Journal of Applied Computational Intelligence

eISSN: 3136-6937

Vol. 4, No. 3, July, 2026

# Density Guided Query Allocation in RT-DETR for Orchard Blossom Density Mapping

Tadas Mikalauskas[1,*], Arnas Butkus[1], Ramūnas Šležas[1]

[1] *Faculty of Technologies, Šiauliai University, Šiauliai, LT-76351, Lithuania*

Received 21 February 2026; Revised 25 June 2026; Accepted 30 June 2026; Published 03 July 2026

**Abstract**

Orchard blossom mapping requires the localization of individual flowers and preservation of spatial variation in densely flowering regions. A fixed decoder query budget can allocate excessive capacity to background while leaving crowded regions underrepresented. This paper proposes a density-guided query allocation framework within a real-time detection transformer for blossom detection and image-plane density mapping. A lightweight density branch estimates regional blossom abundance, while a budget-conserving allocation rule distributes spatial queries and retains a global exploration reserve. Boundary-aware selection and count-consistent supervision connect object detection with normalized spatial maps. Evaluation uses orchard-block partitions to examine detection and mapping errors and allocation stability under varying density, illumination, and visibility conditions. The evaluation set contains 2,400 images, including a held-out partition of 480 images. The complete model achieves an average precision of 0.516 and a mean absolute count error of 8.6 blossoms per image. Normalized density map error decreases from 0.218 to 0.151, while inference latency increases from 18.4 to 20.1 ms. These metrics jointly characterize detection accuracy, spatial mapping error, and computational cost. The framework allocates finite decoder capacity according to regional abundance while incorporating safeguards against density estimation errors. Evaluation across operating conditions connects local query coverage with counting accuracy and spatial map agreement.

*Keywords:* Orchard Blossom Detection, Density Guided Queries, Real Time Detection Transformer, Spatial Density Mapping, Query Budget Allocation

## 1. Introduction

Computer vision and deep learning support orchard phenotyping by linking blossom and fruit observations with apple growth monitoring and production assessment [1]. UAV-acquired RGB images and density-based cluster counting provide a route to tree-level apple flower quantification [2]. Apple flower phenotyping can distinguish blooming stages using visual indicators such as anther color [3]. Dual-view flower counting combines complementary observations from opposite sides of a tree to address the limited visibility of a single image [4]. Flower-cluster identification and discrimination between central and peripheral flowers provide localized units for targeted orchard management [5]. The selection of these measurements will have a different effect depending on whether there is a dense cluster of overlapping flowers in one view or multiple spread-out groups in another. A density map can be employed to separate the local concentrations of different areas and show the support region for each estimate visually. The current problem is to find distinguishable open blossoms and their locations in the image. Each map entry associates an image cell with its visibility mask and counting rule. These associations connect localized blossom instances with canopy-scale flowering patterns.

* Corresponding author: mika.tada@su.lt

http://dx.doi.org/10.68406/jaci.2026.vol4iss3no1:1-16

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

The additional requirement for calibrated information in physical surface density is camera geometry and canopy structure, as the area of a pixel changes with depth and viewing angle.

A detection system provides an interpretable path from photos to flower locations; however, small targets and irregular crowding have different requirements for the detector at various positions in the image. Apple flower detection benefits from multiscale attention and feature fusion under cluttered orchard backgrounds [6]. Apple blossom recognition remains sensitive to illumination changes and complex orchard backgrounds [7]. Bhattarai et al. [8] developed AgRegNet to estimate blossom density, centroid locations, and image counts within a common regression-based framework. DQ-DETR uses count predictions and density maps to adjust the number and positional information of object queries for tiny-object detection [9]. Agricultural adaptations of RT-DETR investigate lightweight feature processing to balance detection accuracy and computational efficiency [10]. A fixed query budget can help control the size of memory and computations. Blossom abundance can vary substantially across image regions. A canopy image can contain extensive background beside compact blossom clusters. Increasing the entire query budget will raise the capacity of all areas, even those with a small demand. The design first estimates regional representation requirements, then distributes a limited query budget while preserving exploration outside the predicted clusters.

This manuscript develops that design as a density guided allocation mechanism for RT-DETR and connects the resulting detections to a spatial density map. The proposed mechanism uses a low-resolution abundance estimate to allocate regional queries, retains a fixed global reserve, and spreads selected queries within crowded cells to avoid spending multiple hypotheses on the same local feature maximum. The density branch is supervised from annotated blossom centers, while a count consistency term discourages disagreement between regional abundance and matched detections. The central research question is whether regional redistribution can reduce dense region omissions while preserving sparse region recall and computational efficiency. A second question concerns the output representation: a count-preserving image-plane density map should communicate flowering intensity and visibility constraints. A controlled experiment was designed to isolate the allocation factor from the backbone capacity and target resolution. Numerical evaluation studies the effect of allocation settings on the spatial structure of prediction errors. It specifies the proposed mechanism, diagnostic tests and evidence requirements for the next implementation and independent verification. Initialization coverage and regional residuals are also recorded in the evaluation to distinguish between errors caused by spatial routing and those resulting from decoding or map reconstruction. The contribution concerns spatial redistribution under a conserved decoder budget, together with explicit coverage and exploration safeguards. Its scope is the coupling of these constraints to count-preserving orchard mapping; density-conditioned queries themselves are an established design direction.

The following sections will present the connection between spatially uneven blossom abundance, the allocation of limited detector capacity, and the interpretation of the resulting density maps in the research context, model construction, and experimental results. Section 2 studies blossom detection, density representation and query-based localization, and introduces the reasons for their joint design. Section 3 introduces the image support, the objective of regional density, a budget-saving allocator and safeguards against spatial selection bias. Section 4 introduces the integration of the allocator with the detection transformer, the training objective, and the conversion of accepted detections into count-preserving maps. Section 5 introduces the experimental plan and conducts comparative tests of the model, component ablation, spatial error analysis and deployment sensitivity. Its figures distinguish aggregate performance changes from density-dependent behavior, while the tables specify the conditions required to interpret the comparisons. Section 6 presents the methodological conclusions, limitations and directions for field validation. The spatial support, confidence threshold and query budget of the comparison models are the same. The design of the experimental records is to distinguish between model output, spatial reconstruction and management interpretation. This organization allows the allocation mechanism to be assessed before the experimental evidence and practical implications are examined. This sequence separates the allocation rule, its integration into the detector, and the evidence used to assess its consequences. Readers can therefore distinguish the method specification from the conditions under which the reported comparison is interpreted.

## 2. Related Work

### 2.1. Blossom localization under canopy variability

Agricultural object detection provides instance locations within canopy imagery. Blossom annotation protocols can explicitly distinguish open flowers from unopened buds and provide corresponding bounding-box labels [11]. This distinction affects both localization labels and count targets, making annotation consistency part of the measurement problem. Box annotations support detection training, whereas center annotations can provide greater stability when petal boundaries are irregular. Box-based evaluation may penalize a plausible blossom center because of ambiguity in the enclosing boundary. Center-based evaluation, in turn, requires sufficient separation between neighboring flowers. The evaluation should retain both perspectives and specify which representation determines the final map.

Combining local feature extraction with transformer-based global context provides complementary spatial representations for object detection [12]. Bright petals can resemble specular leaf highlights, pale branches, support wires, and sky visible through the canopy. Context is used to differentiate these structures, but several pooling operations may lose the separation of adjacent blossoms. A representation suitable for a sparse branch may have different behavior in a dense cluster of flowers. To show this difference, the density strata need to be compared, especially when there are many easy background images in the dataset. An aggregate detection score can conceal substantial undercounting in regions with high flowering intensity.

Geospatial modeling research identifies spatial dependence and the separation of training and validation locations as important considerations in assessing generalization [13]. Adjacent images can share branches, illumination, camera positions, and nearly identical blossom clusters. Random image partitioning can consequently favor recognition of familiar content over transfer to a new block. The proposed evaluation uses orchard blocks as the units of separation and individual images as the units of measurement. The visibility mask is retained independently of the predicted blossom map. This design supports density-stratified error analysis over explicitly defined visible canopy regions.

### 2.2. Density representations and count preservation

Density estimation is the abundance of objects in a particular area and is especially useful when individual boundaries are difficult to distinguish. Density-based counting methods use multibranch representations to address scale variation and background interference [14]. Kernel width controls the trade-off between local detail and spatial smoothing. A broad kernel smooths local abundance but can blur narrow blossom clusters into adjacent regions. A narrow kernel preserves cluster detail while increasing sensitivity to annotation precision and image alignment. Kernel mapping also deals with the boundary of the mask, and truncation without renormalization loses count mass in the vicinity of the visible area.

