Terrestrial, Atmospheric
and Oceanic Sciences

ORIGINAL ARTICLE Open Access

The method and implementation of a Taiwan building recognition model based on YOLOX-S and illustration enhancement

Yung-Yu Zhuang[1], Wei-Hsiang Chen[1], Shao-Kai Wu[1] and Wen-Yao Chang[1*]

Abstract

Earthquakes pose significant risks in Taiwan, necessitating effective risk assessment and preventive measures to reduce damage. Obtaining complete building structure data is crucial for the accurate evaluation of earthquake-induced losses. However, manual annotation of building structures is time-consuming and inefficient, resulting in incomplete data. To address this, we propose YOLOX-CS, an object detection model, combined with the Convolutional Block Attention Module (CBAM), to enhance recognition capabilities for small structures and reduce background interference. Additionally, we introduce the Illustration Enhancement data augmentation method to improve the recognition of obscured buildings. We collected diverse building images and manually annotated them, resulting in a dataset for training the model. YOLOX-CS with CBAM significantly improves recognition accuracy, particularly for small objects, and Illustration Enhancement enhances the recognition of occluded buildings. Our proposed approach advances building structure recognition, contributing to more effective earthquake risk assessment systems in Taiwan and beyond.

Keywords Earthquake risk assessment, Building structure recognition, YOLOX, Object detection, Illustration enhancement

1 Introduction

In Taiwan, earthquakes are a common occurrence, including approximately three events each year surpassing a magnitude of 6.0. These significant seismic events have a profound impact on both the economy and public safety. A key aspect of seismic engineering is to evaluate earthquake risks and implement preventive strategies. This involves analyzing building characteristics such as structure, age, and height to determine their resilience to earthquakes. Consequently, compiling comprehensive building data is vital for effective earthquake risk assessment.

However, acquiring structural details of buildings poses a greater challenge compared to other types of building information. Structural components include the foundation, walls, columns, beams, and trusses, all bearing the load of the building. Manually annotating these details is an expert-intensive and time-consuming task. While governments often provide this data, the level of detail and completeness can vary, especially in developing countries, and may not always keep pace with new construction. Therefore, creating a system for swiftly identifying building structures is critical.

Remote sensing offers a rapid and broad approach to gather building data, but it falls short in identifying specific structural information. Acevedo et al. (2017) have combined satellite imagery with Google Street View to manually identify building heights and types, and to analyze roof shapes for wider data collection. Remote sensing is adept at capturing regional attributes such as the

*Correspondence:
Wen-Yao Chang
wed12345976@g.ncu.edu.tw
[1] Department of Computer Science and Information Engineering, National Central University, Taoyuan City, Taiwan

density of buildings but still depends on manual processes for identifying specific structural details of individual buildings (Huo 2019)

With the rapid evolution of computer technology, machine learning has become increasingly prevalent in image recognition, extending to the identification of building structures. Techniques like SVM, used by researchers such as Pittore and Wieland (2012) in conjunction with satellite images, enable detailed recognition at the pixel level. However, to further refine and automate the identification of building structures, additional imagery detailing the exterior of buildings is necessary. Traditional machine learning approaches, when applied directly to architectural images, face numerous challenges, including varying angles and lighting conditions (Chen et al. 2017; Zhang 2018; Bilal and Hanif 2019).

Convolutional Neural Networks (CNNs) have become a cornerstone in building recognition, leveraging their capability to extract complex, high-dimensional features (Shi 2021). With the advancement of deep learning, CNNs have found applications in areas like object detection, semantic segmentation (Wei 2016), and image classification (Ezat et al. 2020). Since 2017, they have been integrated into building recognition models. For example, Yibo Liu and his team developed a deep learning-based framework for hierarchical building detection using CNNs to identify buildings from remote sensing data (Liu 2018). Similarly, Kang and colleagues employed CNNs with OpenStreetMap data to identify eight building types in North America (Kang 2018). However, CNNs encounter unique challenges in building recognition. They often overemphasize background features in building images, where buildings constitute just about 10% of the image, diminishing the model's accuracy. The architectural diversity and density in certain regions further impact CNNs' efficacy.

CNN-based object detection has proven effective in identifying building structures, efficiently recognizing multiple buildings within a single image and swiftly predicting a range of building structures. This approach is especially beneficial for images containing numerous buildings. However, challenges persist, such as variations in foreground-background ratios, difficulties in detecting small targets, and occlusions.

In response to the incomplete building structure data in Taiwan, we propose using Google Street View and object detection technologies for rapid, automated structure recognition. The high density and variety of buildings in Taiwanese cities, often leading to images with multiple, variably sized, and occluded buildings, pose a significant challenge. To overcome these hurdles, we introduce the YOLOX-CS model. This model incorporates Convolutional Block Attention Module (CBAM) (Woo 2018) convolutional blocks to better detect smaller structures and uses Illustration enhancement for data augmentation, improving the recognition of obstructed buildings.

2 Datasets

Google Street View offers comprehensive street-level imagery across the globe, complemented by an API Google (2021) that allows developers to craft custom applications. This capability is pivotal for us to employ object detection models on Street View images, aiming to swiftly compile a structural map of buildings in Taiwan to facilitate earthquake risk assessment.

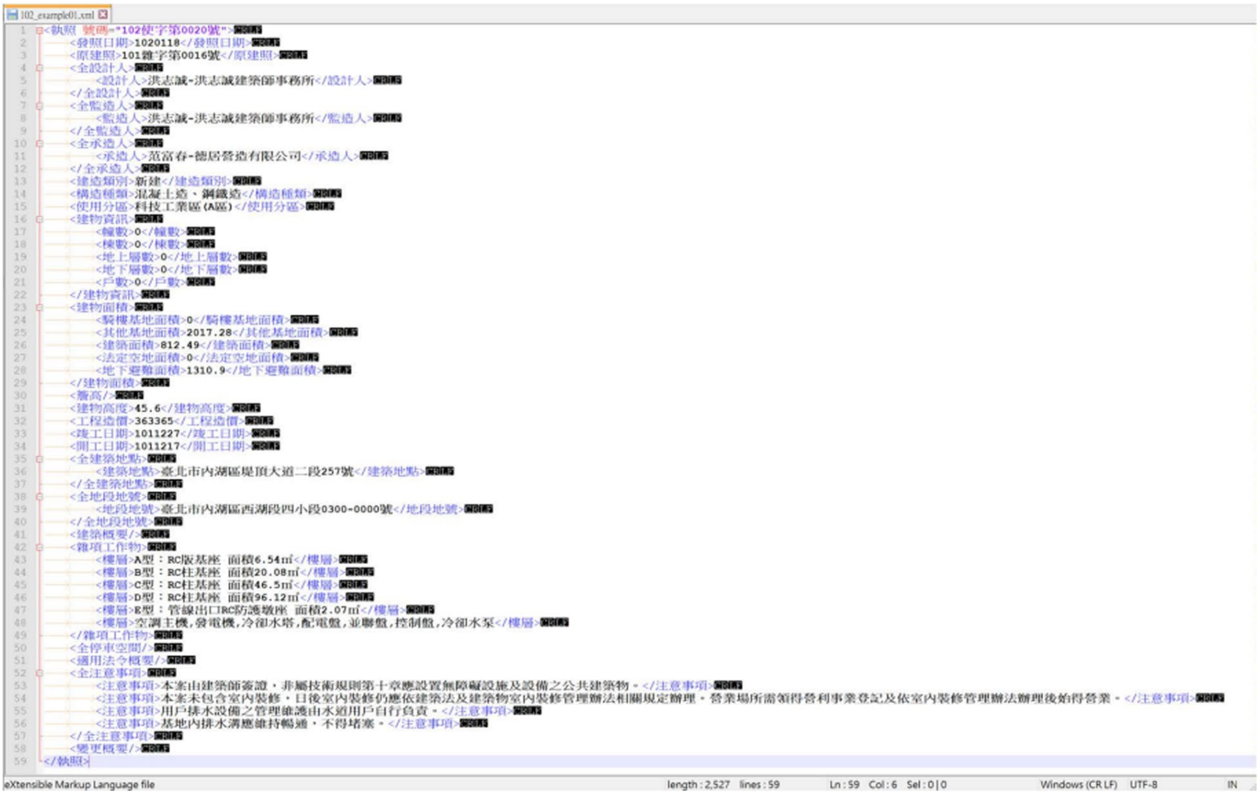
To enhance the accuracy of our model in discerning various building structures, it's crucial to amass a diverse collection of building images for training purposes. Utilizing the Google Street View API, we acquire these images based on specific latitude and longitude coordinates. The subsequent phase involves meticulous manual filtering and annotation to categorize the structural types of the buildings. The coordinates and structural data for this endeavor are sourced from official records like "Taipei City Historical Usage License Summary" (Taipei City Government Open Data Platform 2020) and "Taichung City Buildings_WGS84" (Taichung City Government Open Data Platform 2019) data, provided by the government.

