Residual Intelligence: Deep Residual Learning for Image Recognition in Construction – Applications and Future Directions

Adetayo Onososen[1] and Innocent Musonda [2]
[1, 2]Centre for Applied Research and Innovation in the Built Environment (CARINBE)
Faculty of Engineering and the Built Environment (FEBE)
University of Johannesburg, Johannesburg, South Africa
adetayoo@uj.ac.za

Abstract

Rapid growth in machine learning advancement has positioned computer vision as a linchpin of data-driven construction practice, yet the sector's reliance on deep residual networks remains poorly charted. To clarify the state of the art, we conducted a PRISMA guided bibliometric review of Scopus records published between 2016 and March 2025. A tightly scoped query linking residual-learning terms with image-recognition tasks and built-environment contexts returned 292 items; after language and document-type screening, 258 publications were analysed. Co-citation, keyword and country overlays were generated in VOSviewer and interpreted against domain knowledge. Results reveal a six-fold rise in annual outputs since 2019, with China, the United Kingdom and India accounting for half of all papers. Intellectual structure converges on three mature pillars (semantic segmentation, object detection and image classification) and two emergent strands: resource-aware/federated learning and sustainability-oriented applications. Transformer enhanced detectors, edge-optimised residual hybrids and diffusion-based data augmentation are identified as recent inflexion points. Conversely, limited cross-domain transfer studies, scarce explainability protocols and an absence of carbon footprint assessments expose critical research gaps. The review thus furnishes a data-driven roadmap for shifting residual intelligence from isolated prototypes to trustworthy, resource-efficient and globally transferable vision systems capable of underpinning autonomous, safe and sustainable construction operations.

Keywords

Deep residual learning; Image recognition; Construction engineering; Bibliometric analysis; Transformer detectors

1. Introduction

The construction industry now generates visual data at a volume and velocity unprecedented in its history (Tang and Golparvar-Fard, 2021); (Tonetto et al., 2023). High-resolution drone imagery, laser-scanned point clouds and ubiquitous site cameras together create a continuous digital chronicle of project activities, from earthworks to final hand-over (Onososen and Musonda, 2022); (Sun, 2024). Unlocking the latent value of these unstructured pixels is essential for reducing rework, enhancing safety compliance and shortening project schedules outcomes that speak directly to the sector's long-standing productivity gap and its ambition to meet net-zero and zero-harm targets (Shahadat and Maida, 2025). Early machine vision pipelines, however, struggled with the complexity of real-world construction scenes: irregular geometries, occlusions, dust and highly variable illumination routinely confounded handcrafted feature extractors (Deng et al., 2024) (Muhammed Sufail et al., 2024). Deep convolutional neural networks (CNNs) promised a remedy by learning hierarchical representations directly from data, yet their effectiveness depended on network depth and depth brought its limitations, notably vanishing gradients and accuracy degradation (Lukovic et al., 2024); (Jadhav, 2025). The advent of residual learning removed this ceiling by introducing identity skip connections that stabilised gradient flow, enabling architectures hundreds of layers deep without loss of fidelity (He et al., 2016); (Fahad et al., 2024). In mainstream computer vision, residual networks and their variants quickly superseded earlier backbones across classification, detection and segmentation tasks; their adoption in construction engineering, however, has been uneven, dispersed across disparate application niches. The present study sets out to map and critically appraise the scholarly landscape of deep-residual-learning applications in construction image recognition. By fusing bibliometric techniques with domain-specific interpretation, the study delivers three original contributions. First, it provides the first longitudinal evidence of how residual learning has migrated from generic computer-vision laboratories into construction-specific workflows, quantifying both scale and pace. Second, it reveals the thematic topology of the field, exposing hitherto under-recognised clusters such as circular-economy imaging and federated on-edge learning that are poised to shape the next research wave. Third, it proposes a forward-looking research agenda rooted in the twin imperatives of trustworthy AI and resource-constrained deployment, thereby furnishing scholars and practitioners with a roadmap for translating residual intelligence into tangible project-level value.

