Advanced Intelligent Discovery

WILEY VCH

ADVANCED INTELLIGENT DISCOVERY Open Access

RESEARCH ARTICLE OPEN ACCESS

Computer Vision Pipeline for Image Analysis for Freeze-Fracture Electron Microscopy: Rosette Cellulose Synthase Complexes Case

Siri Mudunuri[1] | Leala Carbonneau[2] | Eric M. Roberts[3] | Alison W. Roberts[4] | Candace H. Haigler[5] | Alexey V. Gulyuk[6] | Yaroslava G. Yingling[6]

[1]Department of Computer Science, North Carolina State University, Raleigh, North Carolina, USA | [2]Department of Computer Science, Worcester Polytechnic Institute, Worcester, Massachusetts, USA | [3]Department of Biology, Rhode Island College, Providence, Rhode Island, USA | [4]Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island, USA | [5]Department of Plant and Microbial Biology and Department of Crop and Soil Sciences, North Carolina State University, Raleigh, North Carolina, USA | [6]Department of Materials Science and Engineering, North Carolina State University, Raleigh, North Carolina, USA

Correspondence: Alexey V. Gulyuk (avgulyuk@ncsu.edu) | Yaroslava G. Yingling (yara_yingling@ncsu.edu)

Received: 11 June 2025 | **Revised:** 7 October 2025 | **Accepted:** 8 October 2025

Funding: National Science Foundation; North Carolina State University, Grant/Award Number: DMR-2150360; Phosphorus Sustainability Center; National Science Foundation Science and Technology Center, Grant/Award Number: CBET 2019435; Center for LignoCellulose Structure and Formation; U.S. Department of Energy; Office of Science; Basic Energy Sciences, Grant/Award Number: DE-SC0001090; National Science Foundation to AWR, Grant/Award Number: IOS-2124176; Research Capacity Fund, Grant/Award Number: 7005416

Keywords: cellulose synthase | computer vision | deep learning | FF-TEM

ABSTRACT

In materials science, plant biology, agriculture, and environmental research, the automated analysis of high-magnification, complex microscopy images, such as those generated by freeze-fracture electron microscopy (FF-TEM), remains a critical challenge that limits the scalability of data interpretation. We present a deep learning computer vision pipeline for high-throughput detection and morphological characterization analysis of cellulose synthase complexes (CSCs, or rosettes) in FF-TEM images. The pipeline integrates preprocessing, detection, human-in-the-loop verification, and semantic segmentation to quantify features such as rosette diameter and inter-lobe spacing. The approach was trained and tested on a curated dataset of high-resolution FF-TEM micrographs of *Physcomitrium patens*, expanded via strategic tiling and augmentation to over 650 images. We compare YOLOv8 and YOLOv9 architectures and demonstrate that YOLOv9 achieves superior performance in both localization accuracy (mAP50-95 = 0.854) and inference speed. The resulting distributions revealed biological variability consistent with prior manual studies, validating the approach for high-throughput applications. Our results show that the pipeline achieves human-expert level accuracy while dramatically reducing analysis time, enabling scalable, reproducible structural characterization of intramembrane protein complexes. The pipeline is broadly applicable to other domains requiring precise interpretation of complex microscopy data and establishes a foundation for future artificial intelligence (AI)-assisted workflows in biological imaging.

1 | Introduction

Digital imaging now plays a crucial role in environmental monitoring and exploration, with research tools like microscopes, drones, satellites, smartphones, robots, and environmental sensor networks producing terabytes of images in a single experiment. Yet, much of this vast data is still analyzed manually or through basic rule-based scripts [1]. For example, in marine science

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Advanced Intelligent Discovery* published by Wiley-VCH GmbH.

monitoring, a single benthic dive can yield tens of thousands of frames that require weeks of expert labeling [2]. Other significant challenges involve dealing with noisy or low-contrast images, complex and heterogeneous backgrounds, overlapping features, and inconsistent illumination conditions, all of which undermine the reliability, reproducibility, and efficiency of manual or rule-based analyses, ultimately introducing bias and variability that can compromise both scientific accuracy and operational decisions. For example, for on-site detection of phosphorus and other environmental contaminants, a portable smartphone-based colorimetric system can quantify phosphate, silicate, and nitrate, but their colorimetric readings drift with turbidity, background hue, and camera optics, underscoring the need for calibration-aware computer-vision pipelines [3]. In materials science, annotating and analyzing atomic force microscopy (AFM) images is particularly challenging due to complex surface morphologies, instrument-induced noise, and artifacts such as thermal drift, all of which can obscure nanoscale features and make accurate, reproducible analysis difficult [4]. In plant sciences, challenges such as segmentation of overlapping plant structures, limited resolution in 2D microscopy hindering 3D morphological analysis, and labor-intensive manual annotation of terabyte-scale image datasets persist, despite advances in tools like whole-slide imaging and fluorescence systems [5, 6] Overall, addressing heterogenous backgrounds, noisy and low-contrast data, scarce labeled data, and rapid domain shifts across these exemplars remains a critical frontier for computer-vision solutions.

Computer vision (CV) workflows, which involve automated extraction, interpretation, and analysis of visual information from images and videos, offer a powerful foundation for scalable, objective, and reproducible analysis. Across different fields such as plant science, phosphorus sustainability, and materials characterization, CV provides a path toward scalable, objective, and reproducible image analysis that can keep pace with today's data-intensive research. Despite recent advances [7], further development of CV pipelines is essential. Current systems often struggle with diverse and noisy real-world data, lack adaptability to new environments, and require large, well-labeled datasets that are difficult to curate at scale. Additionally, there is a growing need for pipelines that are robust, scalable, and easy to maintain. In this work, we develop a CV pipeline for the automated detection of cellulose synthase complexes (CSC) from freeze-fracture transmission electron microscopy (FF-TEM) images, leveraging deep-learning techniques to enable high-throughput, accurate, and reproducible analysis of microscopy images.

Cellulose microfibrils are the primary structural component of plant cell walls, providing mechanical strength, directional growth, and protection against environmental stress and pathogens [8]. Each microfibril is synthesized by a CSC, a six-lobed 'rosette' structure embedded in the plasma membrane, each lobe containing three cellulose synthase (CESA) proteins [8]. Each CESA catalyzes the polymerization of β–1,4-glucan chains, which are exported and coalesce into semi-crystalline microfibrils (Figure 1) [8, 11]. Fluorescently tagged CESAs allow for live-cell imaging of CSC movement, though the full complex appears only as bright spots [11]. While partial CESA structures have been predicted and resolved [9, 12, 13], details of the complete CSC, such as the structure and conformational flexibility of intrinsically disordered domains that may affect its size during active cellulose synthesis, often remain unclear.

CESA proteins are intramembrane proteins with a cytosolic region where glucan chain polymerization occurs and a transmembrane domain that forms a channel for glucan chain export [12]. The transmembrane domain can be visualized using FF-TEM [9, 11, 14]. By fracturing lipid bilayers in rapidly frozen

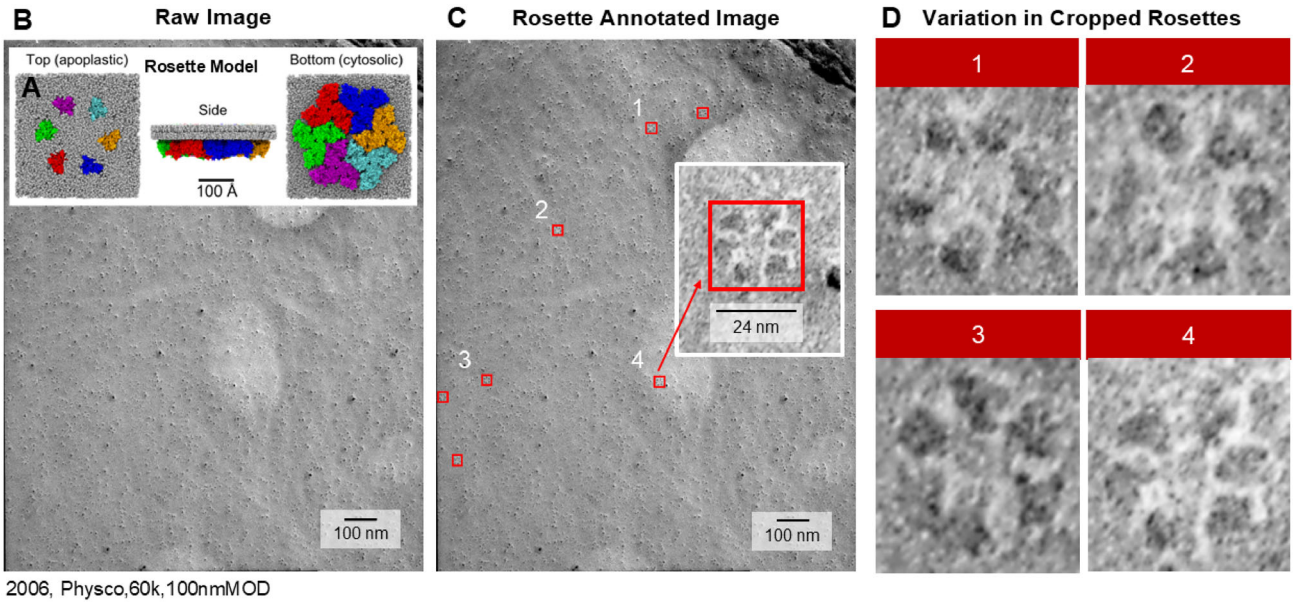


FIGURE 1 | Representative raw TEM micrographs of the plasma membrane with embedded rosette CSCs. (A) (Inset). Top (apoplastic), side, and bottom (cytosolic) views of an all-atom PpCESA5 CSC molecular model, derived using reference sources [9, 10]. (B) A FF-TEM micrograph of a plasma membrane leaflet with transmembrane protein particles, including rosette CSCs. (C) Rosette CSCs are annotated with boxes, and the inset shows an enlargement. (D) Images of potential rosette CSCs illustrating variations in appearance. CSCs = cellulose synthase complexes; TEM = transmission electron microscopy; FF-TEM = freeze-fracture transmission electron microscopy.

cells, this technique exposes the transmembrane portions of proteins like CESAs while minimizing artifacts caused by chemical fixation or dehydration [15]. Platinum/carbon shadowing of the exposed surfaces provides contrast, allowing intramembrane proteins to be visualized as particles on the smooth membrane surface. The TEM micrographs reveal the six-lobed 'rosette' architecture associated with cellulose microfibril formation. An example FF-TEM micrograph is shown at low magnification in Figure 1B,C, with individual rosettes shown at higher magnification in Figure 1D. Manual analysis of rosettes viewed using FF-TEM has recently been used to: provide evidence, in conjunction with computational modeling [11], for a CSC composed of 18 CESAs; select feasible models for CSC structure; and support the ability of other proteins beyond classical CESAs to synthesize cellulose microfibrils [9–11].

