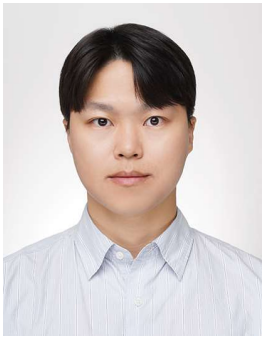
WACV 2026
Tucson, Arizona • Mar 6 - 10, 2026

ADD
Agency for Defense Development

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

Deukryeol Yoon Seonghak Kim Young Hwa Sung Jinho Jung

Defense AI R&D Institute
Agency for Defense Development

What Happens in UAV Detectors?

- Aerial Object Detection (AOD)
  - Infer the bounding box and class of objects from bird's-eye view imagery.

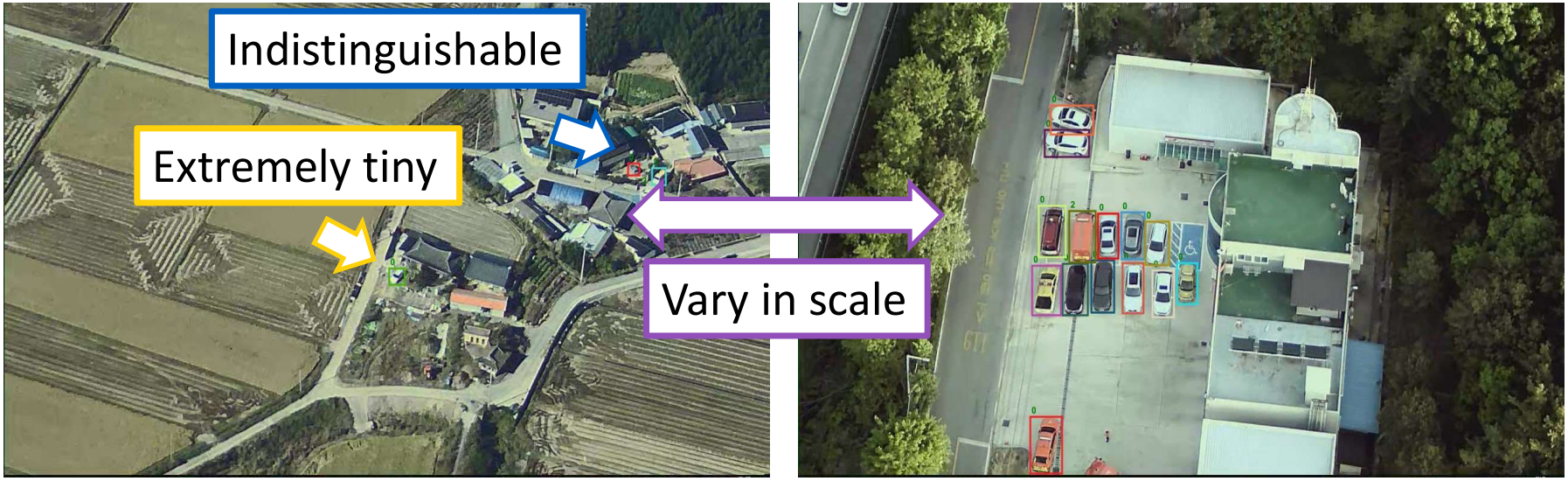
Aerial Imagery
(RGB image)

UAV Detectors

Detections
(bbox, class, conf)

What Happens in UAV Detectors?

- Precise AOD remains a challenge due to several reasons.
  - **Complex Objects:** Objects are extremely tiny, indistinguishable from background, and vary widely in scale.



What Happens in UAV Detectors?

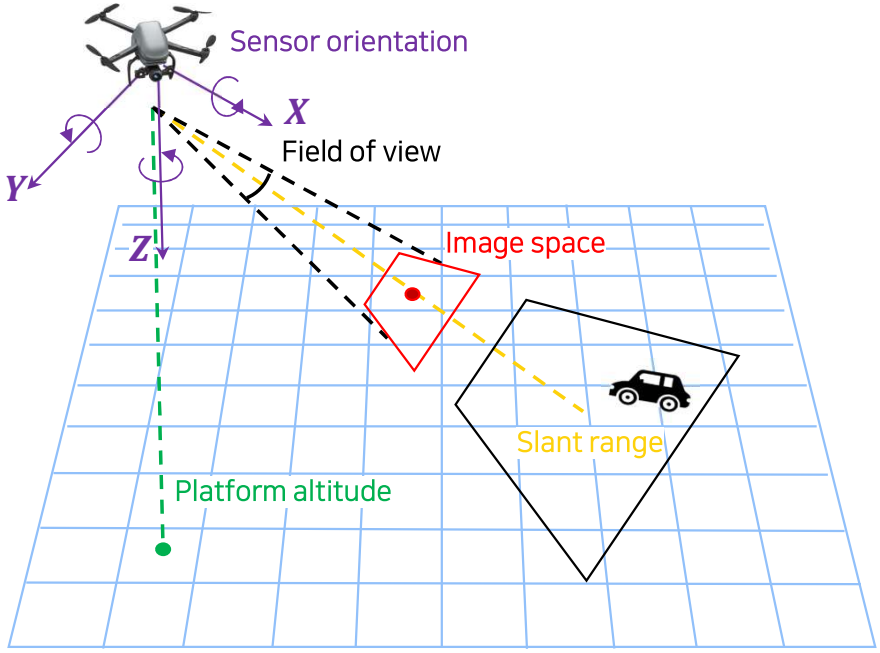
- Precise AOD remains a challenge due to several reasons.
  - **Complex Objects:** Objects are extremely tiny, indistinguishable from background, and vary widely in scale.
  - **Limited Resources:** Detectors on UAV platforms operate within strict computational constraints.

Aerial Imagery
(RGB image)

Detections
(bbox, class, conf)

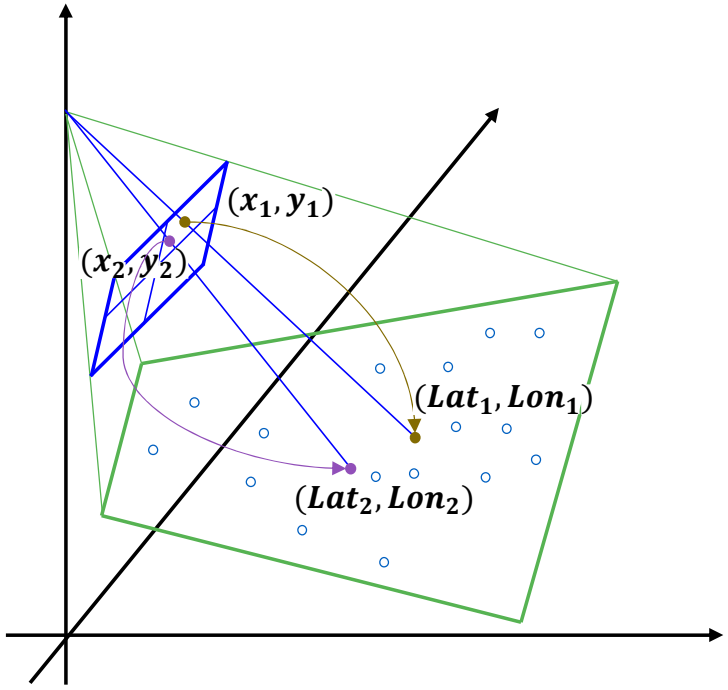
Platform Metadata in Aerial Systems

- Metadata is contextual information recorded during image acquisition.
  - platform altitude, slant range, sensor orientation, field of view (FOV), etc.



