CONSTRUCTION PROGRESS MEASUREMENT USING DEEP LEARNING

A Thesis

by

XUE MOU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Julian Kang
Committee Members,	Wei Yan
	Edelmiro Escamilla
Head of Department,	Patrick Suermann

December 2019

Major Subject: Construction Management

ABSTRACT

Failing to keep track of the construction progress in time and making current construction progress falling behind the schedule will cause a significant loss of time and money. Monitoring the progress of the constructing structure in construction sites is one of the significant challenges today. The work of comparing buildings in construction with drawings requires time-consuming and expensive manual inspections. Computer vision techniques for construction progress monitoring were developed to solve these problems. By using the photographic apparatus, on-site situations can be detected and monitored remotely. However, the complexity of the picture makes it difficult to quantify the progress of the project. Semantic segmentation, as a branch of deep learning, is a method to identify objects in images at the pixel level. With the semantic segmentation technology of computer vision, the scene in the picture can be simplified, and the objects in the picture can be finely segmented. In this paper, to quantify the amount of work in the picture taken in construction sites, a new test was performed. This test is to estimate the completion rate of specific construction structures by comparing the segmented construction structures in the pictures with the BIM model. The objective of the research is to evaluate the rationality of the test.

ACKNOWLEDGEMENTS

I want to thank my committee chair, Dr. Kang, and my committee members, Dr. Yan and Dr. Escamilla, for their guidance and support throughout this research. I would also like to thank Dr. Yabuki at Osaka University for inspiring my research and helping me determining my research direction. I want to thank Neeraj Yadav and Rui Cao for his technical assistance in my experiment.

Besides, I would like to thank Dr. Buck Anderson, engineering manager of Grace Bible Church, for allowing me to use the Creekside Grace Bible Church as the experimental site. He offered me the opportunity to go into the site to collect data and agreed to provide me with the drawings and models of the construction on the site. I want to thank him very much.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my parents for their encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Julian Kang and Professor Edelmiro Escamilla of the Department of Construction Science and Professor Wei Yan of Department of Architecture and Professor Nobuyoshi Yabuki at Osaka University.

All work for the thesis was completed independently by the student.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

NOMENCLATURE

AI	Artificial Intelligence
SVM	Support Vector Machine
FCN	Fully Convolutional Networks
SfM	Structure-from-Motion

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Background

Construction progress assessment is a means of measuring the completion rate of construction projects in development to reflect the relative progress between the project and the schedule. It is an approach that converts the intangible progress on-site to the informative documents that can be read and saved [1]. The problems such as the construction inconsistent with the drawings or construction progress behind schedule are caused by the adverse progress supervision [2]. Therefore, progress assessment is a necessary part of construction management [3].

By definition, "progress" means "advance toward a specific end" [4]. Several methods can measure the degree of progress of construction. Existing research on progress management, listed in Table 1 below, suggested various assessment methods and targets in progress management. In their research, assessment methods are categorized into measuring, estimating, and weighting, while targets are classified into work quantity, milestone, process, and time.

Table 1 Existing construction progress measurement method[5]

Research	Measurement Direction	Measurement Method (CII)	Measurement Target (CII)	Measurement Method (Fleming & Koppleman)	Measurement Target (Fleming & Koppleman)
Fleming & Koppleman & CII	Physical progress measurement	Unit completed	Installed quantity	Installed elements counting	Installed quantity
	Estimated percent complete	Incremental milestone	Milestone	Percent complete & milestones gates	Progress state based on milestone
		Start/finish, supervisor	Start /finish point of work		
		Opinion	Progress state		
	Earned value	Cost ratio	None	None	None
		Weighted or equivalent units	Finish point or progress state of work	Weighted milestones	Finish point of weighted milestone

Among the physical progress assessment, the most common method is comparing the situation at the construction site with the reference drawings to obtain the project progress [6]. The most traditional method for engineers is to visually survey the situation on-site and then contrast the collected information with 2D drawings. With the development of photogrammetry and laser scanning technology, photography modeling and scanning modeling also emerged. Engineers can create photo models and scanning models by taking photos of the site or scanning the site and make these models overlap with the reference 3D models or even BIM models to estimate the completion rate of the project [7].

However, the implementation of these methods remains challenging. For visual inspection, the defect lies in manual observation. First, manual observation is a time-consuming and laborious work. Furthermore, it is difficult to quantify the amount of work by human eyes. For photographic modeling and scanning modeling, high-performance computers, as well as expensive equipment and software, are required. Also, the modeling process is relatively time-consuming and tedious. The time of modeling depends on the volume, quality, and resolution of the project.

1.2. The idea of automatic progress detection

A new test is needed to be proposed to overcome the issues above. That is, without using photographic modeling and scanning modeling, the photos taken in the project can be are compared with the reference drawings or reference models. Different from manual observation, instead of using human eyes, this test relies on a machine to read and analyze the picture, and this test also uses a machine to compare the picture with the model to obtain a precisely calculated completion rate.

However, the new test still faces many problems. First, there is so much information in a photo that it is hard to analyze it without processing it. Moreover, the computer needs to know how to quantify the building elements in the picture. Besides, how to connect and compare pictures with models by machine is also a problem.

Given these problems, some measures are needed to solve these problems. For the first issue, regarding some information in the picture is useless, the computer needs to judge the information and simplify the picture by itself. The computer needs to discard the information that is not needed and retain the information that is needed. For the second issue, the computer needs to capture the building components that engineers need in the image and quantify them by using the area of the building components in the image. Third, the picture is two-dimensional, while the reference model is three-dimensional. If these two things need to be connected, they should establish a contact in the same dimension.

Since photographic modeling is not used in this study, the problem will be solved in two-dimensional scope. That is, the 3D reference BIM model will be disassembled into images and used for comparison.

In the past research of computer vision, many methods of analyzing images by computer have been put forward, and they all have their advantages. However, they also have their limitations. For example, some algorithms can only extract specific characteristics of the target object, such as shape, color, and texture. For some other algorithms, users need to preprocess the image so that the algorithm can extract the object in the image. For other algorithms, users need to adjust the intermediate parameters of the algorithms before using them, such as telling the algorithm what the color of the object is wanted. Therefore, a new algorithm is expected to be applied to extract the object in the picture, which can overcome the above problems. An algorithm that detects the objects with multiple properties in the image without image preprocessing and intermediate modulation is required.

1.3. Deep learning and computer vision

Deep learning is one branch of machine learning, and machine learning belongs to artificial intelligence (AI). AI, in contrast to natural intelligence, is the intelligence demonstrated by machines. AI mimics the "cognitive" functions of humans, such as "learning" and "problem-solving" [8]. AI is divided into three types of intelligence, which are analytical intelligence, human-inspired intelligence, and humanized artificial intelligence [9]. Analytical AI takes advantage of statistical knowledge to predict future decisions based on data provided; human-inspired AI tries to explore and understand human's emotions to guide the decisions making, while humanized AI that combines all types of competencies including cognitive, emotional, and social intelligence establishes a self-aware based on such interactions of intelligence. The technique of deep learning applied in this research is at the analytical intelligence level.

Machine learning is a subset of AI, and it is scientific algorithms based on statistical models. It is applied to solve specific tasks relying on the reasoning obtained from the training data without explicit instructions [10]. The algorithms of machine learning create a mathematical model based on training data, and the model will make decisions or predictions without explicit programming required. Machine learning is classified into four types of learning algorithms: Artificial neural networks, Support vector machines (SVM), Bayesian networks, and Genetic algorithms. Artificial neural networks are also known as deep learning [11].

Deep learning, also known as deep structured learning or hierarchical learning, is a branch of machine learning based on the artificial neural network [12]. Deep learning is divided into supervised learning, semi-supervised learning, and unsupervised learning [13]. It can be used to extract high-level features from raw data by using multiple layers of machine learning networks.

The technique of deep learning is widely used in computer vision. Computer vision is an inter-discipline that focus on how the computer can be used to understand and extract the information from images and videos automatically like a human or better than human do [14]. As a kind of technology applied in the engineering industry, the algorithms were used to build the automated recognition systems and realize batch image classification, object detection, and semantic segmentation.

1.4. Semantic Segmentation and Fully Convolutional Network

In computer vision, deep learning is mainly utilized in image classification, image detection, and semantic segmentation. Image classification is a coarse and relevantly simple image filing system [15]. It can demonstrate what a single object is present in the image. Object detection is a mid-level recognition system higher than image classification, which can locate and classify multiple objects by showing bounding boxes or circles. However, when considering extracting the information of the construction progress in the images, the function of image classification and object detection cannot simplify and quantify the architecture in the pictures. Thus, the techniques above cannot provide a full sense of comprehension of the construction sites.

Semantic segmentation is a high-level image analysis function, which classifies each pixel in the image. It can achieve fine segmentation by inferring labels for every pixel, and each pixel with its labels showing its class composes the objects. Based on the purpose of this study, semantic segmentation technology will be utilized in this research.

1.5. Current Situations of Semantic Segmentation in the Construction Field

Semantic segmentation describes the process of classifying each pixel of an image. In recent years, significant progress has been made to improve the algorithms to enhance the accuracy of image semantic segmentation.

In the last five years, semantic segmentation is often used in automated damages detection, self-driving, and medical microscopic photos analysis [16]. In recent years, it has been put in use in the area of environmental management and urban planning. In the field of construction, the technique of the image-semantic segmentation also attracted attention for more efficient construction management methods.

However, in previous studies, there is no study on quantifying the amount of construction in pictures by using semantic segmentation. Furthermore, most research that uses computer vision to measure construction progress focuses on point cloud or SfM model, that is, to solve problems at the 3D level. Though 3d models are more visible, using 2D images to check the construction process is more available in most situations when taking cost and portability into consideration.

In summary, an efficient, low cost, easy to operate construction progress monitoring method with high precision by using semantic segmentation is in need.

1.6. Research Objective

This research focuses on performing a test that estimates the construction completion rate of some specific construction structures in order to monitor the construction progress by comparing site pictures and reference models on a two-dimensional level.

Our research test has mainly two parts, as follows:

(1) Proposing an image semantic segmentation model based on a deep learning algorithm to detect and segment architectural structure components in images.

(2) Calculating the relevant construction progress by measuring the area difference between the object in the processed image and the object in the planning model.

The research objective of this study is to evaluate the experiment and find the feasibility and limitations of the test.

2. LITERATURE REVIEW

2.1. BIM&photography used in construction progress monitoring

The work of construction progress monitoring is to convert on-site situations to information. Most of the progress tracking technologies currently in use still rely on the intervention of experts and require a considerable workforce through their process of methods [17]. For example, enhanced IT tools do not need much training, and they are less costly compared to other technologies, but they are applicable mainly in small projects [18]. 3D sensing technologies are accepted as the most accurate and efficient tools to acquire data, which can be used for high precision purposed projects and large projects such as landmark development projects. Nonetheless, their high prices made 3D sensing technologies not affordable for a majority of projects [19]. Geospatial Technologies like GPS is often used to track the location of the materials from manufacturing to the construction site [20]. The main limitation of using geospatial technologies is that in crowded environments, positioning accuracy can be severely reduced due to congested and distorted satellite signals. Also, the tracking system becomes uneconomical if the receivers of geospatial technologies need to be attached to each construction object [21].

