

Article

Identification and Extracting Method of Exterior Building Information on 3D Map

Donghwa Shon ¹, Byeongjoon Noh ² and Nahyang Byun ^{1,*}¹ Department of Architecture, Chungbuk National University, Cheongju 28644, Korea; dhshon@cbnu.ac.kr² Applied Science Research Institute, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea; powernoh@kaist.ac.kr

* Correspondence: nhbyun@chungbuk.ac.kr

Abstract: Although the Korean government has provided high-quality architectural building information for a long period of time, its focus on administrative details over three-dimensional (3D) architectural mapping and data collection has hindered progress. This study presents a basic method for extracting exterior building information for the purpose of 3D mapping using deep learning and digital image processing. The method identifies and classifies objects by using the fast regional convolutional neural network model. The results show an accuracy of 93% in the detection of façade and 91% window detection; this could be further improved by more clearly defining the boundaries of windows and reducing data noise. The additional metadata provided by the proposed method could, in the future, be included in building information modeling databases to facilitate structural analyses or reconstruction efforts.

Keywords: deep learning; image processing; 3D map; exterior building information; building façade; image identification; image extraction

Citation: Shon, D.; Noh, B.; Byun, N. Identification and Extracting Method of Exterior Building Information on 3D Map. *Buildings* **2022**, *12*, 452. <https://doi.org/10.3390/buildings12040452>

Academic Editor: David J. Edwards

Received: 31 December 2021

Accepted: 20 March 2022

Published: 6 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Korean government implemented the “Act on the Promotion of the Provision and Use of Public Data” in 2013, which discloses data owned by the national and local governments to the public for the purpose of economic development and quality-of-life improvements [1]. Information owned by government institutions is provided through the Public Data Portal, which, as of Jan 2022, includes 107 national core datasets and 68,323 datasets. The datasets of both quantity and quality are continuously accumulating, and the number of successful business projects that use this public information is also increasing [2]. For urban architecture, 3785 datasets of administrative information are provided, including those for buildings, houses, land, the purpose of using, urban planning and regeneration, transportation, real estate, infrastructure, IoT sensing, spatial information, safety, energy use, etc.; furthermore, the government continues to expand and upgrade the scope of the data collected. Among them, Architectural Administration Information has disclosed 280 million cases of data since 2015 [3–5].

However, compared to other fields, the demand for information related to architecture is limited because such information consists primarily of administrative information. Although the government is making efforts to promote architectural administrative information, such as providing web services, building a cloud-based system, and building a big data hub, there is still not much demand, and there are no successful cases [3,4,6,7]. It is because architectural administration information primarily comprises licensing information, building ledgers, closure and cancellation ledgers, life cycle management and energy use; information such as building location, maintenance, accreditation and permission details, date, size (e.g., area and number of floors, etc.), type (e.g., use of purpose,

structure, etc.), blueprint (e.g., layout drawing that does not violate security), interested parties, and energy (e.g., usage, grade, etc.) is limited for the private use [3,4,7].

For the utilization of architectural administrative information, data expansion through connection with private or other public data can be considered. However, the data that can be considered the most is the exterior building information that directly affects human visual perception. Modern architects such as Le Corbusier and theorists such as Jenks have mentioned the importance of visual information on the urban environment. [8,9]. In addition, visual information has great potential for use because it directly relates to public information, such as aesthetics of buildings, legal regulations, safety, security, regional characteristics, and city images and landscapes. In this way, if the exterior building information is linked with the architectural administration information, the building information's scalability and completeness increase, as does the utility value in the private and public fields.

This point of view is the same as the data utilization policy of the Korean government and the direction of the BIM (building information modeling) policy [10,11]. In July 2020, the government announced the Digital New Deal policy for digital transformation centered on D.N.A. (data, network, AI) [10], in particular, in the construction field, the opening of administrative information and BIM-centered data transformation in the fields of design, construction, and building management [10]. The government announced the Architectural BIM Activation Roadmap (December 2020) and the Roadmap for Digital Transformation of the Construction Industry (June 2021) to expand the application of BIM, and BIM application will be mandatory in all construction industries from 2025 [11]. BIM is a design technology based on a 3D (three-dimensional) shape model and it contains various attributes [12]. Because it has a higher degree of completeness than the existing 2D drawings, it is widely applied to the entire construction process of architectural planning, design, construction, and maintenance [12]. It is also a standard information platform that enables collaboration and communication in various fields [13]. High-quality BIM information can secure economic feasibility, time savings, work efficiency, productivity, and safety in the construction industry [14,15]. In addition, it can promote the development of the construction industry by combining it with the fourth industrial technology, such as AI, big data, digital twin, and automation design and construction [16].

However, since BIM-applied building design or construction only applies to new buildings, there are very few cases where BIM is applied to all buildings in Korea. As of 2021, BIM design is compulsory only for large-scale public projects of more than USD 25 million ordered by the government, and although the scope of application is gradually expanding, it is expected that it will take time for the entire building to be transformed into BIM. However, even if it takes time according to government policies or social changes, most existing buildings will eventually be digitally converted to BIM. In order to respond to these changes, it is necessary to build information gradually by linking attribute information such as building shape, elevation components, and materials based on architectural administrative information constructed as a dataset. First, building exterior information related to shape information should be linked and constructed as a dataset. In particular, since it takes much money, time, and effort to deal with all the buildings in the country one by one, an effective method is needed to overcome it.

Therefore, in this study, images of buildings that have already been built in the private or public are used. In recent years, Korean web sites have improved the quality and quantity of urban data with architectural image information, urban exteriors, street views, and other 3D (three-dimensional) maps. Although the resolution of these data is lower than that of photos taken on site, the potential for future use is very high because of the vast amount of information built nationwide, and the resolution is increasing every year. Studies from various perspectives identify elements using city images provided on the web. For example, No et al. proposed a deep learning algorithm that detects the characteristic information of Braille blocks on the road using street view images to diagnose a

space with insufficient installation [17]. Seiferling et al. (2017) presented a method to detect and quantify the condition of street trees through Street View images for tree management [18]. Yu et al. (2019) tried to build a BIM database after extracting visual information of buildings using Street View and two-dimensional aerial images to identify buildings vulnerable to earthquakes [19]. In structure and construction, images are sometimes used from the viewpoint of construction quality or safety. For example, Santarsiero et al. (2021) presented a method of collecting bridge survey information from bridge images through OpenStreetMap (OSM) and Street View to suggest a bridge risk classification procedure [20].

Unlike other previous studies, this study uses 3D Map to extract building exterior information. This is because image distortion occurs in the Street View image depending on the shooting point, and buildings are sometimes obscured by plantings or automobiles. However, the 3D Map has no obstacles to fully identifying the exterior building information. This is because 3D maps are created by the photogrammetry technique that creates 3D information through overlapping objects after multi-angle aerial photography. However, there are limits to extracting all exterior building information nationwide by human resources. As computer vision (CV) and deep learning technologies advance, time, effort, and cost can be saved.

2. Advances in Image Identification and Extraction Technology

2.1. 3D Maps

National and international private IT corporations, such as Google, Daum (Kakao), and Naver, provide both two-dimensional (2D) and 3D map services—including street views, aerial satellite photos, and spatial building information—that show the exteriors of buildings [21–25]. Three-dimensional maps provide users with a more directly immersive urban experience, and the continuous improvement of autonomous drones and vehicular cameras help to improve the resolution and sophistication of 3D maps. These maps differ from 2D satellite maps: 3D mapping is a method of photographing the exteriors of architectural structures from several angles using drones or aircraft, then synthesizing the photographed images in three dimensions based on coordinate values to create a map (i.e., photogrammetry). The national government employed the spatial information open platform “Vworld” to provide 2D and 3D spatial information (i.e., physical and logical spaces, as well as properties belonging to figures) to connected public administration databases and is currently developing a 3D shape model based on architectural building forms and floor heights [21]. In particular, the Seoul Metropolitan Government has constructed a 3D map (S-Map) of the entire 605 km² of Seoul that features greater location accuracy and higher resolution than government or private maps [22]; furthermore, images generated by S-Map are easily extracted with no obstacles covering the exterior of the buildings (See Figure 1).

