

Ontology-Based Semantic Construction Image Interpretation

Yuan Zheng ^{1,*}, Mustafa Khalid Masood ^{1,2} , Olli Seppänen ¹ , Seppo Törmä ^{3,4} and Antti Aikala ¹

- ¹ Department of Civil Engineering, Aalto University, 02150 Espoo, Finland; mustafa.khalid@siloi.ai (M.K.M.); olli.seppanen@aalto.fi (O.S.); antti.aikala@aalto.fi (A.A.)
- ² Silo AI, 00180 Helsinki, Finland
- ³ School of Real Estate and Construction, Metropolia University of Applied Sciences, 00920 Helsinki, Finland; seppo.torma@metropolia.fi
- ⁴ VisuaLynk Oy, 02150 Espoo, Finland
- * Correspondence: yuan.zheng@aalto.fi

Abstract: Image-based techniques have become integral to the construction sector, aiding in project planning, progress monitoring, quality control, and documentation. In this paper, we address two key challenges that limit our ability to fully exploit the potential of images. The first is the “semantic gap” between low-level image features and high-level semantic descriptions. The second is the lack of principled integration between images and other digital systems used in construction, such as construction schedules and building information modeling (BIM). These challenges make it difficult to effectively incorporate images into digital twins of construction (DTC), a critical concept that addresses the construction industry’s need for more efficient project management and decision-making. To address these challenges, we first propose an ontology-based construction image interpretation (CII) framework to formalize the interpretation and integration workflow. Then, the DiCon-SII ontology is developed to provide a formalized vocabulary for visual construction contents and features. DiCon-SII also acts as a bridge between images and other digital systems to help construct an image-involved DTC. To evaluate the practical application of DiCon-SII and CII in supporting construction management tasks and as a precursor to DTC, we conducted a case study involving drywall installation. Via this case study, we demonstrate how the proposed methods can be used to infer the operational stage of a construction process, estimate labor productivity, and retrieve specific images based on user queries.



Citation: Zheng, Y.; Khalid Masood, M.; Seppänen, O.; Törmä, S.; Aikala, A. Ontology-Based Semantic Construction Image Interpretation. *Buildings* **2023**, *13*, 2812. <https://doi.org/10.3390/buildings13112812>

Academic Editor: Irem Dikmen

Received: 18 September 2023

Revised: 26 October 2023

Accepted: 7 November 2023

Published: 9 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: ontology; semantic image interpretation (SII); construction; digital twin construction (DTC)

1. Introduction

Image-based technologies are increasingly used at construction sites due to their affordability and ability to capture construction process scenes, thus supporting onsite management and the situational awareness of all stakeholders [1,2]. The use of image-based approaches in the construction domain has been explored for various applications, including resource tracking [3], workspace assessment [4], operation process monitoring [5], and productivity evaluation [6].

A key challenge with image-based approaches is interpreting image and video data effectively, such as by inferring the states, progress, or stages of the construction process. Traditionally, this requires experienced labor to manually analyze and interpret the image scenes. To automate this, a machine has to bridge the “semantic gap”, which refers to the challenge of linking low-level image features (such as entities’ visual features, including edges, textures, colors, basic shapes, and relationships with other entities) to high-level concepts (the meaningful information behind the image scene, such as “this is a drywall”, “the drywall is in the framing stage”, etc.). Modern computer vision techniques have made significant advancements in bridging the semantic gap by excelling at tasks like image classification [3], object detection [7,8], and image captioning [9,10], but it is now

increasingly being accepted that correspondence with background knowledge is essential for an effective scene understanding system. When mapped to knowledge of construction processes in general and the specifics of a construction project in particular, CV-based interpretations of image scenes can become more reliable and actionable. Systems that achieve this mapping are rare in the construction domain.

Another challenge with image-based approaches is that they lack interoperability with other construction information systems [11]. As the adoption of the digital twin construction (DTC) concept increases, images could serve as an important element of DTC [11,12]. The strength of images reflects the real-time or semi-real-time progress of physical/actual objects captured in images. However, there are other information communication technology (ICT) systems that could provide enriched contextual digitalized content about the construction process that cannot be obtained from the construction images. These systems include construction schedules, building information modeling (BIM) [13], sensors and the internet of things (IoT) [14,15], and indoor positioning systems (IPSs) [16,17]. These systems are essential elements of a DTC and complement images to comprehensively reflect the actual construction process. For example, the construction schedule could provide the as-planned task information. BIM could provide detailed product information about the building object, while IoT systems could provide the measurement of condition properties, like temperature and relative humidity. Thus, to enable a DTC that involves an image, a principled method of integrating image information with that from other ICT systems is required [18,19].

To address the challenges of *interpretation* and *integration* outlined above, we propose to cast the problem of understanding a construction scene as a semantic image interpretation (SII) [20] task to be achieved in conjunction with a construction ontology. The goal of SII is to generate a semantically rich structured description of an image that describes the objects in the image, along with further information about their types, attributes, and the relations between other objects, thus enabling further image interpretation based on background knowledge [20,21]. The knowledge is contained in an ontology, which provides a formal vocabulary to describe image contents and implicit knowledge to infer higher-level image semantics [20,22]. Once inferred, the image semantics can be integrated with data from other ICT systems.

Recently, a few studies in the construction domain have explored the use of ontologies to interpret image semantics [9,23]. These studies successfully presented specific solutions that use image semantics to address the theme of construction safety to investigate job hazards. However, they only provided a partially structured description of image scene contents, which is insufficient to develop holistic SII. Meanwhile, as SII is a goal-oriented task, the results vary based on the goals of interpretation [24]. To the best of our knowledge, no previous work presents a systematic framework for building SII in the construction domain. These previous works also did not attempt to link image semantics to other digital data sources, thereby narrowing their usability from the point of view of DTC, which requires data fusion. Thus, a generic ontology that can holistically represent and integrate construction image semantics with other data sources is still missing.

Firstly, we propose ontology-based construction SII (CII), a conceptual framework to describe the workflow of interpretation and integration to establish a DTC that involves image semantics. Secondly, based on the requirements of CII, we develop an ontology called Digital Construction Ontology of Semantic Image Interpretation (DiCon-SII). DiCon-SII is an extension of our previous Digital Construction Ontologies (DiCon) [25]. The purpose of developing DiCon-SII is two-fold. First, DiCon-SII aims to provide a generic and formalized vocabulary to specify the objects and features in a construction image for semantically labeled construction images. This ontology could thus trigger further image interpretation by using feature-based inference with the semantic rules defined based on background knowledge to interpret higher-level image semantics. Second, DiCon-SII fills the missing links between construction entities and image contents in DiCon to enable the integration of construction image semantics with other digital information sources. To

demonstrate the use of the proposed framework and ontology, an example case of drywall installation was tested with real project data.

The rest of the paper is organized as follows. Section 2 reviews related works on images in construction, SII, and associated ontologies. Section 3 outlines the framework of CII. In Section 4, the development of DiCon-SII and the ontology itself is illustrated. In Section 5, a case study of implementing CII and DiCon-SII is demonstrated. In Section 6, the discussion of the contributions and limitations of this work are presented, and suggestions for future research are offered. The conclusion of the paper is given in the final section.

2. Background

This section reviews the existing research on image-related works, SII, and related ontology and DTC research to provide a background for the development of DiCon-SII and CII.

2.1. Image-Related Works in the Construction Domain

Image analysis could offer a more direct and consistent way to obtain more in-depth information conveyed by images [26]. Over the past few decades, there has been a growing research trend focused on the implementation of image-related technology in the construction domain. A major focus of this research has been detecting onsite objects, including workers, equipment, and materials for site monitoring [27]. Chi and Caldis [8] proposed a method of using a video camera to automatically identify both onsite personnel and equipment. Park and Brilakis [28] presented a method for detecting construction workers in video frames. Son et al. [29] introduced a vision-based collision warning system based on the automated 3D position estimation of each worker to protect workers from potentially dangerous situations. Some works have focused on the detection of onsite equipment [30–32]. Others have focused on the detection of material and construction progress using CV-based technologies, including Liu et al. [32] and Kim et al. [33]. However, these research efforts did not attempt to acquire detailed semantic information about detected objects, especially their visible attributes.

Additionally, several past works have focused on generating additional semantic features of the detected objects, performing detailed labeling beyond just detecting objects in the scene, with more detailed attributes and the relations among them. Dimitrov and Golparvar-Fard [34] and Han and Golparvar-Fard [35] presented a series of approaches using CV techniques to analyze image sources by monitoring progress and attribute changes in building elements to classify the material. Yang [1] used the interpreted image to check appearance changes for the purpose of indicating the progress and performance of construction. Hamledari [36] extracted the attributes of interior building objects to the reasoning process state, although no formalized structure of image semantics was given. These works explained how construction image contents should be semantically represented based on background domain knowledge. Although the works extracted the image semantics, they did not generalize its structure. Moreover, only the attribute of appearance was considered. Thus, they still lacked a comprehensive semantic representation of the image contents, leading to difficulty in achieving further semantic-based image inference and other utilizations of image semantics.

Many past works employ structured descriptions of image contents to infer and extract tacit information from the image. Liu et al. [10] introduced a method of manifesting construction activity scenes via image captioning by employing CV and natural language generation technologies to extract the semantics of the image scene. In their research, they classified the main scene element, developed a linguistic description schema, and used DL to analyze image contents and generate descriptive captions of the scenes from image contents and their features. Fang et al. [9] presented a knowledge graph-based approach to using an ontological model to identify job hazards via image data. In this research, they developed an ontological model to indicate four core entity classes that can be captured by construction images, including workers, equipment, environments, and materials. This

model can also detect the spatial relationships between different objects. Similarly, Zhong et al. [23] addressed the approach of using an ontology model to represent the potential hazard implied in construction images. The construction images were first semantically annotated, after which the ontology-based inference approach was employed to identify the potential hazard. In their model, they defined worker, equipment, material, and environment as the main classes of the image content objects, together with four attributes: feature, state, measurement, and appearance. These works provide examples of developing structured construction image semantics and further inference applications, but their applications were confined to specific themes. Therefore, they did not provide a generic framework for image interpretation, and no comprehensive image semantic taxonomy or ontology was given. The above works are point solutions that do not integrate image semantics with external data and thus have limited applicability for using the image as a data source for creating the DTC of the process.

2.2. Semantic Image Interpretation (SII)

SII is the task of generating a structured description of the content of images [20]. SII aims to describe more details from the image, including the objects in images and their types, attributes, and relations with a formalized structure [20], and to then conduct the interpretation of the higher-level semantics of the images, such as the context of the image, the theme image is representing, and other tacit information that can be inferred based on the image contents. However, the semantic gap is a major challenge when developing the SII system and involves the lack of direct correspondence between low-level features, including the objects and features in the images, and higher-level semantic information [20,37].

To resolve the semantic gap in the SII structure, background knowledge and semantically labeled images are regarded as two mandatory inputs [20]. Background knowledge refers to the knowledge and logic about the image context and content. A semantically labeled image is an image that has been labeled in detail with its content objects and their features with a structured semantic description. For example, in a semantically labeled image of a construction scene, different regions of the image might be labeled as “building element”, “material”, or “worker”, with their features and relationships also described. This type of labeling provides a more nuanced understanding of the image content, enabling computers to comprehend and interpret scenes in a way that is closer to human perception. The background knowledge can be correlated to the features represented in the semantically labeled image to infer implicit higher-level semantic information from low-level features of the image. For this, challenging requirements are acquiring and representing prior knowledge and developing a structured representation of the content of the image scene [38].

To encode the background knowledge and formalize the semantically labeled image in both a human- and machine-understandable format, Donadello [38] suggested the use of semantic web ontologies. These ontologies represent knowledge in specific domains, with a formal description of their concepts and relationships. Therefore, they can first create a representation of semantically labeled images by representing the low-level features of image contents based on domain knowledge with a formal, explicit structure. Second, they can provide a formalized vocabulary of concepts with explicit definitions to develop rule-represented knowledge that can enable inference from low-level features to obtain higher-level information about the image scene [39].

