HLSS: An Agent-Based Adaptive Layered Perception Framework for Hyperspectral Industrial Environments

Assia Belbachir
NORCE Research AS,
Grimstad, Norway
assb@norceresearch.no

Ahmed-Nabil Belbachir
NORCE Research AS,
Grimstad, Norway
nabe@norceresearch.no

Önder Gürcan
NORCE Research AS,
Kristiansand, Norway
ongu@norceresearch.no

ABSTRACT

We present **Hybrid Layered Spectral Segmentation (HLSS)**, an *agent-based* multimodal framework for material-aware instance segmentation in hyperspectral industrial imagery. HLSS frames perception as a cooperative interaction of **proposal agents** (spatial hypotheses), **verification agents** (spectral coherence checks), and **residual agents** (shared state updates) operating over a residual hyperspectral cube. At each iteration, proposal agents generate candidate masks from pseudo-spectral prompts; verification agents accept candidates using cosine-similarity to a reference library and within-mask variance constraints; residual agents then remove accepted pixels to reveal deeper layers. We formalize this as a discrete multi-agent decision process, characterize monotonic residual dynamics, and interpret HLSS as a greedy maximization of spectral coherence under disjoint-support constraints. Finite termination follows from residual shrinkage. On the TECNALIA WEEE benchmark, HLSS attains high spectral precision and produces interpretable confidence maps while remaining training-light. The agent-based view highlights clear interfaces for learning, decentralization, and adaptive policy design.

KEYWORDS

Adaptive perception, hyperspectral imaging, occlusion reasoning.

1 INTRODUCTION

Modern industrial environments increasingly rely on automated perception for sorting, recycling, and quality control. High-throughput tasks such as waste electrical and electronic equipment (WEEE) sorting demand systems capable of delineating objects in cluttered conveyor scenes while inferring their underlying material composition. Conventional RGB-based research [11] struggle with visually similar materials, partial occlusions, specular reflections, and variable illumination. Hyperspectral imaging (HSI), by contrast, offers rich spectral signatures encoding material properties, yet purely spectral approaches often require dense supervision, careful calibration, and substantial computational resources.

Recent advances in foundation segmentation models such as the Segment Anything Model (SAM) [10] enable high-quality zero-shot spatial mask proposals through prompt-based queries. These models generalize across diverse domains but remain driven by appearance, with limited sensitivity to spectral or material cues. Integrating their strong spatial priors with the physical selectivity of hyperspectral signatures creates an opportunity for interpretable, data-efficient segmentation in complex industrial contexts.

To address these challenges, we propose Hybrid Layered Spectral Segmentation (HLSS), a modular framework that couples promptable foundation segmentation with hyperspectral verification and iterative residual fusion. From a systems perspective, HLSS is naturally interpretable as an *agent-based* layered perception process. Concretely, we view the process as a cooperative population of agents operating over a shared residual state: *proposal agents* generate spatial hypotheses from pseudo-spectral prompts, *verification agents* assess material coherence using hyperspectral cues, and *residual agents* update the global state by removing accepted pixels. This agent-based framing is descriptive rather than prescriptive: it organizes the existing modular components, clarifies interfaces for decentralized or learned policies, and motivates the formalization presented in Section 3.

These agents operate asynchronously yet cooperatively, enabling robust occlusion handling and progressive scene decomposition. The multi-agent perspective allows modular evaluation of each component and simplifies future extensions [2, 7], such as learned verification policies or decentralized decision-making. By explicitly mapping perception modules to agent roles, HLSS provides a clear blueprint for adaptive and scalable industrial vision systems.

The novelty of HLSS stems from its integration of complementary reasoning paradigms. The framework unifies promptable, appearance based segmentation with physics-grounded spectral verification, enabling multimodal inference that benefits from the strengths of each modality. It also introduces an iterative residual strategy that reveals occluded instances by progressively subtracting spectrally verified regions, thereby improving scene interpretability in settings where items overlap or cover one another.