2.1 Taipei City Historical Usage License Summary

The "The Taipei City Historical Usage License Summary," curated by the Taipei City Construction Management Office, encompasses building data of Taipei City spanning from 1949 to 2019. This dataset, available in XML format, includes 24 fields and undergoes annual updates (Fig. 1). However, we had to discard some older records that no longer align with Taiwan's current address system, ultimately retaining 60,387 valid data entries.

In these valid data sets, two critical fields were extracted: the structure of the building and its address. The structural data assists in further annotation tasks, while the address information is converted into geographical coordinates using the geolocation services provided by Taiwan Geospatial One Stop. During this conversion process, we encountered several challenges: 7520 entries couldn't be precisely geolocated (due to multiple possible coordinates), 6347 entries lacked a definitive location, and 17,350 entries faced issues with duplicated locations. After filtering out these discrepancies, 29,170 entries remained viable for use.

A closer analysis of these remaining entries revealed a highly skewed distribution of building types. Reinforced concrete structures dominate, comprising 82.4% of the dataset, followed by strengthened brick constructions at 11.9%, with other types constituting less than 2%. This



```xml
<執照 號碼="102使字第0020號">
    <發照日期>1020118</發照日期>
    <原建照>101雜字第0016號</原建照>
    <全設計人>
        <設計人>洪志誠-洪志誠建築師事務所</設計人>
    </全設計人>
    <全監造人>
        <監造人>洪志誠-洪志誠建築師事務所</監造人>
    </全監造人>
    <全承造人>
        <承造人>范富春-德居營造有限公司</承造人>
    </全承造人>
    <建造類別>新建</建造類別>
    <構造種類>混凝土造、鋼鐵造</構造種類>
    <使用分區>科技工業區(A區) </使用分區>
    <建物資訊>
        <幢數>0</幢數>
        <棟數>0</棟數>
        <地上層數>0</地上層數>
        <地下層數>0</地下層數>
        <戶數>0</戶數>
    </建物資訊>
    <建物面積>
        <騎樓基地面積>0</騎樓基地面積>
        <其他基地面積>2017.28</其他基地面積>
        <建築面積>812.49</建築面積>
        <法定空地面積>0</法定空地面積>
        <地下避難面積>1310.9</地下避難面積>
    </建物面積>
    <簷高/>
    <建物高度>45.6</建物高度>
    <工程造價>363365</工程造價>
    <竣工日期>1011227</竣工日期>
    <開工日期>1011217</開工日期>
    <全建築地點>
        <建築地點>臺北市內湖區堤頂大道二段257號</建築地點>
    </全建築地點>
    <全地段地號>
        <地段地號>臺北市內湖區西湖段四小段0300-0000號</地段地號>
    </全地段地號>
    <建築概要/>
    <雜項工作物>
        <樓層>A型：RC版基座 面積6.54㎡</樓層>
        <樓層>B型：RC柱基座 面積20.08㎡</樓層>
        <樓層>C型：RC柱基座 面積46.5㎡</樓層>
        <樓層>D型：RC柱基座 面積96.12㎡</樓層>
        <樓層>E型：管線出口RC防護墩座 面積2.07㎡</樓層>
        <樓層>空調主機,發電機,冷卻水塔,配電盤,並聯盤,控制盤,冷卻水泵</樓層>
    </雜項工作物>
    <全停車空間/>
    <適用法令概要/>
    <全注意事項>
        <注意事項>本案由建築師簽證，非屬技術規則第十章應設置無障礙設施及設備之公共建築物。</注意事項>
        <注意事項>本案未包含室內裝修，日後室內裝修仍應依建築法及建築物室內裝修管理辦法相關規定辦理。營業場所需領得營利事業登記及依室內裝修管理辦法辦理後始得營業。</注意事項>
        <注意事項>用戶排水設備之管理維護由水道用戶自行負責。</注意事項>
        <注意事項>基地內排水溝應維持暢通，不得堵塞。</注意事項>
    </全注意事項>
    <變更概要/>
</執照>
```

eXtensible Markup Language file
length : 2,527 lines : 59
Ln : 59 Col : 6 Sel : 0 | 0
Windows (CR LF) UTF-8 IN

Fig. 1 Example data from the Taipei City Historical Usage License Summary

imbalance necessitates the acquisition of additional data to enhance the diversity and balance of the dataset.

2.2 Taichung City Buildings_WGS84

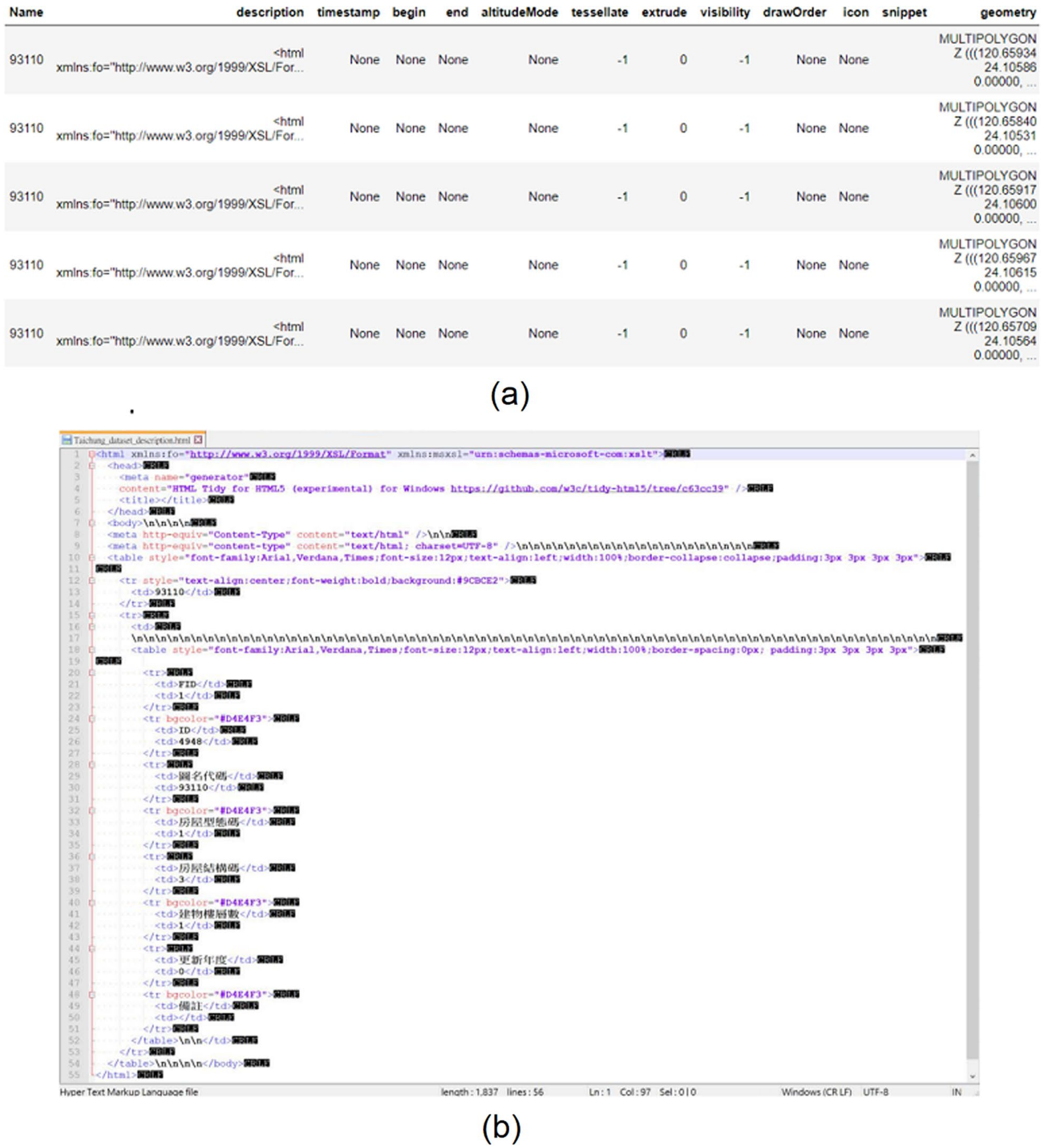
The "Taichung City Buildings_WGS84" dataset, compiled by the Urban Development Bureau of Taichung City Government, offers comprehensive details on the region's buildings and is updated annually. Presented in Shapefile format, each entry in this dataset includes 13 fields (Fig. 2). The description field, formatted in HTML, provides the essential building structure information we require. Additionally, the geometry field in GeoJSON MultiPolygon format delineates the polygonal shapes of the buildings. Given the straightforward nature of building shapes, we opted to use the centroids of these polygons to represent their geographical coordinates. The dataset encompasses a total of 300,183 records.