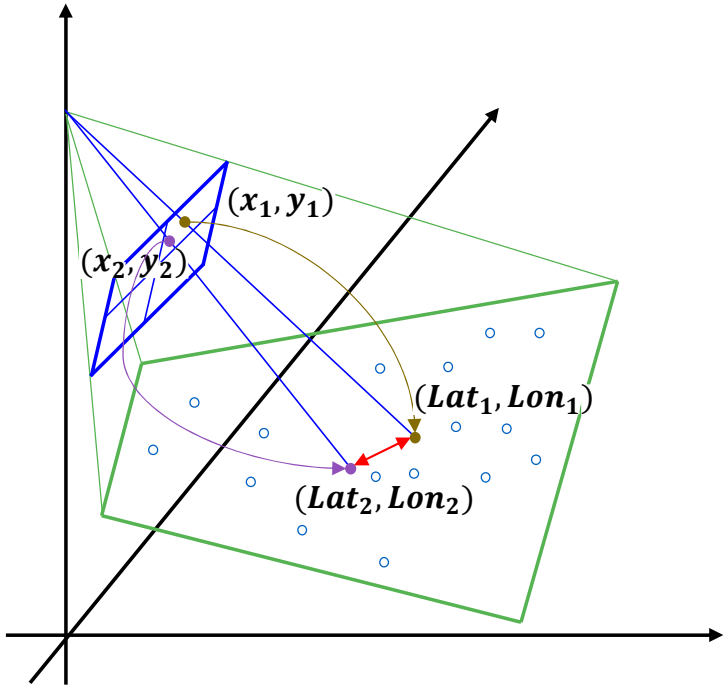
Geometric Prior via Platform Metadata

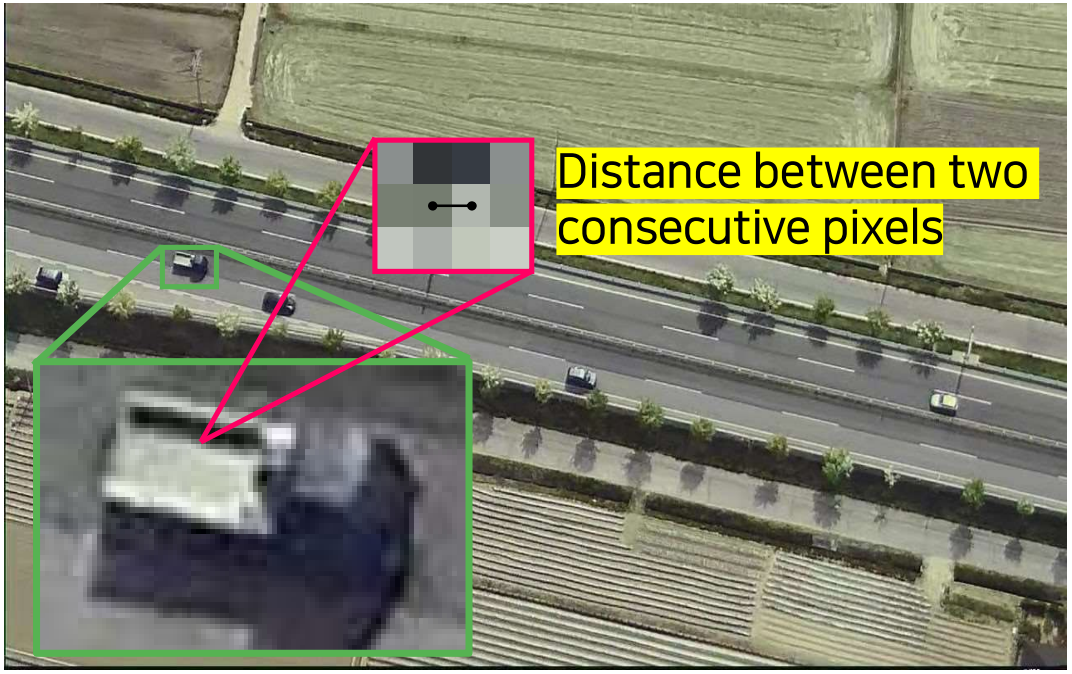
- Metadata is used to **map** image coordinates (x, y) to geographic coordinates (Lat, Lon) through sensor modeling.



Geometric Prior via Platform Metadata

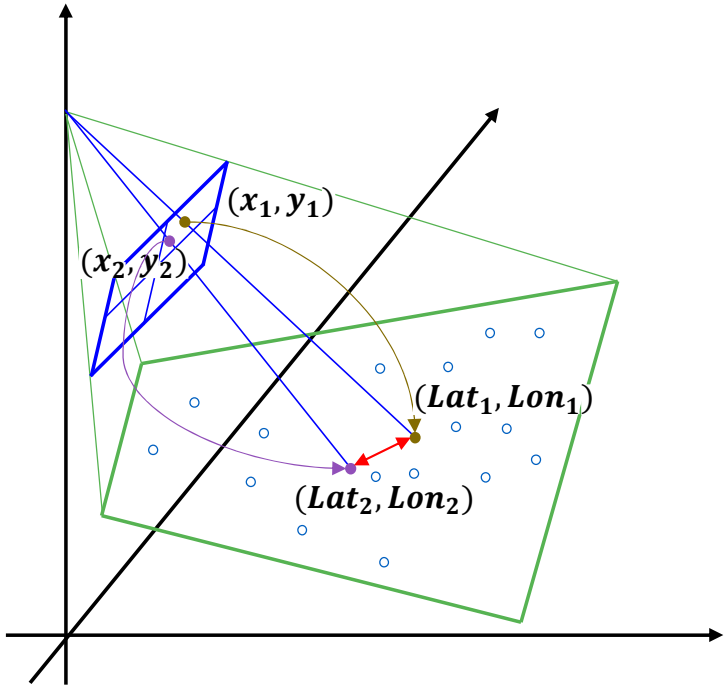
- This coordinate mapping allows us to derive the Ground Sample Distance (GSD), which represents the physical distance between consecutive pixels.

