UAV-enabled Dual-scenario Empowerment for Urban-rural Spaces: A Comparative Study of Immersive E-commerce Platforms and Intelligent Building Facade Inspection Systems

Qi Li[1], Shaoqi Xu[2,*], Wantong Wang[2], Yufei Zhang[2]
[1]College of Life Sciences, Zhejiang Normal University, Jinhua 321000, China
[2]College of Education, Zhejiang Normal University, Jinhua 321000, China
*Corresponding email: 2408355851@qq.com

Abstract

The integration of unmanned aerial vehicles (UAV) remote sensing, virtual reality (VR) panoramic display, and artificial intelligence (AI) has emerged as a transformative paradigm for urban-rural spatial governance. However, existing research predominantly concentrates on single-technology applications within isolated contexts, lacking systematic examination of cross-scenario technology transfer. This paper proposes and validates a tripartite technological framework integrating UAV, VR and AI technologies. It further verifies the practical applicability of the framework via two typical engineering cases. The first case is the "Yunshang E-village", an immersive rural e-commerce empowerment platform. It adopts UAV aerial photography and VR panoramic technology to connect idle rural resources with consumer markets, effectively solving the market docking dilemma of underdeveloped traditional villages. The second case refers to the "Intelligent Wing Building Inspection", an intelligent facade damage detection system tailored for super high-rise buildings. Combined with improved YOLOv5 Large (YOLOv5l) algorithms and edge computing technology, this system realizes automatic building defect identification and standard compliance diagnosis. Empirical findings demonstrate that in the rural context, VR panoramic technology increased visitor arrivals by 45.00%-120.00% and monthly e-commerce sales of agricultural specialty products by 256.00%-339.00% across four partner villages. In the urban building inspection context, the system achieved 95.80% defect recognition accuracy, compressed inspection time for a 150-meter building from 3 days to under 3 hours, reduced costs by 82.40%, and lowered the miss rate to below 2.00%. This study contributes a comparative analytical framework for understanding cross-scenario technology transfer and provides practical pathways for digital transformation in both rural revitalization and urban safety governance.

Keywords

Artificial intelligence, Building facade inspection, Rural e-commerce, Unmanned aerial vehicles, Virtual reality

Introduction

The accelerated advancement of digital China and rural revitalization strategies has positioned the integrated application of UAV, VR, and AI as a critical pathway for empowering urban-rural spatial governance. These technologies collectively form a versatile technological paradigm with theoretical applicability across multiple scales and contexts - from vast rural landscapes requiring digital representation of cultural and geographic resources, to dense urban environments demanding precise perception and intelligent diagnosis of building infrastructure.

Nevertheless, existing scholarship remains largely confined to vertical optimization within single-technology domains or isolated case explorations, lacking systematic theoretical construction and comparative analysis of this technological paradigm across different urban-rural spatial dimensions.

At the policy level, China is experiencing a strategic superposition of urban and rural digital transformation initiatives. In the rural domain, the Rural Revitalization Strategy, the High-quality Development of Rural e-commerce Project, and the Digital Village Construction Initiative are advancing in coordination. China's No. 1 Central Document for 2025, explicitly called for "promoting the deep integration of rural culture and tourism" [1].

In 2026, the Ministry of Commerce and five other departments jointly issued the Guiding Opinions on Better Serving the Real Economy and Promoting the High-quality Development of E-commerce, further specifying the need to "deepen rural e-commerce and implement the high-quality development project for rural e-commerce" [2].

Despite these policy tailwinds, a substantial number of traditional villages continue to struggle with the predicament of having resources but no monetization, having culture but no dissemination. Urban-rural information asymmetry prevents urban consumers from accessing rural cultural tourism resources, while the supply of rural tourism products remains structurally imbalanced. Indigenous rural culture is accelerating toward extinction due to the absence of digital dissemination channels.

In the urban sector, the Ministry of Housing and Urban-rural Development issued the 14th Five-Year Plan for the Development of the Construction Industry. This document sets out clear requirements in two key areas. First, it requires faster research, development and practical application of construction robots. Second, it encouraged the use of UAV and other intelligent devices. Such equipment can take over manual work in scenarios including surveying, monitoring, routine inspection and site safety management [3].

By August 2025, China's total high-rise buildings had reached 347,000, with over 6,000 super high-rise buildings exceeding 100 meters - both figures ranking first globally. Yet traditional manual "spiderman" high-altitude inspection presents a safety accident rate of 1.20%, requires three days to inspect a single 150-meter building, and achieves a micro-crack recognition rate below 65.00% for defects under 0.5 mm [4].

The inspection coverage rate remains below 30.00%, and subjective judgment error in hollow drum detection by manual tapping exceeds 30.00%. Single-technology alternatives such as infrared thermography are constrained by equipment bulk (exceeding 500,000 RMB per unit), strong light dependency, and blindness to cracks below 0.2 mm [5]. These two spatial contexts - rural and urban - share a common underlying challenge: How to overcome physical space constraints through digital technology to achieve efficient matching between resources and demand?

From a technological evolution perspective, UAV remote sensing has expanded from military applications to diverse civilian domains including agricultural surveying, environmental monitoring, and urban planning, offering advantages of high precision, low cost, and flexible deployment. VR technology has migrated from immersive entertainment to real-economy sectors including education, healthcare, cultural tourism, and e-commerce, with 4K high-definition panoramic and hotspot interaction technologies reaching maturity.

Concurrently, deep learning object detection algorithms exemplified by the You Only Look Once (YOLO) series have demonstrated robust capabilities in industrial defect detection and remote sensing image analysis [6]. The convergence of these technological developments enables the construction of a "UAV collection + VR presentation + AI recognition" general-purpose technological framework.

Nevertheless, most existing studies focus largely on the vertical optimization of separate technologies. For instance, research on VR tourism centers on the technical realization of immersive experiences. By contrast, studies concerning building inspections are mainly devoted to boosting the accuracy of relevant algorithms. However, two aspects remain insufficient in current scholarship. First, there is a lack of systematic design for the full-chain technological framework of "perception-presentation-decision".

Second, few works conduct comparative analyses on how this framework is applied in diverse practical scenarios. The core research questions of this paper are therefore: How can a general-purpose technological framework integrating UAV modeling, VR panoramic display, and AI intelligent recognition be constructed? What differentiated empowerment mechanisms does this framework exhibit across different urban-rural spatial dimensions (rural cultural tourism versus urban building inspection)? What commonalities and differences exist between the two scenarios in terms of technological pathways, value logic, and business models?

To address these questions, this paper selects two representative projects with completed implementation validation for in-depth comparative case analysis.

Case one, "Yunshang E-village", is an immersive rural e-commerce empowerment platform based on VR panoramic and UAV modeling technologies. This project targets villages that possess valuable cultural heritage yet suffer from insufficient development. It adopts three technical means: UAV aerial photography, 4K resolution (4K) panoramic image production, and panoramic hotspot-linked e-commerce integration.

These technologies jointly build an all-in-one service platform. The platform supports two core functions: immersive online village sightseeing and agricultural product online shopping. The project has been fully rolled out in four villages. Representative sites include Putang Village in Jinhua and the Former Residence of Shi Guangnan.

Case two: Intelligent Wing Building Inspection is a UAV-powered intelligent system for detecting damages on building exterior facades. Traditional manual inspection is dangerous and inefficient. Detection relying solely on one single technology also delivers unstable outcomes. This system solves these widespread industry problems. It applies vertical layered adaptive path planning with a 10.00% overlap rate, multi-source data fusion and edge computing framework. It achieves precise identification of eight defect categories at an accuracy rate above 95.00% and automatically outputs standardized compliance inspection reports. Practical pilot tests have been completed on 12 super high-rise buildings across Zhejiang and Hubei provinces. For details, see Table 1 and Table 2.

Table 1. Intelligent Wing Building Inspection technical parameters and core metrics.

