A Computer Vision Approach for Building Facade Component Segmentation on 3D Point Cloud Models Reconstructed by Aerial Images

Yu Hou M.Sc.[a*], Zoe Mayer M.Sc.[b], Zhaoyang Li[a], Dr. Rebekka Volk[b], Dr. Lucio Soibelman[a]
[a] University of Southern California, USA, [b] Karlsruhe Institute of Technology, Germany
yuhou@usc.edu

Abstract. Segmenting windows and doors on 3D point cloud models allows for heat loss audits around these areas. Researchers have collected aerial images to reconstruct 3D models for large districts, but easily accessible training datasets with data acquired on ground level cannot be directly used for segmentation on 3D models reconstructed by aerial images. Additionally, building a new dataset is a time-consuming and labour-intensive process. Therefore, we propose a segmentation approach that uses open source training datasets to segment windows and doors on façade images rendered from 3D point clouds. The results show that our approach can make full use of open source datasets to segment windows and doors, and that such trained segmentation models performs differently for different building styles. In addition, different algorithms result in various degrees of accuracy and segmentation on windows performs better than on doors.

1. Introduction

Thermography, a non-destructive inspection technology, is used for heat loss energy audits. However, the most common current data collection approaches only allow individual building energy audit by deploying handheld infrared thermography cameras to collect thermal information from building facades. The biggest downside of current data collection approaches is efficiency. Such approaches also do not consider groups of buildings in large district areas in which interconnected buildings impact each other's thermal behaviors, especially, those connected within the same district heating network. More precisely speaking, if one building that is located in the middle of a heating network has unfixed heat loss issues, it will force buildings located downstream in the network to draw more heat to keep warm, resulting in more energy wasted through the middle-network buildings. Thus, there is a need to investigate novel methods and frameworks for building heat energy audits for large districts. Driven by the need of efficient and thorough energy audits for large districts, researchers have been deploying unmanned aircraft systems (UASs) to improve the data collection process (Hou *et al.*, 2019).

The benefits of using UASs to collect both thermal (infrared spectrum) and RGB (red-green-blue visible light) images include the higher data collection speed and availability of a bird's eye view, which can improve collection efficiency and comprehensively explore high areas of building façades that handheld thermal cameras cannot reach. Thermal and RGB imagery data collected from UASs allow the reconstruction of 3D point cloud models using photogrammetry technology. In order to obtain the 3D point cloud models that can integrate both thermal and RGB information, researchers have deployed different data fusion approaches (Hou *et al.*, 2021; Shahandashti *et al.*, 2010).

Distinguishing windows and other heat loss related building façade elements is an important step for energy audits. Semantic segmentation using 3D point cloud building models fused with thermal information allows researchers to detect heat loss from window and door edges and to monitor thermal bridges and areas of moisture on walls. The first step is to distinguish these façade components. However, in available open source image databases, facade images with their labeled components (the ground truth information) that were taken from the ground cannot be directly used to train a model to segment façade elements either in drone-based aerial images or in point cloud models reconstructed by these aerial images. To manually label newly captured aerial images and then build a new dataset is a potential option. However, conducting ground truth coding on these aerial images is both time-consuming and labor-intensive. Therefore, studies on the use of open source databases obtained from the ground to train artificial neural network (ANN) models for façade components segmentation using aerial images can provide an alternative that does not require the building of a new database.

To reduce labeling time and maintain the benefits of using UAS-based data collection, we propose a framework to train segmentation models using open source terrestrial image datasets taken from the ground to predict semantic information on building façades. In this paper, we introduce the results of our approach that was tested on two different datasets from Karlsruhe, Germany, one from a university campus, and the other from a central business district (Mayer *et al.*, 2021). The research introduced in this paper was designed to answer the following questions: (1) How does the proposed approach perform on different testing datasets with different building styles? and (2) How does the segmentation accuracy vary for different building components? This paper is organized as follows. We introduce and detail our approach in Section 2. Experiment results are described in Section 3, followed by evaluation and discussion in Section 4. Finally, we present our conclusions in Section 5.

