2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

VLM-AUGMENTED RAG FRAMEWORK WITH ONTOLOGY-CONSTRAINED LLM REASONING FOR EXISTING BUILDING MATERIAL DOCUMENTATION

Julia Kaltenegger[1], Stavros Nousias[2], Ekaterina Petrova[1], André Borrmann[2] and Pieter Pauwels[1]
[1]Information Systems Built Environment, TU Eindhoven, Eindhoven, Netherlands
[2]Chair of Computing in Civil and Building Engineering, TU Munich, Munich, Germany

Abstract

Documenting information about materials in existing buildings is essential for circular construction and the reuse of elements. While multimodal Vision-Language Models (VLMs) can classify materials from images, they often lack the contextual semantic depth required for technical documentation. This article introduces a VLM-augmented Retrieval-Augmented Generation (RAG) framework that utilises ontology-constrained LLM reasoning to bridge this gap. By integrating domain-specific text retrieval and formal ontologies, the method enables the identification of building element layers and material types. The method is validated using a reference dataset of Dutch residential buildings and demonstrated by enriching semantic graphs with predicted material data.

Introduction

The built environment contributes significantly to resource depletion and waste generation. Addressing these impacts requires prioritising lifecycle material efficiency (reducing waste at the source), reducing the embodied carbon, and enabling reusability and assembly-disassembly models (European Commission, 2020). Consequently, systematic documentation of the existing building stock is essential to support material assessment and reuse, although such documentation remains challenging at scale. Building engineers typically assess materials in existing buildings through on-site inspections, manually documenting building characteristics, element details, and materials using images and text-based data sources (Smeyers and Mertens, 2022). As existing buildings usually lack Building Information Models (BIM) and diverse technical documentation, information is often retrieved from building codes, construction regulations, and archetype descriptions to infer probable construction types (Byers et al., 2024). Archetypes define building types as reference systems characterised by construction period, thermal and energy performance, average quantities of floor area, building-element surface area, and element thickness, thereby supporting top-down building and material stock modelling (Pei et al., 2024; Rijksdienst voor Ondernemend Nederland, 2022). For example, Dutch apartment buildings from the 1980s typically use prefabricated or cast-in-place concrete with low insulation, whereas single-family homes from the same period are predominantly constructed with insulated brick cavity walls. Therefore, the material composition of buildings varies between construction periods and building types.

With the advancement of digital technologies and Artificial Intelligence, existing buildings and their materials are increasingly documented using information-extraction methods such as image recognition, text-based reasoning, and semantic graph representations (Zeng et al., 2024; Pei et al., 2024). As such, computer vision for detecting and classifying building features, facades and materials has moved beyond solely using traditional deep neural networks (Gordon et al., 2023; Marin-Garcia et al., 2023) and has explored the use of Vision-Language Models (VLMs). By jointly learning image and text embeddings, VLMs enable prompt-based descriptive image classification, allowing, for example, the estimation of building age and energy efficiency from images and text (Zeng et al., 2024; Pan et al., 2025). Reasoning tasks based on other text data have shown success when using Large Language Models (LLM) alongside VLMs (Cooper et al., 2025). In this regard, Retrieval-Augmented Generation (RAG) is typically used to integrate domain-specific text corpora, providing natural-language rationales that support LLM-based reasoning (Oche et al., 2025). Methods for generating rationales include, among others, semantically aware sentence parsing (Chen et al., 2019) and studying textual similarities through text matching aligned with ontology definitions (Harispe et al., 2022). Semantic graphs using ontologies support context-based reasoning while integrating and contextualising heterogeneous, ambiguous data (Luo et al., 2023; Zheng et al., 2023). Research in the building materials domain has investigated a hybrid approach that combines text-based and ontology-based similarity matching to enrich material properties relevant to BIM data (Forth et al., 2025). These capabilities, in turn, support VLM-based image classification and LLM-based reasoning (Li et al., 2026; Chen et al., 2023).

This article recognises the potential of the interplay between VLM-based image classification, LLM inferencing, and ontologies to improve the formal description of visual content and enable further interpretation through linkable vocabularies. The aim of this article is to support the documentation of existing building stock using a semantic graph as the backbone, along with image and textual data enrichments. Therefore, this study investigates a VLM-augmented RAG framework for ontology-constrained LLM reasoning to enable the recognition, identification and inference of material information in existing buildings across archetypes . The proposed method

relies on (1) a VLM-based prediction using on-site images of buildings and structuring the corresponding predicted residential building types and material categories in a semantic graph; (2) a VLM-augmented RAG process that incorporates domain-specific knowledge, including construction regulations and archetype definitions; and (3) an ontology-constrained LLM-based reasoning that incorporates the RAG insights and parses the predicted information as semantic graph statements (Fig. 1). Together, these components enable a structured, semantically consistent representation of materials in existing buildings, thereby supporting reuse-oriented assessment.

