Building Usage Classification in Indian Cities: Utilizing Street View Images and Object Detection Models

Yamini Sahu, Vasu Dhull, Satyajeet Shashwat, and Vaibhav Kumar

Indian Institute of Science Education and Research Bhopal (IISERB)
{yamini21,vasu21,satyajeet21,vaibhav}@iiserb.ac.in

Abstract. Urban land use maps at the building instance level are crucial geo-information for many applications, yet they are challenging to obtain. Land-use classification based on spaceborne or aerial remote sensing images has been extensively studied over the last few decades. Such classification is usually a patch-wise or pixel-wise labeling over the whole image. However, for many applications, such as urban population density estimation or urban utility mapping, a classification map based on individual buildings (residential, commercial, mixed-type, and religious) is much more informative. Nonetheless, this type of semantic classification still poses fundamental challenges, such as retrieving fine boundaries of individual buildings. Street view images (SVI) are highly suited for predicting building functions because building facades provide clear hints. Although SVIs are used in many studies, their application in generating building usage maps is limited.
Furthermore, their application to Indian cities remains void. In this paper, we propose a comprehensive framework for classifying the functionality of individual buildings. Our method leverages the YOLOs model and utilizes SVIs, including those from Google Street View and OpenStreetMap. Geographic information is employed to mask individual buildings and associate them with the corresponding SVIs. We created our own dataset in Indian cities for training and evaluating our model.

Keywords: Building usage classification · Object detection · Street view images · Open street map · Urban planning

1 Introduction

Urban land use classification at the individual level is essential for effective urban planning and management, influencing a wide range of decisions, including solar potential analysis [1], damage identification [2], and infrastructure development. Accurate classification of building usage types helps urban planners optimize resources, improve public services, and enhance community well-being. Traditional methods for building usage classification, such as satellite imagery, OpenStreetMap (OSM) data, and aerial imagery, have shown promise in various contexts; however, the application of SVI remains limited, particularly within the Indian urban landscape, where research is still developing.

This study aims to address the gap in knowledge by developing a robust framework for urban land use classification using SVI, thereby enhancing our understanding of building usage types within complex urban environments. By employing advanced object detection models, specifically YOLOv8s and YOLOv8n [3], we seek to classify buildings into four distinct categories: residential, commercial, religious, and mixed-use. The inclusion of a mixed-use category is particularly significant, as urban buildings often serve multiple functions—commercial spaces on the ground floor and residential units above, for instance.

Our methodology aims to capture the multifaceted nature of urban buildings and improve classification accuracy in mixed-use areas, which poses significant challenges in existing classification frameworks. This research contributes to enhancing urban land use classification methodologies and provides valuable insights for urban planners and policymakers. By offering a more nuanced understanding of urban building types, we facilitate better decision-making in rapidly growing urban areas, ultimately promoting sustainable development and effective urban management.

1.1 Contributions

Our major contribution lies in the creation of a novel dataset tailored explicitly for building detection in Indian cities. This meticulously curated dataset incorporates diverse urban scenarios typical of Indian streetscapes, including varying architectural styles, occlusions by trees and vehicles, and different lighting conditions. Leveraging this dataset, we have developed a pioneering framework for building detection utilizing SVI. Our approach addresses challenges unique to urban environments in India, such as densely packed buildings and irregular structures, ensuring robust and accurate detection. By employing the state-of-the-art YOLOvs object detection model for facade object detection, our framework significantly enhances the precision of building detection. This innovation not only contributes a valuable resource for future research but also sets a new benchmark for urban infrastructure analysis using computer vision methodologies. We are releasing this dataset to support future research and collaboration [4].

1.2 Paper Organization

Our paper is structured as follows. We present the related work in Section 2. Section 3 outlines the proposed methodology, while Section 4 provides a detailed description of the proposed algorithm and implementations. In Section 5, we assess the performance of the YOLOv8s and YOLOv8n object detection models. Section 6 discusses the results and its implications. Furthermore, Section 7 presents the limitations of our study, and Section 8 summarizes our findings. Lastly, Section 9 outlines potential avenues for future research.