2. Methodology

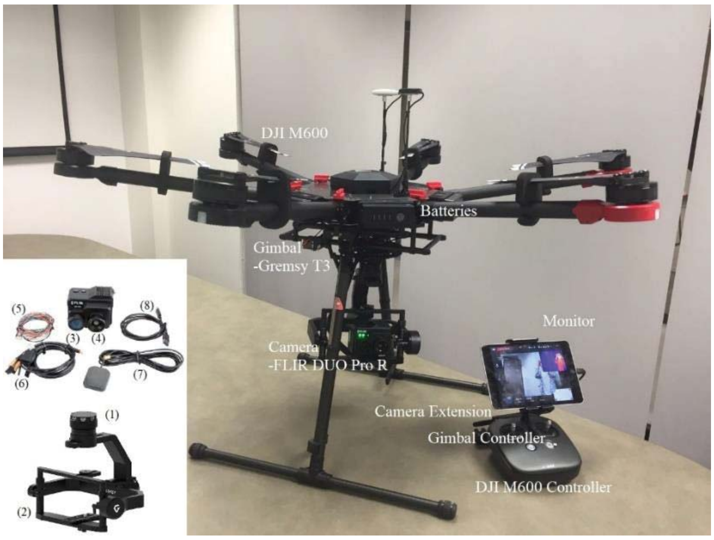
The proposed approach consists of the following four steps: (1) reconstructing a 3D point cloud model with aerial imagery data, (2) rendering 2D images from the 3D model, (3) training a semantic segmentation ANN model with open source datasets, and (4) predicting segmentation results on the rendered 2D images. We also designed the evaluation and validation metrics for the proposed approach.

Note that with the exception of the 3D models that were reconstructed by ContextCapture, a commercial photogrammetry software kit (Shi and Ergan, 2020; Chen *et al.*, 2020), most of the algorithms used in this study (e.g. Thermal-RGB data fusion, ANN model training, image rendering) were implemented using Python. The involved implementing libraries include Open3D (Zhou et al., 2018), OpenCV (Bradski, 2000), scikit-learn (Pedregosa *et al.*, 2019), and PyRender (Matl, Mahler and Goldberg, 2017).

2.1 Photogrammetry and 3D Point Cloud Model Reconstruction

There are many approaches to detecting defects in building envelops, such as fan pressurization (blower door test), ultrasound (tone test), and thermography. Thermography, as a non-destructive technique, is considered the most useful method because it can detect thermal values in envelops allowing for heat loss and moisture detection. However, current thermography methods mostly focus on handheld data collection (Dino *et al.*, 2020; Yang, Su and Lin, 2018), which is not recommended for an energy audit for a group of buildings in a large district. As such, researchers have mounted thermal and RGB cameras on UASs for more efficient large district data collection.

As shown in Figure 1, the data acquisition system used in this study included the drone (DJI M600), camera (FLIR Duo Pro R), control modules, and other equipment. The DJI M600 is a state-of-the-art aerial platform designed for industrial data collection. The FLIR Duo Pro R camera has both photographic and thermal lenses integrated into a single package that enables simultaneous RGB and thermal image data collection. Additionally, the control system allows to remotely operate the drones and the FLIR camera to collect data with the desired flight altitude and camera angles.