2.3. Ontology

2.3.1. Ontology and Digital Twin Construction

There is increasing implementation of semantic web ontologies in the construction domain. Ontology is “an explicit formal specification of a conceptualization” [40], in which the domain knowledge is modeled by Description Logic as concepts, properties,

and the interrelationships between the concepts. Thus, it could represent the structure of information and domain knowledge [41].

Besides providing formal domain knowledge representation, another use of ontologies in the construction domain is to address the problem of data integration [42–44]. Information heterogeneity remains a common domain issue [45] and has huge effects on establishing the digital twin construction (DTC). A DTC refers to a virtual digital replica that integrates various data sources, encompassing information about the construction process [11]. DTC serves as a valuable tool for enhancing construction and operational management. It empowers stakeholders by providing them with integrated insights into the current status, thereby facilitating informed decision-making. The use of ontologies allows the interrelation of different kinds of information [46]. An image can be represented in machine-processable semantics using an ontology, which could enable integration with other data to develop the DTC [12,19]. To realize these objectives, an ontology must describe the image contents in a formalized manner. In the following section, image-related ontology work is reviewed.

2.3.2. Image-Related Ontology Works

Currently, numerous ontologies that are related to image semantics have been developed. In this study, these ontologies are reviewed to explore how they can solve the semantic interpretation problem in the construction domain. The reviewed ontologies are classified into two types: (1) generic ontologies to describe the visual contents from multimedia and (2) other relevant ontologies in the construction domain. Table 1 summarizes the related ontologies.

Table 1. Summary of related ontologies.

Ontologies	Classes of Image Content	Features and Relations of Image Content	Missing Aspects
Multimedia Analysis Ontology [47]	Object, feature, feature parameter, dependency	Directional and topological object spatial relations	Detailed entities of construction domain
Large-Scale Concept Ontology for Multimedia (LSCOM) [48]	Program, location, people, objects, activities, events, and graphics	-	Detailed entities of construction domain
Visual Concept Ontology and Image Processing Ontology [21]	-	Spatial, color, texture	Detailed entities of construction domain
Fang et al. [9]	People, equipment, material, environment	Spatial relationship	Features of different construction entities
Han and Golparvar-Fard [49]	Building components	Physical relationship	Other types of visual objects in construction and features of these entities
Zhong et al. [23]	People, machinery, material, environment	-	Features of different construction entities
DiCon [25]	Image, agent, equipment, building element, material	-	Features of different construction entities

The first type of ontologies includes general visual content description ontologies. The multimedia analysis ontology [47] was designed for knowledge-assisted, domain-specific video analysis that represents the objects and corresponding features in the scene. The large-scale concept ontology for multimedia [48] is an ontology specified to describe video data, especially broadcasts, in which the objects in the scene are associated with corresponding activities, locations, and programs. The visual concept ontology and image processing ontology [24] are upper-level ontologies that match symbolic features with semantic concepts and data.

The second type includes image-related ontologies in the construction domain. Fang et al. [9] developed an ontology for describing a site image to identify job hazards and modeled four entities—people, equipment, material, and environment—along with the spatial relationship between the entities. Zhong et al. [23] also developed an ontology for supporting hazard analysis based on images. This ontology can interpret site images semantically by annotating the entities in the scene and then, based on the image content and domain knowledge, infer the hazard present in the image. Han and Golparvar-Fard [49] introduced an ontology to support automated visual progress monitoring based on point clouds, which describe the building components and the physical relationship between them.

In a previous work, we developed Digital Construction Ontologies (DiCon) [25]. In DiCon, the image class is modeled as an information content entity that contains visual data. Thus, DiCon could represent the metadata of construction images. Meanwhile, DiCon provides a unified representation of the detailed construction workflow with related entities and relations. Although it cannot yet directly represent the content of construction images, it could provide the vocabulary of construction entities as related to the image content. For example, material entities, including building elements, equipment, workers, and materials, could be captured by the image.

There are some limitations of prior ontologies related to interpreting construction images. General visual content description ontologies provide representation for describing visual contents in the images, but they lack the specific domain knowledge of construction. The existing image-related ontologies in the construction domain describe the content entities in the construction image, but two limitations can be identified. First, they have limited relations between the modeled entities—for example, only the spatial relationships are modeled in [35], and only physical relationships are modeled in [23]—and are thus insufficient to represent the detailed features of the objects in the image. Second, these ontologies have limited definitions of image metadata and have not attempted to link with external data sources, thereby diminishing the interoperability of image semantics involved in the construction of DTC. And, in terms of DiCon [25], although it comprehensively models construction workflow-related entities to link with different digital construction data, a detailed description of the visual features and relationships between entities was not included, and thus, it is also insufficient for semantically interpreting image content.

2.4. Research Motivation and Objective

The summary of the reviewed efforts is listed in Table 2. Although state-of-the-art implementations have made important contributions to using images as a source to support construction management, their efforts are mostly focused on resolving the sensory gaps, and therefore the semantic gap remains. Meanwhile, involving image semantics in the DTC environment was neglected by previous efforts. Therefore, the problems of interpretation and integration of construction images remain.

An ontology could provide a formalized semantic representation to describe the construction image contents, including the construction entities, their visible features, and relations among them. First, regarding interpretation, adopting the SII paradigm enables us to infer meaningful higher-level semantics represented in the construction image. For instance, we can determine the process stage of the work shown in the image by examining the process entities and features visible within it. Second, for integration, as the construction entities are semantically represented in the ontology, external information sources about the entities can also be linked [46]. Thus, we posit that there is potential to develop an ontology-based semantic image interpretation system to address both interpretation and integration problems.

From the literature review, we conclude that an ontology that could both serve the SII framework and integrate image semantics with other construction data to establish DTC is still missing. Therefore, this research aims to first propose a conceptual framework to outline the workflow of achieving the semantic interpretation of construction images and

then integrate the image semantics with external digital construction information. Then, we present an ontology that can provide a formalized vocabulary to represent both semantically labeled images and background knowledge for implementing the SII paradigm.

Table 2. Summary and limitations of reviewed works.

Sector	Summary and Limitation
Image-related works in construction	<ul style="list-style-type: none"> • Current implementations mostly are CV approaches that focus on object detection to capture the representation of real construction objects. A semantic description of construction image content is still lacking. • Image data have not been integrated with other digital data sources. Hence, it is difficult to involve the image in the DTC environment.
Semantic image interpretation (SII)	<ul style="list-style-type: none"> • The SII method could support the establishment of a systematic representation of image semantics, but SII depends on domain knowledge to describe the image scenes semantically and to achieve higher-level image semantic inference. • Currently, the concept of SII has limited adoption in the construction domain, and no systematic SII framework has been presented for construction images.
Ontologies	<ul style="list-style-type: none"> • Ontology is one solution to provide a formalized domain knowledge representation to facilitate the interpretation of the construction image and integration with other digital sources to build up DTC. • To date, an adequate ontology that can directly handle the construction of SII and integrate it with other data sources is still missing.

3. Ontology-Based Construction SII (CII)

Considering the reviewed SII paradigm and the requirement of integrating image semantics with external digital sources, we suggested the framework of for CII, as shown in Figure 1. CII consists of four interacting layers designed to indicate the workflow of CII, including (1) data preparation, (2) interpretation, (3) integration, and (4) application. In the following sections, a detailed description of each layer is presented.

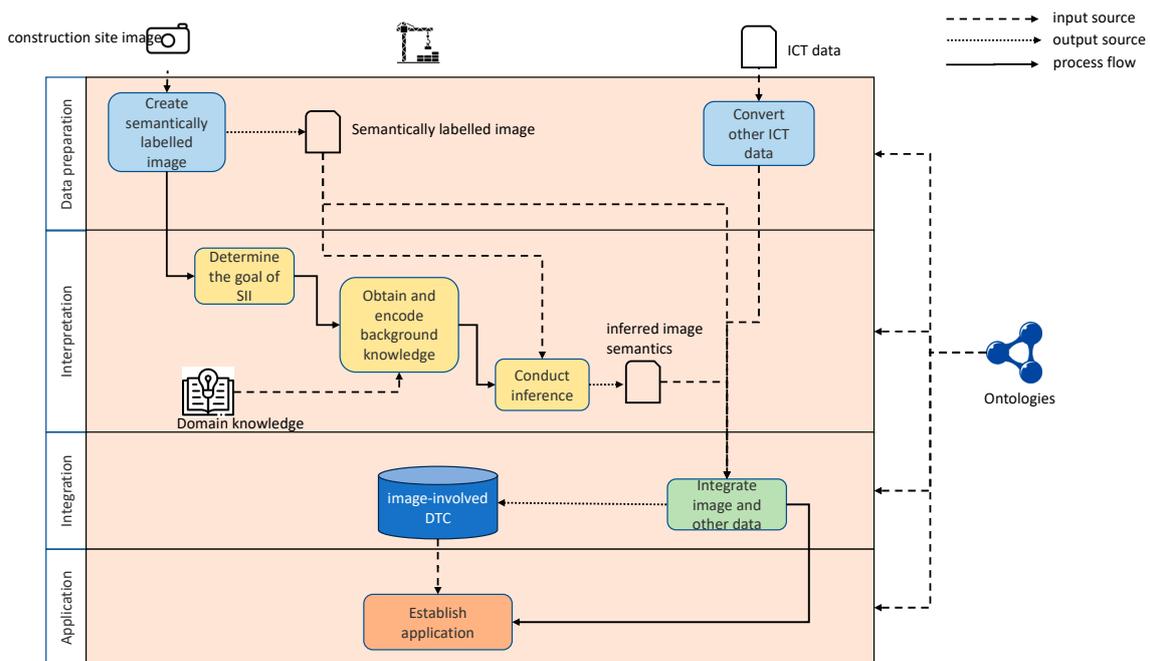


Figure 1. The CII framework.

3.1. Data Preparation Layer

This is the first layer of the framework and aims to obtain the input data of the system. Two different data streams are considered as inputs, which are the construction image and external digital data sources. In terms of the construction image, the initial task is to extract the image semantics from each image, including the related visible construction entities along with their visible features and relationships. Then, the extracted image semantics and the metadata of the image need to be mapped to an ontology that adequately describes the image contents and generate a data graph in the RDF format to build a graph representation of the semantically label image. RDF makes the image semantics machine-readable and ready for inference of the higher-level image semantics and is prepared for further integration with the external construction of digital data sources. The external data sources, including BIM, IoT, IPS, etc., also need to be converted to the RDF format associated with the corresponding ontologies to enable integration.

3.2. Interpretation Layer

Following the data preparation layer, the interpretation layer focuses on achieving the interpretation of the images by using the extracted semantically labeled images. The first task is to set up the goal of SII. According to Hudelot [24], SII is a goal-oriented task, and, for different goals, the interpretation result may differ. For example, given a construction site image, as shown in Figure 2a, the operation management-related personnel may want to know the operation progress states that are represented in the image; they will thus focus on checking the stages of the drywall installation with related entities. However, in terms of the safety-related personnel, they may check whether safety hazards emerge in the images, and as such, they would check whether the worker is wearing personal protection equipment (PPE) appropriately (as shown in Figure 2b). Moreover, for different goals, the knowledge used to interpret the image is also different. Thus, the first step in the interpretation layer is to determine the goal of SII—in other words, to set up the scope of the interpretation—and then obtain the corresponding knowledge needed for use in the interpretation.

