[DC-03-034] Multispectral and Thermal Imagery: RGB and LWIR Sensors

Abstract

Object detection, the task of identifying and localizing objects within images or video, is a rapidly growing field that has become fundamental to numerous applications, including autonomous vehicles, surveillance systems, agricultural monitoring, and infrastructure assessment. While traditional visible-light cameras operating in the Red-Green-Blue (RGB) wavelengths excel at providing detailed color and texture information under good lighting conditions, they become ineffective in challenging conditions, such as low light, nighttime, or adverse weather. Long-Wave Infrared (LWIR) sensors complement these capabilities by detecting thermal radiation naturally emitted by objects, enabling detection regardless of ambient lighting conditions. Because of their complementary strengths, the fusion of RGB and LWIR modalities creates detection systems that maintain robust performance across diverse operational scenarios. Modern object detection leverages deep learning approaches, such as Convolutional Neural Networks (CNNs) and emerging transformer architectures, that have revolutionized the extraction of features and automatic classification of objects. This entry provides an overview of RGB and LWIR sensor technologies, neural network-based object detection methods, and their combined applications across domains, including autonomous driving, precision agriculture, infrastructure monitoring, maritime surveillance, and defense applications.

Keywords: convolutional neural networks (CNN), deep learning, multispectral, object detection, RGB, thermal imaging, transformers, YOLO (You Only Look Once) algorithm

Author & citation

Gallagher, J. E. and Oughton, E. J. (2026). Multispectral and Thermal Imagery: RGB and LWIR Sensors. The Geographic Information Science & Technology Body of Knowledge (Issue 1, 2026 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2026.1.3

Explanation

1. RGB and LWIR Sensor Technologies
2. Object Detection: Tasks and Applications
3. Object Detection Using Neural Networks
4. Sensor Comparison and Fusion Rationale
5. Dataset Availability and Automated Learning
6. Limitations and Future Directions
7. Conclusions

1. RGB and LWIR Sensor Technologies

The electromagnetic spectrum spans wavelengths from short ultraviolet to long microwave. Within this spectrum, different sensing technologies capture specific wavelength ranges for various applications. Two modalities are relevant for robust object detection. The first one is

Red-Green-Blue (RGB) sensors operating in the visible spectrum and the second is Long-Wave Infrared (LWIR) sensors operating in the thermal infrared portion.


Figure 1. The electromagnetic spectrum brooken down by sensor type with their corresponding spectrum range in nanometers (nm). Source: authors.

1.1 RGB Sensing

RGB sensors capture light in the visible spectrum (approximately 380 to 700 nanometers), producing images that closely resemble human vision (Gallagher and Oughton, 2025). These cameras are ubiquitous, appearing in smartphones, surveillance systems, and autonomous vehicles. The primary advantage of RGB sensors is their ability to capture rich color and texture information, making them well-suited for applications where visual features such as color, patterns, and shapes provide discriminative cues. Modern RGB cameras easily achieve megapixel resolutions, providing detailed images even at considerable distances. However, RGB sensors fundamentally depend on ambient lighting. In low-light conditions, at night, or in adverse weather such as fog or smoke, they struggle to provide usable image data.

1.2 LWIR Sensing

LWIR sensors operate in the 8 to 14 micrometer wavelength range, detecting thermal radiation emitted by objects rather than reflected light (Gallagher and Oughton 2025). Thermal cameras typically use microbolometer arrays, in which each pixel changes its electrical resistance when exposed to infrared radiation, creating a thermal map of the scene . The key advantage of LWIR sensors is their independence from ambient lighting. They can operate effectively in complete darkness, through light fog and smoke, and in difficult weather conditions. This makes thermal imaging invaluable for 24/7 surveillance and detection in challenging environments. However, LWIR sensors typically offer lower spatial resolution than RGB cameras, with commercial sensors often limited to 640 by 480 pixels compared to 1920 by 1080 or higher for RGB. Thermal images also tend to appear blurrier due to longer wavelengths and heat diffusion across surfaces. Additionally, LWIR sensors face challenges during thermal crossover periods around dawn and dusk, when ambient and object temperatures equalize, reducing thermal contrast (Gallagher and Oughton, 2023).