Joint density estimation and object detection can combine instance localization with density-map learning to improve dense counting representations [15]. The two branches can share correlated errors, so consistency requires interpretation alongside independent annotation-based accuracy. Such agreement has diagnostic and regularizing value: substantial predicted abundance in a region with few matched detections identifies a potential coverage discrepancy. Conversely, numerous detections in a region with low predicted abundance can indicate abundance estimation error or repeated object hypotheses. Training therefore retains direct supervision for each branch and uses consistency to connect their representations.

The density map requires an explicit denominator. Blossoms per image cell, blossoms per valid pixel area, and blossoms per square meter describe different measurement quantities. The present design provides image-cell counts and valid-pixel-area densities; physical area density additionally requires camera geometry and canopy surface information. Spatial comparison within an image plane depends on preserving the color scale and support mask. This representation distinguishes an empty visible region from a region excluded by the visibility mask. That distinction supports map agreement assessment and the interpretation of regions with limited visual coverage. Comparisons of regional counts require comparable support areas. When visible areas differ, the legend must distinguish counts from area-normalized density, since a small visible region can have a high density while containing few blossoms.

### 2.3. Query allocation in set prediction detectors

Query-based detectors encode a finite set of candidate object hypotheses. Yuan et al. [16] proposed adaptive token selection with dual-teacher supervision to improve the computational efficiency of detection transformers. Confidence-based selection can concentrate candidates around visually prominent blossoms, leaving adjacent, partly occluded flowers underrepresented. A spatial abundance prior provides a complementary signal by estimating the representational capacity required in each region before individual objects are localized.

Query-routing methods allocate decoder processing selectively according to the information carried by individual object queries [17]. A density-guided allocator requires an explicit budget, reported routing overhead, and safeguards against abundance estimation errors. Increasing the total number of queries changes computational capacity and therefore requires separate evaluation from spatial redistribution. Comparing annotation-based routing with predicted-density routing can identify the effect of routing-signal accuracy. The primary inference pathway uses predicted density, while annotation-based routing serves as a diagnostic reference condition. A fixed-budget comparison tests where hypotheses are initialized. A variable-budget comparison additionally changes how many hypotheses can be decoded, so its gains combine allocation and capacity effects.

Studies of detector calibration distinguish confidence reliability from detection accuracy across object detection architectures [18]. Glare, motion blur, and unfamiliar canopy appearance can produce erroneous abundance estimates with high confidence. Regional query reservation and global exploration provide complementary safeguards whose effectiveness depends on the visual evidence available in the image. The central issue concerns the coupling between a

noisy abundance estimate, a finite query set, and the spatial measurement delivered to the user. The proposed architecture makes this coupling explicit through capacity-matched comparisons, stress tests, and independent detection and mapping diagnostics.

## 3. Density Guided Query Allocation

### 3.1. Image support and regional abundance targets

The input is a color canopy image accompanied by a binary valid region mask. Validity indicates that an image location belongs to the visible canopy support selected for measurement. Individual blossoms within valid support can still be partially occluded. Annotation records contain an enclosing box and one center for each distinguishable open blossom. Buds and senescent flowers require separate labels or explicit exclusion because their inclusion changes the meaning of the density target. The image is partitioned into a regular grid in the encoder coordinate system, and each annotation center contributes to one grid cell. Boundary centers follow a consistent half-open indexing convention, ensuring a unique regional assignment. The measurement unit is an individually distinguishable open blossom in the selected view. Fully hidden flowers lie outside this image-based target, and a regional count therefore represents observable flowering rather than total tree abundance.

The target density is formed by nonnegative kernels centered on the annotated flowers. Each kernel is normalized over valid image support, including when its tail intersects an image or mask boundary. This construction preserves count mass while allowing local smoothing. A fixed kernel width in the resized image defines the baseline, while sensitivity analysis examines alternative widths under unchanged detection labels. The target density follows the equation below, with normalization applied separately to each blossom kernel:

$$D^*(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{K_\sigma(\boldsymbol{x} - \boldsymbol{c}_i) M(\boldsymbol{x})}{Z_i}. \tag{1}$$

Here, $\mathbf{x}$ denotes image position, $\mathbf{c}_i$ is the center of the i-th annotated blossom, M is the valid support mask, and $Z_i$ is the integral of the corresponding masked kernel. Consequently, the target integrates to the annotated blossom count whenever each center lies within nonempty valid support. Regional abundance is obtained by integrating this target over each grid cell. A softplus output constrains the density branch to predict nonnegative abundance. Its spatial resolution is intentionally lower than that of the reconstructed map: the branch estimates regional demand for queries, while the decoder separates individual instances. Figure 1 shows how canopy features and valid support guide abundance prediction, integer allocation, and candidate selection. Regional candidates and exploration candidates merge before decoding.

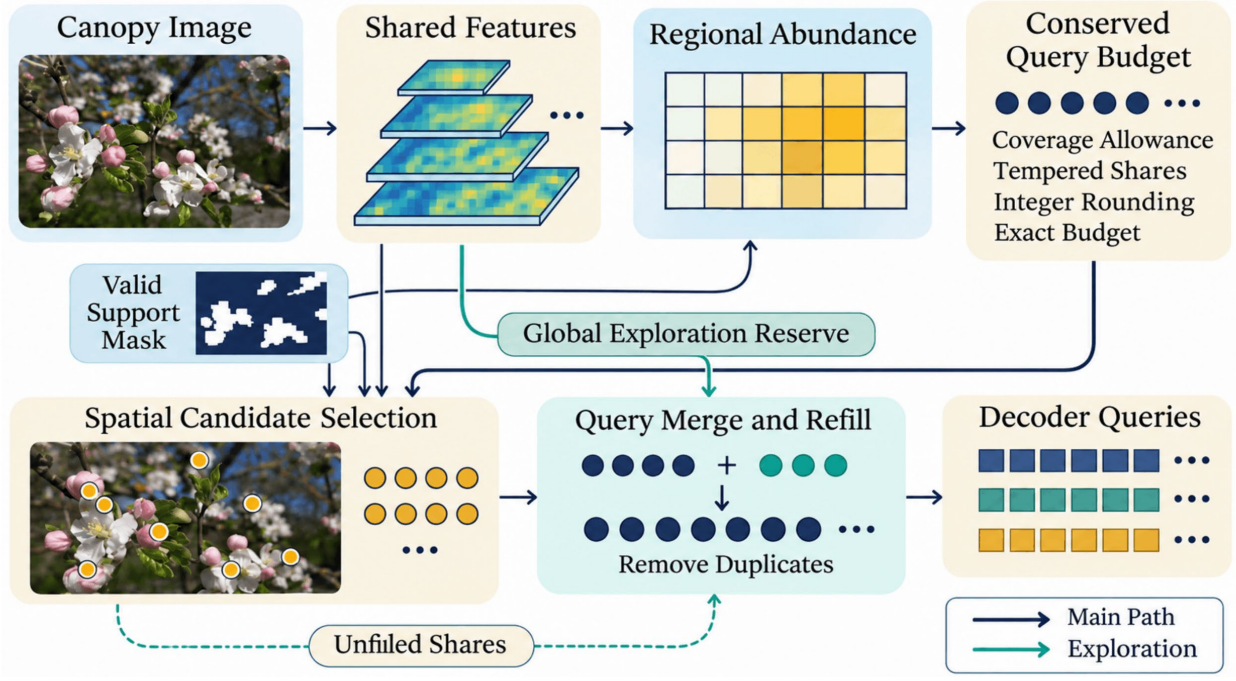



**Figure 1. Density guided allocation pipeline**

### 3.2. Conserved query budgeting and spatial selection

The allocation rule separates spatial coverage from density concentration. Each valid grid cell receives a minimum coverage allowance when the budget permits, while the remaining regional budget follows a tempered abundance distribution. A positive abundance floor preserves allocation opportunities in cells with weak predictions. Tempering moderates the influence of isolated abundance peaks and reduces sensitivity to unusually large branch outputs. The global exploration reserve is separated before regional allocation and retains a fixed share of the total budget. Continuous regional shares are computed using the following equation before integer rounding: Tempering compresses the contrast between high- and low-abundance cells before normalization. This gives neighboring weak-response cells a larger relative share than proportional allocation would provide, while retaining preferential coverage of dense regions.

$$q_g = q_0 + (Q_r - Gq_0)\frac{(d_g + \varepsilon)^{\alpha}}{\sum_{h=1}^{G}(d_h + \varepsilon)^{\alpha}}. \quad (2)$$

In this equation, $Q_r$ is the regional query budget, G is the number of valid cells, $q_0$ is the minimum regional allowance, and $d_g$ is predicted abundance in cell g. The exponent α controls allocation concentration, with values between zero and one providing tempering, while ε supplies a positive abundance floor. The configuration requires $Q_r \geq Gq_0$. When this condition is violated, the implementation coarsens the grid or disables the minimum allowance through a predefined fallback. Largest remainder rounding converts the continuous shares into integers. The allocator first takes their floors and then assigns the remaining slots to the largest fractional remainders, preserving the regional budget exactly. Cell indices resolve ties deterministically. The rounding and fallback choices are part of the inference specification. Recording which fallback is activated is necessary when comparing images whose valid masks produce different numbers of eligible cells.