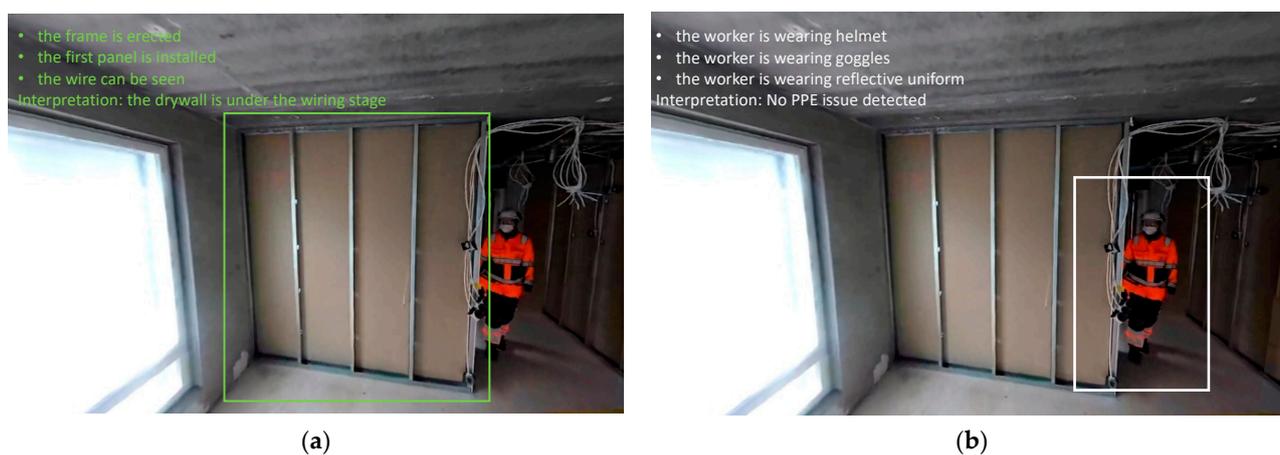


Figure 2. The same image but different interpretations based on different goals. (a) Interpretation of the operation progress states represented in the image; (b) interpretation of checking the worker is wearing personal protection equipment (PPE) appropriately.

To realize further inferencing, besides the semantically labeled image, the background knowledge that specifies the different visual features for the interpretation is required to be input as well. Rules are one kind of knowledge representation and are expressed by formal logic for ontology-based inferencing. They describe an entity that has certain attributes or is related to other entities by certain relations, and thus, implicit knowledge can be derived regarding the entity [50]. Therefore, the second step of this layer is to obtain

and encode the background knowledge into the executable rule sets associated with the ontology that describes the image content. The rules are often expressed by semantic web rule language (SWRL), shapes constraint language (SHACL), or by using the SPARQL protocol and RDF query language rules (SPARQL). Once both the semantically labeled images and rule-represented background knowledge are generated as the two inputs of SII, the last step in this layer is to conduct the inference for higher-level semantics by using the developed rules.

3.3. Integration Layer

This layer aims to integrate the image semantics with the external data to develop a semantic DTC. This process will enrich the database to provide a more comprehensive description of the actual situation and thus trigger further applications that cannot be achieved solely using image semantics. The input of the layer, including the semantically labeled image, inference image semantics in the interpretation phase, and external ICT data sources represented as RDF graphs, will then be integrated into a graph database to build the DTC. A mapping process is required for the integration that connects the entities in the different systems. This mapping process follows the linked data principle that finds the related or corresponding entities in the different systems and creates new links between them. For example, as shown in Figure 3, by using the owl:sameAs statement, we link the drywall (*Wall0602*) represented in the image and with its corresponding element in the BIM model (*GUID: 1ozM5O8GJMGfHwVlxCZkGY*). Similarly, the link between the same location represented in the BIM (*GUID:20FpTZCqJy2vhVJYtjulce*) and IPS (*2nd floor A06*) can also be established.

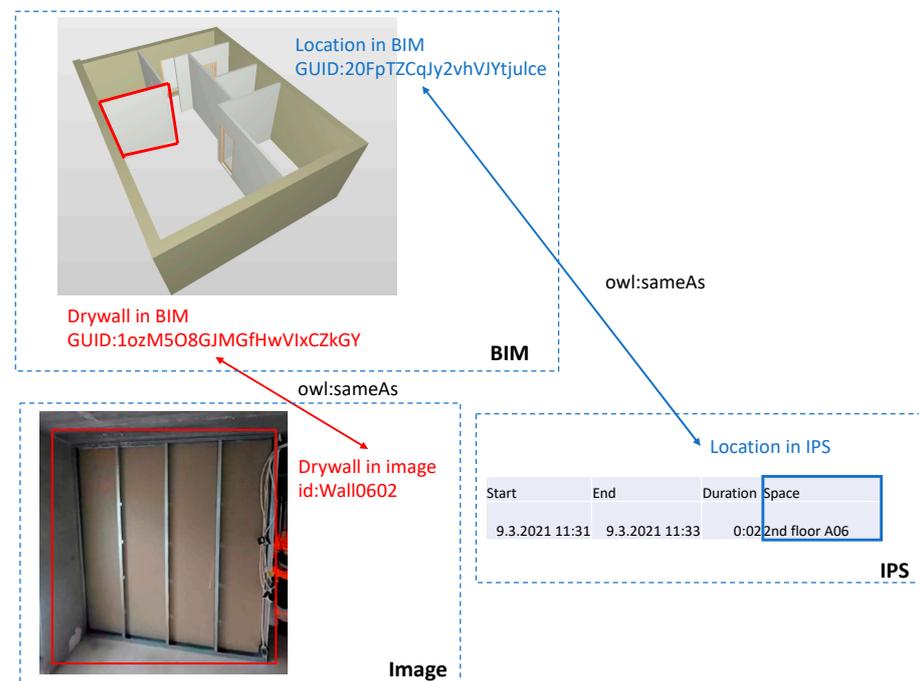


Figure 3. An example of mapping the data in different systems for achieving the integration.

3.4. Application Layer

Once the integration is achieved, useful applications based on the integrated DTC database of the image semantics with the external data can be established to aid the user based on various management demands, for example, complex construction management information retrieval from the integrated database.

4. DiCon-SII Ontology

To further support realizing the CII, an ontology is needed to (1) represent and formalize the construction image semantics of domain concepts and features in the image scene based on construction domain knowledge and (2) create the link between images and other digital data sources required to establish the image-involved DTC. Therefore, in this research, the DiCon-SII ontology is developed. In the following parts of this section, we will introduce the development of ontology.

In our previous research, we developed Digital Construction Ontologies (DiCon), in which we defined construction workflow-related entities and properties and thus achieved the ability to represent and integrate digital construction information from heterogeneous systems [25]. Moreover, in the DiCon Information Module, we defined the image class as an information content entity that has inherited the properties needed to describe its metadata. However, in the DiCon, the detailed description of image contents is missing. Therefore, our aim was to extend DiCon by using the horizontal segmentation approach to yield a detailed description of image semantics. The horizontal segmentation approach is addressed in the SOSA/SNN [51], which aims to develop complementary content from new, related domains by defining classes as well as properties and connecting them to previous ontologies. Meanwhile, to maintain consistency, the development of the extension also follows the same hybrid development methodology of DiCon adopted from Zhou et al. [52], which combines the advantages of the popular methodologies of ontology development, including the Grüninger and Fox approach [53], a system known as “METHONTOLOGY” [54], the “simple knowledge engineering methodology” (SKEM) [55], and the Uschold and Gruninger approach [56]. As shown in Figure 4, several steps were taken to develop DiCon-SII. In the following section, details of each step are described.

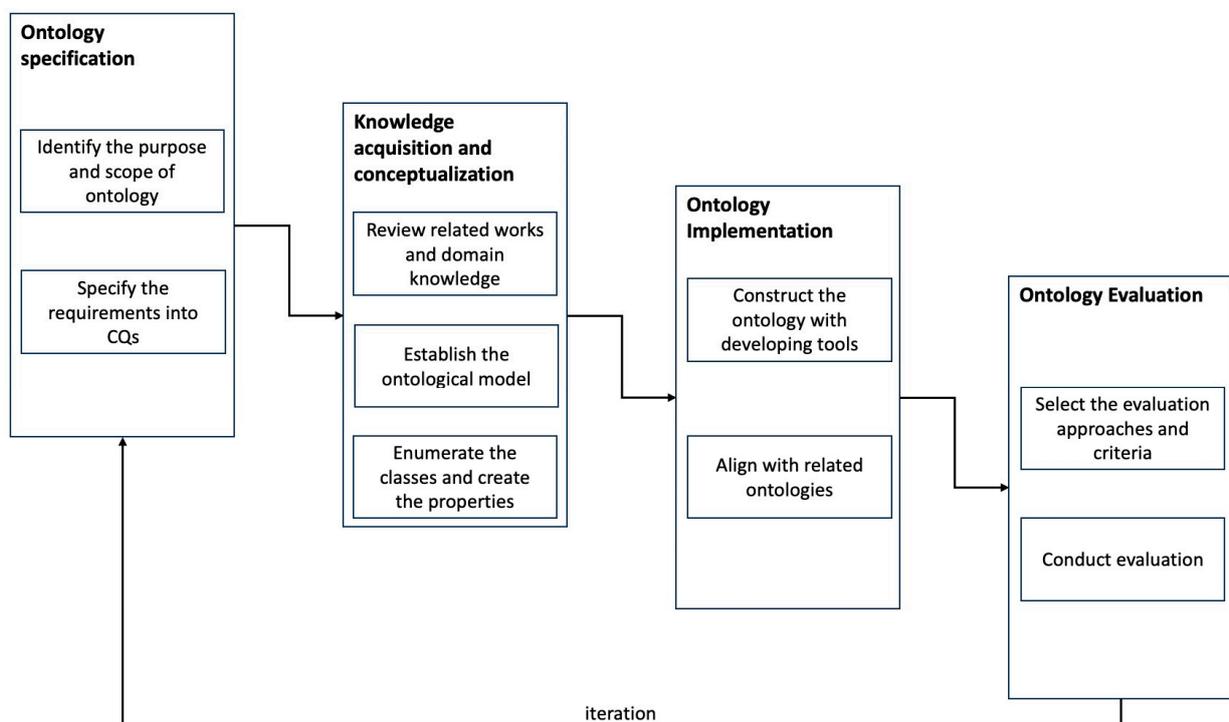


Figure 4. Ontology development approach.

4.1. Ontology Specification

Specification, which aims to determine the purpose and scope of the ontology, is the initial developmental step. The specification process can help the developer to clarify what the ontology should be. To accomplish the specification, three questions are used to determine the purpose, scope, and end users of the ontology.

What is the purpose of ontology? The major purpose of the ontology is to describe contents and low-level features in construction images, including visible entities along with their visual attributes and inter-relations, to provide the vocabulary needed for the interpretation of the image scene and to create the link between images and other digital data sources required to establish the image-involved DTC.

What is the scope of the ontology? Based on the intention of the ontology, DiCon-SII can cover the entities and their visual features and relations that could be captured in the construction images.

Who is the end user of the ontology? The end users of the developed ontology would be from academia, software companies, and industries related to construction images. In terms of academic researchers, they can use ontology to develop image-involved DTC. Software companies could implement ontology as the data structure to develop software solutions. Thus, the industrial user may not directly use the ontology but would benefit from the software developed based on the ontology to utilize the construction image.

The specified ontology scope and purpose are further developed as competency questions (CQs). CQs are a set of question-formulated requirements that contain the tentative terminology of ontology classes and relations. The developed ontology should be able to answer, as an evaluation process, whether the ontology covers the required content [54]. By considering the initial knowledge that must be understood to semantically describe the construction image content within the specified ontology scope, we listed the core CQs for DiCon-SII in Table 3.

Table 3. List of core CQs of DiCon-SII.

Competency Questions
<ul style="list-style-type: none"> • What is the metadata of the image? <ul style="list-style-type: none"> • What is the file of the image? • When is the image created? • What is the scene of the image? • What details are included in the scene of the image? <ul style="list-style-type: none"> • What entities are included in the image? • What visual features and values do the entities have? • What are the visual and physical relationships between entities? • What visual state is the image representing? • What relationships do the entities have between their representation in the image and in other corresponding systems?