2. Object Detection: Tasks and Applications

Object detection is a fundamental challenge in computer vision, requiring systems to simultaneously classify objects and determine their spatial locations within an image (Redmon et al., 2016). Unlike image classification, which assigns a single label to an entire image, object detection must identify multiple objects of varying classes and sizes and provide precise localization using bounding boxes or segmentation masks. This capability has enabled transformative applications across numerous domains.

In autonomous driving, object detection systems must reliably identify pedestrians, vehicles, cyclists, and obstacles under varying lighting and weather conditions (Roszyk et al., 2022). Precision agriculture applications also leverage object detection for crop monitoring, pest detection, weed identification, and yield estimation, often requiring the detection of small objects across large areas (Osco et al., 2020; Osorio et al. 2020).

Infrastructure monitoring employs object detection to automate inspection of power lines, solar panels, and structural components, enabling early detection of defects and preventing costly failures (Chen et al., 2023; Lei et al., 2024). Another application is in maritime surveillance systems, where object detection is used to identify vessel locations and navigation direction across vast ocean areas for security and traffic management [8]. Defense and security, object detection applications range from perimeter surveillance to search-and-rescue operations, often requiring detection capabilities that operate in complete darkness or through obscurants such as smoke and fog (Kristo et al. 2020; Bañuls et al. 2020).

The common theme across these applications is the need for reliable detection regardless of environmental conditions. This requirement has driven the development of multispectral sensing approaches that combine RGB cameras with thermal infrared sensors, leveraging the complementary strengths of each modality to achieve robust, all-weather, day-and-night detection capabilities.

3. Object Detection Methods Using Neural Networks

Deep learning has fundamentally transformed object detection, enabling systems to automatically learn discriminative features directly from data rather than relying on hand-crafted feature extractors. Two primary neural network paradigms now dominate the field: Convolutional Neural Networks (CNNs) and transformer architectures (Sapkota et al., 2025).

3.1 Convolutional Neural Networks

CNNs process images through successive convolutional layers that apply learnable filters to detect increasingly complex features, from simple edges in early layers to complete object parts in deeper layers. CNNs for object detection are typically categorized into two-stage and single-stage detectors. Two-stage approaches, such as Faster R-CNN, first generate region proposals likely to contain objects, then classify and refine these proposals (Gallagher and Oughton, 2025). While accurate, two-stage methods can be computationally intensive.

Single-stage detectors, such as You Only Look Once (YOLO), improve efficiency and speed by performing detection in a single forward pass through the network (Redmon et al., 2016). YOLO divides the input image into a grid, with each cell responsible for predicting objects centered within it. This unified approach enables real-time detection speeds while maintaining competitive accuracy, making YOLO suitable for applications that require real-time detection. Since its introduction in 2015, YOLO has evolved through multiple versions, with YOLOv5 remaining the most widely adopted variant for multispectral applications, accounting for 33% of all modified YOLO models in a recent survey (Gallagher and Oughton, 2025).

3.2 Transformer Architectures

Initially developed for natural language processing, transformer architectures have emerged as powerful alternatives for computer vision tasks (Zhu et al., 2023; Chen et al., 2023; Fang et al., 2022). For object detection, transformer-based approaches like Detection Transformer (DETR) eliminate the need for hand-designed components like anchor boxes by treating detection as a set prediction problem (Saltik et al., 2025; Wang et al., 2025).

Transformers offer advantages for multispectral fusion, as their attention mechanisms can dynamically weight contributions from different spectral bands based on feature extraction. One such model, ViT YOLO, integrates multi-head self-attention into the YOLO backbone to retain global context information while extracting differentiated features for detection (Zhao et al., 2023). TF YOLO incorporates a transformer fusion module to integrate features adaptively between visible and infrared images, demonstrating superior performance under varying illumination conditions (Chen et al., 2023). Similarly, Dual YOLO employs attention fusion and shuffle modules to combine RGB and thermal features effectively Bao et al., 2023). These hybrid approaches leverage the efficiency of CNN backbones with the flexible fusion capabilities of transformer attention, representing a promising direction for multispectral object detection.