Previously, researchers had to invest substantial time and effort systematically zooming in to view small areas while panning across the entire micrograph in a systematic way to identify and annotate potential rosette CSCs. Although experts could work together to decide which putative rosette instances were authentic, the possibility of human bias or error remained (see Roberts et al. 2024 [10]). Figure 1D reflects some of the morphological variation in putative rosette CSCs, with instance 1 being the most uncertain. These variations are inevitable due to the potential of other intramembrane proteins to form clusters, subtle variations in actual shadowing angle and coating thickness due to changing topography of the plasma membrane, and the particulate nature of the platinum/carbon coat with ≥1.25 nm grain size. After an accepted set of instances was available, quantitative analysis of rosette properties, such as diameter, required time-consuming hand measurements, usually resulting in only selected dimensions being measured to save time. These limitations pose challenges for large-scale analysis and introduce variability in data interpretation. Generalizing structural conclusions about CSCs requires statistically robust datasets that capture natural variation in rosette size and substructure. Here, we tackle these issues by developing a deep learning framework that can automate the detection and quantification of CSC rosette features with high precision and reproducibility. This study leverages the You Only Look Once (YOLO) object detection model to automate the identification and characterization of the rosette objects [16, 17, 18]. We describe a comprehensive data pipeline capable of detecting rosette structures within FF-TEM micrographs and performing preliminary quantitative analyses. Our approach to analysis includes metrics on the transmembrane region of the CSC, including inter-lobe distances and rosette diameter, that can be modified for the particular needs of domain experts. This efficient, objective, and quantitative method relying on artificial intelligence tools can aid future characterization of rosette CSCs in diverse natural and engineered cell types as part of research to tailor cellulosic biomaterials for diverse end uses.

2 | Methods

2.1 | Dataset Structure

The dataset utilized for model development and testing was also the basis for prior results using hand measurements [11]. It consisted of 32 FF-TEM micrographs of rosette structures collected from the tips of five wild-type protonemal cells of the moss, *Physcomitrium patens*. These rosettes are now known to include a mixture of rosette CSCs composed of either CESA or cellulose-synthase-like D proteins (CSLDs), but the rosette CSC structure is conserved between the two protein types [10]. The micrographs were captured at 60,000X magnification using a JEOL-1200EXI TEM (Tokyo, Japan) operating at 80 kV, with calibration performed on each imaging day using negatively stained bovine liver catalase crystals (with a known lattice spacing of 172 Å ± 4) to ensure high-accuracy size measurements. Calibration involved adjusting the specimen to its eucentric height and acquiring reference micrographs of the catalase crystals under the same imaging conditions and using the same specimen holder. The scale bar for each micrograph of rosette CSCs was measured 10 times, and the average was converted to nanometers based on the catalase reference. The maximum scale variation between imaging days was 1.2%, introducing a negligible error (0.3 nm for a 25 nm rosette). The photographic negatives were scanned at 1200 dpi. Each micrograph was 3550 pixels by 4270 pixels. All batches (datasets assembled for processing and analysis) were combined for the dataset generation to allow for a larger dataset for model development. However, for testing and validation of the final pipeline, the batches were separately assessed to ensure proper nanometer measurements.

Manual identification of rosette structures across each micrograph of this dataset was conducted to generate data for model training and evaluation using the MATLAB Image Labeling Tool. A structured process was defined for data generation. Rosette identification was performed manually by inspecting the micrograph at 350% magnification. The micrograph was systematically traversed using a raster pattern, beginning at the top-left corner and moving horizontally across to the right edge. Upon reaching the end of a row, the view was vertically shifted downward by a fixed increment to the next row, and the process was repeated until the entire micrograph area was covered. Rosette structures were annotated using rectangular bounding boxes to ensure consistent and reproducible localization of features. The rectangular box format was selected due to its compatibility with common object detection frameworks and its efficiency in enclosing irregular but localized structures such as rosettes. In addition, the rectangular box provided the most efficient bounding box structure for model training and processing for the task of detecting rosette objects. To reduce human bias during manual dataset generation, all 32 micrographs were independently annotated by two separate researchers. The results were compared, and only the rosettes identified by both annotators were retained in the final dataset. This process ensured that only consistently recognized rosettes were included during model development and evaluation.

Figure 2A demonstrates the distribution of the number of rosettes identified across the final annotated dataset. A total of 403 rosettes were identified and annotated. The average number of rosette objects per micrograph was 11, with some micrographs having more than others. Furthermore, Figure 2B illustrates the locations of all manually annotated rosette structures within the dataset, overlaid on a theoretical composite 3550 × 4270-pixel image. Each outlined box corresponds to a single rosette object, with a total of 403 annotations shown. The spatial map provides a visual overview of rosette density and distribution, highlighting

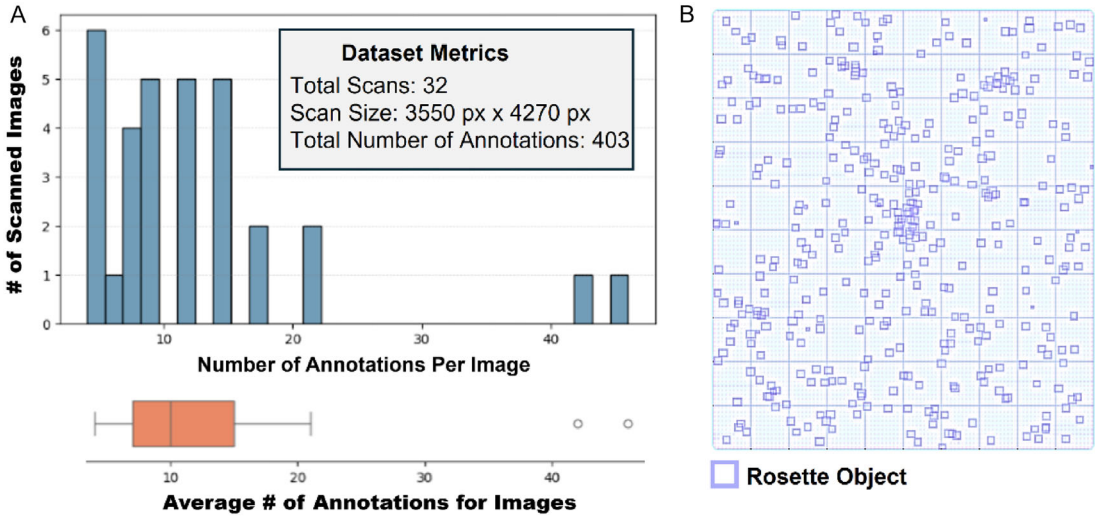


FIGURE 2 | (A) Distribution of a number of identified rosettes across all 32 micrographs used for model development. The whisker plot beneath represents the metrics of the average number of annotations made for the analyzed dataset. (B) Composite image summarizing the positions of 403 annotated rosette CSCs across the entire dataset, illustrating the spatial patterns observed throughout the dataset. CSCs = cellulose synthase complexes.

localized clustering in certain regions of the micrograph while other areas remain sparsely populated. For instance, the center of the micrographs had the most rosette density, which probably arose from the researcher commonly centering the micrograph on a large population of the structures of interest. However, the overall rosette density was relatively uniform across the spatial map, which supported continuing with the analysis.

To enable effective training of object detection models, the initial dataset, which consisted of 32 annotated micrographs, required significant expansion. Deep learning models, particularly those used for object detection, rely on large, diverse datasets to achieve robust performance and generalizability. A limited number of training samples increases the risk of overfitting and restricts the model's ability to accurately detect target structures under varying conditions. Therefore, expanding the dataset was essential to ensure the model could learn meaningful features and perform reliably on unseen data. Given the limited number of original micrographs, both image preprocessing and data augmentation strategies (Figure 3) were employed to increase the number and variability of training samples while preserving the integrity of annotated structures. Each micrograph was first processed using a 10×10 tiling strategy, breaking the full-resolution micrographs into smaller, non-overlapping tiles (Figure 3A). This preprocessing step preserved the structural fidelity of the rosette objects while increasing the dataset to 320 discrete image tiles. Furthermore, the preprocessing step ensured that the input data fed into the model during training was smaller in size, which not only improved computational efficiency but also enabled the use of larger batch sizes and faster training iterations. Reducing the input dimensions also helped lower GPU memory requirements, making the training process more scalable and accessible on standard hardware. These smaller, standardized image inputs are well-suited for deployment in real-time detection pipelines, where speed and memory efficiency are critical. In addition to tiling, the dataset was also filtered to ensure that at least 40% of images contained annotations. This step reduced the dataset size to 260 images and prevented class imbalance through training on a sufficient number of positive examples.