2. Methods

The study adopted a bibliometric review design, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, which has proved effective for mapping research frontiers in construction technology studies (Onososen et al., 2022); (Agapiou, 2024). Scopus was chosen as the bibliographic source because of its broad coverage of engineering, computer science and built-environment journals and its transparent export features that facilitate subsequent network analysis (Gupta et al., 2025). All records were retrieved on 19 March 2025. The search string combined three conceptual blocks—(i) deep residual learning terminology, (ii) image-recognition tasks and (iii) construction-related contexts using Boolean operators and wildcards to minimise omission error. Limiting the search to titles, abstracts and author-supplied keywords. The query combined three conceptual blocks linked with Boolean AND operators. The first block captured the family of residual architectures by searching for the phrases “deep residual learning”, “residual network”, “residual block” and the most common model names (ResNet, ResNet-50, ResNet-101, ResNet-152, ResNeXt). The second block enumerated the principal computer-vision tasks of interest—namely image recognition, image (or object) classification, object detection, semantic segmentation and instance segmentation, together with the broader rubric computer vision. The third block anchored the retrieval to the built-environment domain by specifying construction, built environment, civil engineering, structural engineering, construction site, infrastructure project, building inspection, digital construction, and the twin concepts of building information modelling. The search yielded 292 distinct records spanning 2016–2025. Records were first filtered by language; only English-language publications were retained because they offer the critical mass of citations required for network robustness, leaving 258 documents (Pedral Sampaio et al., 2023). Conference abstracts with fewer than three pages, editorial notes and errata were then excluded to ensure analytic consistency, while full conference proceedings and journal articles were kept since short-format items seldom contain sufficient methodological detail for citation analysis. Full bibliographic information—including authors, affiliations, keywords, abstract, cited references and publication source was exported in CSV format. The dataset was examined in VOSviewer 1.6.20 to generate keyword co-occurrence and citation-coupling maps (Aziminezhad and Taherkhani, 2023). Fractional counting was applied; the minimum threshold for keyword inclusion was set to five occurrences to suppress noise from idiosyncratic phrases. Network visualisations were interpreted iteratively by the authors, drawing on domain knowledge in computer vision and construction engineering to label clusters and derive thematic trajectories.

3. Results

3.1 Publication Trend

Figure 1 depicts the annual volume of publications on deep-residual-learning–based image recognition in construction between 2017 and 2025. A clear, multi-phase trajectory emerges. The field began in 2017–2018 with only a handful of papers, reflecting the lag between (He et al., 2016) deep residual learning for image recognition and its uptake by construction researchers, who generally face longer data-collection and validation cycles than the mainstream computer-vision community. Growth accelerates from 2019 through 2021, rising from roughly ten to almost forty papers. Two factors appear to underpin this expansion: (i) the maturation of low-cost UAV and site-camera ecosystems, which eased large-scale data acquisition for crack, defect and safety monitoring applications; and (ii) the deployment of open-source implementations of ResNet and its variants in widely used deep-learning frameworks, lowering the technical barrier for engineering scholars. The slight contraction observed in 2022 (from about 37 to 34 records) is consistent with broader bibliometric evidence of pandemic-related research-pipeline disruptions, including restricted site access and delayed field trials. This steep ascent coincides with three technological inflexions: the diffusion of transformer–CNN hybrids that improved small-object detection, the emergence of edge-AI accelerators capable of running residual networks in real-time; and the first commercial digital-twin platforms, which created new demand for automated vision-based progress tracking.

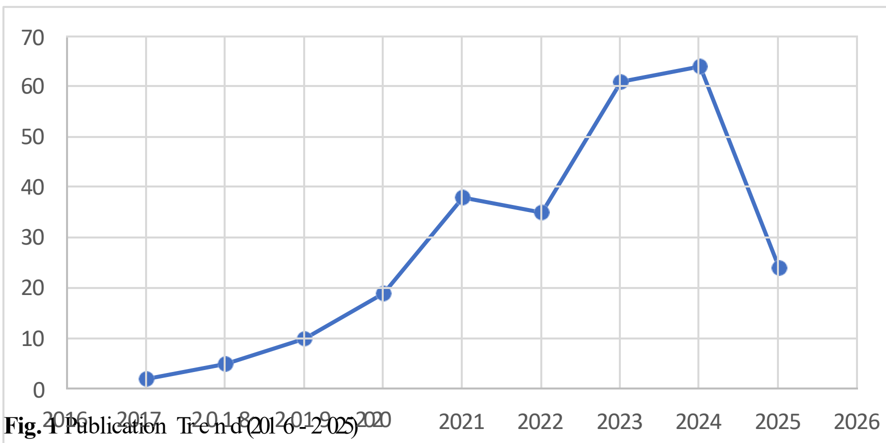

Fig. 1 Publication Trend (2016 - 2025)

3.2 Citation Analysis by Country

Figure 2 presents a country-level overlay map, generated in VOSviewer, where node size reflects the number of publications on residual-learning–based vision in construction and colour denotes the average year of those publications. China occupies a markedly dominant position: its large, light-green node shows that it not only contributes the highest volume of papers but that its activity centres on 2022–2023, aligning with the post-pandemic surge identified in the annual trend. The United Kingdom and India follow at a considerable distance. The British node, shaded in bluish-green, suggests a steadier output beginning earlier in the study period, whereas India's smaller but yellow-tinted node points to a more recent, rapidly growing engagement. Beyond these three actors, the landscape is dispersed. Countries such as Australia, Iran, Thailand and Italy appear in pale green, indicating mid-to-late period entry. Conversely, Canada, Japan and Germany are coloured blue, signalling that their principal contributions pre-date 2022 and that momentum has since slowed. Several emerging economies—including Bangladesh, Saudi Arabia and Kazakhstan—register single-digit outputs only in 2023 (yellow nodes), hinting at nascent research capacity that may expand as data and computation become more accessible. The absence of robust inter-national connections implies that research teams remain largely domestically oriented, despite the global nature of construction supply chains. Strengthening cross-border collaboration—through joint data repositories or coordinated benchmark challenges could speed the diffusion of residual learning advances and balance the presently uneven geographic distribution of expertise.

