Autonomous UAV systems with few-shot visual intelligence for process-aware construction monitoring

Yiquan Zou,[1] Tianxiang Liang,[1,*] Zhendong Xu,[2] Mudan Wang,[3] Ioannis Brilakis,[3] and Biao Xiong[2,4,*]

[1]School of Civil Engineering, Architecture and The Environment, Hubei University of Technology, Wuhan 430068, China

[2]China Construction Third Engineering Bureau Technology Innovation Industry Development Co., Ltd., Wuhan 430050, China

[3]Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom

[4]School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China

*Correspondence: liangliangtovo@gmail.com (T.L.); b.xiong@whut.edu.cn (B.X.)

Received: December 17, 2025; Accepted: January 27, 2026; Published Online: January 30, 2026; https://doi.org/10.59717/j.xinn-inform.2026.100018

Citation: Zou Y., Liang T., Xu Z., et al. (2026). Autonomous UAV systems with few-shot visual intelligence for process-aware construction monitoring. *The Innovation Informatics* **2**:100018.

INTRODUCTION: WHY PROCESS-LEVEL AUTOMATION MATTERS NOW

Accurate and timely monitoring of on-site construction progress is critical for schedule control, risk mitigation, and decision-making in large-scale projects, yet it remains largely dependent on manual inspection and subjective judgment. As construction environments become increasingly dynamic and data-rich, this reliance exposes a fundamental limitation in current progress sensing: the lack of scalable visual intelligence capable of reasoning about which processes are occurring under minimal supervision. This editorial argues that autonomous, dock-enabled UAV systems combined with few-shot learning (FSL) offer a principled solution to this challenge. By integrating repeatable autonomous data acquisition with data-efficient representation learning, process-oriented region-of-interest focusing, and multi-view inference, such systems enable continuous, objective, and automatic progress perception. Real-world deployment demonstrates their potential to reveal early deviations and reduce managerial uncertainty.

Accurate and timely understanding of on-site construction progress is central to schedule control, risk mitigation, and operational decision-making in large-scale projects. Yet, on many construction sites, progress remains only partially observable, relying on sporadic inspections and subjective interpretations. As projects grow in scale and temporal coupling, this lack of continuous, process-level visibility increasingly limits managers' ability to detect emerging risks and intervene effectively.

While recent advances in sensing and computer vision have improved site documentation, they largely focus on identifying visible components rather than inferring which construction processes are actively unfolding. This gap between visual observation and process-level understanding has become a structural bottleneck for digital construction management.

Addressing this gap requires a shift from episodic, human-centered inspection to continuous, autonomous, and learning-driven process inference under limited supervision. Recent studies in construction vision[1] and UAV-based monitoring[2] increasingly recognize that the bottleneck is no longer the sensing hardware, but the lack of scalable visual intelligence that can generalize across evolving scenes with minimal human supervision.[3] In this context, dock-enabled autonomous UAV systems, combined with FSL[4] offer a timely pathway toward persistent, process-aware site perception, positioning process-level automation as a critical frontier for both intelligent construction management and real-world machine intelligence.[5]

A DOCK-ENABLED FEW-SHOT FRAMEWORK FOR PROCESS-LEVEL PROGRESS MONITORING

Autonomous, dock-enabled UAV systems transform visual data acquisition on construction sites from sporadic documentation into persistent