Nowadays, more and more new shooting tools like point-and-shoot cameras, time-lapse cameras, and smartphones are in use in construction sites [22]. With a number of functions embedded into the multi-functional cameras, information within the pictures taken by the cameras can be easily acquired without any unnecessary operations [23]. As a result, many photographic recording services have emerged in recent years to provide project participants with visual records for the construction phase, such as EarthCam, MulitVista, and JobSiteVisitor [24]. It is easier and faster to use these existing software packages for data processing. However, the workload of data processing is still limited by human intervention, which makes such applications both time-consuming and unsuitable for repeated progress monitoring tasks [25]. In construction sites, engineers take a lot of digital photos and record videos to collect information on the current construction situation. After collecting the data, they archive the photos and videos and evaluate the progress by comparing the current photos to the planned schedule to monitor the current rate of progress. The tasks of artificial sorting, annotating, storing, deleting, and distributing these digital images are cumbersome. Also, evaluating the construction progress by navigating the pictures and comparing them with 2D drawings or text version schedules is time-consuming and error-prone work.

Photogrammetry seems like a robust method, but this application in the industry has limitations due to the time-consuming calculation process, and the sensitivity of regions of interest and detectors to different lighting conditions [26]. Moreover, making a model by photogrammetry requires the intervention of humans to adjust the model based on the known dimensions [27]. Also, the quality of images, such as occlusion, has a substantial effect on the quality of the photogrammetry model. For instance, the formworks, scaffolds, and containers that obstruct the object of interest will decrease the quality of the model [28]. For now, many researchers have applied photogrammetry to automatic data collection.

Nowadays, digital pictures are commonly used by construction companies to track the construction process due to its higher efficiency, accuracy, and more natural utilization [29]. Photogrammetry is the technique that can generate three-dimensional or point cloud models of the construction site from digital photographs accurately. By comparing the 3D cloud model made by digital pictures with the 3D CAD model or BIM model, the completion rate of each component can be automatically calculated, and the progress of the construction project can be measured [30].

Structure-from-Motion (SfM) is a specific method to create a model by processing through the digital pictures. SfM is a simple and inexpensive technique based on photogrammetry, which can automatically match the properties in the intersection between images or frames to build a model in detail [31]. This technique is proved to be able to support the reconstruction of the construction site with 3D planning [32] and help to monitor the construction process of the building project [33]. Kevin K developed a method, which assigns correspondences between the 3D point cloud made by SFM and BIM model and classifies materials on extracted image patches of each element for inferring progress [34].

Since analyzing images taken by cameras can be a potential method to track the construction process due to its greater flexibility and availability, it is desirable to automate the process of object detection of images as well as the process of evaluating the progress measurement of construction projects. Scanning -BIM target recognition system combines laser scanning or digital photogrammetric point cloud with the BIM model, providing valuable information for tracking construction projects by comparing the three-dimensional point cloud model with the BIM model. If the system can be simplified to the Image-BIM target recognition system, which can compare the photos with the established BIM model automatically, and store the information of project process into the BIM model for real-time viewing, the system can improve the efficiency of process tracking and reduce errors of tracking made by the human.

2.2. Computer vision technique in construction progress monitoring

In previous studies, many algorithms were invented and used to extract building components in photos. For example, Son, Kim, and Neto use the color as a property to identify the structure [35]. However, this method is only limited to extracting the color of the object. If the object does not have a unique color, the color of the interested object needs to be modified to an obvious one. Moreover, this method can only rely on one specific color recognition, if another color is needed for recognition, then the algorithm is needed to be modified before running the algorithm. Another algorithm uses the shape to identify the column in photos and videos, but it cannot detect objects by other properties like color. Wu used 3D cad-based filtering to complete object recognition, which could not complete just by the image itself [36].

Though these studies have high accuracy in object detection, they lack universality as image object recognition algorithms. With the rise of machine learning and deep learning, the use of image analysis by this technology caught people's attention. Due to the "black box" function of deep learning, that is, in the convolutional layer of the depth model, various properties are learned and remembered by different convolutional layers, the deep learning algorithm enables their trained model to have the ability to identify objects by different kinds of properties.

Image recognition techniques in the area of artificial intelligence can support site practitioners to justify the work done on-site as experts do [37]. In the field of image processing, object detection focuses on judging the types and locations of construction objects in the image [38], while semantic segmentation describes the process of associating each pixel of an image with a class label. These technologies have been applied to civil engineering, such as road crack detection [39], worker detection, heavy equipment detection. AI is an excellent method to automate recognition, and human's eyes can be liberated from labor by using deep learning.

2.3. Semantic segmentation

The methods of Random Forest and Texton Forest for image segmentation were used before the technique of deep learning was developed in the field of computer vision [40]. Then the method of semantic segmentation of deep learning was proposed with the invention and development of convolutional neural networks [41].

The invention of Fully Convolutional Networks (FCN) abandoned the fully connected layer in the last layer of the network to make pixel detection possible [42], which means compared to other convolutional networks, this structure enables to grasp the information of the contours of the objects detected. Besides, the processing speed of the model was also improved. Since then, almost all recent semantic segmentation studies have adopted this structure.

The development of the FCN is a significant feat in the field of deep learning. Many related deep learning networks for semantic segmentation were improved for better performance based on FCN. For example, Badrinarayanan transferred the maximum pooling layer to the decoder to improve the segmentation resolution, which makes it possible to segment the video from a web camera in real-time [43].

Lin proposed an encoder-decoder structure called RefineNet with the encoder resnet-101 module and the decoder RefineNet module [44]. This structure combines the low-resolution features of the previous RefineNet module with the high-resolution features of the encoder. Though the network of RefineNet has an unparalleled precision, the workload of the training process of RefineNet is enormous, which requires a larger dataset. Also, the structure of RefineNet is more complicated than other network structures, which makes its practical application difficult.

In 2015, Olaf designed the deep learning network U-net [45]. The advantage U-net network is that it has a quick connection between the encoder and the decoder, which helps the decoder better determine the details of the target. The high accuracy up to the medical cell segmentation level of the U-net network makes it the most commonly used network since now. It has excellent performance even on small data sets, so it was developed for biomedical image segmentation.

2.4. Related deep learning applications in the field of construction

In the construction industry, several problem resolutions based on deep learning have been proposed. For example, Convolutional Neural Network (Faster R-CNN) based structural visual inspection method is proposed to provide quasi-real-time simultaneous detection of multiple types of concrete spalling damages [46]. A method was proposed to build a new framework to check if a site worker is qualified to work within the constraints of their certification by analyzing the dynamic spatiotemporal relevance between workers and non-worker objects [47].

In this research, the test that extracting the construction structure automatically and compare the extraction with the BIM model to follow up the construction will be proposed, which will considerably improve the efficiency of project process tracking as well as reduce the risks of errors by a human.

3. PROPOSED METHOD

3.1. Overview

As described in Chapter 1, our purpose is to prove if it is possible to extract and quantify the building components in the photos of the construction site by using semantic segmentation technology to measure the project progress. In this part, we will briefly state the proposed test.

The core of the system is to train a deep learning model and complete semantic segmentation tasks with high processing speed and precision. After obtaining the extracted area in the image, we overlapped the extracted structure with the reference model at the same scale and direction to calculate its ratio of the extracted area in the image and the structure area in the reference model, and finally obtain the completion rate.

3.2. Research Originality

For the research to monitor the construction progress on a whole scale, generating 3D point cloud by using photography is more practical since a 3D point cloud model has more advantages when comparing it to the BIM model. For the research to check the progress of a certain kind of construction component, 2-dimensional images are more practical and accurate because we do not need to care about the distortion of an image when it is transformed from two to three dimensions. Therefore, instead of comparing the volume of the construction elements in the 3D point cloud to the same one in the BIM model, in this research the difference of the area of the construction elements between on-site images and virtual images rendered by the as-planned model will be used to check the completion rate.

Besides, instead of using a periodic progress model as the reference, this research utilizes the entire BIM model as a compared baseline to monitor the construction progress so that a number can present the degree of completion as comparing the present situation to the finished product.

3.3. Research Assumptions

In this study, an assumption is needed. That is, the area ratio of the building component in the picture and the BIM model can reflect the completion degree of the building component at present, and the area ratio of the building component in the picture and the BIM model can reflect the completion degree of the building component at present.

3.4. Research Limitations

In this study, due to time constraints, the whole engineering process cannot be tracked. In the experiment, the inspected construction components are sheathing plywood, cement brick wall, and the external wall. Though the whole project progress cannot be reflected by testing these building components, this study can have a glimpse of the whole project progress inspection process through the experimental process by testing the construction progress of the building components above.

3.5. Research Methodology

Since the experiment procedures are generally divided into two steps, the methodology of the study is to break down these two steps. For the deep model training and verification, the steps are divided into three steps, namely, algorithm choosing and the experiment object choosing, model training and photo segmentation, and the usability of models checking. For the step of comparison between the photo and the model, the steps are also divided into three steps, namely reference drawing capturing from the model, image overlapping and correction, and completion rate calculation. The following is the flowchart:

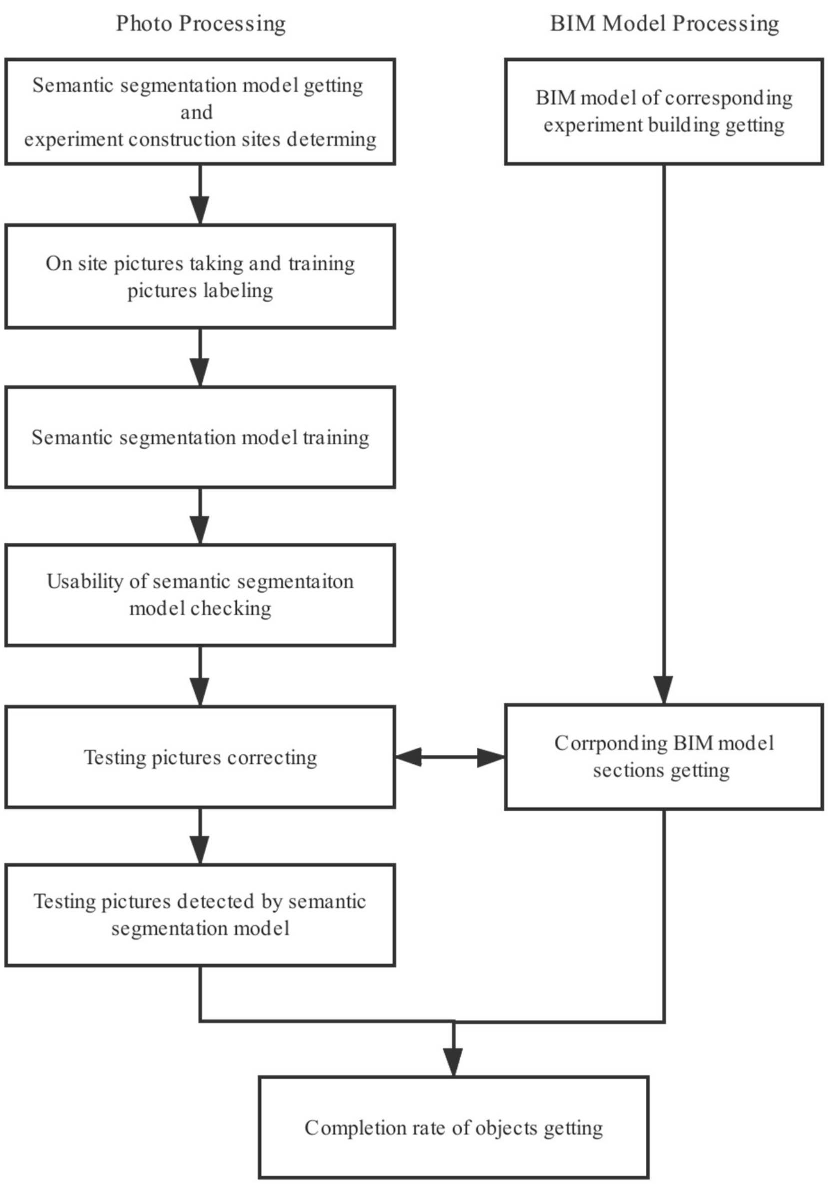

Figure 1 Flow chart of the research methodology

3.5.1. Algorithm choosing and the experiment object choosing

In the above literature review, the invented semantic segmentation algorithms have been described briefly. In this study, U-Net was selected for the research. The reason for choosing U-Net is that it can achieve medical cell level segmentation with a small data set.

