IEA/AIE 2026 — Kuala Lumpur, Malaysia

Paper #160: Industrial Visual Anomaly Detection in Robotics: Methods, Datasets, and Deployable System Architectures

AUTHORS: Thanh-Hai Tran

Xuan-Bach Le*

Ho Chi Minh City University of Technology, VNU-HCM

Outline

Why IVAD for Robotics?

The Problem

- Cobots in **shared workspaces** need context-aware safety [1]
- Binary sensors → **no semantic reasoning**
- Labeled anomaly data is **extremely scarce** [2]

Key Insight

- Learn "**normal**" → flag deviations
- No labeled fault data needed

Gap in Existing Surveys

Focus on **static inspection** [2]

→ **Miss:**

- Robot-specific anomaly types
- Real-time edge deployment
- System design for ROS2

Our Contributions

C1

Robot-centric anomaly taxonomy

C2

Deployment-aware method review

C3

ROS2 design & open challenges

Scope: Survey of deep learning methods for IVAD in robotics, viewed through the lens of real deployment.

Anomaly Taxonomy in Robotic Systems

Structural

Surface deviations

(cracks, scratches, debris)

Logical

Assembly errors

(missing/misplaced parts)

Semantic

Context violations

(safety zones, human proximity)

Temporal

Kinematic drifts

(trajectory errors, vibrations)

Four anomaly types: each requires a different detection strategy – no single method covers all.

Datasets: General vs. Robot-Specific

Dataset	Domain	Modality	Anomaly Type	Dynamic?
MVTec-AD [3]	General	Image	Structural	×
MVTec-LOCO [4]	General	Image	Logical	×
VisA [5]	General	Image	Complex	×
MVTec-AD-2 [6]	General	Image	Structural	Partial
RAD [7]	**Robot**	Multi-view	Struct+Logic	✓
AURSAD [8]	**Robot**	Force/Torque	Temporal	✓
Hazards&Robots [9]	**Robot**	Video	Semantic	✓

Gap: No single benchmark covers dynamic views + semantic safety + multimodal + real-time

Deep Learning Methods: Three Generations

Reconstruction

AE [10], GAN [11]
MemAE, LSTM-VAE [12]

- ✓ Interpretable
- × Over-generalizes
- × No semantic

Embedding

PatchCore [13],
EfficientAD [14]
FastFlow [15], SALAD [16]

- ✓ Fast (2–65 ms)
- ✓ Edge-ready
- × Limited logic

VLM

CLIP [17], WinCLIP [18]
AnomalyGPT [19]

- ✓ Zero-shot
- ✓ Semantic
- × >500 ms latency

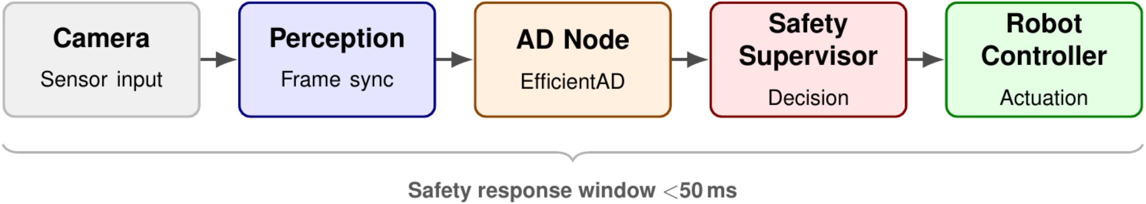
Performance Comparison

Method	Category	Logic AD	Latency	Deploy
PatchCore [13]	Memory Bank	Limited	$\sim$32 ms	Offline/Edge
EfficientAD [14]	Distillation	Moderate†	**2.2 ms**	**Real-time**
FastFlow [15]	Flow-based	Limited	$\sim$17 ms	Near RT
SALAD [16]	Semantic Emb.	**Strong**	$\sim$65 ms	Offline/Near RT
AnomalyGPT [19]	VLM	**Strong**	$>$500 ms	Cloud-only

†85.8% image-level AUROC on MVTec-LOCO (logical)

Trade-off: Better logical/semantic $\Longleftrightarrow$ Higher compute
$\rightarrow$ **Hybrid:** fast local detector (real-time) + async semantic module

