Kalpa Publications in Computing

Volume 22, 2025, Pages 829–840

Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

Vision LLM-Driven Operational Hazard Recognition for Building Fire Safety Compliance Checking

Dayou Chen[1], Long Chen[2], Yiheng Zeng[3], Craig Hancock[4], Russell Lock[5] and Simon Sølvsten[6]

1) Ph.D. Candidate, School of Architecture, Building and Civil Engineering, Loughborough University, United Kingdom. Email: d.chen@lboro.ac.uk

2) Ph.D., Assoc. Prof., Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, SAR. Email: longchen@cityu.edu.hk

3) Department of Computer Science, University College London, United Kingdom. Email: leo.zeng.22@ucl.ac.uk

4) Ph.D., Reader, Architecture, Building and Civil Engineering, Loughborough University, United Kingdom. Email: c.m.hancock@lboro.ac.uk

5) Ph.D., Reader, Department of Computer Science, Loughborough University, United Kingdom. Email: r.lock@lboro.ac.uk

6) Ph.D., Assoc. Prof., European Center for Risk & Resilience Studies, University of Southern Denmark, Denmark. Email: simos@sam.sdu.dk

Abstract

Building fire incidents pose significant risks to human lives and property, making fire safety compliance a critical aspect of building management. Traditional compliance checks are largely manual, relying on expert inspectors to assess and report on fire safety standards. While prior research has explored Automated Compliance Checking (ACC) during the design phase, limited attention has been given to the operational phase, where dynamic risks necessitate continuous monitoring. This study proposes a novel approach that leverages vision Large Language Models (vLLMs) to automate fire safety compliance monitoring in the operational phase. The developed method frames hazard recognition as a Visual Question Answering (VQA) task, enabling the model to analyze visual data and respond to textual queries regarding potential fire hazards. The system employs a Vision Transformer (ViT) for visual encoding and a multimodal fusion process, allowing the vLLM to generate contextually relevant descriptions of observed hazards, along with regulatory references including Occupational Safety and Health Administration (OSHA) standards. Evaluation results demonstrate significant improvements in hazard recognition over a generic vLLM baseline, with an average BLEU score of 0.1355 compared to 0.0410 and higher ROUGE scores reflecting

I. Hkust, J. Cheng and Y. Yantao (eds.), ICCBEI 2025 (Kalpa Publications in Computing, vol. 22), pp. 829–840

superior precision and coherence. The model's ability to automatically generate structured hazard description reports has practical implications for assisting expert-driven inspections, offering a comprehensive and effective solution for long-term fire safety management. This study thus advances ACC research by providing a comprehensive, automated method for continuous fire safety compliance in operational building environments.

1 Introduction

Building fire incidents pose significant threats to human lives and properties, making fire safety compliance an essential aspect of building management. Regulatory frameworks are in place to ensure a minimum standard of safety and performance for built assets, yet these checks predominantly depend on manual inspections. Although prior research has focused on automated compliance checking (ACC) during the design phase, there is a notable gap in automating compliance checking throughout the operational phase of buildings. This gap is especially concerning as fire risks require continuous monitoring to ensure safety measures remain effective beyond initial design compliance.

The operational phase presents unique challenges for fire safety compliance, where risks are dynamic and necessitate ongoing assessment. Fire safety measures, such as the installation of fire doors and maintenance of unobstructed escape routes, are critical during the design and construction phase but may be compromised over time. For instance, fire doors could be propped open, or escape paths could be blocked by stored items, undermining their intended purpose. Thus, continuous monitoring is essential to maintain the integrity of these safety measures, effectively mitigating risks that may emerge during a building's operational lifespan.