This dataset showcases a varied array of building structures: reinforced concrete accounts for 64.9%, steel frame constructions for 22.1%, and brick constructions for 10.6%. After filtering out entries without specified building structures, we utilized the geometry and structure data to generate a distribution map of Taichung City's building structures (Fig. 3).

2.3 Image retrieval and annotation

Upon analyzing the two datasets, we identified that four building structures—Reinforced Concrete (RC), Steel Frame Reinforced Concrete (SRC), Brick Building (BB), and Steel Frame (SB)—are prevalent in Taiwan and represent common construction types. We thus targeted these structures for our model training. Representative examples of these structures are illustrated in Fig. 4.

We then utilized latitude and longitude data to acquire Street View images. Considering the costs associated with the Google Street View API, we employed a systematic approach to randomly extract data from these four structural types to minimize expenses and image requests. Our image capture settings included parameters like size = 640 x 640, field of view (fov) = 120, and pitch = 30, which we validated as optimal for building recognition. The photo orientation was determined based on the closest street to the target building, focusing on capturing the building itself. We discarded images that either did not feature the target building or where background structures predominated, resulting in a dataset of 6,394 records. The breakdown of these records is as follows: 1763 for Reinforced Concrete, 295 for Steel Frame Reinforced Concrete, 1392 for Brick Building, and 2944 for Steel Frame.



Name	description	timestamp	begin	end	altitudeMode	tessellate	extrude	visibility	drawOrder	icon	snippet	geometry
93110	<html xmlns:fo="http://www.w3.org/1999/XSL/For...	None	None	None	None	-1	0	-1	None	None		MULTIPOLYGON Z (((120.65934 24.10586 0.00000, ...
93110	<html xmlns:fo="http://www.w3.org/1999/XSL/For...	None	None	None	None	-1	0	-1	None	None		MULTIPOLYGON Z (((120.65840 24.10531 0.00000, ...
93110	<html xmlns:fo="http://www.w3.org/1999/XSL/For...	None	None	None	None	-1	0	-1	None	None		MULTIPOLYGON Z (((120.65917 24.10600 0.00000, ...
93110	<html xmlns:fo="http://www.w3.org/1999/XSL/For...	None	None	None	None	-1	0	-1	None	None		MULTIPOLYGON Z (((120.65967 24.10615 0.00000, ...
93110	<html xmlns:fo="http://www.w3.org/1999/XSL/For...	None	None	None	None	-1	0	-1	None	None		MULTIPOLYGON Z (((120.65709 24.10564 0.00000, ...

(a)

```
<html xmlns:fo="http://www.w3.org/1999/XSL/Format" xmlns:msxsl="urn:schemas-microsoft-com:xslt">
  <head>
    <meta name="generator"
    content="HTML Tidy for HTML5 (experimental) for Windows https://github.com/w3c/tidy-html5/tree/c63cc39" />
    <title></title>
  </head>
  <body>\n\n\n\n
  <meta http-equiv="Content-Type" content="text/html" />\n\n
  <meta http-equiv="content-type" content="text/html; charset=UTF-8" />\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n
  <table style="font-family:Arial,Verdana,Times;font-size:12px;text-align:left;width:100%;border-collapse:collapse;padding:3px 3px 3px 3px">

    <tr style="text-align:center;font-weight:bold;background:#9CBCE2">
      <td>93110</td>
    </tr>
    <tr>
      <td>
      \n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n
      <table style="font-family:Arial,Verdana,Times;font-size:12px;text-align:left;width:100%;border-spacing:0px; padding:3px 3px 3px 3px">

        <tr>
          <td>FID</td>
          <td>1</td>
        </tr>
        <tr bgcolor="#D4E4F3">
          <td>ID</td>
          <td>4948</td>
        </tr>
        <tr>
          <td>圖名代碼</td>
          <td>93110</td>
        </tr>
        <tr bgcolor="#D4E4F3">
          <td>房屋型態碼</td>
          <td>1</td>
        </tr>
        <tr>
          <td>房屋結構碼</td>
          <td>3</td>
        </tr>
        <tr bgcolor="#D4E4F3">
          <td>建物樓層數</td>
          <td>1</td>
        </tr>
        <tr>
          <td>更新年度</td>
          <td>0</td>
        </tr>
        <tr bgcolor="#D4E4F3">
          <td>備註</td>
          <td></td>
        </tr>
      </table>\n\n</td>
    </tr>
  </table>\n\n\n\n</body>
</html>
```

Hyper Text Markup Language file
length : 1,837 lines : 56
Ln : 1 Col : 97 Sel : 0 | 0
Windows (CR LF) UTF-8
IN

(b)

Fig. 2 Example data from Taichung City Buildings_WGS84 (**a**) Original Shapefile format (**b**) Example of data in the description field in HTML format

In the final step, we handpicked images depicting the target buildings and annotated them with their corresponding building structure types, thereby finalizing our dataset for training the model.

3 YOLOX-CS

YOLO (You Only Look Once) (Redmon et al. 2016), a widely popular object detection model in recent years, has been acclaimed for its rapid prediction speed and