For site selection, Grace Bible Church Creekside was selected for this study. This project is located in College Station and is currently in the stage of installing the facade. A simple BIM model was built using drawings provided by the contractor.

Figure 2 A rendering of the church and corresponding BIM model

3.5.2. Model training and photo segmentation

In this research, two models have been trained. One of the models was used to detect sheathing plywood, and another was used to detect the cement brick wall and external wall. For the image acquisition step of the model detecting sheathing plywood, due to these walls have been built, as shown in figure 3, we chose these walls as the training image collection area.

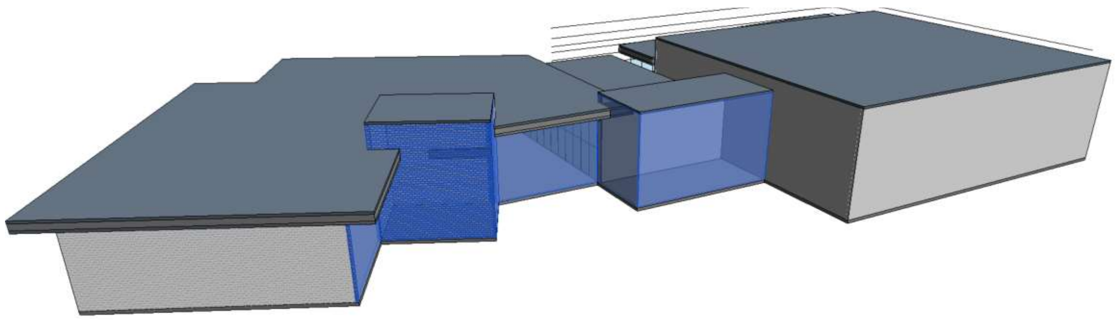

Figure 3 Training image collection area for sheathing plywood

Forty-five original images were collected by taking pictures in these areas above. After the original training pictures were flipped and rotated to a certain degree horizontally, the training data set was expanded to 180 photos. In these photos, the sheathing plywood was labeled. Then the 180 images with their labeled images composed the training group shown in figure 4.

Figure 4 Training images and corresponding labeled images for detecting sheathing plywood

The 180 pairs of photos in the training group were divided into two groups. 70% of pairs of photos are for training the semantic segmentation model, and 30% of them are for testing and supervising the model training. The training epochs is 300, and the size of pictures and labeled pictures is 512*512. The training time is about 8 hours.

After training, the model was used to examine photos taken at the wall shown in Figure 8. This wall was chosen as a test image collection area to monitor the progress of the project because it was under construction.

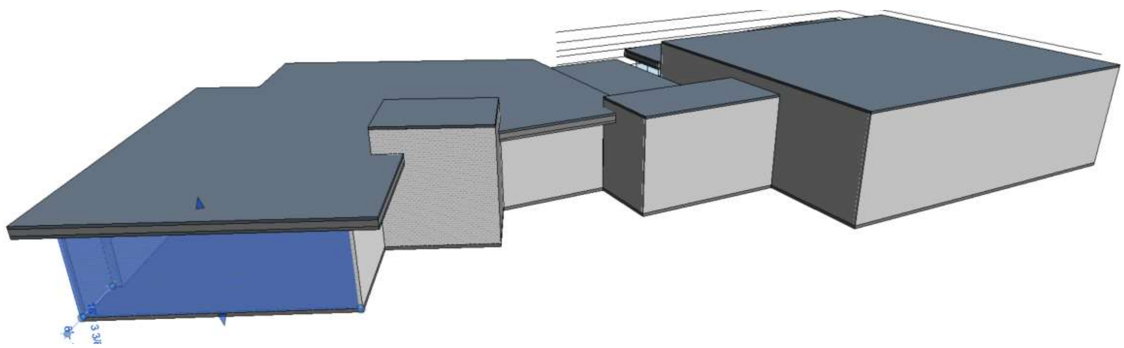

Figure 5 Testing image collection area for sheathing plywood

After training, the model automatically generates charts, which are respectively the average validation accuracy vs. epochs shown in Figure 6, the average IoU vs. epochs shown in Figure 7, and the average loss vs. epochs shown in Figure 8. These charts show the accuracy changes in the model with the increase of epochs. The chart proves that after 250 epochs, the average accuracy, average loss, and average IoU tends to be stable, which proves that 300 epochs can make full use of the data set and train the model to a good state.

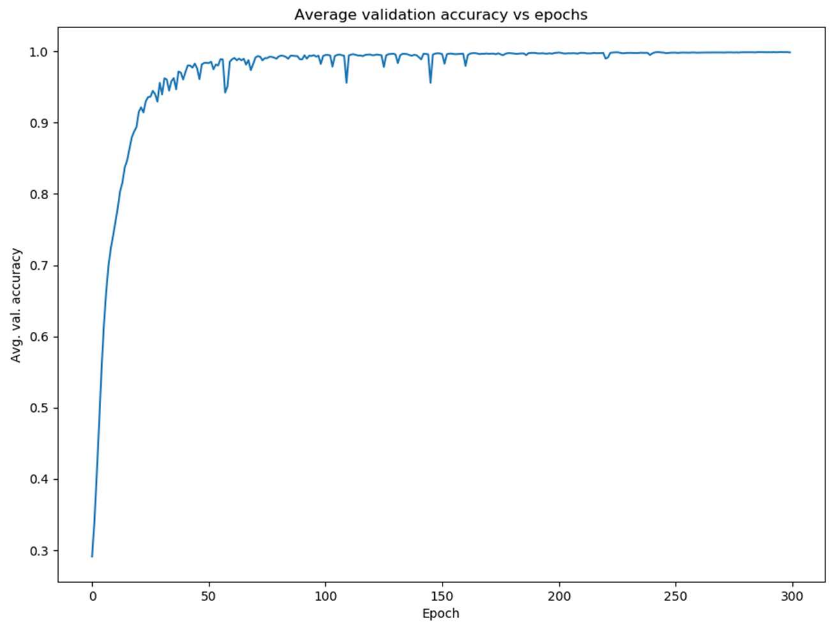

Figure 6 Average validation accuracy vs. epochs of the first model

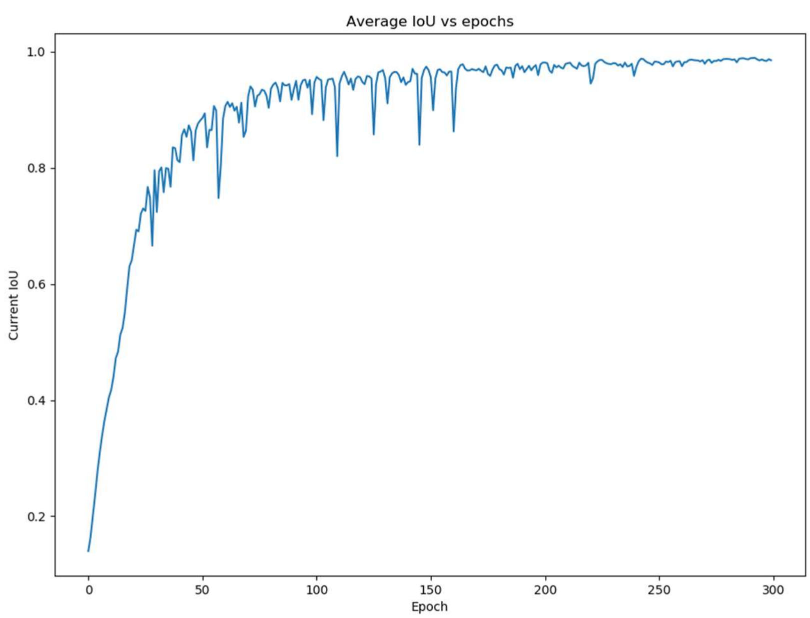

Figure 7 Average IoU vs. epochs of the first model

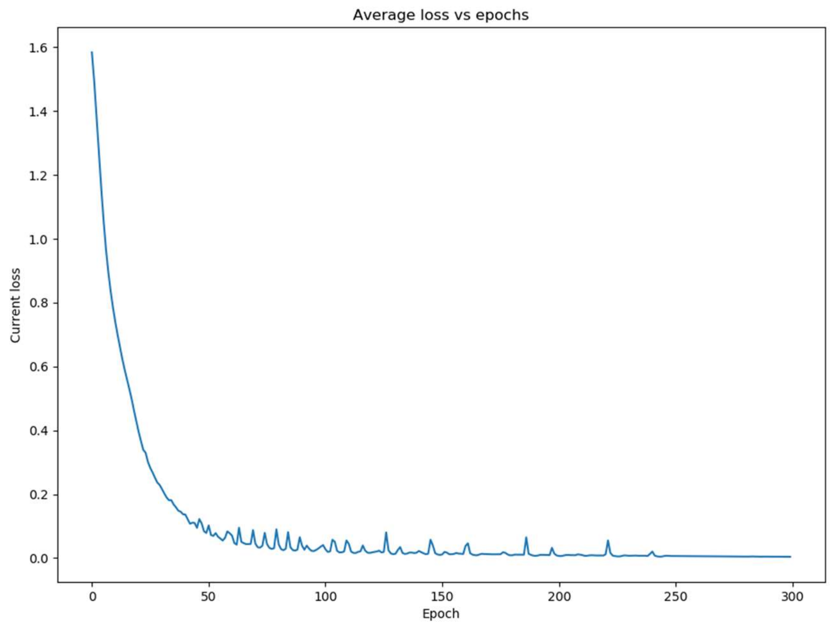

Figure 8 Average loss vs. epochs of the first model

For training the model that can detect cement brick wall and external wall, the method is nearly the same as above. The external wall construction process is shown in Figure 9.

Figure 9 The construction process of the external wall

For the model training picture collection, the wall on the backside was used to collect data. The wall is shown in Figure 9.

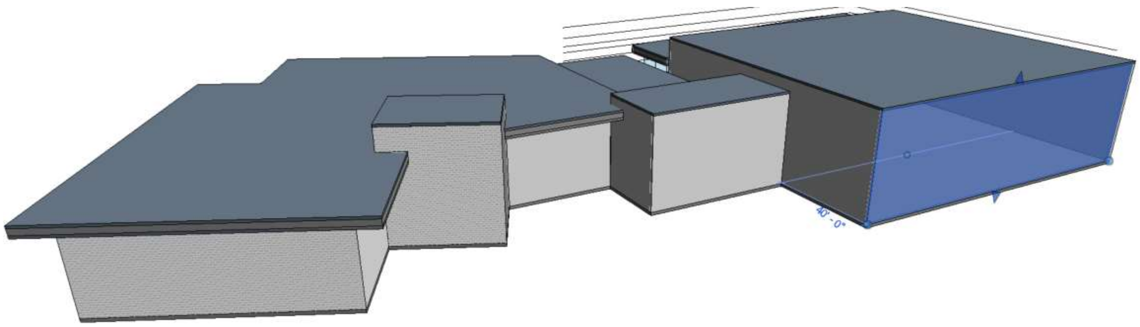

Figure 10 Training image collection area for detecting cement brick wall and external wall

As shown in Figure 11, the red area is the marking of the external wall, and the green area is the marking of the cement brick wall. One hundred thirty-seven images, including original images and annotated images, were used for model training. 70% of them were used for training the running model, and 30% of them are for testing and supervising the model training. The training epochs were still 300, and the size of the images and labeled images was 512*512. The training time is about 6 hours.

