Semiautomatic IFC-Based BIM Reconstruction for As-Designed Bridges from 2D Plans Leveraging Semantic Segmentation and Enrichment

Hang Li, S.M.ASCE[1]; and Jiansong Zhang, Ph.D., A.M.ASCE[2]

Abstract: Building information modeling (BIM) is an innovative approach that enables efficient life-cycle management of construction projects. The increasing application of BIM in the bridge sector can be attributed to its ability to streamline communication, enhance collaboration, and mitigate risks. Consequently, BIM contributes to the enhanced efficiency and effectiveness of bridge project workflows. However, the process of manually constructing a BIM instance model for bridges based on two-dimensional (2D) drawings can be labor-intensive, time-consuming, and error-prone. In this paper, the authors proposed a new framework to semi-automate the reconstruction of bridge BIM based on 2D portable document format (PDF) plans. The proposed framework builds on the state-of-the-art PDF2BIM, a 3D BIM reconstruction tool, and further enriches the model through semantic segmentation, information extraction, and semantic enrichment techniques. The integration of these steps helps address the gap of the output from the PDF2BIM technology in the lack of (1) segmented components; and (2) semantic information. It produces semantically segmented and enriched 3D Industry Foundation Classes (IFC)-based bridge BIM instance models. Semantic segmentation was utilized to break the bridge model into different components, such as piers and decks, leveraging as-designed point cloud data. To extract essential semantic information from the bridge plans, information extraction algorithms were developed iteratively. The developed algorithms leverage optical character recognition (OCR) and natural language processing (NLP), with the aim for extracting Specifications for the National Bridge Inventory (SNBI) items as per federal requirements. The proposed framework was tested on six bridges in the state of Indiana, United States. It achieved 97.7% precision and 94.4% recall in automated information extraction. Additionally, it reduced the overall time consumption on creating a bridge BIM instance model by 94.9% compared to the manual approach. The semantically segmented and enriched bridge IFC models can enhance data interoperability and integrity. The models can also improve stakeholder collaboration and support the life-cycle management of bridges. **DOI: 10.1061/JCCEE5.CPENG-5931.**

Author keywords: Semantic segmentation; Industry foundation classes (IFC); Building information modeling (BIM); Semantic enrichment; Information extraction; Natural language processing (NLP); Optical character recognition (OCR).

Introduction

Contemporary bridge projects primarily rely on portable document format (PDF) plans as the deliverables and documents for storage, communication, and exchange of information among diverse stakeholders. However, the American Association of State Highway and Transportation Officials (AASHTO) has adopted the Industry Foundation Classes (IFC) building information modeling (BIM) standard. This standard is the current national benchmark for modeling bridge and road infrastructure projects. Consequently, it is imperative to upgrade bridge project documentation to three-dimensional (3D) BIMs in conformity with this national standard (Li et al. 2024). The integration of BIM has the potential to significantly improve the retrieval and management of bridge-related data (Sacks et al. 2018). Nonetheless, the manual creation of the BIM instance model (i.e., a specific model under the BIM framework, providing detailed information about its geometry, properties, relationships, and metadata of a specific construction project) and identification of defects and properties for numerous bridges is cost-prohibitive due to the immense workload involved (Sacks et al. 2018). Laser scanning and image recognition were predominant techniques employed in the 3D reconstruction of bridges (Jiang et al. 2023). Nonetheless, these methods primarily yield a geometric model in 3D without capturing the semantic information of individual components and their interrelationships. Furthermore, they are unable to provide vital material performance and structural health information (Belsky et al. 2013). Semantic information refers to the meaningful interpretation of data, including the relationships, functions, and attributes of bridge components. For example, a geometric bridge model can be decomposed into various components such as decks and piers. Each component carries with it semantic attributes (e.g., a deck has deck interaction and deck protective system, etc.). This layer of information goes beyond mere geometry to include the purpose, behavior, and specifications of each part, all of which constitute semantic information. Insufficient semantic information within 3D geometric models hinders their ability to fulfill the practical demands of the architecture, engineering, and construction (AEC) industry (Jiang et al. 2023).

[1]Graduate Student, Automation and Intelligent Construction Lab, School of Construction Management Technology, Purdue Univ., West Lafayette, IN 47907. ORCID: https://orcid.org/0000-0003-0600-1920. Email: li4016@purdue.edu

[2]Associate Professor, Automation and Intelligent Construction Lab, School of Construction Management Technology, Purdue Univ., West Lafayette, IN 47907 (corresponding author). ORCID: https://orcid.org/0000-0001-5225-5943. Email: zhan3062@purdue.edu

Note. This manuscript was submitted on January 5, 2024; approved on August 27, 2024; published online on July 30, 2025. Discussion period open until December 30, 2025; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

The manual creation of BIM instance models for bridges is time-consuming and cumbersome. Akanbi and Zhang (2022a, b) proposed a novel framework for producing 3D bridge BIM based on 2D as-designed bridge plans. This framework fills the existing gap in processing and converting traditional 2D bridge drawings into BIMs. It offers a feasible approach to develop algorithms that enable semiautomated conversion of 2D bridge plan PDFs into 3D information models. These 3D models can be further converted into IFC, an international open standard for building and infrastructure modeling. The IFC standard provides vast flexibility in establishing entities, attributing properties, and defining relationships (Li et al. 2022a, b; Mandava and Zhang 2016; Wu and Zhang 2018, 2019a, b). The proposed approach demonstrates a significant time reduction of 96.67% compared to the current state-of-the-art method in generating 3D models while maintaining comparable model quality. However, each IFC model generated by PDF2BIM contains only one entity instance (*IfcMember*), with no detailed distinction between different bridge parts. In addition, the generated IFC model contains only geometric information and lacks other semantic information (Li and Zhang 2024). These issues limited the potential use of the resulting IFC models in practice, such as in improving BIM interoperability, data integrity, and collaborations among different stakeholders (Li and Zhang 2024).

In this research, we aim to address the gap of the output from the state-of-the-art PDF2BIM (Akanbi and Zhang 2022a, b) technology in the lack of (1) detailed components; and (2) semantic information, by developing and integrating semantic segmentation, information extraction, and semantic enrichment algorithms to refine and augment such output. Semantic segmentation facilitates the decomposition of the IFC model into various segments representative of different bridge components, including, e.g., decks and piers. Concurrently, through the processes of information extraction and semantic enrichment, essential semantic information derived from the bridge plan is seamlessly integrated into the bridge IFC model. The result is a semantically segmented and enriched bridge IFC model. The model includes bridge components (e.g., piers, decks, etc.) segmented and represented by individual *IfcBridgePart* entity instances. Additionally, it is further enriched with semantic information such as span material, deck interaction, etc. This is instrumental in harnessing the full potential of IFC representations.

Existing research primarily focuses on constructing as-built bridge BIM by leveraging scan-to-BIM techniques using as-built point cloud data (Justo et al. 2023; Isailović et al. 2020; Lin and Habib 2022). However, our study shifts the focus toward employing as-designed bridge plans instead of as-built point cloud data. The shift in focus is because of the laborious nature of collecting as-built point cloud data and the inherent limitations of the data collection process, e.g., line-of-sight restrictions. Moreover, the existing bridges already have bridge drawings documented in the governmental database (e.g., Indiana Historic Bridge Inventory) (INDOT 2024). There is a gap in exploring reconstruction of bridge BIMs using as-designed drawings. In the big picture, both scan-to-BIM and our 2D PDF-to-BIM are valuable approaches that supplement each other and provide support to managing bridge projects and assets.

In addition, this study focuses on bridges. Extensive research has been done on BIM reconstruction from PDFs for buildings (Lu et al. 2020; Zhao et al. 2021; Pan et al. 2023; Barreiro et al. 2023; Gimenez et al. 2016). However, bridges have drastically different components from buildings. Instead of doors, walls, windows, beams, and columns in the buildings, bridges comprise unique structural components including piers, bents, abutments, and decks, etc. Consequently, existing approaches or methodologies for semantic segmentation and information extraction developed for buildings cannot be directly adopted for bridge structures. There is a gap in exploring the development of IFC-based BIM for bridges utilizing existing 2D PDF drawings.

Furthermore, existing research concentrated on geometric modeling and as-built point cloud data segmentation (Jiang et al. 2023; Yang et al. 2022; Wang et al. 2023). While the as-built point cloud data are effective in preserving geometric information, they do not address the need for semantic information in the models. For example, as-built point cloud data cannot capture semantic information including object categories (e.g., piers, bents, abutments, railings, and decks, etc.) and component properties (e.g., material type, foundation type). To address this limitation, the authors leverage as-designed point cloud data and propose a framework to generate IFC-based semantically segmented and enriched bridge BIMs using 2D bridge plans.

This transformation from PDF to semantically segmented and enriched bridge BIMs helps enhance multidisciplinary collaboration, facilitate precise clash detection, ensure accurate quantity takeoffs, and promote optimal construction planning and effective asset management (US DOT 2024). Additionally, the IFC-based bridge models can facilitate interoperability among diverse software applications and platforms within the AEC industry (Li and Zhang 2022a, b). The models also support the seamless exchange of data irrespective of the specific software or vendor, thus alleviating potential interoperability issues. The comprehensive utilization of a semantically segmented and enriched bridge IFC model has the potential to deliver sustainable, efficient, and cost-effective results for bridge projects, representing a significant advancement in the field.

Background

BIM for Bridges

BIM plays an essential role in bridge construction projects. In preconstruction applications, 3D visualization of BIM facilitates tasks such as clash detection and detailed design. Conventional bridge drawings separate the illustration of specialized components, complicating the detection of possible infrastructure conflicts in the intricate process of road and bridge construction. Utilizing BIM can align each component with its corresponding parameters using data-driven approaches. The resulting 3D model can undergo a collision check, with the clash counts and specifics automatically displayed (Kaewunruen et al. 2020). A collision detection report is subsequently produced to rectify errors and save time. Beyond component examination, BIM also enables collision detection between machinery, work areas, and structures throughout the construction phase. This contributes to reduced construction expenses, improved risk mitigation, and enhanced construction quality (Song et al. 2023). In addition, the application of BIM extends to post-construction operations, facilitating tasks such as schedule simulation, quality inspections, and emergency/disaster responses. By centralizing the entire bridge construction process on an information-transparent platform, potential construction issues can be predicted, and countermeasures can be developed. This results in a substantial improvement in the efficiency and standard of construction management (Song et al. 2023). Given the extensive lifespan of bridges, the application of BIM is usually in historical bridge contexts (Donato et al. 2017). The goal of such modeling is to create a digital representation that meticulously documents pertinent information. This not only facilitates the preservation of the bridge's cultural importance but also ensures its operation remains safe. Furthermore, this digital model serves as a virtual instrument aiding the formulation of effective remediation strategies (Lee et al. 2020).

IFC, an international open standard for BIM, offers extensive versatility in creating entities, defining properties, and extending relationships (Li et al. 2022a, b; Mandava and Zhang 2016). An illustration of the practicality of the IFC schema is its ability to seamlessly integrate with the building energy modeling (BEM) process, independent of proprietary software platforms (Li and Zhang 2022a, b). Recently, the IFC infrastructure extensions project aims to expand the usage of IFC schemas from the building sector to the infrastructure sector, by providing a unified approach to managing linear infrastructure assets, including roads, railways, bridges, ports, and waterways (BuildingSMART 2023a). This initiative gained urgency by mid-2019, highlighting the need to integrate work across these various domains (BuildingSMART 2023a).

The application of IFC standards to bridge projects becomes crucial. Decisions approved by the AASHTO Board of Directors require all 50 member state DOTs to adopt the IFC specification as the standard for road and bridge design and construction (BuildingSMART 2023d). The adoption ensures a consistent and interoperable framework, simplifying the management and analysis of bridge data associated with other infrastructure elements (BuildingSMART 2023a).

Recent work on using IFC for bridge modeling has taken various approaches. For example, Justo et al. (2023) proposed an automated modeling approach for truss bridges, employing laser scanning data to obtain precise geometric information. Their method utilized a partially instance-segmented point cloud as input and generated an information model adhering to IFC standards, along with a structural graph representing the bridge. Isailović et al. (2020) proposed a novel methodology for detecting spalling damage in point clouds and integrating damaged components into BIM through semantic enrichment. Their approach included generating an as-built BIM, reconstructing damaged point clusters, and enriching the corresponding IFC model. The effectiveness of their multiview classification technique for detecting spalling damage features was evaluated, resulting in a precise as-is IFC model that complies with inspection requirements. However, a common challenge in these studies is the requirement for onsite collection of point cloud data. Sending out people or drones to scan the as-built bridges is labor-intensive and expensive, especially given that there are more than 18,900 bridges in Indiana alone (INDOT 2024). This situation is not unique to Indiana but extends across various states, each grappling with its own set of challenges in bridge data collection and management. In addition, the amount of information that can be collected this way also has limitations, e.g., line-of-sight constraints. However, the existing bridges already have bridge drawings documented in the government bridge information database (e.g., Indiana Historic Bridge Inventory) (INDOT 2024). There is a gap in exploring the development of IFC-based BIM for bridges utilizing existing 2D PDF drawings. In this paper, we address this gap by proposing a novel framework using semantic segmentation, information extraction, and semantic enrichment.