4. Sensor Comparison and Fusion Rationale

Research has demonstrated the dramatic performance differences between modalities under challenging conditions. In adverse weather conditions, LWIR-based object detection models can achieve a mean Average Precision (mAP) of up to 97.9%, whereas RGB-based models may drop to 19.6% under the same conditions (Kristo et al., 2020). Conversely, RGB sensors capture textures, patterns, and color variations that thermal sensors cannot detect. This complementarity motivates sensor fusion approaches that leverage both modalities. When RGB and LWIR sensors are combined, they provide redundant edge information. RGB identifies boundaries through color and brightness transitions, while LWIR detects edges via temperature gradients. This dual source approach maintains detection reliability across all illumination conditions and enhances overall system robustness (Gallagher and Oughton, 2023).

4.1 RGB and LWIR Fusion Applications

The fusion of RGB and LWIR sensors has enabled robust object detection across diverse application domains, each leveraging the complementary strengths of visible and thermal imaging. The most prevalent sensor fusion in the literature is RGB with LWIR, comprising 39% of all multispectral object detection research (Gallagher and Oughton, 2025).

Figure 2. A comparison of RGB, RBG-LWIR, and LWIR images and the unique edges that each sensor extracts. Source: authors.

4.2 Autonomous Driving and Transportation

Pedestrian and vehicle detection are critical capabilities for autonomous vehicles, where detection failures can be fatal. Multispectral approaches have demonstrated significant improvements in challenging conditions. The KAIST dataset, containing approximately 95,000 aligned RGB thermal image pairs, has become a standard benchmark for pedestrian detection research (Hwang, 2024). In one study, YOLOv4 was adapted for low-latency multispectral pedestrian detection, achieving robust performance under varying illumination conditions (Roszyk et al., 2022). MAF YOLO introduced multi-modal attention fusion, designed explicitly for pedestrian detection, demonstrating improved accuracy under complex lighting conditions (Xue et al., 2021), while MRD YOLO addressed complex

road-scene detection by optimizing multispectral feature fusion (Sun et al., 2024).

4.3 Precision Agriculture

Agricultural applications leverage multispectral sensing for crop monitoring, pest detection, and yield estimation. For example, a multispectral approach was developed for counting and geolocating citrus trees in UAV multispectral imagery, demonstrating the potential for automated orchard management (Osco et al., 2020). Another similar application was applying YOLO for weed detection in lettuce crops using multispectral images, enabling targeted herbicide application (Osorio et al., 2020). Plant disease detection has benefited from multispectral approaches, with researchers creating datasets for early detection of tomato diseases and apple scab using deep learning combined with thermal and near infrared imagery (Georgantopoulos et al., 2023; Rouš et al., 2023). YOLOv8 has also been employed with multispectral remote sensing imagery to estimate maize planting densities, supporting precision agriculture planning (Shen et al., 2024).

4.4 Infrastructure Modeling

Multispectral computer vision can also be used in infrastructure inspection, leveraging thermal imaging’s ability to detect heat anomalies that indicate equipment failure or damage (Chen et al., 2023); Stypulkowski et al., 2021; Inam et al., 2023). On study proposed methods based on YOLOv5 and multiscale data augmentation for visual inspection in electrical substations (Chen et al, 2024). Additionally, Deeplab YOLO was built to detect hot-spot defects in infrared images of solar panels, enabling predictive maintenance to prevent costly failures (Lei et al., 2024). These applications showcase the versatility of RGB and LWIR fusion for infrastructure maintenance and safety.