Table 1. Summary of Automated Code Compliance Checking Methods

Literature	Applicable Phases	Technological Approach
Malsane et al. (2015)	Design Phase	Developing object model for automated compliance checking
Jiang et al. (2022)	Design Phase	Using ontology mapping and rule-based reasoning for ACC
Fitkau and Hartmann (2024)	Design Phase	Ontology-based knowledge formalization for ACC in fire safety
Zhang and El-Gohary (2017)	Design Phase	Using NLP and logic reasoning for fully automated code checking
Zhou et al. (2022)	Design Phase	Integrating NLP and CFG for rule interpretation in ACC
Bloch et al. (2023)	Design Phase	Using graph neural networks for ACC
Chen et al. (2024)	Design Phase	Using computer vision and deep generative models for ACC
Bosché (2010)	Construction Phase	Automated recognition of CAD objects in laser scans for compliance control
Cheng et al. (2022)	Construction Phase	Computer vision and deep learning for safety compliance monitoring
Ding et al. (2022)	Construction Phase	Using visual question answering and deep learning for safety compliance checking
Beach et al. (2024)	Operational Phase	Feasibility study on moving ACC to operational phase

Table 1 summarizes the key contributions of previous research in ACC, highlighting the applicable phases and technological approaches employed. Historically, the literature on ACC has emphasized Building Information Modeling (BIM) as a primary tool for compliance verification during the design phase. This emphasis has led to substantial advancements in areas such as structural safety, fire risk assessment, and water distribution systems. These studies typically rely on BIM data combined with rule-based reasoning, aligning design models with regulatory standards. For example, Martins and Monteiro (2013) developed the "LicA" application for automating water distribution network checks using IFC-based BIM models, while Malsane et al. (2015) introduced an object model for fire safety compliance. Zhang et al. (2013) further extended BIM's application to safety rule checking, applying algorithms to prevent fall hazards in construction planning. These design-phase solutions illustrate BIM's capacity to enhance safety and regulatory compliance early in a building's lifecycle.

Recent advances in Natural Language Processing (NLP) and Machine Learning (ML) have further enhanced ACC automation by reducing reliance on manual interpretation of regulatory texts. Zhang and El-Gohary (2017) and Zhou et al. (2022) developed NLP frameworks to extract regulatory requirements directly from text, converting complex legal language into computable rules. These innovations are crucial for addressing one of the most labor-intensive components of ACC, facilitating a more scalable approach to compliance checking. Additionally, Bloch et al. (2023) applied Graph Neural Networks (GNN) to ACC, specifically targeting accessibility requirements in residential designs. By bypassing traditional hard-coded rules, GNNs enable a more flexible and scalable approach, demonstrating the potential of ML techniques in overcoming the limitations of conventional rule-based systems.

While recent studies have started to address the gap in lifecycle-wide compliance with an emphasis on extending ACC into the operational phase, significant work remains to be done. Beach et al. (2024) advocated for automated data capture and analysis to facilitate compliance during building operations. However, fully automated compliance checking and monitoring specifically tailored for the operational phase of buildings is yet to be developed.

The use of computer vision in ACC has become an emerging focus, particularly for compliance monitoring and hazard recognition. Cheng et al. (2022) developed a deep learning model for classifying Personal Protective Equipment (PPE) and tracking worker movement on construction sites. Their approach highlights how vision-based models can support real-time monitoring of safety compliance by recognizing and categorizing visual data relevant to worker safety. In parallel, Ding et al. (2022) leveraged a Vision-and-Language Transformer (ViLT) model for Visual Question Answering (VQA) to detect unsafe behaviors on construction sites. By incorporating both visual and language processing capabilities, their model exemplifies a more advanced integration, enabling nuanced hazard detection through complex reasoning about images and textual queries.

These studies underscore the potential of image-based data for automating hazard recognition in ACC, signaling a shift toward image-driven applications that support operational compliance monitoring. The rapid advancement of vision models—especially with contrastive learning techniques—has further enabled complex scene understanding (Radford et. al, 2021). This recent technique aligns visual data with text for richer interpretation beyond traditional vision-based tasks such as object detection. Recently, Large Language Models (LLMs) have expanded these capabilities, allowing more sophisticated reasoning tasks on image inputs (Touvron et. al, 2023; Liu et. al, 2024). Yet, despite these advancements, the application of such technologies to hazard recognition for image-driven ACCs in building operational environments remains largely unexplored.

In light of these advancements, this study proposes a novel approach using computer vision, particularly vision Large Language Models (vLLM), to automatically identify fire safety non-compliance in buildings during the operational phase. By enabling long-term, dynamic compliance monitoring, this method seeks to assist managers and inspectors in generating compliance reports,

reducing the workload for fire safety experts and improving the efficiency of ongoing fire risk management.