Technical module	Core parameters	Target metric	Achieved metric
UAV platform	DJI Mavic 3 enterprise real-time kinematic (RTK)	RTK error ≤2 cm	2 cm
Dataset scale	8 defect categories, 3 materials ×3 environments	≥3000 images	3,982 images
YOLOv5l optimization	CBAM + BiFPN + CIoU Loss	mAP0.5 ≥75.00%	76.30%
Recognition accuracy	8 defect categories	≥95.00%	95.80%
Inference speed	Edge terminal	≥30 fps	32 fps
Area quantification	MATLAB sobel segmentation	Error ≤5.00%	≤5.00%
Semantic mapping	BERT 8→23 categories	Accuracy ≥95.00%	98.00%
Inspection efficiency	Single 150m building	≤3 hours	2 hours 42 minutes

Table 2. Intelligent Wing Building Inspection damage classification standards.

Damage grade	Defect area (cm^2)	Maintenance recommendation	Recognition accuracy (%)
Minor	<30 cm²	Regular observation, record in database	≥90.00%
Moderate	30-100 cm²	Schedule rectification within quarter	≥95.00%
Severe	>100 cm²	Immediate repair, establish warning zone	≥98.00%

The theoretical contributions of this paper are threefold. First, it proposes a “UAV + VR + AI” tripartite technological paradigm framework (collection layer - presentation layer - decision layer), providing analytical tools for understanding the empowerment logic of multi-technology integration.

Second, through comparative case analysis of two scenarios, it reveals the differentiated empowerment mechanisms of this technological paradigm in rural “resource activation” versus urban “safety maintenance” contexts.

Third, it distills key success factors for cross-scenario technology transfer, offering theoretical support and practical reference for coordinated urban-rural digital transformation.

The remainder of this paper is organized as follows. The section presents the literature review and theoretical foundation, systematically reviewing relevant research in UAV remote sensing, VR panoramic technology, and deep learning object detection, and proposing the collection-presentation-decision tripartite technological paradigm analytical framework.

This section details the research design, including the background, technological architecture, and research methods of the two cases. Another section presents empirical findings from both cases, analyzed across technological implementation, application efficacy, and business value dimensions. And it provides cross-case comparison and discussion, distilling common patterns and differentiated mechanisms. Finally, it concludes with theoretical contributions and practical implications.

Literature review and theoretical foundation

UAV remote sensing applications in urban and rural contexts

UAV remote sensing technology has achieved widespread application across diverse urban and rural settings in recent years. In rural contexts, UAV are primarily deployed for agricultural plant protection, land tenure confirmation, resource surveying, and landscape census.

Compared with satellites, manned aircraft, and ground systems, UAV provide higher spatial and temporal resolution, operational flexibility, and cost-effectiveness, leveraging Red, Green, Blue (RGB), multispectral, hyperspectral, Light Detection and Ranging (LiDAR), and thermal sensors. Empirical evidence demonstrates UAV utility for soil nutrient estimation, early pest and disease detection using vegetation indices, precision spraying, rapid damage assessments, high-accuracy mapping, and enhanced phenotyping and yield estimation when combined with machine learning [7].

Within the cultural tourism domain, UAV aerial photography has become the predominant method for digital acquisition of landscape resources, with its high resolution and multi-perspective capabilities providing high-quality material foundations for subsequent VR panoramic production.

Research has drawn the following conclusion. UAV fitted with 4K high-definition cameras and RTK high-precision positioning modules can perform fixed-point orbiting photography.

The positioning error of these modules is controlled within 2 centimeters. This shooting method realizes full panoramic coverage of three types of targets: village overall scenes, ancient building clusters and natural landscapes. In terms of efficiency, the data collection work for a single village can be finished within 2 to 4 hours [8]. Recent systematic reviews have further synthesized advances in UAV platforms, multisensory payloads, AI analytics, and multisource data fusion, evaluating their combined potential for climate-smart agriculture and smallholder systems.

In urban contexts, UAV remote sensing applications emphasize infrastructure safety monitoring and built environment assessment. Findings from global studies reveal that European and American regions extensively use UAV instead of manual inspectors for fast exterior wall detection.

Equipped with visible-light and infrared cameras, UAV can swiftly collect surface images and thermal abnormal information. Such objective datasets serve as reliable evidence for defect recognition, energy audit and damage evaluation [9]. A comprehensive review of 135 peer-reviewed journal articles (2021-2026) identified four key domains in UAV-based facade inspection: Path planning and data acquisition, multi-modal data fusion, deep learning-driven defect detection algorithms, and three-dimensional (3D) reconstruction with digital twin integration.

Multi-modal data fusion - combining visible-light, thermal infrared, and LiDAR sensing - has been shown to improve detection across multiple defect types, yet edge deployment requires balancing lightweight design with recognition stability. Studies have demonstrated that UAV-mounted visible-light and thermal infrared imaging, when integrated with deep learning algorithms and parametric 3D visualization, enable synergistic identification of cracks and moisture intrusion defects in high-density urban residential areas [10]. For instance, a Knet-based model achieved an mIoU of 87.86% for crack detection and 79.05% for leakage detection across representative residential communities.

A multi-level visual analysis framework has been established for UAV-based crack inspection, organizing existing methods into image-level classification, object-level detection, pixel-level segmentation, geometric quantification, and 3D reconstruction.

Comparative analysis across bridges, pavements, dams, building facades, and wind turbine blades shows that scene-specific differences strongly influence data acquisition strategies, model selection, and method performance. Domestic researchers also verified two effective technical paths. UAV multispectral imaging paired with 3D point cloud curvature analysis raises the accuracy rate of hollow drum detection to 87.10%. In addition, UAV-borne infrared thermography is used to check exterior wall hollowing and leakage problems. Compared with traditional manual surveys, this technology boosts inspection efficiency by a maximum of five times.

However, existing research predominantly focuses on single-scenario or single-technology optimization, with limited investigation of cross-scenario technology transfer patterns. As one review notes, current studies often focus on isolated technical components and lack a systematic evaluation of the entire pipeline, leaving the closed loop from defect detection to maintenance unresolved.

Furthermore, while UAV-based infrastructure inspection has been extensively studied across bridges, buildings, dams, and underground spaces, the methodologies and insights generated remain largely siloed within their respective application domains. A critical research gap persists in the absence of standardized, scalable methodologies that can translate plot-level or building-level UAV findings into reliable, cross-domain monitoring systems. This fragmentation limits the potential for technology transfer and cross-fertilization between rural and urban application domains, a gap that the present study seeks to address through its comparative analytical framework.

VR panoramic technology applications in the experience economy

VR technology, characterized by multi-sensory perception, immersion, and interactivity, has demonstrated unique advantages in cultural tourism experiences, product display, and e-commerce consumption. Technology's capacity to transport users to digitally reconstructed environments has fundamentally altered the relationship between information presentation and consumer engagement, particularly in contexts where physical access is constrained by geographic distance, economic cost, or infrastructural limitations.

In the cultural tourism domain, VR has been deployed not merely as a promotional tool but as a medium for heritage preservation, cultural transmission, and experiential marketing. Existing research demonstrates that cultural experiences delivered via VR can trigger emotional engagement and narrative transportation. These effects are comparable to those produced by physical on-site visits. In certain circumstances, VR experiences achieve even better outcomes.

This advantage becomes more prominent when content incorporates contextual storytelling and interactive components. Cao elaborated on the architecture and development process of rural VR panoramic roaming systems, noting that panoramic roaming applied to rural cultural communication offers advantages of low development cost and favorable user experience [11]. Their work demonstrated that lightweight VR solutions, requiring neither specialized hardware nor extensive technical expertise, could be effectively deployed in resource-constrained rural settings, significantly lowering the barrier to digital presence for remote villages.

Further demonstrated, through a Unity3D virtual rural roaming system, that virtual reality-based tourism models overcome geographic and economic constraints, expanding the developmental space of rural tourism [12]. Their system enabled users to navigate village environments, interact with cultural landmarks, and access multimedia information about local history and customs, effectively compressing the distance between urban consumers and rural destinations.

However, existing research also identifies issues including single-function roaming, insufficient model realism, and missing commercial closed loop, limiting the widespread application of highly immersive and interactive experiences in rural tourism [13]. Several studies have noted that many rural VR systems remain at the demonstration stage, functioning as digital showcases rather than integrated platforms that facilitate actual economic transactions or community engagement.