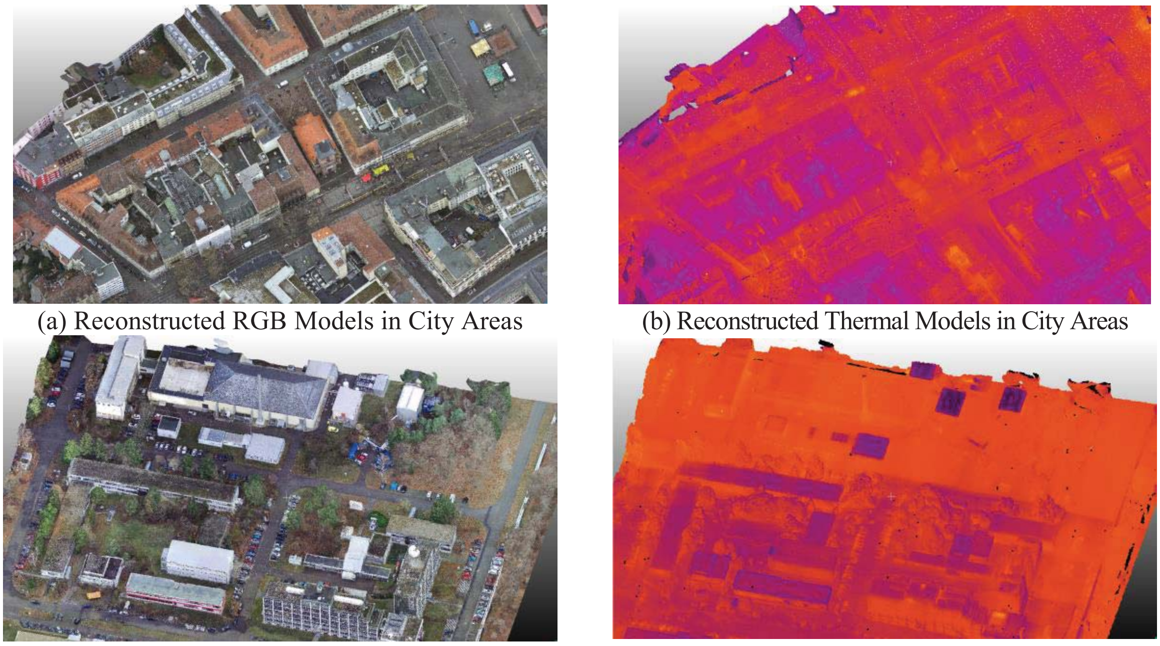


(1) Gimbal - Connection to DJI M600; (2) Gimbal - Frame for Camera; (3) FLIR DUO Pro R – Visible Lens Barrel; (4) FLIR DUO Pro R – IR Lens Barrel; (5) FLIR DUO Pro R – Electric Wires; (6) FLIR DUO Pro R – Integration Cable; (7) FLIR DUO Pro R – GPS Antenna Cable; (8) FLIR DUO Pro R – USB Cable.

Figure 1: Cameras Setup for the Unmanned Aircraft System

After both RGB and thermal images with designed image overlapping rates were collected with the drone, images were used to reconstruct 3D point cloud models over the survey areas using the photogrammetry technique. We collected over 10,000 images for both campus and city areas. There were over 12 buildings included for these two areas. Photogrammetry is the technology for 3D modeling of physical objects such as buildings, infrastructures, and their environment through the process of measuring and interpreting overlapped images. There are many well-established photogrammetry commercial software tools. We chose to use ContextCapture since this software provides an application programming interface (API) that support further extended developments, such as extracting parameters of image-orientation estimations to indicate the relative relationships between images and reconstructed 3D models (Fischer, Dosovitskiy and Brox, 2015; Verykokou *et al.*, 2018).

Photogrammetric modeling reconstructed by aerial images can support the investigation of groups of buildings in large districts. As shown in Figure 2 (a), a 3D point cloud model of some residential buildings was reconstructed by a series of aerial RGB images. To audit the heat-related defects of these residential buildings, researchers can also reconstruct a 3D thermal model. Many current approaches directly use thermal images to build thermal-mapping models. We choose to use high-resolution RGB images to reconstruct a 3D RGB model and then project corresponding thermal information onto the RGB model to create a thermal point cloud model (Hou *et al.*, 2021), as the FLIR camera can simultaneously take thermal and RGB images from the same angle and at the same altitude. Additionally, image-orientation estimations provided by ContextCapture support the data fusion process. Figure 2 (b) represents a 3D thermal model of a group of residential buildings created based on the RGB model in Figure 2 (a). In Figure 2 (b), the dark purple color represents a lower thermal value and a lighter yellow color represents a higher value. Another example is a group of 3D models on a campus shown in Figure 2 (c) and (d).