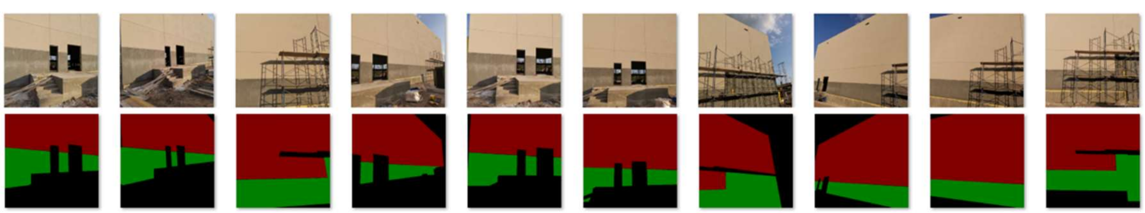

Figure 11 Training images and corresponding labeled images for cement brick wall and external wall

Because the left side of the wall in Figure 12 is in construction from sheathing plywood to external wall, in the experiments, this wall is used for detecting the completion rate of installing sheathing plywood and external wall.

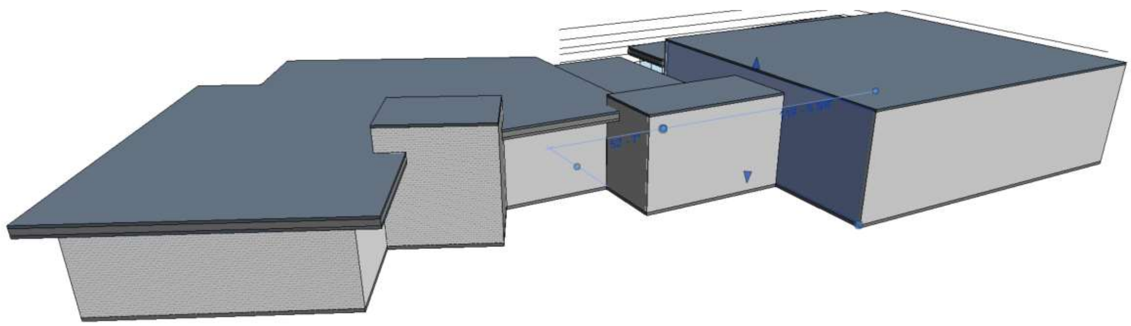

Figure 12 Testing image collection area for cement brick wall and external wall

Again, three charts were generated after training. The charts showed that after 200 epochs, the accuracy tends to be stable. The results showed that 300 epochs have no negative impact on the accuracy of the model.

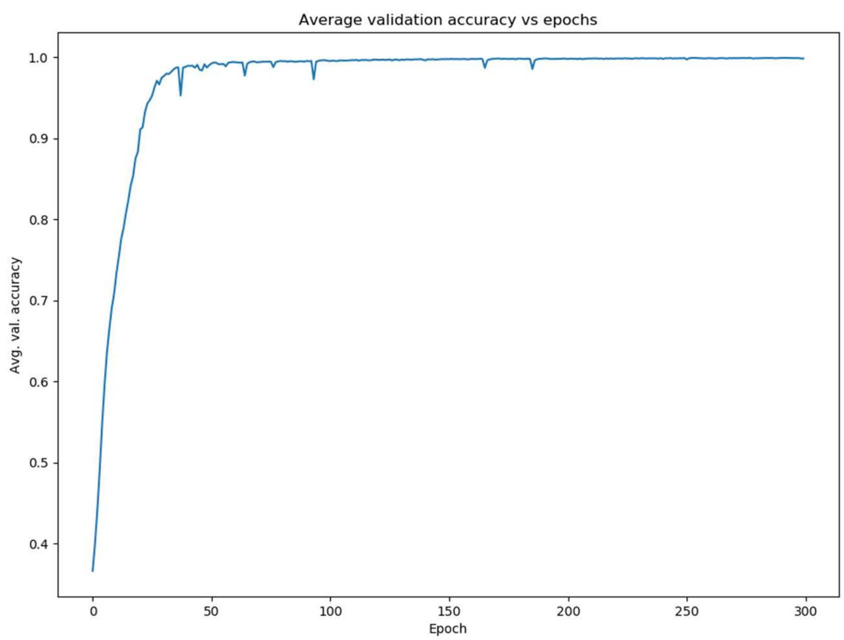

Figure 13 Average validation accuracy vs. epochs of the second model

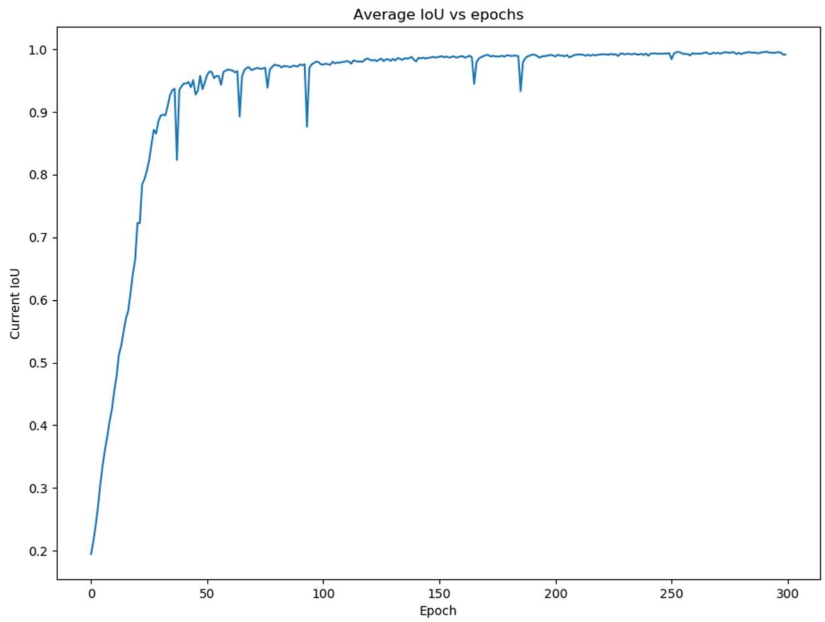

Figure 14 Average IoU vs. epochs of the second model

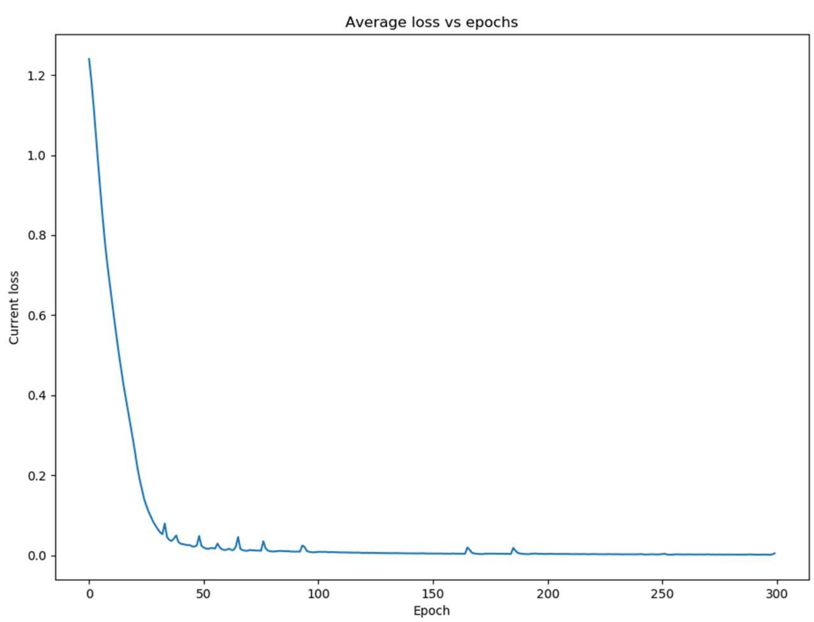

Figure 15 Average loss vs. epochs of the second model

3.5.3. The usability of models checking

After model training, the step is checking the usability of the model. We selected some test images and made their correct segmentation. To check the accuracy of these two models, we used the correct segmentation as the reference, and compare them with the corresponding images segmented by these two models. By calculating the proportion of the marked area segmented by the model, and the real segmentation area in the image, the availability of the model can be assessed before putting these models into use. The result was shown in table 2.

Table 2 The result of the accuracy in these two models

Test_name	Test_accuracy	Precision	Recall	Mean IoU
1	0.975208	0.974974	0.975208	0.901236
2	0.959469	0.959443	0.959469	0.85086
3	0.966537	0.965925	0.966537	0.867597
4	0.964077	0.964132	0.964077	0.861339
5	0.982803	0.982665	0.982803	0.935806
6	0.979565	0.979426	0.979565	0.926168
7	0.981071	0.981036	0.981071	0.932666
8	0.984474	0.984378	0.984474	0.942619
9	0.98801	0.987955	0.98801	0.959254
10	0.980515	0.980552	0.980515	0.9556
11	0.983551	0.983591	0.983551	0.958964
Explanation	Yellow/Green	Yellow/Red	Yellow/Green	Yellow/(Green+Red)

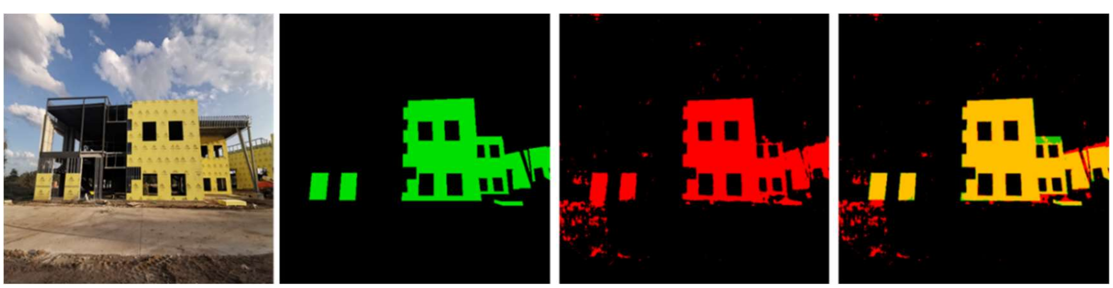

Figure 16 The example of pictures testing the accuracy of these models

For example, in Figure 16, the green area is the correct segmentation of the building component. The red area is the segmented area predicted by the model. The yellow area is the area of the overlap of the two segmented areas, namely the correctly segmented area. Test-accuracy in the table is equal to recall, which means how many percents of the correct pixels in the segmented image are recalled. The IoU is the intersection over Union, which means the ratio of overlapped area and the sum of red and green areas. In this table, test-accuracy is between 95% and 99%, and its IoU is between 85% and 95%. This table indicates that the accuracy of this model is at a high level, and it showed that these two models could be used in the next experiment.

3.5.4. Reference drawing capturing from the BIM model

As mentioned above, the overlapping issue should be solved in 2 dimensions. In this step, the corresponding section in the BIM model is captured into a picture. For the next step of image overlapping and correction, the image should be corrected to 512*512. The color of the image captured from the BIM model was changed to striking green, as shown in Figure 17.