2 METHOD

Ensuring fire safety compliance during the operational phase of a building requires a systematic approach capable of identifying dynamic hazards in real time. This study introduces a novel method leveraging vision-Large Language Models to automate hazard recognition and reporting. The proposed approach frames hazard detection as a Visual Question Answering problem, enabling the generation of detailed and context-aware descriptions of fire safety non-compliance. The framework, referred to as Fire Compliance Visual Question Answering (FCVQA), integrates computer vision and natural language processing techniques to analyze multimodal inputs and produce actionable outputs.

The following subsections describe the problem formulation, the FCVQA framework, the model training process, and the evaluation metrics used to assess the model's performance. Each component contributes to the overall objective of automating compliance checks in operational building environments, offering an effective solution to augment traditional expert-based inspection workflows.

2.1 Problem Formulation

To enable automated assessment of potential fire hazards in the operational phase, we deploy a vision-Large Language Model (vLLM) to perform hazard recognition in building environments. This model uses advanced computer vision techniques to evaluate risks based on visual inputs, offering an advanced method for identifying non-compliance with building regulations in real time.

The hazard recognition task is framed as a Visual Question Answering (VQA) problem (Antol et. al, 2015), where the model receives an image of a building environment alongside a textual query about potential fire hazards. The vLLM integrates these inputs to generate a description of any detected risks. For example, given an image with the query, "What fire hazards do you see?" the model may respond with, "The fire exit is obstructed by a large box," or "The fire door is propped open, compromising its function."

Formally, this task can be described as:

$$M(I, Q) \rightarrow A, \quad (1)$$

where M is the vLLM, I is the input image, Q is a natural language question about possible hazards, and A is the generated answer describing any observed fire safety non-compliance issues.

2.2 Fire Compliance Visual Question Answering (FCVQA) Framework

This work introduces the Fire Compliance Visual Question Answering (FCVQA) framework for effective hazard recognition in building operations. The FCVQA framework, as depicted in Figure 1, integrates visual and textual data through a multimodal encoding and decoding process, leveraging the robust reasoning capabilities of large language models (LLMs) for contextual hazard assessment.

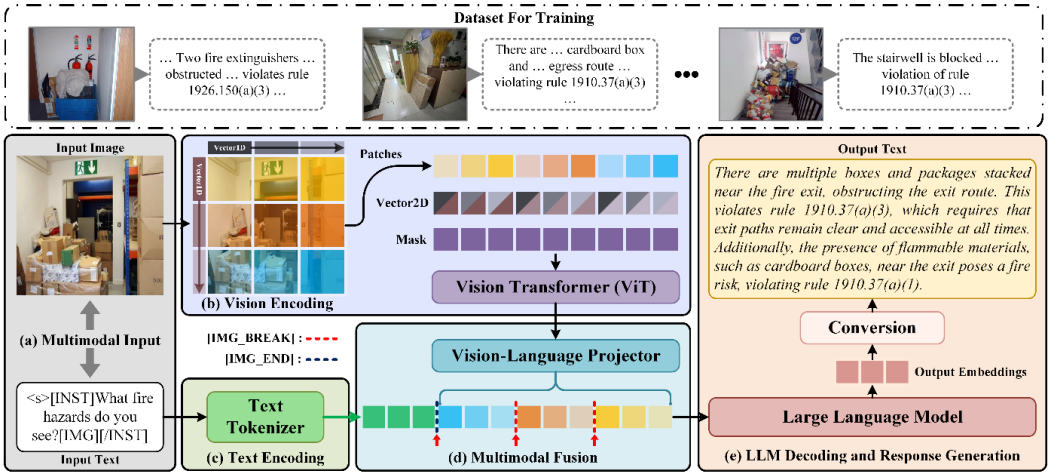

Figure 1. The proposed vLLM-driven FCVQA framework for operational hazards recognition

The FCVQA process includes several key stages:

Vision Encoding: A Vision Transformer (ViT) encodes the visual input, transforming the building environment's image into a high-dimensional latent representation. This representation captures relevant details, such as object shapes and spatial arrangements, essential for identifying fire hazards.