(a) Reconstructed RGB Models in City Areas

(b) Reconstructed Thermal Models in City Areas

(c) Reconstructed RGB Models on a Campus

(d) Reconstructed Thermal Models on a Campus

Figure 2: 3D point clouds reconstructed by overlapped images

2.2 Rendering 2D Images from a Reconstructed 3D model

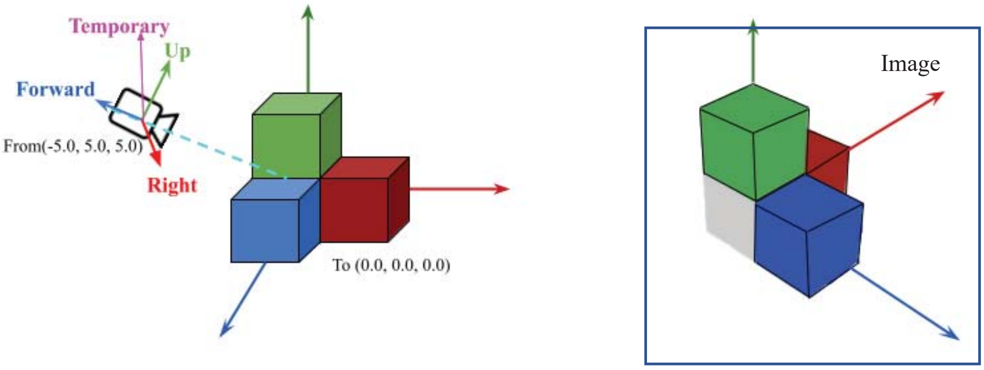
After the development of the 3D point cloud model as described in Section 2.1, the next step focus on how to use the model to audit heat loss. At this step it is important to recognize/classify door and windows elements in the model because those are the most relevant elements when auditing building façade heat loss. Therefore, in this step, we developed a process to render 2D images from the reconstructed 3D models.

We created a virtual camera in the 3D model, which was essential for rendering images that we needed to investigate. In our study, we used the perspective projection, and the default camera position was at the origin and facing the negative Z-axis. To move the camera from its origin position to a position from which the façade image can be rendered, we defined a 4x4 matrix that contains rotation and transformation information, as shown in Eq. (1).

$$\begin{bmatrix} Right_x & Right_y & Right_z & 0 \\ Up_x & Up_y & Up_z & 0 \\ Forward_x & Forward_y & Forward_z & 0 \\ T_x & T_y & T_z & 1 \end{bmatrix} \qquad \text{Eq. (1)}$$

First, we defined the $Forward$ vector. To set a camera position, the computer must know an initial point, which we refer to as the $From$ point. To know the camera's orientation, the computer must also know the point at which the camera looks. We refer to as the To point. As shown in Figure 3 (a), as an example, the $From$ point is (-5.0, 5.0, 5.0), and the To point is (0.0, 0.0, 0.0), and thus we define the $Forward$ vector as $Forward = normalize\ (From - To)$. Next, we define the $Temporary$ vector, which does not have to be precise. The typical value is (0, 0, 1). Thus, the $Right$ vector is perpendicular to the space that $Forward$ and $Temporary$ create. Finally, Cartesian coordinates are defined by three mutually perpendicular vectors, and thus we can calculate the Up vector based on the $Forward$ and $Right$ vectors. Note that $Forward, Right,$ and Up vectors are mutually perpendicular, and they are all normalized unit vectors. Therefore, a rendered image by our current camera settings can be shown in Figure 3 (b). Additionally, we need to define the transformation vector T, which is

$T = From - Origin$. Since the $Origin$ is (0, 0, 0), vector T is the coordinate of the $From$ point.



