YOLOv8 Research Ideas Summary

Mohammad Reza Saadati

November 2025

Introduction

This report is a personal summary of several research papers related to YOLOv8 and similar object detection models. Each section corresponds to one paper, and under it, I will write the main idea, contribution, or interesting technique proposed by the authors. The purpose is to collect and organize new ideas that could potentially improve YOLOv8.

1 Paper : YOLO-APDM — Improved YOLOv8 for Road Target Detection in Infrared Images

Reference: DOI: 10.3390/s24227197

Main idea: The paper proposes several structural improvements to YOLOv8 for infrared road-target detection.

Modifications to YOLOv8:

- Added a **P2 detection layer** to improve detection of very small targets.
- Replaced all **C2f modules** with **C2f-DCNv3** blocks using deformable convolutions.
- Inserted a **Multi-Scale Convolutional Attention (MSCA)** module after the SPPF block to enhance focus on key regions.
- Reconstructed the **neck structure** for more effective multi-scale feature fusion.

2 Paper : CIMB-YOLOv8 — A Lightweight Remote Sensing Object Detection Network Based on Contextual Information and Multiple Branches

Reference: DOI: 10.3390/electronics14132657

Main idea: The paper presents a lightweight and efficient YOLOv8 variant for remote sensing object detection by enhancing contextual understanding and multi-scale feature extraction.

Modifications to YOLOv8:

- Introduced a **Space-to-Depth Multi-Branch Pyramid (SMP)** module to improve multi-scale feature fusion and extract fine details from small targets.
- Designed an **OCSP (Omni Cross Stage Partial)** structure within SMP, combining large-scale, local, and global branches for richer contextual features.

- Proposed a lightweight detection head called **LGDD (Lightweight Generalized Detection Head)** that includes:
  - **Shared convolution layers** to reduce parameters.
  - **GroupNorm Convolution (GNConv)** for improved feature representation.
  - **Detail-Enhanced Convolution (DEConv)** with re-parameterization during inference to boost detail learning without increasing computation.

3 Paper : CPDD-YOLOv8 — A Small Object Detection Model in Aerial Images

Reference: DOI: 10.1038/s41598-024-84938-4

Main idea: The paper improves YOLOv8 for small object detection in aerial imagery by enhancing cross-scale fusion and attention mechanisms.

Modifications to YOLOv8:

- Introduced a **C2f-GAM structure** that integrates a Global Attention Module into the C2f block for better small-object feature extraction.
- Added a **CPDD (Cross-Scale Progressive Dense Detection)** module to enhance multi-scale feature fusion.
- Optimized the detection head with attention-based refinement and improved anchor adaptability for aerial small-object detection.

4 Paper 4: Automatic Object Detection for Behavioural Research Using YOLOv8

Reference: DOI: 10.3758/s13428-024-02420-5

Main idea: The study assesses conditions under which YOLOv8 can be effectively applied in behavioural research for automatic object detection in video recordings.

Modifications / key considerations to YOLOv8:

- Focused on using YOLOv8 in a controlled lab-setting rather than proposing major architectural changes.
- Examined how training on a **small dataset (100–350 images)** still yields high detection quality.
- Highlighted the limited generalisation of the detector to new backgrounds when trained on homogeneous context; recommendation: include diverse backgrounds for robust detection.

5 Paper : Efficient Optimized YOLOv8 Model with Extended Vision

Reference: DOI: 10.3390/s24206506

Main idea: The paper proposes the YOLO-EV model, an enhanced version of YOLOv8 featuring "extended vision" mechanisms tailored for complex scenes.

Modifications to YOLOv8:

- Introduced a **Multi-Branch Group-Enhanced Fusion Attention (MGEFA)** module to improve multi-scale feature extraction and fusion.
- Enhanced the **SPPF layer** (Spatial Pyramid Pooling Fast) by integrating **Large Scale Kernel Attention (LSKA)** for better spatial information processing.
- Replaced the standard IOU loss function with a **Wise-IOU** loss to improve localization across different target sizes.
- Added a **p6 detection layer** to extend the model's ability to handle larger and multi-scale objects.

6 Paper : Enhancing the YOLOv8 Model for Real-Time Object Detection to Ensure Online Platform Safety

Reference: DOI: 10.1038/s41598-025-08413-4

Main idea: The paper builds upon YOLOv8-m to better detect harmful objects in online content by modifying key architectural blocks.

Modifications to YOLOv8:

- Enhanced the **Cross-Stage Partial (CSP) fusion blocks** in the backbone and neck to improve feature propagation and representation across stages.
- Added **three additional convolutional blocks** in the detection head to strengthen localization and classification capabilities in challenging "hard-case" scenarios.
- Increased the kernel size of the final convolution in C2f blocks from 1×1 to 3×3 to better capture spatial context for partially occluded or complex objects.

7 Paper : Improved YOLOv8 Algorithms for Small Object Detection in Aerial Imagery

Reference: DOI: 10.1016/j.jksuci.2024.102113

Main idea: The paper proposes enhancements to YOLOv8 tailored for small object detection in aerial image contexts by introducing modules for attention, scale compensation, and detection head refinement.

