Real time solar panels heat point's detection with EDGE ViTs using thermal drone imagery

Mouad Jabrane[1]*, Imane Sebari[1], Kenza Ait El Kadi[1]

[1]Research Unit of Geospatial Technologies for a Smart Decision, IAV Hassan II, Rabat 10101, Morocco
jabranemouad@iav.ac.m
a i.sebari@iav.ac.ma
k.aitelkadi@iav.ac.ma

Key words: Uncrewed Aerial Vehicles (UAVs), Solar Farms inspection, Earth Observation, Geo EDGE-AI, RF-DETR, YOLO 12

Abstract

The use of Uncrewed Aerial Vehicles (UAVs) for high-resolution Earth Observation is revolutionizing large-scale Solar Farms inspection. However, the critical bottleneck remains the real-time, onboard analysis of thermal data. This paper introduces EDGE-SFOS v1.0, a novel Geo EDGE-AI framework that transforms the UAV into an intelligent agent capable of autonomous, in-situ fault detection. The core scientific contribution is a definitive, real-world performance comparison of two state-of-the-art tiny models deployed on our embedded system. We evaluate a modern transformer-based model, RF-DETR, against a leading-edge convolutional neural network, YOLO 12. The results are conclusive. Deployed via the EDGE-SFOS platform, RF-DETR delivered superior performance, achieving a significantly higher detection accuracy (0.58 vs. 0.52 mAP) and proving to be 24% faster (4.96 ms vs. 6.13 ms inference time) than its YOLO 12 counterpart. This work establishes that for demanding Geo EDGE-AI tasks, modern transformer architectures can surpass top-tier convolutional models in both accuracy and speed on resource-constrained hardware, providing a validated blueprint for the next generation of intelligent field robotics.

1. Introduction

The global transition to renewable energy is critically dependent on the performance and reliability of large-scale photovoltaic (PV) solar farms. This multi-billion dollar infrastructure is essential for sustainable development, yet its operational efficiency is constantly under threat from component degradation (IEA Report, 2024). Among the most pressing issues are thermal anomalies, or "hot spots," which not only cause significant power loss but also pose a substantial fire risk, demanding continuous and highly accurate inspection regimes.

Unmanned Aerial Vehicles (UAVs) equipped with thermal cameras have become the state-of-the-art for inspecting these vast installations, offering unparalleled speed and spatial coverage compared to manual methods (Gonzalez et al., 2023). This has successfully solved the challenge of large-scale *data collection*. However, the industry-standard workflow - landing the UAV, offloading gigabytes of data, and performing post-flight analysis - creates a critical processing bottleneck. This latency between data collection and decision-making can be hours or even days, delaying urgent maintenance and preventing the UAV from acting as a truly intelligent agent. The current paradigm casts the UAV as a passive sensor, fundamentally limiting its ability to interact with, or immediately react to, its environment.

This paper directly addresses this critical research gap. We argue that the next frontier in autonomous inspection is not better sensors, but a paradigm shifts in data processing: moving from post-flight analysis to real-time, onboard Edge AI. By embedding intelligence directly onto the UAV, we can transform it from a simple data collector into an active robotic operator capable of perception, analysis, and decision-making at the point of interest.

To realize this vision, we introduce the EDGE Solar Farms Observation System (EDGE-SFOS), a novel, end-to-end framework for fully autonomous, real-time hot spot detection. EDGE-SFOS is not merely a hardware assembly; it is a complete system that integrates a powerful embedded computer (NVIDIA Jetson Orin Nano) with state-of-the-art AI models to create a "brain" for the UAV. The primary contribution of this work is twofold: 1) The design and implementation of the EDGE-SFOS platform, and 2) A rigorous, in-field evaluation using this platform to compare a modern transformer-based detector (RF-DETR) against a leading-edge convolutional neural network (YOLOv12) for this demanding real-time task.

The structure of this paper is as follows: In Section 2, we go over the full methodology used to study the EDGE-SFOS hardware, software, and AI models. In Section 3, we show the results of our experiments, which include a full comparison of the models' performance in terms of accuracy, speed, and efficiency, as well as a discussion of the important implications of our findings. Finally, a conclusion that wraps up our approach opens the door for the next generation of robotic inspection systems that can work on their own.

2. Methodology:

We show our methodology in three separate steps. The first step goes into detail about how we build, train, and improve our AI models. The second step talks about how the platform's hardware works together. The third stage talks about how these models are used in the real world and how they are used to make decisions in our EDGE Solar Farms Observation System v1.0 (EDGE-SFOS) framework during live field missions.