Beyond tourism, VR technology has found substantial applications in agricultural extension, rural education, and product marketing. In the agricultural domain, VR-based training systems have been developed to demonstrate farming techniques, pest management strategies, and sustainable practices to smallholder farmers, overcoming literacy barriers and geographic isolation through visual, experiential learning.

In rural education, immersive VR environments have been deployed to supplement limited physical infrastructure, enabling students in remote areas to access laboratory simulations, historical reconstructions, and cultural experiences that would otherwise be unavailable. Many rural schools lack well-equipped labs, museum visits and field trip opportunities due to geographic barriers and tight budgets. VR effectively bridges these resource gaps by bringing vivid, interactive scenarios directly into classrooms without costly travel or facility construction. These applications highlight the versatility of VR as a general-purpose medium for information dissemination and experiential learning across rural development domains, laying a solid foundation for balanced and high-quality basic education in underdeveloped regions.

In terms of e-commerce integration, researchers systematically compared traditional e-commerce and immersive 3D VR shopping modes. The comparison focused on how these two distinct digital channels effectively convey multi-dimensional product information to prospective online buyers. Empirical findings reveal that VR shopping consistently performs better in two core aspects closely related to consumer perception and product understanding.

First, it achieves more stable emotional expression of commodities. Second, it lowers the difficulty for users to identify product details. Therefore, VR shopping has great potential for large-scale applications. The sensory richness of VR environments - encompassing visual detail, spatial presence, and interactive engagement - enables consumers to form more accurate mental models of products, reducing uncertainty and enhancing purchase confidence.

Studies on consumer behavior have identified three core mediating variables. They are perceived as product authenticity, emotional engagement and narrative transportation. These psychological factors jointly mediate purchase intentions triggered by immersive VR experience by shaping consumers' inner perceptions of displayed goods. Such mediating effects are especially prominent for experience goods, whose value cannot be fully evaluated merely through static pictures and textual descriptions.

Experienced goods refer to products whose quality cannot be fully judged merely by text descriptions or two-dimensional images. Agricultural specialty products, with their strong place-based associations and quality variability, exemplify such experience goods, making them particularly amenable to VR-enhanced marketing.

Nevertheless, existing research on VR e-commerce is mostly limited to consumer behavior experiments conducted in laboratory environments. Few empirical studies explore the integration of VR panoramic presentation and agricultural product e-commerce within rural tourism scenarios.

Two research gaps are particularly notable. First, scholars have not sufficiently examined how panoramic immersive experiences lower consumers' perceived risks. Second, limited attention is paid to how hotspot link design improves purchase conversion rates [14].

The laboratory-to-field gap is notable: While controlled experiments consistently demonstrate VR's advantages in information processing and attitude formation, there is limited evidence on how these effects translate into actual purchasing behavior in real-world rural e-commerce settings.

Furthermore, the interaction between VR immersion and product storytelling - specifically, how the narrative embedded in rural cultural heritage can be leveraged to enhance product valuation and willingness to pay among consumers with varying cultural backgrounds - remains underexplored. Recent studies have called for more situated, practice-oriented research that examines VR e-commerce in authentic consumption contexts, with attention to the specific mechanisms through which immersive experiences bridge the gap between consumer curiosity and purchase commitment amid the unique constraints of rural online commerce.

This gap is particularly consequential for inclusive rural development, where VR's transformative potential to connect distant urban consumers with geographically remote agricultural producers depends critically on the technology's practical ability to generate not just fleeting consumer interest, but tangible and sustained commercial transactions that support the livelihoods of rural operators.

The present study addresses this gap by examining the Yunshang E-village platform, which operationalizes VR panoramic display with embedded purchasing hotspots in real rural tourism contexts, providing empirical evidence on the mechanisms and magnitudes of VRinduced purchase behavior in authentic consumption settings. Detailed information is presented in Table 3.

Table 3. Yunshang E-village platform functional architecture.

Layer	Core function	Technical implementation	Application scenario
Collection layer	UAV aerial photography	DJI mavic 3/matrice 300 RTK, 4K resolution, RTK error ≤2 cm	Village panorama, ancient buildings, landscape nodes, characteristic streets
Presentation layer	VR panoramic browsing	PTGui stitching + photoshop post-processing, 4K output, 360° interaction	Immersive cloud village tour
Presentation layer	Self-service audio guide	Field collection + professional recording, hotspot-triggered playback	Cultural interpretation and navigation
Decision layer	Panoramic hotspot mall	Scene-product mapping, CDN acceleration, <500 ms jump	Scenario-based e-commerce conversion
Decision layer	Content operation push	WeChat public account, “E-tour the Village” series articles	User engagement and brand dissemination

Deep learning object detection algorithms and engineering applications

One-stage object detection algorithms represented by the You Only Look Once (YOLO) series have become mainstream in industrial defect detection due to their end-to-end regression characteristics and superior inference speed. Unlike two-stage detectors such as Faster Region-based Convolutional Neural Network (Faster R-CNN), which separate region proposal from classification, YOLO frames object detection as a single regression problem, directly predicting bounding box coordinates and class probabilities from full images in a single evaluation.

This architectural choice yields substantial speed advantages while maintaining competitive accuracy, making YOLO particularly suitable for real-time and edge-deployment scenarios where computational resources are constrained. YOLOv5l, as a classic large-sized model in this series, adopts mature technical architecture. It employs CSPDarknet53 as the backbone network for efficient feature extraction and Feature Pyramid Network + Path Aggregation Network (FPN + PAN) for multi-scale feature fusion, achieving a balance between detection accuracy ($mAP_{0.5}$ =65.90%) and inference speed (36.765 FPS on standard hardware).

The Cross Stage Partial (CSP) architecture reduces computational redundancy by splitting gradient flows, while the Feature Pyramid Network combined with Path Aggregation Network (FPN-PAN) combination enables the model to detect objects of substantially varying sizes - from large-area spalling to hairline cracks - within the same image.

Recent research has focused on architectural enhancements to further improve detection performance in challenging scenarios. Common attention mechanisms include the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation Networks (SENet). These mechanisms are adopted to recalibrate feature responses. They guide the model to focus on channels and spatial regions related to defects. Meanwhile, the model suppresses interfering with background clutters. Typical distractions involve wall textures, shadows, and structural joints [15].

Loss function optimization, particularly the adoption of Complete Intersection over Union Loss (CIoU), has improved bounding box localization accuracy by incorporating center-point distance and aspect ratio consistency into the loss calculation, beyond the traditional Intersection over Union (IoU) metric. Feature fusion improvements exemplified by Bidirectional Feature Pyramid Network (BiFPN) have introduced learnable weight scaling across different feature levels, enabling the model to dynamically prioritize the most informative semantic and spatial signals.

These innovations collectively address the specific challenges of building defect detection: extreme scale variation (from sub-millimeter cracks to meter-level spalling), high background complexity (tile patterns, mortar joints, weathering stains), and the need for precise localization to support downstream engineering decisions.

In the building inspection domain, it is said that accuracy optimization for high-rise building facade defect detection based on UAV imagery, validating the feasibility of UAV imagery in detecting cracks and hollow drums with an optimized model achieving 87.10% recognition accuracy for hollow drum defects. Han studied UAV vision-based insulator defect detection, demonstrating the applicability of deep learning approaches to infrastructure inspection in challenging outdoor environments. These studies confirm that UAV-acquired high-resolution imagery, when processed through well-trained CNN-based detectors, can reliably identify the presence and location of surface-level defects.

However, the research predominantly remains at the “image recognition” level - the models output bounding boxes with class labels and confidence scores, but these outputs lack direct correspondence to the technical language and safety classifications used in civil engineering practice.

This creates a significant semantic gap between algorithmic outputs and engineering standards: A model may correctly identify a “crack” in the pixel domain, but engineering standards require determining whether it is a “structural crack” exceeding 0.3 mm in width or a “surface crack” requiring only cosmetic attention. Similarly, a detected “hollow drum” must be classified by area (minor under 30 cm^2, moderate 30-100 cm^2, severe above 100 cm^2) to determine maintenance urgency, yet standard object detection outputs provide only a coarse bounding box without precise area quantification or compliance-grade classification. This mismatch can be summarized as follows. Algorithms are capable of drawing defect bounding boxes, yet they fail to generate maintenance conclusions that meet regulatory requirements. This discrepancy constitutes the core obstacle limiting the industrial application of AI-based inspection in construction. In this sector, regulatory compliance and safety certification have strict prerequisites. They rely on standardized terminology and complete traceable documentation.