Owing to the rapid development of 3D maps, users are provided with convenient and detailed building information, but these are currently used merely as a visual reference. We propose that composing a dataset for exterior building architectural information will contribute to future research and reconstruction efforts.

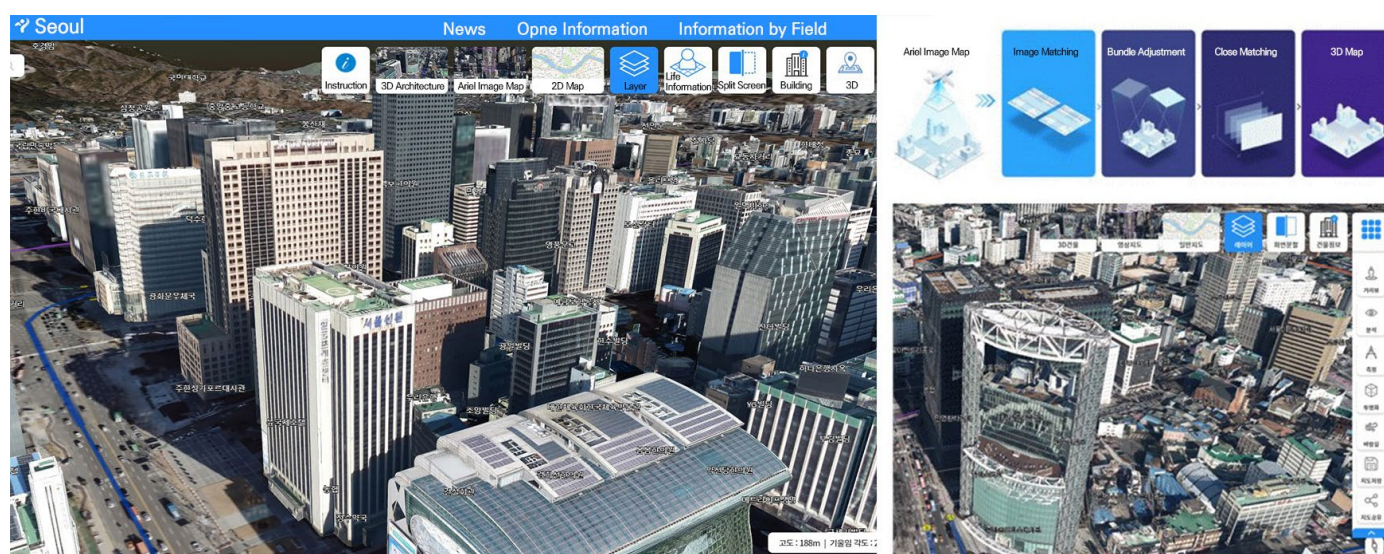


Figure 1. Seoul Metropolitan Government S-Map screen.

2.2. Advances in Image Processing and Deep Learning Technology

CV technology is the current industry standard for image processing. Rapidly improved in the 2000s, CV is primarily used for the identification, division, tracking, mapping, estimation, and exploration of objects [26–28]. Based on knowledge-based methodology, it primarily focuses on feature extraction methods via the detection of contours, corners, and key points within the image. CV's application is divided into image processing (e.g., removing noise and emphasizing features) and background analysis (e.g., extracting or identifying information) [29]. Imaging processing technologies based on CV were able to detect objects but suffered from an inability to properly classify those objects.

After 2010, image processing research developed to a practical utility stage in many fields, such as autonomous cars, fingerprinting, facial recognition, and robot control (with the rapid development of artificial intelligence) [28]. It was considered a significant milestone when a technique for classifying the extracted features was developed. Meanwhile, the importance of data preprocessing and conversion processes such as spatial domain conversion, geometric conversion, and frequency conversion have diminished since the early 2000s, supplanted by the detection and classification of objects and motion detection made possible through image inputs. Based on CNNs (convolutional neuron networks), deep learning technology for image processing is being developed into platforms such as YOLO (You Only Look Once) and Detectron based on several object detection models such as fast and faster-R. These platforms also provide pretrained models and high-capacity image data for commonly used objects such as people, vehicles, chairs, and buildings.

Deep-learning-based image identification technology is used in many fields: in medical fields, these processing techniques are used to analyze chest X-ray images, MRI images, and ophthalmic images, and to develop software that quickly and accurately diagnoses various diseases and aids in effective treatment. Google developed a system that can diagnose diabetic retinopathy using CNN technique-based retinal analysis, and Lunit INSIGHT developed a medical support software system for lung disease and breast cancer diagnosis [30,31]. In the fields of transportation and security, accidents are prevented by tracking and categorizing the movement of objects using closed-circuit television (CCTV) images; systems based on deep learning technology forecast the traffic volume of cars, taxis, bicycles, and other vehicles on roads, then create accident predictions by analyzing the movements of pedestrians and vehicles [26,32,33].

Prior to the application of deep-learning-based technology, any attempts to identify the exteriors of buildings were focused on the recognition of building openings. Several

approaches were developed, such as deciding whether the opening is glass through RGB (red, green, and blue) color values in elevated areas, as well as detecting entrances and structural combinations using lines, colors, and textures [34–36]. This differs from studies prior to deep learning in that identification elements are categorized along with algorithms that extract features from the images; most of the studies primarily focus on extracting building outlines, façade characteristics, and outer wall materials [37–39]. These later studies are similar to previous studies in how they identify the components of the façade; however, they are limited to the identification of shapes, elements, and materials.

Our study presents a multifold method of identifying elements in high-rise buildings by appropriately distinguishing image identification and deep learning techniques on 3D Maps. Similar to other studies, it does not target high-resolution building images; there may have a limit to creating sophisticated results. However, it is meant as basic research because it has the advantage of being scalable to be applied nationwide.

2.3. Deep Neural Networks and Video Detection

Early neural networks suffered from many shortcomings, including slow learning time, data supply and demand problems, and overfitting; the eventual resolution of these issues led to the discovery of a hidden layer of existing artificial neural networks being developed into a deep neural network (DNN) comprising multiple layers [40]. As with conventional neural networks, DNNs comprise input layers, hidden layers, and output layers, and are able to model complex nonlinear relationships. DNNs have been recently developed into different forms, including recurrent neural networks (RNNs) in natural language processing and convolutional neural networks (CNNs) in computer vision: CNNs are primarily used for image identification as artificial neural networks to imitate visual processing methods, whereas RNNs were designed for the purpose of processing time series data such as audio, sensors, and characters [41,42].

Given their utility in image identification, CNNs are the basis for major deep learning platforms involving object detection in images such as the aforementioned YOLO and Detectron2. With Tensorflow, a Python-based deep learning open-source package provided by Google, research can be conducted on several data mining, machine learning, and deep learning models; it also facilitates the construction of large visualization models designed using data flow graphs via the processing power of GPUs (graphics processing units) and CPUs (central processing units). In addition to Tensorflow, platforms such as PyTorch and Caffe2 can be used. Since YOLO is a deep learning model that specializes in detecting objects in images, it improved the computational speed of existing region-based convolutional neural network (RCNN)-series models, allowing one network to extract features, create bounding boxes, and categorize classes simultaneously [43].