Candidate selection operates within each allocated cell. Feature locations are ranked by a combined objectness and abundance compatibility score, then selected under a minimum spatial separation in encoder coordinates. This separation reduces repeated initialization around the same high-contrast petal. The constraint is relaxed progressively when a cell supplies fewer candidates than its allocated share, and any remaining share returns to the global candidate pool. Selection is restricted to valid support. The algorithm records requested, accepted, and returned shares, distinguishing the nominal allocation from the number of spatially distinct candidates supplied to the decoder. The global reserve uses image-wide objectness ranking with a separate spacing constraint. Regional and global candidates are merged before positional embeddings are attached, and duplicate feature locations are removed and refilled deterministically to preserve the final query count. Budget conservation refers to the final number of decoder hypotheses after refill. Requested regional shares describe the preferred distribution; returned slots can change that distribution while the final query count remains fixed.

### 3.3. Safeguards against routing errors

The abundance estimate is an imperfect routing signal, particularly before the density branch learns to distinguish flowers from reflective background. Training therefore begins with uniform regional allocation and gradually increases the contribution of predicted abundance. Ground-truth routing is reserved for a separately identified diagnostic condition, while the main evaluation follows the information pathway available during deployment. This scheduled transition maintains exposure to difficult flowers during early optimization and reduces the influence of unstable abundance predictions on decoder initialization. Annotation-based routing tests the consequence of a more accurate allocation signal under otherwise matched conditions. It represents a diagnostic reference; deployment uses the abundance branch and the same reserve and refill rules.

Spatial coverage and global exploration address different failure modes. Coverage maintains a regional allocation opportunity, whereas exploration can select strong candidates across cell boundaries. Both safeguards are tuned exclusively on the validation partition. The reserve fraction and tempering exponent are selected once and fixed across the comparative evaluation. Allocation entropy measures how strongly the regional budget concentrates within a small subset of cells. Its interpretation depends on scene content because compact blossom clusters can justify concentrated allocation. Missed annotations, candidate diversity, and density residuals provide the additional evidence needed to distinguish appropriate concentration from routing failure.

Mask handling provides a further safeguard. Normalization and candidate selection operate within valid regions, with partially visible cells weighted by their support fractions. When all regional predictions fall below the abundance floor, allocation approaches a uniform distribution over valid cell. An image with empty valid support produces an invalid-support flag and an empty measurement. The decoder retains a fixed maximum query budget, so regions containing more flowers than available distinct hypotheses can experience saturation. Density-stratified recall exposes this condition, while candidate coverage records identify whether missed flowers received an initialization opportunity. The benefit of routing consequently depends on local abundance, candidate diversity, and the total query budget available for the frame. An invalid-support output indicates that the image provides no eligible measurement area. A valid region

with zero accepted detections instead reports an observable area with an estimated count of zero; these states require different downstream handling.

## 4. Detection and Spatial Mapping Model

### 4.1. Architecture integration and computational pathway

The proposed model retains the feature extraction, multiscale encoding, and set prediction structure of the selected RT-DETR implementation. Reproducibility requires recording the implementation version, backbone, feature scales, and decoder depth. The abundance branch is inserted after shared feature fusion and before query initialization. A shallow convolutional projection produces regional abundance estimates. Candidate indices gather the corresponding encoder features, which receive positional embeddings before entering the decoder. Integer allocation is a discrete routing operation. The abundance branch receives direct supervised gradients, while shared features receive gradients from both density estimation and detection objectives. Gradients propagate through the gathered feature values, while the discrete ranking and integer indices define the selected pathway. This separation requires direct abundance supervision to train the routing signal.

The total query budget remains constant across the comparative configurations. The baseline uses its original initialization procedure, whereas the proposed model changes the spatial distribution of the same number of queries. A capacity control adds the abundance branch while retaining fixed initialization, separating the contribution of auxiliary supervision from that of redistribution. Inference timing includes candidate selection, integer allocation, and decoder execution. Comparative measurements use identical batch sizes, input resolutions, arithmetic precision, and synchronization procedures. Hardware-independent operation counts complement device-specific latency measurements by describing computational demand independently of execution conditions.

Figure 2 shows the shared backbone and encoder, the abundance branch, the regional and global selection paths, the common decoder, and density reconstruction. Training connections carry center-density targets, instance-matching targets, and consistency supervision. Inference connections carry predicted abundance and decoded detections. The support mask enters both allocation and reconstruction, establishing a common spatial domain for routing and measurement. A parallel output path carries support validity and confidence diagnostics, allowing downstream software to identify accepted maps and images requiring additional visibility checks.

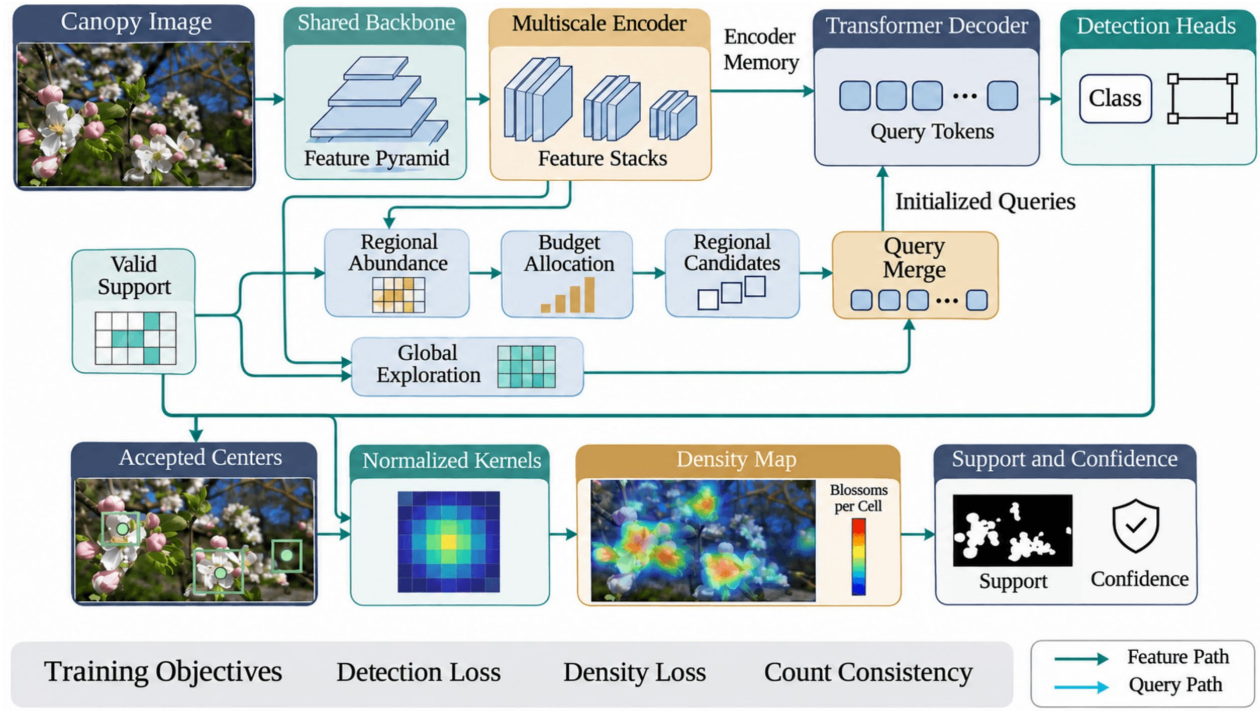



**Figure 2. Integrated detection and mapping architecture**

### 4.2. Joint supervision and count consistency

Training retains the detector's classification and box regression objectives while adding direct density supervision and a count consistency penalty. Validation data determine their weights, and the detection terms remain identical between the capacity control and allocation model. These contributions are combined in the training objective equation:

$$\mathcal{L} = \mathcal{L}_{det} + \lambda_d \mathcal{L}_{density} + \lambda_c \mathcal{L}_{consistency}. \quad (3)$$

The coefficients $\lambda_d$ and $\lambda_c$ control the influence of density supervision and cross-branch agreement. Their weighting preserves independent supervision for both representations because excessively strong agreement can reinforce correlated errors. The density loss averages absolute regional abundance residuals over valid cells, with support weighting available for partially visible cells. One-to-one instance matching assigns each annotation to a single prediction. During training, the blossom probabilities of matched predictions form a regional soft count, using predicted centers to determine cell membership. This matched count limits the influence of background probability distributed across unmatched queries. Regional agreement is then encouraged through the consistency equation: Cross-branch agreement is a training constraint rather than independent evidence of accuracy. Both branches remain anchored to annotations because a shared representation can produce mutually consistent underestimation.

$$\mathcal{L}_{consistency} = \frac{1}{G} \sum_{g=1}^{G} \frac{|d_g - s_g|}{1 + n_g}. \quad (4)$$

The quantities $s_g$ and $n_g$ represent the matched soft count and annotated count in cell $g$, respectively. The denominator moderates the contribution of heavily populated regions. With discrete cell membership, the consistency gradient updates matched confidence and predicted abundance, while box supervision provides the localization gradient. These objectives therefore serve distinct roles in confidence agreement and geometric refinement. A detached decoder count can stabilize the abundance branch during warmup, followed by confidence supervision after validation confirms stable optimization. The training record specifies the warmup schedule and consistency configuration. Matched probabilities enter the loss directly, while inference threshold selection remains a separate validation decision. Controlled ablations vary loss weights, warmup duration, and grid resolution individually to preserve interpretability. A center crossing a grid boundary changes regional membership discretely. The consistency term therefore complements the continuous box loss; it does not provide a smooth localization penalty across cell boundaries.