2 Related Work

Spaceborne or aerial remote sensing images have been extensively studied over the past decades. Typically, classification using these images involves patch-wise or pixel-wise labeling across the entire image. While aerial remote sensing images can provide information about building usage types, they often face challenges in accurately identifying the defined boundaries of building footprints. Figure 1 illustrates building usage classification using aerial remote-sensing images [5–9]. From a top-down view, structures often appear similar, making it difficult to distinguish between different building types. Figure 2 highlights the challenge of mixed-use areas, where commercial, residential, and mixed-type buildings may all appear within the same image patch, complicating the classification process.

Fig. 1: Example of land-use classification using satellite image.

Using OSM data, we can obtain information about building usage types. However, there are several issues with OSM data [10–15]. The accuracy of building footprints in OSM is often questionable, with many building footprints either missing or inaccurately represented. Additionally, there is a lack of labeled data within OSM, which can limit its usefulness for detailed analysis. Another significant concern is the irregularity of updates. OSM data is often updated on a voluntary basis, meaning it cannot always be relied upon for timely or consistent updates. This inconsistency makes it challenging to use OSM data as a sole source for accurate and up-to-date information in urban planning and analysis.

SVI has also been used for building classification, but its application is limited to specific areas. Previous work on building instance classification using SVI primarily focused on datasets from foreign environments [16]. In these studies, buildings were classified into eight categories: apartment, church, garage, house, industrial, office building, retail, and roof. The accuracy of these classifications was relatively low [17–21]. Researchers tested various convolutional Neural Network (CNN) architectures, including AlexNet [22], VGG16 [23], ResNet18 [24], and ResNet34 [24]. Despite the potential of CNNs for building classification, the results indicated significant room for improvement. One of the challenges is the variability in architectural styles and urban layouts across different regions, which can impact the performance of models trained on foreign datasets when applied to local environments. To enhance the accuracy and applicability of building classification using SVI, it is crucial to develop and train models on region-specific datasets and explore advanced techniques to handle the diversity of building types and appearances.

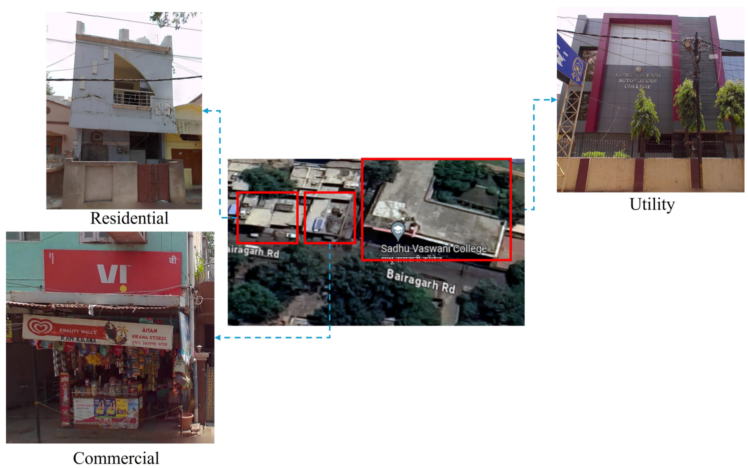

Fig. 2: These buildings do not belong to the same category, even though they are located in the same land-use area. Furthermore, compared to roof structures, the facade structures displayed in SVI provide richer and more sufficient information for building classification using detection methods.

3 Proposed Methodology

In this study, we developed a comprehensive methodology for building detection using geotagged SVI, leveraging a custom dataset tailored to the complexities of urban environments. We categorized buildings into four distinct classes: residential, commercial, religious, and mixed-type, with the mixed-type class representing buildings with multiple functions, such as commercial spaces on the ground floor and residential units above. For the detection phase, we employed a range of object detection models, including YOLOv8s and YOLOv8n, to accurately identify and locate buildings within the images.

Our building detection model provides a detailed understanding of urban land use. Geographic information was used to mask individual buildings and align them with SVIs, ensuring a more precise classification by integrating spatial data with visual features. The effectiveness of this approach was validated using our custom dataset from Indian cities, demonstrating improvements over existing methods and offering valuable insights for urban planning and management through a nuanced classification of diverse building types.