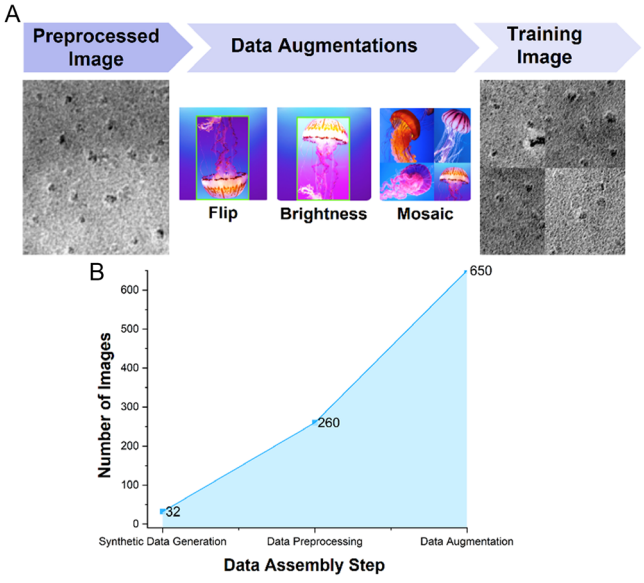


FIGURE 3 | Dataset generation workflow and corresponding increase in dataset size. (A) Schematic representation of the dataset generation process, beginning with the preprocessed or tiled FF-TEM images. These images were augmented using techniques such as flipping, rotation, brightness adjustment, exposure modification, and mosaic composition (adapted from Roboflow [19]) to produce the final training images. (B) Graph illustrating the growth in dataset size across each stage, starting from synthetic data generation, followed by preprocessing, and culminating in data augmentation. FE-TEM = freeze-fracture transmission electron microscopy.

Following preprocessing, the dataset was divided into a training (80%), validation (10%), and testing (10%) splits for detection model development. This split provided a larger training dataset from the limited number of images available after dataset expansion through preprocessing (described below) to better identify patterns in rosette structure, along with deviations, and create a more robust and generalizable model. Furthermore, a comprehensive augmentation pipeline was used to enhance image

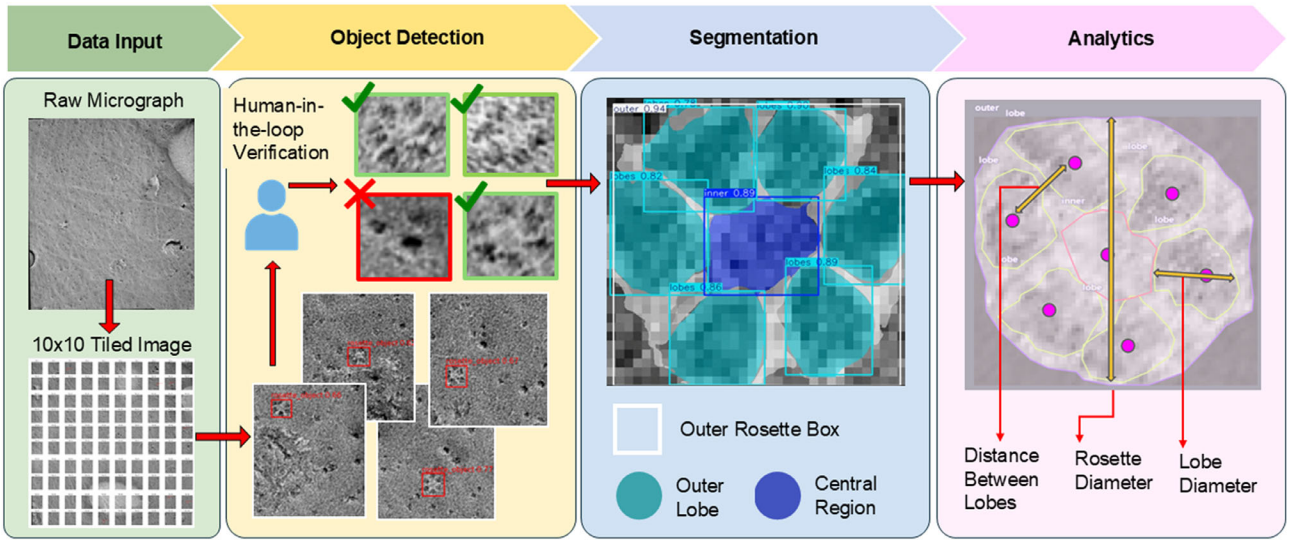


FIGURE 4 | Pipeline for automation of rosette CSC identification and analysis. The pipeline follows 4 main steps: data input, object detection, segmentation, and analytics. The pipeline begins with a raw FF-TEM micrograph showing a top-down view of the plasma membrane and uses the following steps: data-preprocessing through tiling, object detection with a trained model, human-in-the-loop verification of detected objects, segmentation of verified objects, and analytics. Further description and function of each step are described below. CSC = cellulose synthase complex; FE-TEM = freeze-fracture transmission electron microscopy.

variability and improve model generalization. Three augmentation techniques were utilized, including flip, brightness, and mosaic (Figure 4A). Brightness adjustments were used to mimic changes in imaging conditions. As illustrated in Figure 1D, the brightness of original micrographs usually varies, making it essential that rosettes can be detected in different conditions. Horizontal and vertical flipping were utilized to introduce symmetry-based transformations. Algorithmic mosaic augmentation introduced global variability in object positioning, spatial relationships and surrounding conditions. This enhanced the model's robustness to background clutter and improved generalization to unseen test data. Additionally, mosaic augmentation increased the effective object density per image, allowing the model to encounter more annotated objects within a single forward pass. This not only accelerates the learning process but also improves detection performance on cluttered or densely populated regions of interest. Given the relatively small dataset, mosaic augmentation substantially amplified training variability and improved model performance.

These augmentation techniques proved particularly useful [20, 21] in simulating biologically plausible contexts and enhancing the detection model's ability to identify rosettes across a wide range of spatial and visual configurations, while maintaining the visual characteristics critical to accurate rosette identification [22]. As shown in Figure 3B, this multistage data assembly process expanded the dataset from 32 original micrographs to a total of 650 images, ensuring sufficient diversity and volume for effective model training and evaluation for the detection model.

2.2 | Methods

An expanded dataset was used to develop a high-performance object detection model aimed at identifying rosette CSCs in FF-TEM micrographs. For this purpose, we employed two advanced architectures from the YOLO (You Only Look Once) family: YOLOv8 and YOLOv9 [23]. These models were selected based on their strong performance in detecting small, morphologically variable objects [18, 24, 25], similar to CSC rosettes visualized in FF-TEM micrographs. Both YOLOv8 and YOLOv9 represent significant advancements in the YOLO family of models, offering real-time detection capabilities with a favorable trade-off between speed and accuracy.

YOLOv8, developed by Ultralytics, integrates key improvements over prior versions, such as an anchor-free detection mechanism, enhanced convolutional modules (C2f), and more efficient backbone networks based on cross stage partial (CSP) connections [25]. These features contribute to better localization accuracy and enhanced generalization on small-scale and low-contrast objects.

To explore potential performance gains and assess the latest advancements in object detection, we also experimented with YOLOv9 that introduces Generalized Efficient Layer Aggregation Networks (GELAN) to enhance feature reuse and spatial awareness across layers—key for detecting densely packed, low-contrast, and morphologically diverse structures. Moreover, YOLOv9 supports hybrid task learning and drop-path training, both of which contribute to robustness in environments with subtle object variance and imaging noise [26]. These features are particularly valuable given the variability in brightness and lobe configuration observed across our annotated dataset.

3 | Results and Discussion

Figure 4 presents an overview of the proposed CV pipeline, which consists of four key stages: Data Input, Object Detection, Segmentation, and Analytics. The pipeline begins

with high-resolution FF-TEM micrographs of plant cell membranes, which are first preprocessed to reduce image dimensions, improving computational efficiency. The resized images are then analyzed by a YOLO-based object detection model trained to localize rosette structures. Following initial detection, a human-in-the-loop verification step ensures the accuracy of identified objects. The reviewer during this stage verifies or rejects model predictions based on the established practices for manual identification. Specifically, candidate detections are accepted as valid rosettes if the detected object exhibits a six-lobed or near-hexagonal arrangement consistent with CSC morphology and if the rosette can be clearly distinguished from surrounding background particles and irregularities. By integrating this step, the pipeline replicates how experts traditionally validate rosette morphology, while also reducing the likelihood of errors or inconsistencies that can arise in a fully automated system. Once verified, the rosettes are then processed by a segmentation model trained to delineate distinct morphological regions, including the outer rosette boundary, individual lobes, and the central region, as illustrated in Figure 4. These segmented components serve as the foundation for quantitative analytics, enabling precise morphological characterization.

3.1 | Detection Model Training

Both YOLOv8 and YOLOv9 models were trained using a standardized pipeline and evaluated across 25 epochs using mAP at an intersection over union (IoU) threshold of 50%. This metric was chosen because it is widely accepted for object detection tasks, offering a balanced assessment of localization accuracy while tolerating minor misalignments. Training was performed on an expanded dataset across 25 epochs with a batch size of 8. YOLOv8 was trained using an image size of 1280 pixels, while YOLOv9 used 640 pixels due to its architecture being less tolerant of higher input resolutions and exhibiting increased memory demands. Despite the higher computational cost, relatively large image sizes were selected for both models to ensure that the high-resolution features of the rosette structures could be effectively captured during detection.

YOLOv9 demonstrated superior early training performance compared to YOLOv8 (Figure 5), achieving significant learning by epoch 8 when YOLOv8 was still showing minimal progress. This early advantage highlights YOLOv9's more efficient initialization and training dynamics. While YOLOv8 eventually caught up around epoch 10, YOLOv9 maintained consistently strong performance throughout the training process, exhibiting remarkable stability with fewer performance fluctuations. The graph clearly shows YOLOv9 reaching mAP scores above 0.8 as early as epoch 7, while YOLOv8 remained below 0.4 at this stage. By the final epochs, YOLOv9 achieved comparable performance to YOLOv8, demonstrating its ability to reach optimal detection accuracy despite using a smaller input resolution and working faster.