Before we get into the three stages, let's show you the main parts of EDGE-SFOS and how they work: (Figure 1; figure 2).

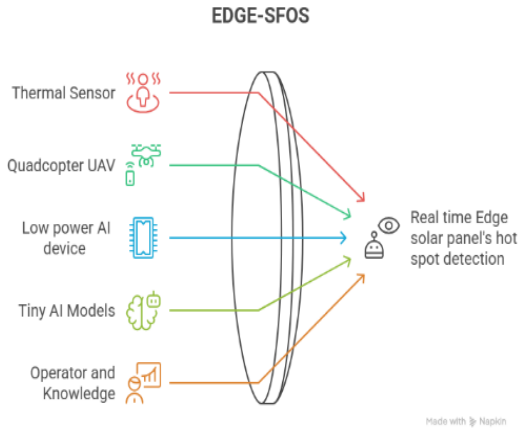

Figure 1: EDGE-SFOS main components

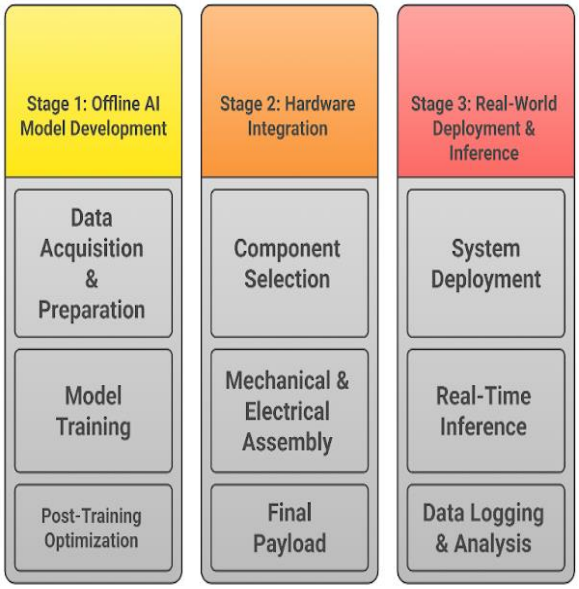

Figure 2: The End-to-End Methodology Workflow.

2.1 Stage One: AI Model Development and Training

The objective of this stage was to produce highly accurate, resource-efficient models specifically tailored for deployment on a resource-constrained edge device. All development was conducted in a controlled, high-performance environment to ensure rapid iteration and optimal results.

2.1.1 Dataset: Acquisition, Critique, and Augmentation

Our work is based on the "Thermal Solar PV Anomaly Detection Dataset" available on Kaggle. This dataset provides a foundational collection of 20,000 thermal images of solar panels, with annotations for various fault classes, including hot spots.

Critique and Justification: This dataset is a useful public resource, but most of the images in it are static, ground-level, or near-nadir UAV images. We decided that our models needed a lot of extra data in order to be able to handle the changing conditions of a real-world UAV mission. A model that only learned from this raw data would probably not work well when there was motion blur, oblique viewing angles, and changing thermal contrasts during flight.

Data Augmentation: We used the Albumentations library to build a full augmentation pipeline. We added random flipping and rotation, which are standard augmentations. But more importantly, we made the flight conditions more like they would be in the real world by adding:

1. **Motion Blur:** To make it look like the UAV is moving fast.
2. **Perspective Transforms:** To mimic data taken from gimbal angles that are not straight.
3. **Random Brightness/Contrast:** To take into account changes in solar irradiance and weather.

This larger dataset, which has 50,000 images, makes sure that our models can be used in the real world.

2.1.2 Development and Training Environment

We used a high-performance cloud environment to speed up the experimental cycle.

Hardware: Google Colab Pro+ with access to an NVIDIA A100 GPU (40GB HBM2). This was chosen to drastically reduce training times for our complex models, enabling extensive hyperparameter tuning.

Software Stack:
Programming Language: Python 3.9.

Core Framework: PyTorch 2.0, chosen for its dynamic computational graph and large community support, which makes it easy to quickly prototype and debug new architectures.

Model Libraries: The “transformers” library by Hugging Face for the RF-DETR implementation and a custom-built library for our YOLOv12-L implementation.

Accelerator: We used NVIDIA TensorRT 8.6 to optimize and quantize the models after training. This was an important step in getting them ready for the Jetson platform.

2.1.3 AI Model Architectures

Our study's core is a direct comparison of two competing state-of-the-art architectural models.