3. Identification and Extracting Exterior Building Information

For our study, we selected a neighborhood in Jongno-gu, Seoul, as the case subject site for the identification of building exteriors and information extraction, because it contains a high concentration of high-rise buildings; any buildings that met the criteria were categorized as “high-rise buildings” (See Figure 2). This is because the images of high-rise buildings are high-resolution, and because it is necessary to establish a model which identifies components (e.g., location, area, shape, texture, material, color of window, entrance, pillar, etc.) within an architectural type of high-rise building that allows a framework to be established for application to other building types as well. Here, the model for detecting building façades is proposed by applying deep learning technology. The model largely comprises (1) an image extraction unit, (2) an image preprocessing unit, (3) a building façade recognition unit, and (4) a data management unit. The model structure is illustrated in Figure 3.

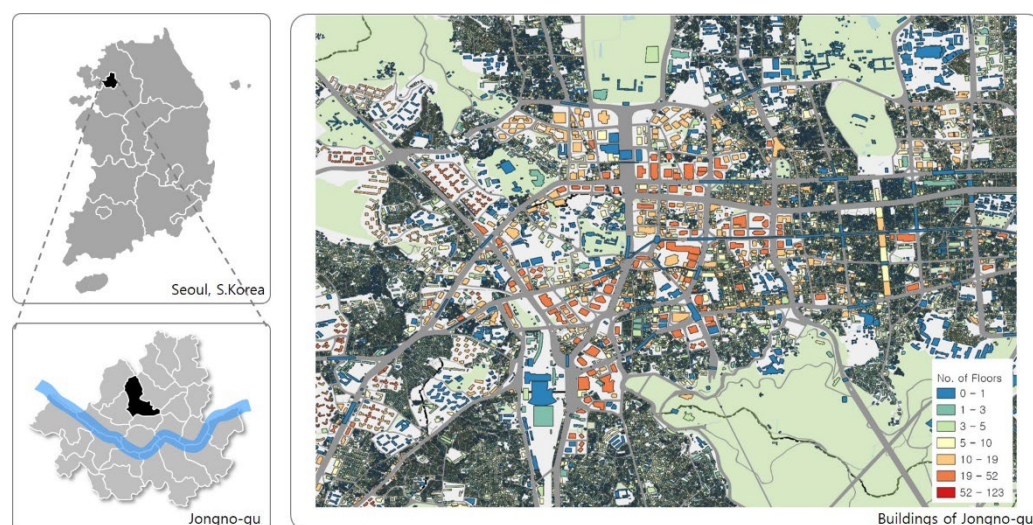


Figure 2. Analysis of target area within Jongno-gu.

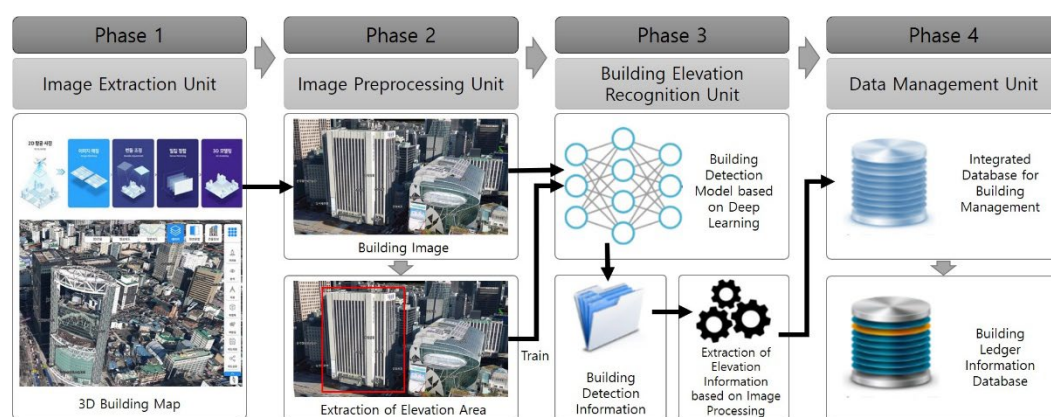


Figure 3. Model for identifying exterior images and extracting façade information using deep learning and image processing techniques.

First, a 3D building map image is used to identify the building exteriors and extract façade information. The image preprocessing unit sets the region of interest (RoI) on the façade of the building to train the deep learning model from the acquired image. The building façade recognition unit detects the building exteriors using a deep learning model wherein the images of building exteriors are learned, and extracts building façade information from the building that was identified. Finally, the identified information related to building façades is added to an integrated database for building management. We propose that this database be used in conjunction with the building ledger information database to provide better datasets for buildings.

3.1. Method Used to Acquire Building 3D Map Image Data

This section describes our method for acquiring images of building exteriors from 3D building maps. First, we used 3D building maps based on the information provided by S-Map [22]. S-Map provides user-friendly 3D-type building maps and enables users to set several layers, including administrative divisions, a state basic district system, road name information, subway information, and building names, all at different resolutions. Additionally, it provides the user with the ability to set altitude and tilt angle settings (see Figure 1).

S-Map supports its own application programming interface (API), but no third-party APIs for the extraction of 3D building images such as those we intend to use. Thus, we implemented a macro program to automatically move and capture the screen via AutoIt

software to acquire a full image of a 3D building [44]. It is trivial to perform these repetitive tasks by automating the mouse and keyboard movements; this software was used to extract approximately 500 façade images of buildings in Jongno-gu, Seoul. Because distortion occurs as façade increases, we focused on the front of each building as much as possible and set the extraction altitude and inclination angle to 114 m and 14°, respectively.

3.2. Preprocessing Method for Training the Deep Learning Model

This section describes our image processing method for training the deep learning model using the 3D building image acquired by the image extraction unit in the previous section. Generally, when a deep learning model is being trained to detect objects, coordinates and images for a section corresponding to an object area within an image—called a bounding box—are used as well. The bounding box for each building image is rectangular and is shown in the form of left upper coordinates (minx, miny) and right lower coordinates (maxx, maxy). Figure 4 illustrates an example of displaying an object area in an image as a bounding box.

We set a bounding box for the building areas from the extracted image information to train the deep learning building detection model. The settings for the bounding box were set manually on all images for the model learning dataset as well as the test dataset, and the VGG Image Annotator (VIA) open-source software [45] was utilized to extract bounding box coordinates. When training the deep learning model, images are required for learning, and it is crucial to include metadata or annotations containing object information in the images. Metadata contain information including image paths, sizes, and bounding box coordinates and are stored in either JavaScript Object Notation (JSON) or eXtensible Markup Language (XML) format; VIA supports object ranging and metadata extraction for training the model. Figure 4 presents the result of setting bounding boxes in 3D building images using the VIA software. In this study, a bounding box was set up to recognize the façade and windows of the building from the 3D building images.

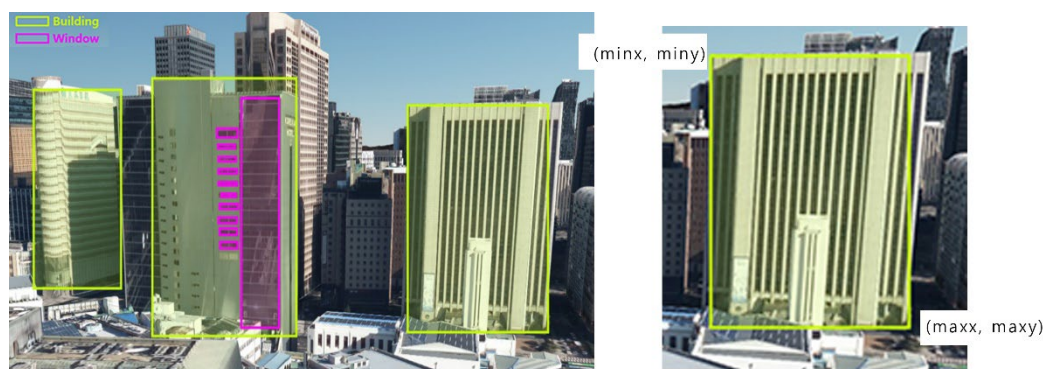


Figure 4. Example of bounding box, window identification, and coordinate metadata in the 3D building image.