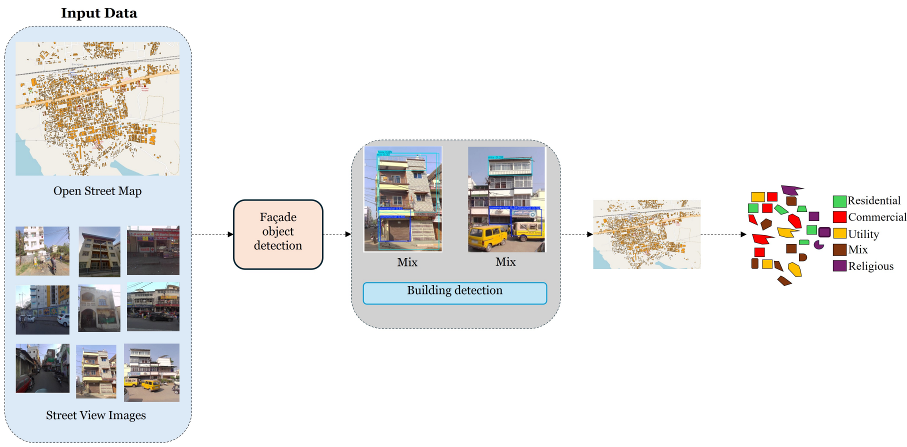

Fig. 3: The proposed framework for building usage classification at a level of individual building.

3.1 Building footprints from GIS map

OSM serves as a valuable resource for obtaining building footprints for a specific study area. By leveraging OSM data and geofabrik [1], we can access detailed geographic information, including building footprints, which are essential for urban planning and spatial analysis. Various tools and platforms are available to extract building footprints from OSM. In our work, we utilized OSMnx[2], a Python library, to efficiently download and visualize OSM data. Begin by defining the study area, either by specifying geographical boundaries or using a location name. OSMnx then queries OSM's servers, enabling the extraction of building footprints and other relevant geographic features.

3.2 Collection of Geotagged SVI

We began by identifying the geographical boundaries of the study area. Using tools like the Google Street View API[3], we captured panoramic images at specified intervals within these boundaries. These panoramic images were then converted into normal images for easier analysis. Each image is inherently tagged with latitude and longitude coordinates, ensuring accurate geolocation.

3.3 Façade object detection model

In this work, we utilized various YOLO models for building facade object detection, including YOLOv8s and YOLOv8n. These models were trained to identify

[1] https://www.geofabrik.de/
[2] https://osmnx.readthedocs.io/en/stable/
[3] https://developers.google.com/maps/documentation/streetview/overview

four classes of buildings: residential, commercial, mixed-type, and religious. By employing these models, we aimed to classify and geolocate buildings within the images accurately. Table 1 provides a brief overview of the different building classes.

Table 1: Building class descriptions from usage types.

Title	Building types
Residential	house, apartment, hostel
Commercial	shop, restaurant, bank, shopping complex, mall, vegetable center, Indian coffee house
Religious	temple, mosque, church
Mixed-type	combination of commercial and residential

4 Proposed Algorithm

Algorithm 1 outlines a systematic framework for detecting buildings in geotagged SVI. The process begins with *Data Acquisition*, where a collection of geotagged SVI, denoted as $\{I_i\}_{i=1}^N$, is gathered and preprocessed. This step ensures that the images are in an optimal format for subsequent analysis. The next phase, *Building Detection*, involves applying the YOLO model to each image to identify buildings. Specifically, for each image I_i, the YOLO model detects buildings and generates bounding boxes B_i that outline these structures. The outcome of this step is a set of bounding boxes for all images, $\{B_i\}_{i=1}^N$. In the final *Output Results* phase, the algorithm displays the detected buildings by presenting these bounding boxes. This structured approach facilitates the accurate identification and visualization of buildings in geotagged SVI.

Algorithm 1 Building Detection Framework

Input: Geotagged SVI $\{I_i\}_{i=1}^N$
Output: Detected buildings $\{B_i\}_{i=1}^N$
Step 1: Data Acquisition
  Obtain and preprocess geotagged SVI $\{I_i\}_{i=1}^N$
Step 2: Building Detection
  Apply the YOLO model to detect buildings in each image:
    $B_i \leftarrow \text{YOLO}(I_i)$
  where B_i denotes the bounding boxes of detected buildings in I_i
Step 3: Output Results
  Generate and present the detected buildings $\{B_i\}_{i=1}^N$