This performance advantage can be attributed to several key architectural innovations in YOLOv9. The GELAN architecture enables more effective feature extraction and integration across multiple scales, allowing YOLOv9 to identify objects more efficiently with less training data. YOLOv9's Programmable Gradient Information (PGI) and Reverse Knowledge Distillation (RKD) techniques

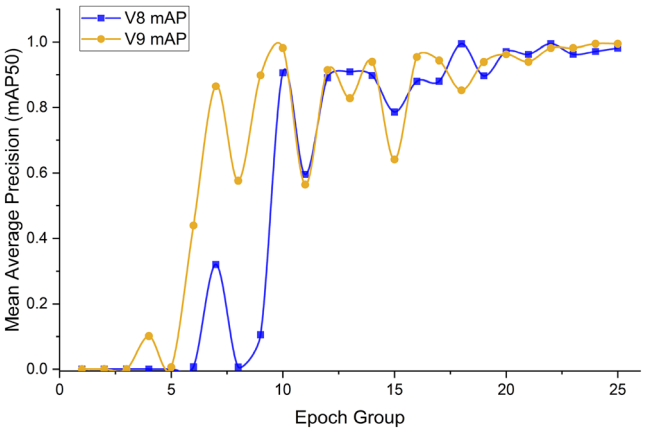


FIGURE 5 | Training performance comparison of YOLOv8 and YOLOv9 detection models through 25 epochs. Mean average precision (mAP50) was used as the measure of comparison, demonstrating similar overall training performance.

contribute to its superior early learning capabilities and stability. Additionally, the implementation of adaptive drop-path regularization helps prevent overfitting while maintaining robust generalization abilities.

Although YOLOv8 eventually achieved similar performance metrics in later epochs, several recent studies highlight YOLOv9's practical advantages. For instance, YOLOv9 consistently outperforms YOLOv8 in terms of inference times, achieving superior accuracy when normalized for computational requirements [23]. The smaller input resolution requirements (640 px vs 1280 px) make YOLOv9 particularly suitable for deployment on less powerful devices without sacrificing detection quality. This efficiency translates to faster inference times and reduced power consumption while maintaining competitive mAP scores, making it the preferred model for the data pipeline.

The inference time and computational complexity advantage offered by YOLOv9 was further confirmed through validation and testing. Both trained models were run on the training set of 32 micrographs. YOLOv8 was able to complete inference for all 32 test micrographs in 203.11 s, while YOLOv9 was able to complete inference for all 32 micrographs in 69.90 s. Furthermore, after running the models on the 32 micrographs on the validation set, YOLOv8 demonstrated an mAP50 of 0.995 and an mAP50–95 of 0.764, while YOLOv9 demonstrated an mAP50 of 0.995 and an mAP50–95 of 0.854. This shows that models can effectively identify objects at moderate localization threshold. However, the difference becomes more pronounced when evaluating with the stricter mAP50–95 metric, which averages precision across a range of IoU thresholds (from 0.50 to 0.95). Here, YOLOv9 achieved a notably higher mAP50–95 score of 0.854 compared to YOLOv8's 0.764. This suggests that YOLOv9 not only matches YOLOv8 in basic detection capability but also surpasses it in precise localization and robustness across varying overlap thresholds.

Collectively, the training results suggest that YOLOv9's architecture is better optimized for tasks involving complex small object localization in noisy or variable imaging conditions while maintaining fast inference speed and low computational complexity.

Therefore, YOLOv9 was selected as the preferred model for further segmentation and quantitative analysis in our CSC rosette detection pipeline.

3.2 | Detection Model Results

After training the YOLOv9 model on the expanded dataset, we evaluated its performance on previously unseen test images to assess detection accuracy and reliability. Figure 6A showcases representative detection results across multiple FF-TEM images, with bounding boxes highlighting predicted rosette structures and their associated confidence scores. The model demonstrated varying degrees of confidence in its predictions, which closely align with human-identifiable patterns in the rosette structures. Higher confidence scores, like 79% and 81%, were consistently assigned to well-defined rosettes with clear hexagonal arrangements and distinct boundaries. These high-confidence detections typically exhibit the characteristic six-lobed structure with optimum contrast against the background, making them readily identifiable even to human experts.

Interestingly, the model assigned lower confidence scores (28–54%) to less distinct rosettes, mirroring the uncertainty that human annotators would experience when examining these structures. This suggests that the model has effectively learned to recognize not only the presence of rosettes but also their quality and structural completeness. The model was the least confident when it made a false-positive detection with a confidence score of 28%, identifying a structure that appears like half a rosette. The 46% confidence detection shown in Figure 6A represents a case where the rosette structure is particularly difficult to discern due to low contrast. Furthermore, Figure 6A also illustrates the model's ability to detect multiple rosettes clustered close together, as demonstrated by the rosettes with confidence scores of 79% and 29%.

To determine the optimal confidence threshold for practical deployment, we analyzed the relationship between model confidence and detection performance. Figure 6B illustrates the F1 score as a function of confidence threshold, revealing a distinct peak at approximately 42.1% confidence. This peak represents the optimal balance between precision and recall, with the model achieving its maximum F1 score of 0.99 at this threshold. The shape of the F1-confidence curve provides valuable insights into model behavior. The steep ascent from 0% to approximately 20% confidence reflects the rapid elimination of obvious false positives as minimum confidence requirements are introduced. The more gradual increase from 20% to 42% demonstrates the model's improving precision as the confidence threshold rises, while still maintaining high recall rates. The plateau near the peak (between approximately 38–45% confidence) indicates a range where the model maintains near-optimal performance, providing some flexibility in threshold selection. The subsequent decline after 50% confidence becomes increasingly steep, signaling that higher thresholds disproportionately reduce recall by eliminating true positives, particularly objects with more ambiguous features or challenging presentations. This optimal 42.1% confidence threshold provides a data-driven parameter for automated object detection, balancing the need to identify all relevant instances while minimizing false detections. The model's ability to achieve an F1 score of 0.99 at this threshold demonstrates excellent detection capability, making it highly reliable for deployment when both precision and recall are critical considerations.

3.3 | Dataset and Preprocessing for Segmentation Model

Following the successful implementation of the detection model, we developed a segmentation framework to detect morphological regions within detected rosette structures. This stage required the creation of a specialized training dataset containing pixel-level

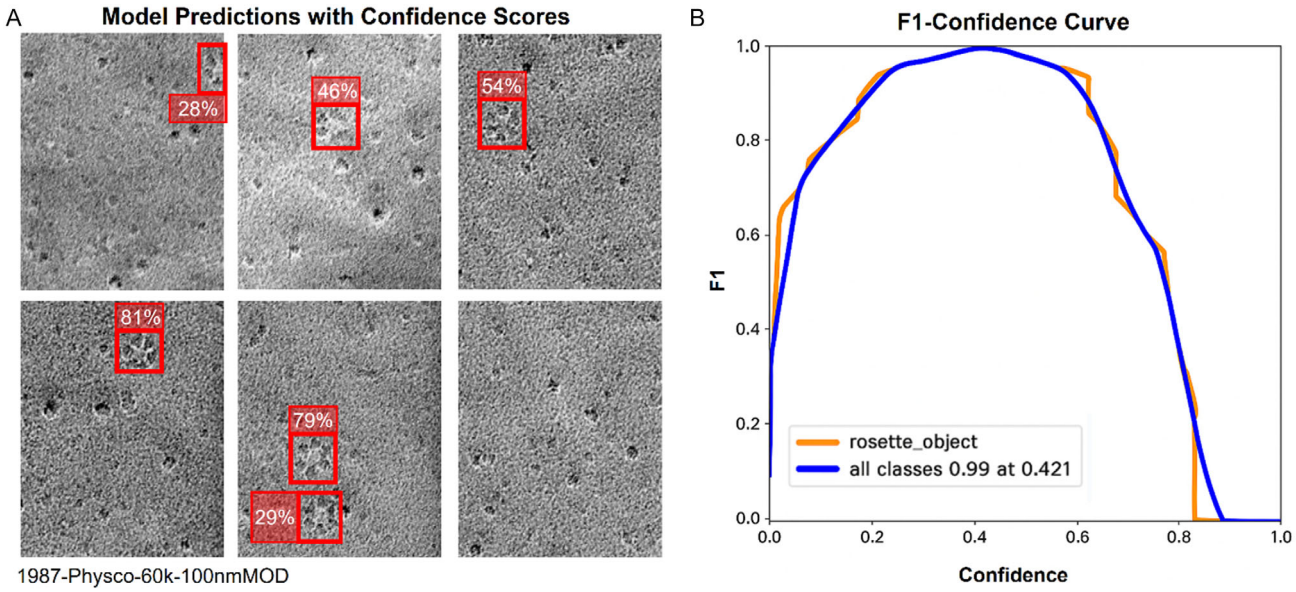


FIGURE 6 | (A) Representation of YOLOv9 performance on test dataset, including model predictions of rosette CSCs with a range of confidence scores. (B) Representation of F1-confidence curve demonstrating model performance during evaluation. The F1-confidence peaks at around 0.42 confidence intervals. CSCs = cellulose synthase complexes.

annotations of rosette subcomponents, including the outer boundary, individual lobes, and the central region. The segmentation dataset was systematically generated using the trained YOLOv8 detection model. A custom Python script extracted bounding boxes from the detection results, capturing individual rosette structures from the original FF-TEM micrographs. Each extracted bounding box underwent human verification to ensure accuracy and quality in detected rosettes, eliminating false positives and maintaining high dataset integrity. This verification step was critical for establishing a connection between the detection and segmentation stages, as the segmentation model's primary function is to process human-verified rosette detections and perform structural analysis.

The cropped rosette images measured approximately 350×350 pixels, with some dimensional variance resulting from differences in bounding box sizes across the dataset. Despite this variation, all images maintained sufficient resolution (ranging from 0.1 to 0.7 megapixels) to preserve the fine scale features necessary for accurate segmentation. Importantly, these cropped images retained all original contrast characteristics and features associated with the platinum/carbon shadowing techniques used in FF-TEM, ensuring that the segmentation model would learn to operate on authentic, unmodified data representative of real-world analysis conditions.