### 4.3. Count preserving reconstruction and uncertainty

Accepted detection centers are converted into a continuous image-plane map through normalized kernels. A single validation-selected confidence threshold defines the accepted set and remains fixed across abundance strata. Kernel width controls display smoothness and is evaluated separately from detection accuracy. Applying per-instance normalization gives the reconstruction equation:

$$\widehat{D}(\boldsymbol{x}) = \sum_{j=1}^{N_a} \frac{K_\tau(\boldsymbol{x} - \hat{\mathbf{c}}_j) M(\boldsymbol{x})}{\hat{Z}_j}. \quad (5)$$

Here, $N_a$ is the number of accepted detections, $\hat{\mathbf{c}}_j$ is an accepted center, and $\hat{Z}_j$ is the valid-support integral of its reconstruction kernel. Normalization preserves each detection's contribution at image and mask boundaries, so the reconstructed map integrates to the accepted count. Confidence scores remain a separate diagnostic field, allowing the map to represent accepted instances consistently. Regional values can be displayed as counts per cell or divided by valid support area, with the legend specifying the chosen quantity. Cells with limited visible support receive explicit flags because area normalization can amplify their density estimates. Interpreting the output as physical surface density additionally requires calibrated canopy geometry. Mass preservation is conditional on the accepted detection set. A false positive adds one unit and a missed blossom removes one unit, so kernel normalization preserves the detector count without correcting detection errors.

Prediction and reference maps are aggregated onto the same evaluation grid before spatial error is calculated. The normalized absolute error equation expresses regional disagreement relative to the annotated image total:

$$E_N = \frac{\sum_{g=1}^{G} |\hat{m}_g - m_g|}{max\left(1, \sum_{g=1}^{G} m_g\right)}. \quad (6)$$

In this equation, $\hat{m}_g$ and $m_g$ denote predicted and reference regional counts. The denominator floor supports evaluation of empty images, which are also reported separately through their false-positive counts. Count error and map error describe different properties: false positives and misses can offset one another in the image total while producing substantial regional disagreement. Uncertainty analysis therefore combines block-level resampling, residual stratification, and kernel-width sensitivity. These procedures characterize sampling variability and reconstruction sensitivity, while a dedicated reliability assessment examines detection confidence. Each delivered map retains its detections, support mask, confidence threshold, and kernel setting, enabling regional estimates to be traced to their contributing instances and visibility conditions. For example, an excess in one cell and an equal deficit in another leave

the image count unchanged while increasing regional disagreement. Reporting both metrics distinguishes total abundance agreement from spatial fidelity.

## 5. Experimental Evaluation and Spatial Error Analysis

### 5.1. Evaluation design and comparative performance

The evaluation ledger defines the dataset partition, comparative model configurations, and performance metrics. It contains 2,400 records from twelve orchard blocks, partitioned into 1,600 training records from eight blocks, 320 validation records from two blocks, and 480 test records from two blocks. Research on spatially correlated samples shows that separating related observations during validation can reduce leakage-induced optimism [19]. The test partition contains low, medium, and high abundance strata with 160 records each. Table 1 specifies the image resolution, query budget, routing settings, and proposed training schedule. Surveys of object detection compare recognition and localization performance across detection architectures [20]. In this evaluation, average precision is calculated across overlap thresholds from 0.50 to 0.95. Count error measures the absolute discrepancy between accepted instances and annotated blossoms, while map error evaluates regional abundance on a common grid. Dense-object counting research evaluates density-map quality alongside count accuracy, reflecting the different information retained by the two outputs [21]. The reconstruction grid remains fixed across model variants, isolating performance differences from changes in map resolution. The same evaluation settings apply to each model variant, supporting paired inspection of counting and spatial errors. Average precision summarizes the precision-recall relationship across score thresholds at the stated overlap thresholds. Count and reconstructed-map errors instead use the accepted detection set at the selected operating threshold; these quantities describe different evaluation decisions.

**Table 1. Dataset partition and evaluation settings**

| Item | Evaluation setting |
|---|---|
| Record scope | Image level detection and mapping evaluation |
| Block split | 8 training / 2 validation / 2 test |
| Image record split | 1600 / 320 / 480 |
| Test strata | Low / medium / high: 160 records each |
| Input and grid | 640 × 640 pixels; 8 × 8 allocation grid |
| Query budget | 300 total; global reserve 0.15 |
| Tempering and minimum | Exponent 0.75; 1 query per valid cell |
| Supervision weights | Density 1.0; consistency 0.2; validation candidates |
| Training protocol | Proposed 120 epochs; 10 epoch routing warmups |
| Runtime status | End to end processing scope |

Figure 3(a) compares four variants under the same 300 query budget. Average precision increases from 46.8% for fixed queries to 47.5% for the capacity control, 49.4% for allocation without consistency, and 51.6% for the complete model; their count errors decrease from 12.8 to 12.1, 10.0, and 8.6 blossoms per image. The capacity control isolates the contribution of auxiliary abundance supervision because its query initialization remains fixed. The larger change associated with allocation identifies the contribution to examine when attributing a performance difference to spatial redistribution. Table 2 places this detection and counting values alongside the map errors, which decrease from 0.218 to 0.207, 0.174, and 0.151. The complete configuration therefore exhibits a 4.8 percentage point precision increase and a 32.8% count error reduction within the comparison. The agreement among the three metrics matters because an improved image count could otherwise conceal spatially compensating false positives and misses. Per image paired errors would be needed to determine whether the aggregate differences are distributed across the test partition or concentrated in a small group of crowded views. These sequential contrasts depend on the configuration in which a component is introduced. Their differences describe conditional contributions and cannot establish that supervision, routing, and consistency have independent additive effects.

**Table 2. Component comparison at a fixed 300 query budget**

| Variant | AP (%) | Count MAE | Map error | Latency (ms) |
|---|---|---|---|---|
| Fixed queries | 46.8 | 12.8 | 0.218 | 18.4 |
| Capacity control | 47.5 | 12.1 | 0.207 | 18.8 |

| | | | | |
|---|---|---|---|---|
| Allocation only | 49.4 | 10.0 | 0.174 | 19.6 |
| Complete model | 51.6 | 8.6 | 0.151 | 20.1 |

Figure 3(b) relates normalized map error to latency: the four variants occupy 18.4, 18.8, 19.6, and 20.1 milliseconds with map errors of 0.218, 0.207, 0.174, and 0.151. The complete model adds 1.7 milliseconds, approximately 9.2% of the fixed query value, while reducing spatial error by approximately 30.7%. Hardware benchmarking shows that detector latency depends on memory behavior, synchronization, and execution configuration as well as nominal model complexity [22]. These considerations guide the execution settings and processing scope of the runtime comparison. Figure 3(c) examines distributional behavior using 36 per image absolute errors per variant. Its boxes expose the median and interquartile range, while individual observations retain the tails that aggregate means conceal. The distribution panel characterizes a diagnostic subset, whereas the table reports the aggregate comparison; these two statistical supports require separate interpretation. Paired comparison requires one error record per test image so that improvements in central tendency can be assessed alongside the frequency of unusually large counting failures. The selection rule for the 36 displayed images is required to assess their representativeness. Without that rule and the complete paired records, the panel supports descriptive inspection rather than an uncertainty estimate for the full test partition.