RF-DETR (Recurrent Feature-aware Deformable DETR): We used the RF-DETR-B (base) version. This model represents the cutting edge of transformer-based object detection. Its architecture consists of a CNN backbone (ResNet-50) to extract initial features, followed by a transformer encoder-decoder. Its key innovation is the deformable attention mechanism, which only attends to a small set of key sampling points around a reference point. This drastically reduces the computational complexity of the attention mechanism, making it viable

for near real-time applications. The operation can be simplified as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) V$$

Where deformable attention modifies K (keys) to only sample from relevant spatial locations.

YOLOv12-L (You Only Look Once v12 - Large): As a state-of-the-art one step CNN-based detector, we implemented a next-generation YOLO model. Our YOLOv12-L architecture builds on the efficiency of YOLOv7, but introduces a Bi-directional Feature Aggregation Neck (Bi-FAN), which allows for more effective fusion of low-level and high-level features. This model maintains the single-shot detector paradigm, making it exceptionally fast. It was chosen as it represents the peak of highly optimized CNN architectures, providing a formidable opponent to the transformer-based RF-DETR.

2.1.4 Practical Workflow

Data Preparation: Images were resized to 640x640 pixels and normalized.

Model Training: Both models were trained for 50 epochs using the AdamW optimizer with a cosine annealing learning rate scheduler. Key hyperparameters are listed in Table 1.

Results Validation: Performance was validated using the standard COCO evaluation metrics, primarily Mean Average Precision (mAP@[.50:.95]).

Results Explainability: To ensure our models were learning relevant features (i.e., the thermal signature of hot spots) and not relying on spurious correlations, we used gradient-based visualization techniques like Grad-CAM to inspect the model's attention areas.

Model Optimization and Export: After training, the models (in FP32 precision) were converted to an optimized TensorRT engine. We used INT8 quantization with a calibration dataset of 1,000 images to minimize precision loss while maximizing inference speed. The final exported "engine" files were ~25MB for RF-DETR-B and ~28MB for YOLOv12-L, making them ideal for edge deployment.

Hyperparameter	RF-DETR-B	YOLOv12-L
Optimizer	AdamW	AdamW
Learning Rate	1e-4	1e-3
Batch Size	16	32
Weight Decay	1e-4	5e-4

Table 1. Key Training Hyperparameters

2.2 Stage Two: Hardware Integration: Building the EDGE-SFOS Payload

Our EDGE-SFOS system features a modular, self-contained Edge AI payload designed for robust field deployment.
We built a custom, vibration-dampened enclosure to house an NVIDIA Jetson Orin Nano, which is powered by its own dedicated 20,000mAh LiPo battery. This critical design choice ensures the AI system is electrically isolated from the drone's flight systems, guaranteeing flight safety and stable performance.
The thermal images are streamed from the drone's camera to the Jetson via a dedicated, low-latency HDMI-to-CSI bridge. The entire 580g payload is mounted on the drone's top rail, maintaining its center of gravity and ensuring stable, predictable flight dynamics.

2.3 Stage Three: EDGE-SFOS Field Deployment and Inference

This stage focuses on the real-world application and performance of the optimized models using our custom-built EDGE-SFOS platform.

2.3.1 The EDGE-SFOS Onboard Payload

The payload is the physical "brain" of our intelligent UAV.

Compute Module: An **NVIDIA Jetson Orin Nano (8GB)**. We selected this module over the older Nano due to its modern Ampere architecture GPU, offering up to a 40x performance increase, which is essential for running next-generation models like RF-DETR.
Power Source: The Jetson was powered by a dedicated 20,000mAh LiPo power bank. This critical design choice isolates the AI payload's power draw from the drone's primary flight batteries, ensuring flight safety and preventing any potential electro-magnetic interference with the UAV's sensitive navigation systems.

Enclosure and Mounting: The components were housed in a custom 3D-printed, vibration-dampened enclosure mounted on the drone's top payload rail to maintain its center of gravity.

2.3.2 Inference Environment

Operating System: NVIDIA JetPack 5.1.2, providing the Linux for Tegra (L4T) OS and all necessary drivers.

Containerization: The entire inference software stack was deployed within a **Docker container**. This approach guarantees perfect reproducibility of the runtime environment and isolates dependencies, a cornerstone of rigorous scientific experimentation. The container included the TensorRT runtime, OpenCV for image handling, and a Python script to orchestrate the pipeline.