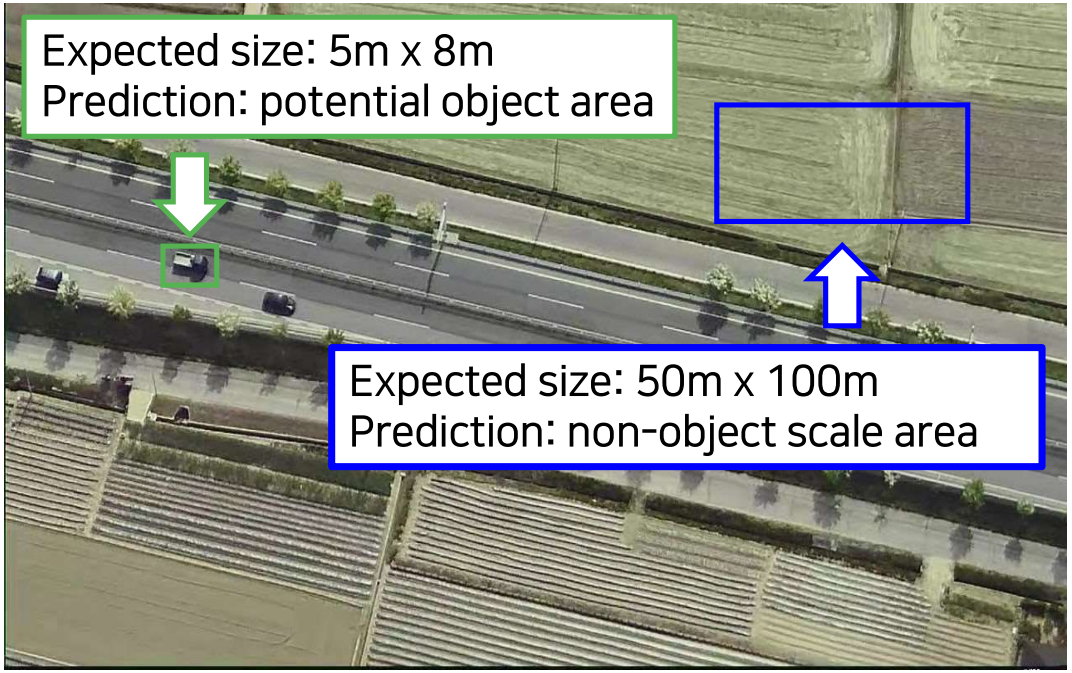




Geometric Prior via Platform Metadata

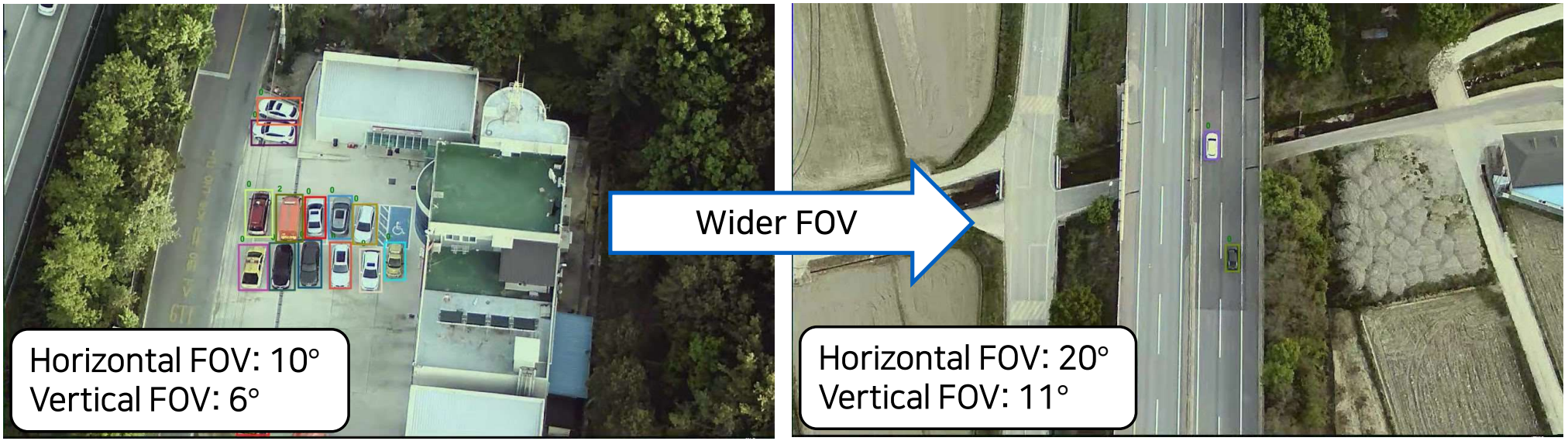
- Finally, metadata determines the actual ground area covered by each bounding box, enabling precise scale estimation.





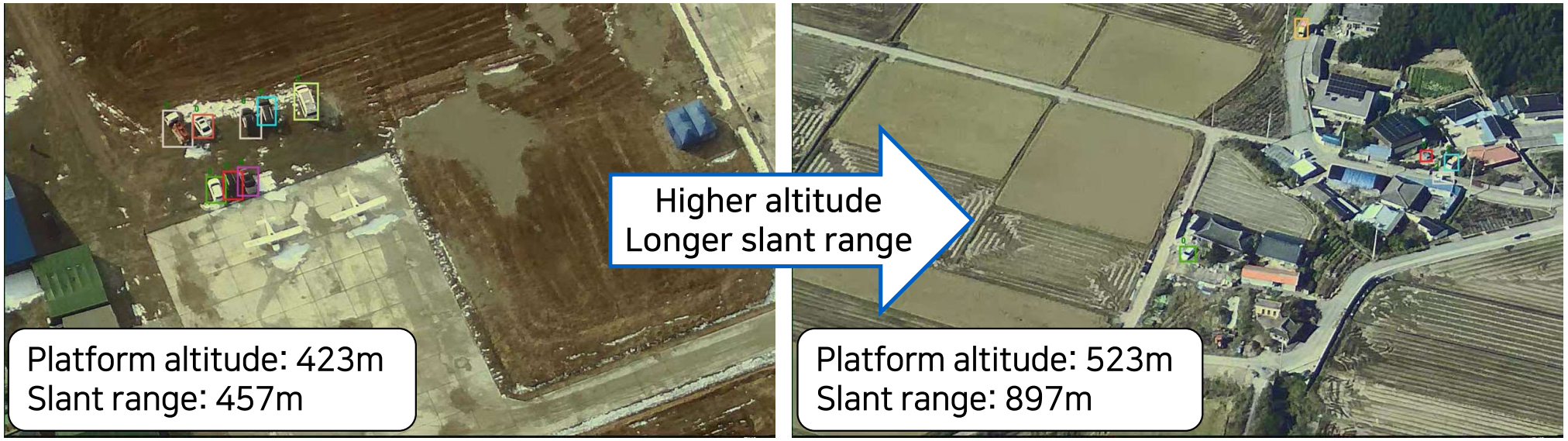
Dataset Analysis: (1) Zoom Effect

- The **wider** the FOV, the **smaller** the object appears.



Dataset Analysis: (2) Distance Effect

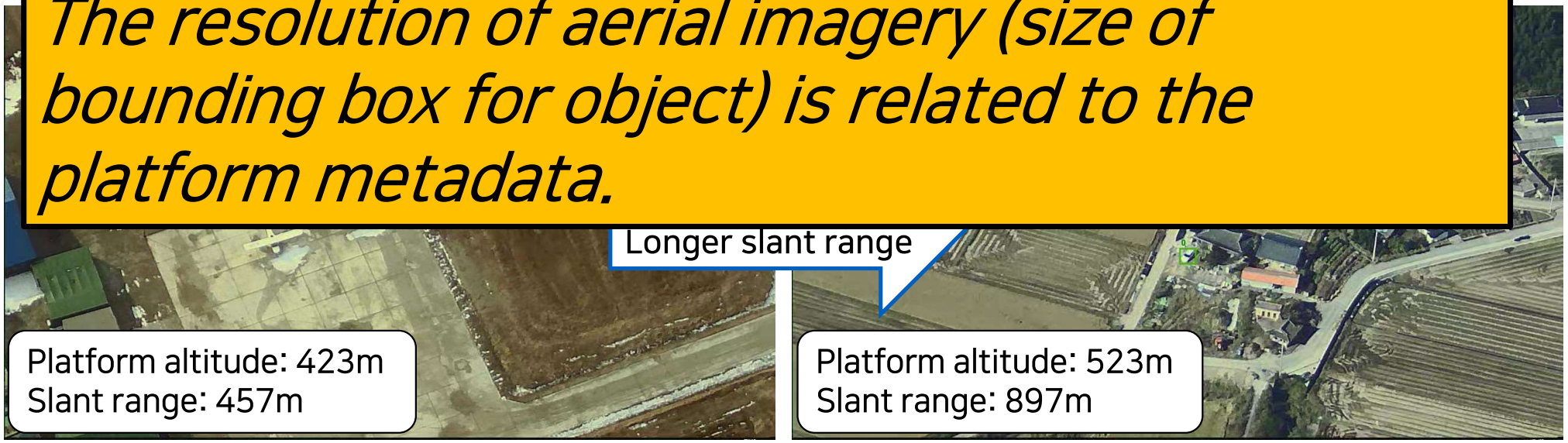
- The **higher** the altitude, the **smaller** the object footprint.
- The **longer** the slant range, the **smaller** the object scale.



Dataset Analysis: (2) Distance Effect

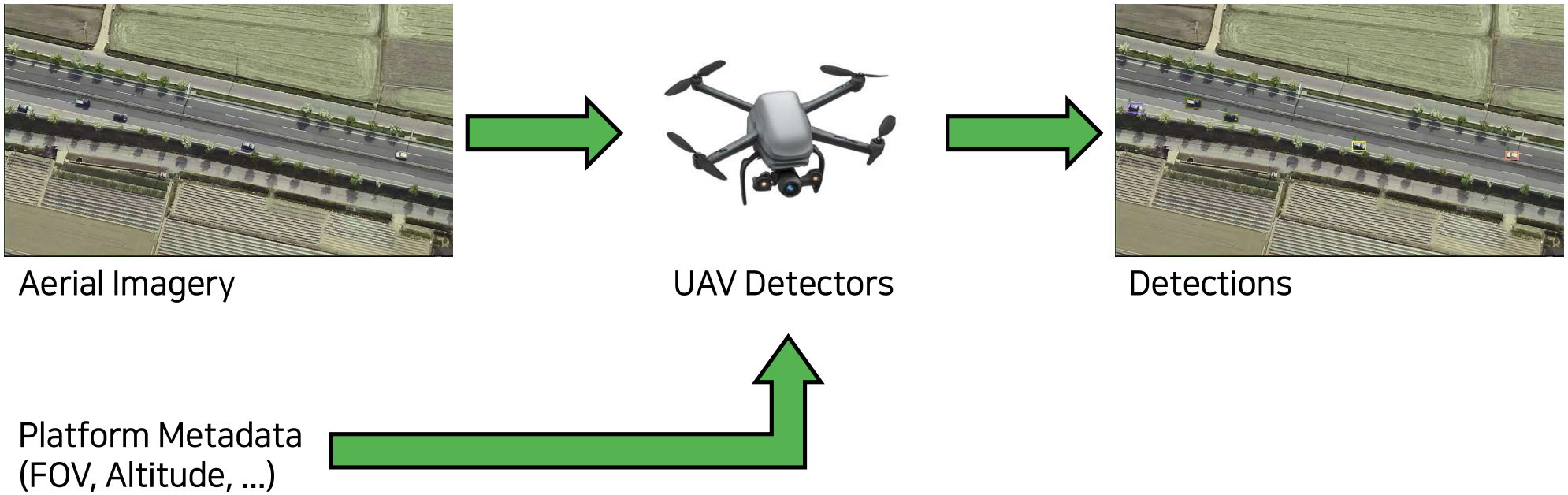
- The **higher** the altitude, the **smaller** the object footprint.
- The **longer** the slant range, the **smaller** the object scale.

