

Cite this as:

Bingquan CHU, Ruiyuan WU, Haijun ZHANG, Haochuan QIN, Zishun PENG, Fengle ZHU, Yong HE, 2026. Embedding of ripening topology into one-stage detection for tomato cluster phenotyping. *Journal of Zhejiang University-SCIENCE B*, 27(5):466-481.

<https://doi.org/10.1631/jzus.B2500647>

Embedding of ripening topology into one-stage detection for tomato cluster phenotyping

Keywords

Tomato ripeness

Phenotype

Object detection

Topology

YOLO

Spatial sequence

Authors

Bingquan CHU¹, Ruiyuan WU¹, Haijun ZHANG¹, Haochuan QIN¹,

Zishun PENG¹, Fengle ZHU², Yong HE³

¹Zhejiang University of Science and Technology

²Zhejiang University of Technology

³Zhejiang University

Research Background & Challenges

Context and motivation for integrating topological priors into agricultural object detection



Global Agricultural Context

Global tomato production reached **186.1 million tons** in 2022, with China contributing nearly **37%**. However, harvesting in major producing regions remains predominantly dependent on manual ripeness assessment, leading to inefficiency and inconsistent product quality.



Current Limitations

Existing computer vision approaches rely solely on appearance-based features, making them susceptible to **variable lighting conditions**, **subjective judgment**, and **color ambiguities** between adjacent ripening stages. These methods overlook critical biological prior knowledge about fruit development patterns.



Key Insight

Tomatoes on the same cluster exhibit a distinct ripening pattern: **fruits at higher positions mature earlier** than those lower on the cluster due to physiological and environmental factors.

This **top-to-bottom ripening topology** provides a powerful spatial prior that can resolve color ambiguities and improve detection robustness.



Research Question

How can we effectively encode this biological prior knowledge into deep learning-based object detection frameworks for accurate in situ phenotyping?

Rank-Aware YOLO: Core Innovation

A novel framework incorporating biological ripening topology into one-stage detection



Framework Overview

The Rank-Aware YOLO framework introduces a plug-and-play "Rank-Aware Module" that integrates the biological prior of top-to-bottom ripening into standard YOLO detectors.

- 1 Feature Extraction: MCAM-enhanced backbone with PAN fusion
- 2 Position-Aware Prediction: Four-branch head with h_{rel} regression
- 3 Rank-Aware Supervision: DM-RankLoss for spatial sequence



Plug-and-Play Design

The Rank-Aware Module is **model-agnostic**, seamlessly integrating with multiple YOLO architectures (v8, v9, v11, v12) without modifying backbone or neck structures.

Architecture Components

A Efficient Position-Aware Head

Novel detection head with **four parallel branches** including h_{rel} prediction, using lightweight group convolutions for edge deployment.

B Dynamic Margin-Aware Ranking Loss

Three-component loss function (L_{rank} , $L_{penalty}$, L_{reg}) enforcing correct spatial sequences with dynamic margins.

C Extended Label Format

Enhanced YOLO labels with **clusterID** and h_{rel} attributes for topological position quantification.

4.96-5.66

mAP50 Gain (%)

>10%

Parameter Reduction

Key Technical Components

Detailed architecture of the two core innovations driving performance improvements



Efficient Position-Aware Head

Novel detection head design

Four Parallel Branches

Unlike standard YOLO heads, our design includes: **Bounding Box Regression**, **Class Prediction**, **Objectness Score**, and a novel **h_{rel} Prediction** branch for relative height regression.

Lightweight Group Convolutions

Shared stem uses **two 3x3 group convolution layers**, splitting input channels into independent groups. This reduces both parameter count and computational complexity (FLOPs) while preserving representation capabilities.

Fully Decoupled Structure

All prediction tasks are structurally decoupled with **single 1x1 convolution layers** per branch. Compared to original multi-layer structures, this achieves **~40% parameter reduction** in the detection head.



Dynamic Margin-Aware Ranking Loss

Three-component hybrid loss function

L_{rank} : Pairwise Ranking Loss

Base Component

Uses **dynamic margin M_{ij}** proportional to ground-truth h_{rel} differences. Enforces that fruits with higher ripeness occupy higher physical positions, promoting discriminative embedding spaces.

$L_{penalty}$: Differentiable Re-weighting

Critical Enhancement

Leverages **Sigmoid function** for smooth, class-aware penalty scaling. Prioritizes correcting critical ranking errors (e.g., fully ripe below ripe) through adaptive gradient weighting.

