Meta-YOLO: Metadata-Guided Real-Time Object Detector in Aerial Imagery

Deukryeol Yoon, Seonghak Kim, Young Hwa Sung, Jinho Jung
{deukryeol, hakk35, yhsungadd, jinhojung}@korea.kr

Agency for Defense Development

WACV 2026 TUCSON, AZ 3/6 – 3/10

Summary

Motivations

- **[Challenge]** In aerial object detection, object scales vary widely in aerial imagery, posing a challenge for precise detection.
- **[Intuition]** Standardized platform metadata (e.g., altitude, slant range, FOV) contains crucial spatial hints to estimate object scale.

Contributions

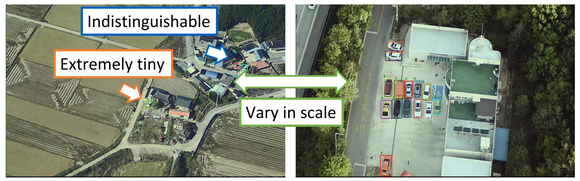
- **[Method]** We propose a Meta-YOLO that integrates metadata into the detection process to leverage metadata-driven spatial hints.
- **[Performance]** We improve detection accuracy while maintaining real-time inference speed, especially in lightweight regimes.
- **[Insight]** We provide practical insights on how standardized platform metadata can resolve scale variance.

What Happens in UAV Detectors?

Aerial Object Detection

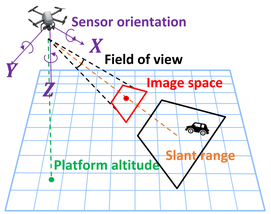
- **[Complex Object]** Objects are extremely tiny, vary widely in scale, and blend into backgrounds.
- **[Limited Capacity]** Detectors on UAV platforms operate within strict computational constraints.

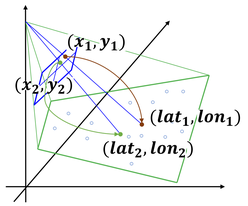
→ **[Problem]** Precise detection remains a significant challenge.



Metadata-driven Context

- **[Platform Metadata]** UAVs record the environmental and flight context of the imagery.
- **[Sensor Modeling]** We can map image to geographic coordinates.

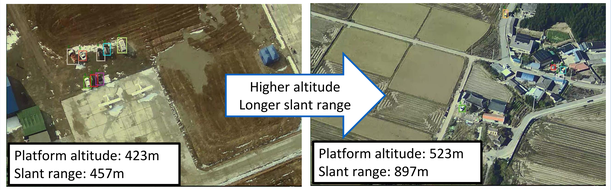


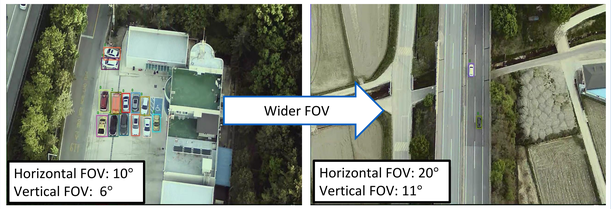


Observation

Relationship between Aerial Imagery and Metadata

- **[Distance Effect]** The higher the altitude (and thus the longer slant range), the smaller the object scale.
- **[Zoom Effect]** The wider the FOV, the smaller the object scale.

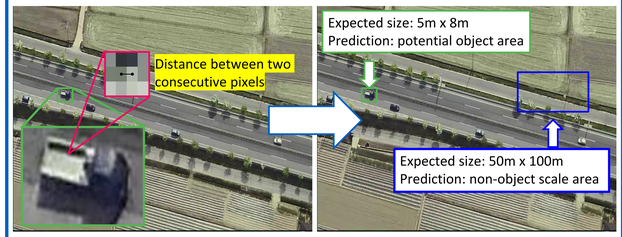




Geometric Prior for Object Scale

- **[Physical Distance Estimation]** Coordinated mapping yields the precise physical distance between two consequtive pixels.
- **[Bounding Box Area Determination]** It follows the actual physical ground area covered by each bounding box.

→ **[Observation]** Metadata provides a geometric prior for estimating object size.



Main Question

- **[Key challenge]** How can we efficiently leverage platform metadata as a geometric prior into the detection process?

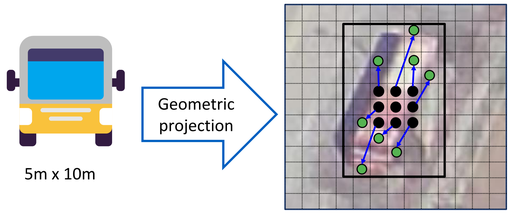
Proposed Method: Meta-YOLO

Meta-YOLO

- Enhance spatial feature representation by aligning receptive fields to object scale through geometric priors via metadata.

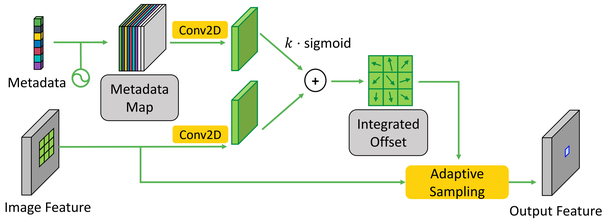
Key Idea

- Metadata provides a geometric prior of object scale in image plane.
- During the convolution operation, we deform sampling points via object-informed spatial guidance.



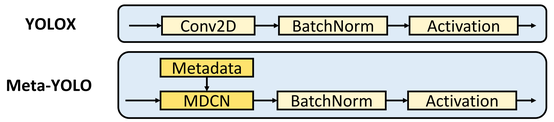
Metadata-Guided Deformable Convolution Network (MDCN)

- **Step 1.** Broadcast metadata and the positional encoding into a metadata map.
- **Step 2.** Generate image and metadata-driven feature maps individually via separate branches.
- **Step 3.** Integrate these feature maps to derive geometric offsets.
- **Step 4.** Apply adaptive sampling using derived offsets through the deformable convolution operator.



Architecture

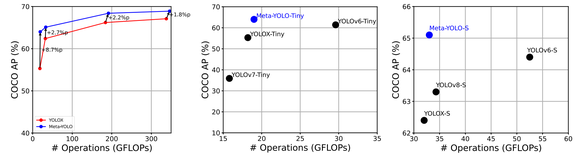
- Built on YOLOX, replace convolution blocks with MDCN layers within the bottleneck layers in feature pyramid network.



Experiments

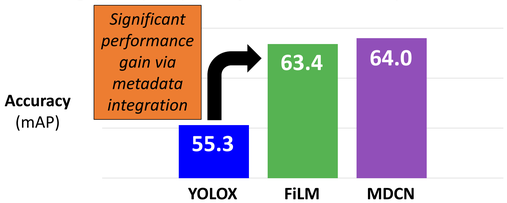
Performance Superiority

- Boost performance significantly (+15.7%) with only a marginal computation.
- Outperform other lightweight detectors, demonstrating superior accuracy and efficiency.



Architecture Effectiveness

- Compare MDCN with alternative metadata modulation method.
  - FiLM (Global Modulation): Prove that metadata is effective even with a simple scaling/shifting strategy.
  - MDCN (Spatial Modulation): Maximize the potential of metadata by enabling spatially-adaptive adjustments to the receptive field.



Qualitative Results

- Localize objects successfully across diverse environments.