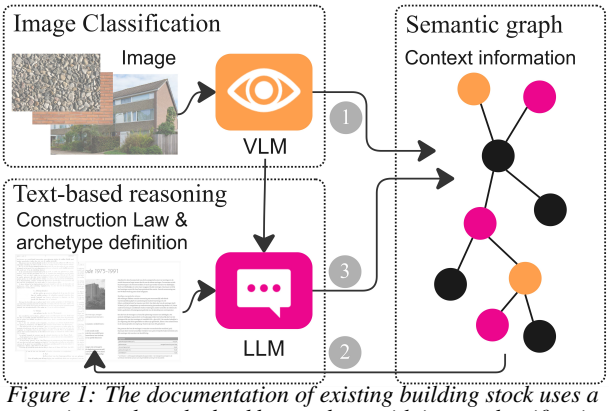

Figure 1: The documentation of existing building stock uses a semantic graph as the backbone, along with image classification and text-based reasoning.

State-of-the-Art

Related works are discussed on the topic of Vision Language Models (VLMs) and semantic-aware text processing in the context of LLM reasoning.

Vision Language models

Computer vision is the fundamental driver for understanding and accessing information from images. Methods such as image classification, semantic segmentation, and object detection are common practices in deep learning, often relying on Convolutional Neural Networks (CNNs). In Architecture, Engineering and Construction (AEC), computer vision for building material recognition supports, e.g., brick inventories by analysing replacement requirements and accounting for deterioration due to efflorescence (Marin-Garcia et al., 2023). Further, automated detection of building elements for deconstruction planning and material reuse has also been studied (Gordon et al., 2023).

Advances in computer vision have led to the development of Vision Language Models (VLMs), which jointly learn from image and text modalities. These generative models embed visual and textual information into a shared representation space, allowing image classification beyond predefined categories. VLMs are typically pre-trained on large-scale image–text pairs using transformer-based architectures. VLMs are based on vision transformers (ViT), using foundation models that transfer knowledge from seen to unseen classes by exploiting cosine similarity between vectors in the shared image–text embedding space (Hassija et al., 2025). Zero-shot VLMs enable object localisation and recognition without labelled target examples by using vision–text transformer architectures with prompt-based image inputs. In contrast, learning with few-shot VLMs predicts classes from a small number of labelled samples by guiding an LLM with limited task examples.

Application-specific studies have explored the use of a zero-shot classification for building-related tasks. For example, Zeng et al. (2024) employ GPT-4 to classify building age from images of buildings in London and predict the age epochs using a structured prompting strategy. Similarly, Pan et al. (2025) utilise GPT-4 as a pre-trained model to infer building energy efficiency levels from descriptive visual attributes, without prior task-specific training on labelled examples. In both cases, custom datasets comprising building images and corresponding labels (e.g., age or energy efficiency) are constructed to provide a baseline for validating the VLM's predictions. As a result, large VLMs can identify previously unseen objects in images using text prompts provided by an operator (Radford et al., 2021). Furthermore, the zero-shot model OWL-Vit (Minderer et al., 2022) is a simple open-vocabulary object-detection model with vision transformers trained on a variety of image-text pairs. It can query an image using one or more text prompts to detect target objects.

Semantically aware text processing for LLM reasoning on top of vision models

VLMs have proven successful in object recognition and classification tasks. However, research shows that using LLMs alongside VLMs can improve VLM performance on tasks that require reasoning from other text data (Cooper et al., 2025). Therefore, text corpora are created that store domain-specific, unstructured datasets, which can be used as natural-language rationale for LLM-based inferencing. The RAG framework is commonly applied in this context as it enables the integration of external data sources into a structured text corpus. Retrieval queries are used to guide the LLM reasoning, thus reducing hallucination (Oche et al., 2025). Methods for generating rationales, as part of the RAG framework, follow prompt engineering, such as Chain of Thought (CoT) (Wei et al., 2022) and semantically aware sentence parsing (Chen et al., 2019).

Semantically aware techniques measure semantic similarity through text matching and ontological alignment to capture conceptual relationships beyond lexical similarity (Harispe et al., 2022). Semantic awareness through text matching has been studied by Zhang et al. (2024). The authors use a RAG framework to enhance LLM-based reasoning on top of VLM image classification. The RAG incorporates context-specific information from an architecture knowledge base, thereby guiding the end user according to the logic of the original document. Semantic awareness through ontology alignments is enabled by Semantic Web technologies. Semantic graphs extend beyond structuring and organisation of domain knowledge

and enable context-based reasoning while ensuring the integration of heterogeneous and ambiguous data (Luo et al., 2023). Recent studies employ semantic graphs to support VLM- and LLM-based inference on construction data. For example, LLM-based question answering enables users to query semantic building data in natural language. Information from semantic graphs is accessed using SPARQL-based RAG queries (Li et al., 2026).