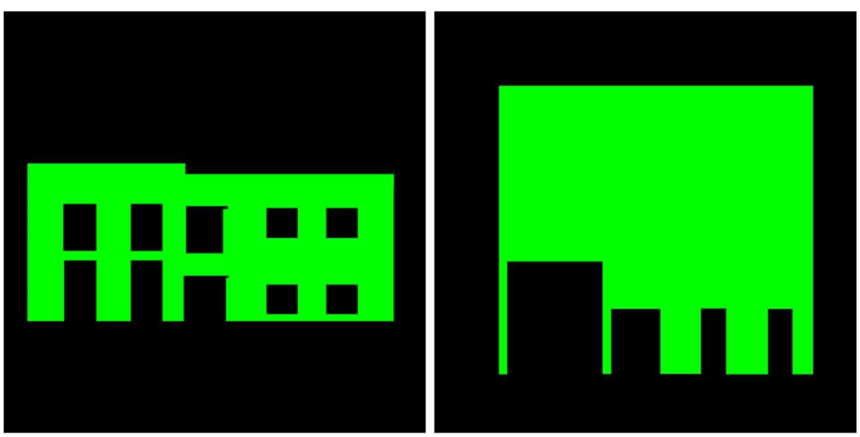

Figure 17 The pictures captured from the BIM model

3.5.5. Image overlapping and correction

Before the photos are superimposed with BIM model section pictures, we need to deal with the perspective of the image. There is a perspective in photos taken with a camera, but it is hard to express the same perspective in the BIM model. Moreover, even if the perspective can be reflected in the BIM model, the overlapped area after calculating cannot represent the progress. In this case, the perspective should be eliminated in this step, at the same time we need to match the size of the building in the images with the building in the BIM model, so that the photos taken on-site can be overlapped entirely with the section pictures in BIM model.

In this experiment, an algorithm of C++ can help to do the perspective correction. The section picture is taken as the reference picture, and the test picture taken on-site is regarded as the pending corrected picture. When the code runs, the code asks the user to select a point in the reference picture, and then it asks the user to select a corresponding point in the pending corrected picture. After selecting four or more points, Images with perspective can be converted to non-perspective images of the same size as the BIM model section. The images before and after adjustment are shown in Figure 18.

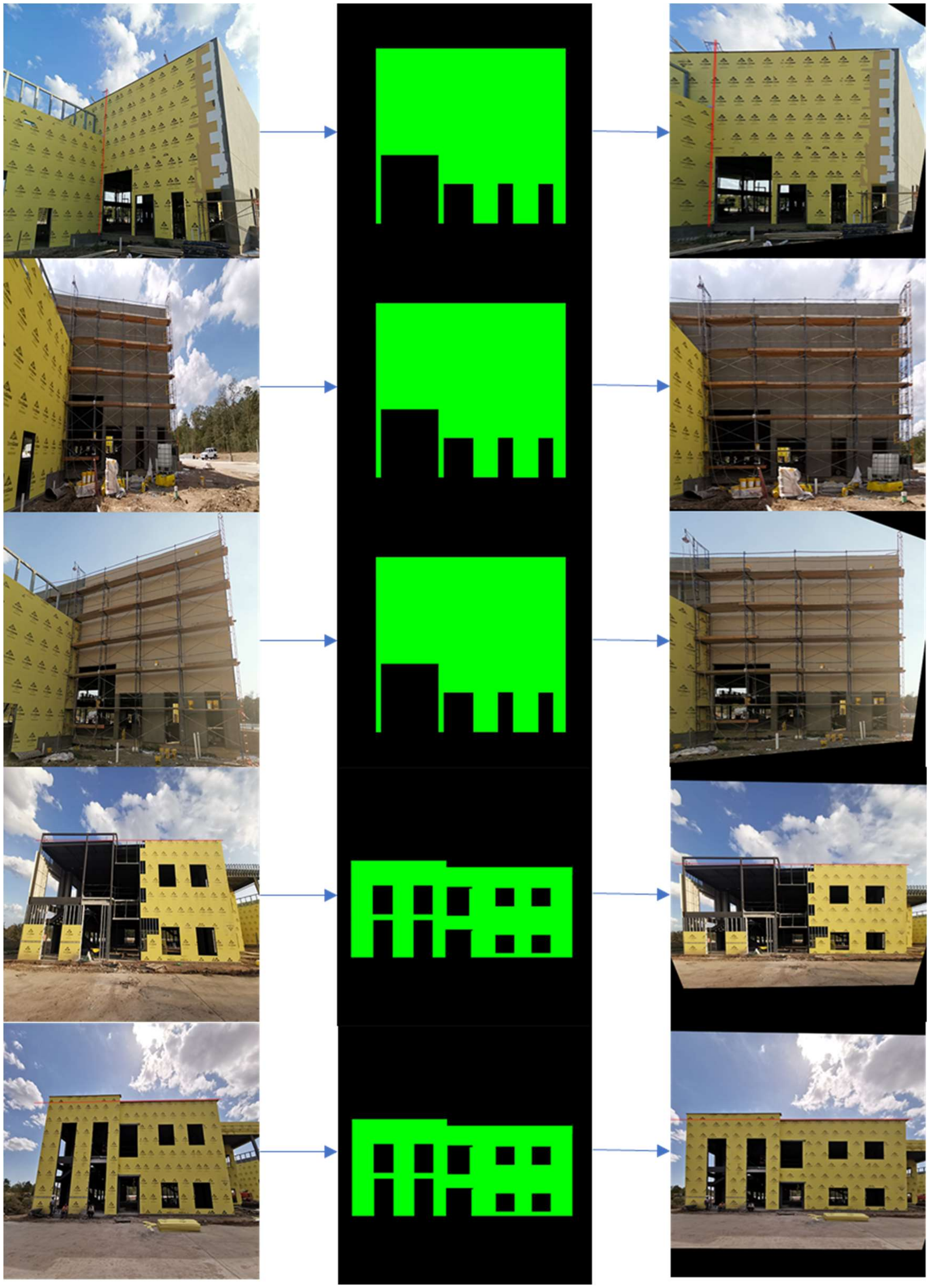

Figure 18 The images before and after adjustment

3.5.6. Completion rate calculation

The final step of the experiment is to calculate the completion rate by calculating the ratio of the predicted pixel value of building components in the tested image to the pixel value of building components in the section image, to obtain the project progress.

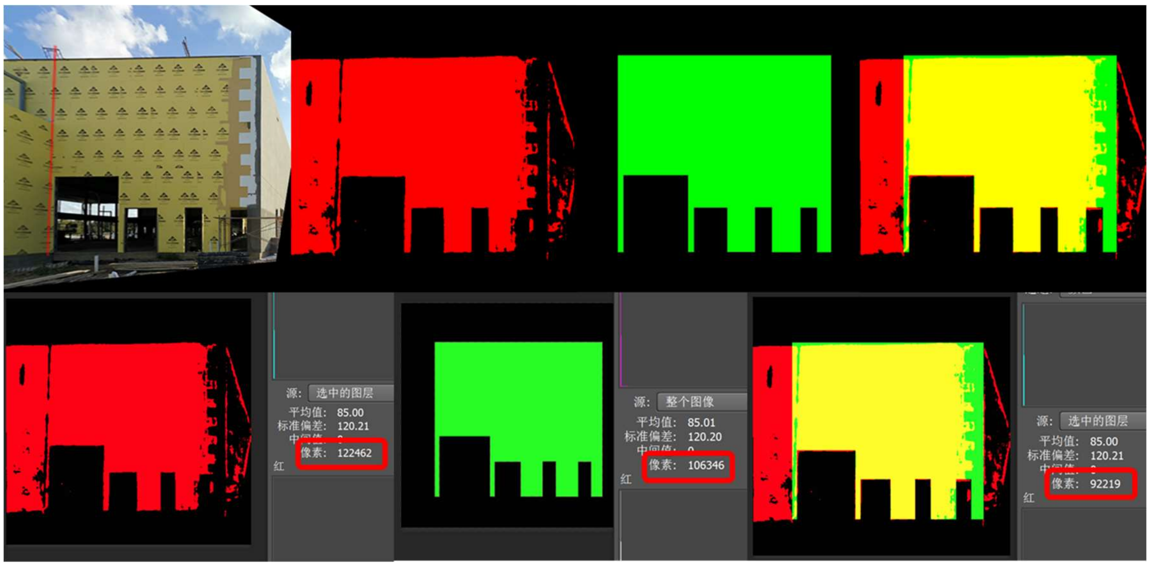

Figure 19 The example of calculation the completion rate of sheathing plywood

As shown in Figure 19, the red area indicates the sheathing plywood area detected in the picture by the deep learning model, and its pixel number is 122462. The green area represents the area of that wall in the BIM model, and its pixel number is 106346. By overlapping the two pictures, and the overlapping yellow area is the intersected area under the reference area. The pixel number of the yellow area is 92219. By making the ratio of the pixel number of the overlapping part and the pixel number of the whole wall, and the predicted completion rate of this wall can be known. According to the calculation, the completion rate is 86.7%. The completion rate of this wall sheathing is nearly 100%, but the right side of the wall has been daubed with some cement bricks, so the predicted completion rate has been reduced, as shown in Table 3.

Table 3 Completion rate of sheathing plywood

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Sheathing	122462	92219	106346	86.72%

Similarly, the completion rate of other types of walls can be calculated. As shown in Figure 20 and Table 4, the blue area is cement brick walls detected, and the sky-blue area was obtained after superimposing the blue detected area with the green BIM section wall. The completion rate was 70.35%. As the scaffold system in front of the wall blocked the wall, the area of the detected wall decreased, thus reducing the estimated completion rate.

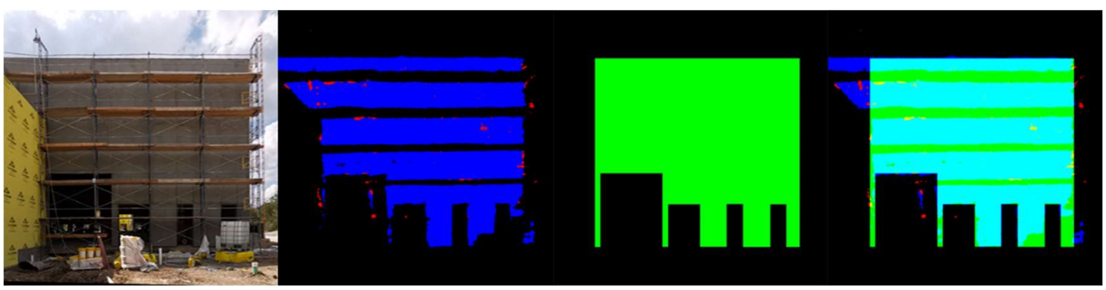

Figure 20 The example of calculation the completion rate of cement brick wall

Table 4 Completion rate of cement brick wall

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Concrete brick	80316	74813	106346	70.35%

Just like the figures above, due to the shielding of scaffolding, the identification of the external wall is not complete, resulting in the inaccurately estimated completion rate. The result is shown in Figure 21 and Table 5.