Semantic Segmentation and Enrichment

Considerable research has been conducted on the semantic segmentation and enrichment of bridge components (Justo et al. 2023; Lin and Habib 2022; Yang et al. 2022; Isailović et al. 2020; Wang et al. 2023). Jiang et al. (2023) categorized previous research on semantic segmentation and enrichment for BIM into two primary areas. Firstly, object recognition and classification techniques are commonly utilized during the reconstruction of BIM instance models, as highlighted by Sacks et al. (2018) and Ismail et al. (2016). Secondly, the enrichment of damage and defect information plays a crucial role in facilitating the detection of bridge issues. This involves specifying the type, deterioration process, position, extent, and severity of the damage, as described by Isailović et al. (2020). For example, Lin and Habib (2022) introduced a semantic segmentation framework for bridge components and road infrastructure using mobile light detection and ranging (LiDAR) point clouds. Their approach incorporated deep learning techniques, including graph convolutions, to achieve accurate semantic segmentation. They also developed cross-labeling and transfer learning techniques to minimize manual annotation efforts and proposed geometric quality control strategies to refine the segmentation results. Yang et al. (2022) presented a weighted superpoint graph method for classifying bridge components from large-scale point cloud data. Their approach involved clustering point clouds into semantically homogeneous superpoints and employing PointNet and graph neural networks for classification. Their model demonstrated an overall accuracy of 99.43%. Lee et al. (2020) developed a system for the automatic extraction of bridge geometric parameters, aimed at minimizing the time and resources required in the scan-to-BIM process. Lee et al.'s (2020) proposed parameter extraction system leveraged the m-estimator sample consensus (MSAC) algorithm to decompose a 3D bridge point cloud model into planar components. In this way, the geometric parameters (e.g., length of the bridge, distance between beams) can be extracted based on different bridge structure types.

Studies on the generation of 3D models from 2D drawings have shown promising progress. For example, Zhang et al. (2023) proposed an autonomous system that can convert 2D orthographic drawings into 3D computer-aided design (CAD) models. The effectiveness of the technique was evaluated against a large-scale public data set comprising thousands of 3D shapes, achieving an F1 score of 99.59% in precise model reconstruction. Poku-Agyemang and Reiterer (2023) developed an inventive method of converting 2D blueprints of bridges into digitized versions, using image processing to pinpoint corners and recreate 3D point clouds of bridge components, which were then assembled into a scaled 3D entity. The produced point cloud model could serve as a supplement to as-built laser scanning or camera data. Nevertheless, their study did not consider semantic information.

However, previous studies have been primarily focused on the geometric aspects of the bridge and point cloud data segmentation. Point cloud data preserve geometric information effectively but lack crucial semantic information. For example, object categories and the composition of construction elements (e.g., span material, span type, and foundation type, etc.), cannot be represented in point cloud data. Consequently, point cloud models fail to provide the necessary contextual details to reconstruct semantic BIM. To address this limitation, the authors leverage as-designed point cloud data and propose a framework to generate IFC-based semantically segmented and enriched bridge BIMs using 2D bridge plans. The generated as-designed BIMs can support engineering and management analysis of design, construction, operation, and maintenance of bridges (Macher et al. 2017).

In the big picture, both as-designed and as-built bridge BIMs are essential; they supplement each other and provide support to managing bridge projects and assets. For example, the as-built model provides precise geometric data that reflect the bridge's current state, capturing any deviations from the original design. The as-designed model offers detailed semantic information, such as the span material and foundation type, essential for understanding the bridge's intended functionality and characteristics. In addition, the as-built model offers real-time data for performance analysis and predictive maintenance. Complementing this, the as-designed model provides historical design data, enabling advanced simulations and analysis for future performance and potential upgrades. The integration ensures a comprehensive and accurate representation of the bridge, and facilitates a deeper understanding and proactive management of the bridge's life cycle.

Specifications for the National Bridge Inventory (SNBI)

This study leverages the Specifications for the National Bridge Inventory (SNBI) as a benchmark for identifying semantic information about bridges, given its provision of standardized definitions and formats for bridge data items (US DOT 2022). The SNBI emphasizes that "standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information," according to US DOT (2022). State DOTs, federal agencies, and tribal governments report their collected bridge inventory information to the Federal Highway Administration (FHWA) in alignment with the National Bridge Inspection Standards (NBIS) reporting requirements. These data are subsequently housed within the National Bridge Inventory (NBI) database. This centralized database not only facilitates state and national analytical endeavors but also facilitates federal funding programs (US DOT 2022).

A critical aspect of SNBI is its focus on accurately evaluating bridge conditions at both the component (such as deck, superstructure, substructure, and culvert) and element levels (US DOT 2022). This ensures the efficient management of bridge inventories and contributes to the safety and efficacy of the highway transportation system.

The SNBI delineate the following seven primary categories for organizing data items essential for constructing and maintaining a comprehensive bridge inventory database (US DOT 2022, 2023b): (1) bridge identification; (2) bridge material and type; (3) bridge geometry; (4) features; (5) loads, load rating, and posting; (6) inspections, and (7) bridge condition. Each category includes multiple data items, and each item exhibits a one-to-one relationship with the individual bridge. Therefore, by leveraging SNBI, studies can ensure consistency in data representation, making it easier for different parties to understand, share, and compare bridge data.

The SNBI provides a robust framework for bridge data representation. However, current research on utilizing its specifications for semantic information enrichment in bridges remains limited. This research aims to address a real need of conforming to the federal SNBI for information extraction, a federal mandate with a deadline of March 15, 2028 (US DOT 2023a).

Optical Character Recognition (OCR)

OCR is a transformative technology that has revolutionized the way machines perceive and process textual data in digital documents. Essentially, OCR is a computerized approach that converts characters present in different types of documents, including PDF documents, scanned papers, or images, into searchable and editable data. It involves processes like noise reduction, feature extraction, and subsequent classification (Zhao et al. 2021). In the context of civil engineering and infrastructure management, extracting accurate data from 2D drawings and PDF plans is crucial. It ensures accurate modeling, analysis, maintenance planning, and quality assurance (QA) and quality control (QC).

Previous research efforts have explored the use of OCR to extract semantic information from 2D building drawings. For example, Lu et al. (2020) introduced a semiautomated methodology aimed at building a systematic, precise, and efficient digital twinning system, based on CAD drawings and images. In Lu et al.'s (2020) methodology, OCR was leveraged to transform typed, handwritten, or printed text stored in image formats into machine-readable text. Zhao et al. (2021) proposed a hybrid method to extract semantic information of building objects from 2D structural drawings. By leveraging OCR, semantic information including annotations tied to specific pixel coordinates of building objects was extracted. Subsequently, these extracted annotations were categorized into dimensions and attributes of the respective objects, contingent on their spatial distinctions. Their methodology achieved high precision and recall across different building object classes, with F1 scores for all object types exceeding 85%. Pan et al. (2023) proposed a method that expeditiously recovered BIMs for two mechanical, electrical, and plumbing (MEP) systems in just 2.85 s and 0.79 s, respectively. Their method achieved a semantic extraction accuracy that exceeds 90%.

Numerous OCR tools are available in the market, including Google Vision AI, Google Document AI, Pytesseract, and Nanonets, among others. One notable recent development is ChatPDF, which is gaining widespread adoption. ChatPDF provides a human-interactable interface for information extraction (ChatPDF 2023). By taking bridge plans as input, users can pose questions to the prompt and receive answers. For example, when asked "What is the span length?," the system returned the correct answer of 38 ft, 6 in. and 69 ft, 6 in. as indicated in the bridge plan. More examples can be found in Fig. 1.

Although ChatPDF provides a user-friendly interface and excels at extracting tabular information, it lacks the capability to extract the specific SNBI items such as deck interaction or minimum span length. These items are not directly indicated in the bridge plans and require specialized civil engineering knowledge for accurate extraction. Furthermore, ChatPDF only provides an end-to-end application programming interface (API), without the ability to tune parameters and further processing of the extraction task. Therefore, a more adaptable and versatile OCR tool is needed in this study to extract SNBI information from bridge plans.

Proposed Framework

Existing research focused on BIM reconstruction for buildings (Lu et al. 2020; Zhao et al. 2021; Pan et al. 2023) and as-built BIM reconstruction for bridges leveraging scan-to-BIM techniques (Justo et al. 2023; Isailović et al. 2020; Lin and Habib 2022). There is a lack of study in as-designed BIM reconstruction for bridges using as-designed drawings. Therefore, in this study, the proposed framework aims to generate semantically segmented and enriched 3D bridge BIM instance models using 2D bridge plans. Fig. 2 shows the overall proposed framework, which comprises four main phases: 3D geometry reconstruction, semantic segmentation, information extraction, and semantic enrichment, respectively. First, the 3D geometry reconstruction phase takes a bridge PDF plan as input and generates a 3D geometric mesh model in the .obj format using the PDF2BIM technology, developed by Akanbi and Zhang (2022a, b). Second, semantic segmentation is carried out to decompose the 3D geometric model into distinct bridge components (e.g., decks and piers). By leveraging the point cloud model and applying segmentation algorithms on it, a semantically segmented mesh model is generated. Third, information extraction is conducted to automatically extract the semantic information associated with the bridge and its components from the bridge PDF plan. Finally, the semantic enrichment phase maps each segmented bridge component to its corresponding IFC representation, generating a segmented IFC model. Extracted semantic information is then integrated

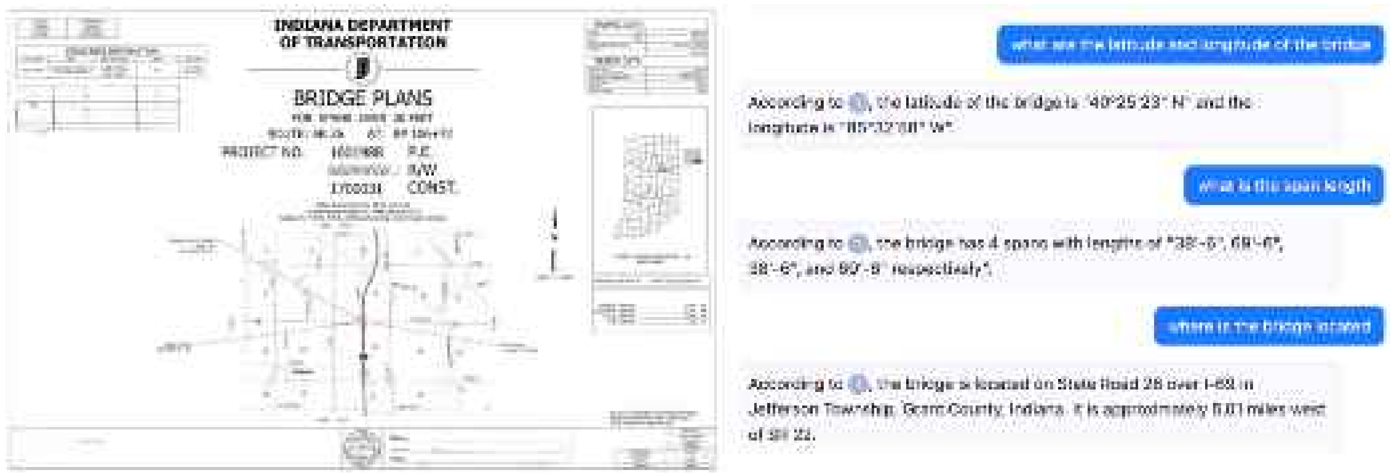

Fig. 1. Interface of ChatPDF and example questions with corresponding answers (ChatPDF 2023). (Base map courtesy of Indiana Department of Transportation, Indianapolis.)

into the IFC model, resulting in a semantically segmented and enriched final model.

3D Geometry Reconstruction

Prior studies have made significant advancement in developing innovative algorithms, such as PDF2BIM algorithms, to facilitate the semiautomated and more efficient reconstruction of 3D models for bridges using 2D PDF drawings (Akanbi and Zhang 2022a, b). PDF2BIM can extract the geometric information in the orthographic drawings (including elevation view, plan view, and typical sections view) from 2D bridge plans and convert it to the 3D mesh model (.obj).

The mesh model serves as the geometric basis for constructing IFC-based BIM instance models of a bridge. A mesh model is composed of vertex data section and face data section. The vertex data section defines the 3D coordinates (x, y, z) for each vertex in the mesh. The face data section defines the index of vertices that constitute the faces or polygons. Fig. 3 illustrates an example IFC geometric model represented by *IfcTriangulatedFaceSet*. Essentially, an *IfcTriangulatedFaceSet* entity consists of an "IfcCartesianPointList3D" attribute and a "CoordIndex" attribute, which correspond to vertex data and face data in the mesh model, respectively. Therefore, *IfcTriangulatedFaceSet* shares the same data structure as meshes, facilitating easy bidirectional transformation.

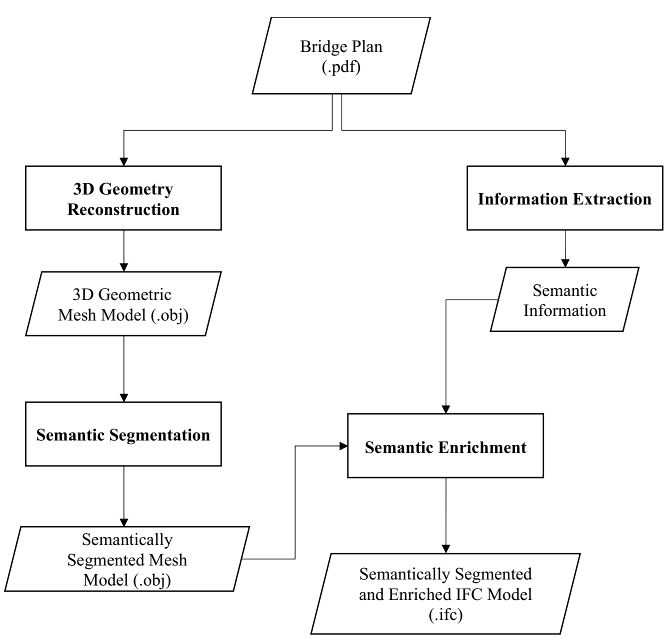

Fig. 2. Proposed framework for generating the 3D semantically segmented and enriched IFC model using 2D bridge plans.