Modifications to YOLOv8:

- Added a **C2f-GAM** structure (Global Attention Module integrated into the C2f block) to enhance feature extraction of small objects.

- Introduced a new **P2 detection layer** to handle very small object scales in aerial images.
- Incorporated a **DSC2f** module (Deformable/Scaled C2f) replacing original C2f units for adaptive receptive fields.
- Employed a refined **DyHead detection head** to better adapt to different object scales, especially small ones.

8 Paper : Real-Time Efficiency of YOLOv5 and YOLOv8 in Human Intrusion Detection Across Diverse Environments

Reference: DOI: 10.1016/j.iot.2025.101707

Main idea: The paper presents a comparative study of YOLOv5 and YOLOv8 for human intrusion detection in real-time under varied environmental conditions to provide deployment recommendations.

Modifications / key observations on YOLOv8:

- Adapted YOLOv8's architecture and inference settings for real-time edge deployment in indoor/outdoor intrusion scenarios.
- Emphasized environmental robustness: adjusted detection thresholds, enhanced preprocessing for varied luminance and simulated weather conditions.
- Recommended configuration tweaks rather than deep structural changes: optimized input size, frame rate, and confidence thresholds to fit real-time intrusion detection constraints.

9 Paper : SOD-YOLOv8 — Enhancing YOLOv8 for Small Object Detection in Aerial Imagery and Traffic Scenes

Reference: DOI: 10.3390/s24196209

Main idea: The paper proposes SOD-YOLOv8, a specialized variant of YOLOv8 designed to boost detection of small objects in aerial imagery and traffic-scenes.

Modifications to YOLOv8:

- Added a **fourth detection layer** to leverage high-resolution spatial information for small objects.
- Enhanced the neck with a **multi-path fusion inspired by GFPN (Generalized Feature Pyramid Network)** to better integrate shallow and deep features.
- Introduced the **C2f-EMA module** (Efficient Multi-Scale Attention within C2f blocks) to redistribute attention weights and emphasize relevant features for small object detection.
- Replaced the standard CIoU loss with a novel **PIoU (Powerful-IoU)** bounding box regression loss, adding corner-difference penalties to improve localization of small objects.

10 Paper : SO-YOLOv8 — A Novel Deep Learning-Based Approach for Small Object Detection with YOLO Beyond COCO

Reference: DOI: 10.1016/j.eswa.2025.127447

Main idea: The paper proposes SO-YOLOv8, an enhanced version of YOLOv8 explicitly targeting small object detection beyond the COCO dataset.

Modifications to YOLOv8:

- Incorporated advanced hyper-parameter optimization and multi-scale training strategies to improve small object detection capability.
- Added a **Squeeze-and-Excitation (SE) block** into the backbone or neck to better capture and enhance features corresponding to small objects.
- Adapted the model to go "beyond COCO" by customizing dataset sampling, augmentation, and detection head configurations for small object-rich domains.

11 Paper : Toward Versatile Small Object Detection with Temporal-YOLOv8

Reference: DOI: 10.3390/s24227387

Main idea: This paper enhances YOLOv8 by incorporating temporal context from video sequences and specialized augmentations to improve detection of small objects across diverse environments.

Modifications to YOLOv8:

- Extensions to accept temporal input (multiple consecutive frames) rather than a single image frame.
- Data augmentations tailored for very small objects, including mosaicking and careful foreground-background balance.
- Emphasis on training with a **diverse and curated dataset** of small objects spanning civilian and military use-cases to improve robustness.

12 Paper: An Improved YOLOv8-Based Lightweight Attention Mechanism for Cross-Scale Feature Fusion

Reference: DOI: 10.3390/rs17061044

Main idea: The paper proposes **LACF-YOLO**, a YOLOv8 variant with a lightweight attention module and improved cross-scale fusion tailored for small object detection in remote sensing imagery.

Modifications to YOLOv8:

- Inserted a **Triplet Attention** module before the backbone outputs to improve context awareness.
- Replaced the standard **C2f concatenation** operations with a **dilated inverted convolution layer** to enhance contextual richness.

- Built the cross-scale feature fusion network using **partial convolution + pointwise convolution blocks** instead of standard fusion blocks.
- Used **Focal EIOU loss** in place of CIoU to speed convergence and better handle sample imbalance.

13 Paper: PKD-YOLOv8 — A Collaborative Pruning and Knowledge Distillation Framework for Lightweight Rapeseed Pest Detection

Reference: DOI: 10.3390/s25165004

Modifications to YOLOv8:

- Applied **structured pruning** on YOLOv8s based on sensitivity analysis (using the mean gamma coefficients of BatchNorm layers) to compress the model.
- Introduced a **dual distillation strategy (LMGD)** combining **logit distillation** and an improved **generative distillation** to transfer knowledge from teacher to student.
- Used a **shared detection head** between teacher and student during distillation to align semantic output (logits) for the student model.