4.2. Knowledge Acquisition and Conceptualization

Following the specification, the knowledge acquisition phase is employed to determine what domain knowledge should be acquired for ontology development [54] and to achieve the conceptualization of the ontology. The relevant domain knowledge was initially reviewed during the literature review phase, including current image implementations, fundamental knowledge of SII, and related ontologies. Reviewing the existing studies helped to determine the terminologies of classes and properties that should be modeled to expand DiCon to DiCon-SII. First, knowledge about the domain concepts in the construction image scene was reviewed from the literature. As discussed previously, recent CV-related works can capture or recognize the construction process involving objects, including workers, equipment, materials, and environment-related objects. All of these objects can be seen and captured in the construction image scene and have different types of visual features. Second, in terms of the different visual features of the objects, previous studies on the SII and image ontologies were reviewed to obtain the mandatory properties needed to describe the visual features of the objects—for example, color, size, shape—or the relationships between different objects, like topological or spatial relationships.

The conceptualization phase for DiCon-SII was carried out following the knowledge acquisition. In this phase, the first step was to list the classes and define the class hierarchy, after which the class properties were defined [55]. The terminologies used for the classes and properties were directly acquired from the reviewed studies or ontologies to ensure unambiguity. After the listing process, a generic ontological model was established as a core conceptual model of the ontology with the definitions of the main classes and properties. Such an ontological model can help to formalize the structure of the ontology and ensure that the vocabularies it uses are coherent [57,58].

Figure 5 illustrates the ontological model of DiCon-SII as a result of the knowledge acquisition and conceptualization phases. All of the new concepts and relations of DiCon-SII have the namespace “dicsii:”. The central concept in the model is *Image*, whose representation has two aspects. The first aspect is the metadata of an image. Since the *Image* class is a subclass of *InformationContentEntity* defined in the DiCon Information Module, it inherits the metadata-related properties from *InformationContentEntity* without the redundant modeling. For instance, an *Image* is carried by a certain *ImageFile* that is stored in devices or on a cloud service. In the second aspect, an *Image* is modeled with its unique *ImageScene*. *ImageScene* is a subclass of the *Context* concept from DiCon and refers to the realm that describes the detailed contents of the image scene, including the objects and their features and properties. Each *ImageScene* has a resource description framework (RDF) named graph to contain a detailed description of its contents. The reason to use named graphs is that a certain domain object could be seen in different images but may have different features or relationships with other objects. RDF-named graphs allow the objects to be associated with different properties in different contexts [59]. Thus, this approach was selected to deal with this dynamic circumstance. In terms of interpretation, each image has a represented state that can be described with a natural language, which can be inferred based on their image semantics and the rule-based background knowledge.

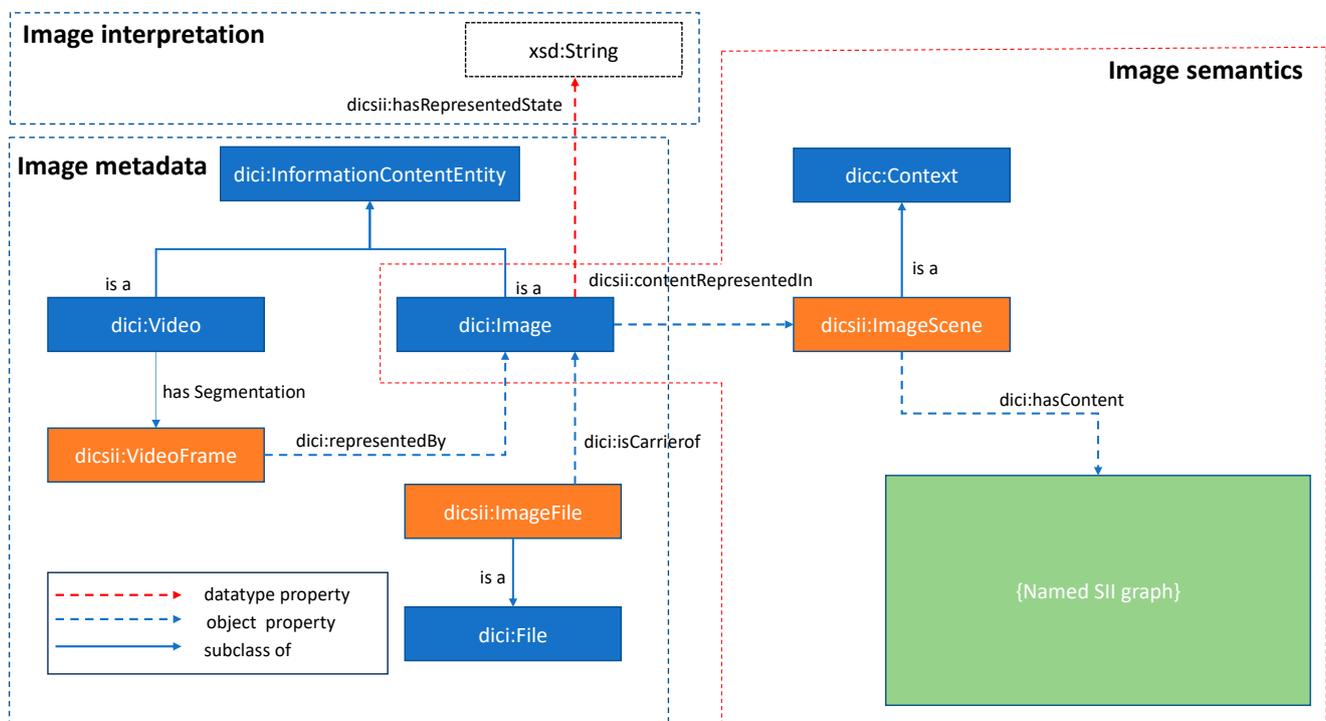


Figure 5. The generic ontological model of DiCon-SII.

For every named graph, there is a more detailed ontological model to represent the contents and features of the image scene. This model absorbs and combines the knowledge from prior works to give a generic semantic description of the image scene contents. As

shown in Figure 6, first, the domain objects, including *BuildingObjects*, *Agent*, *Equipment*, and *MaterialBatch*, are considered subclasses of *VisibleObject*. These classes are essential concepts addressed in previous related works and have already been defined in the DiCon, along with their internal relationships. Therefore, they can link to the corresponding information systems, like BIM, IPS, or the logistic and Enterprise Resource Planning (ERP) system. For a *VisibleObject*, it has different *VisualFeatures* as attributes, including *Shape*, *Color*, *Visibility*, and *Size*, with certain values. These attributes have also been discussed in the related works. Between different objects, they have *VisualPhysicalRelations* to indicate their physical relations—for example, the spatial relation addressed in [9].

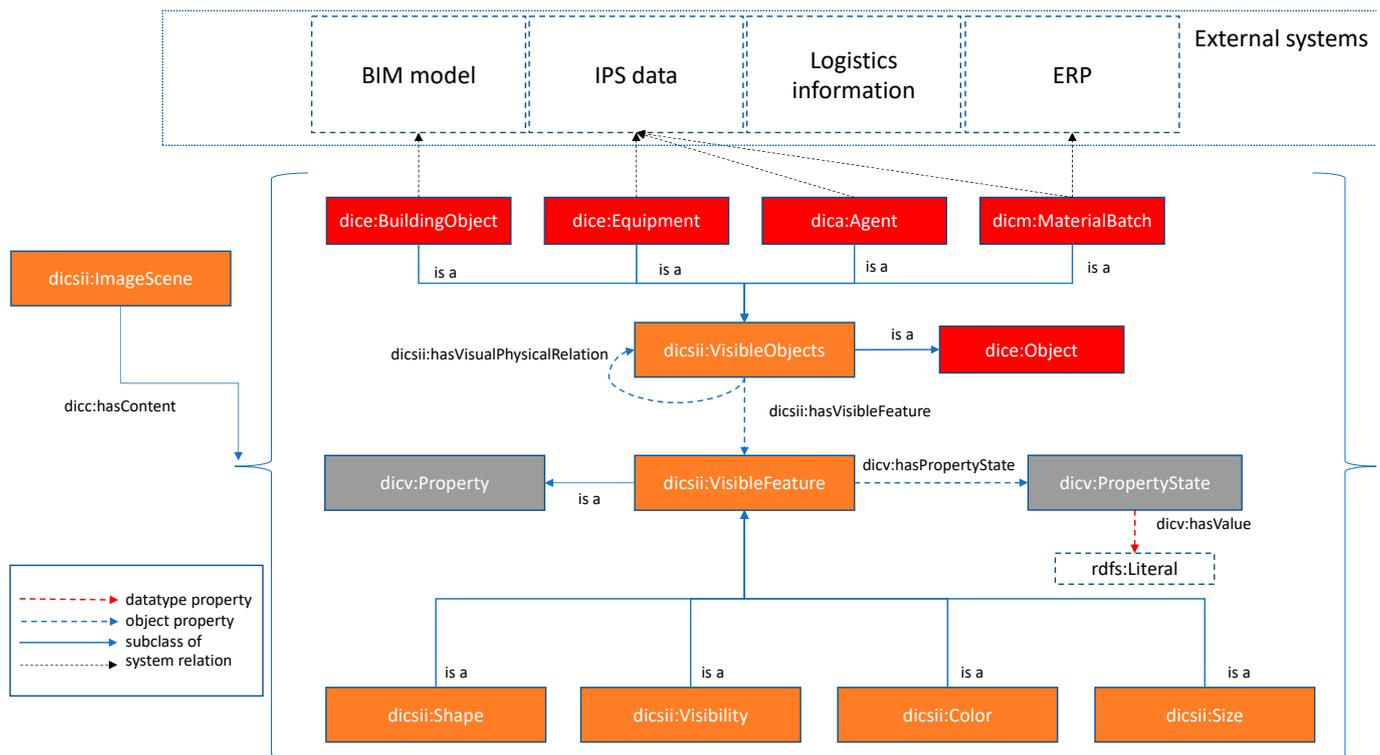


Figure 6. The detailed ontological model for describing construction image contents.

4.3. Ontology Implementation

DiCon-SII was also implemented by encoding in OWL to maintain consistency with DiCon. OWL is a W3C standard language with the advantage of providing richer semantic expression that all concepts, relations, and attributes are modeled as classes, object properties, and data properties, respectively. The OWL-based ontology can be linked to and from other ontologies in the broader ontology ecosystem [60]. The OWL encoding of DiCon-SII was accomplished via Protégé, an open-source and very popular OWL editor that has comprehensive ontology development features [61].

4.4. Ontology Evaluation

Ontology evaluation is an essential process for newly developed ontologies as it can determine the extent to which they satisfy the requirements or are semantically and syntactically correct [62]. The evaluation ensures that the ontologies are correct and usable before implementing them in practical cases. We adopt the criteria-based evaluation method to evaluate the DiCon-SII. Criteria-based evaluation is an analytic process to access the content of an ontology [62]. Based on the purpose of DiCon-SII, we selected coverage, consistency, clarity, and usability as four evaluation criteria. The corresponding evaluation approaches included automated consistency checking, analysis of the clarity, the answering of CQs, and task-based evaluation.

4.4.1. Answering CQs

CQs are the requirement specification of the developed ontology, which should be able to answer these questions [62]. Answering CQs is a simple way for developers to self-check the coverage of the ontology [63]. DiCon-SII was checked to determine whether it contained enough knowledge to answer the questions. The answering process was performed via SPARQL queries toward the instance data from the example case in Section 5. The result of the query is shown in Table 4, which validates that DiCon-SII can cover the requirements we defined in the specification phase.

Table 4. Example of answering the CQs based on the instance data.