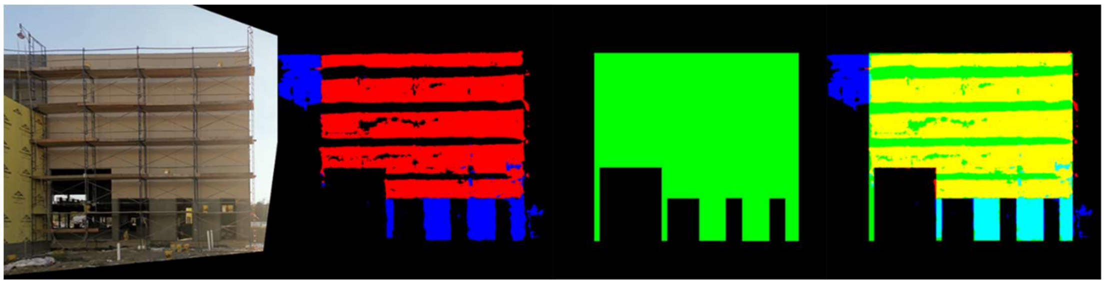

Figure 21 The example of calculation the completion rate of external wall

Table 5 Completion rate of external wall

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Exterior	69107	68052	106346	63.99%

The installation process is captured on the wall in Figure 22, so the half-way progress of the work of plywood installation can be detected. In Table 6, as sheathing plywood showed half installed, the deep learning model predicts a 54.54% completion rate for this photo.

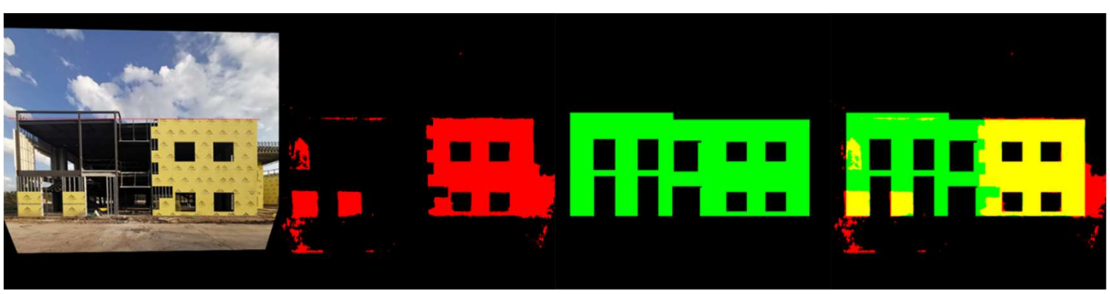

Figure 22 The example of calculation the completion rate of sheathing plywood at stage 1

Table 6 Completion rate of sheathing plywood at stage 1

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Sheathing	38752	32111	58875	54.54%

After two days, the last photo was taken. In this photo, the installation of sheathing plywood has nearly finished, and the completion rate given by the model is 95.93%. In the case of no occlusion, the model can accurately describe the progress of the work.

Figure 23 The example of calculation the completion rate of sheathing plywood at stage 2

Table 7 Completion rate of sheathing plywood at stage 2

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Sheathing	64522	56482	58875	95.93%

4. CONCLUSIONS

4.1. Research Conclusion and Limitations

The construction progress assessment under this methodology is enforceable. However, there are limitations, no matter in the operational processes or the type of detected items.

4.1.1. Accuracy depends on the data set

For detected objects, if the data set is good enough, the model can provide accurate segmentation. The influencing factors include and not limited to:

- 1) The weather or the intensity of sunlight in the photo can affect the accuracy of the model's detection.
- 2) The materials of building components used in different projects are different, which makes it difficult for the model to be used in other projects.

4.1.2. Shielding will reduce the detection accuracy

Shielding will reduce detection accuracy, and it will reduce the credibility of progress detection. The shielding includes not limited to workers, scaffolding systems, uninstalled materials, cranes, and hoists. As shown in Figure 24, in the absence of scaffolds and other obstruction, the accuracy of the model extraction of target building objects is much improved.

Figure 24 The accuracy of images with and without occlusion

4.1.3. The shape may affect the accuracy

Large and massive inspection objects are natural to be detected, such as columns, beams, and walls. However, for some small-size objects, such as steel structure, because they make up a small proportion of the picture, as shown in Figure 25. In this way, the noise in the picture accounts for a large proportion, thus affecting the accuracy of progress detection.

Figure 25 The small-size structure like steel structure

4.1.4. Too much manual intervention

The whole process of manual intervention lacks automaticity, which results in low efficiency.

4.1.5. Data set making

The process of making a dataset, especially making labels for each trained image is time consuming and painful.

4.2. Research Contributions

Though the method of construction progress monitoring in this test cannot be put into use at present, it puts forward the idea of automatic object detection and progress management by using semantic segmentation in the field of construction.

It is hoped that the problems found in this experiment can be resolved in the future.

REFERENCES

[1] Barrie, D. S., & Paulson, B. C. (1992). Professional construction management: including CM, design-construct, and general contracting. McGraw-Hill Science/Engineering/Math.

[2] Gould, F. E. (2005). Managing the construction process. Pearson Education India.

[3] Kerzner, H. (2017). Project management metrics, KPIs, and dashboards: a guide to measuring and monitoring project performance. John Wiley & Sons.

[4] Jung, Y., & Kang, S. (2007). Knowledge-based standard progress measurement for integrated cost and schedule performance control. Journal of Construction Engineering and Management, 133(1), 10-21.

[5] Fleming, Q., & Koppleman, J. (1996). Earned value management system. EIA Publication No. ANSI/EIA-748-1998.

[6] Bosché, F. (2010). Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. Advanced engineering informatics, 24(1), 107-118.

[7] Omar, T., & Nehdi, M. L. (2016). Data acquisition technologies for construction progress tracking. Automation in Construction, 70, 143-155.

[8] Russell Stuart, J., & Norvig, P. (2009). Artificial intelligence: A modern approach. Prentice Hall.

[9] Nilsson, N. J. (2014). Principles of artificial intelligence. Morgan Kaufmann.

[10] Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

[11] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160, 3-24.

[12] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

[13] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436.

[14] Sonka, M., Hlavac, V., & Boyle, R. (2008). Digital image processing and computer vision. Cengage Learning.

[15] Bovik, A. C. (2010). Handbook of image and video processing. Academic press.

[16] Fridman, L., Jenik, B., & Reimer, B. (2017). Arguing machines: Perceptioncontrol system redundancy and edge case discovery in real-world autonomous driving. arXiv preprint arXiv:1710.04459.

[17] Yang, J., Park, M. W., Vela, P. A., & Golparvar-Fard, M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future. Advanced Engineering Informatics, 29(2), 211-224.

[18] El-Omari, S., & Moselhi, O. (2009). Data acquisition from construction sites for tracking purposes. Engineering, Construction and Architectural Management, 16(5), 490-503.

[19] Park, H. S., Lee, H. M., Adeli, H., & Lee, I. (2007). A new approach for health monitoring of structures: terrestrial laser scanning. Computer-Aided Civil and Infrastructure Engineering, 22(1), 19-30.

[20] El-Omari, S., & Moselhi, O. (2011). Integrating automated data acquisition technologies for progress reporting of construction projects. Automation in Construction, 20(6), 699-705

[21] Khoury, H., Chdid, D., Oueis, R., Elhajj, I., & Asmar, D. (2015). Infrastructureless approach for ubiquitous user location tracking in construction environments. Automation in Construction, 56, 47-66.

[22] Pongnumkul, S., Chaovalit, P., & Surasvadi, N. (2015). Applications of smartphone-based sensors in agriculture: a systematic review of research. Journal of Sensors, 2015.

[23] Zhu, Z., & Brilakis, I. (2009). Comparison of optical sensor-based spatial data collection techniques for civil infrastructure modeling. Journal of Computing in Civil Engineering, 23(3), 170-177.

[24] Ahmed, M., Haas, C., Shahi, A., Aryan, A., West, J., & Haas, R. (2011). Rapid tracking of pipe-works progress using digital photogrammetry. In Proceedings of the 9th Construction Specialty Conference, Ottawa, Ontario, Canada (pp. 14-17).

[25] Ibrahim, Y. M., Lukins, T. C., Zhang, X., Trucco, E., & Kaka, A. P. (2009). Towards automated progress assessment of workpackage components in construction projects using computer vision. Advanced Engineering Informatics, 23(1), 93-103.

[26] Golparvar-Fard, M., Peña-Mora, F., Arboleda, C. A., & Lee, S. (2009). Visualization of construction progress monitoring with 4D simulation model overlaid on time-lapsed photographs. Journal of computing in civil engineering, 23(6), 391-404.

[27] Hidaka, N., Michikawa, T., Motamedi, A., Yabuki, N., & Fukuda, T. (2018). Polygonization of point clouds of repetitive components in civil infrastructure based on geometric similarities. Automation in Construction, 86, 99-117.

[28] Braun, A., Tuttas, S., Borrmann, A., & Stilla, U. (2015). A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds. ITcon, 20(8), 68-79.

[29] Hamledari, H., McCabe, B., and Davari, S. (2017). Automated computer vision-based detection of components of under-construction indoor partitions. Automation in Construction, 74, 78-94.

[30] Memon, Z. A., Majid, M. Z. A., & Mustaffar, M. (2005). An automatic project progress monitoring model by integrating auto CAD and digital photos. In Computing in Civil Engineering (2005) (pp. 1-13).

[31] Roh, S., Aziz, Z., and Peña-Mora, F. (2011). An object-based 3D walk-through model for interior construction progress monitoring. Automation in Construction, 20, 66-75.

[32] Xu, Y., Tuttas, S., Hoegner, L. and Stilla, U. (2018). Reconstruction of scaffolds from a photogrammetric point cloud of construction sites using a novel 3D local feature descriptor. Automation in Construction, 85, 76-95.

[33] Bruan, A., Tuttas, A., Borrmann, A., and Stilla, U. (2014). Comparison of Photogrammetric Point Clouds with BIM Building Elements for Construction Progress Monitoring. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(3).

[34] Han, K. K., & Golparvar-Fard, M. (2015). Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs. Automation in construction, 53, 44-57.

[35] Son, H., & Kim, C. (2010). 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring. Automation in Construction, 19(7), 844-854.

[36] Wu, Y., Kim, H., Kim, C., & Han, S. H. (2009). Object recognition in construction-site images using 3D CAD-based filtering. Journal of Computing in Civil Engineering, 24(1), 56-64.

[37] Fang, W., Ding, L., Zhong, B., Love, P.E.D., and Luo, H. (2018). Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach. Advance Engineering Information, 37, 139-149.

[38] Treiber, M.A. (2010). An Introduction to Object Recognition: Selected Algorithms for a Wide Variety of Applications.", Springer Science & Business Media. ISBN-13: 9781849962346.

[39] Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., and Omata, H. (2018). Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone. arXiv, 2018.

[40] Liaw, A., Wiener, M. :2002, Classification and regression by randomForest. R news, 2(3), 18-22.

[41] Krizhevsky, A., Sutskever, I., Hinton, E.: 2012, Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[42] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[43] Badrinarayanan, V., Kendall, A. and Cipolla, R.: 2017, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.

[44] Lin, G., Milan, A., Shen, C., Reid, I. :2017, Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In IEEE conference on computer vision and pattern recognition (CVPR) (Vol. 1, No. 2, p. 3).

[45] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[46] Cha, Y. J., Chen, J. G., & Büyüköztürk, O. (2017). Output-only computer vision-based damage detection using phase-based optical flow and unscented Kalman filters. Engineering Structures, 132, 300-313.

[47] Fang, Q., Li, H., Luo, X., Ding, L., Rose, T. M., An, W., & Yu, Y. (2018). A deep learning-based method for detecting non-certified work on construction sites. Advanced Engineering Informatics, 35, 56-68.