The resolution of aerial imagery (size of bounding box for object) is related to the platform metadata.



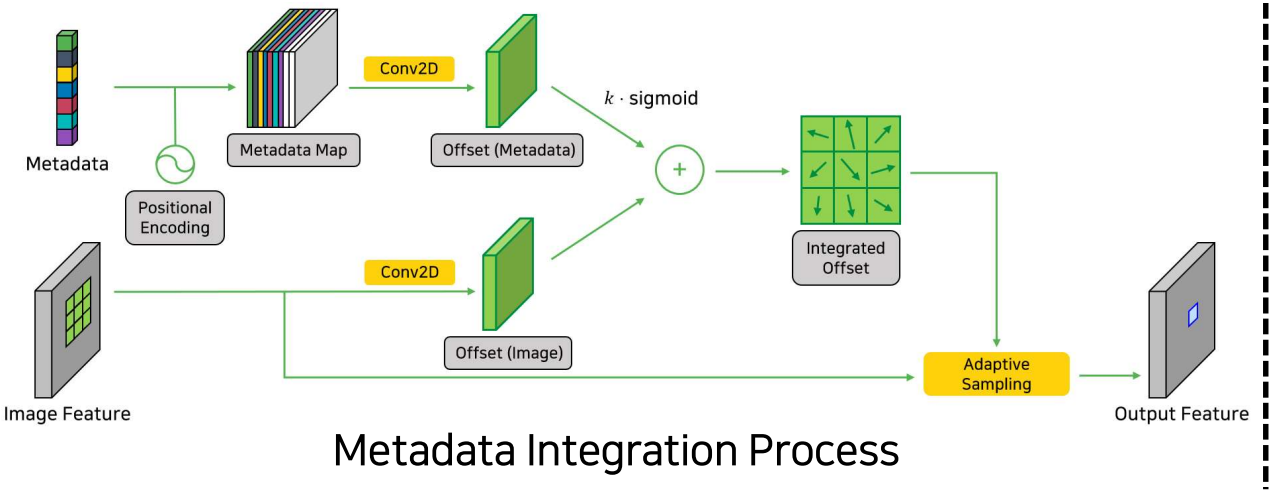
Geometric Prior for Object Detection via Metadata

- Main Question: How can we **leverage platform metadata** as a geometric prior to overcome the challenges of AOD?

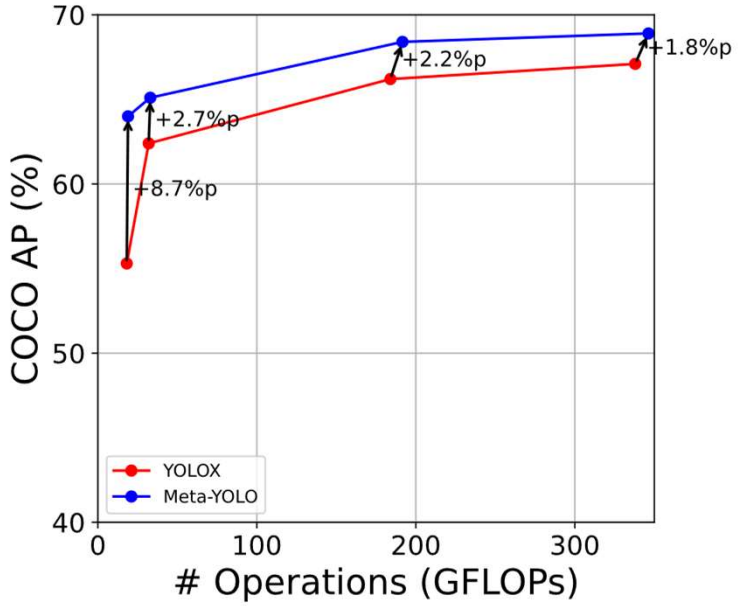


Proposed Method: Meta-YOLO

- Meta-YOLO: Metadata-Guided Real-Time Object Detector in Aerial Imagery
  - Integrate metadata into the detection process efficiently.
  - Enhance spatial feature representation by **aligning receptive fields to object scale** through geometric priors via metadata.
  - Boost the detection performance with minimal computational overhead.