Specified CQs	Title 2
What is the file of the image?	ex:Image20210310_1.png
When is the image created?	2021-03-10T14:30:00^^xsd:dateTime
What is the scene of image 20210310_1?	The scene graph ex:graph20210301_1 ex:Wall1A03 and its sub-elements: ex:Wall1A03Stud, ex:Wall1A03Electricity, ex:Wall1A03FrontPanel, ex:Wall1A03BackPanel.
What entities are included in the image?	dicsii:Visibility ex:Wall1A03Stud "visible"
What visual features and value do the entities have?	ex:Wall1A03Electricity "visible" ex:Wall1A03FrontPanel "visible" ex:Wall1A03BackPanel "not visible"
What are the visual and physical relationships between entities?	ex:Wall1A03Electricity dicsii:within ex:Wall1A03Stud
What visual state is the image representing?	dicsii:hasRepresentedVisualState "Wiring"
What relationships do the entities have between their representation in the image and in other corresponding systems?	ex:Wall1A03 owl:sameAs ex:Wall_1ozM5O8GJMGfHwVlxCZkGy from the BIM model

4.4.2. Automated Consistency Checking

Consistency checking is performed to ensure that no contradictory facts exist in an ontology based on description logic principles, such as logical conflicts or inconsistent classes. Consistency checking is enabled by description logic reasoners, which perform various automated inferencing services [62]. In the present research, consistency checking of the proposed ontology was conducted using the Pellet reasoner [64], which is a built-in Protégé description logics reasoner. After using the consistency checking function in Protégé with the Pellet reasoner, DiCon-SII was confirmed to be consistent and coherent.

4.4.3. Analysis of the Clarity

Clarity refers to whether an ontology effectively communicates the intended meaning of defined terms, which are specified without ambiguity [65]. Since DiCon-SII is the extension of DiCon, the overlapped concepts and relations have already been evaluated. In terms of the extended contents, to ensure their clarity, their definitions in DiCon-SII were extracted from the existing ontologies and reviewed literature. Thus, DiCon-SII was defined formally and unambiguously.

4.4.4. Task-Based Evaluation

The utilization of task-based evaluation served to check the ontology's useability in accomplishing specific tasks aligned with its designed purpose. This guiding principle led us to conduct a case study to implement the DiCon-SII with a CII task. In the following Section 5, the case study is described in detail.

5. Case Study

This section demonstrates a case study that implements the developed ontology and proposed CII framework to interpret drywall installation stages from images and build

up an image-involved DTC for the drywall installation process. The purpose of the case is to conduct a task-based evaluation of the developed ontology to test if it can fulfill the requirement of interpreting and integrating the construction image. We obtained the practical data from the indoor construction phase of a residential building project comprising 86 apartments in Espoo, Finland. The digital data acquired from this project include the following:

1. A total of 100 images acquired from 360-degree videos of weekly drywall installation inspection from a construction project in the drywall installation phase. These images, captured weekly, were taken of drywall partitions and pieces in apartments from the angle of one fixed orientation to distinguish the first side panel from the second side panel.
2. The indoor positioning system (IPS) data of the drywall installers. IPS tracked the presence of installers in different locations, with the stationary gateways deployed in different apartments to capture the signals from portable Bluetooth beacons attached to the installers.
3. The architectural building information model (BIM) to provide the quantity information of the drywalls.

The case study follows the CII architecture discussed earlier in Section 3, with its implementation architecture details shown in Figure 7. Details of each step are explained in the following section.

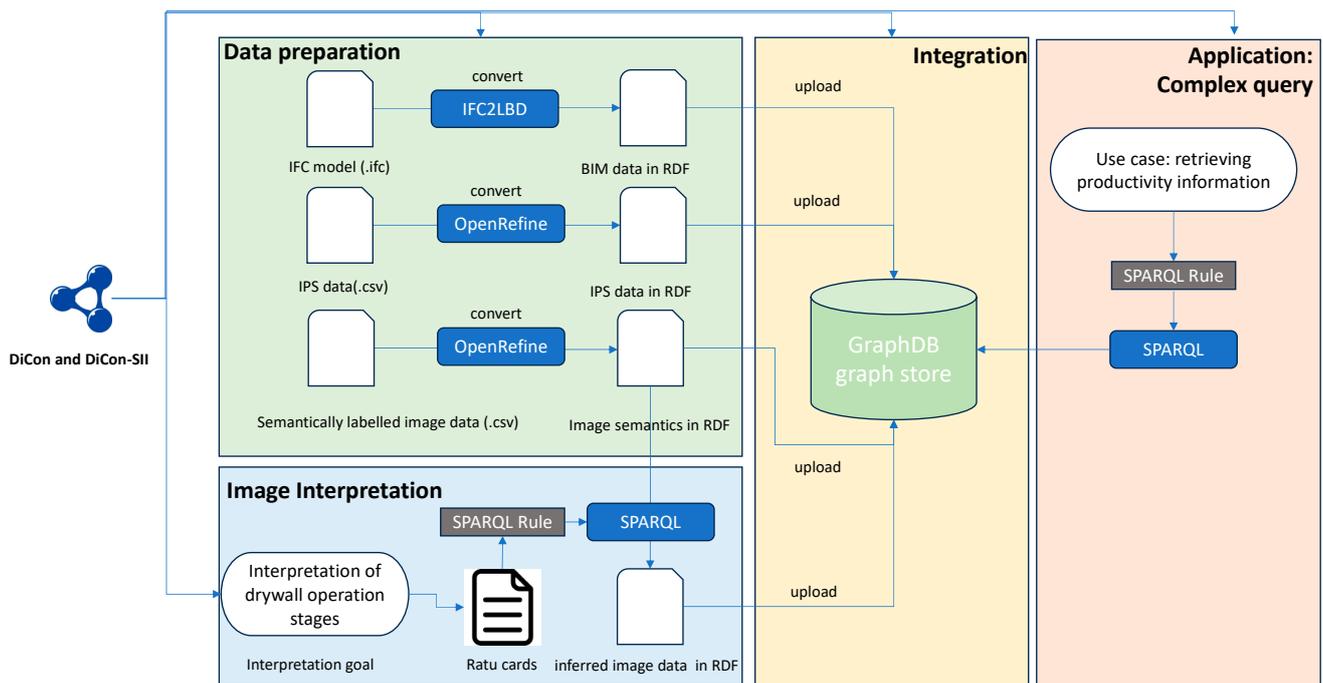


Figure 7. The implementation architecture of the case study.

5.1. Data Preparation

The source image was manually analyzed in order to obtain the ground truth of image scene contents as the semantically labeled images. As an example, shown in Figure 8, the drywall contents (BuildingObject) within the demarcated red box of the image and floorplan were extracted with the visibility features, revealing whether the drywall components were seen being installed or absent. They were subsequently converted to the semantically labeled image as individual RDF-named graphs associated with the DiCon- ontology. IPS data is also in tabular format that comprises the data of workers' presence at the location and time interval. As shown in Figure 7, the tabular data were converted using the DiCon and DiCon SII as the RDF skeleton with the OpenRefine software (version 3.2). In terms

of the architectural BIM model in IFC format, it was converted by using the IFC2LBD converter to RDF.

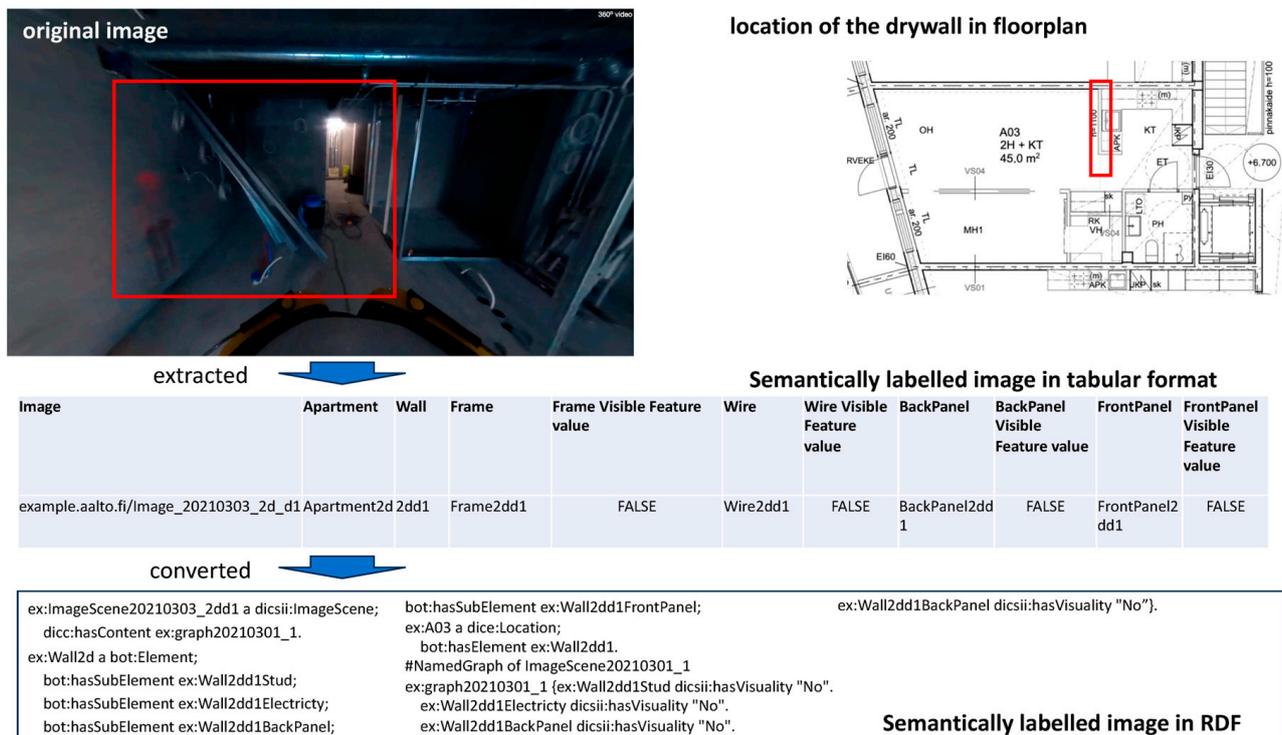


Figure 8. An example of the semantically labeled image.

5.2. Interpretation of Drywall Operation Stages

Installation of drywall is a crucial phase in interior construction, which involves a sequence of stages that require precise flows, specific conditions, and a strong reliance on other related tasks [66]. In each stage, the drywall components have different visible features that can be used to indicate the state of the work stages. In this case, the goal of the interpretation was to infer the operation stages of the drywall installation process represented by the images. Based on this goal, we first acquired the background knowledge of the drywall installation process from a Ratu card (Ratu F52-0327, 0457, and 0452) [67–69] and translated it into rule sets. Ratu cards are Finnish guides of construction operations that describe the detailed workflow of different construction operations, along with the features of different stages. Thus, a Ratu card was used as a knowledge source for developing the rule sets. Meanwhile, the work sequence from the case site began with wiring and then proceeded to panel installation. Therefore, combining the background knowledge and practical information, we classified seven stages in a sequence of drywall installation: “not started”, “framing”, “wiring”, “first panel”, “second panel”, “plastering”, and “painting.” In the different stages, the features of all drywall components are different. For example, during the “not start” stage, there are no drywall components visible. In the “framing” stage, only the frames of drywall are visible. In the “wiring” stage, wires and frames are visible. Since the wall is not enclosed during the “first panel” stage, the frames and wires are still exposed from the open side. In terms of the “second panel” stage, the drywall is enclosed by two side panels; thus, from the same orientation as the “first panel” image, the frames and wires are invisible. For “plastering” and “painting”, the visibility of the panels does not change, but there is a specific value of the color feature. Thus, a logical map of the rules can be set up, as shown in Figure 9.