Some scholars have attempted to address this issue through natural language processing techniques, particularly employing Bidirectional Encoder Representations from Transformers (BERT) models for automatic semantic mapping from AI-generated defect labels to engineering terminology.

The underlying concept is to train transformer-based language models on domain-specific corpora of building codes, inspection standards, and maintenance manuals, enabling the translation of machine vision outputs into human-interpretable, compliance-ready diagnostic statements.

While this approach shows conceptual promise - BERT’s bidirectional attention mechanism can capture contextual relationships between defect descriptors, material specifications, and safety classifications - related research remains at the exploratory stage. Existing studies have demonstrated feasibility on limited, curated datasets but have not yet achieved the accuracy, generalizability, or regulatory acceptance required for systematic engineering deployment.

Several critical challenges still exist. First, large-scale annotated domain datasets with unified engineering terminology are in short supply. Second, the model needs to reliably cope with edge cases and new defect combinations absent from training datasets. Third, liability and certification issues remain unsolved. This concern arises when AI-generated diagnostic conclusions support maintenance decisions closely related to operational safety.

This study fills the gap in engineering integration by developing the Intelligent Wing Building Inspection system. The system implements Bidirectional Encoder Representations from Transformers (BERT)-based semantic mapping. It converts eight types of defects identified by artificial intelligence into 23 engineering terminology categories. The mapping accuracy reaches 98.00% in reference to the Standard for Inspection of Building Structure (GB/T 50344-2019). Besides, the system adopts Sobel edge segmentation algorithms based on Matrix laboratory (MATLAB). This function supports the conversion from pixel area to actual physical area and enables automatic classification of damage grades.

This closed-loop architecture - from UAV image acquisition to defect detection to quantified area estimation to compliance-graded reporting - directly addresses the semantic gap that has limited prior AI-based inspection approaches, offering a pathway from academic algorithms to engineering practice.

Research gaps and theoretical framework

In summary, while current research has achieved notable progress in individual technologies including UAV remote sensing, VR panoramic display, and deep learning object detection, three systematic deficiencies persist.

First, technological dimension fragmentation - existing research predominantly emphasizes vertical optimization of single technologies, lacking systematic design of the "UAV collection → VR presentation → AI decision" full-chain technological architecture. Second, isolated scenario research - rural and urban context studies belong to different academic communities, lacking cross-scenario comparative investigation to distill general patterns and differentiated mechanisms of the technological paradigm across different spatial dimensions. Third, the lack of a value closed loop - technology demonstration predominates over business validation, with AI solutions in both rural e-commerce and building inspection domains particularly lacking effective integration with consumer endpoints or engineering standards. Addressing these research gaps, this paper proposes a "collection-presentation-decision" tripartite technological paradigm analytical framework.

At the collection layer (perception), UAV aerial photography serves as the core technology, equipped with 4K high-definition cameras and RTK high-precision positioning modules, completing standardized spatial data collection.

In rural contexts, emphasis is placed on comprehensive coverage of village panoramas, landscape nodes, and distinctive architecture; in urban contexts, emphasis is placed on precise capture of building facades, structural details, and defect regions. At the presentation layer (interaction), VR panoramic and visualization technologies transform collected spatial data into interactive and experiential digital content. In rural contexts, this manifests as 4K panoramic immersive browsing, self-service audio guides, and hotspot interactive e-commerce. In urban contexts, this manifests as defect heat maps, Building Information Modeling (BIM) overlay, and 3D visualization display.

At the decision layer (transformation), AI intelligent recognition and semantic mapping convert perceptual data and interactive information into executable value outputs. In rural contexts, this manifests as panoramic hotspot-driven e-commerce conversion and cultural tourism decision support. In urban contexts, this manifests as the automated diagnostic closed loop from defect recognition → damage classification → compliance reporting.

The theoretical value of this framework lies in: First, revealing the common logic by which the UAV + VR technology paradigm empowers different scenarios - breaking physical space limitations through the collection layer, establishing cognition and trust through the presentation layer, and achieving value transformation through the decision layer. Second, providing comparable analytical dimensions for cross-scenario technology transfer - technological architecture, empowerment mechanisms, value closed loop, and business models.

Research design and methods

Case selection rationale

This study employs a comparative case study approach following the principle of "theoretical sampling". Two typical cases were selected based on the following criteria.

First, both cases employ UAV as the core data collection tool and VR/AI as the content presentation and intelligent decision-making means, ensuring technological architecture comparability. Second, the two cases represent maximally differentiated application scenarios within urban-rural spaces - rural cultural tourism versus urban building inspection - facilitating revelation of the technological paradigm's differentiated empowerment mechanisms across different spatial scales and service objects. Third, both cases have achieved a complete closed loop from technological validation to practical implementation, with observable and measurable empirical data support.

The rationale for selecting these two cases lies in their representative extremity within the urban-rural continuum. Rural cultural tourism and urban building inspection sit at opposite ends of the spatial governance spectrum: The former deals with expansive, low-density, culturally rich environments where the primary challenge is attracting external attention and converting passive interest into economic activity.

The latter deals with vertical, high-density, safety-critical environments where the primary challenge is systematic coverage, precision detection, and regulatory compliance. This research adopts a maximal variation design on purpose. The aim is to examine the boundary conditions of the UAV-VR-AI technological paradigm. The underlying logic is as follows. If the same core technical components can be effectively adapted to meet such vastly different governance demands, solid evidence can be obtained to verify the generalizability and robustness of this technological paradigm.

Case one, “Yunshang E-village”, represents the “rural → resource activation” scenario, with the core contradiction being resource idleness and value submergence caused by information asymmetry. In this scenario, villages possess cultural heritage, natural landscapes, and agricultural specialty products, yet these assets remain largely invisible to potential urban consumers due to fragmented information channels, lack of professional presentation, and the absence of trusted transaction platforms. The central challenge is thus one of visibility and conversion: How to make rural assets perceptible, desirable, and purchasable to distant audiences.

Case two, “Intelligent wing building inspection”, represents the “urban → safety maintenance” scenario, with the core contradiction being the hazardous inefficiency of manual inspection and the physical limitations of single-technology alternatives. Super high-rise buildings need periodic facade inspections. The fundamental goal is to guarantee structural integrity and public safety. Nevertheless, traditional inspection methods face multiple constraints. They are restricted by human physical limitations and inherent safety hazards. Another drawback lies in subjective judgment. Furthermore, no single sensing technology can capture all categories of defects. The central challenge is thus one of coverage and diagnosis: How to achieve comprehensive, precise, and objective assessment of building health at scale and with speed.

The comparison of these two scenarios facilitates distillation of general patterns and differentiated mechanisms of the technological paradigm across different spatial dimensions. The comparative design keeps the core technical components unchanged. These components include UAV data acquisition, as well as VR/AI presentation and decision-making modules. Meanwhile, the design allows the application scenarios to differ to the greatest extent. This arrangement helps distinguish two types of features within the technological paradigm. Some features are universal and can be applied to both scenarios. Others are context-dependent, which need adjustments to fit specific spatial, social and regulatory constraints.

This analytical structure is essential for developing a nuanced understanding of technology transfer - recognizing that while the same technical building blocks can be redeployed across domains, their configuration, emphasis, and integration must be tailored to the distinctive logic of each application context.

Data sources and research methods

This study comprehensively employs the following research methods. Each selected for its specific capacity to address distinct aspects of the research questions and to enable triangulation across multiple data sources. The mixed-methods design combines qualitative depth - capturing stakeholder perspectives, contextual nuances, and implementation processes - with quantitative breadth - providing statistical generalizability, performance metrics, and comparative benchmarks.

This methodological pluralism is particularly appropriate for comparative case research, where the goal is not only to establish causal relationships but also to understand the mechanisms, conditions, and contextual factors that shape technological outcomes across different settings.