Finally, HLSS emphasizes training-light reproducibility: its modular components, reliance on foundation models, and compatibility with open benchmarks make it straightforward to deploy and adapt within industrial environments. We evaluate HLSS on the TECNALIA WEEE HSI dataset [17], a representative benchmark for industrial hyperspectral perception. Experiments demonstrate that HLSS provides improved precision and interpretable spectral confidence maps while maintaining competitive segmentation metrics relative to supervised HSI- and RGB-based baselines.

The paper is organized as follows: Section 2 reviews related work in hyperspectral perception and multimodal segmentation. Section 3 explains the problem definition. Section 4 details the HLSS agent framework, algorithmic design, and pseudocode. Section 5 presents datasets, baselines, and experimental results. Section 6 discusses the results and future research directions.

Proc. of the Adaptive and Learning Agents Workshop (ALA 2026), Aydeniz, Delgrange, Mohammedalamen, Yang (eds.), May 25 – 26, 2026, Paphos, Cyprus, https://alaworkshop2026.github.io/. 2026.

2 RELATED WORK

Our work lies at the intersection of hyperspectral material recognition, promptable foundation segmentation, and occlusion-aware layered reasoning in industrial vision. We review these areas and expose their limitations to motivate our Hybrid Layered Spectral Segmentation (HLSS) framework as a unifying solution.

Hyperspectral Imaging for Material Classification. Hyperspectral imaging (HSI) captures dense spectral information across narrow wavelength bands, enabling fine-grained material discrimination beyond RGB capabilities [6, 15]. Early methods relied on spectral angle mapping and unmixing, while recent approaches employ 3D CNNs and spectral–spatial Transformers to jointly encode spectral and spatial correlations [11, 12, 17]. Despite high per-pixel accuracy, these approaches require dense supervision and degrade under mixed pixels, specularities, and occlusions—key failure modes in conveyor-based industrial settings.

RGB Segmentation and Promptable Foundation Models. RGB instance segmentation has been dominated by architectures such as Mask R-CNN [8], DeepLab [3], and U-Net, which require large labeled datasets and struggle with domain transfer. Foundation models such as SAM [10] and Grounded-SAM [18] enable zero-shot, promptable segmentation with strong boundary quality. However, these models remain fundamentally appearance-driven and lack sensitivity to intrinsic material properties. Multimodal extensions incorporating thermal or depth cues [4] improve robustness, yet hyperspectral integration remains largely unexplored and underdeveloped.

Layered and Occlusion-Aware Segmentation. Occlusion reasoning remains a central challenge in cluttered scenes. Methods such as BCNet [9], Slot Attention [14], and iterative residual learning [13] decompose scenes into object-centric layers. While effective in spatial disentanglement, these approaches are blind to material composition, limiting their applicability in domains where appearance is insufficient for discrimination. HLSS extends layered reasoning by incorporating spectral verification within an iterative residual fusion process, enabling recovery of materially distinct but visually ambiguous objects.

Hybrid Verification Approaches. Verification and reranking strategies are widely used to combine complementary cues in detection pipelines. HLSS generalizes this paradigm to the spectral domain: foundation models generate high-quality spatial proposals, while a hyperspectral verifier evaluates spectral similarity, variance, and optional pixel-wise classification to accept or reject candidates [17]. This decoupled design improves robustness, reduces false positives, and enables training-light adaptation to new material domains.

Multi-Agent Systems and Cooperative Perception. The agent-based interpretation of HLSS connects to multi-agent systems (MAS), where complex tasks are decomposed into interacting entities with specialized roles. Classical work in distributed AI and cooperative control formalizes coordination through shared state and communication protocols [5, 16, 20]. Recent advances in cooperative multi-agent learning highlight modularity, interpretability, and scalability as key advantages in complex environments [21]. HLSS instantiates these principles through a functional decomposition into proposal, verification, and residual agents that interact via a shared residual representation [1]. Unlike fully learned MAS approaches, this design remains modular, interpretable, and computationally efficient, while still enabling future extensions toward learned coordination or decentralized inference.