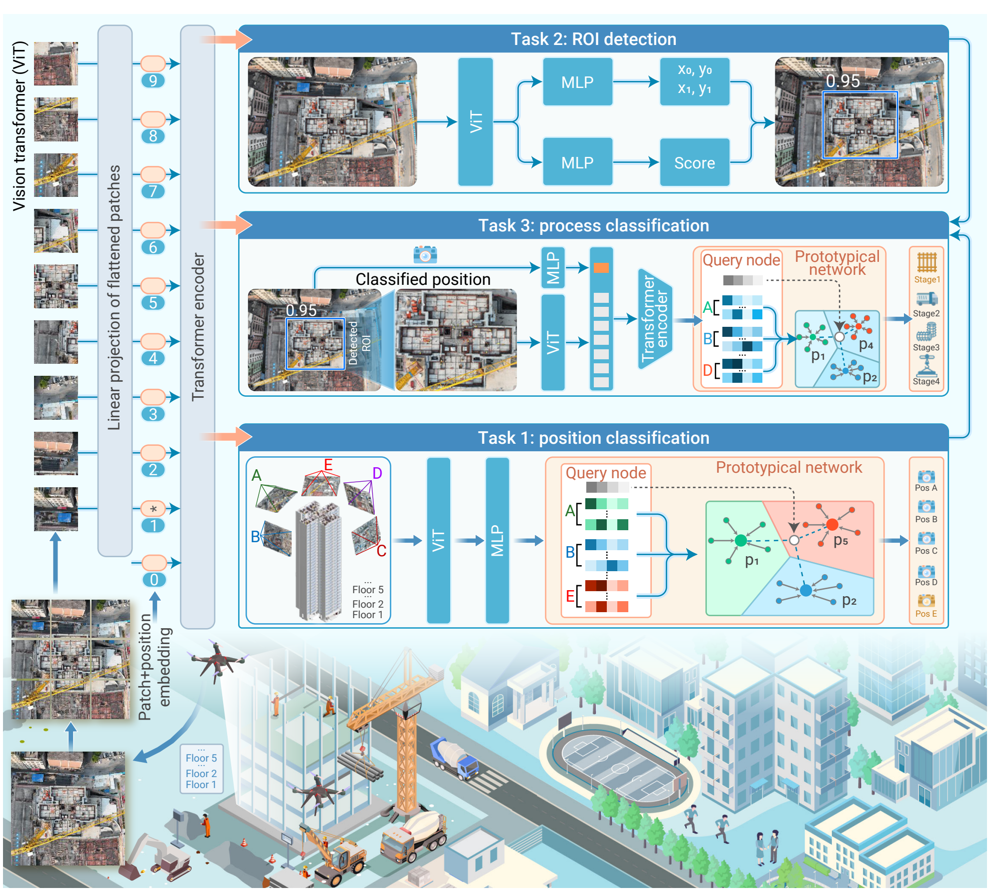

Figure 1. Dock-enabled few-shot framework for process-aware construction monitoring Autonomous UAVs perform repeatable multi-view data acquisition, which is transformed into process-level progress estimates through region-of-interest focusing, exemplar-based few-shot representation learning, and multi-view inference under minimal supervision.

sensing. By automating flight, charging, and data transmission, they enable repeatable image collection from stable viewpoints, providing a reliable foundation for continuous site observation that cannot be sustained manually. However, abundant data does not guarantee understanding. Construction imagery varies rapidly with lighting, occlusion, and shifting workfaces, while nominally identical processes often appear visually different across floors and projects. Interpreting such data requires fine-grained, process-level judgment under conditions of strong visual heterogeneity, making large-scale annotation costly, brittle, and poorly transferable.

These conditions render construction progress monitoring inherently a few-shot learning problem. Only limited exemplars are available at early stages, and later stages follow recurring yet non-identical patterns. Effective monitoring therefore demands models that can generalize from minimal supervision and remain stable under evolving visual distributions, aligning closely with the core goals of few-shot learning.

To address this gap, we propose a dock-enabled few-shot framework that couples autonomous data acquisition with lightweight, process-aware visual reasoning. As illustrated in Figure 1, the framework is built upon a shared Vision Transformer (ViT) backbone that learns global visual representations and supports three tightly coupled tasks: Position Classification, Region-of-Interest (ROI) Detection, and Process Classification.

(1) Position Classification. This task estimates the UAV viewpoint and floor context, producing spatial priors that constrain downstream inference. By accounting for viewpoint-induced appearance variation, it reduces sensitivity to perspective changes which can otherwise degrade recognition performance under limited data.

(2) ROI Detection. This task identifies visually informative regions corresponding to active work areas, suppressing irrelevant background content. Focusing inference on these regions increases the effective signal-to-noise ratio and emphasizes process-discriminative visual cues.

(3) Process Classification. Conditioned on the spatial priors from task 1 and the focused regions from task 2, this task performs metric-based few-shot learning by matching region representations to a small set of process exemplars, enabling robust and data-efficient process recognition with minimal human supervision.

Together, these tasks form a data-efficient perception pipeline that transforms UAV imagery from passive visual records into structured, process-level signals, supporting continuous and objective progress monitoring and offering a practical pathway toward data-efficient machine intelligence in complex real-world environments.

ENGINEERING IMPACT: FROM EARLY DEVIATION DETECTION TO ACTIONABLE SITE INTELLIGENCE

The engineering impact of autonomous few-shot monitoring extends beyond classification performance to the way construction progress is perceived and acted upon. By converting daily UAV imagery into structured, interpretable process-level signals, such systems provide a form of site intelligence that supports timely and evidence-based decision-making, surpassing what can be achieved through manual inspection alone.

A central benefit is early deviation detection. Schedule slippage rarely occurs abruptly; instead, it emerges through small, accumulating inefficiencies in sequencing, turnover, or logistics. Continuous, automated observation makes these weak signals observable at an early stage, enabling proactive intervention before delays become entrenched.

Equally important is the creation of objective and traceable progress records. Daily process labels form a coherent timeline of site activity, reducing interpretive ambiguity across stakeholders and supporting transparent oversight. In real-world deployment, this improved visibility has enabled managers to identify a gradually accumulating delay, adjust resources accordingly, and recover productive time—demonstrating how enhanced observability, rather than additional inputs, can materially improve project outcomes.

