IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Instance Segmentation for Occlusion-Robust Fruit Counting: A Comprehensive Review

[1]Akshata Raut, [2]Mimansha Singh, [3]Foram Shah, [4]Hetal Rana

[1,2,3,4]Assistant Professor

[1,2,3,4]Department of Computer Engineering,

[1,2,3,4]Thakur College of Engineering, Mumbai, India

Abstract: In orchards, where occlusions, overlaps, and fluctuating lighting make traditional detection brittle, instance segmentation—predicting pixel-level masks for individual object instances—has become essential to automated fruit detection and counting. This survey compares representative algorithms (Mask R-CNN, YOLACT/YOLACT++, SOLO/SOLOv2, and recent amodal/occlusion-aware variants), reviews instance-segmentation techniques used for fruit counting under occlusion, categorizes methods into taxonomy (two-stage vs. one-stage/real-time, amodal/multimodal, 3D/view-synthesis), and looks at how each handles partial visibility and tight clusters [9]. We tabulate common datasets, metrics, and evaluation protocols, quantify typical trade-offs (accuracy vs. runtime vs. robustness to occlusion), and synthesize results from agricultural applications (apples, oranges, grapes, tomatoes) that use instance segmentation. Lastly, we propose research directions (multimodal fusion, 3D-aware segmentation, and active learning for rare occlusion cases) and identify open problems (amodal completion, domain shift to field conditions, multi-view fusion, and real-time edge deployment). The purpose of this survey is to serve as a useful manual for researchers selecting segmentation techniques for tasks involving fruit counting and harvest readiness. [1] [7].

***Index Terms* - Instance segmentation, fruit counting, occlusion, Mask R-CNN, YOLACT, amodal segmentation, agritech, counting**

I. INTRODUCTION

Accurate fruit counting plays a crucial role in modern orchard management, supporting essential tasks such as labour planning, harvest scheduling, and yield estimation. However, obtaining reliable counts in natural orchard environments remains challenging due to dense fruit clustering, complex foliage structures, variable lighting, and frequent occlusions. These factors cause traditional object detection methods—particularly those relying solely on bounding boxes—to perform inconsistently in real-world field conditions.

Instance segmentation has emerged as a powerful alternative because it produces pixel-level masks for each individual fruit, enabling precise separation of overlapping or touching fruits and improving size, area, and ripeness-related estimations. Recent advances in instance segmentation, including two-stage architectures such as Mask R-CNN and fast one-stage approaches like YOLACT/YOLACT++, have significantly improved both accuracy and deployability in agricultural settings. Additionally, occlusion-aware methods such as amodal segmentation and 3D/multi-view reconstruction have further advanced the ability to detect partially hidden fruits.

This survey focuses on techniques specifically designed to address occlusion, overlap, and visual complexity in fruit counting across various crops such as apples, oranges, grapes, and tomatoes. We review both classical and state-of-the-art models, highlight their applicability to agricultural scenarios, and discuss recent foundational and emerging works that have shaped the field [8]. The goal is to provide a consolidated understanding of modern instance segmentation strategies and their effectiveness in improving fruit counting under real orchard conditions.

II. Literature Survey

Table 1: Literature Survey

Year	Paper (short title)	Authors	Summary	Results / Key findings	Identified research gap
2017	Mask R-CNN [1]	K. He, G. Gkioxari, P. Dollár, R. Girshick	Two-stage detector (Faster R-CNN) extended with an RoI mask branch to predict instance masks. Widely used baseline for high-fidelity instance segmentation.	Strong mask fidelity and stable training became the standard baseline adapted in many agricultural studies for fruit mask generation.	Moderate runtime for real-time systems; RoI proposals can fail in dense occlusion unless adapted for amodal or cluster scenarios.
2019	YOLACT / YOLACT++ (prototype-based real-time) [9]	D. Bolya et al. (and followups)	One-stage prototype-based approach: produce global prototype masks and per-instance coefficients for fast mask assembly. Designed for speed.	Excellent real-time throughput (>30 FPS on GPU) with reasonable mask quality; attractive for drones/robots.	Prototype linear combination can under-represent fine boundaries or heavily occluded shapes — struggles in dense overlapping fruit clusters unless modified (task-specific variants needed).
2019	Deep learning review — fruit detection & yield (survey) [5]	A. Koirala, K.B. Walsh, Z. Wang, C. McCarthy	Survey of DL methods applied to fruit detection and yield estimation — summarizes detectors, datasets, and practical challenges (lighting, occlusion).	Highlights that DL improved detection/estimation accuracy but real-world robustness (domain shift, occlusion) remains challenging.	Need for instance segmentation focus, mask-level datasets, and standardized occlusion benchmarks for fruit counting tasks.
2020	Fuji-SfM dataset (Fuji apple detection + SfM) [6]	J. Gené-Mola et al.	Dataset of annotated images + point clouds (Structure-from-Motion) for apple detection and localization — enables multi-view / 3D-aware approaches.	Provides paired image + 3D data enabling geometry-based counting/size estimation; useful baseline dataset for testing multi-view/3D methods.	Multi-view data useful but capture complexity and annotation cost limit wide adoption; need more diverse multi-view orchard datasets.