Research Gap and Motivation. Existing approaches address only isolated aspects of the problem: hyperspectral models provide material discrimination but lack occlusion reasoning; foundation models generalize spatially but ignore spectral cues; layered methods recover occlusions but remain materially agnostic. No prior work unifies these capabilities within a coherent, modular framework. HLSS addresses this gap by (i) leveraging foundation-driven proposals for generalization, (ii) enforcing spectral consistency for material-level accuracy, and (iii) employing iterative residual fusion for occlusion recovery. This integration defines a new multimodal, training-efficient paradigm for industrial waste segmentation, directly aligned with real-world deployment constraints.

3 FORMAL PROBLEM DEFINITION

3.1 Multi-Agent Interpretation

We adopt an explicit multi-agent perspective while keeping the original algorithm unchanged. Let $\mathcal{N} = \{1, \ldots, N\}$ denote agents partitioned into three functional classes:

- **Proposal agents** generate candidate masks m from pseudo-spatial prompts derived from $I_{HSI}^{(t)}$ (they operate primarily on I_{RGB} guided by P_t).
- **Verification agents** compute spectral summaries (e.g., μ_m) and score candidates against a reference library $\mathcal{R}$ using similarity and variance tests; their outputs are binary accept/reject decisions or continuous confidence scores.
- **Residual agents** enact accepted decisions by updating the global residual mask $R^{(t)}$, producing $R^{(t+1)}$ and the next residual cube $I_{HSI}^{(t+1)}$.

Formally, the interaction is equivalent to a cooperative discrete decision process where agents share the global residual $R^{(t)}$ and communicate only via proposed masks and acceptance signals. This mapping clarifies the responsibilities of each module and supports modular extensions (e.g., learned verification policies or decentralized consensus) without changing the underlying residual masking dynamics.

3.2 Problem Setup

Let $\Omega \subset \mathbb{Z}^2$ denote the discrete image domain of size $|\Omega| = H \times W$. Let the hyperspectral cube be defined as

$$I_{HSI} : \Omega \to \mathbb{R}^B,$$

where B is the number of spectral bands.
The RGB image is

$$I_{RGB} : \Omega \to \mathbb{R}^3.$$

The objective of HLSS is to recover a finite set of instance masks

$$\mathcal{M} = \{m_1, \ldots, m_K\}, \quad m_k : \Omega \to \{0, 1\},$$

such that:

Table 1: Comparison of representative segmentation paradigms relevant to industrial waste recognition.

Approach	Modality	Supervision	Occlusion	Generalization	Spectral Awareness
Spectral CNNs / Transformers [6, 12, 17]	HSI	Full	✗	Medium	✓
RGB-based Deep Models [3, 8]	RGB	Full	✗	Medium	✗
Multimodal CNNs [19]	RGB-D/T	Full	Partial	Medium	✗
Foundation Models [4, 10, 18]	RGB+Text	Zero-shot	✗	High	✗
Layered/Occlusion-Aware [9, 13, 14]	RGB	Varied	✓	Low	✗
Proposed HLSS	RGB+HSI	Weak/Zero-shot	✓	High	✓

(1) Masks are pairwise disjoint:

$$m_i \odot m_j = 0 \quad \forall i \neq j,$$

(2) Each mask satisfies spectral coherence constraints,
(3) The union of accepted masks maximizes spectral consistency under residual iteration.

3.3 Residual Dynamics

Define the residual hyperspectral cube at iteration t as:

$$I_{HSI}^{(t)} = I_{HSI} \odot R^{(t)},$$

where $R^{(t)} : \Omega \to \{0, 1\}$ is a binary residual mask.
Initialization:

$$R^{(0)}(x) = 1 \quad \forall x \in \Omega.$$

After accepting masks $\mathcal{M}_{acc}^{(t)}$ at iteration t, the residual update is:

$$R^{(t+1)} = R^{(t)} \cdot \left(1 - \sum_{m \in \mathcal{M}_{acc}^{(t)}} m\right).$$