APPENDIX A

APPENDIX A FIGURE

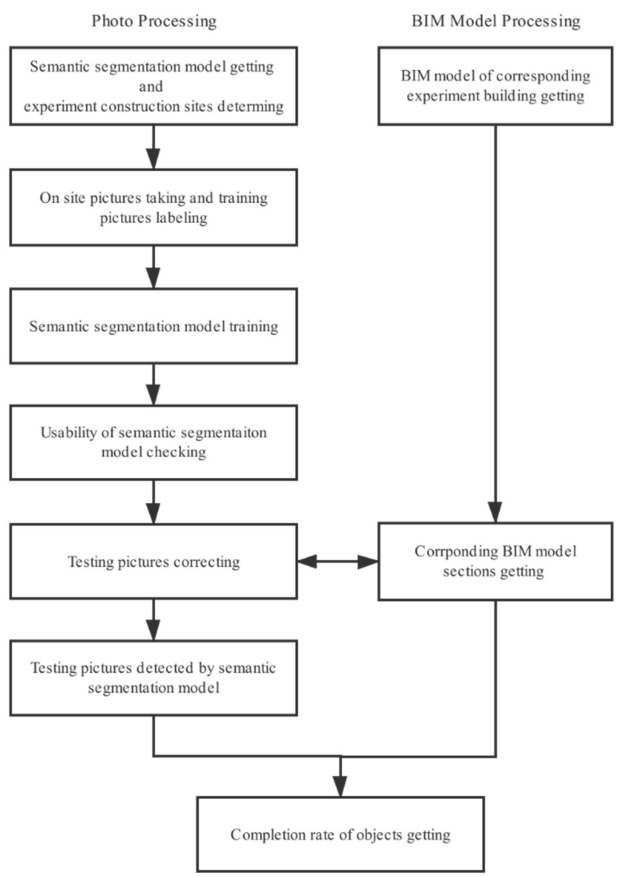

Figure 1 Flow chart of the research methodology

Figure 2 A rendering of the church and corresponding BIM model

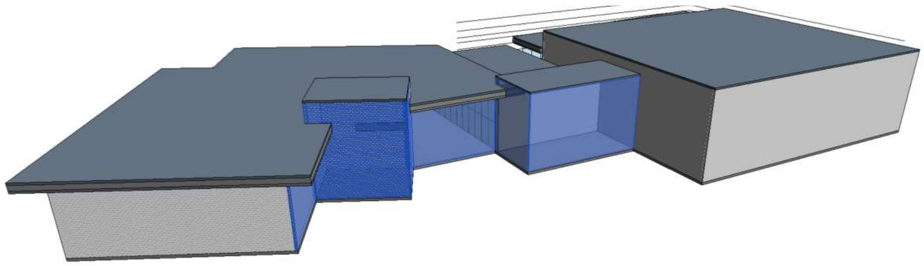

Figure 3 Training image collection area for sheathing plywood

Figure 4 Training images and corresponding labeled images for detecting sheathing plywood

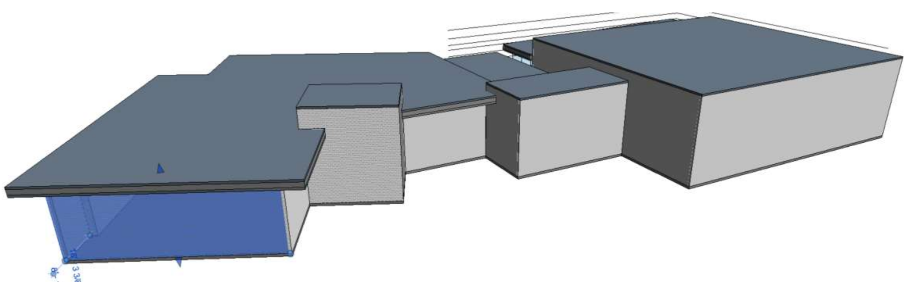

Figure 5 Testing image collection area for sheathing plywood

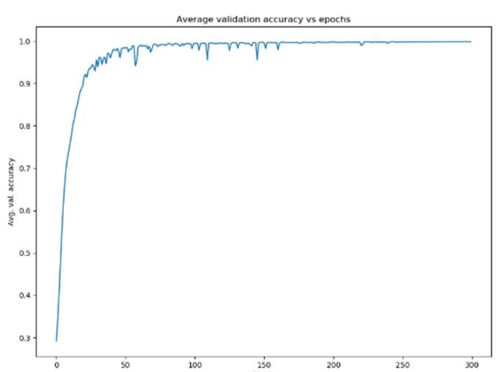

Figure 6 Average validation accuracy vs. epochs of the first model

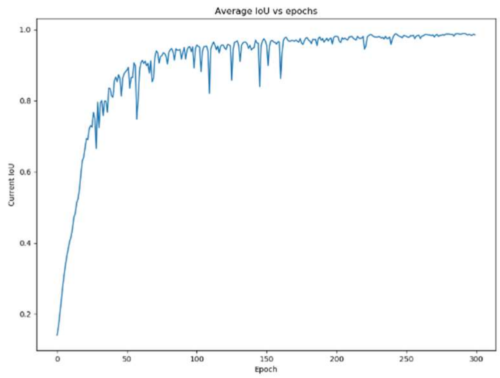

Figure 7 Average IoU vs. epochs of the first model

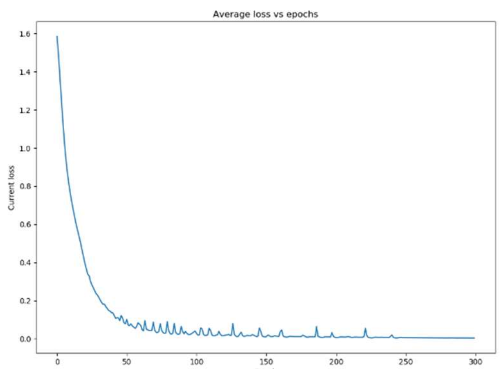

Figure 8 Average loss vs. epochs of the first model

Figure 9 The construction process of the external wall

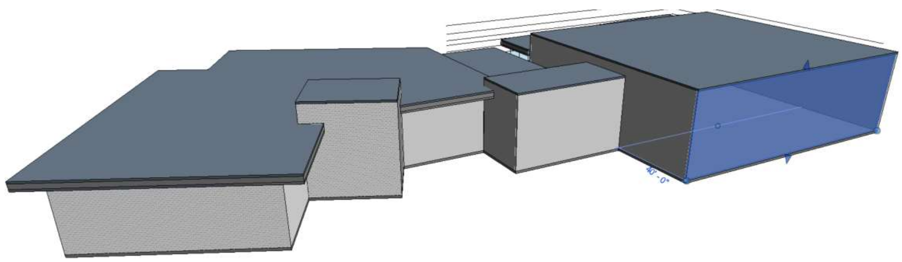

Figure 10 Training image collection area for detecting cement brick wall and external wall

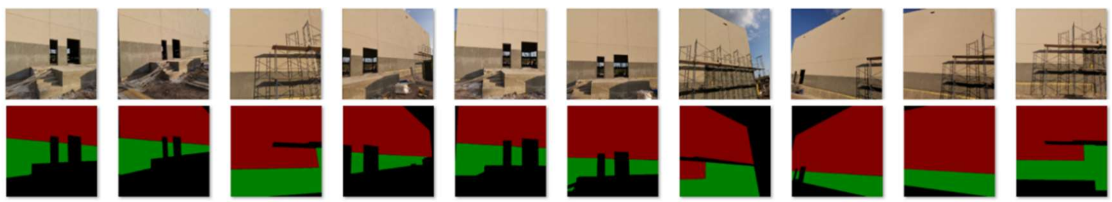

Figure 11 Training images and corresponding labeled images for cement brick wall and external wall

