Proceedings of the Seventh International Conference on
Artificial Intelligence, Soft Computing, Machine Learning and Optimization,
in Civil, Structural and Environmental Engineering
Edited by: P. Iványi, J. Kruis and B.H.V. Topping
Civil-Comp Conferences, Volume 11, Paper 1.3
Civil-Comp Press, Edinburgh, United Kingdom, 2025
ISSN: 2753-3239, doi: 10.4203/ccc.11.1.3

Deep Learning Methods for the Analysis of Townscapes

S. Balestra[1], O. Hänni[2], M.-A. Iten[3], M. Blöchlinger[2], S. Bühler-Krebs[2] and R.-P. Mundani[1]

[1] Institute for Data Analysis, Artificial Intelligence, Visualization and Simulation (DAViS), University of Applied Sciences of the Grisons, Chur, Switzerland
[2] Institute for Construction in Alpine Regions (IBAR), University of Applied Sciences of the Grisons, Chur, Switzerland
[3] Swiss Institute for Information Science (SII), University of Applied Sciences of the Grisons, Chur, Switzerland

Abstract

Finding architectural principles in facades is an important task for urban development. Hence, using AI-based methods for an automated analysis seems obvious; but this also entails certain requirements concerning the training data itself. A new facade dataset with 14 segmentation masks was created and used for the training of deep learning models to semantically segment elements within facades of Swiss buildings. Via a rule-based approach, architectural principles such as rhythm lines and axes of symmetry can be derived from these elements. Those principles, especially in the context of neighbouring buildings, form an architectural pattern that is partially quantifiable to assure quality in design w. r. t. urban development.

Keywords: artificial intelligence, deep learning, architecture, urban development, building culture, semantic segmentation

1 Introduction

Settlements in the alpine region have been experiencing a constant change due to socio-economic and social processes. Urban development and architectural concepts in the Swiss alpine region have been active research areas for many years. Especially the decisions by the legislature have been a focus point in the field of urban planning.

During the last century, many historically valuable centres have been developed that are nowadays in the focus of building activity and undergoing transformation. Hence, the government has a strong motivation to protect, preserve, and carefully develop those centres. For this reason, in the 1970s a unique inventory of the most outstanding settlements was compiled: The Federal Inventory of Heritage Sites of National Importance (ISOS). The aim of ISOS is to collect and describe – on the basis of objective criteria – settlements in Switzerland. It became an important cornerstone in the spatial planning of the Confederation, the cantons, and the municipalities.

Based on the methodology of architectural contextualism, we have developed a tool (Baumemorandum) for architecture offices and municipalities that works as guidelines for assuring quality in design and urban development in critical, historical areas. The Baumemorandum consists of three elements: fundamentals, context memorandum, and facade memorandum. These are available – to date – in printed form only. Since most of the essential information for construction projects are nowadays provided by geographic information systems (GIS), it is therefore beneficial to make this tool digitally available within the latter ones. Figure 1 shows an example of a street with houses as part of a facade memorandum.

By making this tool accessible directly from GIS, the municipality's decision-making process becomes more transparent and the planning entity can take neighbouring buildings into account. Viewing the planned project in the context of its neighbourhood, spatial connections, design relationships, and architectural exceptions can be considered. The representation of spatial relationships and the search for typical characteristics (architectural principles) were done manually (i. e. time-consuming drawing work) and, thus, prone to errors. The processed data serves as a basis for validation and interpretation by domain experts. Integrating the entire approach into GIS finally enables the usage of the Baumemorandum in a larger context and leverages simple analyses of old and new data. Combining regulations, specifications, and information important to urban development into a single information channel is an important step towards simplification of planning processes. It allows deciders to streamline decisions and new developments as well as to support involved experts such as architects. The digital Baumemorandum creates legal certainty, particularly with regard to the question of appropriate integration into the townscape. As partner of our project, Davos Monstein (Canton of Grisons) will be the first municipality in Switzerland to implement the digital Baumemorandum.

