] J. Conceição, B. Poça, J. de Brito, I. Flores-Colen, and A. Castelo. Data analysis of inspection, diagnosis, and rehabilitation of flat roofs. *Journal of Performance of Constructed Facilities*, 33(1):04018100, 2019. doi:10.1061/(ASCE)CF.1943-5509.0001252.

[4] E. Adamopoulos and F. Rinaudo. Close-range sensing and data fusion for built heritage inspection

and monitoring—a review. *Remote Sensing*, 13(19), 2021. doi:10.3390/rs13193936.

[5] D. A. Yudin, V. Adeshkin, A. V. Dolzhenko, A. Polyakov, and A. E. Naumov. Roof defect segmentation on aerial images using neural networks. In *Advances in Neural Computation, Machine Learning, and Cognitive Research IV*, pages 175–183, 2021.

[6] K. Mostafa, T. Hegazy, R. D. Hunsperger, and S. Elias. Using image analysis to quantify defects and prioritize repairs in built-up roofs. *Facilities.*, 41 (7/8), 2023-04-26.

[7] J. Li, B. Tao, F. Bosché, C. X. Lu, and L. Wilson. Automated generation and semantic segmentation of roof orthophoto for digital twin -based monitoring of slated roofs. *Automation in Construction*, 167: 105725, 2024. doi:10.1016/j.autcon.2024.105725.

[8] J. Li, F. Bosché, C. X. Lu, and L. Wilson. Occlusion-free orthophoto generation for building roofs using uav photogrammetric reconstruction and digital twin data. In *Proceedings of the 40th ISARC*, pages 371–378, Chennai, India, July 2023. doi:10.22260/ISARC2023/0051.

[9] J. Li, B. Tao, F. Bosché, C. X. Lu, and L. Wilson. Extracting roof sub-components from orthophotos using deep-learning -based semantic segmentation. In *Proceedings of the 41st ISARC*, pages 675–682, Lille, France, June 2024. doi:10.22260/ISARC2024/0088.

[10] B. Kim and S. Cho. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Structural Control and Health Monitoring*, 26(8):e2381, 2019. doi:10.1002/stc.2381.

[11] Y. Wu, H. Kim, C. Kim, and S. H. Han. Object recognition in construction-site images using 3d cad-based filtering. *Journal of Computing in Civil Engineering*, 24(1):56–64, 2010. doi:10.1061/(ASCE)0887-3801(2010)24:1(56).

[12] M. Z. Shanti, C. Cho, B. Garcia de Soto, Y. Byon, C. Y. Yeun, and T. Y. Kim. Real-time monitoring of work-at-height safety hazards in construction sites using drones and deep learning. *Journal of Safety Research*, 83:364–370, 2022. doi:10.1016/j.jsr.2022.09.011.

[13] H. Kim, H. Kim, Y. W. Hong, and H. Byun. Detecting construction equipment using a region-based fully convolutional network and transfer learning. *Journal of Computing in Civil Engineering*, 32(2):04017082, 2018. doi:10.1061/(ASCE)CP.1943-5487.0000731.

[14] M. Ahmadi, A. G. Lonbar, H. K. Naeini, A. T. Beris, M. Nouri, A. S. Javidi, and A. Sharifi. Application of segment anything model for civil infrastructure defect assessment, 2024.

[15] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. doi:10.1109/TPAMI.1986.4767851.

[16] J. MS. Prewitt et al. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1): 15–19, 1970.

[17] I. Sobel. Neighborhood coding of binary images for fast contour following and general binary array processing. *Computer Graphics and Image Processing*, 8(1):127–135, 1978. doi:10.1016/S0146-664X(78)80020-3.

[18] D. Marr, E. Hildreth, and S. Brenner. Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167): 187–217, 1980. doi:10.1098/rspb.1980.0020.

[19] S. Xie and Z. Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. doi:10.1109/ICCV.2015.164.

[20] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *Computer Vision – ECCV 2020*, pages 213–229, Cham, 2020. doi:10.1007/978-3-030-58452-8_13.

[21] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W. Lo, P. Dollár, and R. Girshick. Segment anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3992–4003, 2023. doi:10.1109/ICCV51070.2023.00371.

[22] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. doi:10.1007/s11263-009-0275-4.