<div align="center">

TABLE III
COMPARISON OF MODEL'S COMPUTATION DEMANDS

</div>

	DETR	SDD	YOLOv5-X	YOLOv9-T	YOLOv9-E	YOLOv10-N	YOLOv10-X
#parameters [M]	46	35	86.7	2.0	57.3	2.3	29.5
GFLOPs	127	34.5	205.7	7.7	189.0	6.7	160.4

characterizing a more precise detection overall. Conversely, the YOLOv5-X model was able to detect the largest number of BMSB reaching the highest recall. Despite achieving particularly notable metrics scores in the ONLY set, the model has depicted the least robustness against background-only patches, namely, in the ALL set. So, YOLOv5-X was prone to recognize $\geq 55\%$ of BMSB instances with a particularly promising precision when a BMSB appears inside the frame. However, this latter capability deteriorates dramatically ($P = 53.5\%$) when images without BMSB are considered. It is noteworthy that each network tends to achieve a larger P than R. This phenomenon is strictly related to the poses experienced during the acquisition, compounded by the noise surrounding the BMSB pixels, such as blur or overexposure. Despite the dataset comprising a diverse range of BMSB samples, certain poses and lighting/blurriness conditions are more frequent than others. Also, since the dataset consists of images captured by a drone operating according to an autonomous protocol, characterizing the entire range of possibilities is extremely challenging. As a result, the networks were able to detect BMSB instances that presented limited variation in pose and color appearance, thereby minimizing FP. However, instances significantly affected by noise were occasionally missed. Both YOLOv9 and YOLOv10-X achieve interesting performance representing a true step ahead of the predecessor, YOLOv5. Although YOLOv10-X obtained the largest R, YOLOv9-T achieves the best performance overall. Specifically, it reaches a satisfactorily R ($\approx 60\%$) and almost the largest P with both the thresholds, 95.0% and 94.0%, respectively.

Looking at mAP$_{0.5}$, all the NNs have achieved $\geq 45\%$ of BMSB, establishing a satisfactory confidence level overall. On the other hand, the NNs have developed limited abilities in contouring the BMSB according to mAP$_{0.5}^{0.95}$ because they are less robust when the IoU increases. In other words, the majority of predictions overlap partially with the ground truth. This represents a predictable outcome due to the limited size of the BMSB with respect to the entire image. For monitoring purposes, recognizing all instances inside the frame, i.e., achieving high recall, is more important than accurately contouring a bug, i.e., achieving high IoU. However, it noteworthy that both YOLOv9 and YOLOv10 demonstrate improved IoU values between prediction and ground truth, suggesting a clear enhancement with previous releases. Moreover, in comparison with the results in [20] where entire images are simply resized and fed to the model, the current implementation of the slicing mechanism allows a boost in the performance, specifically in the recall. Indeed, image slicing prevents information losses since it retains every pixel of the original image.

Table III summarizes model's computation demands listing the number of parameters and the number of giga floating point operations per second (GFLOPs), respectively. We can observe that YOLOv5-X, YOLOv9-E, YOLOv10-X, SSD, and DETR are more "desktop oriented" due to their requirements, rather than YOLOv9-T and YOLOv10-N that appear compatible with current state-of-the-art embedding system boards. According to the computation footprint, both YOLOv9-T and YOLOv10-N are, even more so, the best NNs overall.

IV. BMSB DETECTION USING SPECTRAL IMAGING

In the HALY.ID project, alongside a comprehensive analysis of RGB images primarily captured by UAVs, we also conducted an assessment of spectral imaging as a potential method for field monitoring to detect the presence of BMSB. To this purpose, SWIR-HSI and Vis-NIR MSI systems were evaluated both in the laboratory and in the field, respectively. In fact, unlike cameras operating exclusively within the visible spectrum, they may mitigate misclassifications caused by BMSB's resemblance to vegetal backgrounds with similar colors, such as bark or brown leaves.

A. Evaluation of SWIR-HSI

We captured 35 hyperspectral images of BMSB specimens positioned against various vegetal backgrounds (including bark, branches, grass, soil, green/brown leaves) within the 980–1660 nm spectral range, aiming to replicate real field conditions. We acquired the hyperspectral images using a HSI line-scan system equipped with a desktop NIR Spectral Scanner (DV Optic) incorporating a Specim N17E reflectance imaging spectrometer, coupled with a Xenics XEVA 1.7-320 camera (320×256 pixels) and a Specim Oles 31 $f/2.0$ optical lens. The acquisition software conducted an automatic calibration of the images based on the dark current signal and a high-reflectance standard signal. We excluded pixels unrelated to either bugs or vegetal backgrounds based on a reflectance threshold measured at 1000 nm. In addition, we preprocessed the images using standard normal variate (SNV) and mean center techniques. Subsequently, we applied principal component analysis (PCA) to each image to implement a masking procedure, thereby separating the pixels belonging to the bugs from those belonging to the vegetal backgrounds.[1]

To develop classification models capable of distinguishing between BSBM specimens and vegetal backgrounds based on spectral signatures, we assembled a library comprising 14 000 reference spectra from both classes. We employed the Kennard–Stone algorithm [37] on principal component (PC) scores to select these spectra. Then, we divided the images into training and test sets for model development and validation, respectively. Our classification approach focused on modeling the spectral information contained in the HSI images and identifying the relevant spectral regions to discriminate BMSB from vegetal backgrounds. To do this, we used the soft partial least squares-discriminant analysis (Soft PLS-DA) algorithm, coupled with sparse-based methods for spectral variable selection (s-Soft PLS-DA) [38]. This allowed us to build effective classification models tailored to our dataset.

---

[1] A detailed description of the methodology used for acquiring and processing the hyperspectral images can be found in the work of Ferrari et al. [36].