Image metadata in this study were extracted in JSON format, the general structure of which is shown in Table 1.

Table 1. Object labeling metadata structure for learning.

Key Values		Description	Item Values (Example)	Data Format
	metadata	Image metadata information	{...}	DICTIONARY
└	filename	Image file name	"test_image.png"	STRING
└	size	File size	3202282	INT
└	regions	Bounding box properties information	[...]	LIST
└	shape_attributes	Object size information	{...}	DICTIONARY
└	name	Object area information	"rect"	STRING
└	x	Bounding box top left-hand corner x coordinate	643	INT
└	y	Bounding box top left-hand corner y coordinate	123	INT
└	width	Bounding box width	359	INT
└	height	Bounding box height	371	INT
└	region_attributes	Object attribute information	{...}	DICTIONARY
└	class	Object class	"building"	STRING
	file_attributes	Image file property information.	{...}	DICTIONARY
... (continued)				

Scheme 1 is an example of a JSON file used for learning. In the given example, building façade information was extracted from the 3D image files of "image1.png" and "image2.png", with the file sizes being approximately 3 MB (megabytes). The "region" key value includes attribute information such as a bounding box size, location, class information, and similar component information. The "image1.png" file contains information regarding two buildings and one window as well as location information (top left-hand corner x and y coordinates, length, and height) for each bounding box. Conversely, the "image2.png" file contains only two elements of building information.

```

{
  'image1_info': { 'filename': 'image1.png',
                  'size': 3202282,
                  'regions': [{ 'shape_attributes': { 'name' : 'rect',
                                                       'x'      : 643,
                                                       'y'      : 123,
                                                       'width'   : 359,
                                                       'height'  : 371 },
                               'region_attributes': { 'class' : 'building' }},
                             { 'shape_attributes': { 'name' : 'rect',
                                                       'x'      : 175,
                                                       'y'      : 415,
                                                       'width'   : 30,
                                                       'height'  : 23 },
                               'region_attributes': { 'class' : 'window'}},
                             { 'shape_attributes': { 'name' : 'rect',
                                                       'x'      : 689,
                                                       'y'      : 1048,
                                                       'width'   : 290,
                                                       'height'  : 454 },
                               'region_attributes': { 'class' : 'building' }}}},
    'file_attributes': {}},
  'image2_info': { 'filename': 'image2.png',
                  'size': 390471,
                  'regions': [{ 'shape_attributes': { 'name' : 'rect',
                                                       'x'      : 125,
                                                       'y'      : 213,
                                                       'width'   : 259,
                                                       'height'  : 471 },
                               'region_attributes': { 'class' : 'building' }},
                             { 'shape_attributes': { 'name' : 'rect',
                                                       'x'      : 689,
                                                       'y'      : 1048,
                                                       'width'   : 290,
                                                       'height'  : 454 },
                               'region_attributes': { 'class' : 'building' }}}},
    'file_attributes': {} }
}

```

Scheme 1. An example of JSON including labeled object information.

3.3. Building Façade Recognition Unit

This section describes one method for training a deep learning model and one for recognizing the façade of buildings using preprocessed image information and metadata. It also partially illustrates how various building information can be acquired using image processing techniques from the recognized building façade.

3.3.1. Deep-Learning-Model-Based Method to Recognize Building Façade

Prior to the development of deep learning models, object detection technology was based on in-image algorithms such as the scale-invariant feature transform (SIFT) and sped-up robust features (SURF), or on machine extraction algorithms including K-nearest neighbor (KNN) and support vector machine (SVM) to extract the main features in the image using mathematical estimates. The speed and accuracy of detecting objects in images has improved in conjunction with advances in deep learning technology. The major algorithms for object detection in images are as follows: DetectorNet [46], SPP Net [47],

VGG Net [48], R-CNN [32], fast/faster R-CNN [49–51], YOLO [52–55], and mask R-CNN [56].

In this study, 3D building images were learned by using faster R-CNN—an improved model of R-CNN—and then, these were employed for the detection and extraction of building façade information. The building images are not included in the COCO (Common Object in Context) dataset [57], so we needed to train the object detection model additionally. In general, the faster R-CNN model extracts a “regional proposal” in the image using the selective search algorithm and then extracts features in the image using the CNN model in each candidate area. Classification is performed to categorize objects using SVM based on the extracted features. In the faster R-CNN that emerges, a separate regional proposal network (RPN) is applied to extract feature maps. Additionally, the location of the object can be extracted by performing a bounding box regression operation, which has higher MAP (mean average precision). The structure of faster R-CNN is illustrated in Figure 5.

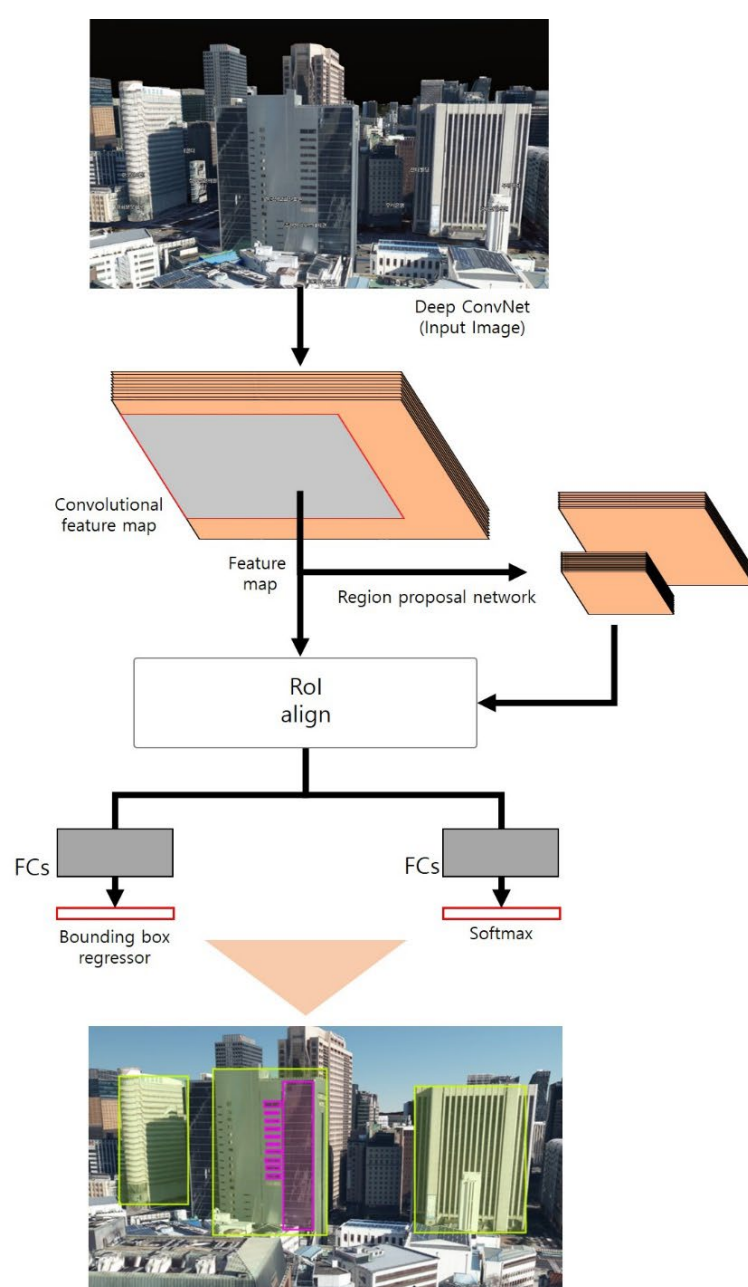


Figure 5. Faster R-CNN model structure.

When learning, label data in images (including images for learning, building façade bounding box information, and window bounding box information) are included in the JSON file for data configuration. In this study, the Detectron2 platform [58–60] provided by Facebook AI Research is used for training.