This defines a monotonic non-increasing sequence of residual supports:

$$\text{supp}(R^{(t+1)}) \subseteq \text{supp}(R^{(t)}).$$

3.4 Spectral Coherence Criterion

For a candidate mask m, define its mean spectral signature:

$$\mu_m = \frac{1}{|m|} \sum_{x \in \Omega} m(x) \cdot I_{HSI}^{(t)}(x).$$

Let $\mathcal{R} = \{r_1, \ldots, r_L\}$ be the spectral reference library.
We define the cosine similarity:

$$\text{sim}(\mu_m, r_l) = \frac{\langle \mu_m, r_l \rangle}{\|\mu_m\| \|r_l\|}.$$

A mask is accepted if:

$$\max_{r_l \in \mathcal{R}} \text{sim}(\mu_m, r_l) \geq \tau_s$$

and

$$\text{Var}_m \leq \tau_u,$$

where Var_m denotes within-mask spectral variance.

3.5 Greedy Spectral Maximization View

HLSS can be interpreted as a greedy optimization process:

$$\max_{\mathcal{M}} \sum_{m \in \mathcal{M}} \left(\max_{r_l \in \mathcal{R}} \text{sim}(\mu_m, r_l) - \lambda \text{Var}_m \right)$$

subject to disjoint support constraints and residual masking dynamics.

At each iteration, HLSS selects spectrally valid masks and removes them from future consideration. This corresponds to a greedy layered decomposition of the hyperspectral cube.

3.6 Finite Termination

Proposition 1. Assume $|\Omega| < \infty$ and that each accepted mask satisfies $|m| > 0$. Then HLSS terminates in at most $|\Omega|$ iterations.

Proof. Each accepted mask removes at least one previously unremoved pixel from the residual domain. Since the domain is finite and removal is monotonic, the process must terminate after a finite number of steps.

4 HLSS AGENT-BASED FRAMEWORK

We introduce **Hybrid Layered Spectral Segmentation (HLSS)**, a multimodal framework designed for material-aware instance segmentation in cluttered industrial environments. HLSS combines the generalization strength of foundation segmentation models with hyperspectral verification and an iterative residual mechanism that progressively uncovers occluded objects while enforcing spectral consistency. The result is a segmentation appraoch capable of operating without task-specific training and robust enough to handle the complex material interactions present in industrial waste sorting (see Figure 1).

4.1 Overview

HLSS operates on an RGB–HSI pair (I_{RGB}, I_{HSI}) and optionally a spectral reference library $\mathcal{R}$. At each iteration t, the method follows a three-stage process. First, a foundation segmentation model, such as SAM, produces a set of candidate instance masks $\mathcal{M}_t$ based on spatial cues computed from the residual hyperspectral cube. Second, each candidate mask is examined using a spectral verification procedure that evaluates its material coherence by comparing hyperspectral signatures, assessing within-mask consistency, and optionally consulting a lightweight per-pixel classifier. Finally, the hyperspectral cube is updated by removing all pixels associated

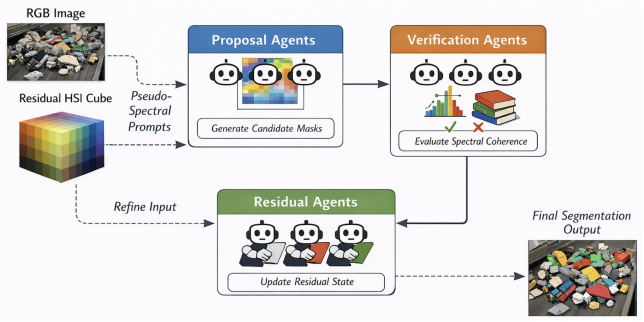


Figure 1: Illustration of the HLSS Agent-based framework.

with accepted masks, exposing deeper layers of the scene for subsequent iterations. This *peel-and-reveal* cycle repeats until the remaining hyperspectral coverage falls below a threshold, preventing over-segmentation and unnecessary iterations.