Text Encoding: The textual query about fire risks is tokenized and encoded to produce a compatible latent representation. This enables the model to interpret specific fire-related questions, such as identifying blocked exits or verifying the functionality of fire doors.

Multimodal Fusion: The encoded image and text representations are combined within a shared latent space, allowing the vLLM to process visual and textual information concurrently. This fusion enables the model to analyze complex visual contexts and correlate them with fire safety questions.

LLM Decoding and Response Generation: Using a large language model decoder, the FCVQA framework processes the combined latent representation to generate a textual response. This output describes any fire hazards detected in the image, providing actionable insights aligned with the original query. For instance, if the image depicts a blocked fire exit, the model would generate a response indicating this specific non-compliance issue.

By leveraging the FCVQA framework, the system can dynamically assess fire safety conditions in real time. This approach enables comprehensive hazard recognition, capturing both obvious and subtle risks within building environments. Examples of hazards that the vLLM can detect include obstructed exits, inadequately stored flammable materials, and disabled fire safety equipment, all of which are essential for ongoing compliance with fire safety regulations. The vLLM's strength lies in its ability to interpret and reason about complex scenes. Unlike conventional object detection models, which only localize predefined objects, vLLMs can understand contextual factors that determine whether an object or scene poses a fire hazard. For instance, a stack of furniture may not be inherently dangerous, but when positioned to obstruct egress, it becomes a regulatory violation. The vLLM's capacity for contextual reasoning is therefore essential for identifying these nuanced risks, allowing for a more sophisticated assessment of fire safety compliance.

2.3 Model Training

The model training approach first follows a vision-language pre-training (VLP) strategy, where larger datasets are utilized to pre-train models on various vision and language tasks prior to specific

Visual Question Answering (VQA) training. This VLP strategy, as discussed by Gan et. al (2022), facilitates the transfer of knowledge from a range of related tasks, thereby enhancing the model's understanding of concepts that go beyond the constraints of limited VQA datasets. For VLP, interpretative VQA datasets were reviewed as potential sources of pre-training data. Several popular VQA datasets representing diverse scenarios were combined to enrich the pre-training phase, specifically VizWiz-VQA (Gurari et. al, 2018) and VQAv2 (Goyal et. al, 2017).

Following the pre-training, the model was fine-tuned using a smaller, specialized dataset developed for this study, referred to as the Fire Compliance VQA dataset. This dataset comprises 135 image-text pairs focused on common operational fire hazards in building environments, such as blocked fire exits. This small dataset serves as the foundation for this preliminary experiment to validate the feasibility of the proposed approach. The dataset was split into training, validation, and test sets with an 8:1:1 ratio. Model training was conducted on a single Nvidia A6000 GPU, utilizing a learning rate of 0.001. Each image is paired with the question, 'What fire hazards do you see?' alongside a human-generated response that describes the scene, identifies non-compliance issues, and references relevant Occupational Safety and Health Administration (OSHA) regulations. In this work, OSHA regulations relating to the maintenance, safeguards, and operational features for exit routes, as outlined in Maintenance, Safeguards, and Operational Features for Exit Routes (29 C.F.R. § 1910.37; OSHA, n.d.), and fire protection, as detailed in Fire Protection (29 C.F.R. § 1926.150; OSHA, n.d.), have been selected and used. The response format mirrors a hazard-report style, offering a practical approach to support or potentially replace expert-led building compliance inspections by providing precise, regulatory-aligned insights.

During the training phase, we employed full supervision to optimize the model. For VQA, the sigmoid function predicted character scores $\hat{s}$ between 0 and 1 as probabilities for the answer. We utilized the widely used binary cross-entropy (Teney et. al, 2018) as the guiding loss function. Given a dataset D having n samples, with image $v \in V$,question $q \in Q$ and answer $a \in A$, the goal is to train a model to optimize a mapping function $f: V \times Q \rightarrow \mathbb{R}^{|A|}$. Hence, the answering loss L_{ans} can be formula as follows:

$$L_{ans} = -\sum_{q}^{Q}\sum_{a}^{A} s_{qa}\log(\hat{s}_{qa}) - (1 - s_{qa})\log(1 - \hat{s}_{qa}), \quad (2)$$

Where s_{qa} is the true score or target probability for answer a given question q. $\hat{s}_{qa}$ is the predicted probability score for answer a given question q, obtained using the sigmoid activation function on the model's output logits. This score represents the model's confidence in predicting a as the correct answer for q. This loss function L_{ans} calculates the cumulative penalty across all questions and possible answers in the dataset, guiding the model to produce probability scores that align with the true answer distribution for each question.