Semantic Segmentation

In this phase, semantic segmentation is performed to segment the 3D geometric mesh model into different components of the bridge. Fig. 4 illustrates the detailed process of semantic segmentation. Essentially, it takes a 3D mesh model as input and outputs the semantically segmented mesh model. The mesh model presents challenges for direct semantic segmentation using clustering algorithms due to its complex structure of interconnected triangular surfaces. Many well-established clustering algorithms work on point clouds directly. Therefore, to facilitate more effective semantic segmentation, the mesh model is initially converted into a point cloud model. The point cloud model represents the object as a dense set of discrete points in the 3D space. Following this conversion, clustering algorithms are applied to the point cloud model to generate a semantically segmented point cloud model.

Once the point cloud model has been semantically segmented, it becomes necessary to transform it back into a mesh model. This phase is critical for the generation of the segmented IFC model, which is the final output of this framework. However, generating a mesh model from a point cloud model can be challenging, which involves surface reconstruction (Wang et al. 2019; Jin et al. 2023). The choice of algorithms depends on the specific characteristics of the point cloud and the desired properties of the resulting mesh. Furthermore, generated meshes may require additional optimization to refine the shape, improve smoothness, and reduce artifacts such as holes or intersections.

Therefore, to transfer the clustering results from the point cloud model to the mesh model, the boundary segmentation technique is applied. Bounding boxes are generated based on the 3D coordinates of the segmented clusters within the point cloud. These bounding boxes are then mapped onto the original mesh model, effectively serving as boundary filters to isolate and segment corresponding

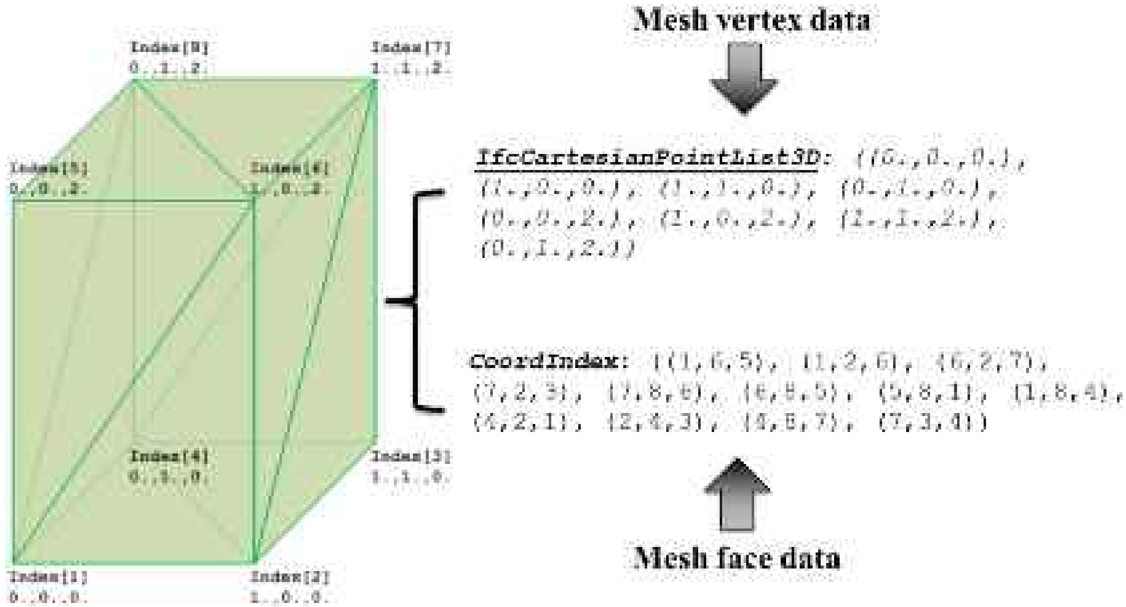

Fig. 3. Example IFC geometric model represented by *IfcTriangulatedFaceSet* (BuildingSMART 2023b).

regions within the mesh. The mesh model is then divided into distinct, semantically meaningful segments, generating a semantically segmented mesh model.

Information Extraction

To produce a semantically enriched bridge model, the 3D geometric model needs to be augmented with semantic information. This information, such as foundation type, concrete strength, and span material, is primarily embedded in the textual data found in bridge plans. In this phase, the SNBI data items are extracted from bridge plans.

To extract SNBI data items of bridges that are described in PDF documents containing textual or numerical data, automatic semantic information extraction algorithms are developed using data-driven approaches. PDF plans contain searchable and selectable text that can be transformed into American Standard Code for Information Interchange (ASCII) data format. These ASCII data can be processed with NLP-based approach leveraging rule-based patterns such as regular expressions (RegEx). However, a challenge in the preservation of table structures arises during the transformation from PDFs to ASCII data. Without these structures, applying RegEx matching to extract tabular information becomes cumbersome due to difficulties in pinpointing the exact location of the target data. This challenge is exemplified in Fig. 5, which presents a table indicating the bridge type as a "continuous composite steel beam bridge." If this PDF plan is directly parsed to ASCII data using PDF parsing tools such as PyPDF, identifying such information under the column name TYPE becomes challenging. This is because the table's structural context is lost. To address this, OCR is leveraged in the development of information extraction algorithms, to discern the column and row patterns in the table. With this approach, the image-based table is then converted to a structured format, like a .csv file, ensuring the preservation and subsequent extraction of the data.

The bridge plans are divided into the development set and the testing set. The plans in the development set are used to iteratively develop the information extraction algorithms using a data-driven approach. The testing set is used to test the accuracy and time efficiency of the developed algorithms. Fig. 6 illustrates the proposed method for the development of information extraction algorithms in the iterative data-driven approach. The detailed step-by-step method is presented in the following subsections.

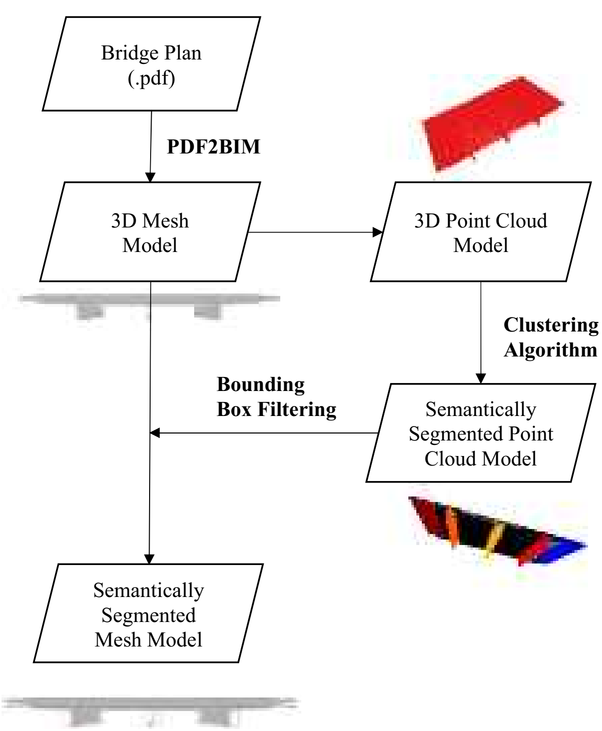

Fig. 4. Detailed process of semantic segmentation.

Target Data Item Identification

The target semantic information that exists in the bridge plan is identified based on SNBI. SNBI is an open standard including seven primary categories for organizing data items essential for constructing and maintaining a comprehensive bridge inventory

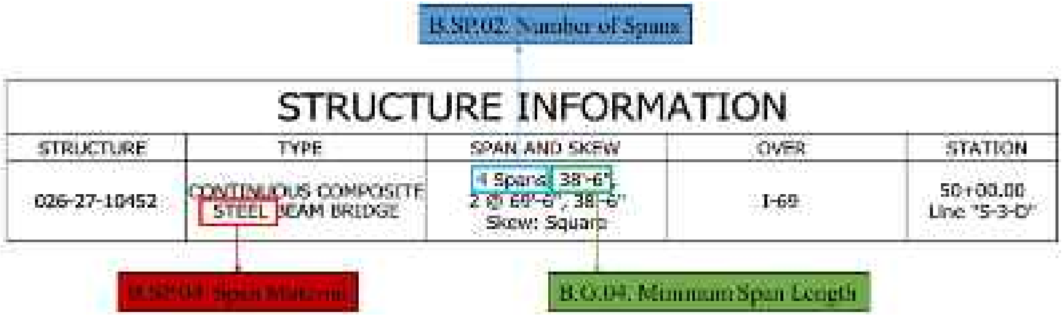

Fig. 5. Example SNBI items and corresponding extraction results from a bridge plan.

database (US DOT 2022, 2023b). Fig. 5 illustrates example SNBI items (span material, number of spans, and minimum span length) and their corresponding locations shown in an example bridge plan.

OCR Integration

A comparative evaluation of various OCR tools is conducted in terms of extraction efficiency, accuracy, and cost. Notably, the OCR tool with an accessible and flexible API is prioritized for selection. This is necessitated by the requirement to customize the extraction results to match specific SNBI data items. Following that, the PDF bridge plan from the development set is converted into an image format to facilitate OCR. Then, the tables and texts containing semantic information of the bridge (and its components) in the image are recognized and extracted leveraging the OCR technique.

Iterative Extraction Rule Development

In this step, the extraction rule is developed iteratively. If the target data item is contained in tables in the bridge plan, the target table and corresponding cell in the table containing target semantic information are identified and extracted. Otherwise, if the target data item is described in the text descriptions of the bridge plan,

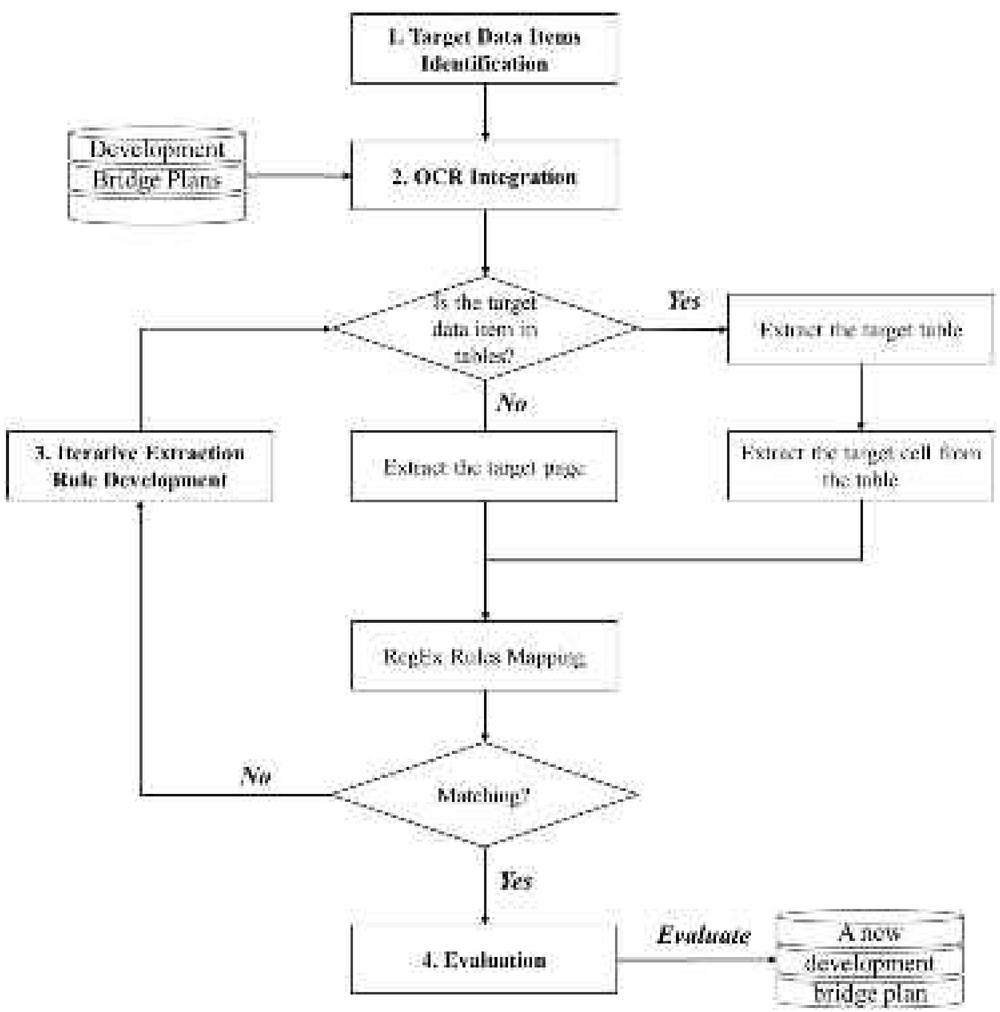

Fig. 6. Proposed data-driven method for iteratively developing information extraction algorithms.

the target page containing the target data item is located. Following that, RegEx rules are developed and applied to identify the exact target data item. If the RegEx rules correctly extract the target data item, the development at this iteration is finished. Otherwise, new RegEx patterns and the corresponding extraction rule are formulated and encoded into the information extraction algorithms based on the target data item. This iterative process is continued until the refined algorithm successfully extracts the target data items from the current bridge plan.