2023	Looking behind occlusions: amodal segmentation for on-tree apple size estimation [2]	J. Gené-Mola et al.	Introduces/amplifies amodal segmentation for apples: model predicts full object silhouette (visible + occluded) to improve size estimation and counting under occlusion.	Amodal predictions significantly improved on-tree apple size estimation and counting accuracy in occluded scenarios vs. visible-only masks.	Amodal methods often rely on object-specific priors (apple shapes); generalization across fruit species with different shapes remains open.
2023	Simultaneous fruit detection & size estimation (multitask networks) [3]	M. Ferrer-Ferrer, J. Ruiz-Hidalgo, E. Gregorio et al.	Multitask deep networks jointly detect fruits and regress size — improves per-instance estimates by sharing features.	Joint detection + size estimation yields better sizing than sequential pipelines; helps in yield estimation tasks.	Multitask pipelines still require dense mask/size labels; annotation cost and robustness to occlusion/lighting are limiting factors.
2023	Instance segmentation & number counting of grape berry images [8]	Y. Chen, X. Li, M. Jia, J. Li, T. Hu, J. Luo	Instance segmentation tailored to small, tightly clustered grape berries and subsequent counting — applies mask models adapted for tiny instances.	Mask-based approaches adapted for small instance scales improve per-berry detection and counting performance on grape datasets.	Small-instance segmentation under heavy occlusion still sees missed detections; needs multi-scale features, super-resolution, or density-map fusion for best results.
2024	FruitNeRF / Neural rendering + multi-view reconstruction (emerging work) [4]	(Representative neural-rendering works cited in survey)	Use neural radiance fields / multi-view synthesis to reconstruct fruits in 3D from images; segmentation/counting performed on reconstructed geometry to mitigate occlusion.	When many views are available, 3D reconstruction reduces missed/occluded instances and can yield very accurate counts; promising proof-of-concepts reported.	High capture and compute costs; impractical for large-scale orchards without efficient capture pipelines. Generalization and runtime for field deployment are open challenges.
2025 (arXiv)	Benchmarking robustness of instance segmentation models [7]	Y. Dalva*, H. Pehlivan*, S.F. Altındiş, A. Dundar	Study/benchmark focusing on robustness of instance segmentation models (domain shifts, occlusion,	Finds significant performance drops under corruptions/domain shift; highlights need for robustness-	Agricultural datasets rarely included in robustness benchmarks; creating domain-

			corruptions) — useful for evaluating agricultural robustness.	focused evaluation and defenses.	specific robustness tests (lighting, foliage occlusion, seasonal shifts) is needed.

III. **STANCE SEGMENTATION APPROACH TAXONOMY (SHORT)**

1. two-stage (based on region proposals); for example, Mask R-CNN [1]: RoI proposals → per-RoI mask branch. Good baseline for agricultural use; generally good accuracy; moderate speed.
2. Prototype masks and per-instance coefficients are produced by one-stage/real-time prototype techniques, such as YOLACT and YOLACT++, which are optimized for speed but may be less effective in situations with significant occlusion.
3. Modern architectures (SOLO, SOLOv2, CenterMask, etc.) that integrate coordinate-aware heads and segmentation designs for efficiency and enhanced mask quality are known as single-stage segmentation/segformer variants.