(1) Field investigation

For the rural scenario, the team visited more than 20 typical villages including Jinhua Putang Village, Shi Guangnan’s Former Residence in Dongye Village, Gewuta Village, and Yuandong Township since September 2025. Each field visit typically lasted between one and three days, allowing for systematic observation of village layouts, documentation of existing tourism infrastructure, and preliminary assessment of landscape assets.

The investigation covered multiple dimensions: Historical and cultural resources, including ancestral halls, ancient dwellings, religious sites, and intangible cultural heritage such as folk traditions, festivals, and local craftsmanship. Natural landscape conditions, encompassing topography, water systems, agricultural fields, and scenic viewpoints; infrastructure development, including road access, parking facilities, signage, accommodation, and sanitation.

Agricultural product characteristics, covering types of specialty crops, production scales, seasonal availability, and existing marketing channels and villagers' e-commerce participation willingness, gauged through informal conversations and observed digital literacy levels. For the urban building inspection scenario, the team conducted UAV flight tests and systematic data collection across 12 super high-rise buildings (150-220 m) in Zhejiang and Hubei provinces.

These pilot buildings were selected to represent a range of facade materials (ceramic tile, cement render, glass curtain wall), building ages (10-30 years since construction), and environmental exposure conditions, ensuring that the resulting dataset would capture the variability encountered in real-world inspection practice.

(2) In-depth interviews

The rural scenario adopts semi-structured interviews as the research method. Interviews targeted three groups of stakeholders. Their viewpoints are critical to revealing the complete operation ecosystem of rural tourism and e-commerce. The first group consists of village cadres, with a total of 12 participants. They offered valuable insights concerning policy execution, collective decision-making mechanisms and infrastructure development priorities.

The political economy of village development; local villagers (28 individuals), who shared their lived experiences with tourism impacts, e-commerce participation, and the challenges and opportunities they perceive in digital transformation. Urban tourists (35 individuals), who articulated their information-seeking behaviors, decision-making criteria for destination choice, purchase motivations, and perceptions of VR-based tourism previews. Each interview followed a semi-structured protocol with open-ended questions, allowing respondents to elaborate on issues they considered important while maintaining comparability across interviews.

Interviews typically lasted 30-60 minutes, were audio-recorded with consent, and subsequently transcribed for thematic analysis. In-depth interviews were implemented for the urban building inspection scenario. Three categories of respondents were recruited for this investigation. The first category covers building safety regulatory authorities. There are 6 agencies and 12 participating officials. These stakeholders clarify compliance rules, safety standards and document specifications for inspection systems.

The second category includes engineering inspection enterprises. A total of 10 companies and 28 practitioners took part in interviews. They illustrated practical operational restrictions, cost burdens, safety risks and productivity bottlenecks of existing inspection work. The third category refers to university research institutions. Seven institutions and 16 researchers participated. They shared insights on cutting-edge technologies, gaps between research and industry practice, and demands for future technical functions.

These interviews cumulatively covered 23 organizations and 56 practitioners, providing a comprehensive mapping of stakeholder needs and constraints across the building inspection value chain.

(3) Questionnaire surveys

In the rural scenario, online questionnaires were distributed to urban tourists through the Yunshang E-village WeChat public account, tourism-focused social media groups, and partner travel agencies. The questionnaire was designed to capture demographic information, prior rural tourism experience, familiarity with VR technology, preferences for information channels, and willingness to engage with immersive previews. A total of 500 questionnaires were distributed, with 427 valid returns (effective response rate of 85.40%).

The high response rate reflects effective distribution channels and the relevance of the topic to the target population. Offline questionnaires were administered to rural villagers in partner villages, focusing on their awareness of and attitudes toward e-commerce, digital platforms, and tourism development, as well as their willingness to participate in the platform as product suppliers or service providers.

These offline surveys were conducted face-to-face with assistance from village cadres, ensuring comprehension and high completion rates. In the urban inspection scenario, questionnaire surveys were conducted with 23 property management companies responsible for the maintenance of high-rise residential and commercial buildings. The survey instrument covered current inspection practices, frequency and costs, perceived effectiveness and limitations, openness to automated solutions, and budget constraints. Twenty-one valid returns were collected, representing a 91.00% response rate.

(4) Experimental validation

In the building inspection scenario, a series of controlled experiments were conducted to rigorously validate the performance of the improved YOLOv5l algorithm. Ablation experiments systematically isolated each optimization component - Mosaic-9 data augmentation, CBAM attention mechanism, BiFPN feature fusion, and CIoU Loss function - to measure its individual contribution to overall detection performance.

These experiments were conducted on a standardized test set of 1,000 annotated images (20.00% of the total dataset), with the performance baseline established using the original YOLOv5l architecture and each optimized variant measured against the same test set to ensure comparability. Comparative experiments pitted the fully optimized system against traditional manual inspection and infrared thermography across the same set of test buildings, measuring not only detection accuracy but also time efficiency, cost, and coverage completeness. These side-by-side comparisons provide the empirical basis for the improvement of metrics reported in Table 2.

(5) Case analysis

In-depth case analysis was conducted on four completed rural projects (Putang Village, Shi Guangnan’s Former Residence, Gewuta Village, Yuandong Township) and 12 super high-rise building inspection projects.

Each case was documented according to a standardized protocol that captures the full project lifecycle: initial conditions (baseline data on visitor numbers, sales, or inspection metrics), intervention characteristics (specific technologies deployed and customization decisions), implementation processes (duration, challenges encountered, and adaptations made), and measured outcomes (quantitative performance data and qualitative stakeholder feedback). The analysis employed a within-case and cross-case analytical strategy: Within-case analysis established the causal narrative and outcome patterns for each project individually; cross-case analysis then compared projects to identify common patterns, outlier cases, and explanatory factors for variation in outcomes.

For the rural projects, the analytical focus was on visitor growth, sales conversion, and platform engagement; for the building inspection projects, the focus was on detection accuracy, time efficiency, cost reduction, and coverage improvement. Quantitative efficacy assessments were derived from operational data collected during project implementation (visitor logs, transaction records, inspection reports), supplemented by follow-up interviews with project stakeholders to interpret performance patterns and contextualize quantitative findings.

This methodological suite - combining field observation, stakeholder interviews, surveys, experiments, and case analysis - ensures that the empirical findings reported in this study are grounded in direct observation, triangulated across multiple data sources, and systematically analyzed using both qualitative and quantitative techniques. The convergence of evidence across these methods strengthens confidence in the validity of the conclusions and provides a rich empirical basis for the comparative analysis that follows.

Technological implementation methods

Rural scenario (Yunshang E-village) technological pathway: UAV aerial surveying is conducted using Da-Jiang Innovations (DJI) Mavic 3 and Matrice 300 RTK platforms following a standardized "eight-direction horizontal shooting plus gimbal downward-tilting supplementary photography" protocol. The eight horizontal angles capture the village from equidistant positions around the perimeter, providing overlapping coverage for seamless stitching, while gimbal-downward shots capture ground-level details and vertical surfaces. Each flight mission typically requires 2-4 hours per village and generates 200-500 raw images.

Panoramic production employs Panorama Tools Graphical User Interface for Panorama Tools (PTGui) and Autopano Giga (APG) for image stitching and fusion, utilizing control-point matching and bundle adjustment optimization to correct for lens distortion, parallax errors, and exposure variations. The fused panorama is exported to Photoshop for sky replacement, color correction, and detail optimization. The final output is 4K ultra-high-resolution VR panoramas (3840×1920 pixels), enabling users to examine architectural features without pixelation.

Platform development comprises the Yunshang E-Village official website and WeChat official account. The website is built on Linux + Apache + MySQL + PHP (LAMP) architecture, supporting VR panoramic browsing, self-service audio guides, panoramic hotspot e-commerce, and content push notifications. The WeChat account provides a mobile-optimized interface, leveraging its integrated payment infrastructure to reduce transaction friction.

Urban Inspection Scenario (Intelligent Wing Building Inspection) Technological Pathway. UAV path planning adopts a "vertical layering +10.00% overlap ratio" adaptive facade flight design. The facade is divided into horizontal layers; the UAV flies along each layer capturing overlapping images with 10.00% overlapping between consecutive images. This approach maintains consistent distance from the facade and avoids rotational instability. Path planning is combined with Real-Time Kinematic (RTK) high-precision positioning (error ≤2 cm) and Simultaneous Localization and Mapping (SLAM) visual navigation for reliable operation in GPS-denied environments.