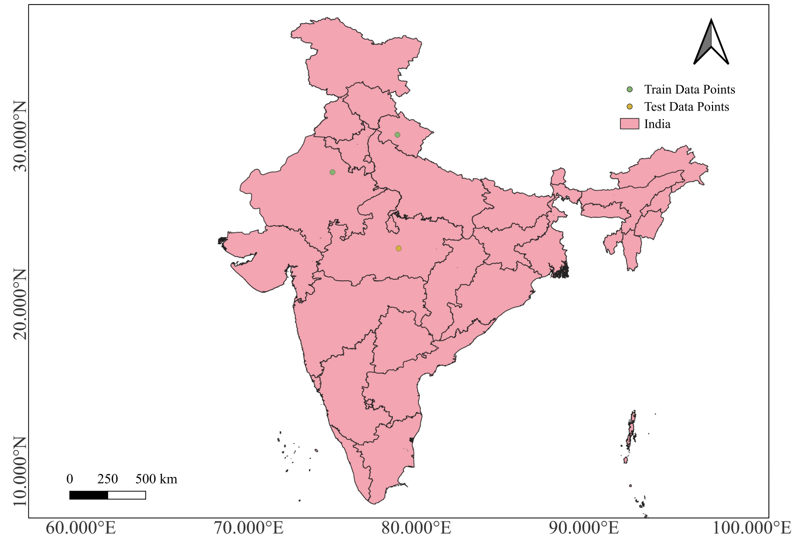

Fig. 4: In this case study, we used images from *Rohtak* and *Mandi* (shown in green dot) for training and images from the *Bhopal* (shown in yellow dot) region for testing purposes.

4.1 Data Collection and Pre-processing

We have selected *Mandi* in *Himachal Pradesh* and *Rohtak* in *Haryana* as the study areas for this research. We used SVI from *Bhopal* in *Madhya Pradesh* for testing purposes. The Figure 4 represents the study area, including the testing and training points. The data collection part is a lengthy and effort-intensive process. The collected dataset consists of geotagged SVI from both cities. It covers different types of built-up areas, including residential, commercial, religious, and mixed-use buildings, as mentioned in Table 1.

During the data preprocessing step, we identified and removed incorrect data, such as blurred or repetitive images. As a result, we had a total of 512 SVI of various buildings, which were divided into training, validation, and test sets. We manually annotated these data in four categories: residential (e.g., apartment, hostel, bungalow), commercial (e.g., shopping complex, mall, stationery shop, restaurant, hotel, tea stall), religious (e.g., mosque, temple, church), and mixed-type buildings (a combination of commercial and residential).

4.2 Detection of Building types from the Data set

The objective of this study is to classify building usage types using geotagged SVI. Figure 5 shows the various building types in the Indian scenario, categorized into residential, commercial, religious, and mixed-use classes. We utilized our proposed methodology for building usage classification, employing object detection techniques to achieve this.

Fig. 5: A visual representation of SVI with buildings belonging to different cities and classes from the dataset.

4.3 Evaluating the Classification Model

In evaluating the performance of our models, we employed several key metrics commonly used in object detection tasks. Box Precision (Box(P)) measures the proportion of true positive bounding boxes out of all predicted bounding boxes, providing insight into the accuracy of the model's predictions. Box Recall (Box(R)) reflects the proportion of true positive bounding boxes out of all actual bounding boxes, indicating the model's ability to detect all relevant objects. To summarize the precision of detections, we used Mean Average Precision at IoU 0.5 (mAP50), which averages the precision across different recall levels at an Intersection over Union (IoU) threshold of 0.5. Additionally, Mean Average Precision across IoU thresholds from 0.5 to 0.95 (mAP50-95) provides a more comprehensive evaluation by averaging precision over a range of IoU thresholds, capturing the model's performance across various levels of detection overlap. These metrics collectively offer a robust assessment of the model's detection capabilities and accuracy.

5 Results

In this research, we developed a framework for building usage classification using SVI. We employed the YOLOv model for facade object detection. Following this, we obtained the results.

To detect building usage types, we initially utilized several YOLO models. These models are well-suited for object detection tasks due to their efficiency and accuracy. By applying YOLO to SVI, we were able to identify various features associated with different building usages effectively. This step was fundamental to ensure that our subsequent classification efforts were based on reliable detection data.

Figure 6a represents the accurate predictions of the model, while Figure 6b shows the model's incorrect predictions. The errors in the model's predictions are primarily due to the limited number of training data points and an unbalanced dataset, which significantly affects its performance.