2.4 Evaluation Metrics

To evaluate the accuracy and effectiveness of hazard identification and compliance flagging, several established metrics were utilized: BLEU-4 (Papineni et. al, 2002), ROUGE-1, ROUGE-2, and ROUGE-L (Lin,2004). These metrics measure the alignment between the generated descriptions and reference texts, focusing on precision in identifying key compliance terms, such as "fire extinguisher," "fire exit," and specific regulatory codes. In this evaluation, we compare the pre-trained and fine-tuned model using our proposed method against the original, generic vLLM model as the baseline to validate the effectiveness of the proposed approach.

BLEU-4 evaluates the precision of four-gram sequences in the generated text, capturing phrase accuracy critical for detailed compliance reporting. ROUGE-1 measures unigram overlap, verifying

that essential terms are included, while ROUGE-2 assesses bigram overlap, ensuring accuracy in phrase sequences. ROUGE-L calculates the longest common subsequence between generated and reference texts, reflecting structural coherence.

Formally, these metrics are defined as follows:

$$\text{BLEU-4} = \exp\left(\sum_{n=1}^{4} \frac{1}{n} \log p_n\right) \times \text{brevity penalty}, \quad (3)$$

$$\text{ROUGE-1} = \frac{\sum \text{matched unigrams}}{\sum \text{reference unigrams}}, \quad (4)$$

$$\text{ROUGE-2} = \frac{\sum \text{matched bigrams}}{\sum \text{reference bigrams}}, \quad (5)$$

$$\text{ROUGE-L} = \frac{\text{LCS(generated, reference)}}{\text{length of reference}}, \quad (6)$$

These metrics provide a comprehensive evaluation framework for the model's ability to generate precise, contextually relevant descriptions, essential for automated hazard detection and compliance monitoring.

3 RESULTS

Table 2. Evaluation Metrics Comparison

Method	Average BLEU	ROUGE-1	ROUGE-2	ROUGE-L
Generic vLLM	0.0410	0.3912	0.1086	0.2435
Proposed Method	**0.1355**	**0.5375**	**0.2590**	**0.3683**

The evaluation metrics for hazard identification performance were compared between the generic vLLM baseline and the proposed method. Table 2 highlights the substantial improvements achieved by the proposed method in both BLEU and ROUGE metrics. The proposed method's average BLEU score reached 0.1355, compared to 0.0410 from the generic vLLM baseline, reflecting enhanced precision in descriptive phrasing. For ROUGE metrics, the proposed method also outperformed the baseline significantly, with ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.5375, 0.2590, and 0.3683, respectively, compared to the baseline's scores. These results underscore the proposed method's greater effectiveness in capturing key terms, maintaining phrase structure, and preserving the overall coherence required for precise hazard identification and compliance monitoring.

Image Input

A bicycle is placed directly in front of the fire extinguisher and fire hose cabinet, obstructing access to both. This violates rule 1926.150(a)(3), which requires that fire safety equipment, such as fire extinguishers and fire safety equipment, remain accessible and unobstructed at all times.