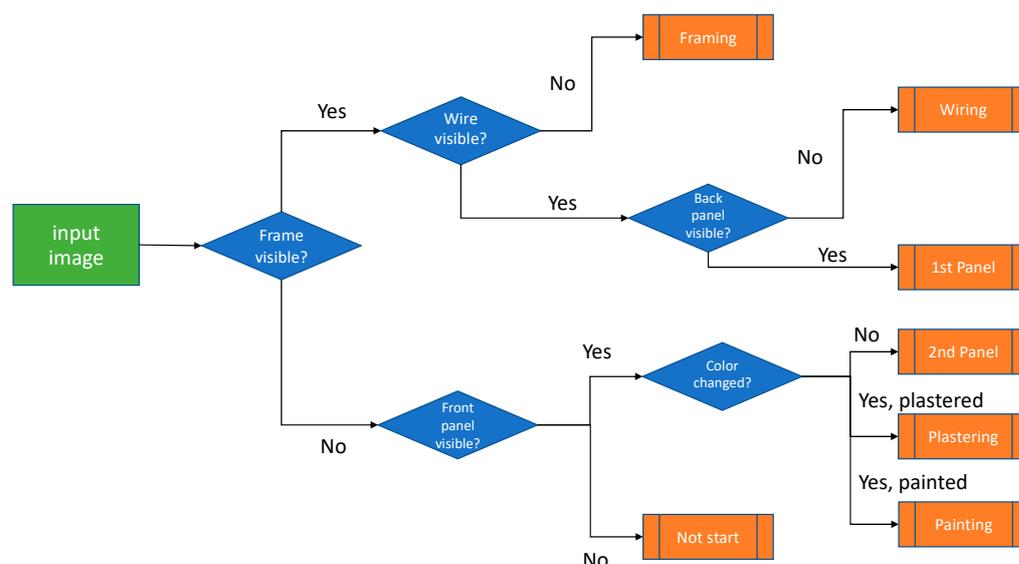


Figure 9. The logical map for identifying the different states of drywall installation.

Based on this logical map, the rules for inferring the drywall installation can be encoded. In this case, we selected SPARQL as the language for expressing the rules due to its better flexibility in handling named graphs. This approach is called materialization inference based on logic, in which the new explicit statements are inferred and added to the repository [70]. As an example, the query and the result for the image with the stage of “wiring” are shown in Table 5. In the query, the INSERT function is used to create a new RDF triple of the image and the stage represented by it when the conditions are satisfied. After the inference process, all of the images are interpreted with respect to which drywall operation stage they are representing.

Table 5. SPARQL query for inference images of target drywall with the “wiring” stage.

```

INSERT {?image dicsii:hasRepresentedVisualState "Wiring"}
WHERE {
  ?image dicsii:contentRepresentedIn ?imagescene.
  ?imagescene dicsii:hasContent ?graph.
  GRAPH ?graph {
    ?1stPanel dicsii:hasVisibleFeature ?1stPanelvisibility.
    ?1stPanelvisibility a dicsii:Visibility.
    ?1stPanelvisibility dicsii:hasPropertyState ?1stPanelvisibilityState.
    ?1stPanelvisibilityState dicsii:hasValue "False".

    ?Wire dicsii:hasVisibleFeature ?Wirevisibility.
    ?Wirevisibility a dicsii:Visibility.
    ?Wire visibility dicsii:hasPropertyState ?WirevisibilityState.
    ?Wire visibilityState dicsii:hasValue "True".

    ?Frame dicsii:hasVisibleFeature ?Framevisibility.
    ?Framevisibility a dicsii:Visibility.
    ?Framevisibility dicsii:hasPropertyState ?FramevisibilityState.
    ?FramevisibilityState dicsii:hasValue "True".

    ?2ndPanel dicsii:hasVisibleFeature ?2ndPanelvisibility.
    ?2ndPanelvisibility a dicsii:Visibility.
    ?2ndPanelvisibility dicsii:hasPropertyState?2ndPanelvisibilityState.
    ?2ndPanelvisibilityState dicsii:hasValue "False".
  }
}

```

5.3. Integration

As mentioned, if the image semantics can be integrated with existing ICT implementations, the images could become important source data for creating the DTC to reflect real or semi-real situations [19]. For this reason, CII and DiCon-SII were implemented to facilitate the fusion of heterogeneous information to build a DTC of drywall operation that involved image, IPS, and BIM data. In the data preparation phase described in Section 5.1, the data inputs have been converted into RDF representation. For the integration, the task is to map the heterogeneous sources by linking the same entities represented in the different systems. In our previous research of the DiCon [25], we achieved similar data mapping of these systems with the DiCon. The mapping of the data is as depicted in Figure 10; the alignment RDF statements were added to the RDF graphs, and the integration was achieved automatically when they were stored in the Graph DB tool.

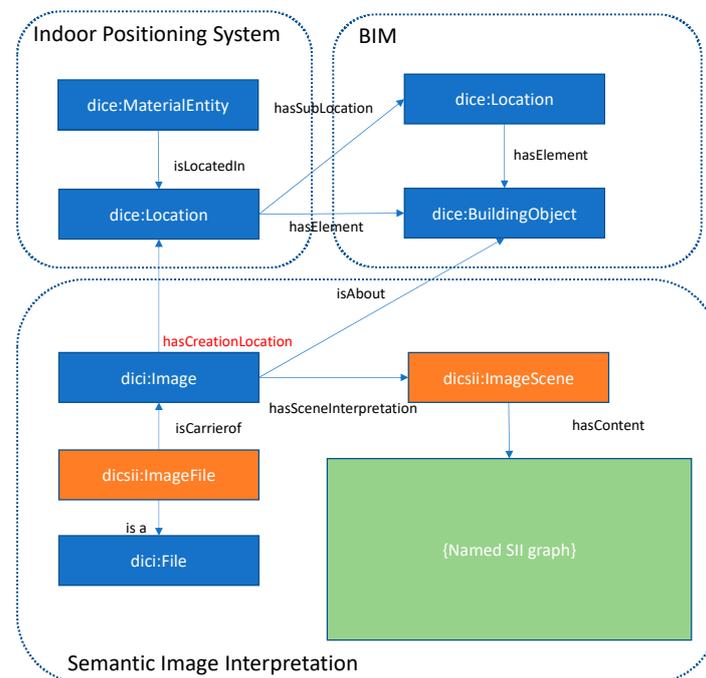


Figure 10. Data mapping of the drywall operation DTC.

5.4. Use Case: Complex Query for Retrieving Productivity Information

To further demonstrate the potential and usage of the image-involved DTC, a use case served to query the integrated data and interpret drywall stage information to achieve complex queries for determining the drywall operation productivity. From the image interpretation, we can infer stage information of the drywall operation but not the actual duration of the worker in the observed location. On the other hand, IPS provides duration information but not the quantity information of the drywalls in the location. Thus, the two data sources are complementary.

To infer productivity, we utilized the inferred stage information from the interpretation phase to query the date interval between two images that contained stage changes in the same drywall. Then, we checked whether other drywalls exhibited stage change in the same location within the date interval and extracted the total area of the wall with a stage change as the workload quantity. Finally, we used this date interval as the time range to find the presence stamps of the drywall partitioner in the location. The first and last presence of the partitioner indicate the actual start and finish time of the stage change. Thus, a productivity index can be calculated as the total area of drywall that has a stage change in the location divided by the actual work duration between actual start and finish times. A set of SPARQL queries were set up for the target information retrieval. As shown in

Table 6, the first query used was the SELECT function to first find the images for a drywall that has inferred the “not started” and “framed” stage and then find the creation time of the images to indicate the time interval between the stage change in the drywall. The second query was employed to determine the area of drywall whose stage changed to “framing” from the BIM model. By using the SELECT function with the logic path of finding the drywall, the image is represented, and its corresponding representation in the BIM model, the area of the drywall can be extracted. The third query was used to detect the presence of the partitioner in Apartment 2D from the IPS.

Table 6. SPARQL queries for images depicting the drywall partition stage changes, work quantity, and worker presence.

```

SELECT ?date1 ?date2
WHERE
{?image1 dicsii:hasRepresentedVisualState "not start".
 ?image2 dicsii:hasRepresentedVisualState "Framing".
 ?image1 dici:isCreatedAt ?date1.
 ?image2 dici:isCreatedAt ?date2.
 ?image1 dici:isAbout <http://example.aalto.fi/example.aalto.fi/Wall/2d/d1>.
 ?image2 dici:isAbout <http://example.aalto.fi/example.aalto.fi/Wall/2d/d1>.
.}

```

```

SELECT distinct ?id ?area
#show the id and area of wall d1 in BIM and images
where
{?image dici:isAbout ?wall.
 ?image dici:isCreatedAt ?date.
 ?image dicsii:hasRepresentedVisualState "Framing".
 ?bimwall owl:sameAs ?wall.
 ?bimwall props:globalIdIfcRoot_attribute_simple ?id.
 ?bimwall props:glazingAreaFraction_simple ?area.
 filter(?date > "2021-03-02"^^<http://www.w3.org/2001/XMLSchema#Date> && ?date
 <"2021-03-11"^^<http://www.w3.org/2001/XMLSchema#Date>)
 }
 }

```

```

SELECT distinct ?observation ?begin ?end
#show the presence of Partitioner in apt2d
where {
 ?observation time:hasBeginning ?begin.
 ?observation time:hasEnd ?end.
 ?observation time:hasDuration ?dura.
 ?observation sosa:isObservedBy ?gateway.
 ?gateway dice:isLocatedIn <http://example.aalto.fi/Apartment/2d>.
 filter(?begin > "2021-03-03T00:00:00"^^<http://www.w3.org/2001/XMLSchema#dateTime> &&
 ?begin <"2021-03-11T00:00:00"^^<http://www.w3.org/2001/XMLSchema#dateTime>)
 }
 }

```

By extracting this information, the time range extracted in the first query was used as the time range to filter the workers’ presence in the location. By running this set of queries, the total area of drywall in Apartment 2D at the “framing” stage was extracted as 54.49 m², and the total work duration was 8 h 58 min. Thus, the productivity of the partitioner for the framing task in Apartment 2D can be calculated as $54.49/8.97 = 6.07$ m²/h.

The case study implemented the proposed DiCon-SII ontology and CII framework with the entire workflow of data preparation, interpretation, integration, and application. First, the case achieved to utilize DiCon-SII represented semantically described image contents and their features and combined with the domain knowledge to infer drywall process stage information represented by the image. Second, it illustrates the capability of DiCon-SII to provide a formalized vocabulary for representing construction image semantics to integrate image and external source data. Moreover, the integrated data

can be used in conjunction with DiCon to conduct a complex query to retrieve drywall productivity information, which is not possible using a single data source alone. In sum, as a task-based evaluation, the case validates the usability of DiCon-SII to provide a formalized vocabulary for representing construction image semantics to complete the task of interpretation and integration.

6. Discussion

6.1. Contribution to Knowledge

In this work, we identified major challenges in making effective use of construction images, namely the problems of effective *interpretation* of images and *integration* with other data sources. We developed the DiCon-SII ontology along with the CII framework to address these challenges. The study makes the following contributions: (1) improving the interpretation of construction images and (2) enabling the integration between construction images and other digital construction data sources.

First, the proposed framework (CII) illustrates the workflow of implementing the SII method based on ontology in the construction domain to support construction image interpretation. Previous efforts have neglected SII as a goal-oriented task since they were confined to single SII goals concerning the specific theme—thus, no generic and systematic SII framework was addressed by these works. Therefore, the previous approaches may not be suitable for other SII goals in the construction sector. CII provides a generic guide from the beginning to set up the SII goals, acquire the related knowledge, and use ontology to generate the semantically labeled image and encode the knowledge to rule sets for the inference of higher-level semantics. Beyond the SII paradigm, the addressed CII framework also takes the integration into account. Since construction involves heterogeneous data/information streams. To further fulfill the aim of CII, we proposed a DiCon-SII ontology that specifically expanded and designed for construction images by thoroughly collecting semantic image knowledge from various previous studies and combining it with holistic construction knowledge. DiCon-SII provides a specific conceptualization of the entities and relations of the construction image. Thus, DiCon-SII can adequately provide the generic vocabulary required to represent the low-level features of construction images to fit with the CII. This also provides a way to represent image information in a structured manner. Compared to existing efforts for interpreting construction image semantics, DiCon-SII obtained specific construction domain knowledge to describe the concepts and features in the construction image scene. Moreover, the DiCon-SII implemented the OWL, which enabled it to conduct the rule-based inference to achieve the interpretation of higher-level information extract, for example, the drywall installation stage extraction in the case study.