In addition, semantic graph entities and relations have been used as input for zero-shot image classification (Chen et al., 2023). Integrating image and text in semantic graphs is studied in the context of multimodal semantic graphs, which link entities that represent semantically similar concepts (Peng et al., 2023). As such, knowledge transfer strategies (Chen et al., 2023) can enable inference of semantic relationships between detected image classes or allow for semantic graph-based object detection (Chen et al., 2024). For instance, picture units that represent several entities such as buildings, trees, and cars, can be detected as classes and linked to the respective semantic entity in a graph database (Peng et al., 2023). Another approach is semantically aware object detection, proposed by Fang et al. (2017). There, the semantic similarity between two concepts is evaluated to determine the likelihood that they co-occur within the same image. Furthermore, ontological entities can be aligned with image-detected objects, thereby enriching image content with semantically related concepts in the semantic graph (Zeng et al., 2024; Zheng et al., 2023).

Method

In this article, images of buildings, taken on-site, and textual data on building archetypes and construction regulations documents are used to extract information about building element layers and possible materials. The paper proposes an image-text information generation process using a VLM-augmented RAG framework for ontology-constrained LLM reasoning, as shown in Fig. 2. A semantic graph is deployed as the backbone to integrate image-text data. The graph uses the Archetype ontology (at:)[1] and the Building Material Performance (bmp:) ontology[2], which describe the semantics of buildings, their elements, possible layer sets, and associated materials. The method is tested and validated on a ground-truth dataset comprising 75 Dutch residential buildings. The proposed method includes:

1. A zero-shot VLM model to classify building images by residential types and material categories.
2. A RAG process that integrates construction regulation and archetype definition.
3. A LLM-based reasoning process that uses the RAG results, Contrastive Language-Image Pre-training (CLIP) predictions, constrained to the ontology vocabulary. The results of the LLM reasoning are then parsed as Resource Description Framework (RDF) triples to enrich the semantic graph.

Zero-shot VLM model for image classification

Contrastive Language-Image Pre-training (CLIP) is used as it is a powerful foundation for zero-shot and few-shot classification. The CLIP framework utilises a dual-encoder architecture to bridge visual and textual data. Unlike traditional classifiers, CLIP is pre-trained on a massive dataset of image-text pairs to learn a shared vector space where related visual and textual concepts are mapped closely together. This article uses 'openai/clip-vit-base-patch32'[3], with an underlying architecture relying on the ViT-G/14 to encode images and a Transformer to encode text. CLIP predicts the residential building type y and the material category z with the function f that uses the input image x and visual prompts P_v to describe context and possible prediction classes in natural language, as shown in Eq. 1.

$$(y,z) = f(x;P_v)(x \in X) \tag{1}$$

The classification tasks for residential building types and material categories follow the formal semantics of the reused ontologies, the AT ontology[1] and the BMP ontology[2]. Prompts are used to introduce the ontology terms and definitions in natural language. Parsing the ontology solely as an RDF file, without descriptions, would result in VLM hallucinations. Therefore, definitions for building types 'Apartment House', 'Single Family House', and 'Terrace House' and for material categories, including 'Brick', 'Concrete', 'Glass', 'Natural stone', and 'Wood', are provided. Although the material categories can provide an estimate of the material origin, the classification of layer-specific material types depends on building-specific contexts, such as construction methods that vary by building type, construction year and country.

Retrieval augmentation process

The RAG pipeline comprises three steps: (2.1) building the text corpus; (2.2) constructing a retrieval query that uses the CLIP predictions and metadata of a building (construction year and country) as input; and (2.3) using textual similarity between the text corpus and the query to extract the most relevant text that the LLM can later use to support the reasoning process (Fig. 3). The text corpus includes building archetype descriptions (Rijksdienst voor Ondernemend Nederland, 2022) and building construction regulations, including building permit codes from 1904 (Schelling and van Gijn, 1904) and 1927 (Peerbolte and der Kaa, 1927). The original data sources are provided in PDF format as embedded bitmaps. Therefore, the optical character recognition (OCR) converter *pytesseract* is used to convert text to text format. The OCR extracts text sections based on the logic of the original document, depicted in Fig. 4. For example, the extracted text can