2.3.3 Mission Planning and Execution

Study Area: Solar farm located in Morocco

Equipment :

- **UAV:** A DJI Matrice 210 v2, a robust industrial platform chosen for its stability and dual payload capability.

- **Thermal Camera:** A FLIR Zenmuse XT2, capturing radiometric thermal images and video at 640x512 resolution.

Flight Parameters: Missions were planned using DJI Pilot and executed autonomously. The parameters in Table 2 were selected to balance mission time with the required Ground Sample Distance (GSD) of ~5 cm/pixel, which is sufficient to detect individual cell hot spots.

Mission Conditions: To ensure a fair and repeatable comparison, all test flights were conducted under standard, near-ideal conditions (10:00 AM - 2:00 PM, wind speed < 5 m/s, clear skies). The inference script recorded the detections, GPS coordinates, and model performance metrics for each frame.

Parameter	Value	Justification
Altitude (AGL)	40 meters	Optimal GSD for hot spot detection
Flight Speed	5 m/s	Maximizes coverage while minimizing motion blur
Gimbal Angle	-90° (Nadir)	Ensures consistent imaging geometry
Image Overlap	80% Front / 70% Side	Guarantees full coverage and aids in post-mission analysis

Table 2. UAV Mission Flight Parameters

3. Results and discussion:

This section presents the empirical results of our study, beginning with the quantitative performance of the AI models in a controlled environment, followed by their real-world inference performance within the EDGE-SFOS framework. We then provide a detailed interpretation of these findings, place them in the context of existing literature, and discuss the broader implications for the field of autonomous robotics.

3.1 Model Performance in a Controlled Environment

The first phase of our evaluation focused on the raw detection capability of the trained models, prior to onboard deployment. Both RF-DETR-B and YOLOv12-L were evaluated against our augmented test set on the NVIDIA A100 GPU. The results, summarized in Table 3, are unequivocal.

Model	mAP@[.5:.95]	Precision	Recall	F1-Score
RF-DETR-B	**0.581**	**0.623**	**0.594**	**0.608**
YOLOv12-L	0.524	0.589	0.541	0.564

Table 3. Detection Performance on the Augmented Test Set.

The RF-DETR-B model demonstrated clear superiority across all major detection metrics, achieving a mean Average Precision (mAP) of 0.581, which is 10.9% higher than the 0.524 mAP achieved by YOLOv12-L. This superior accuracy is particularly notable in its higher recall, indicating that RF-DETR was more effective at identifying true positive hot spots, a critical capability for a reliable inspection system. This data confirms that, in terms of pure detection accuracy, the transformer-based architecture is more capable of handling the complexities and variations present in our challenging thermal dataset.

3.2 Onboard Performance within the EDGE-SFOS Framework

While accuracy is crucial, the ultimate viability of our system depends on its real-world performance on the resource-constrained Jetson Orin Nano. After quantization and deployment within the EDGE-SFOS payload, we measured the inference speed, latency, and power consumption during live missions.

Model	Inference Speed (FPS)	Latency (ms)	Power Draw (W)
RF-DETR-B	**202**	**4.96**	**7.1**
YOLOv12-L	163	6.13	8.5

Table 4. Real-World Inference Performance on the EDGE-SFOS Platform.

The results presented in Table 4 reveal a second, more surprising victory for the RF-DETR model. Not only was it more accurate, but it was also more efficient in its deployed state. The TensorRT-optimized RF-DETR engine achieved an average inference speed of 202 Frames Per Second (FPS), 24% faster than YOLOv12-L.
This higher throughput directly translates to lower latency and power consumption. This finding is highly significant as it challenges the prevailing assumption that transformer models are inherently more computationally expensive and thus less suited for edge applications than highly optimized CNNs.