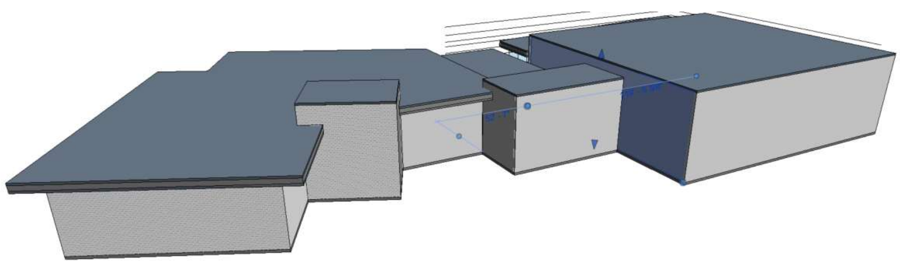

Figure 12 Testing image collection area for cement brick wall and external wall

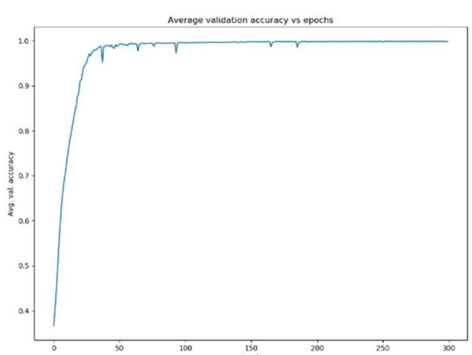

Figure 13 Average validation accuracy vs. epochs of the second model

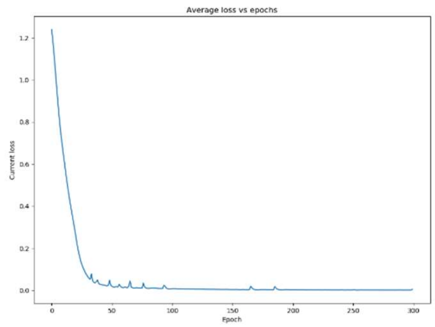

Figure 14 Average IoU vs. epochs of the second model

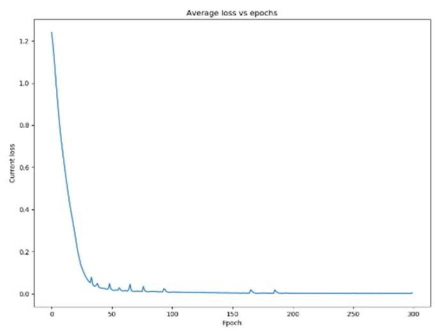

Figure 15 Average loss vs. epochs of the second model

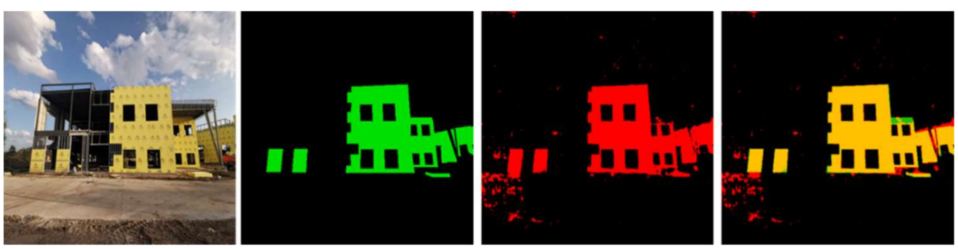

Figure 16 The example of pictures testing the accuracy of these models

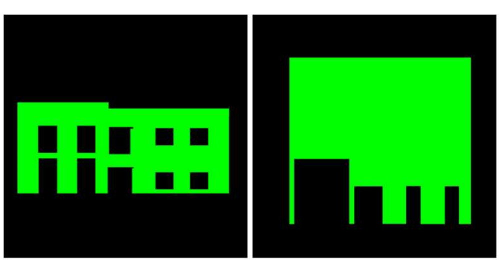

Figure 17 The pictures captured from the BIM model

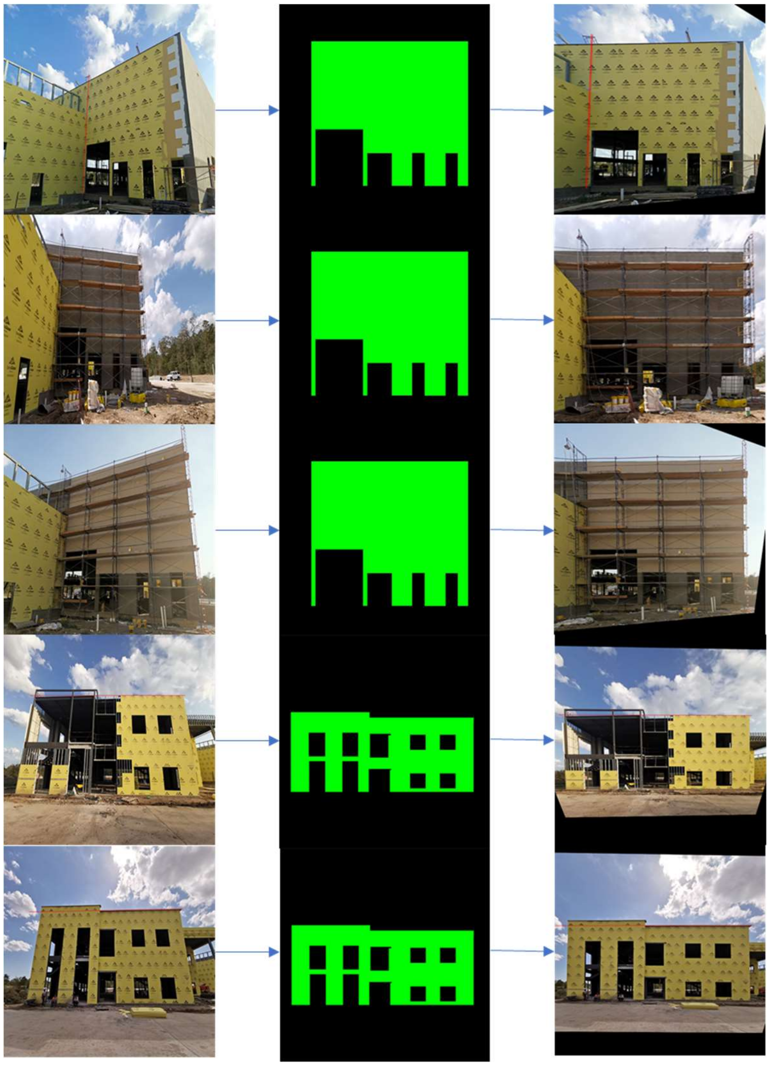

Figure 18 The images before and after adjustment

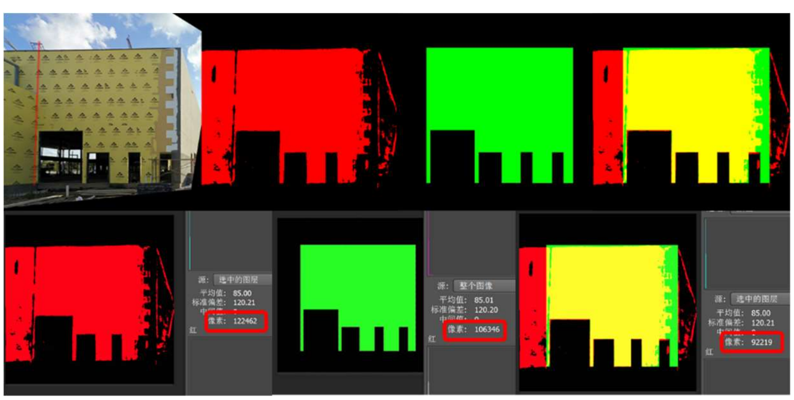

Figure 19 The example of calculation the completion rate of sheathing plywood

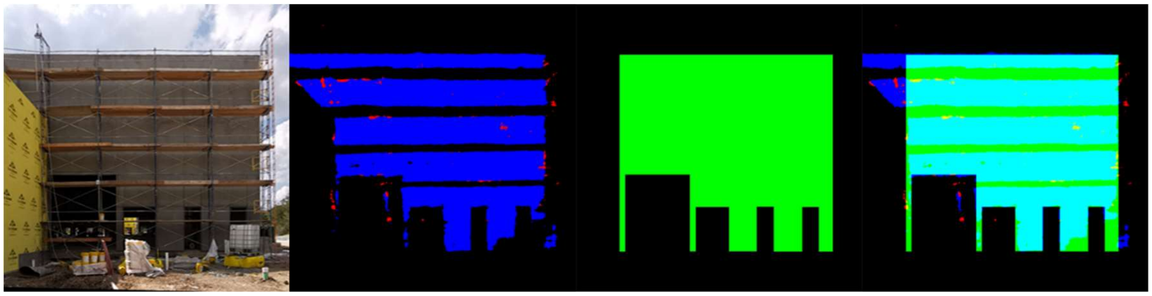

Figure 20 The example of calculation the completion rate of cement brick wall

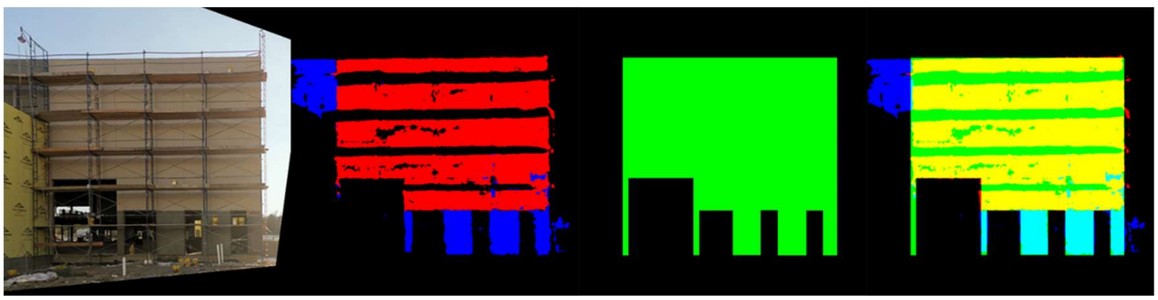

Figure 21 The example of calculation the completion rate of external wall

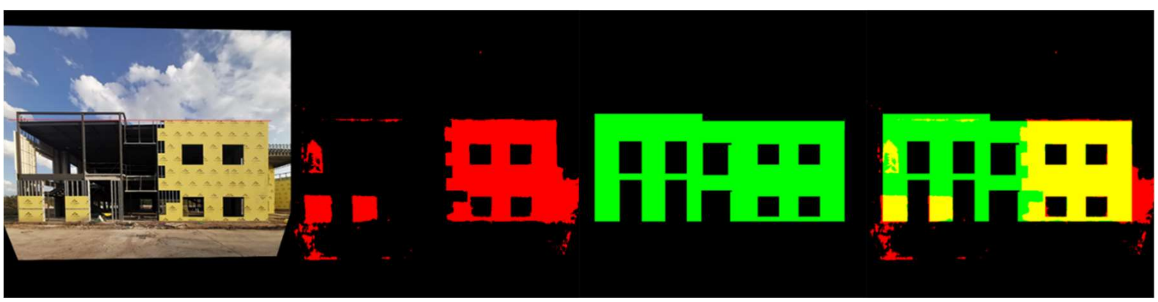

Figure 22 The example of calculation the completion rate of sheathing plywood at stage 1

Figure 23 The example of calculation the completion rate of sheathing plywood at stage 2

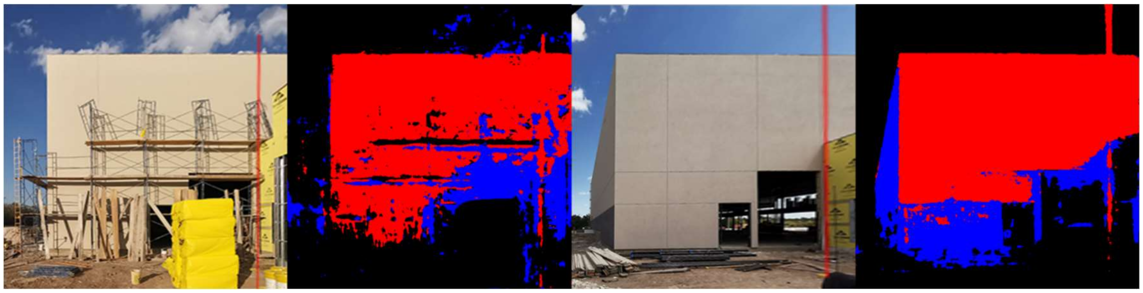

Figure 24 The accuracy of images with and without occlusion

Figure 25 The small-size structure like steel structure

APPENDIX B

APPENDIX B TABLE

Table 1 Existing construction progress measurement method

<table>
<tr><th>Research</th><th>Measurement Direction</th><th>Measurement Method (CII)</th><th>Measurement Target (CII)</th><th>Measurement Method (Fleming & Koppleman)</th><th>Measurement Target (Fleming & Koppleman)</th></tr>
<tr><td rowspan="6">Fleming & Koppleman & CII</td><td>Physical progress measurement</td><td>Unit completed</td><td>Installed quantity</td><td>Installed elements counting</td><td>Installed quantity</td></tr>
<tr><td rowspan="3">Estimated percent complete</td><td>Incremental milestone</td><td>Milestone</td><td rowspan="3">Percent complete & milestones gates</td><td rowspan="3">Progress state based on milestone</td></tr>
<tr><td>Start/finish, supervisor</td><td>Start /finish point of work</td></tr>
<tr><td>Opinion</td><td>Progress state</td></tr>
<tr><td rowspan="2">Earned value</td><td>Cost ratio</td><td>None</td><td>None</td><td>None</td></tr>
<tr><td>Weighted or equivalent units</td><td>Finish point or progress state of work</td><td>Weighted milestones</td><td>Finish point of weighted milestone</td></tr>
</table>

Table 2 The result of the accuracy in these two models

Test_name	Test_accuracy	Precision	Recall	Mean IoU
1	0.975208	0.974974	0.975208	0.901236
2	0.959469	0.959443	0.959469	0.85086
3	0.966537	0.965925	0.966537	0.867597
4	0.964077	0.964132	0.964077	0.861339
5	0.982803	0.982665	0.982803	0.935806
6	0.979565	0.979426	0.979565	0.926168
7	0.981071	0.981036	0.981071	0.932666
8	0.984474	0.984378	0.984474	0.942619
9	0.98801	0.987955	0.98801	0.959254
10	0.980515	0.980552	0.980515	0.9556
11	0.983551	0.983591	0.983551	0.958964
Explanation	Yellow/Green	Yellow/Red	Yellow/Green	Yellow/(Green+Red)

Table 3 Completion rate of sheathing plywood

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Sheathing	122462	92219	106346	86.72%

Table 4 Completion rate of cement brick wall

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Concrete brick	80316	74813	106346	70.35%

Table 5 Completion rate of external wall

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Exterior	69107	68052	106346	63.99%

Table 6 Completion rate of sheathing plywood at stage 1

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Sheathing	38752	32111	58875	54.54%

Table 7 Completion rate of sheathing plywood at stage 2

Number of pixels	Detected	Overlapped	Ground Truth	Completion Rate (%)
Left Wall-Sheathing	64522	56482	58875	95.93%