FCVQA response

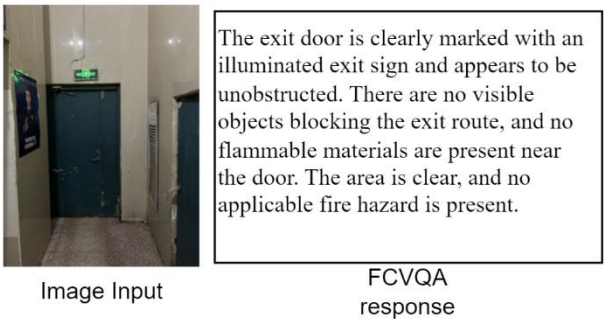

Figure 2. Example output of the developed FCVQA model

Upon closer examination of the model's specific outputs, as illustrated in Figure 2, it is evident that the developed FCVQA model demonstrates strong capabilities in identifying and reporting fire safety compliance issues within building environments. In the first example, the FCVQA model detects that a bicycle is obstructing access to a fire extinguisher and a fire hose cabinet. The model accurately cites OSHA regulation 1926.150(a)(3), which mandates that fire safety equipment must always remain accessible. This output demonstrates the model's ability not only to recognize objects in the scene but also to interpret the situation within the context of fire safety compliance, correlating the obstruction with relevant legal requirements. At the same time, the second example shows the model's capability to identify a compliant scene. The FCVQA model observes that the exit door is clearly marked and unobstructed, with no visible hazards nearby. It concludes that no applicable fire hazard is present, illustrating the model's capacity to affirm compliance in safe environments. This ability to confirm safety in compliant situations is critical for practical deployment, as it enables the model to assist or potentially replace expert-based inspections by offering precise, context-aware assessments of building fire safety compliance.

4 DISCUSSION

This section evaluates the contributions and limitations of the developed model within the broader context of ACC research. First, a comparative analysis highlights the advancements of the proposed method over existing approaches, particularly in terms of complex hazard scene reasoning and automated hazard report generation. Following this, the limitations of the current framework are discussed, along with potential future directions to address these challenges and enhance the model's applicability and robustness. Together, these discussions provide a comprehensive perspective on the strengths and areas for improvement of the proposed method in advancing fire safety compliance monitoring.

4.1 Comparative Analysis

In discussing the capabilities of vision-based approaches for safety compliance monitoring, Table 3 highlights key distinctions between the developed method and prior work by Cheng et al. (2022) and Ding et al. (2022). Specifically, the term "Complex hazard scene reasoning" refers to a model's capacity to interpret intricate hazard situations beyond simple object recognition tasks, such as those related to Personal Protective Equipment (PPE). Traditional CNN-based models, such as those

utilized by Cheng et al. (2022), are limited to identifying isolated objects and lack the interpretative depth necessary for complex hazard assessment. In contrast, vision-language models, as applied in Ding et al. (2022) and the developed approach, provide a more nuanced understanding of contextual safety risks, enabling a richer analysis of compliance within complex scenes.

Table 3. Comparison of Vision-Based Safety Compliance Methods

Literature	Complex hazard scene reasoning	Written report generation
Cheng et. al (2022)	N	N
Ding et. al (2022)	Y	N
Developed method	Y	Y

An additional, critical differentiator is the capability for written report generation. While Ding et al. (2022) employed a vision-language model capable of reasoning about hazard scenes, it did not integrate a mechanism for generating fully automated hazard reports. This limitation necessitates expert intervention to interpret the model's outputs and document findings, which can be both time-consuming and prone to human error. Automated report generation, as realized in the developed method, fills this gap by producing structured, regulatory-aligned hazard descriptions that support compliance inspections.

The ability to generate comprehensive, written hazard reports represents a significant advancement in the automation of safety compliance workflows. This feature is essential for either assisting or potentially replacing traditional expert-driven processes, as it reduces the reliance on human inspectors to document compliance findings. By offering a model that interprets hazards with regulatory context and generates formalized reports, the developed approach not only enhances the efficiency of compliance monitoring but also contributes to automated hazard report generation, which is critical for regulatory adherence and continuous monitoring in safety-critical environments.

In summary, the developed model distinguishes itself from previous methods by combining complex scene reasoning with automated report generation, thereby offering a more comprehensive and effective solution for building fire safety compliance monitoring.

4.2 Limitations and Future Directions

Despite the contributions of the developed model, several limitations must be addressed to ensure broader applicability and robustness. A key limitation lies in the dataset used for training and evaluation. The current dataset of 135 image-text pairs, while sufficient for proof-of-concept, is too small and homogeneous to generalize the model's performance across diverse real-world scenarios. Expanding the dataset to include a wider range of building types, operational conditions, and hazard scenarios, particularly edge cases such as rare fire hazards, would significantly enhance the robustness and applicability of the method. Leveraging larger, publicly available datasets or curated collections tailored for fire safety compliance would further strengthen its predictive capabilities.