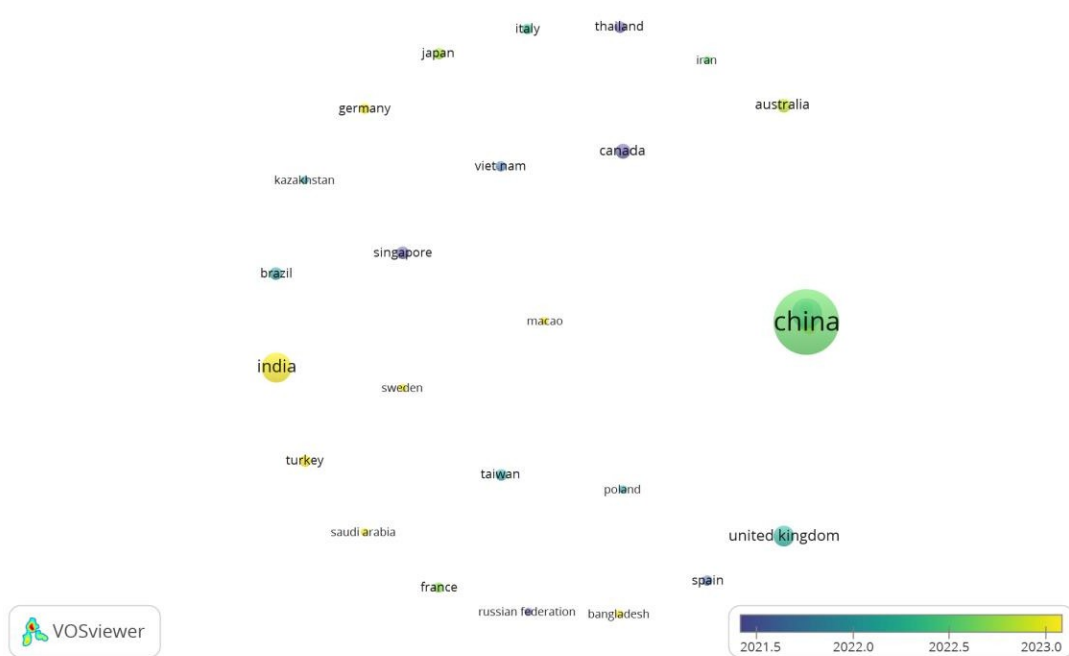

Fig. 2 Citation Analysis by Country

3.3 Reference co-citation network

Figure 3 visualises the reference co-citation network underlying the 258-document corpus. Each node represents a cited source, its diameter proportional to the frequency with which that work is co-cited, and the colour indicating the cluster assigned by VOSviewer's modularity algorithm. Edges reflect the strength of co-citation links; thicker ties denote sources that are frequently referenced together. The network is dominated by a single, multi-author hub labelled (He et al., 2016), Deep Residual Learning for Image Recognition paper that introduced ResNet. This node sits at the geometric centre and anchors virtually every colour cluster, confirming that the residual paradigm remains the pivotal methodological touchstone for vision research in construction. Radiating from this hub are several tightly knit yet partially overlapping communities. One cluster (green) groups seminal object-detection works. – Fast R-CNN (Girshick et al., 2015), FPN (Lin et al., 2017) and YOLO variants – signalling that construction scholars routinely import detection frameworks originally developed for natural-image benchmarks. A second cluster (red) orbits around segmentation classics such as U-Net, reflecting the field's emphasis on crack, defect and material boundary delineation. A third, purple-tinted cluster contains backbone papers on very-deep classification networks (VGG, Xception, Inception-v4), indicating that alternative architectural lineages continue to inform comparative studies even when residual models are the focal point. Several bridging nodes Mask R-CNN (He et al., 2017), Faster R-CNN and Attention-based works forge dense connections across these clusters, illustrating the integrative nature of recent research, where detection and segmentation tasks increasingly share unified multi-head architectures. The overall mesh is compact, with short path lengths between clusters, a structural property that speaks to the rapid cross-pollination of ideas within the broader computer-vision community before their adaptation to construction problems. Methodological developments in the larger AI ecosystem can be expected to propagate quickly into construction research; second, there remains room to establish a distinctive reference set tailored to the idiosyncrasies of the built environment such as irregular geometry, variable lighting and safety-critical constraints—that generic benchmarks seldom capture.