Dataset construction integrates 3,982 high-resolution (4K) images covering three material types (tile, cement, glass curtain wall) and three environmental conditions (sunny, cloudy, light wind). Pixel-level annotation of eight defect categories (cracks, hollow drums, spalling, rust, tile cracks, tile spalling, tile detachment, exposed rebar) is performed using Labeling with a double-pass verification process ensuring annotation accuracy exceeding 98.00%.

Algorithm optimization uses YOLOv5l as the baseline, incorporating CBAM attention mechanism, BiFPN bidirectional feature fusion, CIoU Loss, and Mosaic-9 data augmentation to address defect scale variation and background complexity. Following optimization, the model achieves mAP0.5 of 76.30% on the test set, a 10.4 percentage point improvement over baseline. Edge deployment applies 8-bit Integer quantization (INT8 quantization) and channel pruning, compressing model weights from 32-bit to 8-bit representation and reducing memory footprint by 75.00%. The deployed model achieves 32 fps inference speed.

Damage classification employs MATLAB Sobel edge segmentation for defect area quantification (error ≤5.00%). The system extracts precise defect contours and converts pixel counts to real-world areas using RTK-derived calibration. Defects are automatically classified into three grades (minor <30 cm², moderate 30-100 cm², severe >100 cm²) according to JGJ/T 110-2021 standards, each corresponding to specific maintenance recommendations. This automated classification eliminates subjective variability inherent in manual grading.

Empirical analysis of dual scenarios

Scenario one: Rural resource activation - "E-village" immersive rural e-commerce empowerment platform

(1) Problem diagnosis and needs assessment

Survey findings indicate that 72.60% of urban residents abandon planned rural tourism due to incomplete destination information or a single information acquisition channel. Only 18.30% of respondents reported being able to obtain relatively complete village cultural and tourism resource information through online channels. Investigation of 80 villages in northern Zhejiang revealed that 54.00% have demand for developing local tourism through cultural resource exploitation, yet only 12.00% possess basic digital display capabilities.

Putang Village in Jinhua serves as a case in point. It owns a wealth of ancient buildings and deep historical heritage embodied in the "Wenchang Wuju" culture, meaning civil literary attainment and martial spirit. On Baidu Baike, merely one entry targets commercial tourism. Other materials are limited to news reports and personal blogs. The absence of commercial packaging and marketing support creates obstacles. As a result, the village's native culture cannot be translated into tourism attractiveness.

(2) Technological architecture and platform functions

The "Yunshang E-village" platform adopts a "one core, two wings, three terminals" overall architecture. The core is cultural mining and IP shaping of rural culture. The two wings are the VR panoramic display system and the rural e-commerce trading system. The three terminals are the WeChat public account (mobile), official website (PC), and offline VR experience terminals.

VR panoramic browsing integrates 360° high-definition VR panoramic content (4K resolution) from partner villages, supporting free viewpoint switching and detail magnification, accompanied by rural background music and environmental sound effects.

Self-service audio guides provide standardized self-service audio commentary content based on field visits and historical materials. Panoramic hotspot link e-commerce places interactive hotspots within VR panoramic scenes for agricultural products, homestays, dining, and tourism packages, with clicks directly redirecting to corresponding product pages in the "Yunshang Village" mini-program mall. Mini-program mall direct access centralizes listing of agricultural specialty products, cultural and creative souvenirs, and specialty handicraft products after professional packaging design and cultural value enhancement. Content push notifications regularly release serialized original articles including "E-tour the Village", "Festival Group Purchase" and "Imprint the Village".

(3) Implementation efficacy assessment

As of May 2026, the project had completed landscape design and VR panoramic production for four signature projects - Shi Guangnan's Former Residence, Putang Village, Gewuta Village, and Yuandong Township - with total transaction volume exceeding 1.8 million RMB. Detailed efficacy data are presented in Table 4.

Table 4. Efficacy assessment of Yunshang E-village implemented projects.

Village name	Cultural theme	VR panorama launch	Visitor increase (%)	Monthly e-commerce sales (RMB)	Pre-project monthly sales	Growth rate (%)
Jinhua putang village	Wenchang wuju	Nov 2025	+65.00	87,000	20,700 RMB	+320.00
Shi Guangnan's former residence	Red celebrity culture	Dec 2025	+120.00	46,000	11,000 RMB	+318.00
Gewuta village	She ethnic folk culture	Feb 2026	+85.00	32,000	9,000 RMB	+256.00
Yuandong township	White peach industry culture	Apr 2026	+45.00	123,000	28,000 RMB	+339.00

Total VR panoramic page views reached 68,000, with scenario-based purchasing average order value 42.00% higher than traditional e-commerce. Putang Village's "Wenchang Wuju" - themed cultural landscape was reported by multiple media outlets including Xinhua Net, becoming a regional rural cultural tourism demonstration benchmark. The Shi Guangnan's Former Residence project received the "Excellent Red Cultural Tourism Innovation Project" commendation from the Jinhua Municipal Culture and Tourism Bureau.

Scenario two: Urban safety maintenance - "Intelligent Wing Building Inspection" super high-rise building facade intelligent detection system

(1) Problem diagnosis and needs assessment

As of August 2025, China had over 6,000 super high-rise buildings exceeding 100 meters, with total high-rise buildings reaching 347,000. However, traditional manual "spiderman" high-altitude inspection achieves single-building coverage below 30.00%, micro-crack recognition rate below 65.00% for defects under 0.5 mm, subjective error in manual tapping hollow drum judgment exceeding 30.00%, and a safety accident rate of 1.20%.

Single-technology alternatives such as infrared thermography are constrained by equipment bulk (exceeding 500,000 RMB per unit), strong light dependency, and blindness to cracks below 0.2 mm. Industry surveys indicate that fewer than 30.00% of super high-rise buildings nationwide achieve "annual full coverage inspection", with more than 50 reported exterior wall detachment injury incidents across national media from 2023 to 2025.

(2) Technological architecture and core algorithms

The system adopts a three-layer "data layer - algorithm layer - application layer" architecture. The data layer integrates 3,982 high-resolution (4K) images covering three material types (tile, cement, glass curtain wall) and three environmental conditions (sunny, cloudy, light wind), with pixel-level annotation of eight defect categories (cracks, hollow drums, spalling, rust, tile cracks, tile spalling, tile detachment, exposed rebar) at annotation accuracy ≥98.00%.

The algorithm layer uses YOLOv5l as the baseline ($mAP_{0.5}$ =65.90%, FPS =36.765), with Mosaic-9 data augmentation and adaptive anchor calculation at the end input. CBAM dual attention mechanism embedded in the backbone network; BiFPN bidirectional feature fusion introduced in the neck; and CIoU Loss with weighted Non-Maximum Suppression (NMS) at the output end. Following optimization, $mAP_{0.5}$ reached 76.30% (a 10.4 percentage point improvement), with average recognition accuracy of 95.80% across eight defect categories, 98.00% for hollow drums (area >100 cm²), and single-image inference speed of 32 FPS.

Furthermore, BERT semantic mapping enables automatic conversion between eight AI label types and 23 engineering terminology categories, with a 98.00% accuracy rate.

MATLAB Sobel edge segmentation supports quantitative calculation of defect areas, and the error is no more than 5.00%. Following JGJ/T 110-2021, damage conditions are divided into three grades. Minor damage: below 30 cm²; moderate damage: 30-100 cm²; severe damage: above 100 cm².

The application layer integrates four core modules - flight control, defect detection, 3D positioning (RTK error ≤2 cm), and report generation - automatically outputting standardized reports compliant with the Standard for Inspection of Building Structure (GB/T 50344-2019).

(3) Pilot application efficacy assessment

The system completed pilot applications across 12 super high-rise buildings (150-220 meters) in Zhejiang and Hubei provinces, as summarized in Table 5.

Table 5. Efficacy data of intelligent wing building inspection pilot applications.