[1] https://w3id.org/at

[2] https://w3id.org/bmp

[3] https://github.com/openai/CLIP

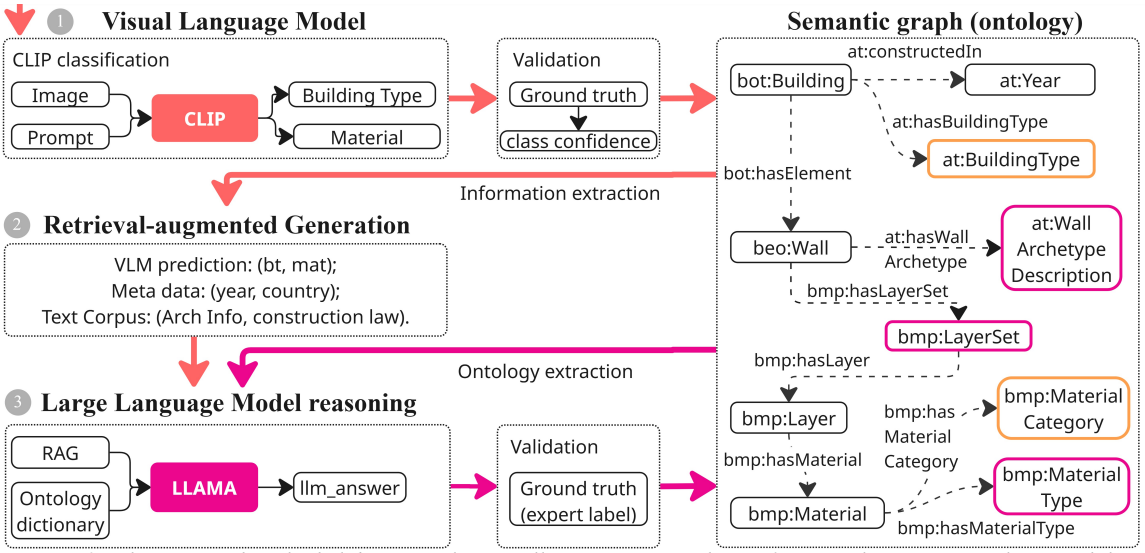

Figure 2: The proposed method elaborating the overall VLM-augmented RAG framework (in orange colour) and the ontology-constrained LLM reasoning (in pink colour).

be structured as a CSV table with columns for 'Building Type', 'Construction Year', 'Archetype Description', and the associated 'Wall Archetype Construction'. Each row of the CSV is extracted and converted to a vector as part of C, the 'Corpus embedding'. The text corpus is processed with the *SentenceTransformer* Python library to provide text embeddings.

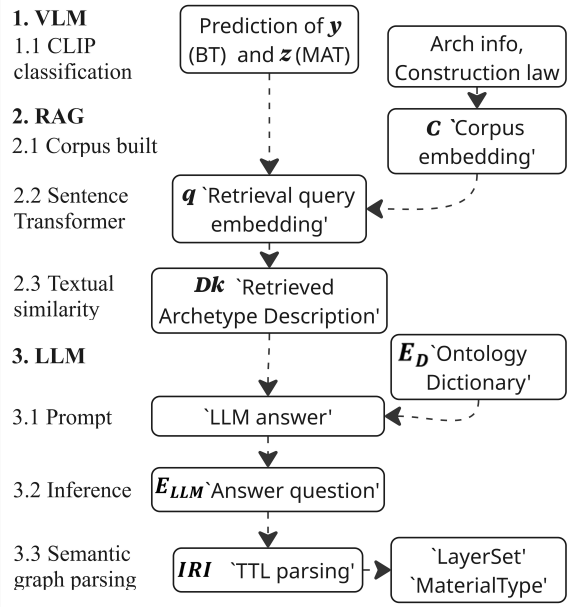

Figure 3: Sequential process of functions that enable VLM informed RAG for ontology-constrained LLM process

During inference, the function 'Retrieval query embedding' builds a query q that fetches the CLIP prediction results for building type y and material category z, and the metadata m (construction year and country) as input, and converts the results to a vector (Eq. 2). The function 'Retrieve archetype description' then extracts the embedded corpus C and the embedded query q, and computes cosine similarity between the two, denoted as Dk (Eq. 3). Dk represents the top-k largest elements of the input tensor (entries that best match the retrieval query with the highest similarity score). The cosine similarity between two vectors a and b is defined in Eq. 4.

$$q = g(y, z, m) \tag{2}$$

$$D_k = Retrieve(C, q) \tag{3}$$

$$S_C(a,b) = \frac{a * b}{||a|| ||b||} \tag{4}$$

The RAG pipeline provides the LLM with the building context information. By applying textual semantic similarity matching against the ontology definitions, the system identifies the most probable entities and integrates them into the semantic graph (Fig. 3).

LLM reasoning and ontology constraining

In the third part of the procedure (Fig.3), the LLM reasoning process, comprises three steps: (3.1) LLM answer template and prompt design; (3.2) LLM model inference of the 'LayerSet' and the 'Material Type'; and (3.3) parsing the LLM results in RDF Triples (TTL). First, the function 'LLM answer' is based on an LLM answering template in JSON notation and is constructed as the LLM prompt input. The prompt input relies on the CLIP prediction results and confidence for the 'building type' and 'material category', the metadata 'construction year' and 'country', the 'retrieved corpus text' from the RAG results and ontology embeddings 'ontology guidance'. The latter contains descriptions of ontology classes (bmp:MaterialType and bmp:LayerSet) represented as vectors. Further, the prompt describes the structure of the TTL file; an excerpt of the input prompt is shown in Listing 1.