L_{reg} : Auxiliary Regression Loss

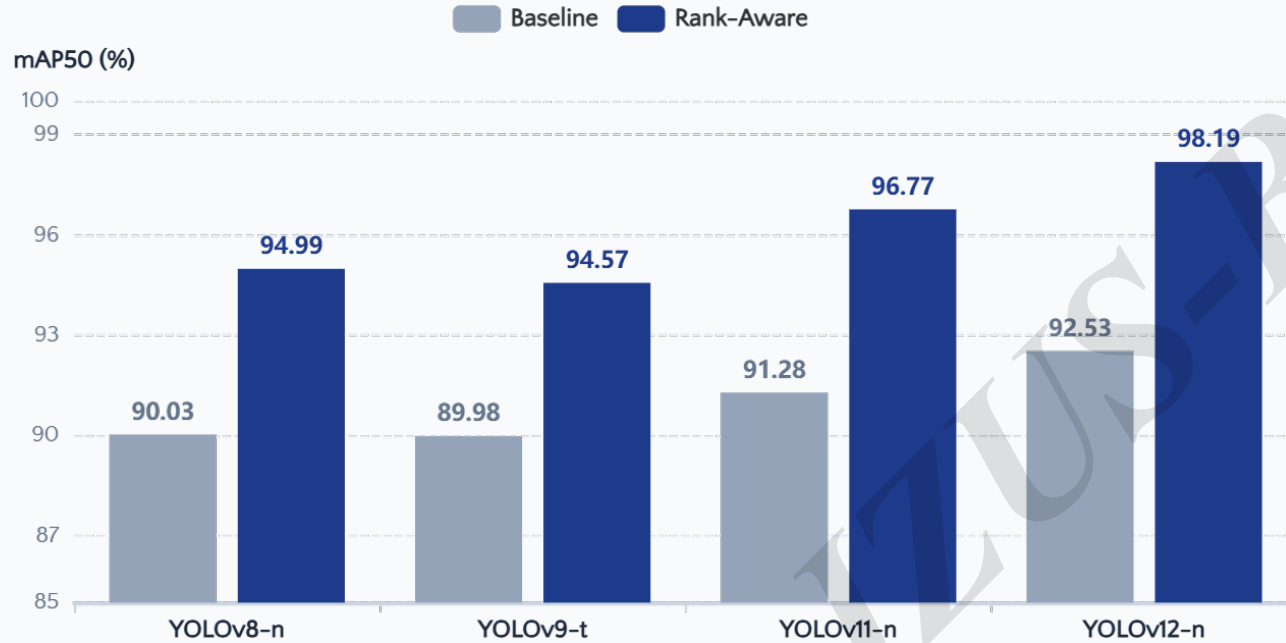
Stability Anchor

Smooth L1 loss anchors predictions within $[0, 1]$ range, preventing global drift. Provides stable gradients and ensures accurate absolute h_{rel} estimation alongside ordinal relationships.

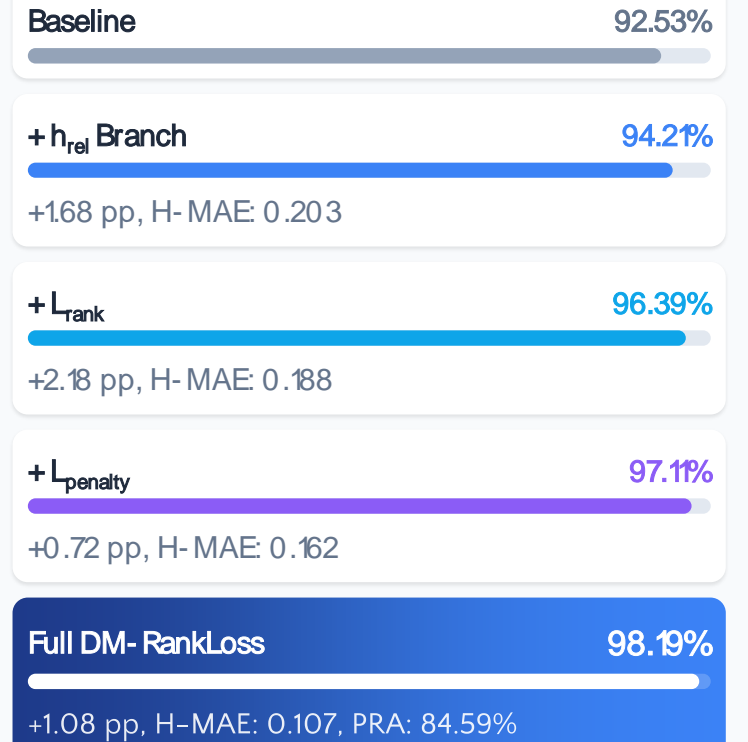
Experimental Validation & Results

Comprehensive evaluation demonstrating framework effectiveness across multiple dimensions

Performance Across YOLO Architectures



Ablation Study Results



Best Performance

98.19%

mAP50 (YOLOv12-n)

+5.66 pp improvement

P = 0.05 significance

Height Accuracy

0.107

H-MAE (normalized)

47.3% reduction

from 0.203 baseline

Robustness: Lighting

Dark Light	+8.33%
Low Light	+6.60%
Strong Light	+5.75%

Key Contributions & Implications

Advancing agricultural computer vision through topology-aware deep learning

★ Primary Contributions

1 First Topology-Embedded Framework

First work to explicitly embed ripening topology into one-stage detection for agricultural phenotyping, demonstrating that spatial-ordinal relationships significantly improve detection accuracy.

2 Software-Centric Paradigm

Eliminates need for expensive multi-modal sensors (depth, hyperspectral) by encoding domain knowledge directly into learning process, offering cost-effective deployment pathway.

3 Model-Agnostic Enhancement

Plug-and-play Rank-Aware Module demonstrates consistent improvements across multiple YOLO architectures, validating generalizability and practical applicability.

🌐 Broader Implications

Agricultural Robotics

Provides robust vision systems for **automated harvesting robots**, enabling accurate in situ phenotyping without additional hardware complexity.

Transferability to Other Crops

Framework applicable to crops with **similar spatial-ordinal patterns** (grapes, peppers, bananas), offering generalizable solution for structured agricultural objects.

Structured Reasoning in AI

Demonstrates value of **integrating physical constraints** into deep learning, shifting from pure pattern recognition to structured reasoning with domain knowledge.

📍 Future Directions

- **3D Spatial Modeling:** Extend to point cloud space using stereo depth data
- **Multi-Crop Extension:** Apply to diverse cultivars and morphologies
- **Complex Topologies:** Model branching and clustering patterns

“ Integrating real-world physical knowledge is a vital step toward advancing AI from perception to cognition ”