The YOLOv8n model demonstrates good performance (as shown in Table 2) in classifying commercial buildings, with high precision (0.718) and recall (0.641). It performs well with mixed-use buildings in terms of recall (0.829) but has lower precision (0.335). The model struggles with religious buildings, showing high precision but no detected instances, and shows moderate results for residential buildings. Overall, YOLOv8n excels in commercial classification while facing challenges in other categories.

Predicted- residential
Ground truth- residential

Predicted- residential
Ground truth- residential

Predicted- residential, commercial
Ground truth- residential

Predicted- residential
Ground truth- residential

(a) The predictions from the YOLOv8s model show that it accurately classifies the residential and commercial classes.

Predicted-residential, commercial
Ground truth - mix

Predicted- commercial
Ground truth - mix

Predicted-residential, commercial
Ground truth - commercial

Predicted-residential, commercial
Ground truth - commercial

(b) The YOLOv8n model demonstrates accurate classification for residential and commercial classes; however, it struggles with mixed-type and religious classes, with incorrect predictions in the mixed-type category and no predictions in the religious class.

Fig. 6: YOLO model performance.

Table 2: Validation Performance Metrics of YOLOv8n.

Class	**Images**	**Box(P)**	**Box(R)**	**mAP50**	**mAP50-95**
all	145	0.657	0.467	0.466	0.213
commercial	85	0.718	0.641	0.729	0.309
mix	34	0.335	0.829	0.600	0.281
religious	11	1.000	0.000	0.094	0.042
residential	15	0.577	0.400	0.441	0.221

From Table 2 and Table 3, we found that YOLOv8n model performed better than YOLOv8s across various validation metrics. The YOLOv8n model demonstrates excellent overall performance, achieving a Box Precision of 0.657, a Box Recall of 0.467, a mAP50 of 0.466, and a mAP50-95 of 0.213. This strong performance is consistent across individual classes. In the commercial category, YOLOv8n achieves impressive results with a Box Precision of 0.718, a Box Recall of 0.641, an mAP50 of 0.729, and an mAP50-95 of 0.309. The model also

Table 3: Validation Performance Metrics of YOLOv8s

Class	Images	Box(P)	Box(R)	mAP50	mAP50-95
all	145	0.309	0.348	0.113	0.0431
commercial	85	0.155	0.504	0.224	0.0852
mix	34	0.0802	0.886	0.167	0.0677
religious	11	0.000	0.000	0.0145	0.00453
residential	15	1.000	0.000	0.0462	0.0149

performs notably well in the mixed class, with a Box Precision of 0.335 and a mAP50 of 0.600, although the Box Recall is slightly lower at 0.829. For the religious class, YOLOv8n achieves a perfect Box Precision of 1.000, with mAP50 and mAP50-95 values of 0.094 and 0.042, respectively. Even in the residential class, YOLOv8n outperforms with a Box Recall of 0.400, a mAP50 of 0.441, and a mAP50-95 of 0.221, despite having a lower Box Precision compared to YOLOv8s. Overall, YOLOv8n's superior performance across these metrics highlights its robustness and effectiveness in handling the validation dataset.

6 Discussion

This work focuses on the classification of building usage at the individual building level. While various techniques exist for building usage classification, such as those utilizing satellite imagery and OSM data, there is limited research leveraging SVI for this purpose. Some studies have centered on identifying specific building features, but our work emphasizes detection at the individual building level. By concentrating on this detection approach, we aim to facilitate the real-time identification of different building types.

This has significant implications, offering a practical and valuable contribution to the field of building usage classification research. Real-time detection of building types can enhance various applications, such as urban planning, navigation, and smart city development. This method can also improve data accuracy and provide timely updates, making it an essential tool for researchers and practitioners working with urban environments and infrastructure.

However, this research also faces challenges, primarily due to the limited availability of geotagged images for building detection. Using geotagged SVI, we encounter numerous occlusions in the images, such as vehicles, vegetation, and humans. Additionally, the dataset is unbalanced, which adversely affects the results. Despite these challenges, our work highlights the potential of using SVI for building usage classification and the need for more comprehensive datasets to improve accuracy and reliability.

6.1 Model Comparison

The YOLOv8n model outperforms YOLOv8s across all metrics, including Box Precision, Box Recall, mAP50, and mAP50-95. Specifically, YOLOv8n shows higher overall performance and better results in individual classes, highlighting its superior capability in detecting buildings in geotagged SVI.

