Enabling RGB Anomaly Detectors for Foreign Object Detection in Lentils via Task-Learned Spectral Selection

Anish Raj, Simon Birkholz, Nima Ghorbani
Cubert GmbH, Ulm, Germany · {raj,birkholz,ghorbani}@cubert-gmbh.de

1 Introduction

Detection of foreign objects (FO) in bulk foods is important for food safety. Contaminants are usually small, subtle, and often look similar to the product. In this aspect, RGB cameras struggle, while hyperspectral imaging (HSI) adds spectral contrast and can be helpful in distinguishing deceptive objects. However, strong off-the-shelf FO/anomaly detectors expect RGB images as input. So, the question arises: which 3 bands should represent an HSI cube? In this work, we test 3 sets of inputs on a public dataset and find that a learned 3 band set outperforms the RGB and handpicked colour-infrared (CIR) triplet.

2 Data

Our data consists of 1136 frames containing Beluga and green marbled lentils. It is captured on a blue conveyor belt material with Cubert Ultris XMR 50. The HSI cube has 61 bands in the 430-910 nm range with a resolution of 1000×1080 px. Each scene is taken under 4 lighting conditions. A total of 696 frames are annotated containing FOs having 7 classes (stems, stones, aluminum shards, blue/white papers, fly, and rubber pieces). The train/val/test split is 808/148/180, where the 4 lighting copies of a scene always stay in one split. The data is released on Hugging Face under Apache 2.0 [1].

3 Method

We create 3 inputs to convert an HSI cube to 3 channel inputs: RGB (650/550/450 nm), CIR (860/670/560 nm) and a learned triplet. The learned triplet is derived through a Gumbel-Softmax mixer that is trained on the train split with an IoU loss backpropagated through a frozen AdaCLIP detector [2]. The hard selection during inference converges the learned bands to 542/902/886 nm. We then feed these 3 input sets to 3 detectors: AdaCLIP zero-shot, finetuned AdaCLIP, and Dinomaly [3], which is a normal-only reconstruction detector with DINOv2 backbone. For fair comparison the Dinomaly models are trained with the same settings (50 epochs, AdamW, seed 42, 308 FO free frames). We set image level thresholds on validation set and test them on 180 test frames of which 112 contain FO.

4 Results and Discussion

3-band view (nm)	Zero-shot AdaCLIP	Finetuned AdaCLIP	Dinomaly
RGB (650/550/450)	0.520 / 0.899	0.649 / 0.906	0.769 / 0.950
CIR (860/670/560)	0.750 / 0.966	0.776 / 0.979	0.907 / 0.990
Learned (542/902/886)	**0.923 / 0.977**	**0.932 / 0.984**	**0.981 / 0.993**

Table 1. Image and pixel AUROC for each 3-band input on the test set. Zero-shot AdaCLIP is pretrained with no training on our dataset. Finetuned AdaCLIP adapts its prompt and adapter layers on train split. Dinomaly is trained on only normal object free frames. Best number per column is highlighted in bold.

We find that the learned bands input outperform RGB and CIR on both image and pixel AUROCs across the 3 models (learned > CIR > RGB). The improvement in results is through the bands and not any single detector. The image AUROC for Dinomaly jumps from 0.77 to 0.98 when RGB is replaced by the learned bands. Similarly, AUPRO@0.30 also improves from 0.72 to 0.94. The RGB (Dinomaly) misses 4 frames with FO, whereas the learned bands (Dinomaly) recover them (0 false negatives). Furthermore, the false positives are reduced from 47 to 14 frames (out of 68 clean frames). Meanwhile, CIR gains most on pixel-level but not at frame level separation. The runtime of the Dinomaly pipeline is approx. 70 ms/frame on an NVIDIA RTX A4000, being 4 times faster than AdaCLIP. Caveats: The channel selector is trained in a supervised manner (with masks), so although Dinomaly needs only normal frames, the bands it uses are chosen with labels. The belt is static during capture, and the 180 test frames come from 45 scene arrangements.

5 Conclusion

For lentils, automatically deriving a task specific 3-band input improves FO detection across all 3 detectors and enables off-the-shelf RGB detectors to be applied to hyperspectral data.

References

[1] A. Raj. Hyperspectral Foreign-Object Detection in Lentils (XMR Industrial FOD). Hugging Face Datasets, 2026. https://huggingface.co/datasets/cubert-gmbh/XMR_Industrial_Foreign_Object_Detection_Lentils

[2] Cao, Yunkang, et al. "Adaclip: Adapting clip with hybrid learnable prompts for zero-shot anomaly detection." European conference on computer vision. Cham: Springer Nature Switzerland, 2024.

[3] Guo, Jia, et al. "Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection." 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2025.