Evaluation

At each step during the iterative development process, a unique bridge plan from the development set is leveraged to support the development and refinement of the algorithm. Diverse sets of target information patterns are then integrated into the algorithm. Following that, the refined algorithm at each step is evaluated with a new, different bridge plan from the development set. For example, in the initial iteration, the algorithm is initially developed based on the first bridge plan and subsequently tested on the second bridge plan. After that, it will be refined based on the second bridge plan and tested on the third, etc. This iterative process is designed to gradually enhance the coverage and adaptability of the algorithm following a data-driven approach.

Upon completing the iterative development using all bridge plans in the development set, the final algorithms are compiled and evaluated using the plans in the testing set.

Semantic Enrichment

Semantic enrichment comprises two substeps: (1) IFC model construction; and (2) semantic information integration. In compliance with BuildingSMART (2023c) IFC standard Version 4×2, the IFC bridge model incorporates *IfcBridge* and *IfcBridgePart* entities.

IFC Model Construction

In the first substep, as illustrated in Fig. 7, the segmented mesh model is transformed to an IFC model leveraging geometric representations in IFC (e.g., *IfcTriangulatedFaceSet*). The geometric information of various components of the bridge is mapped to *IfcBridge* and *IfcBridgePart* instances through the "Representation" attribute. The "PredefinedType" attribute of *IfcBridge* and *IfcBridgePart* specifies the type of the bridge and the type of bridge component, respectively. For example, the bridge types include arched, culvert, or girder, etc., and the bridge component types include deck, foundation, or pier, etc. The attribute "CompositionType" distinguishes whether the predefined spatial structure element is representing itself (denoted by ELEMENT), or if it is a collection or aggregation of similar elements (denoted by COMPLEX), or simply a part of the whole instance (denoted by PARTIAL). The attribute "PropertySet" is in relation to semantic information, which is explained in the next substep.

Semantic Information Integration

In the second substep, the semantic information extracted from the bridge plan is seamlessly integrated with the generated IFC model. As illustrated in Fig. 8, the property sets of an IFC entity instance are established using *IfcRelDefinesByProperties* (BuildingSMART 2023c). For bridges, the *IfcPropertySet* can be categorized into various groups, such as location, span information, and deck information, etc. Each *IfcPropertySet* has the flexibility to include multiple *IfcPropertySingleValue* instances, allowing for customization and assignment of diverse attributes or properties to a specific bridge entity instance in IFC. For instance, the *IfcPropertySet* for location may include information like latitude, longitude, and any other relevant descriptions related to the bridge's location. *IfcRelDefinesByProperties* and *IfcPropertySet* provide a comprehensive and flexible approach to attribute/property customization. As a result, a semantically segmented and enriched IFC model is generated.

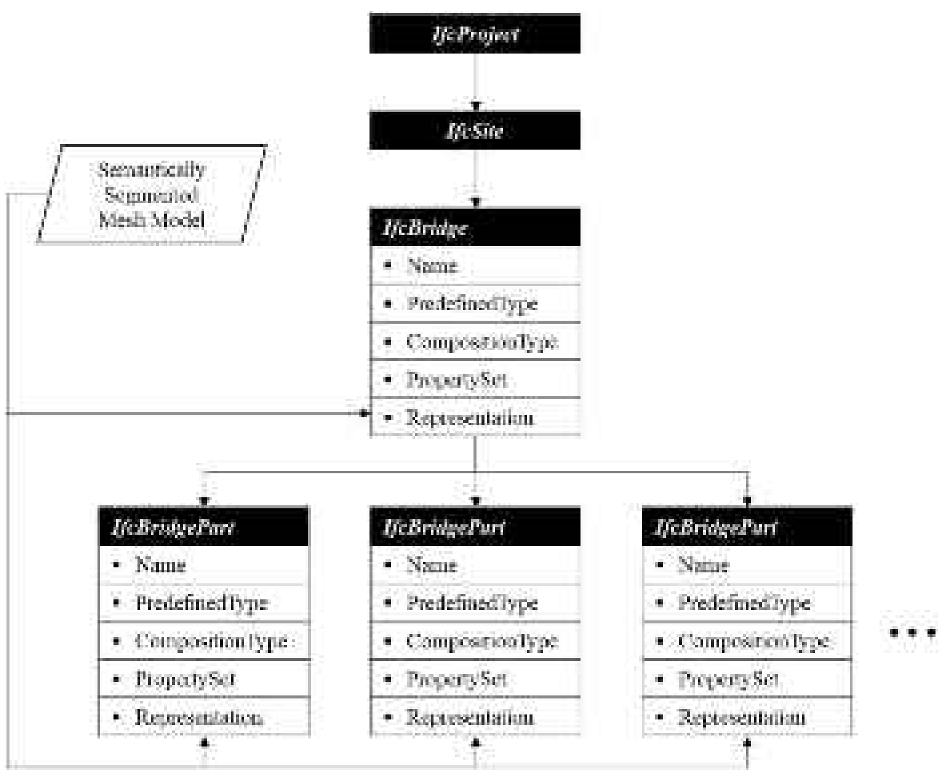

Fig. 7. Entity structure and attributes of *IfcBridge* and *IfcBridgePart* for modeling bridges in IFC.

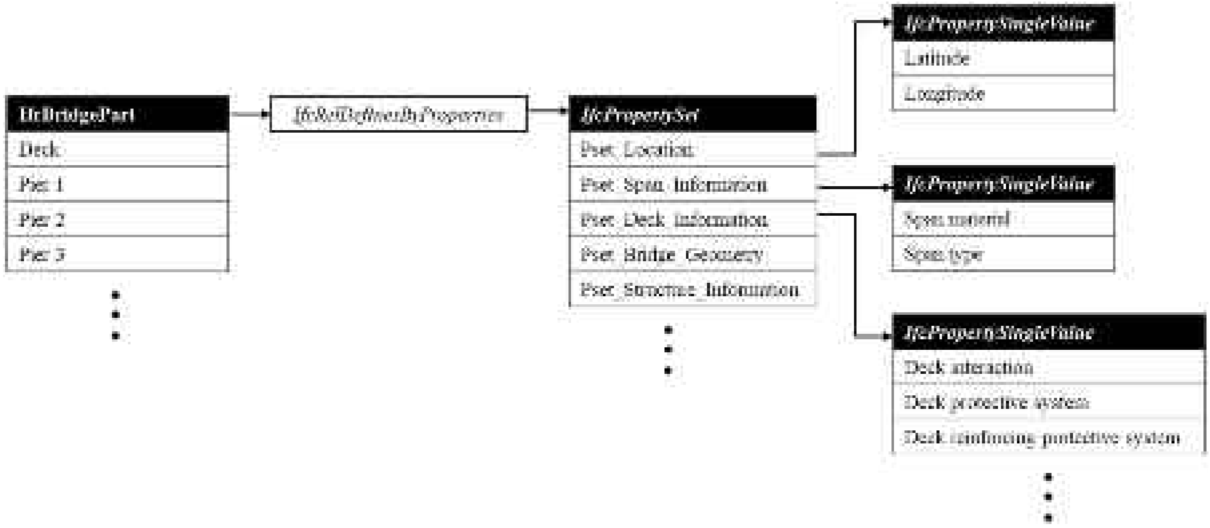

Fig. 8. Semantic information integration leveraging *IfcRelDefinesByProperties* and *IfcPropertySet* in IFC schema (partial).

Experiments

In this section, experiments were carried out to evaluate the proposed framework. PDF drawings for 12 bridges located in various parts of Indiana were collected. The PDF plans contain both geometric drawings (e.g., elevation views, plan views, and typical sections) and semantic information (e.g., span type, span material, and the number of spans, etc.). The collected 12 bridge plans were then divided into a development set and a testing set randomly. Six plans were allocated for the development of information extraction algorithms. The remaining six were reserved for testing the accuracy of the developed algorithms.

3D Geometry Reconstruction

First, 3D geometry reconstruction was conducted, which leveraged PDF2BIM, a tool developed by Akanbi and Zhang (2022a, b). An API was developed within the Flask web development framework (Akanbi and Zhang 2022a). This API accepts 2D orthographic drawings of the bridge as input and proceeds to parse the resulting mesh file for rendering within a web browser using ThreeJS. Illustratively, Figs. 9(a and b) show an elevation drawing example of a bridge situated on Indiana State Road 26 and its corresponding mesh model (.obj) derived from the conversion process. It can be seen that the overall geometry of the bridge is successfully reconstructed. The 3D geometry reconstruction using as-designed drawings laid the foundation for semantic segmentation and semantic enrichment of bridge BIMs. Integrated with the as-built bridge model, it can provide a comprehensive and detailed representation of the bridge, combining precise geometric data with rich semantic information. This integration enhances the overall accuracy and completeness of the BIM, allowing for better maintenance, analysis, and decision-making processes by leveraging both the current state of the bridge and its original design specifications.

Semantic Segmentation

At this phase, semantic segmentation algorithms were developed to segment the 3D geometric mesh model into distinct components of the bridge. First, the mesh model is converted to the point cloud

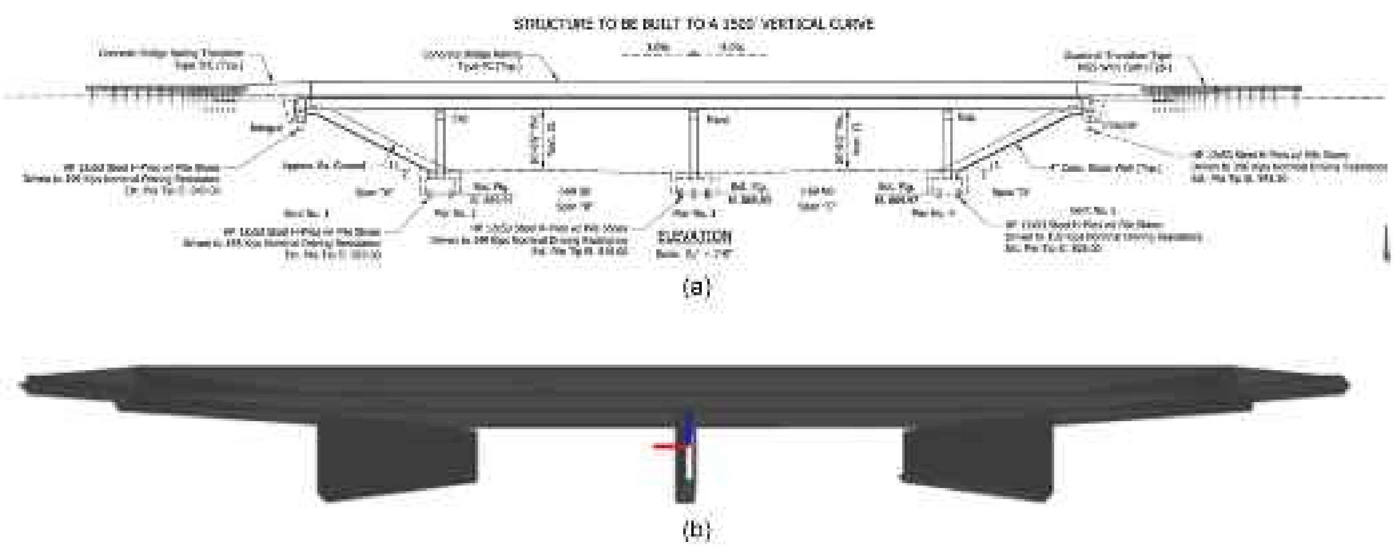

Fig. 9. (a) An example plan of a bridge on Indiana State Road 26; and (b) converted mesh model using PDF2BIM.

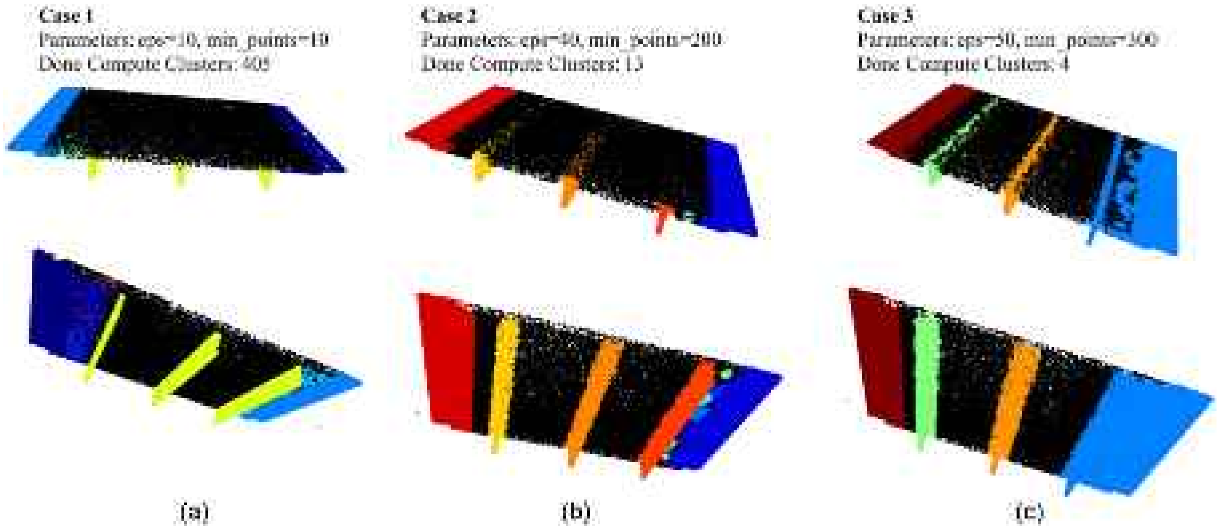

Fig. 10. Example clustering results with parameter tuning using the DBSCAN algorithm.

model by generating sampling points uniformly on each mesh surface in Python. The distributed set of points as a result provides a representative subset of the mesh's geometry, which can be further segmented to provide more reliable results. Then, the density-based spatial clustering of applications with noise (DBSCAN) algorithm is leveraged for semantic segmentation of the point cloud model.