After verification, the final dataset comprising 100 high-quality cropped rosette images was imported into Roboflow [19], a web-based CV platform for annotation and dataset management. To minimize annotation bias and ensure structural consistency, two researchers independently performed pixel-level annotations on each image, focusing on generating annotations that follow the dark regions of the rosette and defining lobes, central region, and CSC boundary (Figure 7A). Using Polygon tools in Roboflow, each of the six lobes was individually traced, inclusive of morphological variations from the native triangular geometry [10, 11] that are likely attributable to the grain of the shadowing metal and potential migration of Pt during the replication process. Next, the central region was traced by connecting the innermost part of each lobe. Finally, the outer boundary of the rosette was traced, inclusive of the outermost part of all the lobes.

Each morphological component was assigned to one of three distinct classes: "lobe" for the six peripheral structures, "central" for the central region, and "outer" for the boundary area. This classification scheme enabled subsequent quantitative analysis of key structural parameters, including inter-lobe distances, lobe size heterogeneity, and overall rosette morphometry. The annotation data were stored in a format compatible with semantic segmentation models, with each pixel assigned a class label corresponding to its morphological region.

The dataset was separated into training, testing, and validation images with the allocated proportions set to 68%, 16%, and 16%, respectively. This split ratio was chosen based on considerations unique to the segmentation task for maximizing training data. The slightly reduced training portion (68% versus the more common 80%) with expanded validation and testing sets reflects the complex nature of rosette segmentation, where precise boundary delineation of morphological features is critical.

Counts for annotations in each lobe category of the dataset for each section of the split are shown in Figure 7B. Preprocessing steps included minimal data augmentation. Because conventional augmentation strategies, like applying random image rotations or adjusting contrast, reduced the overall quality of the images during Roboflow's processing, we used the bare minimum processing pipeline. Each image was preset to a resolution of 640×640 pixels and uniformly resized using aspect-preserving stretching to meet the segmentation model's input requirements. This approach sets

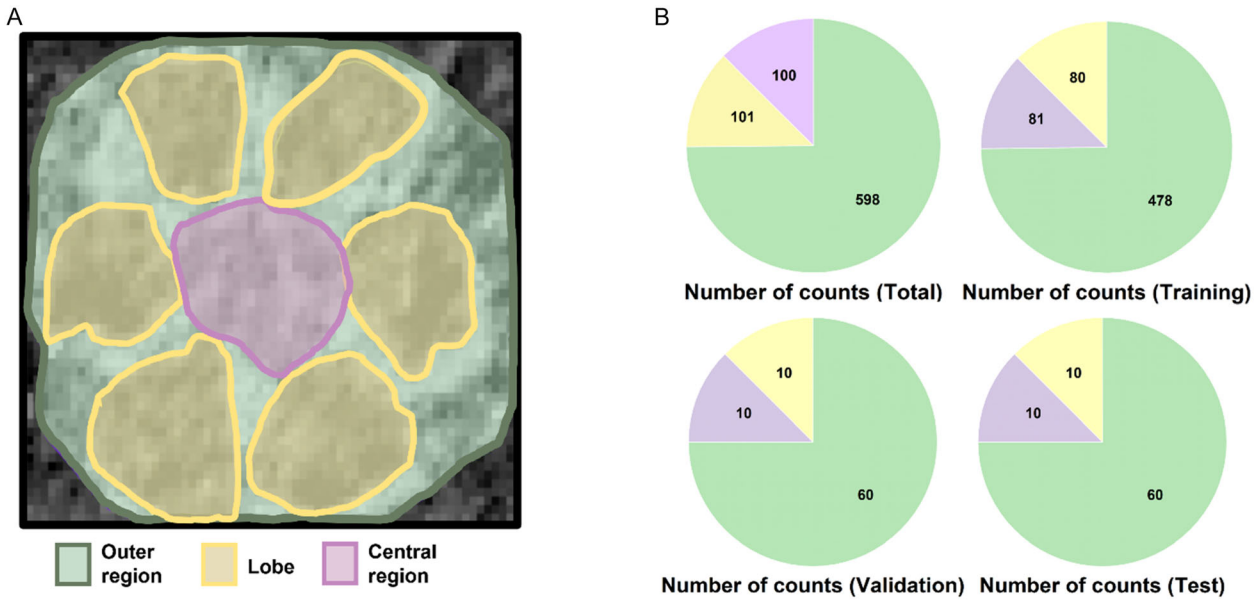


FIGURE 7 | Manual segmentation and data generation for segmentation model training. (A) Representative example of manual segmentation of a rosette CSC into three morphological regions: *outer* (green), *lobe* (yellow), and *central* (purple). These pixel-level annotations were used to generate training data for the segmentation model. (B) Pie charts showing the counts and distribution of annotated classes (outer, lobe, central) across the total dataset and its partitions: training, validation, and test sets. CSC = cellulose synthase complex.

the input dimensions for all images contributing to the model without compromising the important spatial details essential for model learning.

3.4 | Segmentation Model Performance

The annotated dataset was utilized to train two segmentation models that employ different architectural approaches: YOLOv8-seg and YOLOv9-seg. YOLOv8-seg is built upon the proven CSPDarknet53 backbone, which extracts rich and multi-scale features from input images. Unlike earlier versions that relied on traditional neck architectures, YOLOv8-seg replaces this component with a C2f module. The C2f module is designed to improve the flow of information and gradients through the network, ultimately enhancing both learning efficiency and segmentation accuracy. A notable feature of YOLOv8-seg is its dual segmentation heads. This means that, in addition to the standard detection head responsible for identifying object classes and bounding boxes, there is a dedicated head for generating segmentation masks. Central to this process is the Proto module, which consists of several fully connected (FCN) layers. The Proto module produces a set of prototype masks--essentially generic shapes that can be linearly combined to match the contours of detected objects. Each object instance receives its own set of mask coefficients, which are used to blend these prototypes into a final, instance-specific segmentation mask. This approach draws inspiration from the Yolact segmentation framework [27], which pioneered the use of prototype masks and mask coefficients for efficient instance segmentation. YOLOv8-seg maintains five detection modules, as in the original YOLO architecture, but adds an extra output specifically dedicated to mask generation. This design allows YOLOv8-seg to simultaneously deliver high-quality object detection and precise instance segmentation, all within a unified and efficient model architecture.

In contrast, YOLOv9-seg introduces two significant architectural innovations: PGI and GELAN. The PGI system prevents information loss during gradient updates while preserving crucial features throughout the network's depth, addressing a fundamental limitation in deep neural networks. GELAN optimizes parameter utilization and computational efficiency through strategic gradient path planning, allowing for more effective feature extraction, particularly beneficial for complex segmentation tasks. While YOLOv8-seg excels in rapid processing with established segmentation principles, YOLOv9-seg's architecture prioritizes information preservation and gradient flow optimization, specifically designed to maintain fine-grained feature details through deep network propagation.

Both models were trained using identical computational resources and hyperparameter configurations to ensure a valid comparison. The training process spanned 25 epochs. This standardized training protocol enabled direct performance comparisons between the two models and their architectural differences. As illustrated in Figure 8, the learning trajectories of YOLOv8-seg and YOLOv9-seg exhibited markedly different patterns across the training epochs. YOLOv8-seg demonstrated superior initial performance, achieving higher mean average precision (mAP) values during the first 10 epochs. This early advantage can be

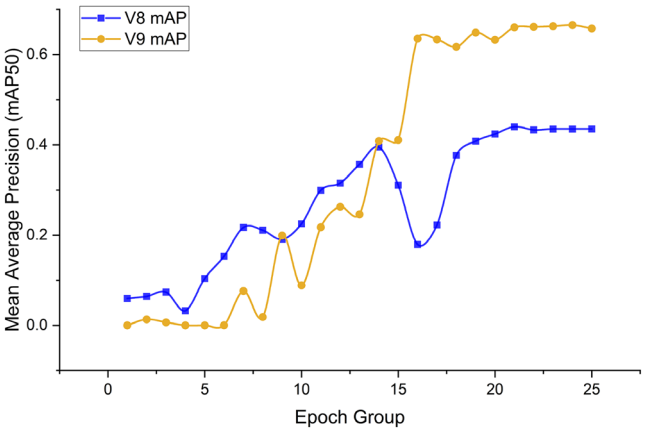


FIGURE 8 | Training performance comparison of YOLOv8 and YOLOv9 segmentation models through 25 epochs using mAP50 score.

attributed to YOLOv8's more mature feature extraction backbone, which excels at capturing general object boundaries even with limited training iterations. However, a decline around epoch 15 suggests potential optimization challenges when learning the intricate morphological features of rosette CSCs. In contrast, YOLOv9-seg displayed initially inferior performance with substantial epoch-to-epoch variability during early training stages (epochs 0–10). This volatility likely stems from YOLOv9's more complex architectural design, requiring additional training iterations to optimize its parameters effectively. The dramatic performance improvement observed around epoch 14 represents a critical inflection point where YOLOv9-seg not only surpassed YOLOv8-seg but also established a substantial performance gap that persisted throughout the remaining training process. By the final epochs, YOLOv9-seg achieved an mAP of approximately 0.69, representing a 64.3% relative improvement over YOLOv8-seg's plateau of 0.42. The mAP of 0.69 represents an average score across all the classes and class-specific performance varied across the three morphological categories. The inner class exhibited exceptional segmentation accuracy with an mAP50 of 0.99, closely followed by the lobes class with an mAP50 of 0.98. These high-precision metrics indicate the model's robust capability to distinguish the distinctive features of both the central region and the six peripheral structures. In contrast, the outer boundary class performed substantially below expectations, achieving only an mAP50 of 0.12. This marked performance discrepancy can be attributed to the outer boundary's unique spatial characteristics within the dataset; specifically, it encompasses nearly the entire cropped image area, providing minimal distinctive boundary features for the model to identify. Segmentation models rely on contrast between object and non-object pixels to differentiate class boundaries. When a structure spans most of the frame, there is minimal external context for comparison. This results in the features inside and outside the boundary appearing similar, making it difficult for the model to accurately localize edges. Furthermore, models such as YOLO-seg optimize overlap-based metrics like IoU, which are highly sensitive to edge deviations. For large regions, even minor misalignments along the boundary can cause large penalties, whereas smaller, high-contrast regions such as the lobes are more tolerant. Overall, unlike the inner and lobe classes, which present clear morphological transitions and contrast boundaries, the outer boundary lacks sufficient discriminative visual cues. This challenge

highlights the inherent difficulty in segmenting structural elements that lack clear delimitation from their surrounding context, particularly when the target region constitutes a majority of the input image dimensions. Using larger box sizes as we have done for prior manual measurements and image averaging [10, 11] could mitigate this issue in the future.