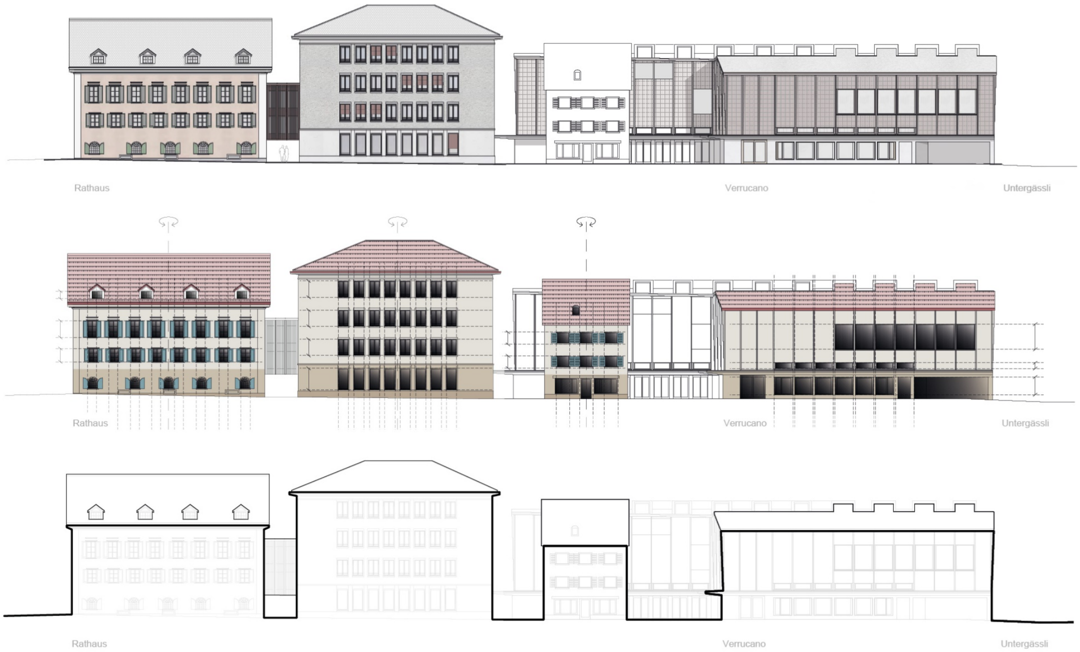


Figure 1: A development view as part of a facade memorandum – top row: true coloured facades, middle row: developed view with reference-lines and axes of symmetry conveying rhythms of facades, bottom row: outlines.

2 Deep learning methods

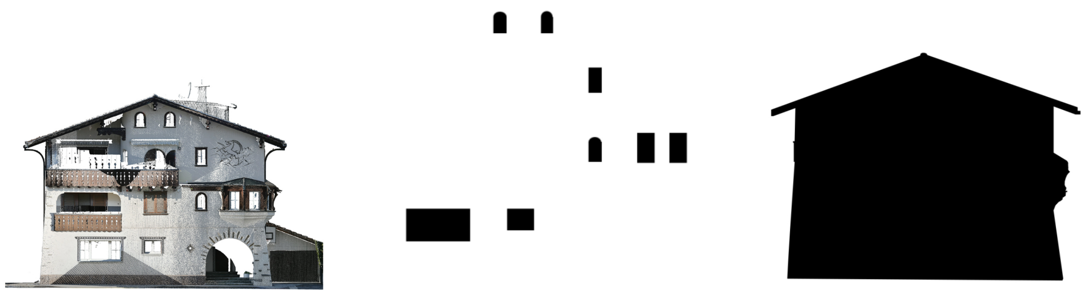
In this section, we briefly overview the process it takes to get from point cloud images to elements of a facade. Discovering patterns and geometrical relationships within image data is a distinct strength of deep learning (DL) methods [1]. Knowledge is transferred via domain experts by annotating multiple masks on point cloud data collected specifically for this use case, thus creating a customized dataset on which the deep learning algorithms can learn to find the specific elements in facades. These elements – in combination with a domain-specific set of rules – are used to establish rhythm and symmetry as seen in Figure 1. Buildings all over the world follow similar structures but facades and the elements within them may look very different. There is a variety of windows, blinds, balconies, and so on. The design of these elements is affected by things such as the time period, the country, or even towns and settlements. This resulted in the creation of a facade dataset for Swiss buildings.