4.2 Pseudo-Spectral Prompting

Because prompt-based foundation models operate exclusively on spatial cues, HLSS converts hyperspectral information into pseudo-spatial prompts. At iteration t, this is achieved by constructing a prompt map

$$P_t = f_{prompt}(I_{HSI}^{residual}),$$

where $I_{HSI}^{residual}$ represents the current residual hyperspectral cube. The mapping function f_{prompt} may arise from spectral saliency measures that highlight regions with distinctive spectral responses, or from learned embeddings in which hyperspectral pixels are clustered and projected into two-dimensional spatial space. These pseudo-spatial representations guide the foundation segmenter toward proposals that align with material boundaries suggested by the underlying spectral data.

4.3 Spectral Verification Module

Each proposed mask $m \in \mathcal{M}_t$ is evaluated through a multi-criteria verification process that ensures spectral coherence. The first component relies on spectral similarity measures such as cosine similarity or the spectral angle mapper (SAM), which compare the aggregated spectral signature of the candidate region to references within $\mathcal{R}$. A second component examines the within-mask spectral variance, since excessive variance typically indicates mixed materials or noisy proposals; masks exceeding a variance threshold are rejected. An optional third component uses a lightweight classifier—such as a shallow CNN or linear spectral model trained on a small labeled subset—to refine the acceptance decision. Only masks that satisfy all verification criteria are incorporated into the final segmentation set, while rejected masks are either discarded or reconsidered in subsequent iterations.

4.4 Iterative Residual Fusion and Stopping Criterion

To handle occlusion and reveal successive object layers, HLSS maintains an updated residual hyperspectral cube. After accepting the set of verified masks $\mathcal{M}_t^{accepted}$, the residual is updated according to

$$I_{HSI}^{residual} = I_{HSI}^{residual} \odot \left(1 - \sum_{m \in \mathcal{M}_t^{accepted}} m\right),$$

where $\odot$ denotes element-wise masking. This mechanism removes already-verified pixels from future consideration, allowing HLSS to progressively expose previously hidden objects. The iterative process continues until either the residual hyperspectral coverage falls below a threshold $\tau_{coverage}$ or a predefined iteration limit T_{max} is reached. To reduce fluctuations caused by noisy spectral responses, HLSS applies exponential smoothing to the spectral confidence maps used in the verification stage.

Algorithm 1 Hybrid Layered Spectral Segmentation (HLSS)

Require: RGB image I_{RGB}, hyperspectral cube I_{HSI}, optional spectral reference library $\mathcal{R}$

Initialize $I_{HSI}^{residual} \leftarrow I_{HSI}$, $t \leftarrow 0$, $\mathcal{M}_{out} \leftarrow \emptyset$
while Residual coverage $> \tau_{coverage}$ and $t < T_{max}$ **do**
  $P_t \leftarrow f_{prompt}(I_{HSI}^{residual})$ ▷ Generate pseudo-spatial prompts
  $\mathcal{M}_t \leftarrow \text{SAM}(I_{RGB}, P_t)$ ▷ Proposal generation
  $\mathcal{M}_t^{accepted} \leftarrow \text{SpectralVerifier}(\mathcal{M}_t, I_{HSI}^{residual}, \mathcal{R})$
  $\mathcal{M}_{out} \leftarrow \mathcal{M}_{out} \cup \mathcal{M}_t^{accepted}$
  $I_{HSI}^{residual} \leftarrow I_{HSI}^{residual} \odot (1 - \sum_{m \in \mathcal{M}_t^{accepted}} m)$
  $t \leftarrow t + 1$
end while
return $\mathcal{M}_{out}$