Using these methods, we were able to acquire a learning model wherein the deep learning model was finally up-to-date. Figure 6 illustrates the result of the 3D building image tested in the learned model. The results show that high-rise buildings are well-detected, but windows are not due to the distortion of the buildings and the overall image size. Section 4 outlines the results of the model learning and its performance.

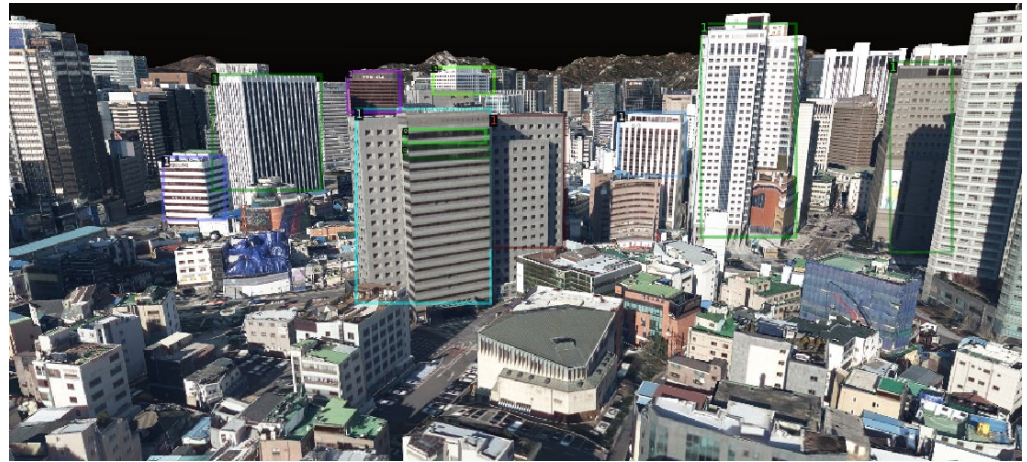


Figure 6. Test result image of the model that learned the 3D building image.

3.3.2. Method for Extraction of Building Façade Information

This section outlines the method for extracting the façade information of a building using the image processing technique based on the results presented through the deep learning model. The exterior details of the building and its façade vary, and the 3D building exterior image used in this study suffers from low image quality and unconventional shape. Thus, the scope of this study was to distinguish windows from walls among the exterior and façade information of the building, the method for which is twofold: we first employ a simple classification method, then a detailed classification method.

In the simple classification method for windows, only the existence or absence of windows is identified based on the façade of the building. As previously mentioned, windows can also be recognized through training window areas during façade training. While window areas exist within the façade area of the building, they are sometimes detected outside the façade area due to noise in the image area (see Figure 7). To prevent this, our study compared the locations between the façade areas and the window areas of the building, thus improving detection accuracy and facilitating the acquisition of a set of window areas (w_j) included in the building façade area (B_i). The equation for comparison between the detected building façade area (B_i) and the detected window area (w_j) is as follows:

$$W_i = \{w_x \mid \text{threshold} \leq |w_x \cap B_i| \leq |w_x|; x = 1, 2, \dots, m\}; i = 1, 2, \dots, n$$

where $|\cdot|$ is a size of area A , m is the number of detected windows, and n is the number of detected buildings.

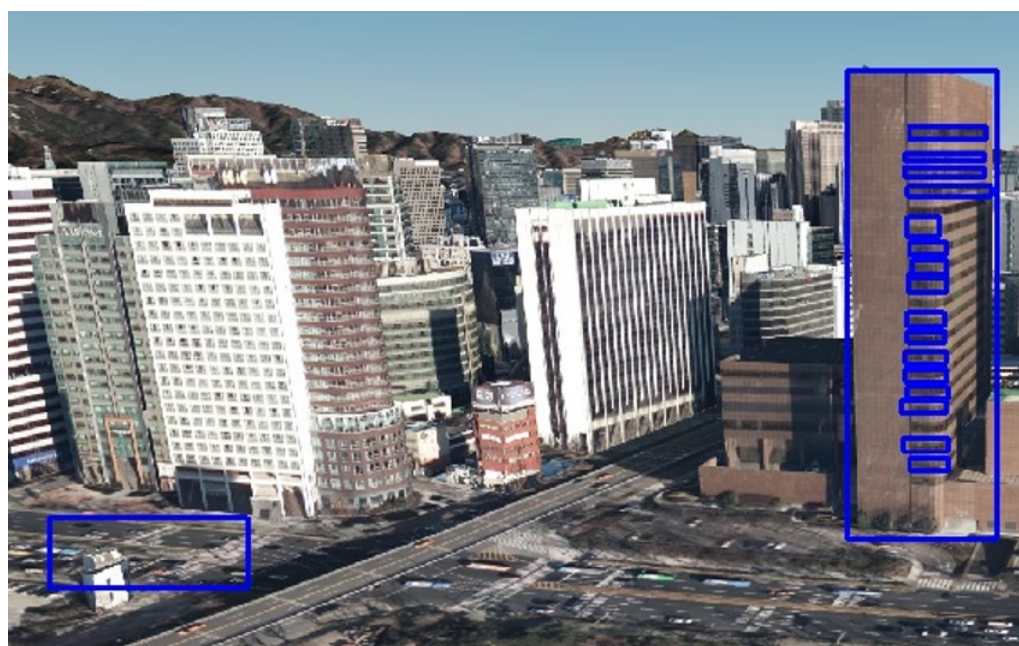


Figure 7. Example of cases in which the window area in the façade of buildings is properly detected (right) and where it is not (left).

When there is a window area within the façade area of the building, the building is categorized as having windows; otherwise, it is categorized as a building with no windows. This is used as part of the preprocessing in the detailed classification method that consists of two parts from which information is extracted: the type of façade windows and the ratio of façade windows. The classified contents for each detail are as follows:

- Type of façade window: front curtain wall, repeated single windows, others (mixed windows, etc.).
- Façade window ratio: ~25%, ~50%, ~75%, and ~100%.

Among the types of façade windows, the front curtain wall describes instances when the window covers the entire building, while single window repetition indicates when the window is repeated every floor in a consistent pattern. Finally, the “other” category includes mixed windows and half-height horizontal windows (see Figure 8).

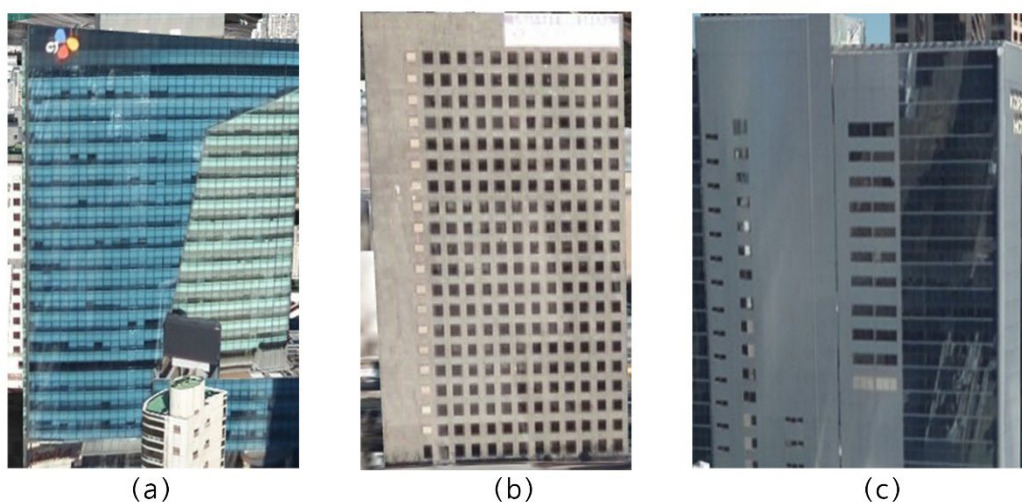


Figure 8. 3D buildings detected on the building map: (a) Front curtain wall building, (b) repeated single window building, (c) “other” (mixed windows).