Essentially, the DBSCAN algorithm takes two parameters that can be adjusted to customize the clustering behavior. The first one is epsilon (*eps*), which specifies the maximum distance between two points for them to be considered as neighbors. Larger values of epsilon result in more points being grouped together in a single cluster. The second parameter is *min_points*, which determines the minimum number of points required to form a dense region. Increasing *min_points* leads to fewer clusters and more noise points (outliers). The semantic segmentation process using the DBSCAN algorithm is semiautomatic. It requires parameter tuning (i.e., incremental adjustments of *eps* and *min_points*) and human visual inspection to obtain the optimal clustering results.

Decks and piers, two essential bridge components, were targeted for semantic segmentation in this study. Fig. 10 illustrates the example clustering results of three parameter sets including different values of *eps* and *min_points* using the DBSCAN algorithm. It can be seen that when *eps* = 10 and *min_points* = 10, the point cloud model was classified into 405 clusters, and three piers were grouped into the same cluster [Fig. 10(a), Case 1]. By increasing the *eps* and *min_points* to 40 and 200, respectively, the number of computed clusters decreased to 13 [Fig. 10(b), Case 2]. Three piers were classified into distinct clusters, respectively. However, the deck was segmented into multiple clusters. To cluster the deck, the pier clusters were excluded from the entire bridge point cloud model and the remaining part was the deck. Continuing to increase the *eps* and *min_points*, the pier on the right side was classified as the same cluster with a part of the deck [Fig. 10(c), Case 3]. As a result, Case 2 was selected as the optimal solution by manually observing the clustering result. The segmented point cloud model was then transformed into the segmented mesh model (.obj).

The overall performance of DBSCAN can be evaluated by computing the average silhouette coefficient across all data points. The silhouette coefficient measures the compactness and separation of clusters. It computes a score for each data point based on its distance to other points within the same cluster (intracluster distance) and its distance to the nearest neighboring cluster (intercluster distance). Higher silhouette coefficients (ranging from -1 to 1) indicate better-defined and well-separated clusters. Table 1 shows different parameter tuning results and their corresponding silhouette scores.

Although the silhouette coefficient is widely used for evaluating the clustering results, it is not sufficient in this case due to the domain-specific requirements for bridge segmentation. Therefore, a visual inspection of the clustering results is still required to obtain the optimal results. Fig. 11 showcases three examples of clustering results from the collected bridge plans using the DBSCAN algorithm with different parameter settings.

Information Extraction

In this phase, information extraction algorithms were developed following the iterative data-driven approach detailed in the "Proposed Framework" section.

Target Data Items Identification

First, SNBI items were identified in the bridge plan in five distinct sections, as shown in Table 2: (1) bridge identification; (2) bridge material and type; (3) bridge geometry; (4) features; and (5) loads, load rating, and posting. These identified SNBI items were set as the foundational targets for developing the information extraction algorithms using the development plan set. For the data items in sections (6) inspections and (7) bridge condition, this information typically requires field inspections or detailed inspection reports and are not typically present in standard bridge plans. Therefore, these sections were not considered in the scope of this study.

OCR Integration

Old bridge plans are created using scanners and represented in raster images. These image-based plans contain textual information

Table 1. Clustering results with different parameters using the DBSCAN algorithm

Case No.	*eps*	*min_points*	Number of clusters	Silhouette score	Successful segmentation by visual inspection
1	10	10	405	−0.667	No
2	40	200	13	−0.022	Yes
3	50	300	4	0.353	No

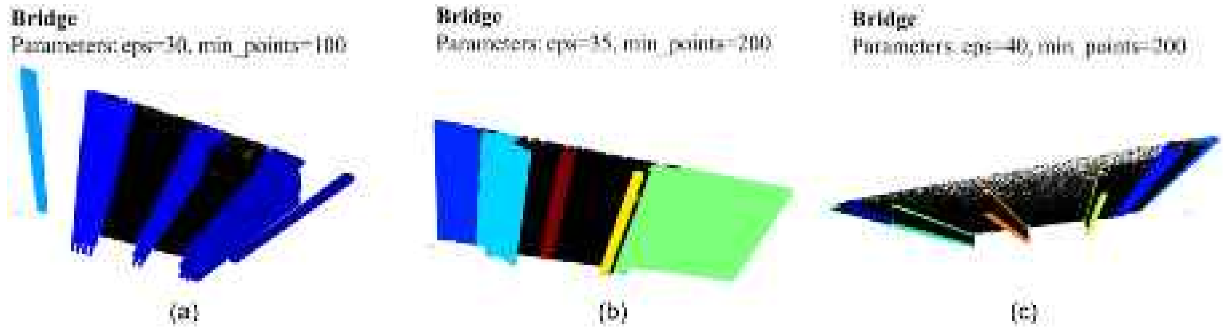

Fig. 11. More examples of clustering results from the collected bridge plans using the DBSCAN algorithm with different parameter settings.

Table 2. SNBI data items and the corresponding RegEx rules developed for the information extraction algorithms

SNBI ID	Data item name	Category	SNBI specification (US DOT 2022)	RegEx patterns or rules
B.L.05	Latitude	Bridge identification	Latitude of the bridge in decimal degrees	(\d{2})(° *)([0-9]+)(')(*)([0-9]+)(.)(*)(W|E)
B.L.06	Longitude	Bridge identification	Longitude of the bridge in decimal degrees	(\d{2})(° *)([0-9]+)(')(*)([0-9]+)(.)(*)(N|S)
B.SP.02	Number of spans	Bridge material and type	The number of spans	\b[0-9] Spans\b
B.SP.04	Span material	Bridge material and type	The principal span material type such as concrete, composite, iron, etc.	Keywords: ['concrete', 'composite', 'iron', 'masonry', 'plastic', 'steel'] RegEx patterns: "W\d+ x \d+" or "C\d+ x \d+"
B.SP.06	Span type	Bridge material and type	The span type such as arch, frame, girder, etc.	Keywords: [arch, frame, girder, beam]
B.SP.08	Deck interaction	Bridge material and type	The specific type of interaction between the superstructure and deck in the span configuration	Keywords: ['composite', 'tee-beam', 'non-composite']
B.SP.11	Deck protective system	Bridge material and type	The protective system implemented for the deck in the given span configuration	Keywords: ['admixture', 'coating', 'membrane', 'patina']
B.SP.12	Deck reinforcing protective system	Bridge material and type	The type of deck reinforcing protective system for the span configuration for concrete decks and slabs	Keywords: ['epoxy coated', 'galvanized', 'metalized']
B.SB.06	Foundation type	Bridge material and type	Foundation type of the bridge such as footing, pile, etc.	Keywords: ['footing', 'pile', 'drilled shaft']
B.G.01	Bridge length	Bridge geometry	The bridge length to the nearest tenth of a foot measured along the roadway centerline	\d+\.\d{3} MI
B.G.04	Minimum span length	Bridge geometry	The length of the minimum span to the nearest tenth of foot, measured from centerline of bearing to centerline of bearing, along the roadway centerline	Pattern 1: \d+-\d+ Pattern 2: \d+(?: '|') - \d+
B.RT.01	Route designation	Features	This item records the designations for routes that utilize the reported highway feature	Data described in the column "OVER"
B.RT.03	Route direction	Features	Record the specified route direction	Keywords: ['NB', 'SB', 'WB', 'EB']
B.LR.01	Design load	Loads, load rating, and posting	Record the live load for which the bridge was designed	Keywords: ['H-10', 'H-15', 'H-20', 'HS-15', 'HS-20', 'HL-93']
N/A	Concrete strength	Loads, load rating, and posting	Concrete strength of the bridge	Pattern 1: (Class ")(A|B|C)(" Concrete) Pattern 2: (f\'c *= *)(\d+,\d+)()(p.s.i.|psi) Pattern 3: (?<=concrete strength)(.*?)(?=psi)

that is not inherently searchable and extractable. Therefore, to extract essential information from both image-based plans and text-based bridge plans, OCR technology was utilized to convert images into tables and texts.

To implement OCR, several OCR tools were evaluated, including Google Vision AI, Google Document AI, OCR with Pytesseract, OpenCV, Tesseract OCR, ChatPDF, and Nanonets. Upon comparative testing, it was determined by the authors that Nanonets outperformed its counterparts in this context because Nanonets provides a highly functional and easily accessible API for the extraction of information from bridge plans (Nanonets 2023). As an OCR tool, Nanonets leverages computer vision algorithms along with artificial intelligence (AI) to extract data from images (Nanonets 2023). When extracting information from tables in PDF plans, it can effectively detect columns, rows, and cells utilizing OCR technology. Consequently, Nanonets was selected for extracting all tabular and textual information contained within bridge plans.

Next, the PDF plans from the development set were converted to images using Python *pdf2image* library for the input of OCR model (Belval 2022). Then, tables and texts in image format were extracted using OCR. The input is the bridge plan in image format (.png) and the output is a spreadsheet file (.csv). An API key

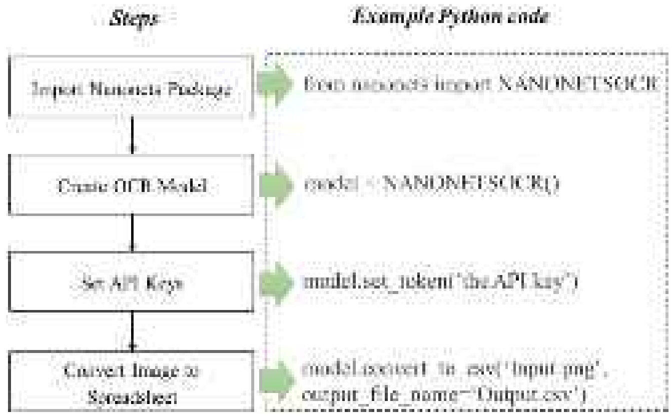

Fig. 12. Steps and example code snippet for using Nanonets API.

can be generated and obtained from the Nanonets account. The output spreadsheet file will be automatically generated including all the tables from the bridge plans. The detailed steps and the code snippet for using the Nanonets API in Python are shown in Fig. 12.

The OCR model in Nanonets is fine-tuned by adjusting the position and size of the grids (including columns and rows) meticulously to align with the tables and contents present in the image. Any superfluous tables that do not contribute to the process of extracting information are manually eliminated from the image. Fig. 13 exemplifies a table from the bridge plan and its transformed version in the spreadsheet, obtained using Nanonets OCR model. It can be seen that the text transformation was accurate and the table structure was correctly preserved (i.e., columns and rows were precisely recognized).

Iterative Extraction Rule Development

After all the information was transformed to searchable text with the table structure preserved, further processing was required to match the extracted data to specific SNBI items. First, the location of the target data item was determined by manual observation or searching through the bridge plan. If the target data item was in a table, the target table and target cell containing the target data item was located by searching the table name and column name. If the target data item was in text descriptions of the plan, the target page containing the target data item was located. Subsequently, RegEx rules were formulated based on the extracted target data item from the six distinct bridge plans from the development set iteratively. For example, the RegEx rule "\b[0-9] Spans\b" was used to extract the number of spans from the table in the bridge plan, as shown in Table 2. This rule shows generality by extracting the number of spans from all six distinct plans. The extraction rules for the 15 SNBI items were then collected and integrated into the information extraction algorithms following the iterative data-driven approach. Table 2 summarizes the developed RegEx rules for extracting the 15 SNBI items using the development set.

Evaluation

The information extraction algorithm developed after each iteration was then evaluated with a new, different bridge plan from the development set. Fig. 14 illustrates the performance improvement of the algorithm at each iteration in the development process. It can be seen that after five iterations, the information extraction algorithm achieved 93.3% validation accuracy on the development set, as shown in Table 3.

Semantic Enrichment

In this phase, semantic enrichment was performed. Algorithms were developed using IfcOpenShell (i.e., an open-source software development library for IFC in Python) for implementing the automation of this process (IfcOpenShell 2023).

IFC Model Construction

First, the IFC model was created based on the segmented mesh model. Fig. 15 illustrates the process of transforming the semantically segmented mesh model to the IFC model. The bridge's different components (as classified into different clusters in the mesh model) were mapped to *IfcBridgePart* instances, including decks and piers. The mapping process leveraged *IfcTriangulatedFaceSet* instances to transform the mesh geometry representations to IFC geometry representations.