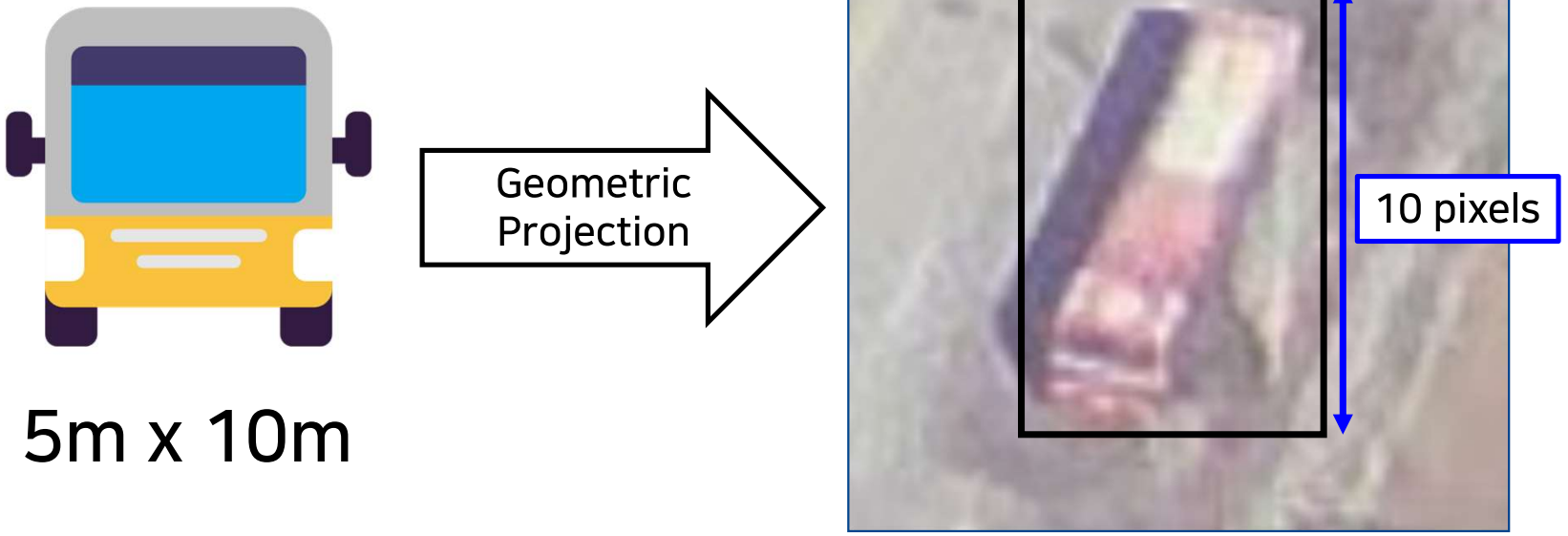


Metadata Integration Process



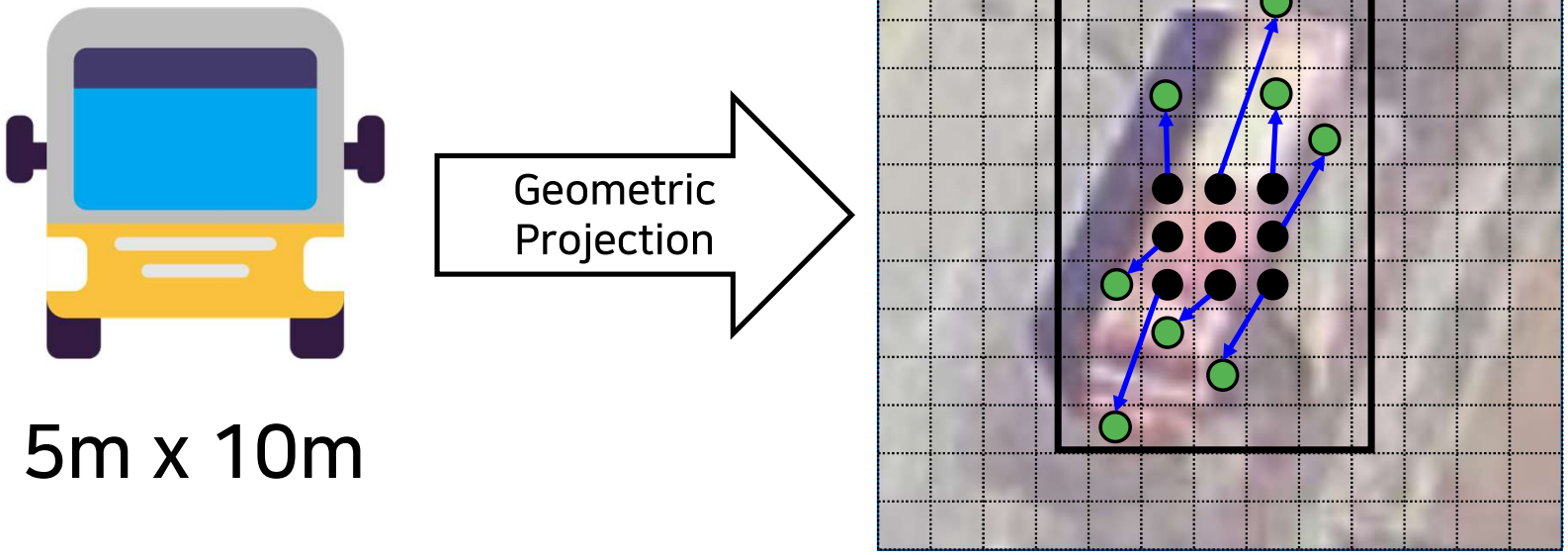
Intuition

- Metadata provides a **geometric prior** of object scale in the image plane.



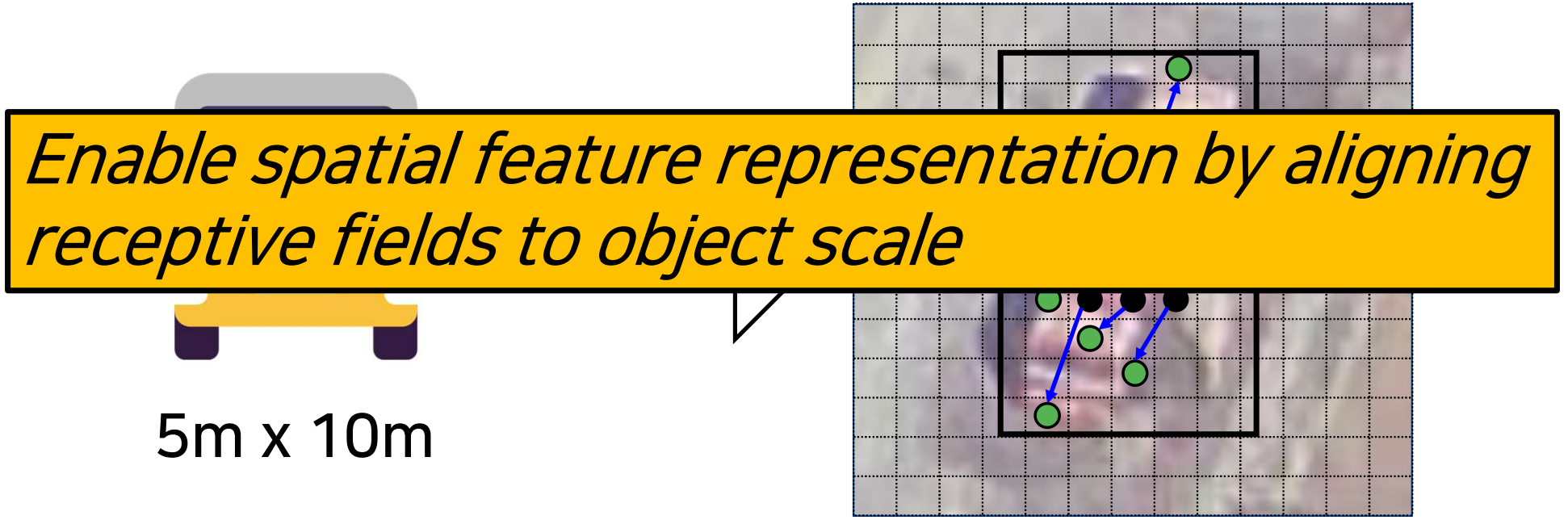
Intuition

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



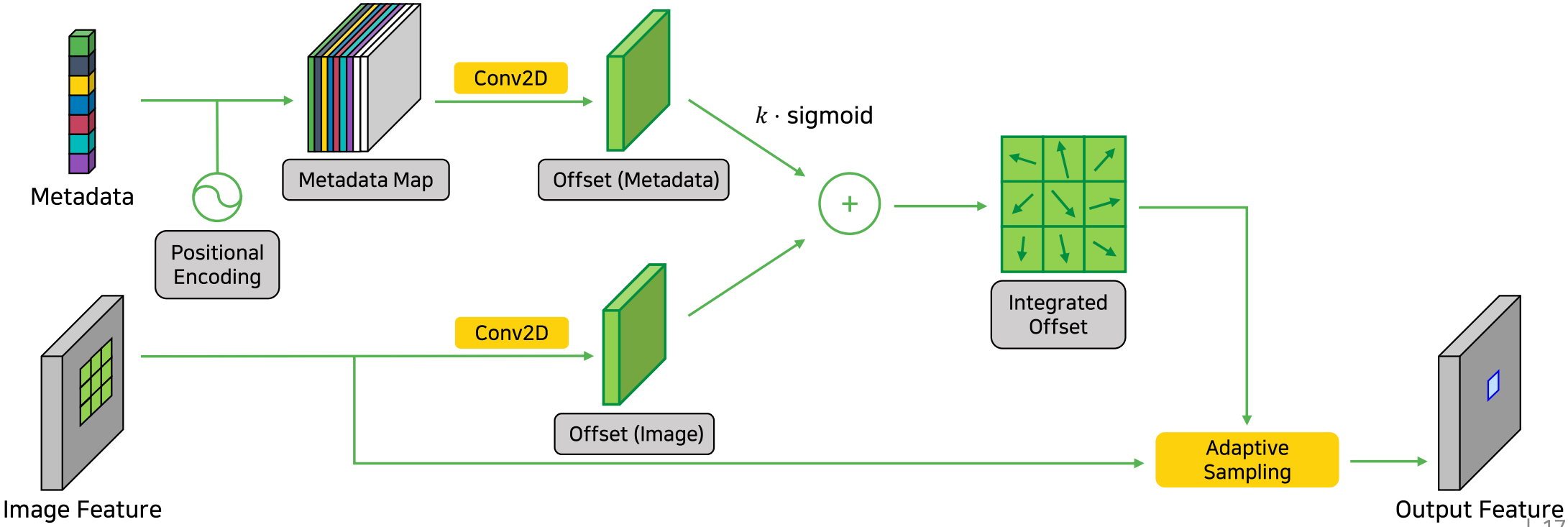
Intuition

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



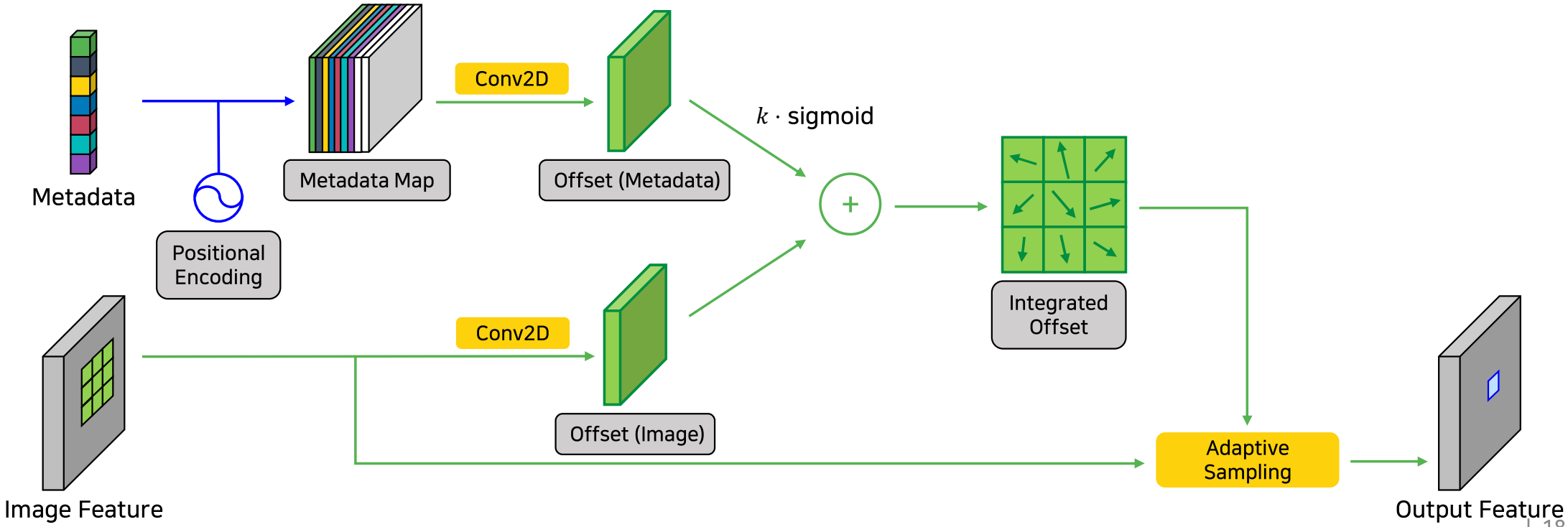
Metadata-Guided Deformable Convolution Network (MDCN)