Despite YOLOv8's computational efficiency, YOLOv9-seg's superior late-stage accuracy and architectural innovations in gradient preservation justified its selection for further analysis. While YOLOv8 remains viable for real-time applications, YOLOv9's ability to maintain feature integrity across deep layers proved critical for resolving subtle rosette morphologies. Consequently, YOLOv9-seg was selected for final segmentation and quantitative analysis in our pipeline, prioritizing precision over inference speed for the precise nature of the structural characterization task.

3.5 | Segmentation Results

The trained YOLOv9-seg model was evaluated against the previously unseen validation test dataset to assess its performance on unseen rosette images. As illustrated in Figure 9A, the model successfully identified the three distinct morphological regions of the rosette structures with varying degrees of confidence. The segmentation results reveal the model's ability to discriminate between different structural components of the rosettes, though with notable variability in confidence scores. The central region (depicted in bright red) shows confidence scores ranging from 0.4 to 0.9 across test samples, rather than consistently high values. Similarly, the six peripheral lobes (shown in lighter pink) exhibit confidence scores varying from 0.3 to 0.9. Despite this variability, the model differentiates the characteristic hexagonal "rosette" pattern critical for understanding the CSC complex structure. Importantly, image quality correlated with prediction confidence as samples with higher contrast consistently produced better overall confidence scores across all structural components. The visible differences in rosette shapes across the sample images are effectively captured in the segmentation masks, demonstrating the model's adaptability to morphological variations that occur in FF-TEM micrographs of CSCs.

The outer boundary region (rendered in orange) encompasses the entire rosette structure, along with some extra surrounding space in some instances, and shows consistently high confidence scores (0.9–1.0) despite its lower mAP50 metrics as discussed during training. This apparent contradiction between visual rendering and quantitative performance metrics substantiates our earlier analysis that, while the model confidently identifies the outer boundary region, its extensive coverage of the image area creates challenges for precise boundary delineation when evaluated using standard IoU-based metrics.

Figure 9B provides critical insights into the model's confidence–precision relationship through the F1-confidence curve. The inner and lobes classes maintain exceptionally high F1 scores (0.9+) across a wide confidence threshold range (0.1–0.8), indicating robust and stable segmentation performance. This plateau region of the curve demonstrates that the model's predictions for these classes remain consistently accurate even when lowering the confidence threshold, suggesting strong feature discrimination capabilities for these morphological components. Notably, this finding contextualizes the variable confidence scores observed in Figure 9A, revealing that the wide range of confidence values (0.3–0.9) displayed during model evaluation does not necessarily indicate inconsistent performance. Rather, the F1-confidence curve confirms that even rosette components identified with lower confidence scores (such as the 0.3–0.4 range seen in some lobes) likely represent accurate segmentations. This resilience to confidence variability validates the model's practical utility for rosette analysis, as it can reliably identify structural components even in challenging FF-TEM micrographs with variable contrast and morphological characteristics. The sharp decline in F1 scores for all classes beyond the 0.8 confidence

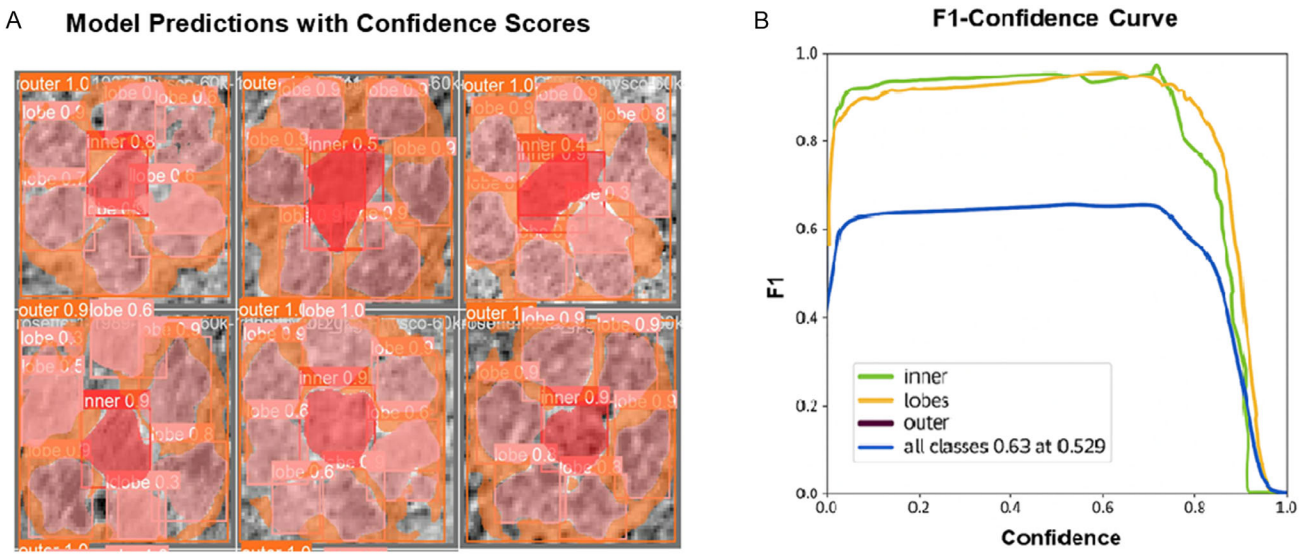


FIGURE 9 | Segmentation results of Yolov9. (A) Representation of Yolov9 performance on test dataset, including model predictions of segmented lobes, central region, and outer lobe area. Associated confidence scores for the predictions are shown. (B) Representation of F1-confidence curve demonstrating model performance during evaluation for all 3 classes. The F1-confidence peaks at around 0.65, indicating optimal model performance.

threshold illustrates a practical challenge in model deployment: setting overly strict confidence requirements causes the model to miss many correct segmentations. This pattern is particularly pronounced for the outer class (blue line), which maintains substantially lower F1 scores (∼0.6) across all confidence thresholds, confirming the challenges in precise outer boundary segmentation as discussed previously. The aggregate F1 score across all classes (dark blue line) peaks at 0.63 with an optimal confidence threshold of 0.529, providing valuable guidance for model deployment in automated analysis pipelines. This moderate optimal threshold represents a balance between precision and recall, particularly important when analyzing the complex morphological features of rosette structures, where both false positives and false negatives can impact structural interpretation.

3.6 | Performance

The trained YOLOv9 detection model and YOLOv9 segmentation model were successfully integrated into the comprehensive data pipeline described in Figure 2, creating a robust system for automated rosette characterization. To evaluate the pipeline's real-world performance and reliability, we processed three batches of FF-TEM images of rosettes in the original dataset (Table 1). The data for every micrograph (image) analyzed by the model are present, where 'Image number' refers to the particular micrograph from our dataset(s). The magnification on each imaging day was precisely calibrated using catalase crystal standards. Careful measurement of scale bars revealed a maximum variance across samples of 1.2% (equivalent to 0.3 nm uncertainty for a typical 25 nm diameter rosette).

Table 1 shows that the pipeline successfully identified and analyzed multiple rosettes per micrograph in the *original dataset*. The average number of rosettes detected per micrograph by the pipeline (11.8) is similar to the vetted dataset's average (13.9, raw data available in the dataset), indicating that the detection sensitivity of the automated system is on par with manual curation. For example, images 1985–1991 contained 11, 11, 4, 6, and 5 rosettes identified by human evaluators, thus showing a close match with the model in batch 1. The greater number of rosettes identified in batch 3 likely indicates more active cellulose synthesis in the analyzed cell (or region of the cell) when it was frozen.

Numerical analysis of rosette geometry may provide one means of identifying functional differences between different groups of CSCs. The size and morphology of the rosette CSC enable the formation of a semi-crystalline cellulose microfibril from 18 glucan chains exported in close proximity. The individual glucan chains must quickly coalesce with their neighbors to avoid self-folding in order to form native, parallel-chain cellulose [8]. This is accomplished within the composite CSC as three glucan chains exported about 3 nm apart within one lobe first form a protofibril, followed by the coalescence of six protofibrils exported about 10 nm apart by the six lobes of the CSC [10–12, 28]. Therefore, researchers can compare the observed dimensions of CSC structures within new datasets with prior observations to detect deviations that could be biologically significant. The average diameter values determined by CV for each batch were 22.5, 23.8, and 24.0 nm, respectively; whether biological, experimental, or stochastic effects explain the difference requires further investigation. The overall average for the rosette CSC diameter for all batches was 23.76 ± 2.23 nm, similar to the previous average estimate generated by hand and image averaging for the same image set (23.4 nm) [11]. Given the >1.25 nm grain size that is inherent in FF-TEM replicas prepared by our methods, the two average values can be considered to be the same. Figure 10A reveals a normal distribution of rosette diameters in the test set, centered around 23–24 nm. The highest frequency bin (22.09–23.77 nm) accounts for nearly 30% of all detected rosettes, with the distribution tails extending from ≈18 to 26 nm. The variations could represent different activity states of the CSCs, which require further investigation. Distances between lobes (Figure 4) remained consistent across batches with an overall average of 12.20 ± 2.46 nm (Figure 10B), which was larger than the 10 nm predicted based on theoretical assembly of individual lobes [28].

The original dataset used for model development, evaluation, and comparison with manual object identification was curated with

TABLE 1 | YOLOv9 model analysis results for assembled original dataset (in batches) and validation dataset.

