The research described in the paper **"Combining Visible and Infrared Spectrum Imagery using Machine Learning for Small Unmanned Aerial System Detection"** proposes a method that leverages both visible spectrum (RGB) and long-wave infrared (LWIR) imagery for the detection of small unmanned aerial systems (sUAS). Here's an overview of the methodology used:

---

1. Hardware Setup

- **Cameras Used:**
  - **Visible Spectrum (RGB):** FLIR CM3-U3-50S5C-CS Chameleon3 USB3 camera with a 3.5mm lens.
  - **Infrared Spectrum (LWIR):** FLIR BOSON 640 camera with a 4.9mm lens.
- The two cameras are mounted side-by-side on a custom mount system to ensure alignment and synchronization.
- Both cameras are calibrated and synchronized using the **Robot Operating System (ROS)** , operating at **30 Hz** .

---

2. Data Collection

- Two drones were used for data collection:
  - DJI Mavic Pro
  - DJI Phantom 4
- Data was collected under various conditions:
  - Different times of day (e.g., sunrise/sunset)
  - Weather conditions (e.g., cloudy vs. sunny)
  - Scenarios such as flying above/below treelines, against bright light sources, and near heat-emitting objects like streetlights.
- Total recorded data: **30 minutes at 30 FPS** , resulting in **52,800 frames per camera** .
- Only **1,275 frames** were manually annotated with bounding boxes around the sUAS.

---

3. Image Fusion

- The RGB and LWIR images were aligned using **homography transformation** , which maps pixels from one image space to another.
- Pixel-level fusion was performed by computing an **equally weighted sum** of corresponding pixels from the RGB and LWIR images.
- This results in a fused image that combines:

    - The **high contrast** of LWIR for distinguishing the sUAS from backgrounds,
    - The **higher resolution** of RGB for better spatial detail.

---

4. Machine Learning Model

- **YOLOv3 (You Only Look Once version 3)** was used as the object detection model.
- Three models were trained separately:
    1. **RGB-only model** : Trained on RGB images.
    2. **LWIR-only model** : Trained on LWIR images.
    3. **Fused (LWIR+RGB) model** : Trained on the combined images.
- No changes were made to the YOLO architecture since the input format remained consistent (3 channels), regardless of whether the input was RGB, LWIR (converted to grayscale or pseudo-color), or fused.

---

5. Performance Metrics

The performance was evaluated based on:

- Detection Rate (DR):

$$DR = \frac{TP}{TP + FN}$$

- False Alarm Rate (FAR):

$$FAR = \frac{FP}{TP + FP}$$

- TP = True Positives
- FN = False Negatives
- FP = False Positives

Models were tested across different confidence thresholds: **25%, 50%, 75%, and 90%** .

---

6. Scenarios Tested

Four key scenarios were evaluated:

1. **Drone crossing above and below horizon/treeline**
2. **Drone flying directly against sunlight during sunrise**
3. **Presence of other heat sources (e.g., streetlights)**
4. **Long-range detection (~300 meters away)**

---

7. Key Results

- At **50% confidence threshold** , the proposed **fused model** achieved:
  - **Detection Rate (DR): 71.2 ± 8.3%**
    - **69% improvement over LWIR alone**
    - **30.4% improvement over RGB alone**
  - **False Alarm Rate (FAR): 2.7 ± 2.6%**
    - **74.1% reduction compared to LWIR alone**
    - **47.1% reduction compared to RGB alone**

✅ Why Did This Method Work Well?

1. Complementary Sensors

- RGB provides spatial detail
- LWIR provides contrast against background
- Fused data combines both advantages

2. Robustness Across Conditions

- Sun glare? LWIR handles it.
- Low visibility? LWIR sees heat.
- Long-range detection? RGB helps with detail.

3. No Extra Complexity

- No custom CNN architecture needed
- Just fuse images before feeding into standard YOLOv3

8. Conclusion

The study demonstrates that combining visible and infrared imaging through pixel-level fusion significantly enhances the accuracy and robustness of sUAS detection. It outperforms models relying solely on either spectral modality, especially in challenging conditions like low visibility, glare, or presence of similar heat sources.

Summary of Method Steps:

1. **Capture synchronized RGB and LWIR video streams**
2. **Calibrate and align images via homography**
3. **Fuse RGB and LWIR at the pixel level**
4. **Train YOLOv3 detector on individual and fused datasets**
5. **Evaluate performance across multiple real-world scenarios**

This multi-sensor approach is promising for counter-drone systems requiring reliable detection in diverse environmental conditions.

—---

there are better and more modern models that can potentially improve performance on the task of **small unmanned aerial system (sUAS) detection using RGB + LWIR fusion** , especially compared to **YOLOv3** , which was used in the original paper [File].

Below is a breakdown of **better-suited models** for this task, based on recent advancements and research cited in the web_search content:

✅ 1. YOLOv4 and YOLOv5 / YOLOv7 / YOLOv8

Why Better?