The distinction between the front curtain wall and the repeated single windows can be acquired by using the number of window areas recognized in the façade area of the building. Buildings in the “other” category refer to the form to which both factors can be applied (mixed windows). When the exterior information of buildings is extracted based on a deep learning model, the distinction of buildings and windows is manually defined in advance. Provided the size of the area between the building façade area (B_i) and the window area ($w_x \in W_i$) within the building façade area is similar or equal, the building is classified as a front curtain wall building. If two or more window areas are detected, the building is classified as a repeated single window building. Buildings that are not classified as either type are classified as “other”. The classification function, D , to detect the type of façade windows is as follows:

$$D = \begin{cases} \text{front curtain wall,} & |B_i| - |w_x| \leq \text{threshold}; i = 1, 2, \dots, n; x = 1, 2, \dots, |W_i|, w_x \in W_i \\ \text{repeated single windows,} & x \geq 2 \\ \text{others,} & \text{otherwise} \end{cases}$$

Next, to classify the façade window ratio, the ratio of the size of the window area to the size of the façade area of the building is employed. The window ratio is classified into four categories, namely, ~25%, ~50%, ~75%, and ~100%, and the window ratio (R) can be acquired using the following equation:

$$R = 100 * \frac{\sum_{x=1}^{|W_i|} |w_x|}{|B_i|}; w_x \in W_i; i = 1, 2, \dots, n$$

Next, we employ a simple classification method for the walls using RGB channel values to distinguish colors in the image so the color of the wall is easily detected. First, RGB channel values in the façade area of the building were extracted, excluding the window areas, and colors were classified as black, maroon, silver, orange, green, and blue. The following RGB channel values for each color are defined as reference channel values:

- Black: $K(r, g, b) = (0, 0, 0)$.
- Maroon: $R(r, g, b) = (90, 0, 0)$.
- Silver: $S(r, g, b) = (192, 192, 192)$.
- Orange: $Y(r, g, b) = (255, 127, 0)$.
- Green: $G(r, g, b) = (0, 255, 0)$.
- Blue: $B(r, g, b) = (0, 0, 255)$.

The average of the RGB channel values for the parts, excluding the window area among the façade areas of buildings, was compared to the RGB channel values for each color as presented above. The color of the façade of buildings is determined by comparing the average RGB channel value and the reference RGB channel value in the corresponding area. The function for comparing RGB channel values is as follows [60]:

$$\operatorname{argmin}_{i; U_i \in U} l(U, A); U = \{K, R, S, Y, G, B\}$$

$$l(U, A) = \sqrt{((512 + rm) * r_i * r_i \gg 8) + 4 * g_i * g_i + (767 - rm) * b_i * b_i \gg 8}; rm = (r_i + \bar{r})/2$$

where $r_i(\bar{r}), g_i(\bar{g}), b_i(\bar{b})$ are the symbols of pixel value for red, green, and blue, respectively. The average RGB channel value of the corresponding building façade is $A(\bar{r}, \bar{g}, \bar{b}) = (\frac{\sum_{i=1}^n r_i}{n}, \frac{\sum_{i=1}^n g_i}{n}, \frac{\sum_{i=1}^n b_i}{n})$, where n refers to the number of pixels in the detection region. Figure 9 illustrates the RGB channel and the assigned color of each building façade.

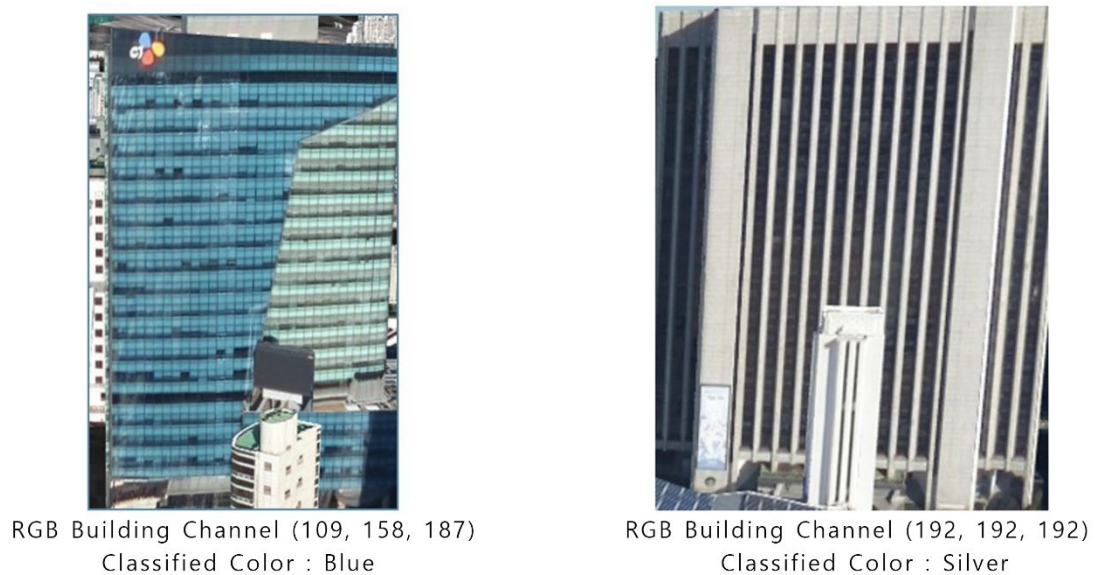


Figure 9. Example of allocating façade colors based on RGB building channels.

Next, we utilize a detailed method to classify the walls by classifying the materials themselves (glass, cement, etc.) on the façade of the building. To recognize and classify these materials, they must be manually relabeled and classified by viewing the detected façade of the building in the image; we did not conduct a detailed classification of the walls separately. Nonetheless, given the results acquired after detecting the façade and windows, there is sufficient reason to consider that future classifications can be performed separately. In addition, materials that are not clearly distinguished with the naked eye (e.g., cement, bricks, and tiles) could be considered together, using the façade color classification method for buildings that is used in the simple classification method.

4. Experiment and Results

4.1. Experiment Design

This section briefly describes the experimental performance plan for identifying the façade of buildings to which deep learning technology was applied, as described in Section 3, and verifying the performance of the façade information extraction methodology. First, the experimental process and results for verifying the performance of the model are described. Next, a process for extracting building façade information and an experimental result is outlined, using an image processing technique from the extracted image.

4.2. Results of Building Façade Recognition Using the Deep Learning Model

This section explains the results of building façade recognition using the deep learning model. We used the faster R-CNN model in the Detectron2 environment to recognize the façade of the building from the façade image. The total number of acquired images was 554, and the approximate building façade and windows numbered, respectively, 800 and 900. Among the acquired images, approximately 388 (70%) images were used for learning, and approximately 166 (30%) images were used for the test. The building façade image used here was that of a high-rise building in Seoul. Regarding the hyperparameters for learning, the learning rate was set to 0.0005 and the maximum number of iterations was set to 5000 times. Additionally, by using the early stopping function of learning, the learning time was reduced since it halted when no further reduction in loss occurred on subsequent iterations; in our study, iteration occurred approximately 1200 times before stopping. The graph in Figure 10 illustrates this point. The loss function utilized the root mean squared error (RMSE), and it is crucial to note that the model learned from the decreasing trend of loss values on each iteration, as opposed to absolute values.