Second, DiCon-SII and CII also maximize the utilization of image semantics, not only providing a guideline for higher-level semantics inference but also enabling integration with external data sources to build the DTC. Although images contain massive amounts of information to reflect the construction process, they are also a point solution that can only capture the information from a visual aspect. To build up a holistic situation comprehension, the image should collaborate with other digital systems to establish a digital twin. CII describes the process of integrating image semantics and external information, thus enabling the building of the semantic DTC and presenting ways to conduct further complex queries to utilize the integrated data in the DTC. Meanwhile, leveraging by semantic web ontology, DiCon-SII uses the structured description of image semantics and metadata based on DiCon to achieve systematic integration with external ICT data. This makes the image one part of the DTC. Therefore, the proposed ontology also addresses the data fusion problem of DTC, which is known as one bottleneck of establishing the DTC [11,19]. Hence, DiCon-SII contributes a solution that could support the establishment of a comprehensive DTC with the images, further enabling more complex queries to retrieve interesting information.

By addressing the interpretation and integration challenges in construction images, DiCon-SII and CII can offer several practical benefits:

1. Improving automatic image analysis: Resolving interpretation issues ensures that the computer comprehends the content of construction images, enabling automatic inference for image analysis. This leads to more precise and automated analysis, enhancing the stakeholders understanding the situation represented in construction images.
2. Enhancing situational awareness onsite: Integrated and interpreted construction image data forms a robust foundation for establishing the image-involved DTC. Such a DTC provides project managers and stakeholders with comprehensive insights into onsite situations, fostering more informed and successful decision-making.
3. Providing a versatile tool for construction image analysis and management: The presented ontology and framework serve as a versatile tool for researchers and industrial users in the construction domain. It accommodates those with related works requiring the analysis or management of construction images. The ontology's extendibility allows users to further expand DiCon-SII with specific concepts and properties, addressing different SII goals, such as quality inspection, safety issues, and progress checking. Additionally, CV/DL researchers in the construction domain can leverage the proposed ontology to link with their works and explore its potential to enhance their approaches via ontology-based inference, improving the accuracy of object classification and higher-level image semantic inference.

6.2. Limitations and Future Works

Despite the novelty of the research presented here, it has several limitations that must be addressed in future work.

First, this research is our initial step in exploring the use of image semantics for supporting construction management. No CV-/AI-based approaches were used in this research for object recognition and feature retrieval. The scope of this research starts from the semantic aspect to provide an ontology-based framework to describe the construction of image contents and achieve the interpretation and integration of image information. Thus, we used the ground truth data as the facts for achieving the inference. Automated image processing is currently out of the scope of this research. As future work, we are currently developing an approach that combines CV/AI algorithms to recognize objects and extract their key features from onsite images with DiCon-SII for the purpose of generating semantically labeled images in an RDF format with large language models (LLMs). This upcoming research will provide further validation of the DiCon-SII and achieve the automatic generation of semantically labeled images and inferences with the rule-represented background knowledge, as shown in this work.

Second, the case study is only tested on specific drywall installation trade. In terms of the construction process inference of the drywall installation case, we applied Ratu cards, a Finnish standardized construction operation guide, as the background knowledge to determine the drywall installation operation sequence and the key features of the process states. In reality, drywall installation has a certain degree of freedom of work sequence, although, in other contexts, the work sequence might be different. Thus, the developed rule sets, in this case, were confined to Finland and may not be applicable in other cases. In general, this is a problem of the rule's genericity. To address this problem, our research group is developing a conceptual platform called the construction process library (CPL), which is a general collection and representation of construction process knowledge. The CPL is intended to be an open platform for representing construction domain knowledge into executable semantic rules and allowing users throughout the world to create, reuse, and share the rules to support information retrieval in the construction sector [71]. The image feature rules are one of the core themes of the CPL for collecting the key visual features of different types of the construction process and encoding the knowledge to ontological rule representation. By using DiCon and DiCon-SII as the unique vocabulary to build the rules, users in the construction sector worldwide can contribute to building the rules or reuse existing rules in the library for semantic image inferencing.

Finally, the proposed ontology requires iterative maintenance and refinement [62]. It should be emphasized that no “perfect” ontologies exist because ontologies are developed based on knowledge. Their contents should be synchronized with dynamic changes in domain knowledge and interests [72]. Therefore, DiCon-SII needs continuous refinement to account for the latest knowledge and user interests.

7. Conclusions

This paper presents a novel ontology-based approach that adopts the semantic image interpretation (SII) paradigm to interpret construction images and an ontology to describe the detailed construction image contents semantically. The proposed CII framework and DiCon-SII address the inherent challenges of construction image interpretation and integration. First, the CII framework outlines the workflow of achieving image interpretation and integration. To support CII, DiCon-SII was developed to semantically describe formalized and linkable vocabulary for visual construction contents and features. These modeled contents and features enable DiCon-SII to be used to interpret higher-level semantics about the construction image scene and bridge images and other digital systems to help construct an image-involved DTC. Overall, DiCon-SII and CII provide the basis for representing construction image semantics and thus enable further interpretation. Novel applications of construction image semantics can be developed and applied to support information retrieval and image management to improve situational awareness of the construction process.

To achieve automated image analysis and exploit the potential of DiCon-SII and CII, further research is needed. Future work will focus on further exploiting the combination of CV/DL and LLMs with ontologies and semantic web rules to build a system for automatically analyzing construction images. This direction will be explored with larger and standard construction image datasets.

Author Contributions: Conceptualization, Y.Z. and M.K.M.; methodology, Y.Z.; validation, Y.Z.; data curation, A.A.; writing—original draft preparation, Y.Z.; writing—review and editing, M.K.M., O.S., S.T. and A.A.; supervision, O.S.; funding acquisition, O.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the ACTOR project of Business Finland’s Low Carbon Built Environment Program that receives funding from EU’s Recovery and Resilience Facility (from 2022 onward) and the Building 2030 consortium of 21 Finnish companies (until 2021).

Data Availability Statement: Data are available upon request from the authors. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, J.; Park, M.W.; Vela, P.A.; Golparvar-Fard, M. Construction Performance Monitoring via Still Images, Time-Lapse Photos, and Video Streams: Now, Tomorrow, and the Future. *Adv. Eng. Inform.* **2015**, *29*, 211–224. [[CrossRef](#)]
2. Zhong, B.; Wu, H.; Ding, L.; Love, P.E.D.; Li, H.; Luo, H.; Jiao, L. Mapping Computer Vision Research in Construction: Developments, Knowledge Gaps and Implications for Research. *Autom. Constr.* **2019**, *107*, 102919. [[CrossRef](#)]
3. Yang, J.; Arif, O.; Vela, P.A.; Teizer, J.; Shi, Z. Tracking Multiple Workers on Construction Sites Using Video Cameras. In *Advanced Engineering Informatics*; Elsevier: Amsterdam, The Netherlands, 2010; Volume 24, pp. 428–434. [[CrossRef](#)]
4. Khosrowpour, A.; Niebles, J.C.; Golparvar-Fard, M. Vision-Based Workforce Assessment Using Depth Images for Activity Analysis of Interior Construction Operations. *Autom. Constr.* **2014**, *48*, 74–87. [[CrossRef](#)]
5. Martinez, P.; Barkokebas, B.; Hamzeh, F.; Al-Hussein, M.; Ahmad, R. A Vision-Based Approach for Automatic Progress Tracking of Floor Paneling in Offsite Construction Facilities. *Autom. Constr.* **2021**, *125*, 103620. [[CrossRef](#)]
6. Gong, J.; Caldas, C.H. Computer Vision-Based Video Interpretation Model for Automated Productivity Analysis of Construction Operations. *J. Comput. Civ. Eng.* **2010**, *24*, 252–263. [[CrossRef](#)]
7. Brilakis, I.; Park, M.W.; Jog, G. Automated Vision Tracking of Project Related Entities. *Adv. Eng. Inform.* **2011**, *25*, 713–724. [[CrossRef](#)]
8. Chi, S.; Caldas, C.H. Automated Object Identification Using Optical Video Cameras on Construction Sites. *Comput. Civ. Infrastruct. Eng.* **2011**, *26*, 368–380. [[CrossRef](#)]

9. Fang, W.; Ma, L.; Love, P.E.D.; Luo, H.; Ding, L.; Zhou, A. Knowledge Graph for Identifying Hazards on Construction Sites: Integrating Computer Vision with Ontology. *Autom. Constr.* **2020**, *119*, 103310. [CrossRef]
10. Liu, H.; Wang, G.; Huang, T.; He, P.; Skitmore, M.; Luo, X. Manifesting Construction Activity Scenes via Image Captioning. *Autom. Constr.* **2020**, *119*, 103334. [CrossRef]
11. Sacks, R.; Brilakis, I.; Pikas, E.; Xie, H.S.; Girolami, M. Construction with Digital Twin Information Systems. *Data-Centric Eng.* **2020**, *1*, e14. [CrossRef]
12. Jiang, F.; Ma, L.; Broyd, T.; Chen, K. Digital Twin and Its Implementations in the Civil Engineering Sector. *Autom. Constr.* **2021**, *130*, 103838. [CrossRef]
13. Akinci, B.; Akinci, B. Situational Awareness in Construction and Facility Management. *Front. Eng. Manag.* **2015**, *1*, 283–289. [CrossRef]
14. Dave, B.; Kubler, S.; Främpling, K.; Koskela, L. Opportunities for Enhanced Lean Construction Management Using Internet of Things Standards. Available online: <http://our-plan.com/about-page> (accessed on 28 October 2019).
15. Ghimire, S.; Luis-Ferreira, F.; Nodehi, T.; Jardim-Goncalves, R. IoT Based Situational Awareness Framework for Real-Time Project Management. *Int. J. Comput. Integr. Manuf.* **2017**, *30*, 74–83. [CrossRef]
16. Teizer, J.; Cheng, T.; Fang, Y. Location Tracking and Data Visualization Technology to Advance Construction Ironworkers' Education and Training in Safety and Productivity. *Autom. Constr.* **2013**, *35*, 53–68. [CrossRef]
17. Zhao, J.; Seppänen, O.; Peltokorpi, A.; Badihi, B.; Olivieri, H. Real-Time Resource Tracking for Analyzing Value-Adding Time in Construction. *Autom. Constr.* **2019**, *104*, 52–65. [CrossRef]
18. Opoku, D.G.J.; Perera, S.; Osei-Kyei, R.; Rashidi, M. Digital Twin Application in the Construction Industry: A Literature Review. *J. Build. Eng.* **2021**, *40*, 102726. [CrossRef]
19. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a Semantic Construction Digital Twin: Directions for Future Research. *Autom. Constr.* **2020**, *114*, 103179. [CrossRef]
20. Donadello, I. Semantic Image Interpretation—Integration of Numerical Data and Logical Knowledge for Cognitive Vision. Ph.D. Thesis, University of Trento, Trento, Italy, 2018. Available online: <http://eprints-phd.biblio.unitn.it/2888/> (accessed on 10 August 2021).
21. Hudelot, C.; Maillot, N.; Thonnat, M. Symbol Grounding for Semantic Image Interpretation: From Image Data to Semantics. In Proceedings of the Tenth IEEE International Conference on Computer Vision Workshops (ICCVW'05), Beijing, China, 17–20 October 2006; p. 1875. [CrossRef]
22. Town, C. Ontological Inference for Image and Video Analysis. *Mach. Vis. Appl.* **2006**, *17*, 94–115. [CrossRef]
23. Zhong, B.; Li, H.; Luo, H.; Zhou, J.; Fang, W.; Xing, X. Ontology-Based Semantic Modeling of Knowledge in Construction: Classification and Identification of Hazards Implied in Images. *J. Constr. Eng. Manag.* **2020**, *146*, 04020013. [CrossRef]
24. Hudelot, C. Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of Biological Organisms, de l'Université de Nice—Sophia Antipolis. 2005. Available online: <https://www-sop.inria.fr/orion/Publications/Articles/THESES/TheseCelineHudelot.pdf> (accessed on 22 March 2022).
25. Zheng, Y.; Törmä, S.; Seppänen, O. A Shared Ontology Suite for Digital Construction Workflow. *Autom. Constr.* **2021**, *132*, 103930. [CrossRef]
26. Mostafa, K.; Hegazy, T. Review of Image-Based Analysis and Applications in Construction. *Autom. Constr.* **2021**, *122*, 103516. [CrossRef]
27. Kim, H.; Kim, H.; Won Hong, Y.; Byun, H. Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. *J. Comput. Civ. Eng.* **2017**, *32*, 04017082. [CrossRef]
28. Park, M.-W.; Koch, C.; Brilakis, I. Three-Dimensional Tracking of Construction Resources Using an On-Site Camera System. *J. Comput. Civ. Eng.* **2012**, *26*, 541–549. [CrossRef]
29. Son, H.; Seong, H.; Choi, H.; Kim, C. Real-Time Vision-Based Warning System for Prevention of Collisions between Workers and Heavy Equipment. *J. Comput. Civ. Eng.* **2019**, *33*, 04019029. [CrossRef]
30. Memarzadeh, M.; Golparvar-Fard, M.; Niebles, J.C. Automated 2D Detection of Construction Equipment and Workers from Site Video Streams Using Histograms of Oriented Gradients and Colors. *Autom. Constr.* **2013**, *32*, 24–37. [CrossRef]
31. Azar, E.R. Construction Equipment Identification Using Marker-Based Recognition and an Active Zoom Camera. *J. Comput. Civ. Eng.* **2015**, *30*, 04015033. [CrossRef]
32. Liu, C.W.; Wu, T.H.; Tsai, M.H.; Kang, S.C. Image-Based Semantic Construction Reconstruction. *Autom. Constr.* **2018**, *90*, 67–78. [CrossRef]
33. Kim, J.; Hwang, J.; Chi, S.; Seo, J.O. Towards Database-Free Vision-Based Monitoring on Construction Sites: A Deep Active Learning Approach. *Autom. Constr.* **2020**, *120*, 103376. [CrossRef]
34. Dimitrov, A.; Golparvar-Fard, M. Vision-Based Material Recognition for Automated Monitoring of Construction Progress and Generating Building Information Modeling from Unordered Site Image Collections. *Adv. Eng. Inform.* **2014**, *28*, 37–49. [CrossRef]
35. Han, K.K.; Golparvar-Fard, M. Appearance-Based Material Classification for Monitoring of Operation-Level Construction Progress Using 4D BIM and Site Photologs. *Autom. Constr.* **2015**, *53*, 44–57. [CrossRef]
36. Hamledari, H.; McCabe, B.; Davari, S. Automated Computer Vision-Based Detection of Components of under-Construction Indoor Partitions. *Autom. Constr.* **2017**, *74*, 78–94. [CrossRef]