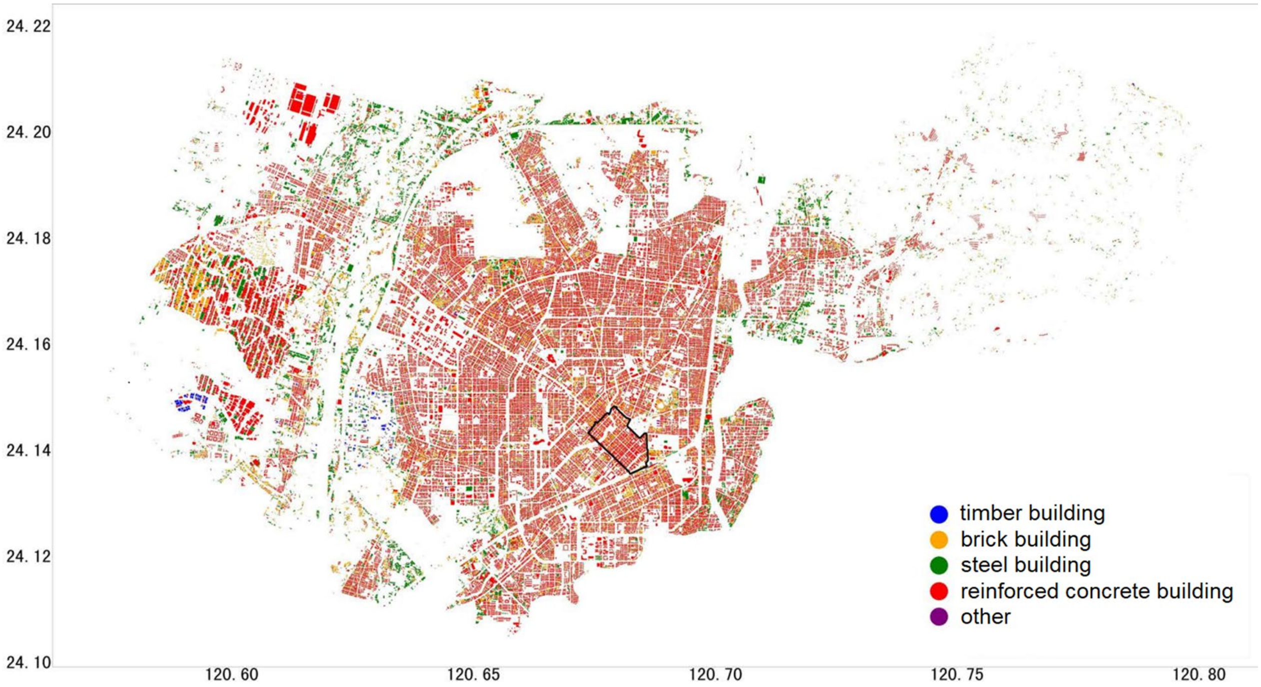


Fig. 3 Taichung City building structure distribution map. The X-axis represents longitude, and the Y-axis represents latitude

high accuracy, making significant strides in various fields. Unique in its approach, YOLO employs a single neural network to tackle regression problems instead of classification, enhancing its training efficiency and ability to detect a diverse range of objects. Notably, YOLO's swift inference capabilities, operating at milliseconds, make it ideal for real-time image recognition, vastly broadening its application scope.

From its inception, YOLO has evolved through multiple iterations. Versions like YOLOv4 and beyond have shown outstanding performance in both speed and accuracy, with YOLOX (Ge 2021) particularly excelling. YOLOX amalgamates elements from YOLOv3 (Redmon and Farhadi 2018), YOLOv4 (Bochkovskiy et al. 2020), YOLOv5 (Jocher et al. 2021), and incorporates techniques such as the Anchor-free framework and Simplified Optimal Transport Assignment (SimOTA), optimizing the model's effectiveness. Catering to different requirements, YOLOX offers a range of models—YOLOX-S, YOLOX-L, YOLOX-M, etc., based on YOLOv5. For our research, we selected YOLOX-S as our principal training model for its excellent balance between speed and accuracy.

However, challenges arise due to the varied angles of street view images and the dense building structures in Taiwan. Target buildings in these images are frequently obscured by elements like trees and streetlights, resulting in incomplete and inconsistent building outlines. Moreover, the presence of multiple buildings in a single image, some occupying smaller portions with less prominent features, poses a risk of being overlooked by the model, thus affecting its accuracy.

To boost the model's proficiency in detecting smaller targets, we introduced YOLOX-CS. This innovation integrates the CBAM module into the YOLOX-S network structure, preserving its performance while enhancing its capability to discern smaller objects. Figure 5 illustrates the network architecture of YOLOX-CS.

3.1 YOLOX-S

YOLOX-S, built upon YOLOv5-S, incorporates a standard network structure with three primary components: Cross Stage Partial Darknet (CSPDarknet), Feature Pyramid Network (FPN), and YOLOHead. The CSPDarknet serves as the primary feature extraction network in YOLOX-S, utilizing residual convolutions to boost accuracy effectively. This network also integrates Cross Stage Partial Network (CSPnet) and Focus structures, along with the Spatial Pyramid Pooling (SPP) network, to broaden its receptive field. The FPN enhances feature extraction by amalgamating effective feature maps from the core network, enriching the representation of multi-scale features. YOLOHead, functioning as YOLOX-S's classifier, employs a Decoupled Head architecture and introduces an Anchor-free approach, improving both convergence speed and accuracy. The

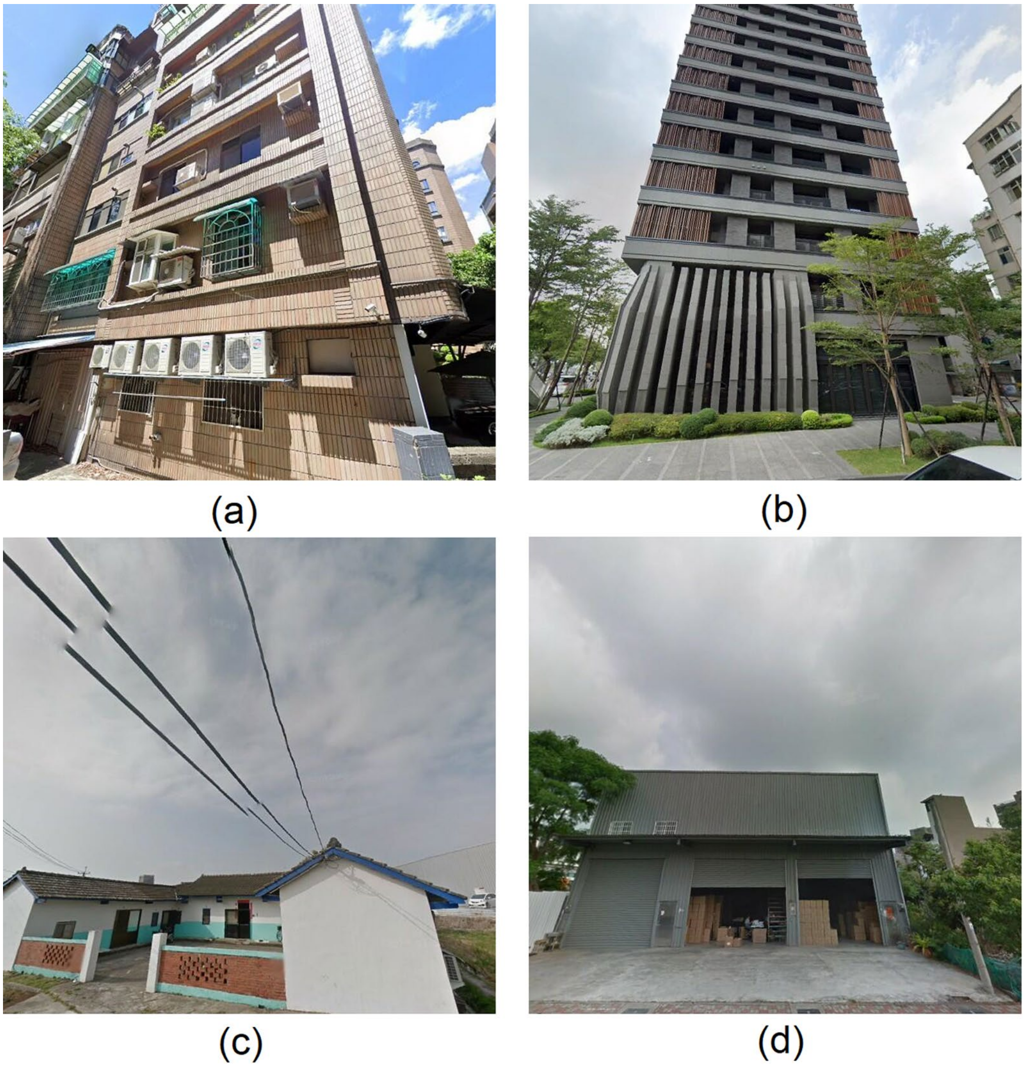


Fig. 4 Representative images of the four building structure types in the dataset: **a** reinforced concrete **b** steel reinforced concrete **c** brick building **d** steel building

SimOTA method is employed to ensure the best possible prediction outcomes.