- MDCN is a core module of Meta-YOLO.
  - Extension of Deformable Convolution Network v2 (DCNv2).
  - Integrate metadata into convolution process to **adaptively align the receptive field** with the object scale.



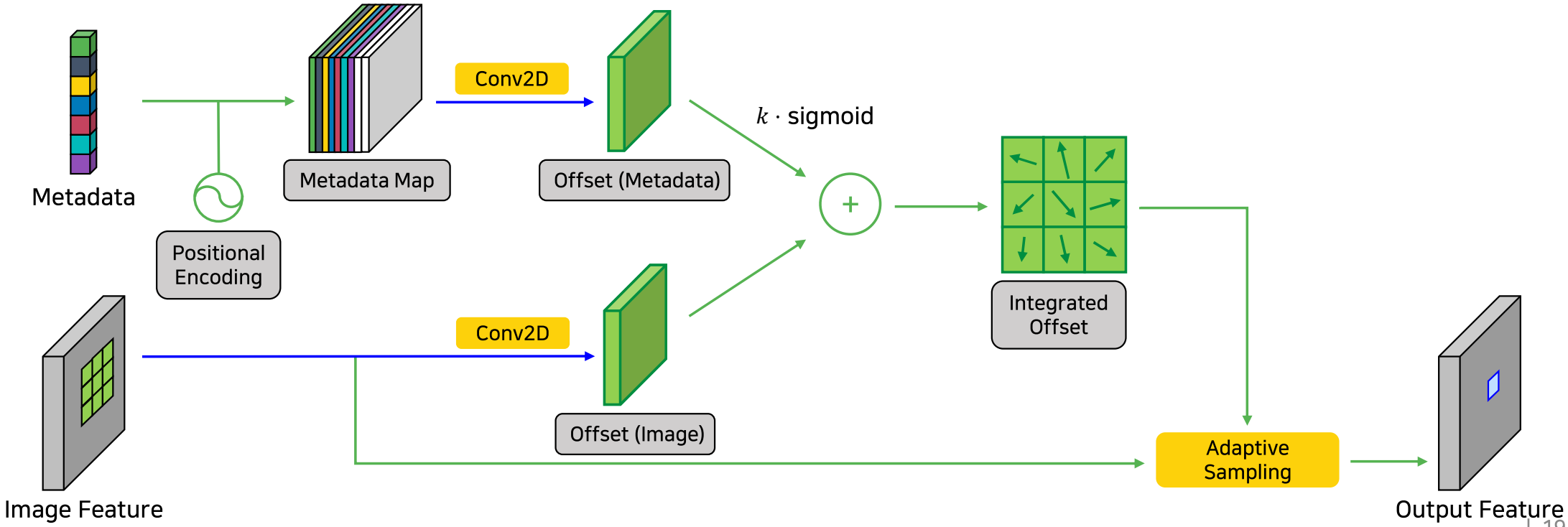
Metadata-Guided Deformable Convolution Network (MDCN)

- **Step 1.** Generate a spatial metadata map by broadcasting the metadata and appending positional encodings.



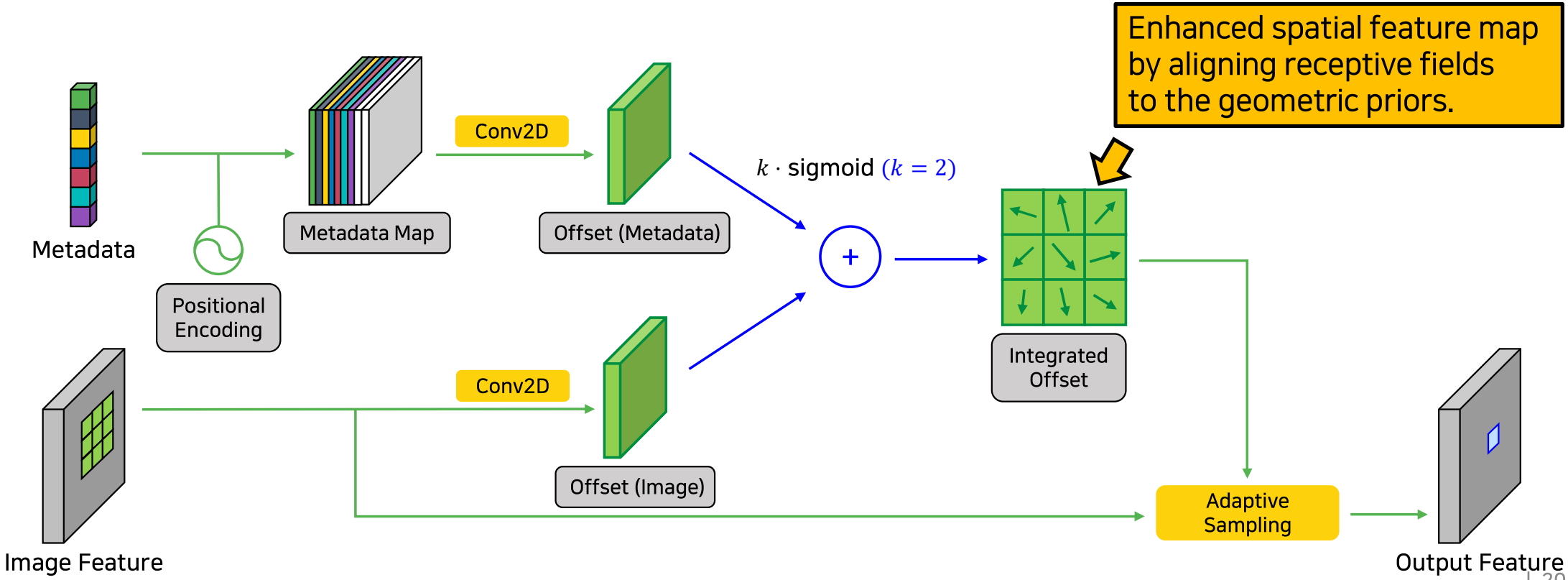
Metadata-Guided Deformable Convolution Network (MDCN)

- **Step 2.** Generate image and metadata-driven feature maps individually via separate branches.