37. Donadello, I.; Kessler, B.; Fondazione, L.S.; D'Avila Garcez, A. Logic Tensor Networks for Semantic Image Interpretation. *arXiv* **2017**, arXiv:1705.08968.
38. Donadello, I. Ontology Based Semantic Image Interpretation. In Proceedings of the Doctoral Consortium (DC) Co-Located with the 14th Conference of the Italian Association for Artificial Intelligence (AI*IA 2015), Ferrara, Italy, 23–24 September 2015.
39. Atif, J.; Hudelot, C.; Bloch, I. Explanatory Reasoning for Image Understanding Using Formal Concept Analysis and Description Logics. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *44*, 552–570. [[CrossRef](#)]
40. Gruber, T.R. A Translation Approach to Portable Ontology Specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [[CrossRef](#)]
41. El-Diraby, T.E.; Kashif, K.F. Distributed Ontology Architecture for Knowledge Management in Highway Construction. *J. Constr. Eng. Manag.* **2005**, *131*, 591–603. [[CrossRef](#)]
42. Anumba, C.J.; Issa, R.R.A.; Pan, J.; Mutis, I. Ontology-Based Information and Knowledge Management in Construction. *Constr. Innov.* **2008**, *8*, 218–239. [[CrossRef](#)]
43. Beetz, J.; Borrmann, A. Benefits and limitations of linked data approaches for road modeling and data exchange. In *Advanced Computing Strategies for Engineering: 25th EG-ICE International Workshop 2018, Lausanne, Switzerland, 10–13 June 2018; Proceedings, Part II 25*; Springer International Publishing: Cham, Switzerland, 2018; pp. 245–261.
44. Akinyemi, A.; Sun, M.; Gray, A.J.G. An Ontology-Based Data Integration Framework for Construction Information Management. *Proc. Inst. Civ. Eng.-Manag. Procure. Law* **2018**, *171*, 111–125. [[CrossRef](#)]
45. Kosovac, B.; Froese, T.M.; Vanier, D.J. Integrating Heterogeneous Data Representations in Model-Based AEC/FM Systems. *Proc. CIT* **2000**, *2*, 556–567.
46. Pauwels, P.; Zhang, S.; Lee, Y.C. Semantic Web Technologies in AEC Industry: A Literature Overview. *Autom. Constr.* **2017**, *73*, 145–165. [[CrossRef](#)]
47. Dasiopoulou, S.; Mezaris, V.; Kompatsiaris, I.; Papastathis, V.K.; Strintzis, M.G. Knowledge-Assisted Semantic Video Object Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2005**, *15*, 1210–1224. [[CrossRef](#)]
48. Naphade, M.; Smith, J.R.; Tesic, J.; Chang, S.F.; Hsu, W.; Kennedy, L.; Hauptmann, A.; Curtis, J. Large-Scale Concept Ontology for Multimedia. *IEEE Multimed.* **2006**, *13*, 86–91. [[CrossRef](#)]
49. Han, K.K.; Cline, D.; Golparvar-Fard, M. Formalized Knowledge of Construction Sequencing for Visual Monitoring of Work-in-Progress via Incomplete Point Clouds and Low-LoD 4D BIMs. *Adv. Eng. Inform.* **2015**, *29*, 889–901. [[CrossRef](#)]
50. Wu, C.; Wu, P.; Wang, J.; Jiang, R.; Chen, M.; Wang, X. Ontological Knowledge Base for Concrete Bridge Rehabilitation Project Management. *Autom. Constr.* **2021**, *121*, 103428. [[CrossRef](#)]
51. Haller, A.; Janowicz, K.; Cox, S.J.D.; Lefrançois, M.; Taylor, K.; Le Phuoc, D.; Lieberman, J.; García-Castro, R.; Atkinson, R.; Stadler, C. The SOSA/SSN Ontology: A Joint W3C and OGC Standard Specifying the Semantics of Sensors, Observations, Actuation, and Sampling. *Semant. Web* **2018**, *10*, 9–32. [[CrossRef](#)]
52. Zhou, Z.; Goh, Y.M.; Shen, L. Overview and Analysis of Ontology Studies Supporting Development of the Construction Industry. *J. Comput. Civ. Eng.* **2016**, *30*, 04016026. [[CrossRef](#)]
53. Grüninger, M.; Fox, M.S. Methodology for the Design and Evaluation of Ontologies. 1995. Available online: https://www.researchgate.net/publication/2288533_Methodology_for_the_Design_and_Evaluation_of_Ontologies (accessed on 6 November 2023).
54. Fernández-López, M.; Gómez-Pérez, A.; Juristo, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series, Palo Alto, CA, USA, 24–26 March 1997; Facultad de Informática (UPM), Stanford University: Stanford, CA, USA, 1997.
55. Noy, N.F.; McGuinness, D.L. Ontology Development 101: A Guide to Creating Your First Ontology. 2001. Available online: www.unspsc.org (accessed on 14 December 2020).
56. Uschold, M.; Gruninger, M. Ontologies: Principles, Methods and Applications. *Knowl. Eng. Rev.* **1996**, *11*, 93–136. [[CrossRef](#)]
57. Holsapple, C.W.; Joshi, K.D. A Collaborative Approach to Ontology Design. *Commun. ACM* **2002**, *45*, 42–47. [[CrossRef](#)]
58. France-Mensah, J.; O'Brien, W.J. A Shared Ontology for Integrated Highway Planning. *Adv. Eng. Inform.* **2019**, *41*, 100929. [[CrossRef](#)]
59. Groth, P.; Gibson, A.; Velterop, J. The Anatomy of a Nanopublication. *Inf. Serv. Use* **2010**, *30*, 51–56. [[CrossRef](#)]
60. Kalibatiene, D.; Vasilecas, O. Survey on Ontology Languages. In *Perspectives in Business Informatics Research; Lecture Notes in Business Information Processing*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 90, pp. 124–141. [[CrossRef](#)]
61. Horridge, M.; Knublauch, H.; Rector, A.; Stevens, R.; Wroe, C. *A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools Edition 1.0*; University of Manchester: Manchester, UK, 2004.
62. El-Gohary, N.M.; El-Diraby, T.E. Domain Ontology for Processes in Infrastructure and Construction. *J. Constr. Eng. Manag.* **2010**, *136*, 730–744. [[CrossRef](#)]
63. El-Diraby, T.E.; Osman, H. A Domain Ontology for Construction Concepts in Urban Infrastructure Products. *Autom. Constr.* **2011**, *20*, 1120–1132. [[CrossRef](#)]
64. Sirin, E.; Parsia, B.; Grau, B.C.; Kalyanpur, A.; Katz, Y. Pellet: A Practical OWL-DL Reasoner. *J. Web Semant.* **2007**, *5*, 51–53. [[CrossRef](#)]
65. Gomez-Perez, A. Some Ideas and Examples to Evaluate Ontologies. In *Proceedings the 11th Conference on Artificial Intelligence for Applications, CAIA 1995, Los Angeles, CA, USA, 20–23 February 1995*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 1995; pp. 299–305. [[CrossRef](#)]

66. Liu, H.; Lu, M.; Al-Hussein, M. Ontology-Based Semantic Approach for Construction-Oriented Quantity Take-off from BIM Models in the Light-Frame Building Industry. *Adv. Eng. Inform.* **2016**, *30*, 190–207. [[CrossRef](#)]
67. Rakennustieto. Ratu F52-0327 Kevyen Väliseinän Purku ja Uusiminen. *Levyseinät. Menekit ja Menetelmät (Demolition and Replacement of a Light Partition. Panel Walls. Processes and Methods)*. Rakennustieto. Available online: <https://www.rakennustietokauppa.fi/sivu/tuote/ratu-f52-0327-kevyen-valiseinan-purku-ja-uusiminen-levyseinat-menekit-ja-menetelmat/2743115> (accessed on 6 November 2023).
68. Rakennustieto. Ratu 0457 Rappaus (Plastering). *Rakennustieto*. Available online: <https://www.rakennustietokauppa.fi/sivu/tuote/ratu-0457-rappaus/2742605> (accessed on 30 November 2021).
69. Rakennustieto. Ratu 0452 Sisämaalauk. *Menekit ja Menetelmät (Interior Painting Processes and Methods)*. Rakennustieto. Available online: <https://www.rakennustietokauppa.fi/sivu/tuote/ratu-0452-sisamaalaus-menekit-ja-menetelmat/2742626> (accessed on 30 November 2021).
70. Reasoning—GraphDB SE 9.11.0 Documentation. Available online: <https://graphdb.ontotext.com/documentation/standard/reasoning.html> (accessed on 26 April 2022).
71. Zheng, Y.; Seppänen, O.; Masood, M.; Törmä, S. Ontology-Based Construction Process Library for Process States Inference. In Proceedings of the International Conference on Computing in Civil and Building Engineering (ICCCBE), São Paulo, Brazil, 18–20 August 2022; Springer: Berlin/Heidelberg, Germany, 2022.
72. Sharman, R.; Kishore, R.; Ramesh, R. (Eds.) *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems; Integrated Series in Information Systems*; Springer: Boston, MA, USA, 2007; Volume 14. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.