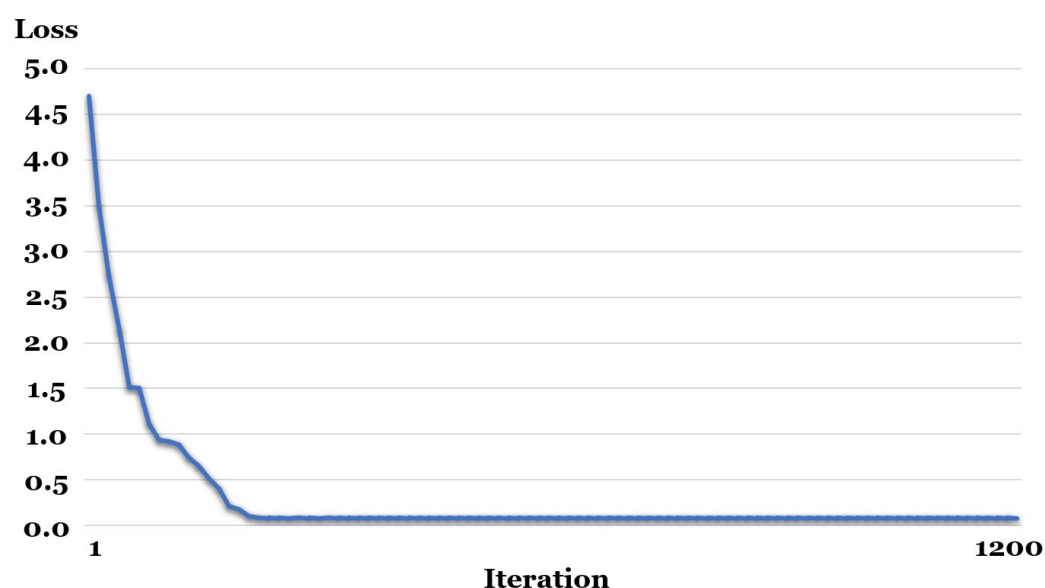


Figure 10. Graph showing losses by iteration.

The result of the model's learning indicated that the total loss value was 0.03624. As an indicator for the evaluation of the learned model, accuracy was used. The method to derive it is as follows:

$$Accuracy = \frac{\text{Number of detected objects}}{\text{Number of total objects}}$$

The building façade detection accuracy and window detection accuracy were 93% and 91%, respectively.

4.3. Result of Extracting Building Façade Information Based on Image Processing Techniques

This section describes the results of extracting the building façade information using image processing techniques from the recognized building façade images while using a deep learning model. We divided building façade information into (1) the simple classification method for windows, (2) the detailed classification method for windows, and (3) simple classification for walls. The first part (simple classification method for windows) was already conducted in the previous section, so we handle the remaining parts in this section.

The detailed classification of windows focused on the types of façade windows (front curtain walls, repeated single windows, other) and the classification of façade window ratio (~25%, ~50%, ~75%, and ~100%). To verify the suggested method, 43 front curtain wall façade images, 31 repeated single window façade images, and 38 other images were used as test data. As a result, the accuracy in recognizing front curtain wall, repeated single windows, and other building types was 95% (41 images), 90% (28 images), and 100% (38 images), respectively.

Next, the simple classification for walls focused on the RGB channel value of the walls. After we removed the window areas from the façade areas prior to extracting the channel values, the average of the channel values of the remaining area was calculated. "Matching" was performed based on the channels, which were calculated according to six previously defined channels. As a result, all 30 images for each color were well matched. This shows that the color of the building can be distinguished when RGB channel values are used. This methodology can be used in preprocessing for the detailed classification method for walls in future studies.

5. Conclusions

This study aimed to detect the façade of a building using deep learning and image processing techniques to recognize and classify façade information such as wall colors and the window ratio of the building. The ways in which the study contributes to the field are threefold. First, a model was built to learn the façade and window images of the building using a 3D building façade map, which was never previously attempted. In this study, data were acquired by automating the 3D building façade map extraction process, whereas previous studies were conducted with actual images in the form of street views. Unlike street views that were photographed from limited angles, this model has the advantage of learning more sophisticated building façade images from several angles. Second, the current study acquired better detection accuracy using the latest object detection model. In previous studies, basic image processing techniques or deep learning models that are not optimized for object detection were used, whereas, in this study, faster R-CNN models optimized for object detection and segmentation within images were employed. Third, a methodology for recognizing and extracting information regarding the façade of buildings using image processing techniques was proposed in the current study. The proposed methodologies could be adopted in future studies to acquire fundamental building façade information automatically. Moreover, they can be used in preprocessing to recognize more detailed information regarding building façade.

As a result of detecting the exterior buildings and extracting information, the techniques were 93% and 91% accurate in correctly detecting the façade and windows, respectively, while showing excellent performance in experiments in which façade information was extracted. However, in the process of detecting façade and windows using deep learning techniques, the false detection rate during learning was higher than that of building façade detection due to unclear window boundaries. In fact, data noise is likely to be variable due to experimenter errors when generating learning data using VIA software. If window detection can be improved, then learning results will be more satisfactory.

In addition, the reliability of the accuracy may be reduced due to the small size of the experimental dataset when extracting building façade information based on processing techniques. Despite this, verifying performance with minimal data is possible since it is based on rules rather than the use of a learning model or statistical methodology. Moreover, this can be used in preprocessing for determining specific information such as the façade material of a building in the future.

This methodology can be used to collect exterior building information while saving time, cost, and effort on a national level. It may then be added to public building databases (e.g., those containing utility ledgers) to further increase value. In particular, exterior building information such as the beauty of urban architecture, regional characteristics, city images, or landscape can be constructed as a BIM dataset, and a “digital twin” can be implemented to build real-time information with the development of smart architecture and urban technology.

Author Contributions: Conceptualization and methodology, D.S. and N.B.; software and formal analysis and data curation, B.N.; writing—original draft preparation, D.S.; writing—review and editing, N.B.; visualization and supervision, D.S.; funding acquisition, D.S. and B.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the research grant of the Chungbuk National University in 2020. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2021R1F1A1061804).

Acknowledgments: The authors wish to express their gratitude to the four anonymous reviewers for the comments and suggestions provided, which have been very helpful in improving the paper.