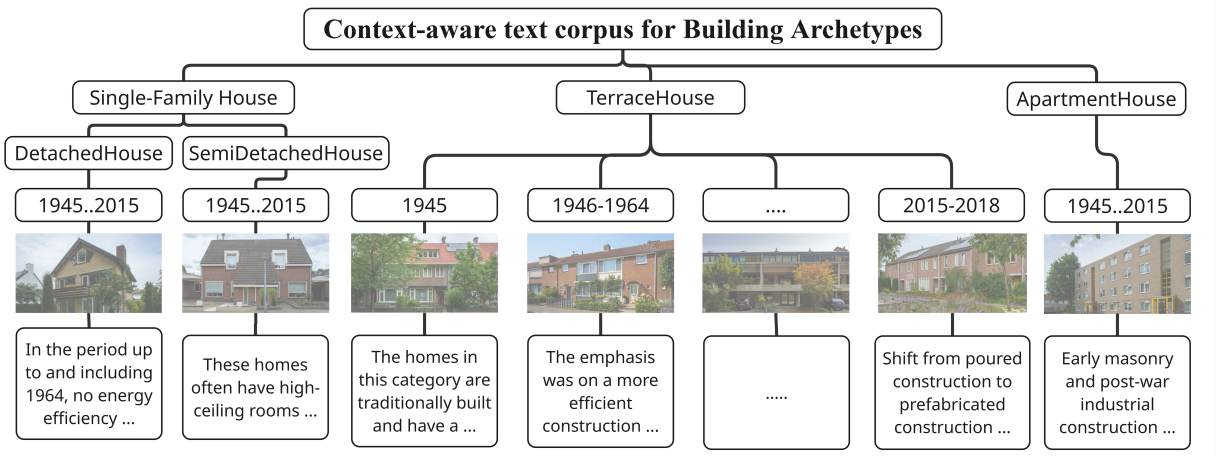

Figure 4: Archetype information structure as preparation for the RAG corpus embedding

Listing 1: LLM Input Prompt for TTL Description

```
# INPUT EVIDENCE:
Metadata Year: {year}
Metadata Region: {region}
Visual Prediction (Type): {bt_pred} (conf: {bt_conf:.2f})
Visual Prediction (Material Category): {mat_cat_pred} (conf: {mat_cat_conf:.2f})
# ARCHETYPE CORPUS EVIDENCE:
{retrieved_corpus_text}
# ONTOLOGY
{ontology_guidance}
# TASK:
# Generate RDF/TTL output
# for the building using these rules and
# proposed ttl layout:
inst:building01
  at:BuildingType [at:hasValue
    {building_types}; {bt_confidence}];
  at:ConstructedIn {year} ;
  at:Country {country} .
inst:wall01
  rdf:type beo:Wall ;
  bmp:hasLayerSet{context['ontology']
    ['wall_layer_option']} ;
  bmp:hasOuterLayer inst:Layer01 ;
  bmp:WallArchetypeDescription
    {retrieve_archdescription} .
inst:Layer01
  rdf:type bmp:Layer ;
  bmp:hasMaterialCategory [at:hasValue
    {material_cat}; {mat_confidence}]; ;
  bmp:hasMaterialType {context['ontology']
    ['material_type']}
```

Next, the 'Answer question' function relies on the context information and the prompt; the LLM model infers the appropriate 'LayerSet' and, if applicable, the 'Material Type' for building archetypes. In the experiments, the LLM model 'llama3:8b' has been used; however, the methods can be applied with any LLM. The 'LayerSet' and 'MaterialType' are selected based on the highest cosine similarity between the retrieved text-based 'Wall description' and the ontology embedding. The final ontological grounding is achieved through a maximum-likelihood selection over the embedding space interpreted by Eq. 5. In the function 'TTL parsing', the IRI is the assigned OWL syntax based on cosine similarity between the embedded LLM results E_{LLM} and the embedded ontology class description E_D. D is the ontology description as a subset of the ontology classes O_E. A fallback logic is implemented, which, if the evidence (the provided text corpus) does not match an ontology-based IRI, causes the LLM to provide a descriptive string instead of generating a new ontology entity. As a result, the LLM writes a semantic graph in Terse RDF Triple Language (TTL) format that assigns the 'LayerSet' with the associated 'MaterialType' to the observed component. The full code and datasets are available on GitHub [3].

$$IRI = \underset{D \in \mathcal{O}\mathcal{E}}{\operatorname{argmax}} \left(\frac{E_{LLM} \cdot E_D}{\|E_{LLM}\| \cdot \|E_D\|} \right) \tag{5}$$

Validation corpus

For validation, a dataset is constructed that contains images and corresponding labels describing 75 Dutch residential buildings by country, construction year, building type, material types, and definitions of wall-layer sets. The data set is manually labelled in collaboration with domain experts; see an example entry in Listing 2. The validation process validates the zero-shot VLM predictions and second, the LLM reasoning using Precision, Recall and F1-Score.

Listing 2: JSON validation dataset

```
"buildingImage":"1965-SFH.png",
"materialImage":"1965-SFH-mat.jpg",
"yearFrom":"1965",
"yearTo":"1974",
"Country":"Netherlands",
"BuildingType":"Single Family House",
"FacadeDescription":"Outer layer is sand
    limestone, in-between layer is insulated
    with eps, inside layer is brick masonry. ",
"materialCategory": "Brick",
"materialType": "Sand limestone",
"LayerSet": "Cavity Wall"
```

Results and validation

The results discuss zero-shot VLM classification, LLM reasoning on the validation dataset, and finally demonstrate

[3] https://github.com/JulKaltenegger/VLM-RAG-LLM-reasoning

the proposed methods on a use case.