The semantic segmentation of the facade is the most critical step in the process of finding characteristics of buildings. Every error in the segmentation result needs to be handled (manually) by later stages of the process. For this reason, the development and training of the deep learning model is essential for finding relevant characteristics. Semantic segmentation of images is a distinct strength of deep learning models [2]. There are multiple advantages w. r. t. classical image processing. For example, DL-based approaches are able to recognise different types of objects that occur in a facade. This is especially useful for objects that tend to look slightly different each time. In architecture, objects such as windows, doors, or balconies never look exactly the same, hence the detection and segmentation of those objects becomes more difficult.

Today, there exist many different approaches for window detection in facades, such as convolutional neural networks (CNN) [3], Mask- [4] and Faster region-based convolutional neural networks (R-CNN) [5, 6], convolutional generative adversarial networks [7], hybrid approaches [8] , cascaded classifiers [9–11], YOLO [6], single shot detectors (SSD) [7], atrous large kernel (ALK) networks [12] or models with prior information [13]. Many of those show competitive performances on the task of window detection. Nevertheless, for our use case it was decided to concentrate on Mask R-CNN and YOLO.

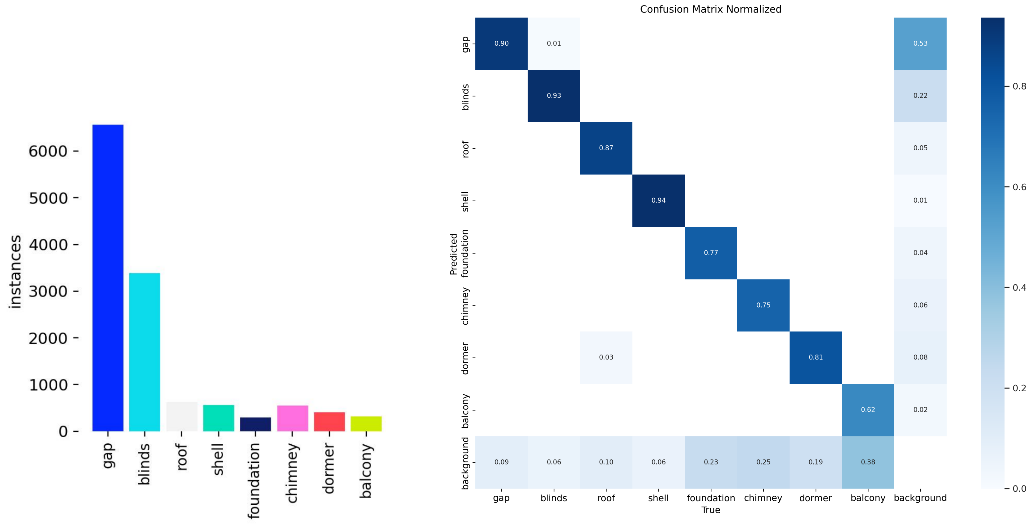
2.1 Dataset

The dataset was created by domain experts. Point clouds were captured by drones and drivable camera systems and exported as images. For each image, a set of up to 14 masks was created. Out of the 14 masks, 8 were used for the training of the models. Those were: gaps in the facade (windows and doors), window blinds, roofs, building shell, building foundation, chimneys, dormers, and balconies. Figure 2 shows an original point cloud image with the respective segmentation masks for windows and the building shell. The masks were translated into labels and then used for training, validation, and testing of the different DL models. In total, there were 769 images captured, these were split into a training set (577 images), a validation set (128 images), and a test set (64 images). Obviously, the 14 segmentation masks are not equally distributed over the 769 images (e. g. there are much more windows than dormers). Figure 3 (a) shows the class imbalance in the test set representatively for the whole data set.



(a) Original image (b) Window mask (c) Building shell mask

Figure 2: Original point cloud image with respective segmentation masks.



(a) Label distribution of the test set.

(b) Normalized confusion matrix of the test set.

Figure 3: Characteristics of the test set.

2.2 Mask R-CNN

Mask R-CNN is a two-stage deep learning model that can be used for semantic segmentation. It extends Faster R-CNN and is today state-of-the-art for a variety of tasks including object detection and instance segmentation [14]. Mask R-CNN was part of a series of experiments conducted with the new facade dataset. Even though the model initially showed some promising results, the YOLO model finally convinced due to its ease of use when in the future integrated into a product, coupled with its faster training and inference times [15].