Compared to YOLOv5-S, YOLOX-S introduces several key enhancements:

1. Activation Function: YOLOX-S adopts the Sigmoid Linear Unit (SiLU) activation function in its backbone and neck, slightly slowing inference speed but providing enhanced non-linear fitting capabilities.
2. Decoupled Head: YOLOX-S's YOLOHead uses a Decoupled structure, segregating the classification (Cls), bounding box localization (Reg), and foreground-background differentiation (Obj) branches for operations prior to concat. This design expedites model convergence and boosts overall performance.
3. Loss Function: YOLOHead has three distinct branches, each with its specific loss function: Binary Cross Entropy (BCE) for Cls and Obj branches and Intersection Over Union (IoU) for the Reg branch. The final Loss function is as follows:

$$L = \frac{L_{\text{cls}} + \text{reg}_{\text{weight}} \cdot L_{\text{reg}} + L_{\text{obj}}}{N_{\text{pos}}} \tag{1}$$

The aggregate loss, balanced by a reg_weight for the reg loss, averages across positive samples.

4. Anchor-free: Post-Decoupled Head, YOLOX-S generates feature vectors, replacing the original feature maps and markedly reducing parameter needs. It also incorporates scale data from the original feature maps via downsampling.

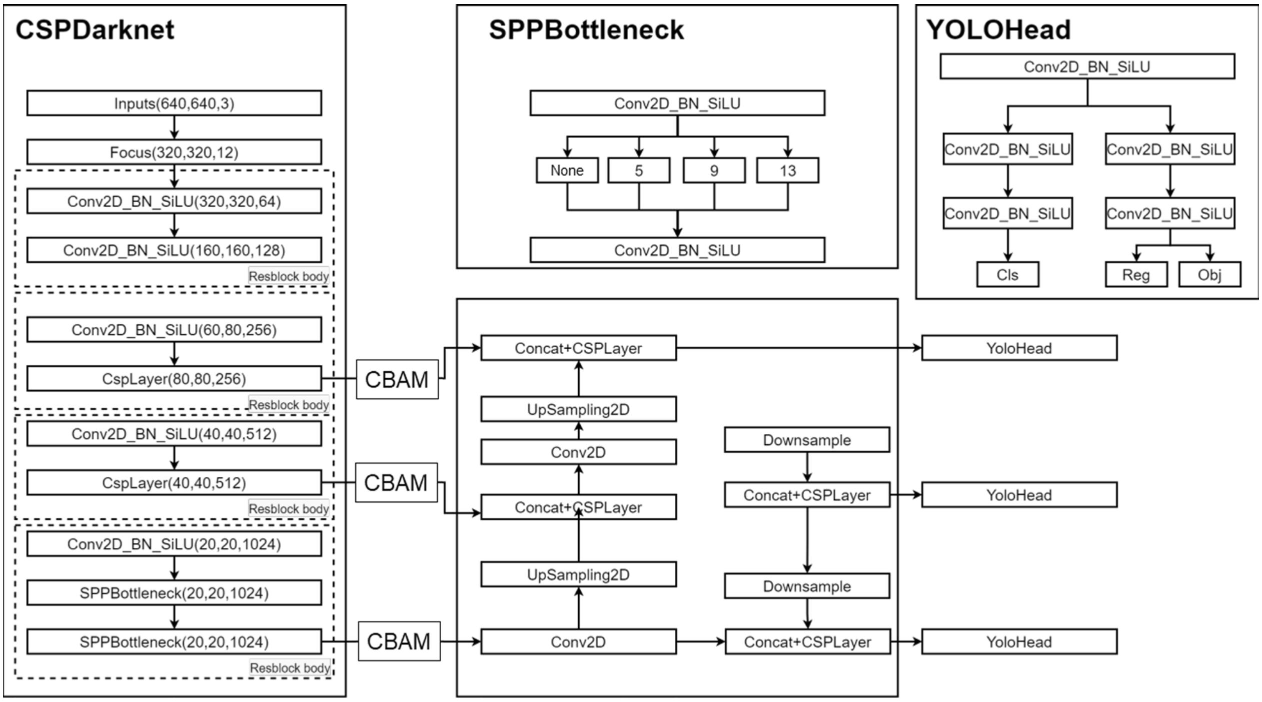


Fig. 5 YOLOX-CS network structure, with the CBAM module embedded in the original YOLOX-S

5. SimOTA: This technique assigns labels to potential positive samples and pinpoints the predicted boxes closest to the label boxes. By transforming label assignment into an optimal transport problem, SimOTA enhances the detection algorithm's inference speed and training efficiency without compromising accuracy, and it does so without requiring extra parameters.

3.2 Convolutional block attention module

CBAM, a versatile and lightweight model, seamlessly integrates into various CNN architectures, effectively enhancing the model's focus on significant regions within the feature maps. Du et al. (2021) Ding and Zhang (2021) In our research, we incorporated CBAM into the YOLOX-S framework to bolster its capacity for detecting smaller architectural targets. Figure 6 depicts the structural layout of the CBAM model.

At its core, CBAM is composed of two distinct modules: the channel attention module and the spatial attention module. The channel attention module is designed to pinpoint and accentuate meaningful features, whereas the spatial attention module focuses on identifying the specific locations of these significant features within the feature map. The computation of feature maps in Fig. 6 can be expressed by formula (2). A feature map $F \in R^{C\times H\times W}$ is given as input, and CBAM infers a 1D

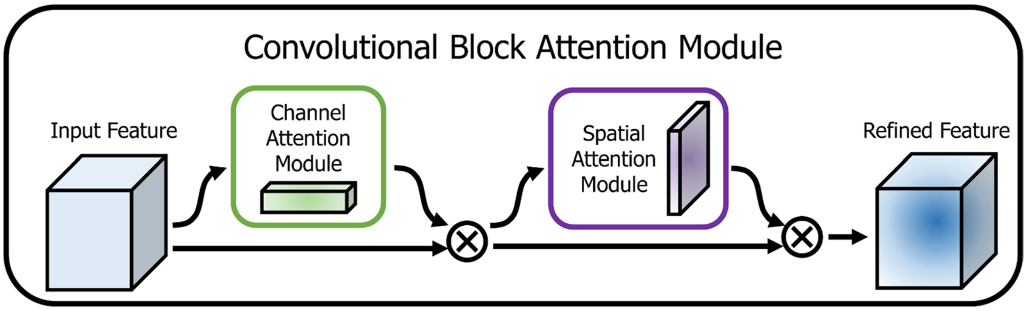


Fig. 6 Convolutional block attention module (CBAM) module structure

channel attention map $M_c \in R^{C\times 1\times 1}$ as well as a 2D spatial attention map $M_s \in R^{1\times H\times W}$ in a sequential manner. Where $\otimes$ represents element-wise multiplication.

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \tag{2}$$

4 Illustration enhancement

To boost the model's proficiency in detecting buildings partially concealed by obstructions, we introduced the "Illustration Enhancement" data augmentation technique. This method introduces obstructions in some training data and uses post-processing to synthesize images of buildings obscured by these obstructions, thus improving the model's ability to identify occluded objects. Our analysis of the YOLOX-S model's recognition abilities revealed its shortcomings in identifying buildings obscured by trees. Consequently, we aim to use the "Illustration Enhancement" technique to improve the model's detection of buildings masked by foliage. We have decided to forgo additional data augmentation methods on images that have already been processed using Illustration Enhancement to accurately assess its effectiveness.

Tree images sourced from textures.com (Texture 2022) were used as synthetic obstructions. These were then digitally composited with building images using Photoshop CS6. To ensure a balanced representation of obscured buildings across the four building structure types, we randomly selected images from each type for this compositing process. Figure 7 illustrates the before-and-after effects of applying Illustration Enhancement.