Metric	Traditional manual inspection	This system	Improvement
Single 150 m building inspection time	3 days (72 hours)	≤3 hours	96.50% efficiency improvement
Single 200 m building cost	68,000 RMB	12,000 RMB	82.40% cost reduction
Defect recognition accuracy	<65.00% (micro-cracks)	≥95.00% (8 categories)	+30 percentage points
Miss rate (%)	15.00-20.00	<2.00	About 90.00 reduction
Inspection coverage	<30.00%	100.00%	Blind spot elimination
Report generation time	4 hours	40 minutes	83.30% efficiency improvement

The system identified 326 hidden defects including micro-cracks and hollow drums during on-site building inspections, successfully preventing two potential high-altitude falling accidents triggered by structural deterioration. The system has officially received "Smart Inspection Recommended Technology" certification from the Zhejiang Provincial Department of Housing and Urban-rural Development and has been included in the Zhejiang Provincial Building Safety Supervision Informatization Construction Guide case library to promote widespread engineering application across the province.

Cross-scenario comparison and discussion

Common patterns: General logic of the tripartite technological paradigm

Empirical analysis of the two cases demonstrates that the "UAV + VR + AI" tripartite technological paradigm follows a common empowerment logic across different urban and rural contexts.

The commonality at the collection layer is "breaking physical boundaries". Whether the vast rural fields and dispersed villages, or the urban super high-rise building facades and irregular structures, UAV technology enables low-cost, high-efficiency spatial data collection that is unattainable by manual means.

"Yunshang E-village" transforms rural landscapes previously "hidden deep in the mountains unknown to the world" into transmissible digital assets; "Intelligent Wing Building Inspection" completely eliminates the safety risks of traditional "spiderman" high-altitude operations. RTK high-precision positioning (error ≤2 cm) and 4K high-definition imaging constitute the technological foundation in both scenarios.

The commonality at the presentation layer is "establishing cognitive trust". VR panoramic technology, with its intuitiveness, authenticity, and immersion, bridges the cognitive gap between information supply and demand. Within rural application scenarios, VR preview serves urban consumers. It delivers comprehensive information about village landscapes and cultural traits. Thus, it overcomes the travel barrier caused by inadequate information.

For urban building inspection, BIM heat maps and 3D visualization are deployed. They help property managers intuitively find defect locations and judge damage severity. Traditional inspection often produces unintelligible reports with unclear defect coordinates, and this system effectively resolves the problem.

The commonality at the decision layer is "achieving value transformation". Scenario-based e-commerce hotspots transform rural "viewing" into "purchasing"; damage classification and compliance reports transform urban "recognition" into "diagnosis". Both scenarios achieve a complete closed loop from data collection to value output, rather than remaining at the technology demonstration level.

Differentiated mechanisms: Spatial scale and service logic comparison

Despite following common logic, the two scenarios exhibit significant differences in operational mechanisms. Distinctions can be observed on a spatial scale and data acquisition demands. The rural case adopts area-oriented data collection. It covers extensive ranges such as village panoramic views, street networks and rural scenery. The aerial survey scope for one village is normally 0.5-5 km². The urban inspection case integrates line and area collection modes. High-resolution facade details need to be captured per square meter. The sampling area for one building facade is about 5,000-10,000 m². Moreover, ultra-high precision is required, enabling crack recognition at the 0.1 mm scale.

The two scenarios differ in service objects and value logic. The rural scenario faces consumer-side users. End users are urban tourists. VR immersive experiences reduce information costs for consumption decisions. Scenario-based e-commerce further facilitates sales of agricultural products. In essence, it combines the attention economy and experience economy. The urban inspection scenario focuses on management-side users. Core users include property management institutions and regulatory authorities. Automated diagnosis improves building safety maintenance efficiency and mitigates accident hazards. Its development relies on the safety economy and compliance economy.

Distinctions exist between the two scenarios in business modes and revenue pathways. The rural case applies the "platform + village" collaborative framework. Revenue is generated from landscape design services (50,000-150,000 RMB for each village), VR panoramic production, and mall sales commission ranging from 8.00% to 15.00%. The urban inspection case provides an all-in-one "hardware + software + service" solution. Income streams cover equipment sales (the full set ≤200,000 RMB), per-building inspection fees (12,000 RMB per building), and yearly maintenance service fees.

Theoretical contributions and practical implications

The theoretical contributions of this research can be summarized below. For the first time, the "UAV + VR + AI" technological paradigm is adopted for urban-rural comparative analysis. The study establishes the "collection-presentation-decision" tripartite analytical framework. It uncovers both the universal logic and differentiated mechanisms of the paradigm. The paradigm functions in two contexts: resource activation in rural areas and safety maintenance in cities. This work fills the existing theoretical vacancy in research on cross-scenario technology transfer.

Additionally, this study provides empirical validation of the effectiveness of VR scenario-based consumption and AI engineering semantic mapping, contributing new evidence to theoretical development in rural e-commerce and intelligent inspection domains.

The practical implications are as follows. First, urban and rural digital transformation can share the same technological foundation - UAV platforms, VR engines, and AI algorithms - providing possibilities for technology suppliers to explore cross-scenario market expansion. Second, technology transfer requires attention to scenario-specific value logic and user needs, avoiding one-size-fits-all solution replication. Third, standardized data collection processes and modular system architectures are key to ensuring replicability and scalability of technological solutions.

Limitations and future directions

This study has several limitations. First, the geographic scope of both cases is primarily concentrated in central Zhejiang Province; the applicability of findings to broader regions requires further validation. Second, the operational period of the rural scenario is relatively short (approximately 12 months), and long-term sustainability data for the urban inspection scenario remain to be accumulated. Third, the current research emphasizes technological validation and efficacy assessment within individual scenarios, with limited exploration of technological synergy and data interoperability between the two scenarios.

Several directions are proposed for follow-up research. First, expand case coverage to the Yangtze River Delta and nationwide. It enables verification of the cross-regional replicability of the technological paradigm. Second, explore interoperability between rural VR panoramic datasets and urban inspection BIM data. The goal is to construct a unified digital twin foundation for urban and rural areas. Third, incorporate 5G real-time transmission and edge cloud computing. These technologies enhance real-time interaction and instant decision support in both scenarios. Fourth, develop long-term tracking and assessment frameworks. They can quantify comprehensive contributions of technical solutions, such as increasing rural residents' income and safeguarding urban public security.

Conclusion

This paper constructs a tripartite "UAV + VR + AI" technological paradigm, namely the collection-presentation-decision framework. Empirical verification is conducted via a comparative case study. Two typical application systems are selected as research objects. One is the "Yunshang E-village", an immersive platform for rural e-commerce empowerment. The other is "Intelligent Wing Building Inspection", which realizes intelligent detection for super high-rise building facades. This research investigates two highly distinct urban and rural application contexts. It systematically verifies three core attributes of this integrated technological framework. They are cross-scenario applicability, differentiated empowerment mechanisms and practical value. Four major conclusions can be drawn from this study. First, the integration of UAV remote sensing, VR panoramic display, and AI recognition constitutes a general-purpose technological paradigm with robust cross-scenario applicability. The "collection-presentation-decision" tripartite architecture provides a coherent analytical lens for understanding how multi-technology integration empowers spatially distributed governance challenges.

At the collection layer, UAV platforms equipped with RTK high-precision positioning and 4K imaging sensors overcome the physical access constraints that have historically limited both rural resource documentation and urban infrastructure inspection. At the presentation layer, VR and visualization technologies transform raw spatial data into cognitively accessible, experientially rich content that bridges the gap between information supply and user comprehension. At the decision layer, AI-driven recognition and semantic mapping convert perceptual inputs into actionable value - whether as e-commerce transactions in the rural context or as compliance-graded diagnostic reports in the urban context. This tripartite logic transcends individual application scenarios, offering a reusable architectural template for digital empowerment across diverse spatial governance domains.

Second, in the rural context, VR panoramic technology effectively reduces information asymmetry and unlocks the economic potential of cultural tourism and agricultural specialty markets in remote villages. The empirical findings from the Yunshang E-village platform provide compelling evidence for this proposition.