4.5 Maritime Surveillance and Defense

Maritime and defense applications demand multispectral detection capabilities that operate reliably under varying lighting conditions, often over extended ranges. Many of these sensors operate on different platforms and altitudes, ranging from low-altitude to space-borne. A novel study proposed BiFA YOLO for arbitrary-oriented ship detection in high-resolution SAR images, demonstrating robust detection for naval surveillance (Sun et al., 2021). Another study addressed target detection in long-range, low-quality infrared videos, reaching 95% vehicle-detection accuracy at extended ranges (Kwan and Gribben, 2024). These developments highlight the critical role of multispectral sensing for security applications where environmental conditions are complex and non-conducive for detection using RGB sensors.

5. Dataset Availability and Automated Labeling

A significant challenge in advancing RGB and LWIR fusion is the limited availability of comprehensive datasets of thermal imagery. While RGB datasets like COCO, ImageNet, and Pascal VOC contain millions of annotated images across thousands of classes, LWIR datasets remain comparatively scarce. Existing thermal datasets, such as KAIST and FLIR ADAS, focus on specific applications, such as pedestrian detection and autonomous driving, thereby limiting the development of general-purpose thermal object detection models (Hwang, 2024; Teledyne FLIR, n.d.). This scarcity stems from the higher cost of thermal sensors, niche applications, and the labor-intensive nature of manual annotation.

The Multispectral Automated Transfer Technique (MATT) offers a promising solution by leveraging the Segment Anything Model (SAM) to automatically generate annotations for LWIR images based on corresponding RGB images (Gallagher, Gogia, and Oughton, 2025). Research has demonstrated that models trained on MATT-generated datasets achieve performance within 6.7% of that of models trained on fully manual annotations, while reducing labeling time by 87.8%. This approach significantly lowers barriers to creating multispectral datasets, enabling broader research participation.

6. Limitations and Future Directions

Despite significant advances, several challenges limit current RGB and LWIR fusion approaches. Firstly, sensor alignment between RGB and thermal cameras requires precise calibration, as misalignment degrades fusion effectiveness and may introduce additional edges that do not exist. The resolution disparity between modalities complicates feature-level fusion, requiring interpolation or super-resolution techniques that may introduce artifacts. Real-time processing of multiple spectral streams also demands significant computational resources, limiting deployment on edge devices.

Future research directions include developing adaptive fusion architectures that automatically optimize spectral weighting based on environmental conditions, rather than using fixed fusion ratios. Transformer-based approaches show promise for dynamic, context-aware fusion, while synthetic data generation with generative models may address data scarcity by producing realistic multispectral training data (Zhu et al., 2023; Fang et al., 2022; Guo et al., 2024; Blythman et al., 2020). Transfer learning techniques tailored for multispectral domains could also enable models trained on limited thermal data to generalize effectively. Finally, extending fusion research beyond RGB and LWIR to incorporate additional modalities, such as near-infrared, short-wave infrared, or synthetic aperture radar, may unlock new application possibilities where current approaches are insufficient.

7. Conclusion

RGB and LWIR sensors offer complementary capabilities for a myriad of object detection applications. RGB sensors excel at capturing detailed color and texture information under good lighting, while failing in challenging visibility conditions. Conversely, LWIR sensors provide consistent detection regardless of ambient lighting but offer lower resolution and cannot capture color information. The fusion of these modalities creates detection systems that maintain robust performance across diverse environmental conditions, leveraging each technology’s strengths to compensate for the other’s weaknesses. Modern object detection approaches, including CNN-based methods such as YOLO and emerging transformer architectures, have enabled effective multispectral fusion and detection for applications ranging from autonomous driving and precision agriculture to infrastructure monitoring and defense. Understanding the fundamental principles, advantages, and limitations of these sensor technologies and detection methods provides a necessary foundation for developing effective fusion strategies tailored to specific object detection application requirements. As research addresses current challenges in dataset availability, computational efficiency, and adaptive fusion, RGB and LWIR detection systems will continue to advance toward more

robust and reliable performance in complex real-world environments.