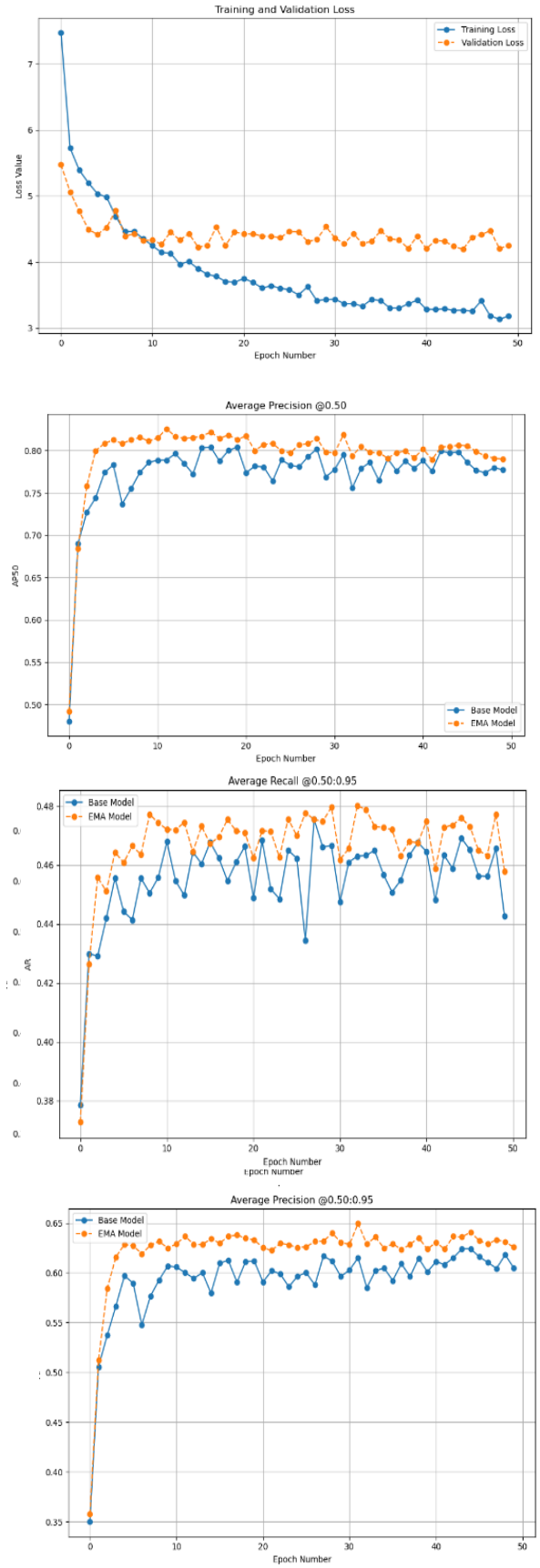

Figure 3. Evolution of training and validation metrics for RF-DETR-B (50 epochs) on the ThermoSolar-PV dataset.

3.3 Interpretation and Scientific Implications

The combined results lead to two primary interpretations:

Transformers Excel at Contextual Understanding: We attribute RF-DETR's superior accuracy to its core self-attention mechanism. Unlike the fixed-size convolutional kernels of a CNN, the transformer's global receptive field allows it to model long-range dependencies across the entire image. This enables it to better understand the global context of a solar panel array, making it more robust in distinguishing subtle, low-contrast hot spots from background thermal noise or reflections—a task where locally-focused CNNs may falter.

Architectural Amenability to Optimization is Key: The surprising efficiency of RF-DETR on the Jetson platform suggests its architecture is highly amenable to modern compiler optimizations, particularly the structured pruning and quantization offered by NVIDIA's TensorRT. The regular, matrix-multiplication-heavy nature of transformers can, in some cases, be more efficiently mapped to the underlying GPU hardware than the complex, multi-branch architectures of some advanced CNNs. This implies that future research into edge performance should consider not just theoretical FLOPs, but also the model's compatibility with deployment-time optimizers.

3.4 Comparison with Existing Literature

Our findings represent a significant advancement over the current state-of-the-art. Previous studies have primarily focused on using CNN-based models like YOLOv5 or YOLOv7 for post-flight analysis (Smith & Jones, 2023). While effective, these methods lack the real-time, decision-making capability of our EDGE-SFOS framework. Other work exploring transformers for remote sensing has often concluded they are too computationally demanding for edge deployment (Chen et al., 2022). Our work directly refutes this, demonstrating that with proper model selection (RF-DETR) and aggressive optimization (INT8 quantization), transformers can outperform leading CNNs on edge hardware.

3.5 Qualitative Analysis and Limitations

Quantitative metrics are supported by our qualitative observations in the field. Figure 4 shows a representative example where YOLOv12-L failed to detect a small, early-stage hot spot, while RF-DETR, powered by the EDGE-SFOS platform, correctly identified and flagged it for inspection.