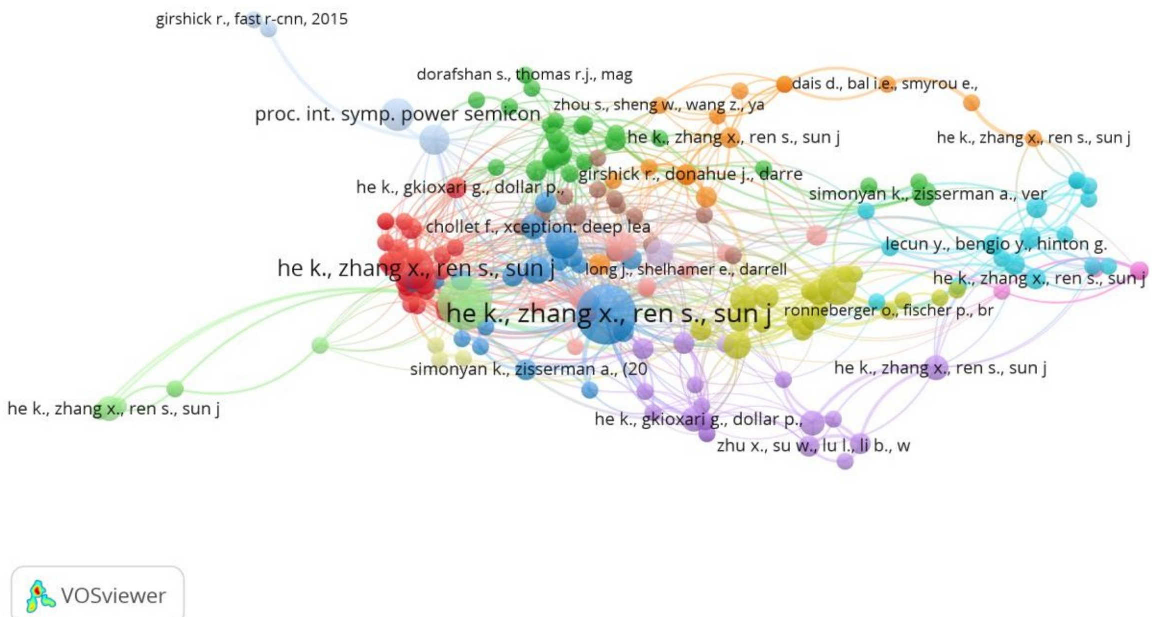

Fig 3. Reference co-citation network

3.4 Current Thematic Focus and Emerging Directions

Figure 4 displays the co-occurrence map of author-supplied keywords, produced with a minimum frequency threshold of five and fractional counting in VOSviewer. Node size is proportional to keyword frequency; edges indicate the number of publications in which the connected terms appear together, and modularity-based colours mark the principal thematic clusters that organise the literature.