IV. **REPRESENTATIVE ALGORITHMS: CONCISE EXPLANATIONS AND PERTINENCE TO OCCLUSION**

3.1 Two-stage Mask R-CNN	Mechanism: adds a RoI "mask branch" to the Faster R-CNN to predict a binary mask for each detected instance. strong mask quality and a high degree of adoption as a starting point.
	High mask fidelity, stable training, and extensibility (backbones, FPN, GN, attention) are among its strengths. If RoIs are accurate, they are good at separating adjacent objects.
	Weaknesses: moderate runtime; unless customized, RoI proposals may fail in extreme cluster density or occlusion.
	Agricultural evidence: Mask R-CNN has been used extensively for grapes, oranges, apples, and sweet peppers with good outcomes in numerous field studies.
3.2 YOLACT/YOLACT++ (real-time, prototype-based)	Mechanism: generate linear combination coefficients and global prototype masks for each detected instance very quickly (>30 fps on GPU).
	Strengths: easier training and inference pipeline; real-time inference appropriate for drones and robots.
	Weaknesses: Overlapping fruits in dense clusters may be under-segmented by linear prototypes, which may find it difficult to depict delicate boundaries or highly obscured shapes. To detect fruit bunches, agricultural variants (like MTA-YOLACT) modify prototypes and multitask heads.
3.3 Single-stage segmentation, including SOLO and SOLOv2	Mechanism: avoids the RoI stage by spatially classifying pixels into instance "locations" and predicting masks for each location. Better at small object instances and frequently offers a good trade-off between accuracy and speed. (Many orchard studies adapt these concepts, and representative modern work demonstrates improvements in speed/accuracy.)
3.4 Occlusion-aware and Amodal models	Mechanism: either use priors (shape, spherical assumptions for many fruits) to recover occluded instances or predict the entire object silhouette, including occluded parts. particularly advantageous for on-tree fruits that are severely obscured.
3.5 Neural, multi-view, and three-dimensional rendering techniques	Mechanism: create a 3D-aware segmentation by combining several views (or synthesizing views). Fruits can be counted by reconstructing them in three dimensions, which reduces missed detections due to occlusion, according to recent work (e.g., FruitNeRF). These are promising, but they require a lot of data and computation.

V. THE LITERATURE'S EXPLICIT TREATMENT OF OCCLUSION

Higher resolution features combined with mask refinement: FPNs and stronger backbones (ResNeXt, VoVNetV2) improve mask fidelity in near-occlusion areas.

Amodal segmentation and shape priors: predicting complete shapes using explicit occlusion modeling or fruit shape priors (roughly spherical). Research on pomegranates and apples grown on trees demonstrates significant improvements in counting when amodal [2] masks are applied.

Temporal/multi-view fusion reduces missed fruits and double-counts by tracking instances across frames (video) to combine partial observations from various angles. A number of fruit-counting projects take advantage of video continuity by using detection + tracking pipelines.

Multimodal sensors: employ RGB+NIR/depth to identify partially hidden areas and distinguish fruit from foliage; multimodal instance segmentation techniques report better recall under occlusion.

VI. EVALUATION METRICS AND DATASETS (OFTEN USED)

Custom field datasets (apples, oranges, grapes, tomatoes) are used in numerous agricultural publications. Core models are evaluated using public benchmarks for general instance segmentation (COCO, Cityscapes); each paper typically includes domain-specific datasets (e.g., grape berry datasets, pomegranate datasets). Grapes (instance segmentation per berry) and apples/pomegranates (on-tree segmentation) are examples of agricultural dataset papers. [8]

VII. COMPARATIVE SUMMARY

Table 2: Comparative Summary

Method family	Typical models	Mask quality	Speed	Occlusion robustness	Suitability for fruit counting
Two-stage	Mask R-CNN (+ variants)	High	Medium	Good (with amodal / shape priors)	Strong baseline; widely used in apple/orange/pepper studies.
One-stage prototype	YOLACT / YOLACT++	Medium	High (real-time)	Moderate struggles with heavy overlap unless adapted.	-
SOLO/ SOLOv2, CenterMask	Single-stage segmentation	Medium–High	Medium–High	Better for small instances; competitive in dense clusters	Useful for small clustered fruits (berries)
Amodal/ occlusion-aware	Amodal Mask R-CNN variants	High (predicts occluded shape)	Medium	Strong explicitly models occlusion.	-
3D / Multi-view (NeRF)	FruitNeRF / multiview fusion	Very High (if many views)	Low (compute heavy)	Excellent resolves occlusion via geometry.	-

VIII. SELECTED CASE STUDIES THAT ARE REPRESENTATIVE OF AGRICULTURE

Orange detection and Mask R-CNN [1]: Combining RGB+HSV channels increased F1 in field tests; Mask R-CNN [1] generated reliable masks for orange detection.

Grape berry segmentation and Mask R-CNN enhancements: Research suggests Mask R-CNN variations specifically designed for small, closely spaced berries in order to minimize missed detections under occlusion. [8]

Amodal on-tree apple segmentation: Counting accuracy on highly occluded on-tree apples was enhanced by amodal [2] techniques that predict occluded object shapes.

Fruit bunch real-time YOLACT variants: MTA-YOLACT and other multitask adaptations aim to detect clustered cherry tomatoes and fruit bunches for robotic harvesting in real-time. [3]

For mask fidelity, start with the Mask R-CNN baseline; when occlusion is frequent, add shape priors or amodal [2] heads.

Use task-specific adaptations (multitask heads, cluster NMS) and assess YOLACT++ or optimized SOLOv2 variants if real-time (robot or drone) is needed. [3]

Try instance-segmentation techniques designed for small object masks for dense small fruits (berries), and take into account multi-scale features and super-resolution inputs.