Another limitation is the lack of real-world testing under dynamic operational conditions. The framework has not yet been deployed in live building environments, such as high-occupancy residential buildings or industrial facilities. A pilot study in these settings would provide crucial insights into the practical performance, scalability, and utility of the model. Real-world testing would also enable the identification of potential challenges, such as adapting to variations in lighting conditions, spatial layouts, or obstruction types, which are critical for ensuring reliable hazard detection in practice.

Additionally, the computational demands of the model, particularly for real-time hazard detection and report generation, warrant further exploration. It is unclear how the system performs under

constrained hardware environments or in challenging conditions, such as low-light settings or poor image quality. Addressing these concerns through hardware optimization or lightweight model adaptations would enhance the framework's usability in resource-constrained scenarios.

The lack of contextual integration also limits the system's ability to fully understand hazards within the broader building-level context. Currently, the developed method focuses on hazard recognition, which, while critical, represents only part of the overall ACC workflow. Hazards such as blocked escape routes or improperly maintained fire doors often require contextual understanding of building design specifications, spatial relationships, and evacuation strategies to assess their true severity. Future work should address these limitations by integrating operational hazards with comprehensive building information, such as BIM or other digital twin technologies. This would enable a transition from isolated hazard detection to end-to-end ACC workflows, offering a more holistic solution for fire safety compliance.

5 CONCLUSIONS

The study presents a novel approach to automated fire safety compliance monitoring for building operational phases, addressing a critical gap in the field of ACC. While previous research has predominantly focused on the design phase, this work emphasizes the importance of continuous monitoring to manage dynamic fire risks during the operational phase. The proposed method leverages advanced computer vision techniques, specifically a vLLM, to assess fire safety compliance in the building operational phase, thus supporting long-term, dynamic fire risk management.

The developed model frames hazard recognition as a VQA task, wherein an image of the building environment is processed along with a text query regarding potential fire hazards. The model effectively identifies hazards, generates descriptive responses aligned with regulatory standards, and provides references that align with OSHA regulations. This dual capability of recognizing hazards and producing detailed, regulatory-referenced reports is essential for enhancing the efficiency of fire safety inspections. By generating automated hazard reports, the model reduces the reliance on human experts to manually document compliance issues, offering a scalable solution for compliance monitoring.

Quantitative evaluation further demonstrates the effectiveness of the proposed method. Significant improvements over a generic vLLM baseline were observed across BLEU and ROUGE metrics, indicating enhanced precision and coherence in hazard descriptions. These metrics underscore the model's ability to capture critical terminology, maintain structured phrasing, and generate accurate, comprehensive responses. Such performance highlights the model's potential to support or even replace expert-driven fire safety compliance inspections.

Nevertheless, limitations remain. The small dataset used for fine-tuning constrains the model's ability to generalize across diverse building types, operational scenarios, and edge cases, such as rare or atypical fire hazards. Expanding the dataset to encompass a broader range of real-world conditions would significantly enhance the robustness and applicability of the model. Additionally, the lack of contextual integration prevents the system from fully understanding hazards within the broader building-level context, such as design specifications, spatial relationships, and evacuation strategies. Currently, the developed method focuses on hazard recognition, which, while critical, is only part of the overall ACC workflow. Future work should address these limitations by expanding the dataset and developing systems that integrate operational hazards with comprehensive building information. Such advancements would enable transition from isolated hazard detection to complete ACC workflows.

Overall, this study contributes an automated and effective solution for fire safety compliance monitoring in operational building environments. By combining complex scene understanding with

automated reporting, the proposed method enables a more efficient approach to managing operational building fire risks, offering building managers and risk inspectors a powerful tool to ensure continuous regulatory compliance and ultimately safeguard human lives and property.

ACKNOWLEDGMENTS

This work was supported by the Willis Towers Watson Research Network TECHNGI-CDT Scholarship.

References

Agrawal, P., Antoniak, S., Hanna, E. B., Chaplot, D., Chudnovsky, J., Garg, S., ... & Wang, T. (2024). Pixtral 12B. arXiv preprint arXiv:2410.07073.

Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015). VQA: Visual question answering. In Proceedings of the IEEE international conference on computer vision (pp. 2425-2433).