Image Number	Rosette Count	Average Rosette Diameter (nm)	Average Lobe to Lobe Distance (nm)
1985	10	22.61 ± 2.51	11.68 ± 3.11
1988	11	23.46 ± 1.11	12.68 ± 2.43
1989	4	21.97 ± 1.10	12.62 ± 2.73
1990	6	23.71 ± 2.15	12.87 ± 2.58
1991	5	20.81 ± 2.52	11.09 ± 3.84
Batch 1 Averages	**7.2**	**22.51 ± 1.88**	**12.19 ± 2.94**
2006	7	23.69 ± 1.57	12.18 ± 2.07
2007	11	23.90 ± 1.62	12.58 ± 1.98
2010	12	23.38 ± 1.73	11.56 ± 2.52
2025	7	23.63 ± 3.25	11.27 ± 2.52
2026	11	24.26 ± 1.37	11.43 ± 2.17
Batch 2 Averages	**9.6**	**23.77 ± 1.91**	**11.80 ± 2.25**
2035	35	24.03 ± 2.60	12.59 ± 2.40
2036	33	24.45 ± 1.78	12.47 ± 2.24
2037	9	24.10 ± 2.68	12.39 ± 1.22
2040	10	23.97 ± 2.32	12.13 ± 2.63
2041	6	23.53 ± 1.32	11.19 ± 2.08
Batch 3 Averages	**18.6**	**24.02 ± 2.14**	**12.15 ± 2.11**
***Original Dataset* Averages**	**11.8**	**23.76 ± 2.23**	**12.20 ± 2.46**
***Validation* Dataset Averages**	**6.49**	**23.67 ± 3.18**	**11.87 ± 3.24**

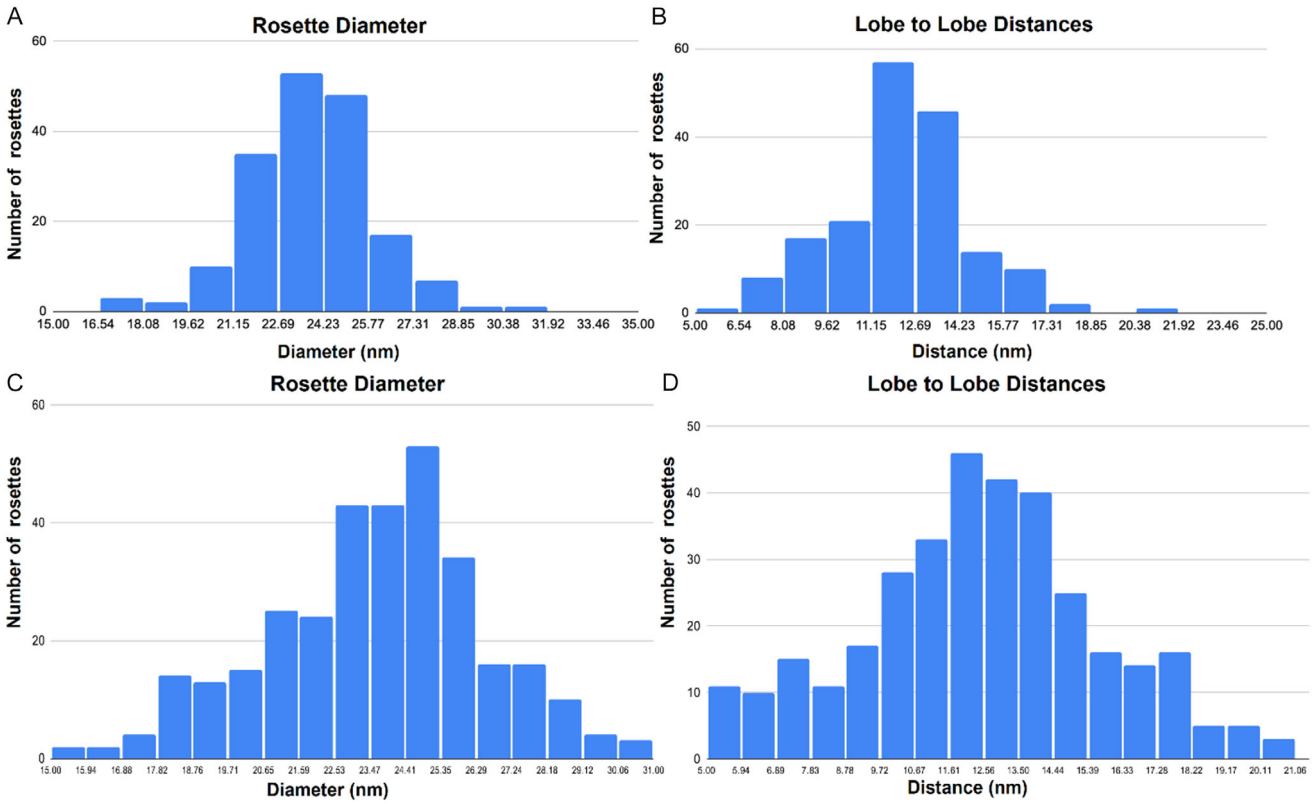


FIGURE 10 | Representation of the distribution of (A) rosette diameters and (B) lobe-to-lobe distances of the produced data of the original dataset, and (C) rosette diameters and (D) lobe-to-lobe distances for the secondary validation dataset.

regions with relatively flat topography and high contrast to enable reliable geometric measurements. As an additional validation step, a *new validation dataset* was assembled to assess the pipeline's ability to generalize to imaging conditions not represented in this training dataset. This new validation dataset was not included in the original training, validation, or test splits and includes 73 FF-TEM micrographs (images) with greater variation in contrast, shadowing, and topography compared to the original images. Furthermore, the validation dataset was recorded directly with a digital camera rather than scanned from film to further assess the pipeline's ability to adapt to different and more variable image characteristics. As a result, the new dataset spans a broader range of image attributes and provides a stronger test for model generalizability. One can observe the new dataset's normal distribution of rosette diameters (Figure 10C) and lobe-to-lobe distances (Figure 10D), as well as summarized averages for rosette diameters and lobe-to-lobe distances in Table 1. The extracted quantitative features (average rosette diameter = 23.67 ± 3.18 nm; average lobe-to-lobe distance = 11.87 ± 3.24 nm) closely matched both the outputs of the primary dataset and previously reported manual measurements. This supports the reliability of the segmentation process even when imaging conditions differ substantially from those included during model training. However, some additional validation and fine-tuning steps were necessary to adjust the model performance, thus indicating the importance of human-in-the-loop and the necessity of domain experts to be involved in the model assembly and accuracy evaluation.

Freeze-fracture TEM images arising from rotary shadowing of platinum-carbon may appear low contrast, but the intramembrane particles that compose rosette CSCs are typically darker against the lighter gray background that results when shadowing the membrane lipids. However, the degree of contrast between a CSC and the background may vary due to factors such as unavoidable small differences in membrane topography. Analysis of the detection model on the new dataset indicates that the model identified rosettes under variable imaging conditions with moderate success. For example, comparing the model's performance against human experts evaluating the micrographs, we were able to detect moderate (up to 30% in some cases) variability in the results, with the model consistently 'dropping' objects in areas of images with low contrast-to-brightness ratio. To illustrate this, we can take several micrographs with 40 and 50 rosettes identified upon manual inspection, where only 23 and 34 rosettes were detected by the model (with omitted objects located in low-contrast areas of the image). At the same time, in areas of images that were slightly out of focus, the model outperformed a human expert by detecting rosettes with higher sensitivity (for example, 11 identified rosettes vs 8 detected by a human expert originally). We also observed that fine-tuning boundary conditions for the lower and upper object size limits improved the pipeline performance. As an additional human-in-the–loop interaction with the model, lowering the confidence threshold increased sensitivity, particularly for low-contrast micrographs. While the model did not perform as well when applied to low-contrast images with varying topography, it performed more reliably on images with higher contrast and less clustering (lower protein particle density). This is consistent with the absence of such challenging conditions in the training set and demonstrates some of the model's limitations in generalizing to lower contrast and variable images. However, the model's ability to detect a significant number of biologically meaningful objects

under these circumstances highlights the robustness of its learned feature representations.

These findings underscore the potential value of transfer learning. The model's ability to detect rosettes in low-contrast and variable images indicates that it has already learned meaningful structural representations of the rosette features, allowing it to generalize beyond the conditions present in the training set. This provides a strong foundation for further refinement by fine-tuning the model with additional training data curated to capture a broader range of imaging conditions. Transfer learning could enhance its generalization capability and improve detection confidence, particularly in settings where variability is large. Such an approach would be especially advantageous for applications requiring reliable performance on suboptimal image conditions and quality, further extending the model's utility.

Overall, CV and manual results are similar while also expanding empirical knowledge of lobe-to-lobe distance that can be considered in future models of cellulose microfibril formation.

4 | Conclusion

This study demonstrates the successful implementation of a deep learning-based pipeline for automated detection and characterization of cellulose-synthesizing rosette CSCs in FF-TEM micrographs. Our results establish that modern object detection and segmentation architectures can effectively analyze nanoscale biological structures with accuracy comparable to manual annotation. A pipeline for automated analysis of FF-TEM micrographs represents a substantial advancement in structural biology research. Manual annotation of rosette structures is labor-intensive and subject to potential inconsistencies between operators. Our pipeline reduces processing time to approximately 2 min per micrograph (including human-in-the-loop verification), representing a 90% reduction in analysis time. The higher throughput will enable comprehensive analysis of substantially larger datasets, facilitating more statistically robust characterization of structural variations and population-level properties that would be impractical through manual methods alone.

Beyond time efficiency, automated analysis offers several methodological advantages. The consistent application of detection and segmentation criteria eliminates variations that may arise from unconscious human bias during manual detection and analysis. The framework also provides objective confidence metrics for both detection and segmentation, enabling researchers to implement data-driven quality thresholds. Furthermore, the segmentation model's ability to delineate specific morphological regions (outer boundary, peripheral lobes, and central region) enables detailed quantitative analysis of the structural organization of an intramembrane protein complex.