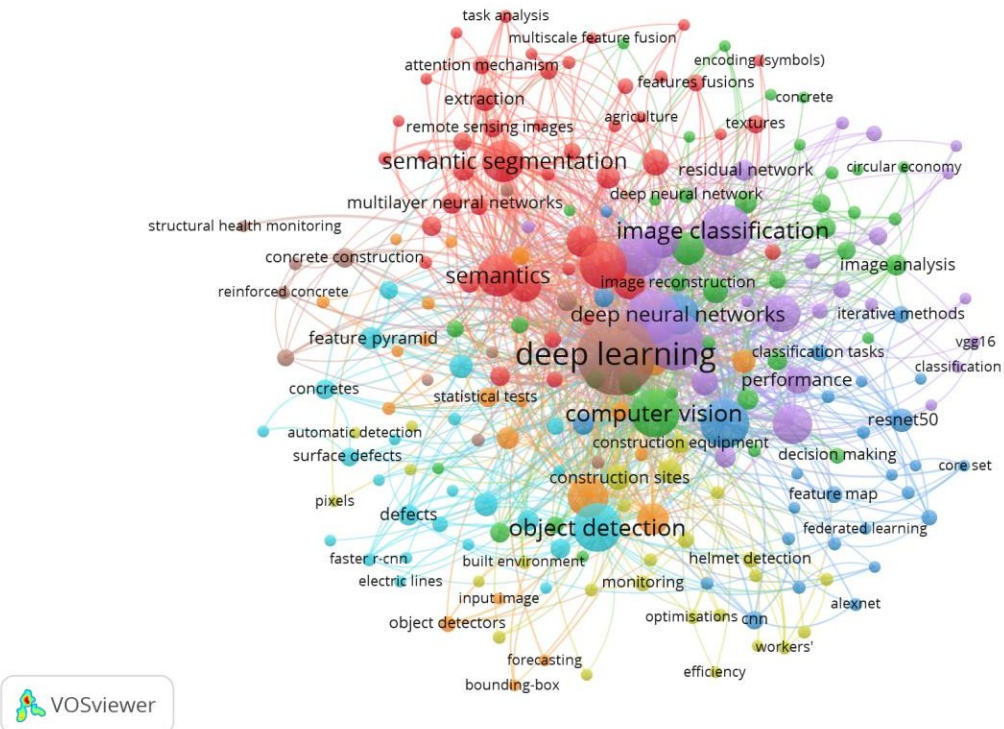

Fig. 4. Co-occurring Keywords

At the centre of the network, “deep learning” forms a gravitational hub and is densely linked to “computer vision,” “deep neural networks,” and “performance.” Their proximity reflects the methodological consensus that residual networks are conceived first and foremost as generic vision engines before being tailored to construction-specific tasks. Radiating from this core are three tightly knit yet conceptually distinct regions. The upper-left quadrant (red) is dominated by “semantic segmentation,” “semantics,” and ancillary phrases such as “feature pyramid,” “attention mechanism,” and “multiscale feature fusion.” The prominence of these terms suggests that fine-grained, pixel-level interpretation required for crack tracing, spalling measurement and remote-sensing damage assessment has become a mature sub-field. The inclusion of “structural health monitoring,” “reinforced concrete,” and “concrete construction” within the same cluster indicates that segmentation research is no longer confined to laboratory imagery; it is being deployed on genuine civil infrastructure datasets. In the lower-left quadrant (blue), “object detection” anchors a constellation of keywords that include “bounding box,” “YOLO,” “Faster R-CNN,” and practical terms such as “construction sites,” “defects,” “workers,” and “robotics.” The edges linking “helmet detection,” “monitoring,” and “forecasting” to this node illustrate how detection pipelines underpin automated safety management and progress-tracking tools. The comparatively shorter path lengths between detection-related keywords and “deep learning” underscore the field’s shift toward end-to-end frameworks rather than handcrafted feature pipelines. A third region, at the top of the map (green), revolves around “image classification,” “residual network,” and specific architectures such as “ResNet50” and “VGG16.” This cluster captures studies that benchmark backbone variants or apply classification to material identification, biophilic art curation, or circular economy investigations. The presence of a “circular economy” on the periphery reveals an emerging sustainability discourse in which image classifiers quantify waste streams or material stocks, an application space that has not been prominently featured in earlier reviews. Toward the right-hand side (purple), a compact group connects “federated learning,” “coreset,” “efficient,” and “optimization.” These keywords mark a nascent but rapidly growing interested in resource-aware and privacy-preserving training strategies—imperatives for on-site inference where bandwidth is limited and data sensitivity is high. Their adjacency to the core methodological terms but relative isolation from domain-specific phrases suggests that algorithmic efficiency research is still largely generic, awaiting deeper integration with construction use cases. Finally, the lower-right fringe (yellow) contains application-oriented labels such as “helmet detection,” “safety,” “workers,” and “forecasting.” The looser connectivity of this fringe indicates that many safety and logistics studies remain case-specific and have yet to converge on standardised taxonomies or benchmark datasets, a gap that could impede cross-project generalisability.

Table 1. Current Thematic Focus and Emerging Directions

Cluster	Typical vision task	Applications	Value proposition
Safety & risk mitigation	PPE / helmet compliance, crowd-distance, violence & fatigue detection	*Deep learning for site safety: Real-time detection of personal protective equipment* (Nath et al., 2020); *PPE-Glove Detection for Construction Safety Enhancement* (Gugssa et al., 2021)	Real-time alarms, lower incident rates, workforce analytics
Structural & material health	Crack/spall/weld flaw segmentation, concrete bug-hole grading, rail-steel inspection	*Automatic detection of building surface cracks using UAV and DL* (Wang et al., 2024); *Swish-ResNet for faulty-weld detection*	Rapid, objective QC and life-cycle documentation
Productivity & logistics	Dump-truck load counts, excavator pose, autonomous earth-work classification.	*Automatic Vision-Based Dump-Truck Productivity Measurement*; *Vision-based end-pose estimation for excavator manipulators.*	Data-driven progress tracking, idle-time diagnostics
Remote-sensing & large-area mapping	Waste-site localisation, open-pit road extraction, post-earthquake damage mapping	*T-SSD: Transformer-based single-stage multi-scale detector* (Huang et al., 2023)	Strategic planning, disaster assessment, carbon-stock audits
Digital-twin & BIM integration	Point-cloud segmentation (ResPointNet++), automatic façade material stocktake, SLAM for robot navigation	*Automated semantic segmentation of industrial point clouds using ResPointNet++*	Continuous, as-built model updates; progress-vs-plan analytics