6.2 Implications

The challenges faced in categorizing mixed-use and religious buildings indicate the importance of diversifying the training data to ensure balanced performance across all building categories. A more comprehensive dataset will improve the model's classification accuracy and enhance its generalizability to different urban settings. Furthermore, the limitations in retrieving fine boundaries highlight the need for more precise segmentation techniques, which are crucial for reliable land-use classification. Semi-supervised learning for the annotation process will significantly reduce manual efforts and improve label quality, particularly in large-scale datasets. This approach can accelerate the framework's scalability, making it more efficient and applicable to various contexts, including cities in India. Future model iterations can offer more accurate and globally applicable insights by addressing these issues, contributing to better urban planning and decision-making.

7 Limitation

This research work contributes to building usage classification, aiding various applications in urban planning at the individual building level. Using our method, we achieved excellent results overall; however, it did not perform accurately for the mixed-type class. This is because buildings often have commercial use on the ground floor and residential use on the remaining floors. Another limitation is the limited number of data points. In future work, we aim to collect more data points to improve the results. Currently, around 50% of roadside buildings fall into the mixed-type category, but unfortunately, we have a limited number of data points for this class.

The performance of the YOLOv model is hindered by an unbalanced dataset and limited data points in specific classes, such as mixed and religious buildings. This data imbalance impacts the model's ability to accurately classify these less-represented categories. Additionally, the manual labeling process is time-consuming and costly.

8 Conclusion

The classification of building usage is a crucial tool that serves a wide range of urban planning applications. Providing detailed insights into how buildings are utilized across a city enables various agencies and organizations to implement solutions more effectively. For instance, it can be instrumental in conducting solar potential analysis, which assesses the feasibility and efficiency of solar panel installations on different buildings. Additionally, it aids in population estimation by offering data on residential building usage, thereby supporting infrastructure development and resource allocation.

The framework underpinning this work has been delivering promising results. Its potential extends beyond static analysis; these methods can also be adapted for real-time building type prediction. Such predictive capabilities are invaluable for dynamic urban management and emergency response scenarios. The framework exhibits strong performance in categorizing commercial and residential buildings. However, it encounters challenges when dealing with mixed-use and religious buildings. This difficulty largely stems from the limited or insufficient data available for these categories, which hinders the accuracy and reliability of the predictions.

In addressing these challenges, the YOLOv8n model has demonstrated exceptional performance, delivering some of the best results observed so far. Its advanced capabilities in object detection and classification make it a powerful tool for refining building usage classification. As we continue to enhance the dataset and model accuracy, the potential applications of this framework in urban planning and management will only grow, paving the way for smarter, more efficient cities.

9 Future Work

In future research, we aim to address the limitations encountered in this study by enhancing our data collection strategy. Specifically, we plan to expand the dataset to include a more diverse and balanced representation of building types, mainly focusing on underrepresented categories such as mixed-use and religious buildings. By collecting more data across varied urban environments, we aim to improve the model's robustness and accuracy in categorizing these challenging types. Additionally, improving the boundary retrieval process is essential for obtaining precise land-use classification. We plan to explore advanced segmentation techniques to delineate fine boundaries of individual buildings more effectively, ensuring higher precision in object detection and classification tasks. Employing techniques like semantic segmentation, alongside integrating semi-supervised learning methods, will help streamline the annotation process. These enhancements can facilitate more accurate and scalable data labeling, improving the model's performance across diverse building categories. Finally, applying this framework to Indian cities will be a significant future focus. Extending the framework's applicability to diverse geographic and cultural contexts will offer valuable insights into the global adaptability of the model. This will allow us to evaluate its efficacy in regions with unique urban patterns, ensuring broader relevance and utility.

References

1. Zuoqi Chen, Bailang Yu, Yong Li, Qiusheng Wu, Bin Wu, Yan Huang, Siyuan Wu, Siyi Yu, Weiqing Mao, Feng Zhao, et al. Assessing the potential and utilization