System Design: Edge Deployment & ROS2



Edge Platform

- NVIDIA Jetson (Nano → Orin)
- TensorRT [20]: FP32 → INT8 (2–5×)

Design Principles

- Modular ROS2 pub-sub nodes [21]
- Cloud offload for heavy VLM

Open Challenges & Future Directions

Research Gaps

- **Logical AD** — no real-time method
- **Sim-to-real** [22] — domain gap
- **Explainability** — ISO/TS 15066
- **Dynamic backgrounds** — false alarms
- **Continual learning** [23] at edge
- **Multi-robot** coordination

Proposed Directions

- **Hybrid VLM + lightweight**
  VLM $\rightarrow$ pseudo-labels $\rightarrow$ edge
- **Multimodal fusion**
  Vision + force + audio
- **Few-shot logical AD**
  Meta-learning + synthetic [24]
- **Full-stack reference**
  ROS2-compatible open system

References

[1] Amaya-Mejia, L.M., et al.: Vision-based safety for barrierless HRC. arXiv:2208.02010 (2022)

[2] Shukla, V., et al.: Survey: DL-based image AD in industrial contexts. Front. Rob. AI 12 (2025)

[3] Bergmann, P., et al.: MVTec AD dataset for UAD. In: CVPR, pp. 9592-9600 (2019)

[4] Bergmann, P., et al.: Logical constraints in UAD. IJCV 130, 947-969 (2022)

[5] Zou, Y., et al.: SPot-the-Difference self-supervised AD. In: ECCV, pp. 392-408 (2022)

[6] Heckler-Kram, L., et al.: The MVTec AD 2 Dataset. arXiv:2503.21622 (2025)

[7] Zhang, K., et al.: RAD: Benchmark for AD in robotics. arXiv:2410.00713 (2024)

[8] Leporowski, B., et al.: AURSAD: Screwdriving AD Dataset. DiB 35, 106880 (2021)

[9] Mantegazza, D., et al.: Hazards&Robots dataset for AD in robotics. DiB 48, 109264 (2023)

[10] Gong, D., et al.: Memorizing normality to detect anomaly: MemAE. In: ICCV (2019)

[11] Schlegl, T., et al.: Unsupervised AD with GANs. In: IPMI, pp. 146-157 (2017)

[12] Park, D., et al.: Multimodal AD for robot-assisted feeding. IEEE RA-L 3(3) (2018)

[13] Roth, K., et al.: Towards total recall in industrial AD. In: CVPR (2022)

[14] Batzner, K., et al.: EfficientAD: Visual AD at millisecond latencies. In: WACV (2024)

[15] Yu, J., et al.: FastFlow: Unsupervised AD via 2D Normalizing Flows. arXiv (2021)

[16] Fučka, M., et al.: SALAD: Semantics-Aware Logical AD. In: ICCV (2023)

[17] Radford, A., et al.: Learning transferable visual models (CLIP). In: ICML (2021)

[18] Jeong, J., et al.: WinCLIP: Zero-/Few-Shot anomaly classification. In: CVPR (2023)

[19] Gu, Z., et al.: AnomalyGPT: Industrial AD using large VLMs. In: AAAI (2024)

[20] NVIDIA: TensorRT – High performance DL inference. Tech Doc (2023)

[21] Antmicro: Dedicated ROS 2 nodes for AI computer vision. (2023)

[22] James, S., et al.: Sim-to-real via sim-to-sim. In: CVPR, pp. 12627-12637 (2019)

[23] Kirkpatrick, J., et al.: Overcoming catastrophic forgetting. PNAS 114(13) (2017)

[24] Tong, X., et al.: Component-aware unsupervised logical anomaly gen. arXiv (2025)

Key Takeaways

- **Benchmark $\neq$ Real-world**

High AUROC $\not\Rightarrow$ robustness in dynamic robot scenes

- **Safety needs semantics**

Context-aware understanding, not just local features

- **Real-time is non-negotiable**

Low-latency edge design = first-class constraint

Thank You!

Questions & Discussion

Thanh-Hai Tran

tthai.sdh241@hcmut.edu.vn

Xuan-Bach Le

lexuanbach@hcmut.edu.vn

Ho Chi Minh City University of Technology