2.3 YOLO

YOLOv8 and YOLOv11 [16] are one-stage models for different computer vision tasks such as object detection, oriented object detection, pose/keypoints detection, and instance or semantic segmentation. They are based on previous YOLO architectures like YOLOv3 [17]. Both models were tested in a series of experiments. The experiments included changes in model size, model parameters, and model hyper-parameters. Table 1 shows the metrics for each of the best model configurations. This includes hyper-parameter tuning for multiple models. Compared to R-CNN models, YOLO not only looks at regions of interest but at the whole image which helps speeding up the process of object detection [18].

Model	F1 (box)	F1 (mask)	mAP50 (box)	mAP50-95 (mask)
yolov8l-seg	0.833	0.817	0.844	0.697
yolo11l-seg	0.828	0.801	0.840	0.687

Table 1: Performance metrics for YOLO models.

Eventually, a pre-trained large YOLOv8 model was fine-tuned with the facade dataset to achieve accurate and efficient facade segmentation. Table 2 shows the complete performance metrics on the test set for the chosen model. The small number of balconies and the corresponding low value for the Box mAP50 is noteworthy. So, too, are the very good values for the metrics on the building shell, despite the low number of instances.

1	2	3	4	5	6	7	8	9	10	11
all	64	1364	0.896	0.778	0.844	0.697	0.877	0.764	0.820	0.458
gap	64	696	0.893	0.848	0.893	0.748	0.887	0.849	0.885	0.467
blinds	37	358	0.924	0.889	0.921	0.782	0.903	0.872	0.897	0.336
roof	64	69	0.945	0.841	0.919	0.823	0.941	0.841	0.893	0.535
shell	63	63	0.999	0.937	0.963	0.883	0.964	0.905	0.934	0.596
foundation	35	35	0.836	0.743	0.805	0.690	0.803	0.714	0.777	0.486
chimney	43	73	0.917	0.685	0.788	0.535	0.892	0.671	0.731	0.357
dormer	29	36	0.745	0.694	0.745	0.639	0.713	0.667	0.706	0.504
balcony	17	34	0.908	0.588	0.716	0.474	0.910	0.596	0.737	0.388

Table 2: Performance metrics for the best YOLOv8 model – 1: object class, 2: # images, 3: # instances, 4: precision bounding box (BB), 5: recall BB, 6: mean average precision at IoU=0.50 BB, 7: mean average precision IoU=0.50 to IoU=0.95 BB, 8: precision mask, 9: recall mask, 10: mean average precision at IoU=0.50 mask, 11: mean average precision IoU=0.50 to IoU=0.95 mask

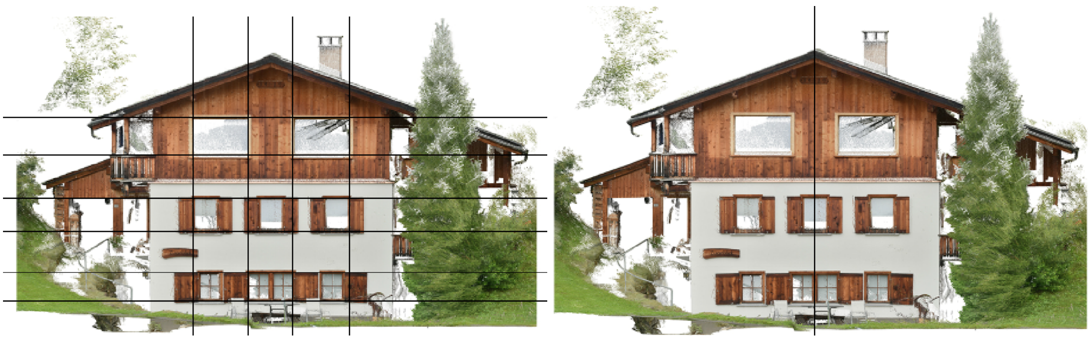
The confusion matrix of a multi class object segmentation result is a useful tool for cross class comparisons and to find mix-ups between classes. Figure 3 (b) shows the confusion matrix for the test result. Figure 4 shows an original image and the respective image with the segmentation result. The segmentation masks show the classes of building shell, roof, chimney, balcony, gap, and blinds. The model can, after the training, be deployed to a server where it can be used to segment facades. Inference for the semantic segmentation can be done in reasonable time on CPU and does not require an expensive GPU.

(a) Original image (b) Segmentation result

Figure 4: Original point cloud image with respective segmentation result.