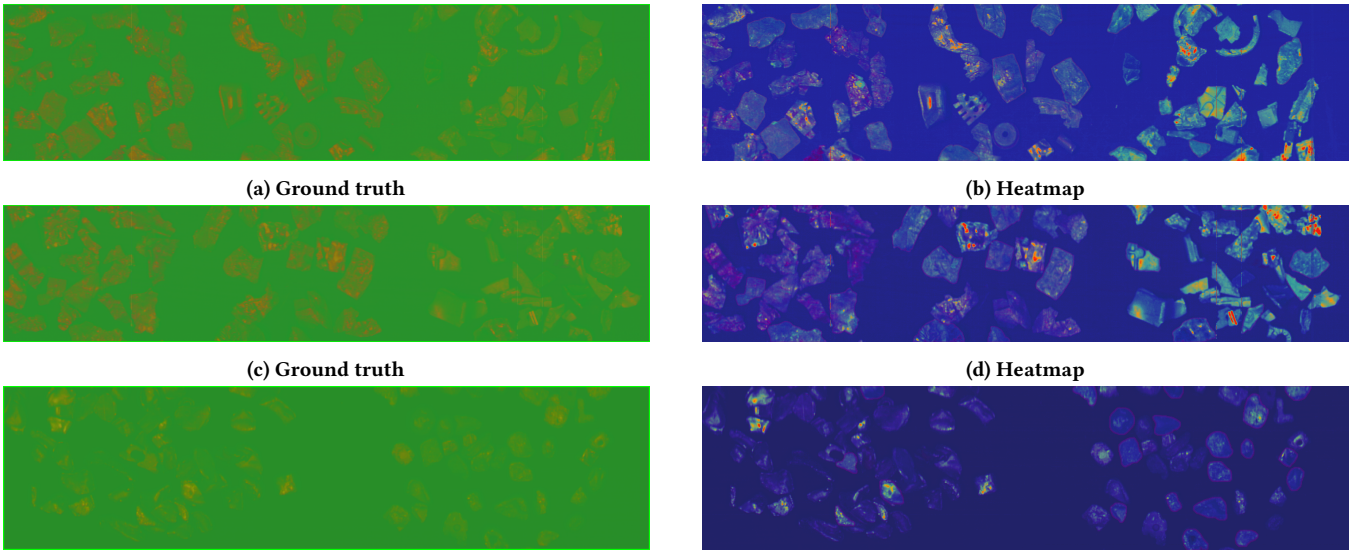
4.5 Algorithm Summary

Algorithm 1 presents the full HLSS in pseudocode form, summarizing the interplay between prompt generation, mask proposal, spectral verification, and iterative residual fusion.

HLSS operates through an iterative *peel-and-reveal* process. Starting with the full hyperspectral cube I_{HSI}, the method generates pseudo-spatial prompts $P_t = f_{prompt}(I_{HSI}^{residual})$ that guide a foundation segmenter (such as SAM) to propose instance masks. These candidates $\mathcal{M}t$ are then filtered by the spectral verification module using spectral similarity to $\mathcal{R}$, within-mask variance, and optional lightweight classification. Accepted masks $\mathcal{M}t^{accepted}$ are added to the output set and removed from the residual cube via $I_{HSI}^{residual} \leftarrow I_{HSI}^{residual} \odot (1 - \sum_{m \in \mathcal{M}t^{accepted}} m)$, progressively exposing deeper occluded objects. Iterations continue until the residual coverage drops below $\tau_{coverage}$ or T_{max} is reached, yielding a spectrally consistent, layer-aware segmentation of the scene.

5 EXPERIMENTAL RESULTS

Prior to segmentation, the hyperspectral data undergoes radiometric calibration, edge-band removal, and per-band normalization using the median absolute deviation. Dimensionality reduction via PCA is optionally applied. The spectral reference library can be either manually annotated spectra or automatically generated by clustering high-saliency pixels using k-means with k between 8 and 16. Model hyperparameters are set as follows: pseudo-spectral prompt threshold $\tau_p = 0.6$, spectral verification threshold $\tau_s = 0.5$, uncertainty threshold $\tau_u = 0.15$, exponential smoothing factor $\alpha = 0.8$, maximum number of iterations $L_{\max} = 6$, and convergence



(a) Ground truth (b) Heatmap

(c) Ground truth (d) Heatmap

(e) Ground truth (f) Heatmap

Figure 2: Qualitative HLSS results. Red: high spectral confidence; Blue: low confidence. HLSS preserves spectral precision while localizing confident regions across heterogeneous materials.

Table 2: Comparison on TECNALIA WEEE HSI dataset. Supervised HSI baselines are reported from [17].