Figure 3(d) separates recall across five abundance bins. Fixed query recall declines through 0.91, 0.86, 0.79, 0.68, and 0.56, whereas the complete model follows 0.92, 0.90, 0.86, 0.80, and 0.71. The gap expands from 0.01 in the lowest bin to 0.15 in the highest, identifying crowded regions as the intended source of allocation benefit. The remaining decline also identifies an unresolved capacity or visibility constraint that regional redistribution alone may leave intact. Recommendations for machine-learning-based science distinguish model development and selection from independent evaluation to support valid performance assessment [23]. Accordingly, a measured implementation would select its confidence threshold using a declared validation objective and preserve that threshold across abundance strata. Candidate coverage, target visibility, and nearest initialization distance should be recorded for missed flowers in the highest bin. These records distinguish a flower that received insufficient representation from a flower whose initialized hypothesis failed during decoding, and permit the observed recall curve to be interpreted in terms of the computational stage responsible for each omission. The recall curve localizes the abundance range associated with omissions. Attribution to routing, decoding, or visibility requires the candidate and annotation diagnostics, since the same recall loss can arise at different stages.

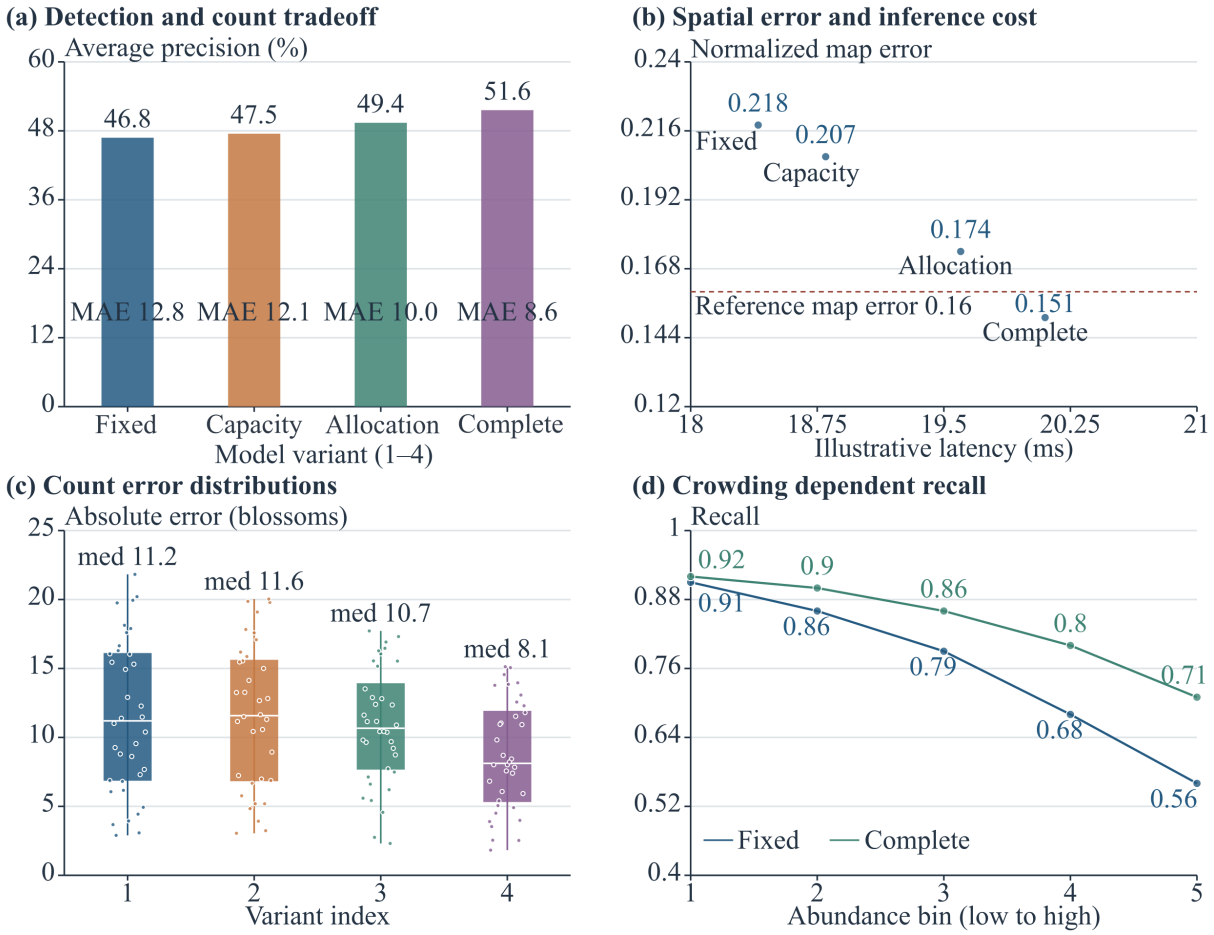



**Figure 3. Comparative evaluation. (a) Detection average precision and count error for four model variants. (b) Map error versus latency. (c) Per-image count error distributions. (d) Recall across five abundance bins**

### 5.2. Allocation ablations and operating sensitivity

The component comparison separates abundance supervision, query redistribution, spatial diversity, and count consistency. Orchard machine-vision research evaluates bud detection, counting, and branch measurement as distinct components of a crop-load management workflow [24]. The fixed query, capacity control, allocation only, and complete configurations have average precision values of 46.8%, 47.5%, 49.4%, and 51.6%. An additional configuration with spatial diversity disabled has precision of 50.0% and count error of 9.5 blossoms, compared with 51.6% and 8.6 blossoms for the complete configuration. This diagnostic asks whether distributing queries among cells is sufficient or whether the selected locations within each cell must also represent distinct flowers. Repeated initialization locations, returned regional shares, and the fraction of targets lacking a nearby candidate would provide the relevant intermediate evidence. The interpretation also depends on how frequently the minimum separation rule is relaxed. Frequent relaxation in dense cells would indicate that the candidate representation, rather than the integer budget itself, limits the availability of distinct hypotheses. Relative to the complete configuration, disabling spatial diversity reduces average precision by 1.6 percentage points and increases count error by 0.9 blossoms per image. This contrast concerns within-cell candidate separation under the reported settings.

Figure 4(a) examines reserve fractions of 0.05, 0.10, 0.15, 0.20, and 0.25, with precision values of 48.5%, 50.6%, 51.6%, 51.0%, and 49.7%. The interior peak reflects the intended balance between protection against underestimated abundance and concentration of queries in crowded cells. Increasing the reserve from 0.05 to 0.15 raises precision by 3.1 percentage points, whereas increasing it further to 0.25 reduces precision by 1.9 points. Figure 4(b) examines tempering exponents of 0.25, 0.50, 0.75, 1.00, and 1.25, yielding count errors of 10.8, 9.1, 8.6, 9.0, and 10.2 blossoms. The settings above one deliberately extend the test toward stronger concentration than the proposed tempered operating range. Their higher error motivates inspecting whether isolated abundance peaks absorb queries at the expense of neighboring flowers. Hyperparameter optimization studies compare search strategies and parameter sensitivity to identify configurations suited to the target dataset [25]. Both settings should therefore be chosen on validation data before their interaction is examined on held out orchard blocks. The interior optima characterize the reported parameter sweep. Transfer of these settings to another orchard remains conditional on image scale, abundance distribution, and the reliability of the predicted density signal.

Figure 4(c) displays dense stratum recall over five reserve fractions and five query budgets, with cell values from 0.580 to 0.810. Each row corresponds to a fixed budget and therefore separates the reserve effect from the increase in total decoder capacity. At 200 queries, the five recall values are 0.580, 0.605, 0.670, 0.640, and 0.595; at 400 queries, they become 0.720, 0.745, 0.810, 0.780, and 0.735. The repeated interior peak indicates that additional capacity improves coverage while preserving a preference for moderate exploration. The response surface can change if the density branch becomes less reliable at a different image scale or if sparse regions consume a larger share of the scene. The allocator grid and map evaluation grid should therefore remain separate. Coarsening the latter can absorb localization errors into larger cells, while changing the former alters the routing decision itself. Evaluation at several fixed map scales would reveal whether a gain persists at the spatial resolution needed for a canopy management decision. The allocation grid controls where decoder candidates are initialized. The evaluation grid controls the spatial scale at which map disagreement is measured, so changing the latter can alter the reported error even when detections are unchanged.

Figure 4(d) presents the efficiency frontier for budgets of 200, 250, 300, 350, and 400 queries. Their precision values are 48.9%, 50.5%, 51.6%, 52.0%, and 52.2%, at latencies of 17.5, 18.8, 20.1, 21.8, and 23.7 milliseconds. Increasing the budget from 200 to 300 adds 2.7 percentage points for 2.6 milliseconds; increasing it from 300 to 400 adds 0.6 points for 3.6 milliseconds. The changing slope identifies the range in which additional hypotheses provide diminishing detection benefit. Reviews of FPGA-based detection emphasize latency, throughput, and accuracy as joint criteria for real-time deployment [26]. The 22-millisecond reference is an operating boundary: the 350-query setting falls below it, while the 400-query setting exceeds it. The runtime record should include image decoding, mask generation, candidate sorting, allocation, decoder execution, kernel reconstruction, and serialization. Latency percentiles and peak memory would help determine whether occasional dense frames produce delays that the mean latency conceals. The reported latency contrast describes the listed configuration. A guaranteed frame-rate claim additionally requires a declared device, software stack, timing boundary, warmup procedure, and tail-latency record under the intended acquisition stream.