Zero-shot VLM classification validation

The zero-shot VLM classification validates the CLIP-based predictions for building type and material category. For each image in the validation dataset, the CLIP generates top-3 predictions. Both classes are directly compared against the corresponding labels in the JSON validation dataset. The 'Single-family Housing' class exhibits the lowest stand-alone CLIP performance, largely due to its typological similarity to 'Terrace Houses'. In terms of material classification, CLIP achieves high precision (0.98) for 'Brick' but fails on 'Concrete' with a precision of only 0.33. This indicates a high rate of false positives, with the model correctly identifying concrete only one out of three times. This type of hallucination is likely caused by visual similarities with cement stucco or large-format stone elements. The confusion matrix for the CLIP prediction for building types is represented in Fig. 5. The precision, recall and the F1 score are presented in Tables 1 and 2.

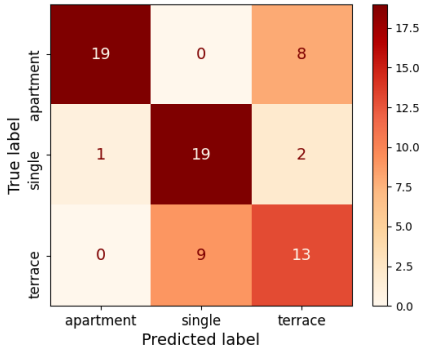

Figure 5: Confusion Matrix Clip Building Type classification

Table 1: Building type classification (CLIP-only)

Class	P	R	F1	Sup.
Apartment house	0.96	0.81	0.88	27
Single family house	0.53	0.83	0.65	12
Terrace house	0.63	0.55	0.59	22
Accuracy		0.72		61
Macro avg	0.70	0.73	0.70	61
Weighted avg	0.75	0.72	0.73	61

Table 2: Material category classification (CLIP-only)

Class	P	R	F1	Sup.
Brick	0.98	0.81	0.88	52
Concrete	0.33	0.86	0.48	7
Accuracy		0.79		61
Macro avg	0.44	0.55	0.45	61
Weighted avg	0.87	0.79	0.81	61

Detecting building type solely from facade images is often ambiguous, motivating the integration of contextual metadata and ontology-constrained reasoning in subsequent pipeline stages.

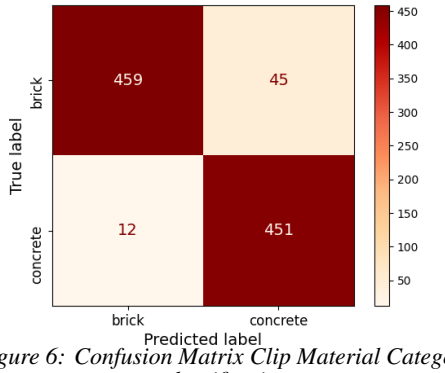

Figure 6: Confusion Matrix Clip Material Category classification

LLM reasoning validation

The VLM-augmented RAG and ontology-constrained LLM pipeline is validated on the validation dataset. Fig. 7 shows the validation performance for each image. This includes the CLIP predictions and context data (construction year and country), which are used to extract archetype descriptions from the text corpus. These retrieved descriptions are then used by the LLM to infer wall construction characteristics and to generate RDF/Turtle statements constrained by the ontology. The resulting RDF triples are parsed and compared against the JSON validation dataset. The cosine similarity is computed between the inferred information (the 'LayerSet' and the 'MaterialType') and the labelled JSON dataset, resulting in average similarity and confidence scores of 0.45 and 0.56, respectively.

Demonstration use case

The use case demonstrates how the proposed method can help construction engineers with large-scale building documentation, particularly in detecting wall layer sets and associated material definitions. A Dutch residential building, a terraced house from 1965, is selected to illustrate how insights from the reasoning process can enrich a semantic graph with entities and their associated confidence values (Fig. 8). A semantic graph is used to describe a building instance inst:Building001, and its associated elements, layer sets and materials. Fist, the semantic graph is enriched by the detected building type and material category, derived from the VLM model. The CLIP classification results with a confidence of 0.87 for 'Building type' at:TerraceHouse, and 0.98 for 'Material Category' bmp:Brick of the observed wall. Furthermore, the LLM's reasoning extracts the archetype description for each building element, in this case, the wall archetype. Furthermore, the LLM reasons that the 'WallLayerSet' and the outer layer is assigned by the 'MaterialType', with a confidence of 0.59 for bmp:CavityWall and 0.34 for bmp:ClayUnit.