Emerging thematic directions

The most conspicuous shift in the recent vision literature for construction engineering is the migration from purely convolutional backbones to transformer-centred detectors. Models such as DST-DETR, which couples a dehazing front-end with deformable DETR, achieve markedly higher recall for safety-helmet detection under adverse visibility than YOLOv5 or Faster-RCNN, illustrating the ability of multi-head attention to retaining small, highly occluded objects in cluttered scenes (Liu et al., 2024). Because PPE compliance and proximity monitoring are foundational to ISO 45001 safety regimes, the superior scale-robustness and long-range context modelling of transformer detectors represent a substantive advance in site-safety automation. A parallel development concerns resource-aware architectures optimised for edge deployment. Lightweight residual hybrids such as Mobile-DenseNet compress parameter counts to below five million while sustaining F1-scores above 0.90 in millimetre-scale crack segmentation of concrete façades (Akgül, 2023). Running on handheld cameras, smartphones or embedded GPUs, these networks close the latency gap between data capture and decision-making, enabling inspectors or drones to generate quality control reports without cloud connectivity which is an essential capability on remote or power-constrained sites. Data scarcity and privacy constraints have stimulated interest in weak, self- and federated-learning paradigms. A recent YOLOv8-based surveillance framework trains through hierarchical federated learning at edge nodes with only a three-per-cent mean average-precision loss relative to centralised training, while keeping all raw imagery in situ (Sun et al., 2023). Such approaches allow independent contractors or asset owners to share gradient updates rather than images, lowering the institutional barriers to AI adoption across fragmented project consortia.

Concurrently, residual learning is being extended to multimodal perception and digital-twin integration. ResPointNet++, which fuses residual blocks with PointNet++ set abstractions, delivers 87 % mean intersection-over-union in the semantic segmentation of industrial point clouds—14 percentage points above classic benchmarks (Yin et al., 2021); (Savino and Tondolo, 2023). By piping these labelled 3-D data streams into BIM or digital-twin platforms, practitioners can automate progress quantification, clash detection and inventory updates, thereby shortening the scan-to-BIM feedback loop that underpins schedule and cost control. As computer-vision outputs begin to inform contractual decisions, explainability and robustness have become critical research foci. Integration of Grad-CAM++ with residual crack classifiers has allowed engineers to visualise the pixel regions driving defect predictions, facilitating model debugging and engendering trust among regulators and insurers (Zhang et al., 2019); (Kumar et al., 2023); (Mayya and Alkayem, 2024). Complementary work on adversarial attack audits reveals the vulnerability of CNNs to imperceptible perturbations, underscoring the need for certifiable robustness before AI assessments can be embedded in acceptance criteria. Finally, diffusion-based synthetic data pipelines are emerging as powerful tools for alleviating annotation bottlenecks. DefectFill, for example, exploits latent-diffusion inpainting to generate high-fidelity concrete- and steel-surface defects that raise inspection accuracy on MVTec-AD by up to eighteen percentage points without additional field photography (Alkannad et al., 2025). Because the method can be conditioned on material, texture and lighting parameters, it enables balanced, privacy-preserving datasets that democratise model training for smaller contractors.

Future research opportunities

Despite these advances, several gaps remain. First, the cross-domain transferability of transformer and residual detectors is poorly understood: models trained on high-rise imagery perform unpredictably when applied to tunnels, dams or offshore platforms. Longitudinal studies that quantify domain-shift impacts and devise on-device, few-shot fine-tuning routines are needed to ensure reliable generalisation across the diverse topologies of civil infrastructure. Second, true cyber-physical safety loops will require orchestrating edge-accelerated inference—already achievable in sub-50 ms on FPGA implementations of residual blocks with 5 G/6 G uplinks and digital-twin application-programming interfaces; research on scheduling, bandwidth allocation and control co-design is thus imperative. Third, the fusion of RGB imagery with LiDAR, thermography and vibration signals remains largely heuristic. Unified multimodal encoders that learn shared embeddings could support holistic structural health monitoring and resilient infrastructure objectives in line with SDG 9. Fourth, as regulatory frameworks evolve, explainable AI dashboards must translate saliency maps into actionable maintenance or safety directives and integrate seamlessly with ISO 19650 information-management workflows. Fifth, simulation-in-the-loop methodologies that combine physics-based rendering with diffusion-based texture synthesis could stress-test perception systems under rare or hazardous scenarios—such as scaffold collapse or extreme weather without endangering personnel. Finally, the sustainability footprint of pervasive site vision remains under-explored; future work should investigate neural-network-on-chip design, adaptive frame skipping and quantised residual blocks to align energy consumption with net-zero construction targets. Collectively, addressing these challenges will move the field from task-specific prototypes toward a cohesive, trustworthy and resource-efficient vision ecosystem capable of underpinning the next generation of autonomous, safe and sustainable construction engineering.