Beach, T., & others. (2024). Moving automated compliance checking to the operational phase of the building life-cycle: Analysis and feasibility study in the UK. International Journal of Construction Management, 1–10. https://doi.org/10.1080/15623599.2024.2366727

Bloch, Y., Zhao, R., & Zhang, J. (2023). Graph-based learning for automated code checking – Exploring the application of graph neural networks for design review. Advanced Engineering Informatics, 58, 102137. https://doi.org/10.1016/j.aei.2023.102137

Bosché, F. (2010). Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. Advanced Engineering Informatics, 24(1), 107-118. https://doi.org/10.1016/j.aei.2009.08.006

Chen, D., Chen, L., Zhang, Y., Lin, S., Ye, M., & Sølvsten, S. (2024). Automated fire risk assessment and mitigation in building blueprints using computer vision and deep generative models. Advanced Engineering Informatics, 62, 102614. https://doi.org/10.1016/j.aei.2024.102614

Cheng, J. C. P., Wong, P. K.-Y., Luo, H., Wang, M., & Leung, P. H. (2022). Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. Automation in Construction, 139, 104312. https://doi.org/10.1016/j.autcon.2022.104312

Ding, Y., Liu, M., & Luo, X. (2022). Safety compliance checking of construction behaviors using visual question answering. Automation in Construction, 144, 104580. https://doi.org/10.1016/j.autcon.2022.104580

Gan, Z., Li, L., Li, C., Wang, L., Liu, Z., & Gao, J. (2022). Vision-language pre-training: Basics, recent advances, and future trends. Foundations and Trends® in Computer Graphics and Vision, 14(3–4), 163-352.

Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., & Parikh, D. (2017). Making the V in VQA matter: Elevating the role of image understanding in visual question answering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6904-6913).

Gurari, D., Li, Q., Stangl, A. J., Guo, A., Lin, C., Grauman, K., ... & Bigham, J. P. (2018). Vizwiz grand challenge: Answering visual questions from blind people. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3608-3617).

Jiang, L., Shi, J., & Wang, C. (2022). Multi-ontology fusion and rule development to facilitate automated code compliance checking using BIM and rule-based reasoning. Advanced Engineering Informatics, 51, 101449. https://doi.org/10.1016/j.aei.2021.101449

Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2024). Visual instruction tuning. Advances in neural information processing systems, 36.

Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. In Text summarization branches out (pp. 74-81).

Martins, J. P., & Monteiro, A. (2013). LicA: A BIM based automated code-checking application for water distribution systems. Automation in Construction, 29, 12-23. https://doi.org/10.1016/j.autcon.2012.08.008

Malsane, S., Matthews, J., Lockley, S., Love, P. E. D., & Greenwood, D. (2015). Development of an object model for automated compliance checking. Automation in Construction, 49, 51-58. https://doi.org/10.1016/j.autcon.2014.10.004

Mistral AI team. (2024). Mistral nemo. Retrieved from https://mistral.ai/news/mistral-nemo

Occupational Safety and Health Administration. (n.d.). Maintenance, safeguards, and operational features for exit routes (29 C.F.R. § 1910.37). U.S. Department of Labor.

Occupational Safety and Health Administration. (n.d.). Fire protection (29 C.F.R. § 1926.150).

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: A method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning. Retrieved from https://api.semanticscholar.org/CorpusID:231591445

Teney, D., Anderson, P., He, X., & Van Den Hengel, A. (2018). Tips and tricks for visual question answering: Learnings from the 2017 challenge. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4223-4232).

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models. ArXiv, abs/2302.13971.

Zhang, J., & El-Gohary, N. M. (2017). Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. Automation in Construction, 73, 45-57. https://doi.org/10.1016/j.autcon.2016.08.027

Zhou, Y.-C., Zheng, Z., Lin, J.-R., & Lu, X.-Z. (2022). Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. Computers in Industry, 142, 103746. https://doi.org/10.1016/j.compind.2022.103746

Fitkau, I., & Hartmann, T. (2024). An ontology-based approach of automatic compliance checking for structural fire safety requirements. Advanced Engineering Informatics, 59, 102314. https://doi.org/10.1016/j.aei.2023.102314