Semantic Information Integration

Second, the extracted semantic information from the bridge plan was automatically integrated with the 3D IFC model through the use of *IfcPropertySet*, as depicted in Fig. 8. This step completed the process of producing a semantically segmented and enriched 3D bridge IFC model. Fig. 16 provides a visualization of this enriched bridge IFC model using BIMvision. In the depiction, the green

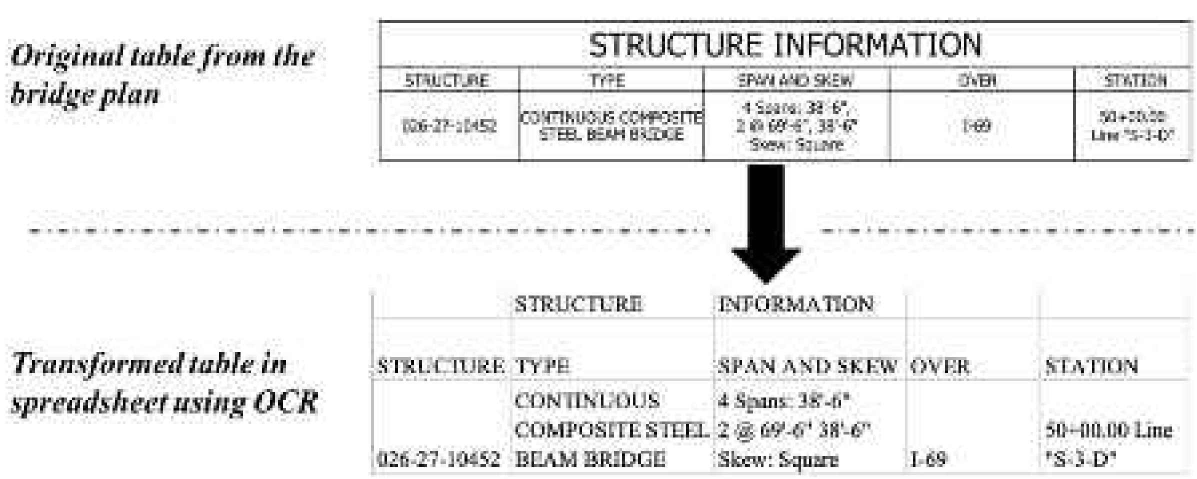

Fig. 13. Example table from the bridge plan and transformed table in spreadsheet format using Nanonets API.

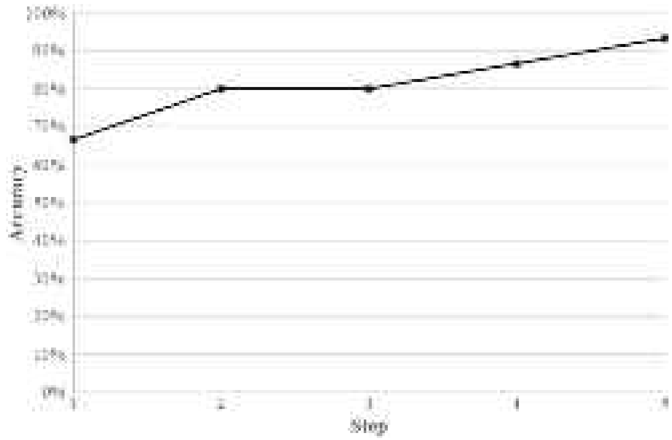

Fig. 14. Performance improvement of the information extraction algorithm through iterative training using the development set.

color highlight indicates the selected entity instance. Fig. 16(a) shows the *IfcBridge* instance, with its pertinent semantic information displayed on the right-hand panel of the interface. Fig. 16(b) highlights the *IfcBridgePart* instance with its "PredefinedType" as deck. Lastly, Figs. 16(c–e) show the three distinct piers.

Results and Discussion

The results were evaluated using the testing bridge plan set following these two criteria: (1) the accuracy performance of the developed information extraction algorithms; and (2) the time savings of the proposed automated approach.

Table 3. Performance improvement of the information extraction algorithm through iterative training using the development set

Step	Total number of information elements	Number of correctly extracted elements	Accuracy (%)
1	15	10	66.7
2	15	12	80.0
3	15	12	80.0
4	15	13	86.7
5	15	14	93.3

Information Extraction

The information extraction algorithm developed through the iterative approach using the six different bridge plans was subsequently tested on the testing set (i.e., six independent bridge plans). The algorithm demonstrated an overall 97.7% precision and 94.4% recall on extracting 90 SNBI data items from the six new bridge plans, as detailed in Table 4. Within this table, specific extraction results are provided for each SNBI category, along with the corresponding accuracy metrics for the six new bridge plans.

Error Analysis

One possible cause of error is the distinct patterns for representing certain SNBI items in bridge plans. For example, Fig. 17 shows the minimum span length displayed in three distinct bridge plans with their corresponding RegEx patterns, respectively. Each style of rectangular box highlights a unique pattern, with its associated RegEx rules indicated in the matching legend item on the right. For example, pattern 1 represents any number followed by a hyphen, and then another number. In Case 1, only pattern 3 is applicable. In Case 2, both pattern 1 and pattern 2 apply. In Case 3, both pattern 2 and pattern 3 are applicable. Therefore, a robust and comprehensive information extraction algorithm needs to cover all possible patterns. Another example is that keywords including "I-beam" and "bulb-tee beam" can both represent the "I-shaped" span type.

Another potential source of error stems from imperfections of the OCR model. For instance, as illustrated by pattern 2 in Fig. 17, the OCR sometimes misinterprets an apostrophe (') as a superscript character "1" (^{1}). This error can be handled by encoding specially tailored rules in the algorithm.

Time Savings

The time savings were evaluated using a bridge plan from the testing set. Table 5 provides a comparative analysis of the time consumption for the manual approach and the proposed automated approach for each phase, including 3D geometry reconstruction, semantic segmentation, information extraction, and semantic enrichment.

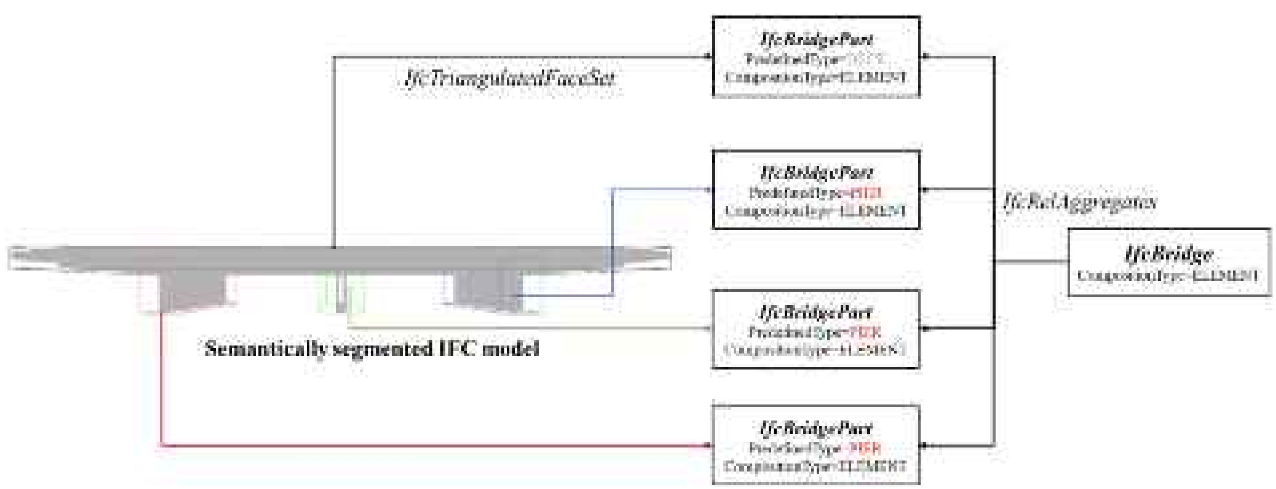

Fig. 15. Illustration of entity mapping for the IFC model construction.

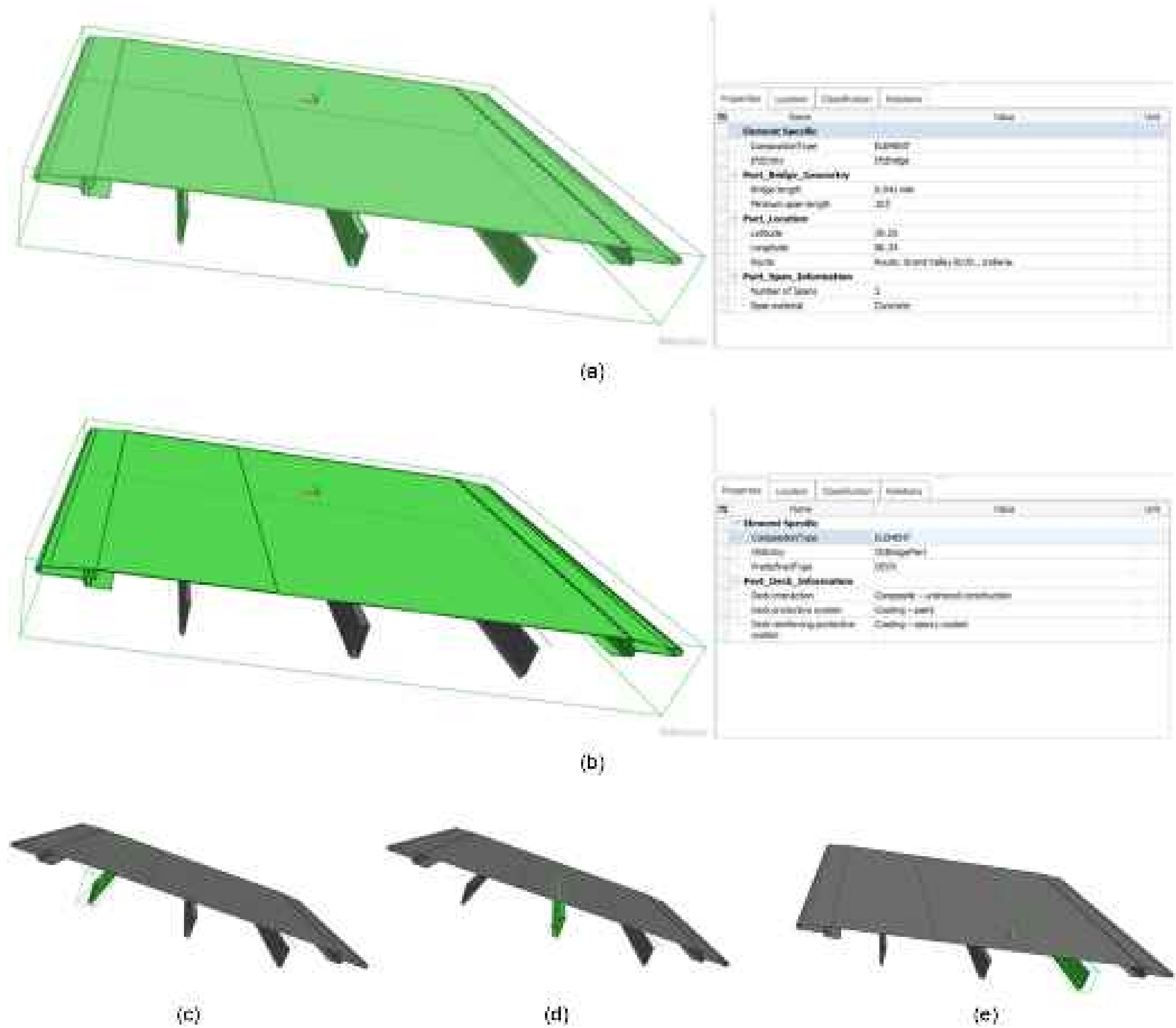

Fig. 16. An example of a semantically segmented and enriched IFC bridge model: (a) bridge; (b) deck; (c) Pier 1; (d) Pier 2; and (e) Pier 3.

Table 4. Testing results of the information extraction algorithms on the six new bridge plans

Category	Total number of information elements from the six bridge plans	Number of extracted elements using the proposed approach	Number of correctly extracted elements using the proposed approach	Precision (%)	Recall (%)
Bridge identification	12	12	12	100.0	100.0
Bridge material and type	42	40	39	97.5	92.9
Bridge geometry	12	12	11	91.7	91.7
Features	12	12	12	100.0	100.0
Loads, load rating, and posting	12	11	11	100.0	91.7
Total	90	87	85	97.7	94.4

3D geometry reconstruction: According to Akanbi and Zhang (2022a), constructing a 3D bridge model from 2D plans manually required 5 h, while the PDF2BIM algorithm significantly reduced this time to 0.179 h.

Semantic segmentation: Manual semantic segmentation involved observing the point cloud model and segmenting it by drawing bounding boxes, which took 0.128 h. In comparison, the proposed approach took 0.055 h leveraging the DBSCAN algorithm with parameter tuning and visual inspection to segment the point cloud model semiautomatically.

Information extraction: Regarding information extraction, manually extracting the 15 SNBI data items from a single bridge plan

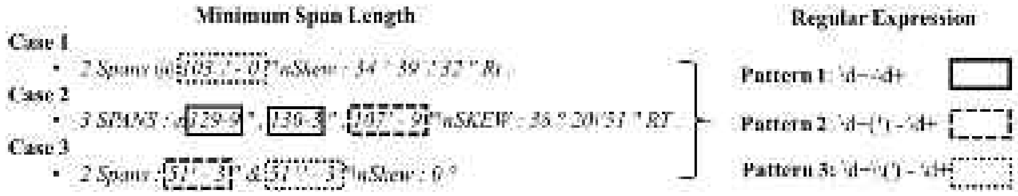

Fig. 17. Different RegEx patterns for extracting minimum span length in distinct cases.

Table 5. Time consumption of the manual approach and the proposed automated approach for processing a single bridge plan from the testing set

Phase	Manual approach (h)	Proposed approach (h)	Time savings (%)
3D geometry reconstruction	5.000	0.179	96.4
Semantic segmentation	0.128	0.055	57.0
Information extraction	0.195	0.037	81.0
Semantic enrichment	0.097	0.004	95.9
Total	5.420	0.275	94.9

took approximately 0.195 h, whereas the proposed automated approach took 0.037 h using the developed information extraction algorithms.