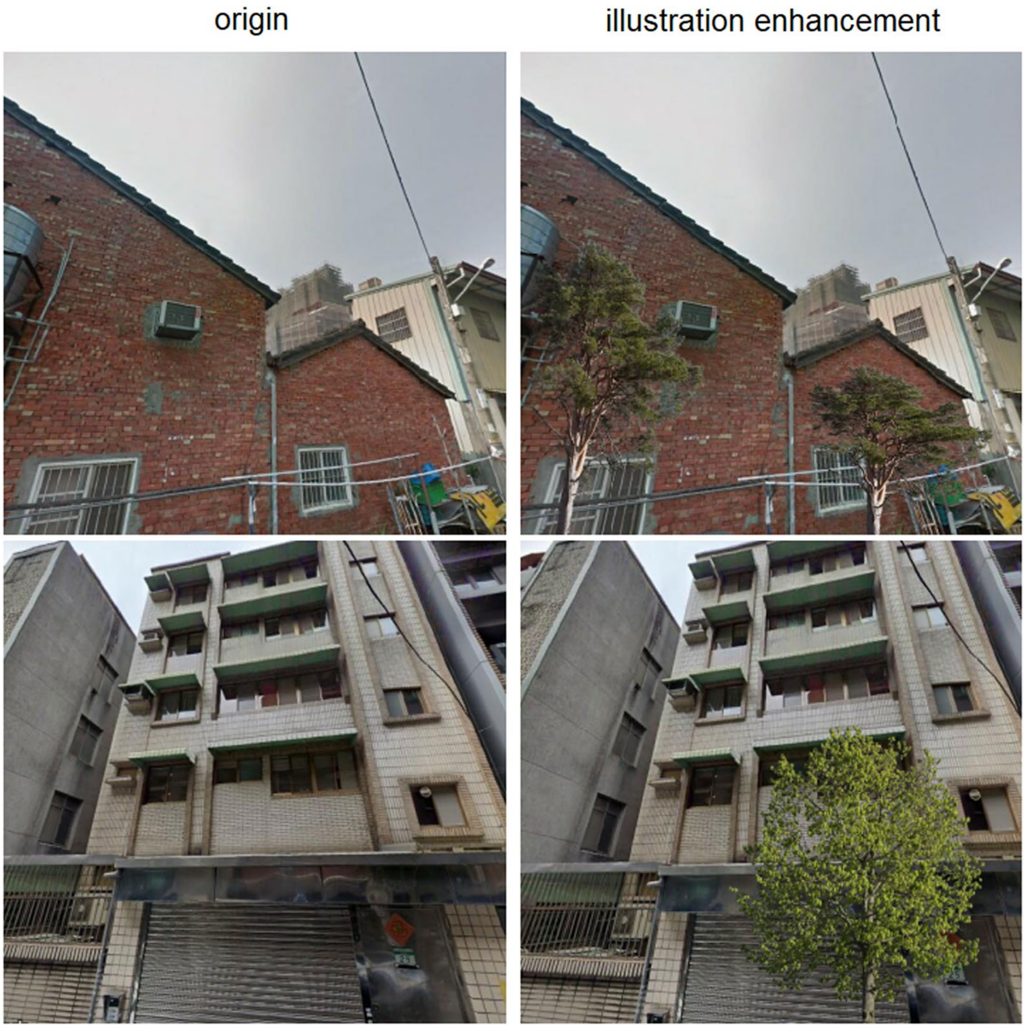


Fig. 7 Example of the illustration enhancement method

Table 1 Detailed training strategies of the models

Model	Batch size	Optimizer	LR	Momentum	Weight decay	Epoch
YOLOX-CS	4	SGD	1×10^{-2}	0.937	5×10^{-4}	100
YOLOX-S	4	SGD	1×10^{-2}	0.937	5×10^{-4}	100
YOLOv4	4	SGD	1×10^{-2}	0.937	5×10^{-4}	100

LR learning rate, *SGD* stochastic gradient decent

Table 2 Training results of YOLOX-CS using each data augmentation

Data augmentation	mAP (%)
Illustration Enhancement	76.32
Horizontal Flip	74.89
Random Noise	73.96
Gaussian Blur	73.45

5 Results

5.1 Training

Initially, we set out to evaluate the impact of our “Illustration Enhancement” data augmentation method. This involved applying different data augmentation techniques, namely Illustration Enhancement, horizontal flipping, random noise, and Gaussian blur, to the original dataset. We then assessed the efficacy of each method in terms of its ability to improve building structure recognition. Following this, our focus shifted to assessing the proposed YOLOX-CS model. This evaluation entailed training the YOLOv4, YOLOX-S, and YOLOX-CS models with both the original dataset (without Illustration Enhancement) and the dataset processed with Illustration Enhancement.

The metric chosen for evaluation was mAP (mean Average Precision), a standard measure in object detection models. This metric calculates the average of the Average Precision (AP) for all classes. The AP is computed by measuring the area under the precision/recall curve across 11 recall intervals, as shown below:

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} P_{interp}(r) \tag{3}$$

All model training sessions were conducted on an NVIDIA GEFORCEÂ® GTX 2080Ti 11GB graphics card. For a detailed overview of our training methodologies, refer to Table 1.

5.2 Comparison

The results from Table 2 reveal the distinct advantage of Illustration Enhancement in building structure identification compared to other common data augmentation techniques. From Table 3, it’s evident that embedding CBAM in YOLOX-CS led to a marginal decrease in precision for the BB and SRC types, yet overall, YOLOX-CS outshined YOLOX-S. This advantage is visually represented in Fig. 8, where YOLOX-CS detected many small-scale architectural targets that eluded YOLOX-S. Significantly, as depicted in Fig. 9, all three models demonstrated improved mAP after being trained on the Illustration Enhancement-enhanced dataset.

6 Discussion

The initial section of this chapter delves into a comprehensive analysis of the experimental outcomes associated with the Illustration Enhancement data augmentation technique and the YOLOX-CS object detection model. It emphasizes the enhancements these methodologies bring to the identification of architectural structures. The latter section addresses the inherent limitations in our study’s automated approach to architectural structure

Table 3 The precision of the three models

Model	BB[1]	RC[1]	SB[1]	SRC[1]	Overall mAP (%)
YOLOv4	75.68	80.53	75.93	100	54.58
YOLOX-S	72.96	81.52	73.75	76.92	73.87
YOLOX-CS	**70.47**	**82.62**	**76.31**	**72.22**	**74.48**
YOLOv4 with Illustration Enhancement	80.13	85.98	77.22	100	59.83
YOLOX-S with Illustration Enhancement	73.10	82.39	76.41	72.34	75.51
YOLOX-CS with Illustration Enhancement	**73.01**	**84.27**	**75.61**	**84.62**	**76.32**

The precision of the models for four building structures, expressed as percentages

To highlight the performance of YOLOX-CS, the model we have proposed, by using bold text for emphasis