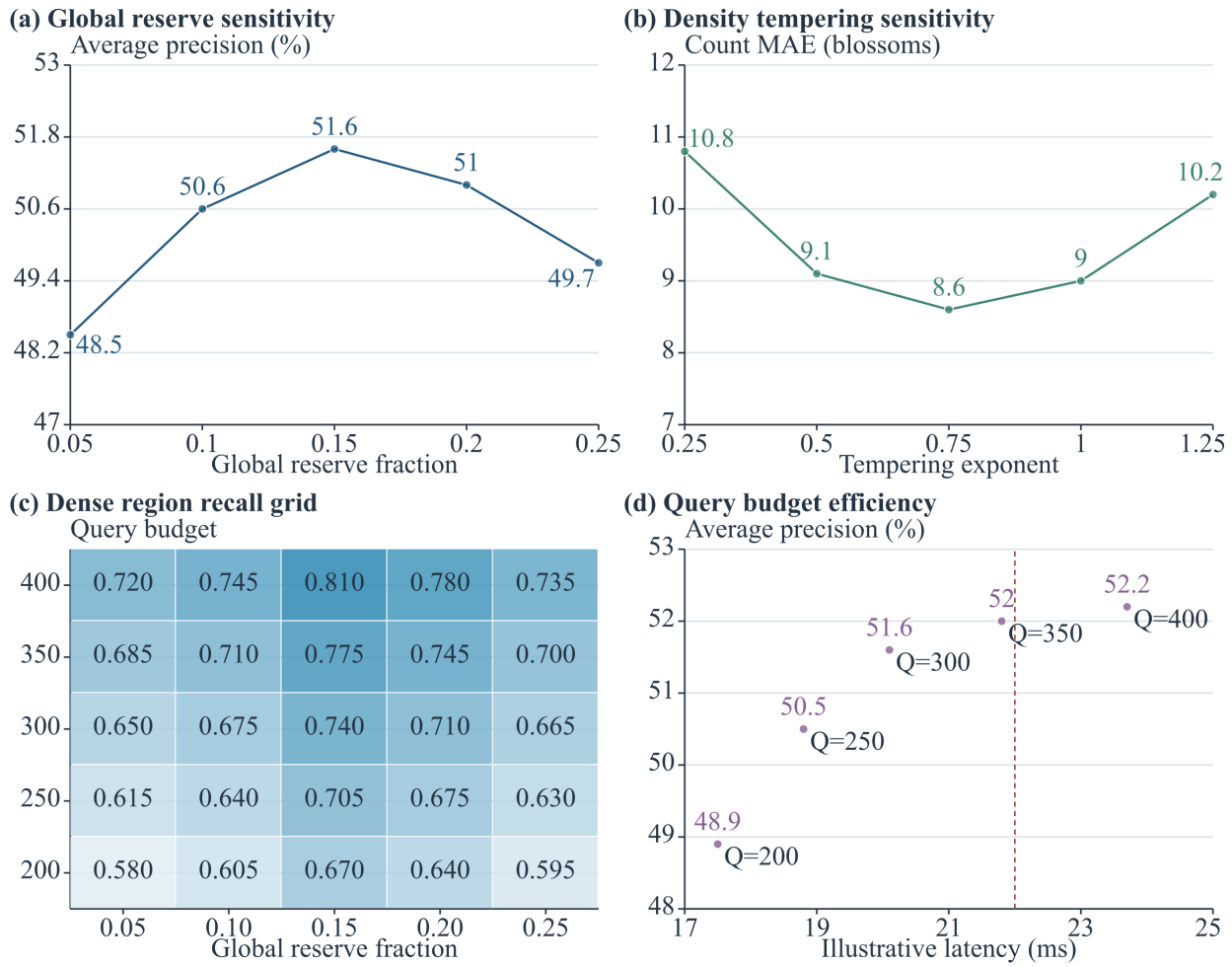



**Figure 4. Allocation sensitivity. (a) Reserve fraction response. (b) Tempering exponent response. (c) Dense stratum recalls over reserve fractions and query budgets. (d) Query budget efficiency frontier**

### 5.3. Spatial residuals and acquisition robustness

Orchard phenotyping pipelines combine instance localization with depth-based measurements to characterize flower corymbs and developing fruitlets [27]. Figure 5(a) compares annotated regional counts of 20, 40, 60, 80, and 100 with complete model predictions of 22, 39, 57, 76, and 94 and fixed query predictions of 23, 35, 51, 66, and 79. The complete model residual changes from an excess of two blossoms to a deficit of six; the fixed query deficit reaches 21 at the largest count. This abundance dependence is consistent with a residual crowded region limitation across the regional count settings. Figure 5(b) expands the comparison to five regional residual distributions, each containing 30 regional observations centered around an abundance-dependent center. The spread indicates that regions with the same nominal abundance can experience different error magnitudes, which a mean trajectory can obscure. Figure 5(c) displays the empirical cumulative distributions of absolute error for the fixed and complete configurations. The ten-blossom reference permits a tolerance-based comparison, while the upper tail identifies regions that may require manual inspection. The tolerance itself should follow the intended mapping use and the spatial support over which a discrepancy is consequential.

Figure 5(d) ranks five acquisition conditions with complete model count errors of 6.9, 8.2, 9.0, 11.5, and 14.0 blossoms under clear, shaded, backlit, blurred, and heavily occluded imagery. Small-object detection remains sensitive to low resolution, occlusion, and background interference [28]. The 7.1 blossom difference between clear and heavily occluded conditions motivates separate inspection of the density signal and decoder response. Blur can merge nearby features before allocation, while occlusion can reduce the visual evidence available to both branches. Figure 5(e) shows normalized map errors of 0.142, 0.151, 0.160, 0.184, and 0.221 at valid support fractions of 1.0, 0.8, 0.6, 0.4, and 0.2. The error increase accelerates as support becomes more restricted, identifying the interaction between fragmented masks and reconstruction boundaries. Figure 5(f) plots 150 regional residuals against abundance and retains the zero-error reference. The increasingly negative centers indicate systematic underestimation, whereas the vertical spread reflects additional regional variation around that bias. Residual analysis should associate these points with visibility and boundary proximity to locate the conditions producing the largest deficits. At small valid fractions, a few detections can

dominate an area-normalized estimate. Fragmented regions should therefore be interpreted together with their visible area and raw count, since the same absolute count discrepancy can imply very different density errors.