Semantic enrichment: The semantic enrichment process involved mapping the segmented mesh model into the IFC model and manually inputting the extracted data items into the IFC bridge model using an IFC authoring platform such as the BlenderBIM Add-on (Blender 2023). This manual process took approximately 0.097 h, compared to 0.004 h using the proposed approach, showing once again that substantial time savings were achieved.

Combining the four phases above, the proposed approach achieved an overall time reduction by 94.9% compared to the manual approach for processing a testing plan, as shown in Table 5. Consequently, the results allow for a more efficient reconstruction of a semantically segmented and enriched 3D bridge IFC model.

Contributions to the Body of Knowledge

This research contributes to the human body of knowledge in the following ways. First, this work extended the existing PDF2BIM (Akanbi and Zhang 2022a, b) technology. The authors leverage as-designed point cloud data and propose a novel framework to generate semantically segmented and enriched IFC-based bridge BIM instance models using 2D drawings. The proposed framework integrates semantic segmentation, information extraction, and semantic enrichment techniques on top of PDF2BIM, to address the gap of the output from the PDF2BIM technology in the lack of (1) segmented components; and (2) semantic information. The previous IFC model generated by PDF2BIM has only one main IFC entity instance (*IfcMember*) to represent the whole bridge, which lacks distinction among different parts of the bridge. In addition, the generated IFC model is only a geometric shell without semantic information (Li et al. 2024; Li and Zhang 2024). Moreover, existing research concentrated on geometric modeling and as-built point cloud data segmentation (Jiang et al. 2023; Yang et al. 2022; Wang et al. 2023). While the as-built point cloud data are effective in preserving geometric information, they do not address the need for semantic information in the models. For example, as-built point cloud data cannot capture semantic information including object categories (e.g., piers, bents, abutments, railings, and decks, etc.) and component properties (e.g., material type, foundation type). These issues limit the potential benefits of 3D bridge BIMs following IFC standards such as improved interoperability, data integrity, and collaborations among different stakeholders. Through semantic segmentation leveraging the as-designed point cloud, the IFC model is broken down into different pieces corresponding to the bridge's components (e.g., decks and piers). Furthermore, information extraction and semantic enrichment enrich the bridge IFC model with necessary semantic information extracted from the bridge plan. The semantically segmented and enriched bridge IFC model can better unlock the capabilities offered by IFC representations. For example, it can improve collaboration among different parties (Akanbi et al. 2020) and optimize construction planning and effective asset management. In addition, it will enable interoperability between various software applications and platforms used in the AEC industry by exchanging data seamlessly. This is because IFC schema is independent of the software vendor or the specific software version being used. Overall, the model will contribute to more sustainable, efficient, and cost-effective bridge projects.

Second, existing research primarily focuses on constructing as-built bridge BIMs by leveraging scan-to-BIM techniques using as-built point cloud data (Justo et al. 2023; Isailović et al. 2020; Lin and Habib 2022). However, our study shifts the focus toward employing as-designed bridge plans instead of as-built point cloud data due to the laborious nature of collecting as-built point cloud data and the inherent limitations of the data collection process, e.g., line-of-sight restrictions. Moreover, the existing bridges already have bridge drawings documented in the governmental database (e.g., Indiana Historic Bridge Inventory) (INDOT 2024). There is a gap in exploring reconstruction of bridge BIMs using as-designed drawings. In addition, certain fundamental semantic information, like material type and foundation type, is typically derived from as-designed drawings. These data form the core of the initial BIM instance model, which can be coded into the as-designed bridge BIM using the proposed approach. This is crucial for establishing a baseline understanding of the bridge's design and intended function. However, such information could be difficult to collect by only using the on-site data collection approaches. In the big picture, both scan-to-BIM and our 2D PDF-to-BIM are valuable approaches that supplement each other and provide support to managing bridge projects and assets. For example, the as-designed bridge BIM can serve as the foundation for the in-service bridge BIM. The as-designed BIM can be integrated into asset management systems to monitor the bridge's condition over time. By updating the BIM instance model with real-time data collected from sensors or periodic inspections, transportation agencies can have an up-to-date view of the bridge's health and anticipate maintenance needs before issues become critical. In addition, the as-designed BIM can be used to simulate and analyze the expected life cycle of the bridge (Meng et al. 2023). The model provides detailed semantic information about the bridge's components, materials, and design specifications. It can facilitate more targeted and efficient maintenance activities, as teams can plan and execute tasks with a clear understanding of the bridge's structure and components' locations. By incorporating factors like material degradation and environmental impacts, stakeholders can predict future maintenance needs and costs, helping with long-term budgeting and resource allocation. Furthermore, the as-designed BIM can help ensure that the bridge remains compliant with evolving regulations. The model can help track changes, manage documentation, and streamline the verification process to ensure that the bridge meets all regulatory requirements.

Third, this study focuses on bridges. Extensive research has been done on BIM reconstruction from PDFs for buildings (Lu et al. 2020; Zhao et al. 2021; Pan et al. 2023). However, bridges have drastically different components from buildings. Instead of doors, walls, windows, beams, and columns in the buildings, bridges comprise unique structural components including piers, bents, abutments, and decks, etc. Consequently, existing approaches or methodologies for semantic segmentation and information extraction developed for buildings cannot be directly adopted for bridge structures. The gap for constructing IFC-based BIM utilizing existing 2D PDF drawings for bridges is addressed in this study.

Fourth, the SNBI provides a comprehensive and systematic framework for representing bridge data. The federal government has mandated the adoption of SNBI for collecting and describing bridge data with a deadline of March 15, 2028 (US DOT 2023a). However, the research on applying SNBI to enhance semantic information in bridges is still limited. This study aims to address a real need of conforming to the federal SNBI, by developing information extraction algorithms to map the information from bridge plans to the SNBI code. In addition, leveraging the wealth of information available in bridge plans significantly accelerates the process of enriching IFC-based BIM instance models for bridges. The information extraction algorithms developed under the proposed framework in this study effectively exploit the rich data embedded in bridge plans. This results in an efficient integration of such extensive information into the IFC-based BIM instance models. Moreover, the time efficiency of the automated process scales well with the increased number of SNBI data items and the complexity of the bridge (e.g., more piers and decks and sophisticated geometries). Consequently, this renders the process of enriching the BIM instance models for bridges more time-efficient, thereby enhancing overall bridge project productivity.

Fifth, the proposed method for developing the information extraction algorithm is iterative and data-driven, designed to adapt and improve with the incorporation of additional training data. This approach allows the algorithm to accommodate various bridge plan formats, including those from different states. By integrating diverse plan formats into our training data set, the robustness and flexibility of the algorithm can be iteratively enhanced, ensuring its applicability across a wide range of standards and formats. This adaptability is a core strength of our proposed method, enabling it to generate the semantically segmented and enriched bridge BIMs effectively in other states by learning from the plan formats encountered.

Finally, the deployment of semantically segmented and enriched BIM instance models for bridges brings about considerable benefits in enhanced data integrity. By cross-validating the extracted semantic data with other bridge databases like the National Bridge Inventory (NBI) and Indiana Bridge Inspection Application System (BIAS), the accuracy and reliability of the bridge-related data can be checked and enhanced. Furthermore, the integration of semantic enrichment in IFC bridge models ensures an accurate representation of asset details, which are routinely updated to aid efficient life-cycle data management (US DOT 2024). This process fosters unimpeded collaboration among various stakeholders and simplifies QA and QC methodologies. In essence, the integration of semantically segmented and enriched bridge IFC models is likely to boost data trustworthiness, promote effective collaboration, and make QA/QC processes more efficient. Therefore, it has the potential to support improved decision-making and management of bridge projects and assets (US DOT 2024).

Conclusion

In this study, the authors proposed a new four-phase framework to generate semantically segmented and enriched bridge IFC models using 2D bridge plans. The framework consists of (1) 3D geometry reconstruction, (2) semantic segmentation, (3) information extraction, and (4) semantic enrichment. The integration of these steps helps address the gap of the output from the PDF2BIM technology in the lack of (1) segmented components; and (2) semantic information. Algorithms to streamline and automate this process were developed following the proposed framework leveraging open and transparent data structures, including the mesh model (.obj) and point cloud model (.pcd), to facilitate seamless transformation to bridge BIM (.ifc). One of the challenges in the semantic segmentation is that the mesh model has a complex structure of interconnected triangular surfaces. To facilitate smooth and accurate segmentation, the mesh model is initially converted into a point cloud model. Therefore, the DBSCAN clustering algorithm coupled with the boundary segmentation technique can be leveraged to semiautomatically (i.e., with parameter tuning and human visual inspection) decompose the bridge model into its constituent parts (e.g., piers and decks). Moreover, the challenges for information extraction are summarized as follows. First, some bridge plans are image-based. OCR technology is thus required to convert these plans to searchable data and extract tabular information. Several OCR tools were evaluated and compared for the development of information extraction algorithms. Nanonets was selected due to its highly functional and easily accessible API. Second, the information extraction algorithms are uniquely tailored to the SNBI standard, which requires extensive bridge engineering knowledge and RegEx technology to map the information found in bridge plans to the specific SNBI code. For example, in the bridge plan, designations such as "W12×35" or "C12×35" indicate the span material as *S01 Steel–rolled shapes* according to the SNBI code. The letter "W" refers to wide flange beams, and "C" refers to channel sections. The number followed by the letter "W" or "C" indicates the dimensions of the shapes. Because the dimensions can vary across different bridges, RegEx rules "W\d+ x \d+" and "C\d+ x \d+" were developed to identify these designations. However, one exception is that "W8×31" needs to be excluded, which refers to the railing structure. An iterative data-driven approach was therefore employed to develop the information extraction algorithms (using six development bridge plans). The developed algorithms were then evaluated with six testing bridge plans, which achieved 97.7% precision and 94.4% recall. The semantic enrichment phase mapped the segmented mesh model to the IFC bridge model, following IFC standard Version 4 × 2 (BuildingSMART 2023c). The extracted semantic information was then integrated with the IFC model automatically.

The end result is an IFC-based bridge model that goes beyond simple geometric representation to include additional semantic information and segmented parts. This study offers an effective approach to unlock the full potential of IFC standard in bridge engineering. Furthermore, the proposed automated approach reduces the overall bridge modeling time consumption by 94.9% compared to the manual approach for processing a testing plan.

By automating a traditionally labor-intensive, time-consuming, and error-prone process, the authors provide a promising new framework for constructing BIM instance models using 2D plans more efficiently. This work sets the stage for further innovations in BIM for bridges.

Limitations and Future Work

The authors acknowledge five limitations of this study as follows. First, from the perspective of semantic segmentation, only the DBSCAN algorithm was tested. Although the process of semantic segmentation is semiautomatic, it still requires parameter tuning and human visual inspection to obtain optimal segmentation results. In the future, other algorithms, such as the m-estimator sample consensus (MSAC), random sample consensus (RANSAC), and region growing segmentation (Khaloo and Lattanzi 2017), will be further tested. A more advanced and comprehensive algorithm needs to be developed toward full automation of semantic segmentation.

Second, in terms of semantic segmentation for modeling bridges in IFC, one of the limitations of this work is that only the pier and deck were identified as the bridge component types and mapped to *IfcBridgePart* instances. Other *IfcBridgePart* types, such as those in superstructure, substructure, and foundation, were not included. According to the IFC Bridge fast track project for conceptual bridge models proposed by BuildingSMART (2018), hierarchy levels among *IfcBridge* and *IfcBridgePart* are richer compared to the hierarchy presented in Fig. 15. Specifically, the "pier" instances in *IfcBridgePart* are part of the "substructure" instances, and these relationships are defined through the *IfcRelAggregates* entity. Furthermore, *IfcBridgePart* can be further decomposed into *IfcElementAssembly* instances including *IfcColumn* and *IfcFooting*, etc. (BuildingSMART 2018).

Third, from the perspective of information extraction, one of the limitations of this work is that the information extraction algorithm can only extract semantic information that is explicitly represented as textual descriptions in the plan. It does not currently handle more sophisticated extractions that require interpretation of the bridge's geometry. For example, span types with I-shaped beams can be further classified into either adjacent or spread beams according to SNBI guidelines (US DOT 2022). However, such information is not explicitly described in text or tables. Although PDF2BIM can already transform the geometry of the bridge from 2D to 3D, semantic interpretation ability, such as to interpret the distribution of the beams from framing plans, is lacking. This highlights future work directions in leveraging computer vision algorithms to understand the geometry of the plan to further help distinguish and identify SNBI items such as the span type. In addition, although the semantic information extraction and enrichment achieved a high precision of 97.7% in the context of meeting the SNBI, it is not error-free yet. Future research should strive to remove the remaining errors through adding more background knowledge and/or with the help of cutting-edge AI.

Fourth, the efficiency of the information extraction algorithm needs to be improved. As shown in Table 5, the information extraction algorithms took 0.037 h to process a single bridge plan on average. This inefficiency stems from the scattered nature of the target information within the PDF plan, necessitating an exhaustive search. It significantly increases the time needed for the information extraction algorithms to operate. For example, information about the span type is usually found in the "Type" column of the structure information table, typically located on the first page of the plan. However, when this information is absent, the algorithm must search the entire document using keywords like "bulb-tee," substantially extending the processing time. To enhance efficiency, more heuristic rules could be formulated to improve the logic of the algorithm. For example, utilizing the table of contents in the drawing to pinpoint the page numbers where target data items should be located could eliminate the need for an exhaustive search. Such refinement and optimization will be pursued in future work.