The success of the YOLO architecture in detecting and characterizing nanoscale structures has broader implications for biological imaging. While YOLO models were initially developed for macroscale CV applications, our results demonstrate their adaptability to nanoscale objects in TEM micrographs. The ability of these models to recognize subtle morphological patterns despite the low contrast and high noise (e.g., due to finite platinum-carbon grain size) in FF-TEM micrographs suggests high potential for application to other challenging microscopy modalities. The presented results indicate that the proposed pipeline demonstrates transferability and stability without requiring retraining for high-quality and high-contrast data. At the same time, the incorporation of the domain experts in the model assembly, fine-tuning, and validation significantly increases its performance. Thus, for low-contrast images with varying topography, such 'transfer learning' may provide an effective strategy to further optimize detection sensitivity and extend the pipeline's applicability.

The architectural principles demonstrated in this pipeline, separating detection from detailed segmentation and incorporating human-in-the-loop verification to maintain data quality, establish a framework that can be transferred to other biological structure analysis tasks. Potential applications include the characterization of other membrane protein complexes, viral capsids, cytoskeletal assemblies, and other nanoscale biological structures where manual analysis currently creates bottlenecks in research progress. The recent advances in foundation models and transfer learning further enhance the transferability of this approach, potentially reducing the annotation requirements for new structural targets.

4.1 | Supplemental

The code and datasets used for the model are available in the GitHub repository. The repository contains a README that provides a detailed description of the contents included. In summary, the repository includes code used for training the YOLOv8 and YOLOv9 detection models and for training the YOLOv8 and YOLOv9 segmentation models. It also includes test and validation code for validating and testing all the models described above. Further, supplemental information including model training, validation, and test metrics, such as box loss, precision, and recall, and detection result output for all test datasets. Three different datasets are included in the repository. The first is the original micrograph dataset with annotated rosette objects, previously used for manual analysis of CSC morphology [11]. This dataset contains 32 images, along with labels that are recorded in the YOLOv8 format. Further information on the dataset is included in the README file included within the original dataset folder. The second dataset contains the dataset used for object detection training, validation, and testing. This dataset includes 650 images and is split into a training, validation, and testing set. The third dataset contains the dataset used for the segmentation training, validation, and testing.

Acknowledgments

S.M. and L.C. were supported by the National Science Foundation through the Research Experiences for Undergraduates (REU) Site in Materials Research with Data Science (MAT-DAT) at North Carolina State University (DMR-2150360). A.V.G. and Y.G.Y. were supported by the Science and Technologies for Phosphorus Sustainability (STEPS) Center, a National Science Foundation Science and Technology Center (CBET 2019435). We thank Dr. Albert L. Kwansa for providing the guidance and Figure 1 inset. The freeze fracture electron microscopy work of

A.W.R. and C.H.H. was supported by the Center for LignoCellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0001090. Collection of the new validation dataset was funded by a grant from the National Science Foundation to AWR (IOS-2124176). We thank Dr. Katayoun Mansouri for taking the 32 TEM micrographs in the baseline dataset of this study, as also supported by the Award DE-SC0001090. The writing contributions of C.H.H. were supported by the Research Capacity Fund (HATCH), project award no. 7005416, from the U.S. Department of Agriculture's National Institute of Food and Agriculture. An AI language model (ChatGPT, OpenAI) was used during the manuscript preparation process to rephrase and improve readability. All edits were verified by the authors.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The codebase and datasets used for training and evaluation of the YOLOv8 and YOLOv9 detection and segmentation models are publicly available at GitHub: https://github.com/AI-and-Materials-Simulations-Laboratory/RosetteCV.

References

1. M. Zurowietz and T. W. Nattkemper, "Current Trends and Future Directions of Large Scale Image and Video Annotation: Observations From 4 Years of BIIGLE 2.0," *Frontiers in Marine Science* 8, no. 760036 (2021): 1–16.

2. M. Zurowietz, D. Langenkämper, B. Hosking, H. A. Ruhl, and T. W. Nattkemper, "MAIA—A Machine Learning Assisted Image Annotation Method for Environmental Monitoring and Exploration," *PLOS ONE* 13, no. 11 (2018): 1–18.

3. A. Manbohi and S. H. Ahmadi, "Portable Smartphone-Based Colorimetric System for Simultaneous on-Site Microfluidic Paper-Based Determination and Mapping of Phosphate, Nitrite and Silicate in Coastal Waters," *Environmental Monitoring and Assessment* 194, no. 3 (2022): 190.

4. S. V. Kalinin and A. Gruverman, Scanning Probe Microscopy: Electrical and Electromechanical Phenomena at the Nanoscale, (Springer New York, 2007).

5. G. Lobet, "Image Analysis in Plant Sciences: Publish Then Perish," *Trends in Plant Science* 22, no. 7 (2017): 559–566.

6. T. Vamerali, M. Bandiera, and G. Mosca, "Minirhizotrons in Modern Root Studies," in Measuring Roots: An Updated Approach, ed. S. Mancuso (Springer, Berlin, 2012), 341–361.

7. E. A. Holm, R. Cohn, N. Gao, et al., "Computer Vision and Machine Learning for Microstructural Characterization and Analysis," *Metallurgical and Materials Transactions A* 51, no. 12 (2020): 5985–5999.

8. C. H. Haigler and A. W. Roberts, "Structure/Function Relationships in the Rosette Cellulose Synthesis Complex Illuminated by an Evolutionary Perspective," *Cellulose* 26, no. 1 (2019): 227–247.

9. A. L. Kwansa, A. Singh, J. T. Williams, C. H. Haigler, A. W. Roberts, and Y. G. Yingling, "Structural Determination of a Full-Length Plant Cellulose Synthase Informed by Experimental and in Silico Methods," *Cellulose* 31, no. 3 (2024): 1429–1447.

10. E. M. Roberts, K. Yuan, A. M. Chaves, et al., "An Alternate Route for Cellulose Microfibril Biosynthesis in Plants," *Science Advances* 10, no. 50 (2024): eadr5188.

11. B. T. Nixon, K. Mansouri, A. Singh, et al., "Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex," *Scientific Reports* 6, no. 1 (2016): 28696.

12. P. Purushotham, R. Ho, and J. Zimmer, "Architecture of a Catalytically Active Homotrimeric Plant Cellulose Synthase Complex," *Science* 369, no. 6507 (2020): 1089–1094.

13. L. Sethaphong, C. H. Haigler, J. D. Kubicki, et al., Model of a Plant Cellulose Synthase, *Proceedings of the National Academy of Sciences* 110, no. 18 (2013): 7512–7517.

14. S. C. Mueller and R. M. Brown Jr., "Evidence for an Intramembrane Component Associated with a Cellulose Microfibril-Synthesizing Complex in Higher Plants," *The Journal of Cell Biology* 84, no. 2 (1980): 315–326.

15. N. J. Severs and D. M. Shotton, "Chapter 25 - Rapid Freezing of Biological Specimens for Freeze Fracture and Deep Etching," Cell Biology, ed. J. E. Celis, 3rd ed. (Academic Press, Burlington, 2006), 249–255.

16. C. Ma, Y. Fu, D. Wang, R. Guo, X. Zhao, and J. Fang, "YOLO-UAV: Object Detection Method of Unmanned Aerial Vehicle Imagery Based on Efficient Multi-Scale Feature Fusion," *IEEE Access* 11 (2023): 126857–126878.

17. S. Kang, Z. Hu, L. Liu, K. Zhang, and Z. Cao, "Object Detection YOLO Algorithms and Their Industrial Applications: Overview and Comparative Analysis," *Electronics* 14, no. 6 (2025): 1104–1140.

18. M. Mao and M. Hong, "YOLO Object Detection for Real-Time Fabric Defect Inspection in the Textile Industry: A Review of YOLOv1 to YOLOv11," *Sensors* 25, no. 7 (2025): 2270-2313.

19. S. Alexandrova, Z. Tatlock, and M. Cakmak, "RoboFlow: A Flow-Based Visual Programming Language for Mobile Manipulation Tasks," *2015 IEEE International Conference on Robotics and Automation (ICRA)* (2015): 5537–5544.

20. Y. Tong, X. Luo, L. Ma, S. Xie, W. Yang, and Y. Guo, "Saliency Information and Mosaic Based Data Augmentation Method for Densely Occluded Object Recognition," *Pattern Analysis and Applications* 27, no. 2 (2024): 34.

21. H. Zhang, S. Zhang, and R. Zou, "*Select-Mosaic: Data Augmentation Method for Dense Small Object Scenes*," 2024, https://doi.org/10.48550/arXiv.2406.05412.

22. P. Li, Y. Fan, Z. Cai, Z. Lyu, and W. Ren, "Detection Method of Marine Biological Objects Based on Image Enhancement and Improved YOLOv5S," *Journal of Marine Science and Engineering* 10, no. 10 (2022): https://doi.org/10.3390/jmse10101503.

23. R. Sapkota, M. Flores-Calero, R. Qureshi, et al., "Advances to its Genesis: A Decadal and Comprehensive Review of the You Only Look Once (YOLO) series" *Artificial Intelligence Review* 58, no. 9 (2025): 274.

24. R. Varghese and M. S., "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)* (2024): 1–6, https://doi.org/10.1109/ADICS58448.2024.10533619.

25. J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction* 5, no. 4 (2023): 1680–1716.

26. M. Yaseen, "*What is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector*," 2024, https://doi.org/10.48550/arXiv.2409.07813.

27. D. Bolya, C. Zhou, F. Xiao, Y. J. Lee, "YOLACT: Real-Time Instance Segmentation," *2019 IEEECVF International Conference on Computer Vision (ICCV)*, (2019), 9156–9165.

28. E. Cosgrove, P. Dupree, E. Gomez, C. Haigler, J. Kubicki, and J. Zimmer, "How many glucan chains form plant cellulose microfibrils?," *Biomacromolecules* 25 (2024): 6357–6366.