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

References

1. Korean Law Information Center. Act on Promotion of the Provision and Use of Public Data. Available online: <https://www.law.go.kr/> (accessed on 15 December 2021).
2. Public Data Portal. List of Data. Available online: <https://www.data.go.kr/tcs/dss/selectDataSetList.do> (accessed on 15 December 2021).
3. Seumteo, Architectural Administrative Information System. Available online: <https://cloud.eais.go.kr/> (accessed on 25 January 2022).
4. Architecture Data Private Open System. Available online: <https://open.eais.go.kr/main/main.do> (accessed on 25 January 2022).
5. Policy Briefing of S. Korean Government (28 January 2015). Available online: <https://www.korea.kr/news/pressReleaseView.do?newsId=156034648> (accessed on 25 January 2022).
6. Korea Land Daily, (11 June 2019), Ministry of Land, Infrastructure and Transport, Operation of Building Information System Innovation T/F... Presenting the Future of the Building Industry. Available online: <http://www.ikld.kr/news/articleView.html?idxno=205119> (accessed on 25 January 2022).
7. BLCM. Building Life Cycle Management System. Available online: <https://blcm.go.kr/> (accessed on 25 January 2022).
8. Corbusier, L. *Toward a New Architecture*; Preager: New York, NY, USA, 1959.
9. Jenks, C. *Visual Culture*; Routledge: New York, NY, USA, 1995.
10. Digital New Deal Policy of S. Korean Government. Available online: <https://digital.go.kr/> (accessed on 25 January 2022).
11. Ministry of Land, Infrastructure and Transport (June 2021). BIM-Based Construction Industry Digital Transformation Roadmap. Available online: <https://arxiv.org/abs/1910.06391> (accessed on 20 January 2022).
12. Estman, C.; Estman, C.; Teicholz, P.; Sacks, R.; Liston, K. *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
13. BS EN ISO 19650; Organization and Digitisation of Information about Buildings and Civil Engineering Works, Including Building Information Modelling—Information Management Using Building Information Modelling. British Standards Institution: London, UK, 2019.
14. AUTODESK BIM 360, Top 10 Benefits of BIM in Construction. Available online: <https://bim360resources.autodesk.com/connect-construct/top-10-benefits-of-bim-in-construction> (accessed on 25 January 2022).
15. Al-Ashmori, Y.; Othman, I.; Rahmawati, Y.; Amran, Y.; Sabah, S.; Rafindadi, A.; Mikić, M. BIM benefits and its influence on the BIM implementation in Malaysia. *Ain Shams Eng. J.* **2020**, *11*, 1013–1019. Available online: <https://www.sciencedirect.com/science/article/pii/S2090447920300344> (accessed on 20 January 2022).
16. Deng, M.; Menassa, C.; Kamat, V. From BIM to digital twins: A systematic review of the evolution of intelligent building representations in the AEC-FM industry. *J. Inf. Technol. Constr.* **2021**, *26*, 58–83. Available online: <http://www.itcon.org/2021/5> (accessed on 20 January 2022).
17. No, W.; Lee, D. A Deep Learning-Based Braille Blocks Detection System from Street View Images for the Visually Impaired. In Proceedings of the 31st KKHTCNN Symposium on Civil Engineering, Kyoto, Japan, 20 November 2018.
18. Seiferling, I.; Naik, N.; Ratti, C.; Proulx, R. Green streets-Quantifying and mapping urban trees with street-level imagery and computer vision. *Landsc. Urban Plan.* **2017**, *165*, 93–101.
19. Yu, Q.; Wang, C.; Cetiner, B.; Yu, S.; McKenna, F.; Taciroglu, E.; Law, K. Building Information Modeling and Classification by Visual Learning At A City Scale. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 18 October 2019.
20. Santarsiero, G.; Masi, A.; Picciano, V.; Digrisolo, A. The Italian Guidelines on Risk Classification and Management of Bridges: Applications and Remarks on Large Scale Risk Assessments. *Infrastructures* **2021**, *6*, 111. <https://doi.org/10.3390/infrastructures6080111>.
21. Vworld Map. Available online: https://vworld.kr/v4po_main.do (accessed on 15 December 2021).
22. Seoul City 3D Map. S-MAP. Available online: <https://smap.seoul.go.kr/> (accessed on 15 December 2021).
23. Google Earth, 3D Map. Available online: <https://earth.google.com/web/> (accessed on 15 December 2021).
24. Daum Kakao Map. Available online: <http://map.daum.net> (accessed on 15 January 2022).
25. Naver Map. Available online: <http://map.naver.com> (accessed on 25 January 2022).
26. Noh, B.; No, W.; Lee, J.; Lee, D. Vision-based potential pedestrian risk analysis on unsignalized crosswalk using data mining techniques. *Appl. Sci.* **2020**, *10*, 1057.
27. Huang, T. Computer vision: Evolution and promise. 1996 CERN School of Computing, Egmond aan Zee, The Netherlands, 8–21 September 1996.
28. Voulodimos, A.; Doulamis, N.; Doulamis, N.; Protipapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. <https://doi.org/10.1155/2018/7068349>.
29. Sabe, N.; Cohen, I.; Garg, A. Huang. *Machine Learning in Computer Vision*; Springer Science & Business Media: Berlin, Germany, 2005.
30. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410.
31. Lunit Insight. Available online: <https://insight.lunit.io/> (accessed on 15 December 2021).

32. Huang, W.; Song, G.; Hong, H.; Xie, K. Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 2191–2201.
33. Huang, H.; Tang, Q.; Liu, Z. Adaptive Correction Forecasting Approach for Urban Traffic Flow Based on Fuzzy -Mean Clustering and Advanced Neural Network. *J. Appl. Math.* **2013**, *2013*, 7.
34. Koo, W.; Yokota, T.; Takizawa, A.; Katoh, N. Image Recognition Method on Architectural Components from Architectural Photographs Glass Openings Recognition Based on Bayes Classification. *Archit. Inst. Jpn.* **2006**, *4*, 123–128. (In Japanese)
35. Seo, D. Study on the Method for Visual Perception of Architectural Form through Digital Image Processing. Ph.D. Thesis, Yonsei University, Seoul, Korea, 2013. (In Korean)
36. Talebi, M.; Vafaei, A.; Monadjemi, A. Vision-based entrance detection in outdoor scenes. *Multimed. Tools Appl.* **2018**, *77*, 26219–26238. <https://doi.org/10.1007/s11042-018-5846-3>.
37. Armagan, A.; Hirzer, M.; Roth, P.M.; Lepetit, V. Accurate Camera Registration in Urban Environments Using High-Level Feature Matching. In Proceedings of the British Machine Vision Conference (BMVC), London, UK, 2 September 2017.
38. Seong, H.; Choi, H.; Son, H.; Kim, C. Image-based 3D Building Reconstruction Using A-KAZE Feature Extraction Algorithm. In Proceedings of the 35th International Symposium on Automation and Robotics in Construction, Berlin, Germany, 20–25 July 2018.
39. Yuan, L.; Guo, J.; Wang, Q. Automatic classification of common building materials from 3D terrestrial laser scan data. *Autom. Constr.* **2020**, *110*, 103017.
40. Rocio, V.; Mosavi, A.; Ruiz, R. Deep Learning: A Review. Available online: <https://eprints.qut.edu.au/127354/> (accessed on 20 January 2022).
41. Li, D.; Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Processing* **2013**, *7*, 197–387.
42. Khan, T.; Sherazi, H.H.R.; Ali, M.; Letchmunan, S.; Butt, U.M. Deep Learning-Based Growth Prediction System: A Use Case of China Agriculture. *Agronomy* **2021**, *11*, 1551.
43. YOLO. You Only Look Once: Unified, Real-Time Object Detection. Available online: <https://www.arxiv-vanity.com/papers/1506.02640/> (accessed on 15 December 2021).
44. AUTOIT. Available online: <https://www.autoitscript.com/site/> (accessed on 15 December 2021).
45. VGG Image Annotator (VIA). Available online: <https://www.robots.ox.ac.uk/~vgg/software/via/via.html/> (accessed on 15 December 2021).
46. Szegedy, C.; Toshev, A.; Erhan, D. Deep Neural Networks for Object Detection. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 1–9.
47. Purkait, P.; Cheng, Z.; Christopher, Z. SPP-Net: Deep Absolute Pose Regression with Synthetic Views. In Proceedings of the British Machine Vision Conference (BMVC 2018), London, UK, 9 December 2017.
48. Wang, L.; Guo, S.; Huang, W.; Qiao, Y. Places205-VGGNet Models for Scene Recognition. *arXiv* **2015**, arXiv: Abs/1508.01667.
49. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 22 October 2014.
50. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 27 September 2015.
51. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. *arXiv* **2016**, arXiv:1506.01497.
52. Tian, Y.; Yang, G.; Wang, Z.; Wang, L.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426.
53. Shafiee, M.; Chywl, B.; Li, F.; Wong, A. Fast YOLO: A fast you only look once system for real-time embedded object detection in video. *arXiv* **2017**, arXiv:1709.05943.
54. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
55. Bochkovskiy, A.; Wang, C.; Liao, H. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
56. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask r-cnn. *arXiv* **2018**, arXiv:1703.06870.
57. COCO. Common Objects in Context. Available online: <https://cocodataset.org/#home> (accessed on 25 January 2022).
58. Pham, V.; Pham, C.; Dang, T. Road Damage Detection and Classification with Detectron2 and Faster R-CNN. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 5592–5601. <https://doi.org/10.1109/BigData50022.2020.9378027>.
59. Detectron2. Available online: <https://github.com/facebookresearch/detectron2> (accessed on 15 December 2021).
60. Chernov, V.; Alander, J.; Bochkov, V. Integer-based accurate conversion between RGB and HSV color spaces. *Comput. Electr. Eng.* **2015**, *46*, 328–337.