References

Bañuls, A., Mandow, A., Vázquez-Martín, R., Morales, J., and García-Cerezo, A. (2020). Object Detection from Thermal Infrared and Visible Light Cameras in Search and Rescue Scenes," 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Abu Dhabi, United Arab Emirates, pp. 380-386.

Bao, C., Cao, J., Hao, Q., Cheng, Y., Ning, Y., & Zhao, T. (2023). Dual-YOLO Architecture from Infrared and Visible Images for Object Detection. Sensors, 23(6), 2934.

Blythman, R., Elrasad, A., O'Connell, E., Kielty, P., O'Byrne, M., and Moustafa, M. (2020). "Synthetic Thermal Image Generation for Human-Machine Interaction in Vehicles," 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), Athlone, Ireland, 2020, pp. 1-6

Chen, C., Yu, J., Lin, Y., Lai, F., Zheng, G., and Lin, Y. (2023, June). “Fire detection based on improved PP-YOLO,” SIViP, vol. 17, no. 4, pp. 1061–1067.

Chen, J., Pan, S., Chan, Y. et al. (2024). A new method based on YOLOv5 and multiscale data augmentation for visual inspection in substation. Sci Rep 14, 9362.

Chen, Y. -Y., Jhong, S. -Y., and Lo, Y. -J. (2023). Reinforcement-and-Alignment Multispectral Object Detection Using Visible–Thermal Vision Sensors in Intelligent Vehicles," in IEEE Sensors Journal, vol. 23, no. 21, pp. 26873-26886

Chen, Y., Ye, J., & Wan, X. (2023). TF-YOLO: A Transformer–Fusion-Based YOLO Detector for Multimodal Pedestrian Detection in Autonomous Driving Scenes. World Electric Vehicle Journal, 14(12), 352.

Fang, Q., Han, D. and Wang, Z. (2022). Cross-Modality Fusion Transformer for Multispectral Object Detection. SSRN.

Gallagher, J. E. and Oughton, E. J. (2025). “Surveying You Only Look Once (YOLO) Multispectral Object Detection Advancements, Applications, and Challenges,” IEEE Access, vol. 13, pp. 7366–7395.

Gallagher, J. E., Gogia, A., and Oughton, E. J. (2025). “A Multispectral Automated Transfer Technique (MATT) for Machine-Driven Image Labeling Utilizing the Segment Anything Model (SAM),” IEEE Access, vol. 13, pp. 4499–4516.

Gallagher, J.E., and Oughton, E.J. (2023). Assessing thermal imagery integration into object detection methods on air-based collection platforms. Sci Rep 13, 8491.

Georgantopoulos, P. S., Papadimitriou, D., Constantinopoulos, C., Manios, T., Daliakopoulos, I. N., and Kosmopoulos, D. (2023, August). “A Multispectral Dataset for the Detection of Tuta Absoluta and Leveillula Taurica in Tomato Plants,” Smart Agricultural Technology, vol. 4, p. 100146.

Guo, J., Gao, C., Liu, F., Meng, D., and Gao, X. (2024, March 7). “DAMSDet: Dynamic Adaptive Multispectral Detection Transformer with Competitive Query Selection and Adaptive Feature Fusion,” Mar. 07, 2024, arXiv: arXiv:2403.00326.

Hwang, S. (n.d.). KAIST Multispectral Pedestrian Detection Benchmark Leaderboard. KAIST Multispectral Pedestrian Detection Benchmark [CVPR '15].

Inam, H., Islam, N. U., Akram, M. U., & Ullah, F. (2023). Smart and Automated Infrastructure Management: A Deep Learning Approach for Crack Detection in Bridge Images. Sustainability, 15(3), 1866.

Krišto, M., vasic-Kos, M., Pobar, M. (2020). Thermal Object Detection in Difficult Weather Conditions Using YOLO," in IEEE Access, vol. 8, pp. 125459-125476.

Kwan, C. and Gribben, D. (2021). Practical Approaches to Target Detection in Long Range and Low Quality Infrared Videos. Signal & Image Processing: An International Journal (SIPIJ) Vol.12, No.3.