4. Discussion

This bibliometric study offers the first systematic synthesis of how deep-residual architectures are being mobilised for image-based analytics in construction engineering. Three interlocking insights emerge. First, the sharp rise in publications after 2019 confirms that residual learning has moved beyond exploratory trials to become an accepted work-horse for site vision. The acceleration coincides with the commercial availability of cost-effective UAVs, edge GPUs and open-source model zoos, all of which lowered the entry barrier for construction scholars. Similar inflexion points were noted when 3-D laser scanning became affordable a decade earlier, suggesting that hardware democratisation is a recurrent catalyst for digital innovation in the sector (Li & Wen, 2017); (Zhu et al., 2023). The co-citation and keyword maps expose an epistemic asymmetry: methodological anchors remain almost exclusively rooted in generic computer-vision literature—ResNet, U-Net, YOLO—whereas domain-specific references from leading construction journals are sparse. This pattern signals the sector's heavy dependence on exogenous algorithmic advances. The thematic clustering further delineates the field into three mature pillars; semantic segmentation, object detection and image classification and two emergent strands: resource-aware/federated learning and sustainability-oriented applications such as circular economy assessments.

Transferability studies should quantify how residual and transformer detectors trained on high-rise imagery perform in tunnels, dams or offshore platforms, mirroring recent cross-domain evaluations in autonomous driving (Chen et al., 2023). Edge-to-cloud orchestration frameworks must be devised to fuse sub-50 ms FPGA inference with 5 G/6 G uplinks and BIM application programming interfaces. Multimodal encoders capable of learning joint embeddings from RGB, LiDAR, thermography and vibration data could advance holistic asset-health diagnostics and realise the resilient infrastructure targets of SDG 9. Equally crucial is the development of explainable AI dashboards that map saliency information onto actionable maintenance directives—an essential step for meeting the transparency clauses of ISO 19650 and forthcoming AI governance regulations.

5. Conclusions

This review shows that deep residual learning has evolved from a conceptual breakthrough to an indispensable engine for image analytics across the construction life cycle. Publication outputs grew almost six-fold between 2019 and 2024, driven by affordable sensing hardware, open-source model repositories and the advent of edge accelerators. The knowledge base, while dominated numerically by Chinese institutions, remains globally diffuse and weakly interconnected, suggesting that cross-border collaboration has yet to match the sector's transnational supply chains. Co-citation analysis confirms that construction researchers continue to draw methodological guidance from mainstream computer-vision literature—ResNet, U-Net, YOLO, while adapting these backbones to safety monitoring, structural diagnostics and productivity tracking. Keyword clustering highlights a mature segmentation and detection core, a growing preoccupation with resource-aware and privacy-preserving training, and an emergent sustainability narrative wherein residual networks aid material circularity assessments. The synthesis also uncovers substantive blind spots. Few studies interrogate model robustness across divergent asset classes, climatic zones or sensor modalities; fewer still integrate explainable inference with ISO 19650-compliant information workflows or measure the embodied carbon of pervasive vision pipelines. Addressing these gaps will require longitudinal, cross-domain benchmarks, neuromorphic or quantised deployments tailored to site-level power budgets, and simulation-in-the-loop validation capable of stress-testing algorithms under safety-critical extremes. Realising such an agenda will move the field beyond isolated proofs-of-concept toward a coherent, trustworthy and resource-efficient vision ecosystem, an essential foundation for the autonomous, data-driven and sustainable construction industry envisioned for the decade ahead.

References

Akgül, İ., 2023. Mobile-DenseNet: Detection of building concrete surface cracks using a new fusion technique based on deep learning. Heliyon 9.

Alkannad, A.A., Smadi, A.A.L., Al-Makhlafi, M., Yang, S., Feng, Z., 2025. CrackVisionX: A Fine-Tuned Framework for Efficient Binary Concrete Crack Detection. IEEE Trans. Intell. Transp. Syst.

Aziminezhad, M., Taherkhani, R., 2023. BIM for deconstruction: A review and bibliometric analysis. J. Build. Eng. 73, 106683.

Chen, R., Hei, L., Lai, Y., 2023. Object detection in optical imaging of the Internet of Things based on deep Learning. PeerJ Comput. Sci. 9.

Deng, T., Sharafat, A., Lee, S., Seo, J., 2024. Automatic Vision-Based Dump Truck Productivity Measurement Based on Deep-Learning Illumination Enhancement for Low-Visibility Harsh Construction Environment. J. Constr. Eng. Manag. 150.

Fahad, A., Nayem, N.H., Hossain, M.N., Rabbani, M.L., Opu, R.K., Al Shuaeb, S.M.A., 2024. Sand fineness modulus prediction in construction sector using convolutional neural network. Asian J. Civ. Eng. 25, 443–450.

Gugssa, M., Gurbuz, A., Wang, J., Ma, J., Bourgouin, J., 2021. PPE-Glove Detection for Construction Safety Enhancement Based on Transfer Learning. In: R.R.A., I. (Ed.), Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021. American Society of Civil Engineers (ASCE), Richard A. Rule School of Civil and Environmental Engineering, Mississippi State Univ., Mississippi State, MS, United States, pp. 58–65.