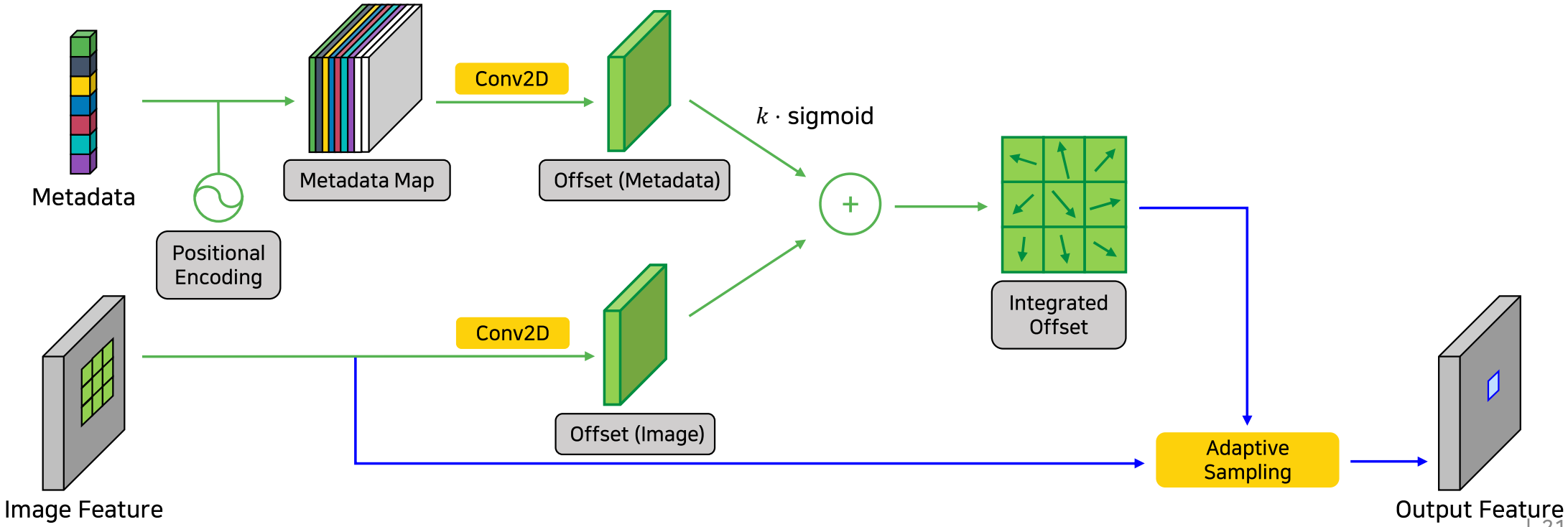
Metadata-Guided Deformable Convolution Network (MDCN)

- **Step 3.** Integrate image and metadata-driven features to derive geometric offsets, constrained by $k \cdot sigmoid$ activation.



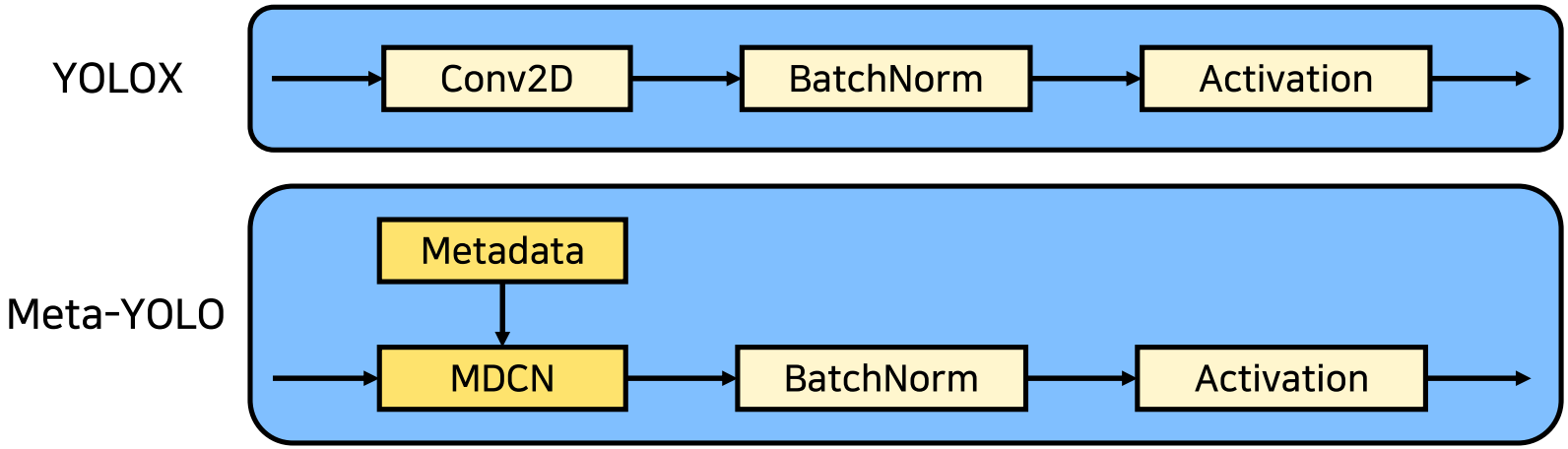
Metadata-Guided Deformable Convolution Network (MDCN)

- **Step 4.** Apply adaptive sampling using the derived offsets through the DCNv2 operator.



Meta-YOLO Architecture

- The Meta-YOLO architecture built on YOLOX.
- YOLOX → Meta-YOLO
  - Convolution blocks in the bottlenecks of the Feature Pyramid Network (FPN) are replaced with MDCN layers.



Experimental Questions

- RQ1. Performance Superiority: Does Meta-YOLO outperform recent detectors in lightweight regimes?
- RQ2. Architecture Effectiveness: Is MDCN superior to the existing modulation strategy?

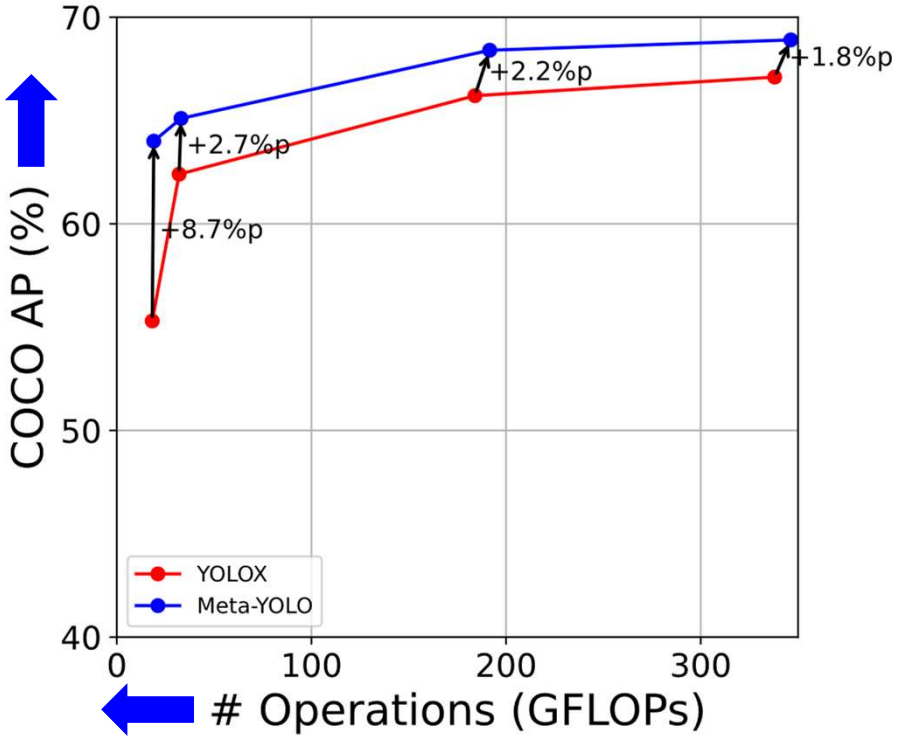
Performance Superiority

RQ1. Does Meta-YOLO outperform recent detectors in lightweight regimes?

- Experiment 1
  - Compare Meta-YOLO against the YOLOX baseline and recent state-of-the-art lightweight detectors.