14 Paper: LKD-YOLOv8 — A Lightweight Knowledge Distillation-Based Method for Infrared Object Detection

Reference: DOI: 10.3390/s25134054

Modifications to YOLOv8:

- Applied **masked generative distillation (MGD)** to transfer knowledge from a large teacher (YOLOv8s) to a smaller student (YOLOv8n).
- Integrated **LDConv (Linear Deformable Convolution)** modules to dynamically adjust kernel offsets and better extract spatial features.
- Included **Coordinate Attention (CA)** modules after C2f blocks to align features via channel-wise interactions.
- Replaced standard CBS modules with a hybrid **CLBS module** consisting of Convolution + LDConv + BatchNorm + activation.

15 Paper: EgoVision — A YOLO-ViT Hybrid for Robust Egocentric Object Recognition

Reference: DOI: 10.1038/s41598-025-18341-y

Main idea: The paper proposes a hybrid architecture combining YOLOv8 with Vision Transformer (ViT) modules to improve contextual reasoning in egocentric (first-person) object recognition.

Modifications to YOLOv8:

- Fused YOLOv8's spatial detection ability with a **Vision Transformer (ViT)** branch to capture global context and long-range dependencies.
- Employed a **key-frame extraction strategy** to reduce computational cost while preserving relevant frames for recognition.
- Integrated a **Feature Pyramid Network (FPN)** to harmonize multi-scale features before feeding into ViT modules.

16 Paper: Object Detection in Real-Time Video Surveillance Using Attention-Based Transformer-YOLOv8

Reference: DOI: 10.1016/j.aej.2025.01.032

Main idea: The paper integrates attention and Transformer mechanisms into YOLOv8 to enhance real-time object detection in video surveillance scenarios.

Modifications to YOLOv8:

- Combined the YOLOv8 backbone with an **attention mechanism module (IPA: Integrated Perceptual Attention)** to emphasize regions of interest.
- Introduced an **MSCCR module (Multiscale Spatial Channel Reconstruction)** to rebuild features across scales without significantly increasing parameter count.
- Replaced the standard detection head with a **Transformer-based detection head** to leverage long-range dependencies and global context.

17 Paper: ViT-YOLO — Transformer-Based YOLO for Object Detection

Reference: ICCV Workshop 2021

Main idea: ViT-YOLO integrates Transformer attention mechanisms into the YOLO architecture to capture global contextual information and enhance object detection accuracy, especially for complex scenes.

Modifications to YOLO:

- Replaced the standard backbone with **MHSA-Darknet**, embedding multi-head self-attention layers into CSP-Darknet for global context modeling.
- Substituted the PANet neck with a **weighted Bi-directional Feature Pyramid Network (BiFPN)** to improve cross-scale feature fusion efficiency.
- Implemented **test-time augmentation (TTA)** and **Weighted Boxes Fusion (WBF)** in the inference stage for enhanced detection robustness.

18 Paper: Enhancing Small Object Detection in Remote Sensing Images Using Mixed Local Channel Attention with YOLOv8

Reference: DOI: 10.5281/zenodo.10986298

Main idea: The paper proposes adding a Mixed Local Channel Attention (MLCA) mechanism into YOLOv8 to better detect small objects in remote sensing images.

Modifications to YOLOv8:

- Incorporated the **MLCA (Mixed Local Channel Attention)** module, which combines local and global channel attention and spatial attention, to enhance the representational power.
- Applied both **local average pooling (LAP)** and **global average pooling (GAP)** within the attention mechanism to capture fine details and global context simultaneously.
- Replaced or augmented relevant feature blocks to integrate MLCA without significantly increasing model complexity.

19 Paper: BGF-YOLO — Enhanced YOLOv8 with Multiscale Attentional Feature Fusion for Brain Tumor Detection

Reference: DOI: 10.1007/978-3-031-72111-3_4

Modifications to YOLOv8:

- Introduced **Bi-level Routing Attention (BRA)** to focus on salient features and route information more effectively.
- Replaced the neck with a **Generalized Feature Pyramid Network (GFPN)** for enhanced multiscale feature fusion.
- Added a **fourth detection head** to extend detection across more scales.

20 Paper: SUD-YOLO — Semi-Supervised Method for Underwater Object Detection Based on Improved YOLOv8

Reference: DOI: 10.3390/app15031065

Main idea: The paper introduces SUD-YOLO, which integrates a semi-supervised learning strategy and architectural enhancements to YOLOv8 for underwater detection.

Modifications to YOLOv8:

- Adopted a **Mean Teacher** framework to combine labeled and unlabeled data, using a teacher model to guide pseudo-label generation.
- Developed a **multi-scale pseudo-label enhancement module** to mitigate low-quality pseudo labeling.
- Added a **receptive-field attention module** merging local spatial features to improve feature extraction under underwater conditions.

- Designed a **lightweight detection head** based on task alignment and shared convolutions to preserve detail while reducing parameter count.