3 Architectural principles

The results from the facade segmentation are fed into a classical computer vision algorithm. This algorithm is based on rules derived from two domain experts. These rules reflect architectural principles and based on type, position, and shape of objects within a facade some reference-lines and axes of symmetry are constructed. Reference-lines highlight rhythmic patterns in a facade and together with an axis of symmetry (if present) they form the facade's aesthetic coherence. Figure 5 shows a facade with the reference-lines and the axis of symmetry which were constructed on the basis of the segmentation result in Figure 4.



(a) Facade with reference-lines (b) Facade with axis of symmetry

Figure 5: Facade with reference-lines and axis of symmetry.

3.1 Rules

Architecture and design can be difficult to formalize. The rules to reflect the architectural principles are based on experience and domain knowledge. One of the main contributions to find formalized domain-specific rules to be applied to automatically segment facade images was the assessment and review of separate facades by an expert.

The computer vision algorithm for finding rhythmic and symmetry patterns follows a multi-stage process. Reference-lines that build the rhythmic patterns are based on the positions and shapes of the objects in the facade. Windows with open, half closed, or completely closed blinds, with blinds that close from the sides or above, jalousies, or sun blinds all contribute equally to the rhythmic pattern of a facade.

The algorithm to find a potential axis of symmetry in a facade consists of multiple and recursive checks on axis candidates while simultaneously moving these candidates based on the results of the different checks. The axis candidates depend on the roof, the building shell, the position and shape of windows, doors and blinds as well as the positions of these elements mutually to each other. If the axis candidate fails to fulfil a criteria (e. g. the distance of a window is not the same as the one to the corresponding window on the same floor on the other side of the candidate) the candidate is moved by a margin depending on the distance with which the test was failed.

To accommodate the domain experts working with this tool, an editor was developed. This human in the loop approach of the process to find a facades' aesthetic coherence allows to rectify any potential error of the deep learning method. Finally, the software for finding facade objects as well as rhythmic and symmetrical patterns was integrated (together with the editor) into a GIS application for future usage by domain experts.

4 Conclusions

A new facade dataset consisting of facade images from point clouds collected by drones and drivable camera system was constructed. This dataset includes segmentation masks for up to 14 classes of elements which can typically be found in facades. The dataset was used in the training of different deep learning models for semantic segmentation. Furthermore, a set of rules was derived for finding architectural patterns in facades. Based on the segmentation result and the set of rules, reference-lines and axes of symmetry were constructed to establish the rhythm and help to convey architectural principles of a facade. As an outlook, our new DL model can be used to (semi-)automatically create developed views, add new information such as building textures to GIS, and show planned buildings in the context of the specific neighbourhood to help communicate specific characteristics of the established and the planned buildings. Figure 6 shows an example of such a use case.

Nevertheless, there are some limitations to this technology. For instance, the detection of structures important for spatial connections and design relationships is highly dependent on the quality of data. Moreover, is it very difficult to find aesthetically pleasing architectural patterns if they have not been trained before. Therefore, artificial intelligence will never decide on its own what is beautiful. However, artificial intelligence will be able to recognize the extent to which a new building is able to take up the typical characteristics of a specific village or settlement. In the end, decisions will still be made by the building authorities and the population.