FROM AUTOMATION TO INTELLIGENCE: TOWARD MINIMAL-SUPERVISION CONSTRUCTION

The adoption of autonomous UAV monitoring signals a shift from automating data collection to enabling intelligent decision-making in construction management. This transition is driven by three converging trends:

Autonomous sensing as infrastructure. Dock-enabled UAVs provide continuous, repeatable visual data without the need for on-site pilots.

FSL for data-efficient inference. By leveraging inductive bias and similarity structure, FSL supports process inference under limited supervision.

Multi-view fusion. Integrating complementary viewpoints helps stabilize predictions in dynamic, three-dimensional environments.

Together, these trends enable continuous, objective progress monitoring with minimal human intervention, replacing fragmented visual inspection with structured, process-level perception. As such capabilities integrate with digital twins and model-driven planning frameworks, progress monitoring can evolve from retrospective reporting toward anticipatory reasoning, supporting earlier intervention and more resilient project control.

FUTURE RESEARCH DIRECTIONS: FROM RECOGNITION TO DEEPER UNDERSTANDING

While the dock-enabled few-shot paradigm provides a practical starting point, advancing beyond process recognition raises several focused research challenges. A primary issue is cross-project generalization. Future work should systematically examine how process representations learned from one site can transfer across different building types, construction methods, and organizational practices, and should identify which visual and spatial priors remain stable under domain shift. A second challenge lies in temporal and transitional-state modeling. Because construction activities exhibit strong sequential dependencies, incorporating information from preceding stages as temporal priors may improve process inference. Explicitly modeling transitional and overlapping states would more temporally consistent reasoning under uncertainty. Finally, greater impact requires tighter integration with planning and decision models. Linking process-level predictions with digital twins and 4D BIM scheduling frameworks could help close the loop between perception and planning, allowing detected deviations to inform schedule forecasts and intervention strategies. Together, these directions move progress monitoring from static recognition toward explanatory and anticipatory intelligence.

CONCLUSION: A PRACTICAL PATH TOWARD INTELLIGENT SITE MONITORING

Autonomous few-shot UAV monitoring represents more than an incremental technological advance; it also reflects a shift in how construction sites are perceived and reasoned about. By combining autonomous sensing with data-efficient process inference, this paradigm addresses long-standing limitations in progress monitoring arising from visual variability, annotation scarcity, and reliance on subjective inspection. Crucially, the approach is practical and scalable. It reduces dependence on manual supervision, improves the objectivity and continuity of progress perception, and transforms raw visual data into actionable insights for decision-making. As construction continues its digital evolution, such autonomous and minimally supervised monitoring systems are likely to become a foundational component of modern project delivery, while simultaneously serving as a compelling real-world testbed for data-efficient machine intelligence.

REFERENCES

1. Liu Y., Liu C., Wang B., et al. (2025). CAGE: Continuity-aware edge network unlocks robust floorplan reconstruction. *arXiv* preprint. DOI:10.48550/arXiv.2509.15459
2. Meng Y., Bing Z., Yao X., et al. (2025). Preserving and combining knowledge in robotic lifelong reinforcement learning. *Nat. Mach. Intell.* **7**:256–269. DOI:10.1038/s42256-025-00983-2
3. Yu H., Mineyev I., Varshney L.R., et al. (2025). Learning from one and only one shot. *npj Artif. Intell.* **1**:13. DOI:10.1038/s44387-025-00017-7
4. Snell J., Swersky K. and Zemel R. (2017). Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **30**. https://proceedings.neurips.cc/paper_files/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html
5. Zhang J., Li A., Qi Y., et al. (2025). Embodied navigation foundation model. *arXiv* preprint. DOI:10.48550/arXiv.2509.12129

FUNDING AND ACKNOWLEDGMENTS

This work was supported by the Major Research Project of Hubei Province (Grant No. (JD)2023BAA007), the National Natural Science Foundation of China (Grant No. 61901311), and the Hubei Provincial Science and Technology Program (Grant No. 2025BCB033). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Y.Q.Z. and T.X.L. conceived and designed the study. T.X.L. and I.B. developed the methodology. T.X.L. and M.W. were responsible for the software development and visualization. T.X.L. and I.B. performed the formal analysis. T.X.L. carried out the investigation and data curation with support from Z.D.X. and M.W. T.X.L. wrote the original draft of the manuscript. B.X., I.B., and Y.Q.Z. supervised the research and provided critical revisions. Y.Q.Z., Z.D.X., and B.X. were responsible for project administration and funding acquisition. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AND CODE AVAILABILITY

Data are available from the corresponding author upon reasonable request.