Method	Input type	Precision	Recall	F1	Mean IoU
U-Net (VGG-style), 76 bands [17]	HSI	0.83	0.87	0.84	0.73
Encoder–Decoder, 76 bands [17]	HSI	0.84	0.84	0.84	0.72
U-Net, 7 bands [17]	HSI	0.81	**0.89**	0.84	**0.74**
HLSS (ours)	HSI + SAM	**0.988**	0.750	**0.85**	0.64

tolerance $\epsilon = 0.005$. Experiments are conducted on the TECNALIA WEEE HSI dataset [17], which represents industrial conveyor waste with heterogeneous materials. Performance is evaluated using both instance-level metrics, including precision, recall, F1 score, and mean Intersection-over-Union (IoU) per cluster, and pixel-level metrics for precision, recall, F1, and mean IoU. For comparison, we include fully supervised hyperspectral baselines such as U-Net (VGG-style, 76 bands), Encoder–Decoder (76 bands), and U-Net (7 bands) [17], as well as RGB-only approaches like Mask R-CNN and PCA-RGB [8]. Additionally, we report results for SAM-only, in which promptable proposals are generated without any spectral verification [10]. Ablation studies investigate the contributions of pseudo-spectral prompting, per-pixel classifier refinement, geometric refinement, and residual removal policy.

Quantitative Results. Table 2 summarizes instance-level segmentation results on the TECNALIA WEEE HSI dataset. HLSS achieves the highest spectral precision due to its selective acceptance of proposals, though recall is slightly lower than fully supervised HSI networks, reflecting a conservative detection strategy. This highlights the framework's ability to prioritize material coherence while avoiding false positives. While HLSS attains competitive F1 scores, its lower mean IoU and recall stem from its conservative spectral verification stage, which intentionally rejects ambiguous or spectrally mixed regions. This behavior improves precision but may fragment object masks or omit low-confidence regions, thereby reducing overlap-based metrics such as IoU. In particular, partial acceptance of object regions can lead to under-segmentation, which directly impacts IoU despite correct material identification.

To address this limitation, several strategies can be considered. First, adaptive thresholding of the spectral similarity and variance criteria could relax acceptance conditions in later iterations, improving recall while preserving early-stage precision. Second, incorporating region-growing or mask refinement steps after verification may recover missing object extents. Third, learned verification policies (e.g., lightweight neural classifiers) could replace fixed thresholds, enabling better calibration between precision and recall. Finally, multi-scale or ensemble proposal strategies may increase candidate diversity, reducing missed detections. These directions represent promising extensions for improving overlap-based performance without sacrificing HLSS's material-awareness.

Qualitative Results. Figure 2 presents representative qualitative results. HLSS preserves spectral precision while effectively localizing confident regions across heterogeneous materials. High spectral confidence areas are indicated in red, while low-confidence regions appear in blue. Across multiple scenes, HLSS demonstrates the

ability to segment complex material arrangements and recover occluded objects, consistent with the quantitative evaluation.

Runtime and Complexity. The theoretical per-iteration complexity of HLSS is $O(KT_M + \sum_m A_m B)$, where T_M denotes the computational cost of the foundation model, B the number of spectral bands, A_m the area of each mask, and K the number of proposals. In practice, runtime benchmarks show significant speedups on GPU compared to CPU execution. Table 3 summarizes runtime and estimated energy consumption per image for different devices.

Table 3: Runtime and estimated energy per image.

Device	Time (s)	Speedup	Energy (Wh)
CPU (12 cores)	2.47	1.0×	0.17
GPU (estimated)	0.21	12.0×	0.03

6 CONCLUSION AND FUTURE WORK

We presented HLSS as an **agent-based** layered perception framework that combines foundation-model spatial proposals with hyperspectral verification and monotonic residual updates. By organizing the pipeline into proposal, verification, and residual agents operating over a shared residual hyperspectral cube, HLSS provides a modular, interpretable decomposition of cluttered scenes and a clear interface for future learning or decentralization.

We formalized the problem, characterized residual masking dynamics, and showed that the greedy layered selection yields finite termination in a discrete domain. Empirically, HLSS attains high spectral precision and produces interpretable confidence maps on industrial WEEE imagery while remaining training-light.