This enrichment process relies on text-based documentation of building archetypes and ontology classes, including descriptive annotations. The confidence of the textual semantic matching depends on the degree of align-

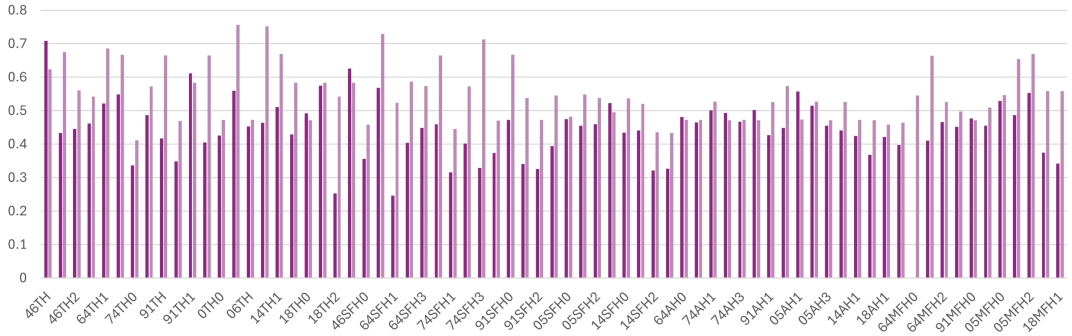

Figure 7: Validation of the complete pipeline with the validation dataset (N=75) showing the semantic similarity (dark purple) and the confidence score (light purple)

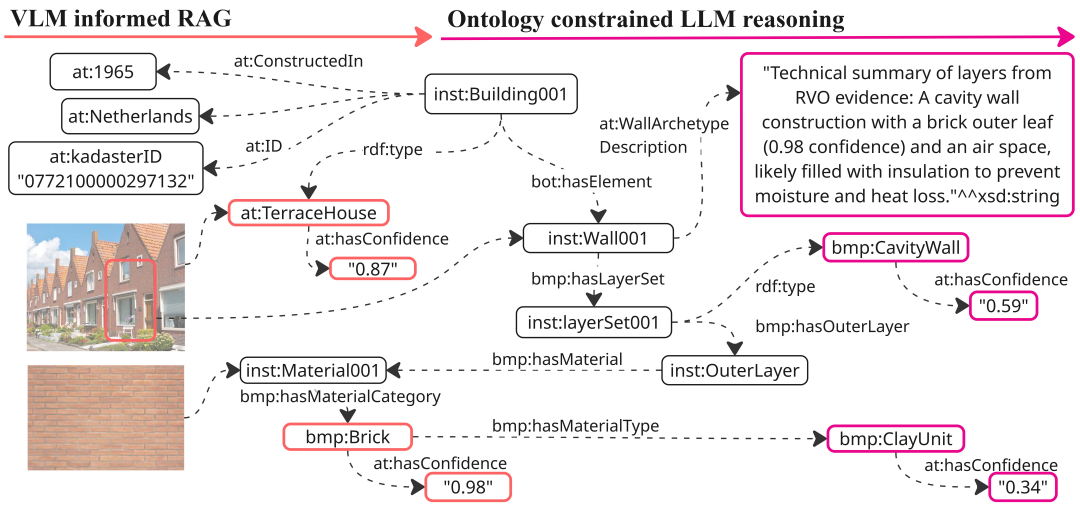

Figure 8: Results representing the enriched semantic graph with the deducted information from the VLM-augmented RAG framework (in orange colour) and the ontology-constrained LLM reasoning (in pink colour)

ment between these sources. The greater the alignment of sources in their descriptive annotations, the greater the likelihood of identifying meaningful semantic textual similarity and correctly assigning ontology classes. A key limitation of the proposed method is that the archetype descriptions require partial manual adjustment and verification to ensure terminological consistency with the ontology, which directly influences the cosine similarity scores in the embedding-based matching process. Therefore, a human-in-the-loop approach is strongly recommended, allowing building engineering experts to iteratively validate and refine LLM-generated results to ensure domain accuracy and methodological robustness.

Conclusion

Effective material-related reuse decisions in existing buildings depend on robust and well-structured documentation, which remains a challenge in current practice. Information retrieval technologies, including VLMs, LLMs, and semantic graphs, have emerged as viable approaches to modelling material information. In this context, the article explores a VLM-augmented RAG framework and ontology-constrained LLM reasoning to integrate unstructured text from building documents. The results show that the proposed VLM-informed RAG pipeline successfully extracts building-archetype text data, whereas ontology-constrained LLM reasoning yields only moderate performance. Therefore, future work will investigate advanced ontology embedding methods to fine-tune LLM reasoning.

References

Byers, B. S., Raghu, D., Olumo, A., De Wolf, C., and Haas, C. (2024). From research to practice: A review on technologies for addressing the information gap for building material reuse in circular construction. Sustainable Production and Consumption, 45:177–191.

Chen, J., Geng, Y., Chen, Z., Pan, J. Z., He, Y., Zhang, W., Horrocks, I., and Chen, H. (2023). Zero-shot and few-shot learning with knowledge graphs: A comprehensive survey. Proceedings of the IEEE, 111(6):653–685.

Chen, X., Liang, C., Yu, A. W., Zhou, D., Song, D., and Le, Q. V. (2019). Neural symbolic reader: Scalable integration of distributed and symbolic representations for reading comprehension. In International Conference on Learning Representations.