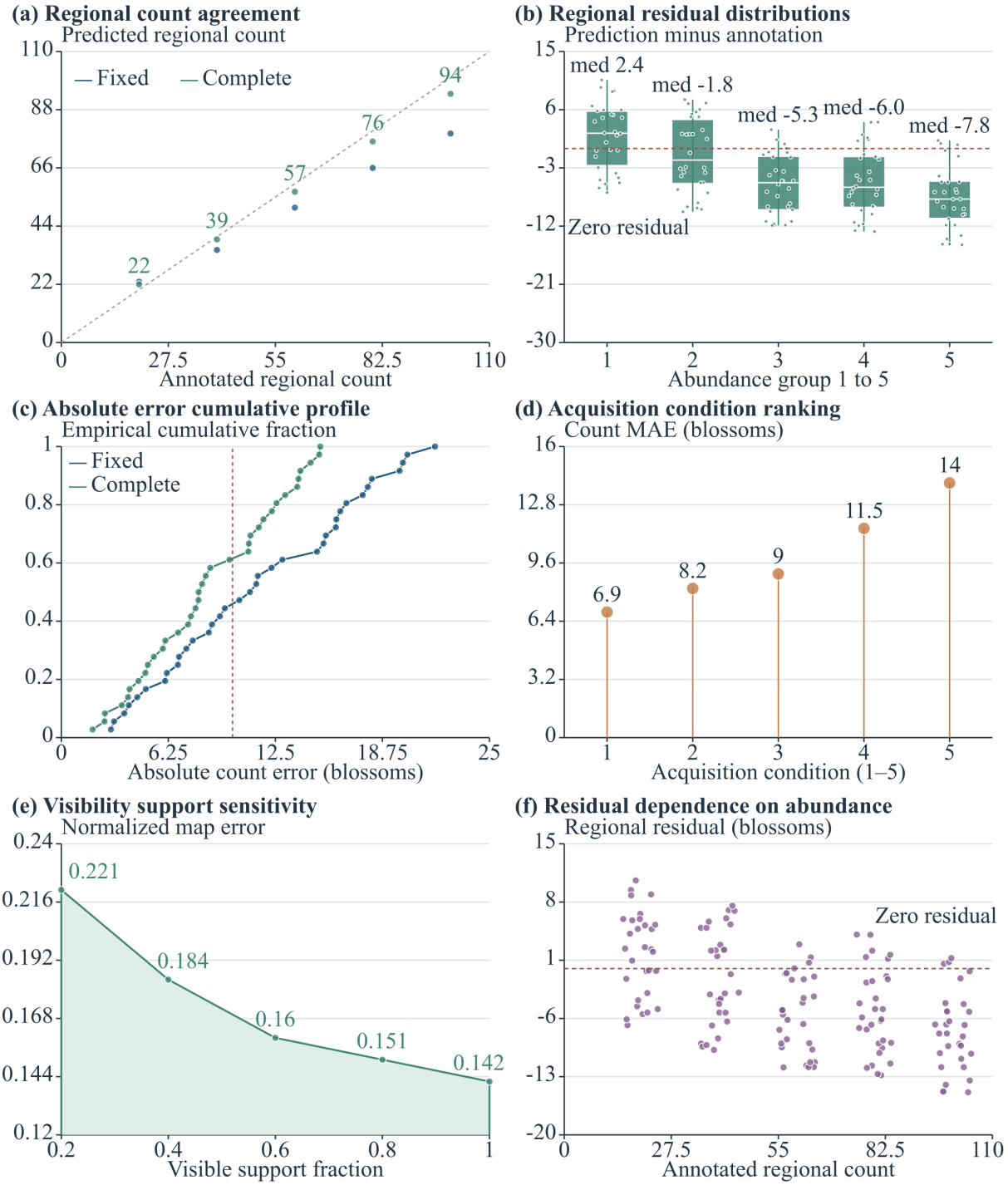



**Figure 5. Spatial and robustness diagnostics. (a) Annotated versus predicted regional count. (b) Regional residual distributions. (c) Absolute error cumulative distributions. (d) Acquisition condition ranking. (e) Map error across support fractions. (f) Residual dependence on abundance**

Confidence-consistent detection research identifies discrepancies between classification confidence and localization quality as a reliability problem [29]. Figure 6 uses confidence bin centers of 0.55, 0.65, 0.75, 0.85, and 0.95, with observed precisions of 0.50, 0.61, 0.72, 0.82, and 0.93. Their confidence gaps decrease from 0.05 to 0.02, indicating mild overconfidence across the confidence range. Detection supports of 150, 240, 330, 280, and 180 determine marker area and show that the middle bins contribute more observations than either extreme. The declared threshold of 0.60 lies between the first two bin centers, so the displayed bins provide a coarse diagnostic around the operating decision. Finer bins or individual score records would be required to characterize predictions immediately above the threshold. Calibration should also be inspected within shaded, blurred, and occluded subsets, because an acceptable pooled relationship can coexist with a subgroup whose high confidence detections have lower precision. Recalibration would

require its own fitting subset and an independent evaluation subset to preserve the meaning of the displayed reliability relationship. Weighting the five confidence gaps by their displayed supports gives an average gap of approximately 0.032. This descriptive value applies to the displayed bins; the restricted score range and subgroup aggregation limit its interpretation as an overall calibration measure.

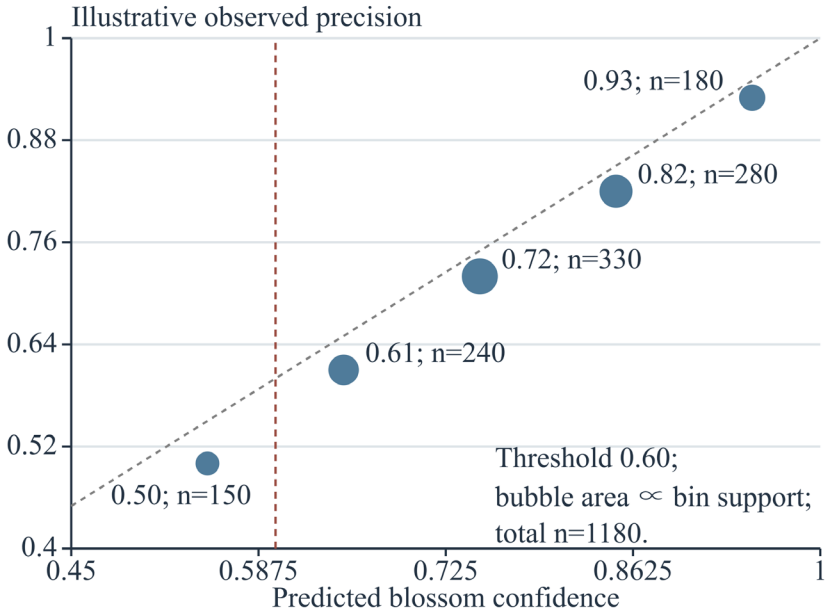



**Figure 6. Confidence reliability diagnostic**

Uncertainty analysis should preserve the sampling structure of the evaluation. Bootstrap inference for clustered mixed-effects models provides a framework for quantifying uncertainty in cluster-level estimates under distributional misspecification [30]. Therefore, in a field experiment, orchard blocks and their corresponding image data were resampled, and a separate measure of optimization variability was provided by training seeds. The two test blocks in the proposed split provide limited evidence of transfer across locations or seasons. A broader acquisition design should include additional orchards and assess whether the abundance-dependent residual pattern persists within each block. Reproducibility studies identify documentation, code and data availability, and experimental conditions as complementary requirements for reproducible machine-learning research [31]. A plot ledger is used to store the aggregated data of regional sample distributions and their corresponding display statistics. Each plotted observation should retain an image identifier, valid support fraction, annotation count, accepted detection count, and regional reconstruction setting. These fields allow large residuals to be traced to their source images and permit a reviewer to examine whether mask fragmentation, threshold crossings, or candidate saturation contributed to a particular spatial discrepancy. With two held-out blocks, resampling cannot characterize broad between-orchard variability reliably. The present point estimates should therefore be interpreted within the reported partition until block-level records, seed variation, and a larger external sample support uncertainty intervals.

## 6. Conclusion

This study develops a density guided query allocation framework for RT-DETR, linking regional blossom abundance with a conserved decoder budget to improve detection accuracy and image-plane density mapping.

1. Detection and mapping performance. In the reported comparison, the complete model achieves an average precision of 51.6%, exceeding the fixed query baseline by 4.8 percentage points. Mean absolute count error decreases from 12.8 to 8.6 blossoms per image, a reduction of 32.8%, while normalized map error decreases from 0.218 to 0.151. These improvements connect more accurate instance localization with better preservation of regional blossom abundance.

2. Density dependence and computational tradeoff. The recall advantage increases with blossom abundance, expanding from 0.01 in the lowest abundance bin to 0.15 in the highest. This pattern identifies crowded canopy regions as the principal beneficiaries of spatial query redistribution. Latency increases from 18.4 to 20.1 milliseconds, approximately 9.2%. Increasing the query budget beyond 300 produces progressively smaller precision gains, indicating that allocation efficiency becomes increasingly important as decoder capacity grows.

3. Innovation and applicability. The framework combines budget-conserving regional allocation, spatially diverse candidate selection, and protected global exploration within a common detection pathway. Its effectiveness depends on the accuracy of abundance predictions and the availability of visible blossom features. Residual underestimation persists

in crowded or heavily occluded regions, while interpreting image-plane maps as physical surface densities requires calibrated canopy geometry. The immediate application is prioritizing inspection of visible canopy regions. Effects on thinning decisions, yield forecasting, or labor requirements require separate task-level validation with field management outcomes.

## 7. Declarations

### 7.1 Author Contributions

Tadas Mikalauskas contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Arnas Butkus contributes to validation, analysis, investigation, data collection, draft preparation. Ramūnas Šležas contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

### 7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

### 7.3 Institutional Review Board Statement

Not applicable.

### 7.4 Informed Consent Statement

Not applicable.

### 7.5 Date Availability

Not applicable.

