

Dear Editor-in-Chief and Editorial Team of the Journal of Real-Time Image Processing,

We are pleased to submit our manuscript entitled:

“Drone-Based Early Forest Fire Detection Using Vision-Language Models on NVIDIA Jetson Orin Nano”

for consideration for publication in the Journal of Real-Time Image Processing.

Forest fires require rapid and reliable detection, particularly during the early stages when intervention can significantly reduce environmental and economic damage. In this work, we propose an edge-AI framework for early forest-fire detection using unmanned aerial vehicles (UAVs), combining vision-language models (VLMs) with the NVIDIA Jetson Orin Nano embedded computing platform.

The main contribution of this study is the integration of the semantic and contextual reasoning capabilities of vision-language models with real-time edge inference on a resource-constrained embedded platform. Unlike conventional fire-detection approaches that rely primarily on bounding-box object detectors, our approach exploits multimodal visual understanding to identify and interpret early fire and smoke manifestations in forest environments. The proposed system is designed specifically for deployment on UAV platforms, where computational resources, latency, energy consumption, and responsiveness are critical considerations.

The manuscript presents the complete system architecture, including image acquisition, preprocessing, vision-language inference, edge deployment, and decision-making components. Particular attention is given to real-time processing on the NVIDIA Jetson Orin Nano, demonstrating the feasibility of deploying advanced vision-language models directly at the edge rather than relying exclusively on cloud-based processing. This edge-based architecture reduces communication requirements and enables rapid detection and response in remote forest environments.

We believe that this work is well aligned with the scope of the Journal of Real-Time Image Processing because it addresses several important aspects of real-time image processing, including:

- real-time visual analysis for early fire and smoke detection;
- deployment of advanced vision models on resource-constrained edge hardware;

- optimization of vision-language models for embedded inference
- UAV-based image acquisition and autonomous environmental monitoring;
- low-latency image understanding and decision making; and
- practical integration of computer vision and edge-AI technologies.

Beyond the specific application of forest-fire detection, the proposed framework demonstrates a broader approach for deploying vision-language-based image understanding systems in real-time and resource-constrained environments.

We confirm that this manuscript is original, has not been published previously, and is not currently under consideration for publication elsewhere. All authors have contributed to the work and have approved the submitted manuscript and its submission to the Journal of Real-Time Image Processing.

We respectfully ask you to consider our manuscript for peer review. We believe that the proposed combination of vision-language models, UAV-based image acquisition, and embedded real-time inference represents a timely contribution to the field of real-time image processing and edge artificial intelligence.

Thank you for your time and consideration. We look forward to hearing from you.

Sincerely,

Mohamed Amine Jebli

Mehrez Abdellaoui

Ali Douik

ENISO -- National Engineering School of Sousse

University of Sousse

Sousse, Tunisia

NOCS Laboratory

Corresponding Author:

Mohamed Amine Jebli

Email: jeblimmedamine@gmail.com