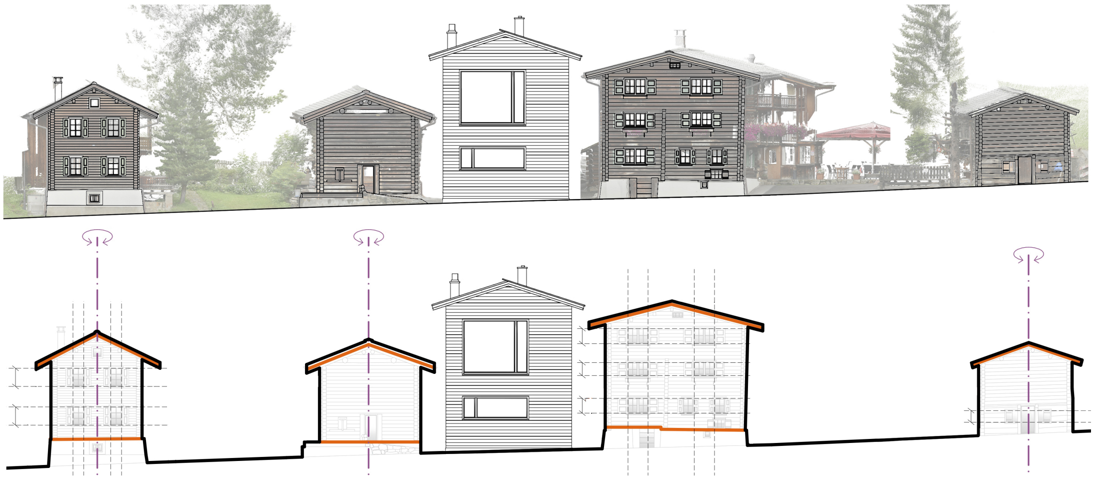


Figure 6: A developed view with a CAD model of a planned building in the context of the neighbourhood.

Acknowledgements

We specifically would like to thank the Swiss Innovation Agency Innosuisse for providing funding under grant 103.722 IP-SBM.

References

[1] K. He, X. Zhang, S. Ren, & J. Sun, "Deep residual learning for image recognition". In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778)., 2016.

[2] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, & D. Terzopoulos, "Image segmentation using deep learning: A survey". IEEE transactions on pattern analysis and machine intelligence, 44(7), 3523-3542., 2021.

[3] G. Kong and H. Fan, "Enhanced Facade Parsing for Street-Level Images Using Convolutional Neural Networks." in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 12, pp. 10519-10531, 2021.

[4] N. Nordmark, & M. Ayenew, "Window detection in facade imagery: a deep learning approach using mask R-CNN". arXiv preprint arXiv:2107.10006., 2021.

[5] W. Ma and W. Ma, "Deep Window Detection in Street Scenes," KSII Transactions on Internet and Information Systems, vol. 14, no. 2, pp. 855-870, DOI: 10.3837/tiis.2020.02.022., 2020.

[6] G. Sezen, M. Cakir, M. E. Atik, & Z. Duran, "DEEP LEARNING-BASED DOOR AND WINDOW DETECTION FROM BUILDING FAÇADE", Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B4-2022, 315–320, 2022.

[7] X. Guo, Q. Liu, Y. Ma; "Facades and Openings Detection Based on Different Deep Learning Models", 2021.

[8] K. Rahmani, H. Mayer, "A Hybrid Method for Window Detection on High Resolution Facade Images". In: Bennour, A., Ensari, T., Kessentini, Y., Eom, S. (eds) Intelligent Systems and Pattern Recognition. ISPR 2022. Communications in Computer and Information Science, vol 1589. Springer, 2022.

[9] M. Neuhausen, M. König; "Automatic window detection in facade images", In: Automation in Construction, Volume 96, 2018.

[10] M. Neuhausen, M. König, "Improved Window Detection in Facade Images". In: Mutis, I., Hartmann, T. (eds) Advances in Informatics and Computing in Civil and Construction Engineering. Springer, 2019.

[11] M. Neuhausen, M. Obel, A. Martin, et al., "Window detection in facade images for risk assessment in tunneling". Vis. in Eng. 6, 1 , 2018.

[12] W. Ma, W. Ma, S. Xu and H. Zha, "Pyramid ALKNet for Semantic Parsing of Building Facade Image", in IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 6, pp. 1009-1013, 2021.

[13] G. Zhang, Y. Pan, L. Zhang, "Deep learning for detecting building façade elements from images considering prior knowledge", Automation in Construction, Volume 133, 104016, ISSN 0926-5805, 2022.

[14] K. He, G. Gkioxari, P. Dollár, & R. Girshick, "Mask r-cnn". In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969)., 2017.

[15] R. Sapkota, D. Ahmed, M. Karkee; "Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments", Artificial Intelligence in Agriculture, Volume 13, Pages 84-99, ISSN 2589-7217, 2024.

[16] G. Jocher, J. Qiu, & A. Chaurasia, Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics, 2023.

[17] J. Redmon, & A. Farhadi, "Yolov3: An incremental improvement". arXiv preprint arXiv:1804.02767., 2018.

[18] Y. Wang, Un. Ahsan, H., Li & M. Hagen, "A Comprehensive Review of Modern Object Segmentation Approaches", Foundations and Trends in Computer Graphics and Vision: Vol. 13, No. 2-3, pp 111–283, 2023.