of solar energy at the building-scale in shanghai. *Sustainable Cities and Society*, 82:103917, 2022. 1
2. Vahid Reza Gharehbaghi, Ehsan Noroozinejad Farsangi, Tony Y Yang, and Iman Hajirasouliha. Deterioration and damage identification in building structures using a novel feature selection method. In *Structures*, volume 29, pages 458–470. Elsevier, 2021. 1
3. Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. 2
4. Yamini Sahu, Vasu Dhull, Satyajeet Shashwat, and Vaibhav Kumar. Indian Street View Images of buildings. https://osf.io/4e2bd/, 2024. 2
5. Christian Geiß, Patrick Aravena Pelizari, Mattia Marconcini, Wayan Sengara, Mark Edwards, Tobia Lakes, and Hannes Taubenböck. Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *ISPRS journal of photogrammetry and remote sensing*, 104:175–188, 2015. 3
6. Eike Jens Hoffmann, Yuanyuan Wang, Martin Werner, Jian Kang, and Xiao Xiang Zhu. Model fusion for building type classification from aerial and street view images. *Remote Sensing*, 11(11):1259, 2019. 3
7. Jiayi Li, Xin Huang, Lilin Tu, Tao Zhang, and Leiguang Wang. A review of building detection from very high resolution optical remote sensing images. *GIScience & Remote Sensing*, 59(1):1199–1225, 2022. 3
8. Zhenyu Lu, Jungho Im, Jinyoung Rhee, and Michael Hodgson. Building type classification using spatial and landscape attributes derived from lidar remote sensing data. *Landscape and Urban Planning*, 130:134–148, 2014. 3
9. Junfei Xie and Jianhua Zhou. Classification of urban building type from high spatial resolution remote sensing imagery using extended mrs and soft bp network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(8):3515–3528, 2017. 3
10. Abhilash Bandam, Eedris Busari, Chloi Syranidou, Jochen Linssen, and Detlef Stolten. Classification of building types in germany: A data-driven modeling approach. *Data*, 7(4):45, 2022. 3
11. Kuldip Singh Atwal, Taylor Anderson, Dieter Pfoser, and Andreas Züfle. Predicting building types using openstreetmap. *Scientific Reports*, 12(1):19976, 2022. 3
12. Yuheng Zhang, Qi Zhou, Maria Antonia Brovelli, and Wanjing Li. Assessing osm building completeness using population data. *International Journal of Geographical Information Science*, 36(7):1443–1466, 2022. 3
13. Hongchao Fan, Alexander Zipf, and Qing Fu. Estimation of building types on openstreetmap based on urban morphology analysis. *Connecting a digital Europe through location and place*, pages 19–35, 2014. 3
14. Cidalia C Fonte, Patrícia Lopes, Linda See, and Benjamin Bechtel. Using openstreetmap (osm) to enhance the classification of local climate zones in the framework of wudapt. *Urban Climate*, 28:100456, 2019. 3
15. Qi Zhou, Yuheng Zhang, Ke Chang, and Maria Antonia Brovelli. Assessing osm building completeness for almost 13,000 cities globally. *International Journal of Digital Earth*, 15(1):2400–2421, 2022. 3
16. Jian Kang, Marco Körner, Yuanyuan Wang, Hannes Taubenböck, and Xiao Xiang Zhu. Building instance classification using street view images. *ISPRS journal of photogrammetry and remote sensing*, 145:44–59, 2018. 3
17. Dominik Laupheimer, Patrick Tutzauer, Norbert Haala, and Marc Spicker. Neural networks for the classification of building use from street-view imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:177–184, 2018. 3

18. Kun Zhao, Yongkun Liu, Siyuan Hao, Shaoxing Lu, Hongbin Liu, and Lijian Zhou. Bounding boxes are all we need: street view image classification via context encoding of detected buildings. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2021. 3
19. Rui Cao, Jiasong Zhu, Wei Tu, Qingquan Li, Jinzhou Cao, Bozhi Liu, Qian Zhang, and Guoping Qiu. Integrating aerial and street view images for urban land use classification. *Remote Sensing*, 10(10):1553, 2018. 3
20. Yoshiki Ogawa, Chenbo Zhao, Takuya Oki, Shenglong Chen, and Yoshihide Sekimoto. Deep learning approach for classifying the built year and structure of individual buildings by automatically linking street view images and gis building data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1740–1755, 2023. 3
21. Surya Prasath Ramalingam and Vaibhav Kumar. Automatizing the generation of building usage maps from geotagged street view images using deep learning. *Building and Environment*, 235:110215, 2023. 3
22. Md Zahangir Alom, Tarek M Taha, Christopher Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Brian C Van Esesn, Abdul A S Awwal, and Vijayan K Asari. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*, 2018. 3
23. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 3
24. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3