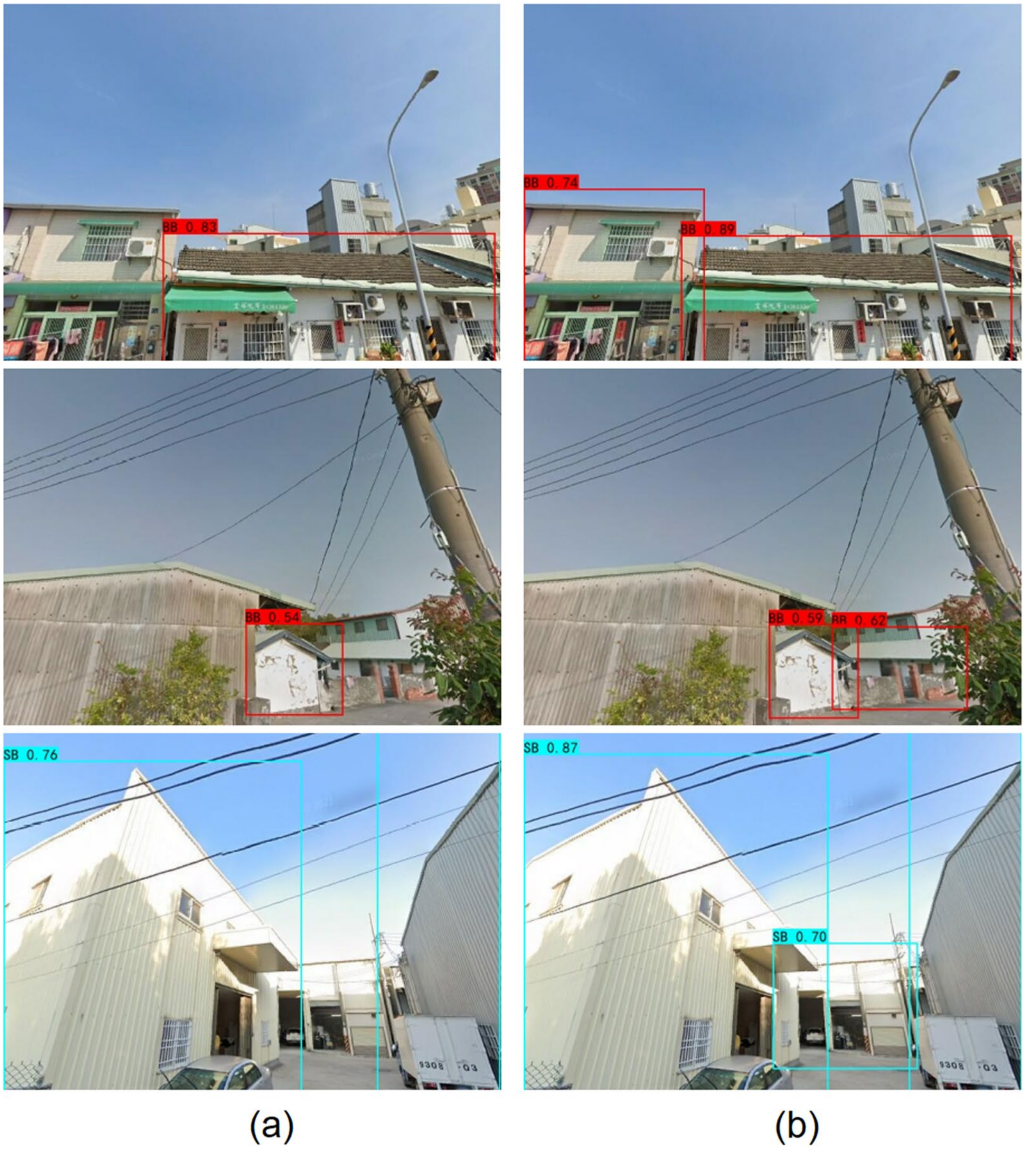


Fig. 8 Improved recognition of small objects after embedding the CBAM module. **a** Buildings detected by YOLOX-S. **b** Buildings detected by YOLOX-CS

recognition, which could potentially impact the model's performance and its applicability in real-world scenarios.

6.1 Analysis of empirical results

Primarily, the recognizability of buildings in images is often hindered by obstructions from other objects, a challenge more pronounced in the case of smaller structures. The Illustration Enhancement approach effectively mitigates this issue. In comparison with alternative data augmentation strategies, Illustration Enhancement demonstrates superior performance in recognizing obscured buildings. Although there's a slight decrement in recognizing unobstructed buildings, the overall accuracy rate sees a notable improvement. Integrating Illustration Enhancement with other augmentation methods could potentially further amplify the model's capabilities. The marked enhancement in recognition performance, as evidenced by training three distinct models on the Illustration Enhancement dataset, underscores its efficacy for Taiwanese architectural datasets.

In the specific context of small-scale building recognition, the incorporation of CBAM into YOLOX-S led to an improvement in the model's mean Average Precision (mAP), albeit with a decline in precision for certain

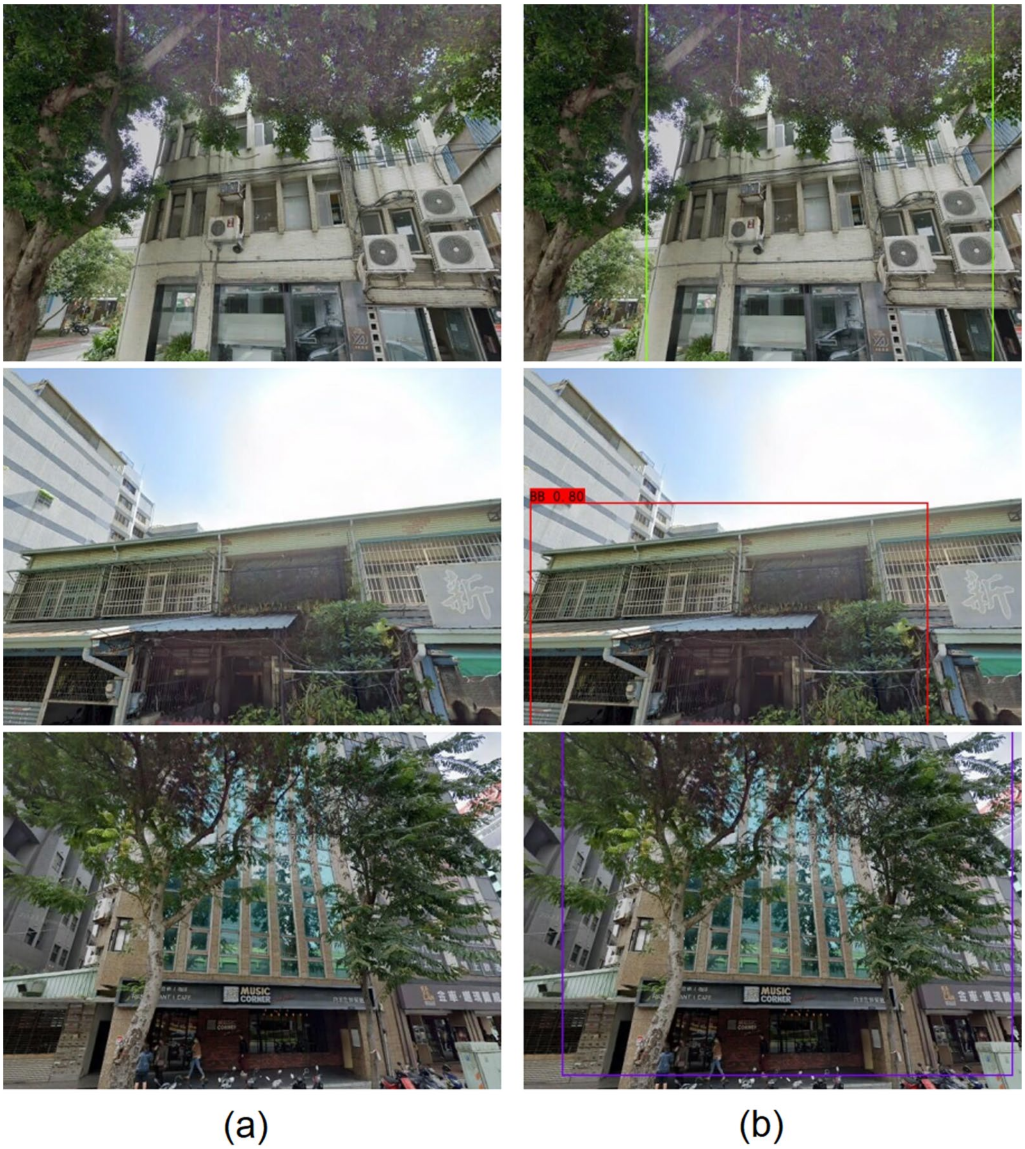


Fig. 9 Enhanced recognition of buildings obstructed by trees after using Illustration Enhancement. From top to bottom, the building structures are RC, BB, and SRC. **a** YOLOX-CS prediction results on the original dataset. **b** YOLOX-CS prediction results on the Illustration Enhancement dataset

categories (refer to Table 3). Further analysis revealed that YOLOX-CS identified numerous unmarked smaller buildings, contributing to a relative precision drop compared to YOLOv4. This issue likely stems from the inclusion of unmarked structures in the dataset generated from street view imagery. To address this, more stringent dataset processing is needed to minimize errors in marking and omissions, particularly in underrepresented categories.