Chen, Z., Zhang, Y., Fang, Y., Geng, Y., Guo, L., Chen, X., Li, Q., Zhang, W., Chen, J., Zhu, Y., et al. (2024). Knowledge graphs meet multi-modal learning: A com-

prehensive survey. arXiv preprint arXiv:2402.05391.

Cooper, A., Kato, K., Shih, C.-H., Yamane, H., Vinken, K., Takemoto, K., Sunagawa, T., Yeh, H.-W., Yamanaka, J., Mason, I., et al. (2025). Rethinking vlms and llms for image classification. Scientific Reports, 15(1):19692.

European Commission (2020). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions: A new circular economy action plan — for a cleaner and more competitive europe. Communication, COM(2020) 98 final CELEX: 52020DC0098, European Commission, Brussels.

Fang, Y., Kuan, K., Lin, J., Tan, C., and Chandrasekhar, V. (2017). Object detection meets knowledge graphs. International Joint Conferences on Artificial Intelligence.

Forth, K., Kaltenegger, J., Petrova, E., and De Wolf, C. (2025). Bim-based material property enrichment using text-based and ontology-based semantic similarity matching.

Gordon, M., Batallé, A., De Wolf, C., Sollazzo, A., Dubor, A., and Wang, T. (2023). Automating building element detection for deconstruction planning and material reuse: a case study. Automation in Construction, 146:104697.

Harispe, S., Ranwez, S., Montmain, J., et al. (2022). Semantic similarity from natural language and ontology analysis. Springer Nature.

Hassija, V., Palanisamy, B., Chatterjee, A., Mandal, A., Chakraborty, D., Pandey, A., Chalapathi, G., and Kumar, D. (2025). Transformers for vision: A survey on innovative methods for computer vision. IEEE Access.

Li, M., Hu, Z., Mohebi, P., Li, S., and Wang, Z. (2026). Enhancing llm-based building data query with chain-of-thought, retrieval-augmented generation, and fine-tuning. Automation in Construction, 182:106738.

Luo, L., Li, Y.-F., Haffari, G., and Pan, S. (2023). Reasoning on graphs: Faithful and interpretable large language model reasoning. arXiv preprint arXiv:2310.01061.

Marin-Garcia, D., Bienvenido-Huertas, D., Carretero-Ayuso, M. J., and Della Torre, S. (2023). Deep learning model for automated detection of efflorescence and its possible treatment in images of brick facades. Automation in Construction, 145:104658.

Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z., et al. (2022). Simple open-vocabulary object detection. In European conference on computer vision, pages 728–755. Springer.

Oche, A. J., Folashade, A. G., Ghosal, T., and Biswas, A. (2025). A systematic review of key retrieval-augmented generation (rag) systems: Progress, gaps, and future directions. arXiv preprint arXiv:2507.18910.

Pan, Y., Lu, L., Wang, M., Reja, V. K., Noichl, F., Bormann, A., and Brilakis, I. (2025). Zero-shot learning with vision-language models for estimating building energy efficiency from street view images. In Computing in Civil Engineering 2025: Resilient, Robotic, and Educational Systems, pages 254–262. n.a.

Peerbolte, L. and der Kaa, V. (1927). Leidraad bij het samenstellen of herzien van bouwverordeningen. N. Samsom, Alphen aan den Rijn.

Pei, W., Biljecki, F., and Stouffs, R. (2024). Techniques and tools for integrating building material stock analysis and life cycle assessment at the urban scale: A systematic literature review. Building and Environment, 262:111741.

Peng, J., Hu, X., Huang, W., and Yang, J. (2023). What is a multi-modal knowledge graph: A survey. Big Data Research, 32:100380.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In International conference on machine learning, pages 8748–8763. PmLR.

Rijksdienst voor Ondernemend Nederland (2022). Voorbeeldwoningen 2022 bestaande bouw. Technical report, Rijksdienst voor Ondernemend Nederland (RVO).

Schelling, B. and van Gijn, A. (1904). Leidraad bij het samenstellen van eene Verordening, ter uitvoering van artikel 1 de Woningwet. n.p., 3 edition.

Smeyers, T. and Mertens, M. (2022). Reuse toolkit: The reclamation audit. Technical report, Facilitating the Circulation of Reclaimed Building Elements.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35:24824–24837.

Zeng, Z., Goo, J. M., Wang, X., Chi, B., Wang, M., and Boehm, J. (2024). Zero-shot building age classification from facade image using gpt-4. arXiv preprint arXiv:2404.09921.

Zhang, J., Xiang, R., Kuang, Z., Wang, B., and Li, Y. (2024). Archgpt: harnessing large language models for supporting renovation and conservation of traditional architectural heritage. Heritage Science, 12(1):220.

Zheng, Y., Khalid Masood, M., Seppänen, O., Törmä, S., and Aikala, A. (2023). Ontology-based semantic construction image interpretation. Buildings, 13(11):2812.