He, K., Xiangyu Zhang, Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. arXiv.1512.03385v1.

Huang, K., Wen, M., Wang, C., Ling, L., 2023. T-SSD: A Transformer-based Single-Stage Multi-Scale Sampling Object Detector. In: 13th International Workshop on Computer Science and Engineering, WCSE 2023. International Workshop on Computer Science and Engineering (WCSE), College of Computer Science and Technology, Shanghai University of Electric Power, China, pp. 156–163.

Jadhav, L.R., 2025. IoT networking for worker safety monitoring using the construction site images and worker health records. Int. J. Intell. Inf. Database Syst. 17, 236–267.

Kumar, A., Martin, H., Leon, L., 2023. CONCRETE DAMAGE IDENTIFICATION FOR STRUCTURAL HEALTH MONITORING USING COMPUTER VISION. In: CICE 2023 - 11th International Conference on FRP Composites in Civil Engineering. International Institute for FRP in Construction (IIFC), Queen's University, Belfast, United Kingdom.

Liu, Z., Sun, C., Wang, X., 2024. DST-DETR: Image Dehazing RT-DETR for Safety Helmet Detection in Foggy Weather. Sensors 24.

Lukovic, M., Ciernik, L., Müller, G., Kluser, D., Pham, T., Burgert, I., Schubert, M., 2024. Probing the complexity of wood with computer vision: from pixels to properties. J. R. Soc. Interface 21.

Mayya, A.M., Alkayem, N.F., 2024. Enhance the Concrete Crack Classification Based on a Novel Multi-Stage YOLOV10-ViT Framework. Sensors 24.

Muhammed Sufail, M.K., Jacob, C., Ponnan, C.D., 2024. An Integrated Deep Learning Approach for Crack Detection and Localization in Tyre Images. In: RAICS - IEEE Recent Advances in Intelligent Computational Systems. Institute of Electrical and Electronics Engineers Inc., Centre for Artificial Intelligence, Tkm College of Engineering, Kerala, Kollam, India.

Nath, N.D., Behzadan, A.H., Paal, S.G., 2020. Deep learning for site safety: Real-time detection of personal protective equipment. Autom. Constr. 112, 103085.

Onososen, A.O., Musonda, I., 2022. Perceived Benefits of Automation and Artificial Intelligence in the AEC Sector: An Interpretive Structural Modeling Approach. Front. Built Environ. 8, 1–16.

Pedral Sampaio, R., Aguiar Costa, A., Flores-Colen, I., 2023. A discussion of digital transition impact on facility management of hospital buildings. Facilities 41, 389–406.

Savino, P., Tondolo, F., 2023. Civil infrastructure defect assessment using pixel-wise segmentation based on deep learning. J. Civ. Struct. Heal. Monit. 13, 35–48.
Shahadat, N., Maida, A.S., 2025. Deep Residual 1-Dimensional Convolutional Neural Networks in Vision. In: A., A., S., C., R., C., C.-L., L., S., B., U., P. (Eds.), Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) . Springer Science and Business Media Deutschland GmbH, Truman State University, Kirksville, MO, United States, pp. 129–144.
Sun, K., 2024. DMFF: Deep multimodel feature fusion for building occupancy detection. Build. Environ. 253.
Sun, S., Peng, T., Huang, H., 2023. Machinery Prognostics and High-Dimensional Data Feature Extraction Based on a Transformer Self-Attention Transfer Network. Sensors 23.
Tang, S., Golparvar-Fard, M., 2021. Machine Learning-Based Risk Analysis for Construction Worker Safety from Ubiquitous Site Photos and Videos. J. Comput. Civ. Eng. 35.
Tonetto, M.S., Formoso, C.T., Saurin, T.A., Bonesi De Luca, F., Lora, F.P., Lantelme, E., 2023. Resilient performance on construction projects in the post-pandemic era: an organizational perspective. Eng. Constr. Archit. Manag.
Wang, J., Wang, P., Qu, L., Pei, Z., Ueda, T., 2024. Automatic detection of building surface cracks using UAV and deep learning-combined approach. Struct. Concr. 25, 2302–2322.
Yin, C., Wang, B., Gan, V.J.L., Wang, M., Cheng, J.C.P., 2021. Automated semantic segmentation of industrial point clouds using ResPointNet++. Autom. Constr. 130.
Zhang, J., Lu, C., Wang, J., Wang, L., Yue, X.-G., 2019. Concrete cracks detection based on FCN with dilated convolution. Appl. Sci. 9.
Zhu, J., Shi, Q., Li, Q., Shou, W., Li, H., Wu, P., 2023. Developing predictive models of construction fatality characteristics using machine learning. Saf. Sci. 164.