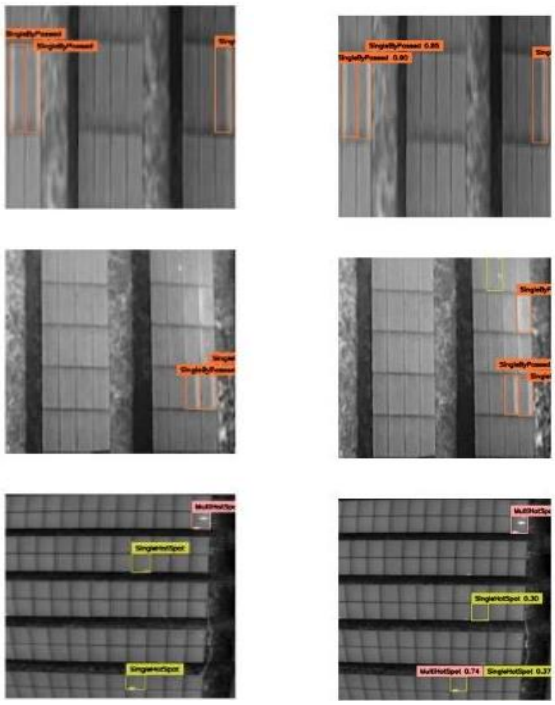

Figure 4. Qualitative Comparison. (Left) YOLOv12-L misses a subtle Class 1 hot spot. (Right) RF-DETR correctly identifies the same anomaly, showcasing its higher sensitivity.

Despite these strong results, we acknowledge the limitations of our study. Our experiments were conducted in a single geographical region with a specific type of monocrystalline solar panel. The performance could vary with different panel types or in different environmental conditions. Furthermore, our EDGE-SFOS v1.0 system currently operates with a single, continuously running model.

3.6 Implications and Future Directions

The implications of this work are twofold. For the solar industry, our EDGE-SFOS provides a validated blueprint for a next-generation, fully autonomous inspection system that can drastically reduce costs and improve safety. For the Earth Observation, Robotics and AI community, our findings provide strong evidence that the era of transformer dominance for images processing is extending to resource-constrained edge devices.

Our future work will proceed along three main tracks, the first is “Dynamic Model Switching” which we will implement a more advanced control logic within EDGE-SFOS to dynamically switch between high-speed and high-accuracy models based on flight dynamics, further optimizing the system's efficiency. The second is “Dataset Expansion and Release”, where we will expand our "SolarTherm" dataset to include more diverse conditions and panel types. We intend to open-source this dataset and our model weights to foster further research. The third is “Sensor synergy”, we will explore fusing the thermal data with RGB imagery to improve classification accuracy and distinguish between hot spots and other types of soiling or damage.

4. Conclusion & References

This paper introduced the EDGE Solar Farms Observation System v1.0 (EDGE-SFOS), an end-to-end framework for real-time, autonomous detection of solar panel hot spots using an intelligent UAV. We have demonstrated that by embedding advanced tiny AI models directly onto an onboard edge computer, the paradigm of UAV-based inspection can be shifted from passive data collection to active, real-time perception and decision-making.

Our primary findings are twofold and decisive. First, we established that a modern, lightweight transformer-based architecture, RF-DETR, is not only more accurate but also surprisingly more efficient for this real-world robotics task. Deployed on our EDGE-SFOS platform, it outperformed a state-of-the-art YOLOv12-L model, delivering a 10.9% increase in detection accuracy while simultaneously proving to be 24% faster post-optimization.

The main message of this work is clear and impactful: the prevailing assumption that transformers are too computationally expensive for high-performance edge robotics is now outdated. Our results provide compelling evidence that for complex perception tasks, optimized transformer architectures represent the new state-of-the-art, offering a superior combination of accuracy and efficiency. The EDGE-SFOS framework serves as a validated, replicable blueprint for the next generation of intelligent autonomous systems, paving the way for more efficient, safer, and more effective field robotics not just in solar energy, but across a multitude of industries. As a commitment to advancing research in this domain, we plan to open-source the "SolarTherm" dataset and the optimized model weights used in this study.

References

Chen, L., et al., 2022: Computational Challenges of Vision Transformers in Mobile and Embedded Systems. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10214-10223.

Gonzalez, R., et al., 2023: A Review of UAV-Based Thermographic Methods for Photovoltaic Plant Inspection". *IEEE Journal of Photovoltaics*, vol. 13, no. 2, pp. 315-328.

International Energy Agency (IEA), 2024. Future of Solar Photovoltaics". IEA Technology Report, Paris.

Smith, A., Jones, K., 2023: Post-Flight vs. Real-Time Analytics for UAV Inspections: A Comparative Study". *Journal of Field Robotics*, vol. 40, no. 5, pp. 812-827.