### 7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### 7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

## 8. References

[1] Lv, M., Xu, Y.-X., Miao, Y.-H., & Su, W.-H. (2025). A comprehensive review of deep learning in computer vision for monitoring apple tree growth and fruit production. Sensors, 25(8), Article 2433. https://doi.org/10.3390/s25082433

[2] Shang, Y., Geng, M., Fang, H., Cai, M., Wang, J., & Song, H. (2024). Using unmanned aerial vehicle acquired RGB images and Density-Cluster-Count model for tree-level apple flower quantification. Computers and Electronics in Agriculture, 226, Article 109389. https://doi.org/10.1016/j.compag.2024.109389

[3] Yan, J., Duan, H., Xu, R., Sun, M., Yang, L., & Jia, W. (2026). DyRFA-Net: Apple flower blooming stage detection based on anther color features. Computers and Electronics in Agriculture, 248, Article 111772. https://doi.org/10.1016/j.compag.2026.111772

[4] Vitousek, M., Vrba, J., Cejnek, M., Jura, J., & Schankova, K. (2026). Dual-view weakly-supervised learning for apple tree flower counting. Frontiers in Plant Science, 17, Article 1869941. https://doi.org/10.3389/fpls.2026.1869941

[5] Geng, M., Shang, Y., Xiang, S., Wang, J., Wang, L., & Song, H. (2025). Using improved density peak clustering algorithm for flower cluster identification and apple central and peripheral flower detection. Computers and Electronics in Agriculture, 232, Article 110095. https://doi.org/10.1016/j.compag.2025.110095

[6] Wang, D., Song, H., & Wang, B. (2025). YO-AFD: An improved YOLOv8-based deep learning approach for rapid and accurate apple flower detection. Frontiers in Plant Science, 16, Article 1541266. https://doi.org/10.3389/fpls.2025.1541266

[7] Yang, Y., Li, Y., Cao, K., Chen, X., & Jia, W. (2025). LEAF-Net: A multi-scale frequency-aware framework for automated apple

blossom monitoring in complex orchard. Horticulturae, 11(11), Article 1382. https://doi.org/10.3390/horticulturae11111382

[8] Bhattarai, U., Bhusal, S., Zhang, Q., & Karkee, M. (2024). AgRegNet: A deep regression network for flower and fruit density estimation, localization, and counting in orchards. Computers and Electronics in Agriculture, 227, Article 109534. https://doi.org/10.1016/j.compag.2024.109534

[9] Huang, Y.-X., Liu, H.-I., Shuai, H.-H., & Cheng, W.-H. (2025). DQ-DETR: DETR with dynamic query for tiny object detection. In A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, & G. Varol (Eds.), Computer vision – ECCV 2024 (Lecture Notes in Computer Science, Vol. 15134, pp. 290–305). Springer. https://doi.org/10.1007/978-3-031-73116-7_17

[10] Wang, S., Liu, M., Dong, S., & Chen, S. (2025). ERL-RTDETR: A lightweight transformer-based framework for high-accuracy apple disease detection in precision agriculture. Concurrency and Computation: Practice and Experience, 37(23–24), Article e70276. https://doi.org/10.1002/cpe.70276

[11] Wang, D., & Wang, B. (2026). A high-quality apple flower detection dataset for precision agriculture. Frontiers in Plant Science, 17, Article 1927541. https://doi.org/10.3389/fpls.2026.1927541

[12] Li, T., Zhang, Z., Zhu, M., Cui, Z., & Wei, D. (2024). Combining transformer global and local feature extraction for object detection. Complex & Intelligent Systems, 10(4), 4897–4920. https://doi.org/10.1007/s40747-024-01409-z

[13] Koldasbayeva, D., Tregubova, P., Gasanov, M., Zaytsev, A., Petrovskaia, A., & Burnaev, E. (2024). Challenges in data-driven geospatial modeling for environmental research and practice. Nature Communications, 15(1), Article 10700. https://doi.org/10.1038/s41467-024-55240-8

[14] Zhou, L., Rao, S., Li, W., Hu, B., & Sun, B. (2024). Multi-branch progressive embedding network for crowd counting. Image and Vision Computing, 148, Article 105140. https://doi.org/10.1016/j.imavis.2024.105140

[15] Phan, D. T., Gao, J., Lu, Y., Yap, K.-H., Garg, K., & Han, B. S. (2024). DEO-Net: Joint density estimation and object detection for crowd counting. IEEE Transactions on Instrumentation and Measurement, 73, 1–11. https://doi.org/10.1109/tim.2024.3441018

[16] Yuan, M., Zhang, W., Yan, C., Gong, T., Zhang, Y., & Ying, J. (2024). Adaptive token selection for efficient detection transformer with dual teacher supervision. Knowledge-Based Systems, 300, Article 112036. https://doi.org/10.1016/j.knosys.2024.112036

[17] Senthivel, T., & Vu, N.-S. (2025). QR-DETR: Query routing for detection transformer. In M. Cho, I. Laptev, D. Tran, A. Yao, & H. Zha (Eds.), Computer vision – ACCV 2024 (Lecture Notes in Computer Science, Vol. 15477, pp. 395–412). Springer. https://doi.org/10.1007/978-981-96-0960-4_24

[18] Huseljic, D., Herde, M., Hahn, P., Müjde, M., & Sick, B. (2025). Systematic evaluation of uncertainty calibration in pretrained object detectors. International Journal of Computer Vision, 133(3), 1033–1047. https://doi.org/10.1007/s11263-024-02219-z

[19] John, K., Saurette, D. D., & Heung, B. (2025). The problematic case of data leakage: A case for leave-profile-out cross-validation in 3-dimensional digital soil mapping. Geoderma, 455, Article 117223. https://doi.org/10.1016/j.geoderma.2025.117223

[20] Li, Z., Dong, Y., Shen, L., Liu, Y., Pei, Y., Yang, H., Zheng, L., & Ma, J. (2024). Development and challenges of object detection: A survey. Neurocomputing, 598, Article 128102. https://doi.org/10.1016/j.neucom.2024.128102

[21] Zhang, Z., Zhang, L., Zhang, H., Guo, Y., Wang, H., & Lu, X. (2024). AMSA-CAFF Net: Counting and high-quality density map estimation from X-ray images of electronic components. Expert Systems with Applications, 237, Article 121602. https://doi.org/10.1016/j.eswa.2023.121602

[22] Kong, C., Li, F., Yan, X., Yang, J., Mo, P., Luo, Q., & Mao, R. (2026). Object detection on low-compute edge SoCs: A reproducible benchmark and deployment guidelines. Scientific Reports, 16(1), Article 5875. https://doi.org/10.1038/s41598-026-36862-y

[23] Kapoor, S., Cantrell, E. M., Peng, K., Pham, T. H., Bail, C. A., Gundersen, O. E., Hofman, J. M., Hullman, J., Lones, M. A., Malik, M. M., Nanayakkara, P., Poldrack, R. A., Raji, I. D., Roberts, M., Salganik, M. J., Serra-Garcia, M., Stewart, B. M., Vandewiele, G., & Narayanan, A. (2024). REFORMS: Consensus-based recommendations for machine-learning-based science. Science Advances, 10(18), Article eadk3452. https://doi.org/10.1126/sciadv.adk3452

[24] Pawikhum, K., Yang, Y., He, L., & Heinemann, P. (2025). Development of a machine vision system for apple bud thinning in precision crop load management. Computers and Electronics in Agriculture, 236, Article 110479. https://doi.org/10.1016/j.compag.2025.110479

[25] Raiaan, M. A. K., Sakib, S., Fahad, N. M., Mamun, A. A., Rahman, M. A., Shatabda, S., & Mukta, M. S. H. (2024). A systematic review of hyperparameter optimization techniques in convolutional neural networks. Decision Analytics Journal, 11, Article 100470. https://doi.org/10.1016/j.dajour.2024.100470

[26] Hozhabr, S. H., & Giorgi, R. (2025). A survey on real-time object detection on FPGAs. IEEE Access, 13, 38195–38238.

https://doi.org/10.1109/access.2025.3544515

[27] Checola, G., Moser, D., Sonego, P., Iob, C., Micheli, F., & Franceschi, P. (2025). Apple phenotyping using deep learning and 3D depth analysis: An experimental study on fruitlet sizing during early development. Smart Agricultural Technology, 11, Article 100964. https://doi.org/10.1016/j.atech.2025.100964

[28] Nikouei, M., Baroutian, B., Nabavi, S., Taraghi, F., Aghaei, A., Sajedi, A., & Moghaddam, M. E. (2025). Small object detection: A comprehensive survey on challenges, techniques and real-world applications. Intelligent Systems with Applications, 27, Article 200561. https://doi.org/10.1016/j.iswa.2025.200561

[29] Liu, C., Li, X., Xiao, W., & Xie, S. (2024). CCDet: Confidence-consistent learning for dense object detection. IEEE Transactions on Image Processing, 33, 2746–2758. https://doi.org/10.1109/tip.2024.3378457

[30] Reluga, K., Lombardía, M.-J., & Sperlich, S. (2024). Bootstrap-based statistical inference for linear mixed effects under misspecifications. Computational Statistics & Data Analysis, 199, Article 108014. https://doi.org/10.1016/j.csda.2024.108014

[31] Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., & Kowald, D. (2025). Reproducibility in machine-learning-based research: Overview, barriers, and drivers. AI Magazine, 46(2), Article e70002. https://doi.org/10.1002/aaai.70002