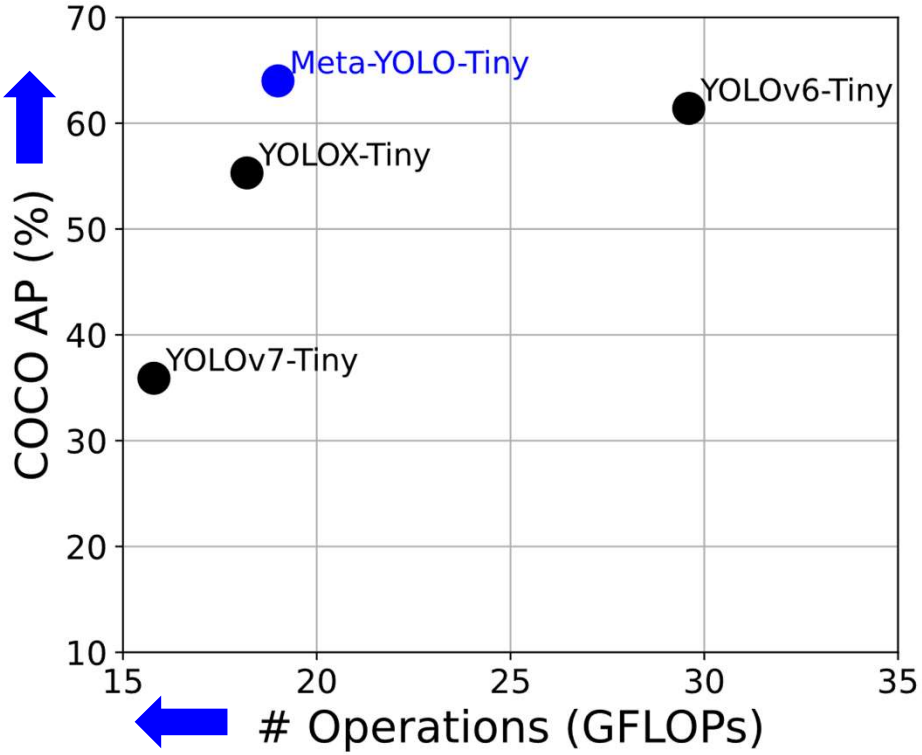
Performance Superiority

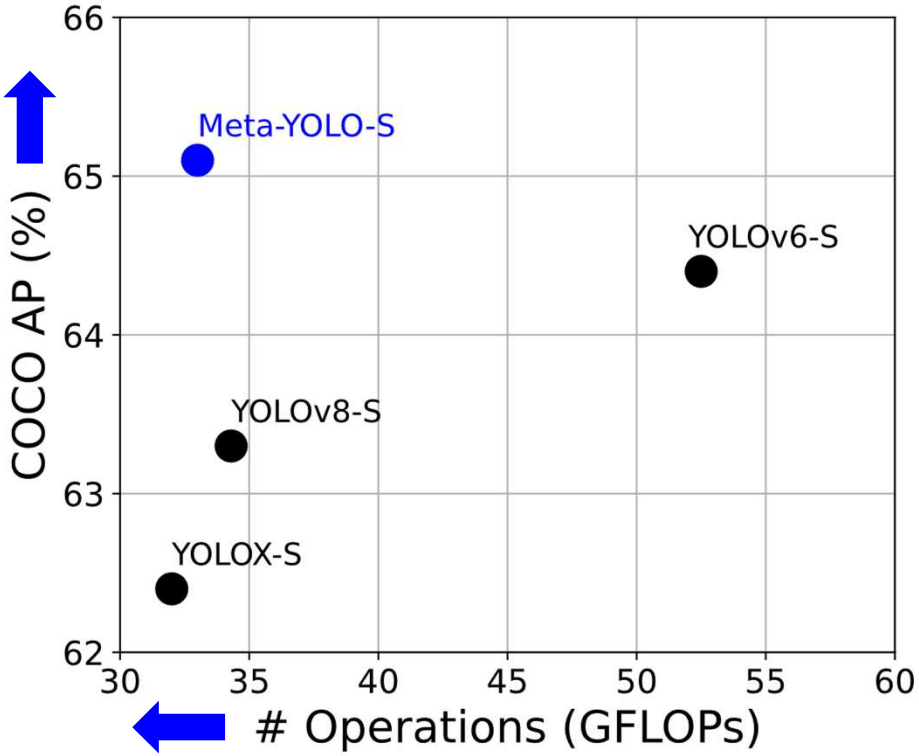
- A1. Meta-YOLO significantly **boosts performance** with **only a marginal computational overhead.**



Performance Superiority

- A1. Meta-YOLO shows **superior performance** over other lightweight detectors.



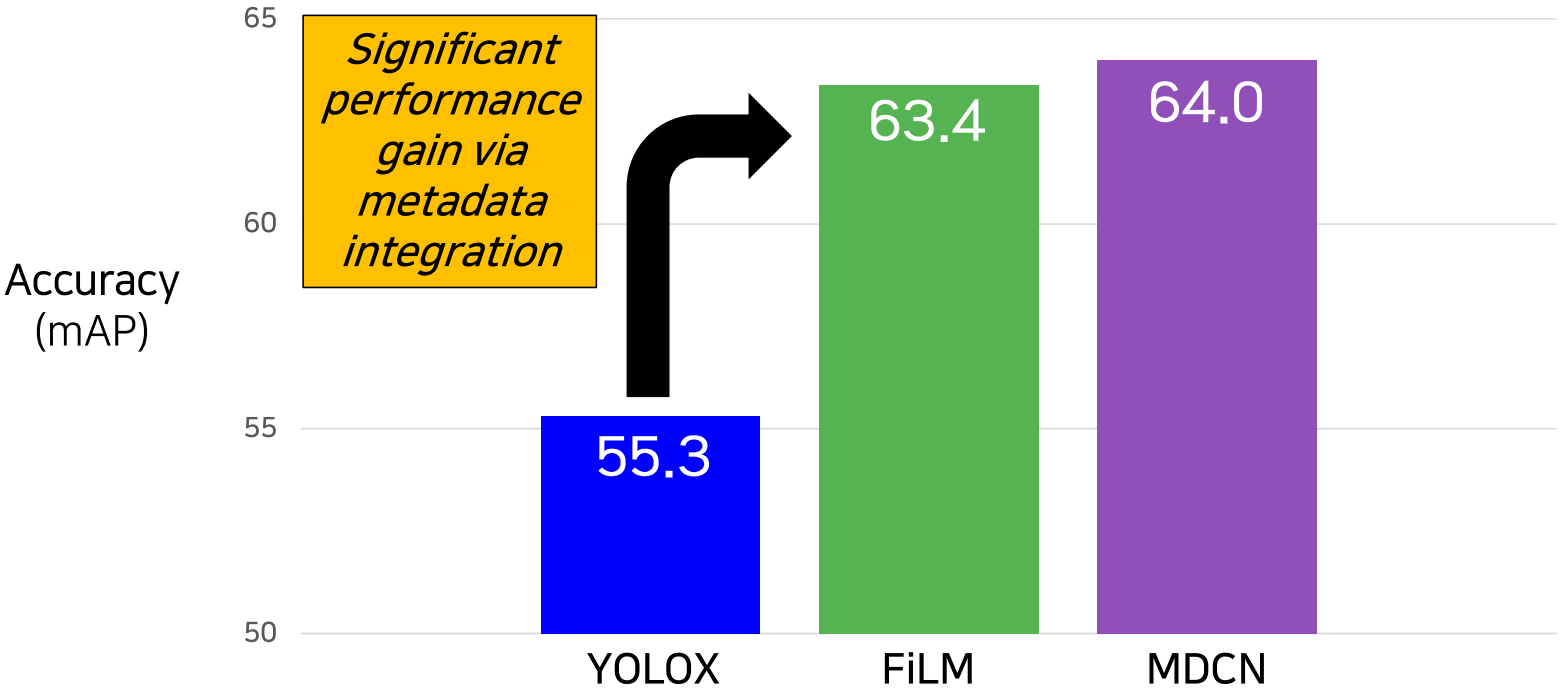


RQ2. Is MDCN superior to the existing modulation strategy?

- Experiment 2
  - Compare MDCN with alternative metadata modulation method (FiLM).
  - Evaluate the impact of global vs. spatially-adaptive modulation.

Architecture Effectiveness

- A2. MDCN provides the **most effective** metadata integration.
  - **FiLM (Global Modulation):** Demonstrates that metadata is **effective even with a simple scaling/shifting strategy.**
  - **MDCN (Spatial Modulation):** Maximizes the potential of metadata by enabling **spatially-adaptive adjustments** to the receptive field.



Qualitative Results

- Meta-YOLO **successfully localizes** objects across diverse environments.

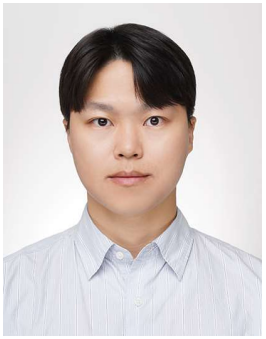
Conclusion

- Research Question
  - How can we **effectively leverage platform metadata** as a geometric prior into the detection process?
- Method
  - Introduce **Meta-YOLO**, which enhances spatial feature representation by **aligning receptive fields to object scale** through geometric priors via metadata.
- Key Outcomes
  - **Powerful Performance:** Significantly improves the detection accuracy.
  - **High Efficiency:** Maintains real-time inference speed suitable for aerial platforms.

WACV 2026
Tucson, Arizona • Mar 6 - 10, 2026

ADD
Agency for Defense Development

Thank you for listening!

Deukryeol Yoon Seonghak Kim Young Hwa Sung Jinho Jung

Defense AI R&D Institute
Agency for Defense Development