- YOLOv4 introduced advanced techniques like:
  - **CSPDarknet53 backbone**
  - **PANet (Path Aggregation Network)** for feature fusion
  - **Mosaic data augmentation**
- These improvements enhance accuracy and robustness, especially in detecting small or low-contrast objects — ideal for sUAS detection
- .

Performance:

- In experiments comparing object detection models, **YOLOv4 achieved 99.50% precision and recall** , outperforming YOLOv3
- .
- YOLOv5 and YOLOv7 further improve speed and accuracy through better architecture scaling and training strategies.

✅ **Recommendation:** Use **YOLOv5 or YOLOv7** with fused RGB+LWIR input for improved accuracy and inference speed.

✅ 2. EfficientDet / EfficientNet-based Detectors

Why Better?

- EfficientDet uses a **compound scaling method** that uniformly scales depth, width, and resolution.
- It offers a better trade-off between **accuracy and model size** , making it suitable for embedded or real-time drone detection systems.

Comparison:

- A realistic implementation of **EfficientDet outperformed YOLOv3** on custom image detection tasks in terms of **training time, model size, and inference performance**
- .

✅ **Recommendation:** If deploying on edge devices or needing lightweight yet accurate models, try **EfficientDet-D3 or D4** .

✅ 3. YOLOv10 (or Real-Time Detectors with Efficient Feature Fusion)

Why Better?

- Recent work titled *“Efficient Feature Fusion for UAV Object Detection”*
- introduces improvements in feature pyramid networks and attention mechanisms tailored for **UAV-based object detection** .
- These models optimize how information flows across different layers, helping detect small targets like drones at long range.

✅ **Recommendation:** Look into **YOLOv10 or modified YOLO architectures with enhanced feature fusion** for better small-object detection.

✅ 4. Two-Stage Detectors: Faster R-CNN with Multi-Spectral Backbone

Why Better?

- While slower than one-stage detectors like YOLO, two-stage models offer **higher accuracy** , especially when dealing with:
  - Occlusions
  - Similar-looking false positives (e.g., birds vs. drones)
- You can combine Faster R-CNN with backbones trained on both RGB and thermal images.

✅ **Recommendation:** For high-accuracy scenarios (e.g., military or security applications), use **Faster R-CNN with ResNet + thermal fusion backbone** .

✅ 5. DeepSORT or ByteTrack for Tracking + Detection

Why Better?

- As mentioned in the paper's limitations, distinguishing **drones from birds** remains an issue.
- Adding a **tracking layer like DeepSORT or ByteTrack** allows you to analyze motion patterns over time, improving classification
- .

✅ **Recommendation:** Combine detection models with **DeepSORT or ByteTrack** for tracking-based filtering of false positives.

📊 Summary Table: Model Options for Drone Detection Using RGB+LWIR Fusion

---

📊 Summary Table: Model Options for Drone Detection Using RGB+LWIR Fusion

MODEL	TYPE	STRENGTHS	LIMITATIONS
YOLOv3	One-stage	Fast, widely supported	Lower accuracy on small/low-contrast targets
YOLOv4/v5/v7/v8	One-stage	Higher accuracy, better feature extraction	Slightly heavier than YOLOv3
EfficientDet	One-stage	Good accuracy-speed tradeoff, scalable	More complex setup
YOLOv10	One-stage	Improved feature fusion, optimized for UAVs	Less mature ecosystem
Faster R-CNN	Two-stage	High accuracy, good for similar objects	Slower inference
DeepSORT / ByteTrack	Tracker	Reduces false positives by analyzing motion	Requires post-processing integration

🔧 Practical Tips for Choosing a Model

✅ Best for Speed & Accuracy:

- **YOLOv5 or YOLOv8** with fused RGB+LWIR input
- Use Mosaic augmentation and pretrained COCO weights

✅ Best for Long-Range Detection:

- **YOLOv10** or **EfficientDet** with attention modules

✅ Best for Reducing False Positives:

- Combine any detector with **DeepSORT or ByteTrack**

✅ Best for High-Accuracy Scenarios:

- **Faster R-CNN with multi-spectral backbone**

📚 How to Get Started

1. **Use Pretrained Models:**

    - [YOLOv5 GitHub]
    - [YOLOv8 GitHub]
    - [EfficientDet PyTorch]
2. **Dataset Preparation:**
    - Align and fuse RGB + LWIR frames using homography
    - Annotate with bounding boxes using LabelImg or CVAT
3. **Training:**
    - Train with confidence threshold around 50%
    - Evaluate using **DR (Detection Rate)** and **FAR (False Alarm Rate)** as in the paper
4. **Evaluation Tools:**
    - Use Hugging Face leaderboards to find top-performing models for your specific dataset
    -

🧠 Final Thought

While **YOLOv3** was effective for the original study, **more modern models offer significant improvements in accuracy, efficiency, and robustness** . Especially for challenging conditions like glare, long-range detection, and bird-like false positives, upgrading to **YOLOv5+, EfficientDet, or integrating tracking algorithms** will likely yield better results.