Strategies for further enhancing model performance include expanding the architecture of the backbone network and tailoring models specifically for architectural structure recognition. Beyond architectural imagery, incorporating additional attributes such as building height and age could provide supplementary insights for the modelâ€™s predictions. The overarching aim is to develop a rapid, automated system for architectural structure recognition, enabling efficient assessment of building-related risks prior to seismic events.

6.2 Limitations and improvement

This study encounters certain constraints and outlines avenues for future enhancements:

1. Image Dependence: Our methodology, tailored for Google Street View imagery, necessitates specific standards for image resolution, field of view (fov), and angle of elevation. For analogous outcomes using this model, it's imperative that the images align closely with those from Taiwan's Google Street View, including aspects like exposure and shooting direction.
2. Challenges in Image Selection: In sourcing images through the Google Street View API, issues such as absence of the target building, incomplete structures, or excessive background buildings were common. This necessitated manual image selection, a process not in line with the automation ethos. Future enhancements might include capturing multiple images of a target building from varying horizontal angles, thereby elevating the success rate of recognition.
3. Incorporating Additional Factors: While the study successfully established an architectural structure map, a comprehensive seismic risk assessment requires factoring in additional elements, such as building height and unauthorized extensions. Subsequent research could explore integration with remote sensing techniques to swiftly construct a detailed seismic risk assessment map, considering a broader range of influential factors.

7 Conclusion

The relentless evolution of object detection technology has unlocked new avenues for swiftly identifying architectural structures. However, current object detection models grapple with certain challenges, notably in processing small and occluded targets. In response to these challenges, our study introduces the Illustration Enhancement data augmentation method and the YOLOX-CS model. The former significantly bolsters the model's capacity to identify occluded targets, while the latter not only retains superior performance but also enhances the detection of small-scale targets. The experimental findings indicate that both approaches exhibit commendable performance on the Taiwanese building dataset.

A key contribution of our research lies in the development and validation of novel methods tailored for architectural structure recognition, confirming the efficacy of both Illustration Enhancement and YOLOX-CS in boosting recognition precision. These advancements pave the way for the creation of a fast and accurate system for architectural structure identification, aimed at enhancing the assessment of potential seismic risks. Furthermore, we envisage the applicability of these methodologies in other regions with architectural styles akin to those in Taiwan, thus broadening their scope of practical deployment.

Acknowledgements

We acknowledge the support of the National Science and Technology Council in Taiwan for this research, under the grant MOST 110-2221-E-008-020.

Author contributions

YYZ conceptualized and supervised the research. WHC and SKW conducted primary experiments, collected and analyzed data. WYC drafted and edited the manuscript, and served as the corresponding author. All authors read and approved the final manuscript.

Availability of data and materials

The datasets used in this study were collected from publicly available sources, as detailed in the 'Datasets' section. The data sources include the Taipei City Historical Usage License Summary and the Taichung City Buildings WGS84 dataset. Additional information about these datasets and their availability can be found in the 'Datasets' section of this paper.

Declarations

Competing interests

The authors declare that they have no competing interests

Received: 29 August 2023 Accepted: 21 February 2024
Published online: 28 February 2024

References

Acevedo AB (2017) Evaluation of the seismic risk of the unreinforced masonry building stock in antioquia, colombia. Nat Hazards 86:31–54. https://doi.org/10.1007/s11069-016-2647-8

Bilal M, Hanif MS (2019) Benchmark revision for hog-svm pedestrian detector through reinvigorated training and evaluation methodologies. IEEE Trans Intell Transp Syst 21(3):1277–1287. https://doi.org/10.1109/TITS.2019.2906132

Bochkovskiy A, Wang C-Y, Liao H-YM (2020) YOLOv4: optimal speed and accuracy of object detection. https://doi.org/10.48550/arXiv.2004.10934

Chen J, Takiguchi T, Ariki Y (2017) Rotation-reversal invariant hog cascade for facial expression recognition. SIViP 11:1485–1492. https://doi.org/10.1007/s11760-017-1111-x2

Ding W, Zhang L (2021) Building detection in remote sensing image based on improved yolov5. In: 2021 17th International Conference on Computational Intelligence and Security (CIS). https://doi.org/10.1109/CIS54983.2021.00036 . IEEE

Du S et al (2021) Fa-yolo: an improved yolo model for infrared occlusion object detection under confusing background. Wirel Commun Mob Comput. https://doi.org/10.1155/2021/1896029

Ezat WA, Dessouky MM, Ismail NA (2020) Multi-class image classification using deep learning algorithm. J Phys Conf Ser. 1441(1):012021. https://doi.org/10.1088/1742-6596/1447/1/012021

Ge Z (2021) Yolox: exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430

Google: Google Street View Static API (2021) Google Maps Platform. https://developers.google.com/maps/documentation/streetview

Huo FF (2019) Automatic extraction of building rooftop outlines using airborne lidar: a review. Geomat World 26(5):1–13. https://doi.org/10.3969/j.issn.1672-1586.2019.05.001

Jocher G, Stoken A, Borovec J, NanoCode012, Chaurasia A, TaoXie Liu C, VA, Laughing, tkianai, yxNONG, Hogan A, lorenzomammana, AlexWang1900, Hajek J, Diaconu L, Marc Kwon Y, oleg wanghaoyang0106, Defretin Y, Lohia A, ml5ah Milanko B, Fineran B, Khromov D, Yiwei D, Doug Durgesh Ingham F (2021) ultralytics/yolov5: v5.0—YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations. https://github.com/ultralytics/yolov5

Kang J (2018) Building instance classification using street view images. ISPRS J Photogramm Remote Sens 145:44–59. https://doi.org/10.1016/j.isprsjprs.2018.02.006

Liu Y (2018) Multilevel building detection framework in remote sensing images based on convolutional neural networks. IEEE J Sel Top Appl Earth Observ Rem Sens 11(10):3688–3700. https://doi.org/10.1109/JSTARS.2018.2866284

Redmon J, Farhadi A (2018) YOLOv3: an Incremental Improvement. https://doi.org/10.48550/arXiv.1804.02767

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

Shi L (2021) Identifying damaged buildings in aerial images using the object detection method. Rem Sens 13(21):4213. https://doi.org/10.3390/rs13214213

Taichung City Government Open Data Platform: Taichung City Buildings_WGS84. https://opendata.taichung.gov.tw/dataset/bdaa52e5-b5d6-4a62-81b6-d4d5e9728c45. Accessed Sept 2019

Taipei City Government Open Data Platform: Taipei City Historical Usage License Summary. https://data.taipei/dataset/detail?id=c876ff02-af2e-4eb8-bd33-3d444f5052733. Accessed Dec 2020

Texture, Textures for 3d, graphic design and photoshop! https://www.textures.com/library. Accessed 1 Feb 2022

Wei Y (2016) Stc: a simple to complex framework for weakly-supervised semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(11):2314–2320. https://doi.org/10.1109/TPAMI.2016.2636150

Wieland M (2012) Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging. Soil Dyn Earthq Eng 36:70–83. https://doi.org/10.1016/j.soildyn.2012.01.003

Woo S (2018) Cbam: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV). https://doi.org/10.1007/978-3-030-01234-2_1

Zhang D (2018) Vehicle target detection methods based on color fusion deformable part model. EURASIP J Wirel Commun Netw 2018(1):1–6. https://doi.org/10.1186/s13638-018-1111-8

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.