Lei, Y., Wang, X., An, A. et al. (2024). Deeplab-YOLO: a method for detecting hot-spot defects in infrared image PV panels by combining segmentation and detection. J Real-Time Image Proc 21, 52.

Osco, L. P., dos Santos de Arruda, M., Marcato Junior, J., Buceli da Silva, N., Ramos, A. P. M., Moryia, E., A., S., Imai, N. N., Pereira, D. P., Creste, J. E., Matsubara, E. T., Li, J., Gonçalves, W. N. (2020). A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 160:97-106.

Osorio, K., Puerto, A., Pedraza, C., Jamaica, D., & Rodríguez, L. (2020). A Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images. AgriEngineering, 2(3), 471-488.

Redmon, J. Divvala, S., R. Girshick, R., and Farhadi, A. (2016). "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788.

Roszyk, K., Nowicki, M. R., & Skrzypczyński, P. (2022). Adopting the YOLOv4 Architecture for Low-Latency Multispectral Pedestrian Detection in Autonomous Driving. Sensors, 22(3), 1082.

Rouš, R., Peller, J., Polder, G., Hageraats, S., Ruigrok, T., and Blok, P. M. (2023, February 17). Apple scab detection in orchards using deep learning on colour and multispectral images,” arXiv: arXiv:2302.08818.

Saltık, A.O., Allmendinger, A., Stein, A. (2025). Comparative Analysis of YOLOv9, YOLOv10 and RT-DETR for Real-Time Weed Detection. In: Del Bue, A., Canton, C., Pont-Tuset, J., Tommasi, T. (eds) Computer Vision – ECCV 2024 Workshops. ECCV 2024. Lecture Notes in Computer Science, vol 15625. Springer, Cham.

Sapkota, R., Flores-Calero, M., Qureshi, R. et al. (2025). YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series. Artificial Intelligence Review 58, 274.

Shen, J., Wang, Q., Zhao, M., Hu, J., Wang, J., Shu, M., Liu, Y., Guo, W., Qiao, H., Niu, Q., & Yue, J. (2024). Mapping Maize Planting Densities Using Unmanned Aerial Vehicles, Multispectral Remote Sensing, and Deep Learning Technology. Drones, 8(4), 140.

Stypułkowski, K., Gołda, P., Lewczuk, K., & Tomaszewska, J. (2021). Monitoring System for Railway Infrastructure Elements Based on Thermal Imaging Analysis. Sensors, 21(11), 3819.

Sun, C., Chen, Y., Qiu, X., Li, R., & You, L. (2024). MRD-YOLO: A Multispectral Object Detection Algorithm for Complex Road Scenes. Sensors, 24(10), 3222.

Sun, Z., Leng, X., Lei, Y., Xiong, B., Ji, K., & Kuang, G. (2021). BiFA-YOLO: A Novel YOLO-Based Method for Arbitrary-Oriented Ship Detection in High-Resolution SAR Images. Remote Sensing, 13(21), 4209.

Teledyne FLIR. (n.d.) Teledyne FLIR Free ADAS Thermal Dataset v2. Accessed March 10, 2026.

Xue, Y., Ju, Z., Li, Y., and Zhang, W. (2021, November). “MAF-YOLO: Multi-modal attention fusion based YOLO for pedestrian detection,” Infrared Physics & Technology, vol. 118, p. 103906.

Zhao, X., Xia, Y., Zhang, W., Zheng, C., & Zhang, Z. (2023). YOLO-ViT-Based Method for Unmanned Aerial Vehicle Infrared Vehicle Target Detection. Remote Sensing, 15(15), 3778.

Zhu, J., Chen, X., Zhang, H., Tan, Z., Wang , S., and Ma, H. (2023). "Transformer Based Remote Sensing Object Detection With Enhanced Multispectral Feature Extraction," in IEEE Geoscience and Remote Sensing Letters, vol. 20, pp. 1-5, 2023, Art no. 5001405.