For challenging occlusion, take into account multimodal sensors or multi-view capture; depth/NIR or view synthesis can yield notable improvements but complicate acquisition.

IX. UNRESOLVED ISSUES AND POTENTIAL PATHS

Amodal generalization: It is still difficult to generalize to numerous fruit species (different shapes) because existing amodal [2] methods frequently require object-specific priors.

Domain shift and field robustness: rigorous robustness benchmarks and adaptation techniques are required because models trained on small orchard datasets frequently fail across farms, seasons, or camera setups. [7]

Data efficiency and active learning: weak supervision (box→mask, synthetic augmentation) and active learning are promising; labeling masks is costly.

3D and neural rendering techniques: incorporating multi-view reconstructions and NeRF for accurate counting under occlusion appears promising, but it requires workable capture pipelines.

Edge deployment and energy constraints: hardware co-design and model compression are required for real-time instance segmentation on drones and harvesters.

X. CONCLUSION

A thorough analysis of instance segmentation methods used for fruit counting under occlusion has been provided by this survey. This problem is still crucial to obtaining accurate yield estimation and automated orchard monitoring. Due to their high mask accuracy, classical two-stage frameworks—especially Mask R-CNN and its variations—continue to offer strong baseline performance; however, their efficacy decreases when there is significant leaf coverage or densely clustered fruits. Although one-stage and real-time architectures provide the computational efficiency required for systems that can be deployed in the field, they typically show lower boundary precision in occluded scenarios. Though they present difficulties with generalization, annotation cost, and computational overhead, recent developments in amodal [2] segmentation and multi-view or 3D reconstruction techniques show great promise for resolving occlusion by predicting entire fruit geometries.

Inadequate generalization across fruit species and environmental conditions, a lack of standardized robustness benchmarks specific to agricultural contexts, and the scarcity of occlusion-rich annotated datasets are just a few of the ongoing gaps found in the literature. Future research on domain adaptation, amodal [2] generalization across various fruit types, multimodal sensor integration, and the creation of lightweight models appropriate for real-time deployment on edge devices is necessary, as these gaps highlight. In order to develop dependable and scalable fruit-counting systems that can function well in intricate and naturally changing orchard environments, it will be essential to address these research directions. [7]

REFERENCES

[1] Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick, Mask R-CNN, In CVPR, 2017.

[2] J. Gené-Mola, M. Ferrer-Ferrer, E. Gregorio, P.M. Blok, J. Hemming, J.R. Morros, J.R. Rosell-Polo, V. Vilaplana, J. RuizHidalgo, Looking behind occlusions: a study on amodal segmentation for robust on-tree apple fruit size estimation, Comput. Electron Agric. 209 (2023) 107854, doi:10.1016/j.compag.2023.107854.

[3] M. Ferrer-Ferrer, J. Ruiz-Hidalgo, E. Gregorio, V. Vilaplana, J.R. Morros, J. Gené-Mola, Simultaneous fruit detection and size estimation using multitask deep neural networks, Biosyst. Eng. 233 (2023) 63–75.

[4] J.C. Miranda, J. Gené-Mola, M. Zude-Sasse, N. Tsoulias, A. Escolà, J. Arnó, J.R. Rosell-Polo, R. Sanz-Cortiella, J.A. Martínez-Casasnovas, E. Gregorio, Fruit sizing using AI: a review of methods and challenges, Postharvest Biol. Technol. 206 (2023) 112587, doi:10.1016/j.postharvbio.2023.112587.

[5] A. Koirala, K.B. Walsh, Z. Wang, C. McCarthy, Deep learning – method overview and review of use for fruit detection and yield estimation, Comput. Electron Agric. 162 (2019) 219–234, doi:10.1016/j.compag.2019.04.017.

[6] J. Gené-Mola, R. Sanz-Cortiella, J.R. Rosell-Polo, J.R. Morros, J. Ruiz-Hidalgo, V. Vilaplana, E. Gregorio, Fuji-SfM dataset: a collection of annotated images and point clouds for Fuji apple detection and location using structure from-motion photogrammetry, Data Br 30 (2020), doi:10.1016/j.dib.2020.105591.

[7] Y. Dalva∗, H. Pehlivan∗, S. Fahri Altındiş and A. Dundar, Benchmarking the Robustness of Instance Segmentation Models, arXiv preprint arXiv:2303.03471, 28 Mar 2025

[8] Yanmin Chen, Xiu Li, Mei Jia, Jiuliang Li, Tianyang Hu and Jun Luo, Instance Segmentation and Number Counting of Grape Berry Images Based on Deep Learning, 2023.

[9] Bolya, Daniel; Zhou, Chong; Xiao, Fanyi; Lee, Yong Jae., YOLACT: Real-Time Instance Segmentation, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9157–9166.