Another limitation is that the collected bridge plans are only from Indiana. Bridge plans from other US states have not been used for training the information extraction algorithm yet. In this study, the authors proposed an iterative data-driven approach to develop the information extraction algorithm. This approach allows the algorithm to accommodate various bridge plan formats, including those from different states. By integrating diverse plan formats from other states into our training data set, the robustness and flexibility of the algorithm can be iteratively improved. Eventually, the algorithm can become comprehensive enough to cover a wide range of standards and formats of bridge plans.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the award SPR 4622 from the Joint Transportation Research Program administered by the Indiana Department of Transportation and Purdue University. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.

References

Akanbi, T., and J. Zhang. 2022a. "Framework for developing IFC-based 3D documentation from 2D bridge drawings." *J. Comput. Civ. Eng.* 36 (1): 04021031. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000986.

Akanbi, T., and J. Zhang. 2022b. "Semi-automated generation of 3D Bridge models from 2D PDF bridge drawings." In *Proc., ASCE Construction Research Congress*, 1347–1354. Reston, VA: ASCE.

Akanbi, T., J. Zhang, and Y.-C. Lee. 2020. "Data-driven reverse engineering algorithm development method for developing interoperable quantity takeoff algorithms using IFC-based BIM." *J. Comput. Civ. Eng.* 34 (5): 04020036. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000909.

Barreiro, A. C., M. Trzeciakiewicz, A. Hilsmann, and P. Eisert. 2023. "Automatic reconstruction of semantic 3D models from 2D floor plans." Preprint, submitted June 2, 2023. https://arxiv.org/abs/2306.01642.

Belsky, M., R. Sacks, and I. Brilakis. 2013. "A framework for semantic enrichment of IFC building models." In *Proc., 30th Int. Council for Research and Innovation in Building and Construction (CIB) W78 Int. Conf., 2013*, 514–523. Kanata, ON, Canada: International Council for Research and Innovation in Building and Construction.

Belval, E. 2022. "pdf2image: A Python module that wraps the PDFTOPPM utility to convert PDF to PIL image object." *GitHub*. Accessed September 12, 2023. https://github.com/Belval/pdf2image.

Blender. 2023. "An beautiful, detailed, and data-rich OpenBIM." *BlenderBIM Add-on*. Accessed September 14, 2023. https://blenderbim.org/.

BuildingSMART. 2018. "IFC Bridge fast track project." Accessed August 7, 2023. https://app.box.com/s/w3r53huy4srhfg8t2vr0o12ot93hflsa.

BuildingSMART. 2023a. "IFC infrastructure deployments." Accessed August 15, 2023. https://www.buildingsmart.org/standards/domains/infrastructure/ifc-bridge/.

BuildingSMART. 2023b. "IfcTriangulatedFaceSet." Accessed September 14, 2023. https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD1/HTML/schema/ifcgeometricmodelresource/lexical/ifctriangulatedfaceset.htm.

BuildingSMART. 2023c. "Industry foundation classes version 4.2 BSI candidate standard." Accessed August 7, 2023. https://standards.buildingsmart.org/IFC/DEV/IFC4_2/FINAL/HTML/.

BuildingSMART. 2023d. "OpenBIM in the US." Accessed March 7, 2024. https://www.buildingsmartusa.org/about/openbim/openbim-in-the-us/.

ChatPDF. 2023. "Chat with any PDF." Accessed October 16, 2023. https://www.chatpdf.com/.

Donato, V., C. Biagini, G. Bertini, and F. Marsugli. 2017. "Challenges and opportunities for the implementation of H-BIM with regards to historical infrastructures: A case study of the Ponte Giorgini in Castiglione della Pescaia (Grosseto–Italy)." *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 42 (May): 253–260. https://doi.org/10.5194/isprs-archives-XLII-5-W1-253-2017.

Gimenez, L., S. Robert, F. Suard, and K. Zreik. 2016. "Automatic reconstruction of 3D building models from scanned 2D floor plans." *Autom. Constr.* 63 (Mar): 48–56. https://doi.org/10.1016/j.autcon.2015.12.008.

Song, H., G. Yang, H. Li, T. Zhang, and A. Jiang. 2023. "Digital twin enhanced BIM to shape full life cycle digital transformation for bridge engineering." *Autom. Constr.* 147 (Mar): 104736. https://doi.org/10.1016/j.autcon.2022.104736.

IfcOpenShell. 2023. "IfcOpenShell." *Ifcopenshell.org*. Accessed February 1, 2023. http://ifcopenshell.org/.

INDOT (Indiana DOT). 2024. "Bridge inspection." *Doing Business with INDOT*. Accessed April 4, 2024. https://www.in.gov/indot/doing-business-with-indot/consultants/bridges/bridge-inspection/.

Isailović, D., V. Stojanovic, M. Trapp, R. Richter, R. Hajdin, and J. Döllner. 2020. "Bridge damage: Detection, IFC-based semantic enrichment and visualization." *Autom. Constr.* 112 (Apr): 103088. https://doi.org/10.1016/j.autcon.2020.103088.

Ismail, A., Y. Srewil, R. Scherer, and T. Mansperger. 2016. "Semantic enrichment and multimodel data exchange approach for CFD analysis of bridges." In *Proc., 23rd Int. Workshop of the European Group for Intelligent Computing in Engineering*. Berlin: Springer.

Jiang, S., X. Feng, B. Zhang, and J. Shi. 2023. "Semantic enrichment for BIM: Enabling technologies and applications." *Adv. Eng. Inf.* 56 (Apr): 101961. https://doi.org/10.1016/j.aei.2023.101961.

Jin, C., T. Wu, and J. Zhou. 2023. "Multi-grid representation with field regularization for self-supervised surface reconstruction from point clouds." *Comput. Graphics* 114 (Aug): 379–386. https://doi.org/10.1016/j.cag.2023.06.016.

Justo, A., D. Lamas, A. Sánchez-Rodríguez, M. Soilán, and B. Riveiro. 2023. "Generating IFC-compliant models and structural graphs of truss bridges from dense point clouds." *Autom. Constr.* 149 (May): 104786. https://doi.org/10.1016/j.autcon.2023.104786.

Kaewunruen, S., J. Sresakoolchai, and Z. Zhou. 2020. "Sustainability-based lifecycle management for bridge infrastructure using 6D BIM." *Sustainability* 12 (6): 2436. https://doi.org/10.3390/su12062436.

Khaloo, A., and D. Lattanzi. 2017. "Robust normal estimation and region growing segmentation of infrastructure 3D point cloud models." *Adv. Eng. Inf.* 34 (Oct): 1–16. https://doi.org/10.1016/j.aei.2017.07.002.

Lee, J. H., J. J. Park, and H. Yoon. 2020. "Automatic bridge design parameter extraction for scan-to-BIM." *Appl. Sci.* 10 (20): 7346. https://doi.org/10.3390/app10207346.

Li, H., F. Yang, and J. Zhang. 2024. "IFC-based semantic segmentation and semantic enrichment of BIM for bridges." In *Proc., Construction Research Congress 2024: Advanced Technologies, Automation, and Computer Applications in Construction*, 597–606. Reston, VA: ASCE.

Li, H., and J. Zhang. 2022a. "IFC-based information extraction and analysis of HVAC objects to support building energy modeling." In *Proc., 39th Int. Symp. on Automation and Robotics in Construction (ISARC 2022)*, 159–166. Enschede, Netherlands: Univ. of Twente.

Li, H., and J. Zhang. 2022b. "Improving IFC-based interoperability between BIM and BEM using invariant signatures of HVAC objects." *J. Comput. Civ. Eng.* 37 (2): 04022059. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001063.

Li, H., and J. Zhang. 2024. "Information extraction for semantic enrichment of BIM for bridge." In *Proc., Construction Research Congress: Advanced Technologies, Automation, and Computer Applications in Construction*. Reston, VA: ASCE.

Li, H., J. Zhang, S. Chang, and A. Sparkling. 2022a. "BIM-based object mapping using invariant signatures of AEC objects." *Autom. Constr.* 145 (Jan): 104616. https://doi.org/10.1016/j.autcon.2022.104616.

Li, H., J. Zhang, X. Xue, L. Debs, S. Chang, M. Qu, A. Sparkling, and D. Goldwasser. 2022b. "Issues in bi-directional interoperability between BIM and BEM." In *Proc., Construction Research Congress 2022: Computer Applications, Automation, and Data Analytics*, 1355–1364. Reston, VA: ASCE.

Lin, Y.-C., and A. Habib. 2022. "Semantic segmentation of bridge components and road infrastructure from mobile LiDAR data." *ISPRS Open J. Photogramm. Remote Sens.* 6 (Dec): 100023. https://doi.org/10.1016/j.ophoto.2022.100023.

Lu, Q., L. Chen, S. Li, and M. Pitt. 2020. "Semi-automatic geometric digital twinning for existing buildings based on images and CAD drawings." *Autom. Constr.* 115 (Jul): 103183. https://doi.org/10.1016/j.autcon.2020.103183.

Macher, H., T. Landes, and P. Grussenmeyer. 2017. "From point clouds to building information models: 3D semi-automatic reconstruction of indoors of existing buildings." *Appl. Sci.* 7 (10): 1030. https://doi.org/10.3390/app7101030.

Mandava, B., and J. Zhang. 2016. "A new automated quantity takeoff method for BIM-based bridge designs." In *Proc., CIB W78*. Rotterdam, Netherlands: Conseil International du Bâtiment.

Meng, W., H. Zhang, Q. Ai, T. Bao, and J. Yan. 2023. "CBR-RBR fusion based parametric rapid construction method of bridge BIM model." *Adv. Eng. Inf.* 57 (Aug): 102086. https://doi.org/10.1016/j.aei.2023.102086.

Nanonets. 2023. "Intelligent document processing with AI." Accessed April 27, 2023. https://nanonets.com/.

Pan, Z., Y. Yu, F. Xiao, and J. Zhang. 2023. "Recovering building information model from 2D drawings for mechanical, electrical and plumbing systems of ageing buildings." *Autom. Constr.* 152 (Aug): 104914. https://doi.org/10.1016/j.autcon.2023.104914.

Poku-Agyemang, K. N., and A. Reiterer. 2023. "3D reconstruction from 2D plans exemplified by bridge structures." *Remote Sens.* 15 (3): 677. https://doi.org/10.3390/rs15030677.

Sacks, R., et al. 2018. "SeeBridge as next generation bridge inspection: Overview, information delivery manual and model view definition." *Autom. Constr.* 90 (Jun): 134–145. https://doi.org/10.1016/j.autcon.2018.02.033.

US DOT. 2022. "Specifications for the national bridge inventory." *Office of Bridges and Structures*. Accessed July 11, 2023. https://www.fhwa.dot.gov/bridge/snbi/snbi_march_2022_publication.pdf.

US DOT. 2023a. "Bridges & structures." *Office of Bridges and Structures*. Accessed April 3, 2024. https://www.fhwa.dot.gov/bridge/snbi/qanda/02.cfm.

US DOT. 2023b. "Bridges & structures—National bridge inventory." Accessed July 11, 2023. https://www.fhwa.dot.gov/bridge/snbi.cfm.

US DOT. 2024. "Building information modeling (BIM) for bridges and structures." Accessed March 7, 2024. https://www.fhwa.dot.gov/bridge/bim/.

Wang, W., T. Su, H. Liu, X. Li, Z. Jia, L. Zhou, Z. Song, and M. Ding. 2019. "Surface reconstruction from unoriented point clouds by a new triangle selection strategy." *Comput. Graphics* 84 (Nov): 144–159. https://doi.org/10.1016/j.cag.2019.08.002.

Wang, Z., H. Ying, R. Sacks, and A. Borrmann. 2023. "CBIM: A graph-based approach to enhance interoperability using semantic enrichment." Preprint, submitted April 23, 2023. https://arxiv.org/abs/2304.11672.

Wu, J., and J. Zhang. 2018. “Automated BIM object classification to support BIM interoperability.” *Proc., 2018 ASCE Int. Conf. on Computing in Civil Engineering*, 706–715. Reston, VA: ASCE.

Wu, J., and J. Zhang. 2019a. “Introducing geometric signatures of architecture, engineering, and construction objects and a new BIM dataset.” In *Proc., 2019 ASCE Int. Conf. on Computing in Civil Engineering*, 264–271. Reston, VA: ASCE.

Wu, J., and J. Zhang. 2019b. “New automated BIM object classification method to support BIM interoperability.” *J. Comput. Civ. Eng.* 33 (5): 04019033. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000858.

Yang, X., E. del Rey Castillo, Y. Zou, L. Wotherspoon, and Y. Tan. 2022. “Automated semantic segmentation of bridge components from large-scale point clouds using a weighted superpoint graph.” *Autom. Constr.* 142 (Oct): 104519. https://doi.org/10.1016/j.autcon.2022.104519.

Zhang, C., R. Pinquié, A. Polette, G. Carasi, H. De Charnace, and J.-P. Pernot. 2023. “Automatic 3D CAD models reconstruction from 2D orthographic drawings.” *Comput. Graphics* 114 (Aug): 179–189. https://doi.org/10.1016/j.cag.2023.05.021.

Zhao, Y., X. Deng, and H. Lai. 2021. “Reconstructing BIM from 2D structural drawings for existing buildings.” *Autom. Constr.* 128 (Aug): 103750. https://doi.org/10.1016/j.autcon.2021.103750.