Four partner sites are investigated: Putang Village, Shi Guangnan's Former Residence, Gewuta Village and Yuandong Township. VR immersive displays helped boost visitor numbers. The growth of tourist arrivals varied from 45.00% to 120.00%. In addition, after the platform was launched, monthly sales of characteristic agricultural goods rose sharply. The sales growth reached 256.00%-339.00%.

These results confirm that the "immersive experience → immediate purchase" closed loop is not merely a theoretical construct but a commercially viable mechanism for rural resource monetization. By enabling urban consumers to visually and experientially engage with village environments prior to travel, VR technology lowers the perceived risk of destination choice, enhances product authenticity perception, and converts passive browsing into active purchasing behavior. This finding has significant implications for rural revitalization policy, suggesting that digital display infrastructure - rather than physical transportation infrastructure alone - can serve as a catalytic investment for poverty alleviation and regional development.

Third, in the urban building inspection context, the improved YOLOv5l algorithm integrated with UAV adaptive path planning and edge computing architecture elevates the efficiency, accuracy, and safety of manual inspection to fundamentally new levels. The Intelligent Wing Building Inspection system achieved 95.80% defect recognition accuracy across eight defect categories, compressed single 150 m building inspection from 3 days to under 3 hours, reduced inspection costs by 82.40%, and lowered the miss rate to below 2.00%.

These performance metrics represent a paradigm shift from experience-driven, hazardous manual operations - with their 1.20% accident rate, 30.00% subjective error margin, and 70.00% coverage blind spots - to data-driven, automated intelligent diagnosis. The system's capacity to detect cracks as fine as 0.1 mm and to generate compliance-ready diagnostic reports through BERT-based semantic mapping addresses the long-standing industry pain point of "algorithmic detection without engineering interpretation". This technological leap not only enhances building safety maintenance efficiency but also fundamentally redefines the occupational safety profile of the inspection industry, eliminating the need for high-altitude "spiderman" operations.

Fourth, cross-scenario comparison reveals systematically differentiated operational mechanisms across spatial scale, service objects, value logic, and business models. The rural scenario follows an "attention-experience-conversion" value chain characterized by area-oriented spatial coverage and consumer-end services, where the primary value driver is reducing information costs to facilitate market transactions.

The urban inspection scenario follows a "perception-diagnosis-early warning" safety logic characterized by fine-scale "line-area" collection and management-end services, where the primary value driver is reducing safety risks and compliance costs through automated, quantified assessment. These differences are not incidental but are structurally determined by the distinct spatial geometries, stakeholder configurations, and regulatory environments of the two contexts.

It is vital to recognize these differentiated mechanisms to achieve successful technology transfer. Identical core technical components may be reused in different scenarios. However, researchers must carefully reconfigure multiple elements. These include data collection specifications, presentation modes, and decision-support interfaces. Such adjustments ensure the technology matches scenario-specific user demands and value propositions.

Theoretical contributions and practical implications. This study makes several contributions to the literature on digital transformation, urban-rural governance, and technology transfer. Theoretically, it advances beyond single-technology, single-scenario studies to articulate a general-purpose tripartite paradigm that can be instantiated across diverse application domains. The "collection-presentation-decision" framework provides an analytical bridge between the engineering literature on UAV sensing, the human-computer interaction literature on VR experience design, and the policy literature on digital governance. Practically, the findings suggest that urban and rural digital transformation initiatives - often treated as separate policy domains - can benefit from a shared technological infrastructure.

For technology suppliers, the cross-scenario applicability of the UAV-VR-AI paradigm offers pathways for market diversification beyond single-sector specialization. For policymakers, the demonstrated efficacy of VR-based information provision in rural contexts and AI-based automated inspection in urban contexts suggests that continued investment in these enabling technologies merits priority within national digital development agendas.

Limitations and future research. This study has several limitations that point to promising directions for future investigation. First, the geographic scope of both cases is concentrated in central Zhejiang Province; broader regional validation - particularly in western China, where rural conditions differ significantly - is needed to establish external validity. Second, the rural scenario has operated for approximately 12 months, precluding assessment of long-term sustainability and path-dependent effects; longitudinal studies are required to track whether the observed economic benefits persist, accelerate, or diminish over time. Third, the current research treats the two scenarios as separate cases, with limited exploration of potential synergies - for instance, whether UAV-captured rural landscape data could feed into urban planning models, or whether urban BIM data could inform rural infrastructure development. Future research should investigate these interoperability mechanisms as a foundation for integrated urban-rural digital twin platforms.

As 5G network deployment deepens, UAV and VR equipment costs continue to decline, and AI algorithms undergo rapid iteration, the “UAV + VR + AI” technological paradigm will play an increasingly critical and expansive role in urban-rural governance. This study provides both a systematic theoretical framework and robust empirical evidence for understanding and guiding this transformation. We hope the findings and analytical framework offered here will serve as a useful reference for subsequent researchers and practitioners working at the intersection of digital technology, spatial governance, and sustainable development.

Funding

This work was not supported by any funds.

Acknowledgments

The authors would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Li, F., Gan, Y. (2025) Research on the sustainable development capability of Chinese rural e-commerce based on multidimensional perspective. *Scientific Reports*, 15(1), 11547.

[2] Zhang, H., Millan, E., Money, K., Guo, P. (2024) E-commerce development, poverty reduction and income growth in rural China. *Journal of Strategy and Management*, 18(1), 148-176.

[3] Javaid, M., Khan, I. H., Singh, R. P., Rab, S., Suman, R. (2022) Exploring contributions of drones towards Industry 4.0. *Industrial Robot: The International Journal of Robotics Research and Application*, 49(3), 476-490.

[4] Jia, Y., Wang, Y., Tang, L., Wang, C. (2024) Review of identification methods for early-stage cracking in solids. *Journal of Testing and Evaluation*, 52(5), 3123-3148.

[5] Ai, W., Zou, J., Liu, Z., Wang, S., Teng, S. (2025) Light propagation and multi-scale enhanced DeepLabV3+ for underwater crack detection. *Algorithms*, 18(8), 462.

[6] Afanaseva, O. V., Tulyakov, T. F., Shaimardanov, A. A. (2026) Deep learning-based visual analytics for efficiency and safety optimization in power infrastructure. *Engineering*, 7(3), 135.

[7] Roy Choudhury, M., Das, S., Christopher, J., Apan, A., Chapman, S., Menzies, N. W., Dang, Y. P. (2021) Improving biomass and grain yield prediction of wheat genotypes on sodic soil using integrated high-resolution multispectral, hyperspectral, 3D point cloud, and machine learning techniques. *Remote Sensing*, 13(17), 3482.

[8] Li, Z., Wang, T., Sun, S. (2024) Research on quantitative analysis methods for the spatial characteristics of traditional villages based on three-dimensional point cloud data: a case study of Liukeng Village, Jiangxi, China. *Land*, 13(8), 1261.

[9] Kim, H., Lamichhane, N., Kim, C., Shrestha, R. (2023) Innovations in building diagnostics and condition monitoring: a comprehensive review of infrared thermography applications. *Buildings*, 13(11), 2829.

[10] Fan, Y., Jinghua, M., Fei, X., Yu, L. S. S., Jiang, S., Yiqi, T., Tsang, W. C. (2025) UAV and deep learning for automated detection and visualization of façade defects in existing residential buildings. *Sensors*, 25(23), 7118.

[11] Cao, K. (2022) Development and design case function comparison of panoramic roaming system of virtual museum based on Pano2VR. *Mobile Information Systems*, 2022(1), 7363221.

[12] Jude, O. C., Ukekwe, C. (2020) Tourism and virtual reality (VR) in developing nations. *African Journal of Hospitality, Tourism and Leisure*, 9(2), 1-16.

[13] Han, B., Yu, J., Bi, M. (2025) A rural tourism promotion mechanism based on virtual reality technology for real-time interactive experience.

Systems and Soft Computing, 7, 200299.
[14] Israel, K., Zerres, C., Tscheulin, D. K. (2019) Presenting hotels in virtual reality: Does it influence the booking intention? *Journal of Hospitality and Tourism Technology*, 10(3), 443-463.
[15] Maruthi, M., Devi, M. S., Choi, Y., Yi, C. Y. (2026) Damage attention-aware dense layered framework for surface crack classification. *Buildings*, 16(12), 2313.