Future work will address current limitations and extend HLSS capabilities. Improvements may include learned spectral verifiers (e.g., transformer-based), tighter multimodal fusion within foundation models, and self-supervised pretraining on large-scale hyperspectral datasets to reduce dependence on reference libraries. Further large-scale evaluations across diverse industrial HSI scenarios will assess robustness, generalization, and efficiency, supporting the integration of HLSS into high-throughput recycling in an ethically responsible manner.

ACKNOWLEDGEMENTS

The iBot4CRMs project, leading to this paper, has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101189783.

REFERENCES

[1] Assia Belbachir, Amal El Fallah-Seghrouchni, Arthur Casals, and Marcia Pasin. 2019. Smart Mobility Using Multi-Agent System. *Procedia Computer Science* 151 (2019), 447–454. https://doi.org/10.1016/j.procs.2019.04.061 The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops.

[2] Salima Bella, Assia Belbachir, and Ghalem Belalem. 2020. A hybrid air-sea cooperative approach combined with a swarm trajectory planning method. *Paladyn, Journal of Behavioral Robotics* 11, 1 (2020), 118–139. https://doi.org/doi:10.1515/pjbr-2020-0006

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 4 (2018), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

[4] Zhiwei Chen, Yuan Li, and Fang Xu. 2024. Multimodal Segment Anything Model for Infrared and RGB Fusion Segmentation. *Sensors* 24, 3 (2024), 982.

[5] Edmund H. Durfee, Victor R. Lesser, and Daniel D. Corkill. 1989. Distributed problem solving and planning. *Distributed Artificial Intelligence* 2 (1989), 121–164.

[6] Jiayao Feng, Dongli Sun, and Hongbo Zhang. 2017. Hyperspectral imaging for detection and classification of plastic waste. *Waste Management* 60 (2017), 56–65.

[7] Önder Gürcan, Oguz Dikenelli, and Carole Bernon. 2013. A generic testing framework for agent-based simulation models. *Journal of Simulation* 7, 3 (01 Aug 2013), 183–201. https://doi.org/10.1057/jos.2012.26

[8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2980–2988. https://doi.org/10.1109/ICCV.2017.322

[9] Liang Ke, Yu-Wing Tai, and Chi-Keung Tang. 2021. BCNet: Boundary-aware Instance Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[10] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371

[11] Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and {Jon Atli} Benediktsson. 2019. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing* 57, 9 (Sept. 2019), 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932

[12] Y. Li, M. Zhang, and Q. Wang. 2022. Spectral-Spatial Transformer Networks for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–12.

[13] Sheng Liu, Lin Wang, and Rui Zhang. 2023. Occlusion-aware Instance Segmentation using Iterative Residual Learning. *Pattern Recognition* 141 (2023), 109530.

[14] Francesco Locatello, Dirk Weissenborn, and et al. 2020. Object-Centric Learning with Slot Attention. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[15] Daniel Lorente and et al. 2012. Hyperspectral image analysis for food quality and safety assessment: Review. *Trends in Food Science & Technology* 31, 1 (2012), 34–47.

[16] Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray. 2007. Consensus and cooperation in networked multi-agent systems. *Proc. IEEE* 95, 1 (2007), 215–233.

[17] Artzai Picon, Pablo Galán, Arantza Bereciartua-Perez, Leire Benito del Valle, et al. 2025. On the analysis of adapting deep learning methods to hyperspectral imaging: Use case for WEEE recycling and dataset. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 330 (2025), 125665.

[18] Jie Ren, Hao Zhang, and Weijian Liu. 2024. Grounded SAM: Accurate Visual Grounding with Segment Anything Model. *IEEE Transactions on Image Processing* (2024).

[19] Tao Wang, Yuanzheng Cai, Lingyu Liang, and Dongyi Ye. 2020. A Multi-Level Approach to Waste Object Segmentation. *Sensors* 20, 14 (2020), 3816. https://doi.org/10.3390/s20143816

[20] Michael Wooldridge. 2009. *An Introduction to MultiAgent Systems*. Wiley.

[21] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control* (2021), 321–384.