(a) The Camera Aiming at a Point (b) The Image Can Be Rendered by Such Settings

Figure 3: The Local Coordinate System of the Camera Aiming at a Point

As we have defined the 4x4 rotation and transformation matrix, we can render façade images by the given pairs of $From$ and To points. After we selected the $From$ points on streets and the To points inside of buildings, the façade images can be then rendered.

2.3 Training a Semantic Segmentation ANN Model

In this step, we used an open source database to train a segmentation ANN model based on different algorithms. This open source dataset is annotated into eight classes (e.g. Loft, Top, Wall, window, Shop, Door, and Balcony), which is available from the studies of Mathias, *et al.*, 2016 and Simon *et al.*, 2011 and can be freely downloaded from the webpage of *Ecole Centrale Paris Facades Database* (Teboul, 2008). The data contains 400 images for training and 100 images for testing. The images of facades are taken from different cities including Paris, Barcelona, and San Francisco, among others.

Many state-of-the-art ANN algorithms exist to train the segmentation models, including DeepLab, MaskRCNN, and Generative Adversarial Networks (GAN) (Goodfellow *et al.*, 2014). Among these algorithms, GAN can learn density distributions of imagery datasets and explore their internal representations (Hou, *et al.*, 2021). Additionally, as the detailed architecture of a GAN shows in Figure 4, the main difference between the GAN and other ANNs is that the GAN has two separated networks including a generator network and discriminator network; therefore, the GAN architecture is more flexible than other neural network approaches. The function of the discriminator network is to decide if the generated samples are similar to the ground truth samples, and the differences are calculated by the loss function. Further, the backpropagation improves the parameters in generator and discriminator networks based on the loss function. After several epochs, the samples generated by the generator network evolve from random noise to predicted results, and then the model is trained for use in testing datasets. As previously discussed, the GAN architecture is flexible. Thus, it is easy for us to replace the network architecture. We choose to use two different network architectures to build the generator network including “Resnet+9 blocks” and “Unet256”.

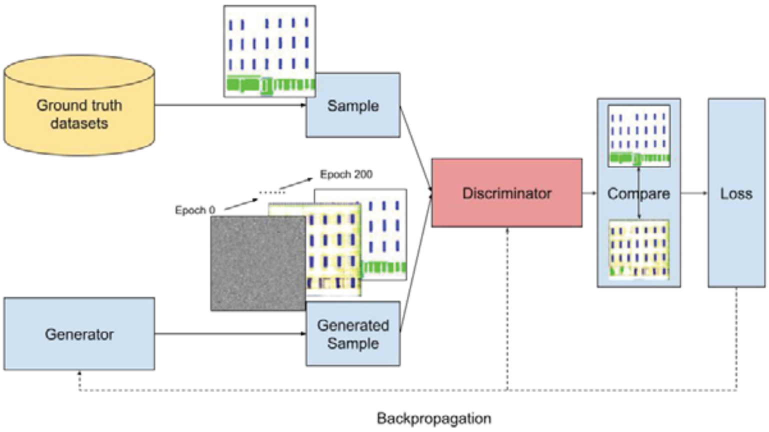


Figure 4: The Detailed Architecture of a GAN

2.4 Segmentation Results and Evaluation of the Proposed Approach

As rendering the façade images and building the semantic segmentation ANN model, we were able to use the trained model to evaluate the segmentation results of the rendered images. We applied trained ANN models (both “Resnet+9 blocks” and “Unet256” versions) on two datasets, including the campus and city areas as shown in Figure 2. As for the evaluation metrics, we chose two evaluation criteria to analyze the performance of the proposed method: (1) an accuracy analysis of the segmentation performance on the open source datasets, and (2) a performance analysis on the rendered images.

We applied four methods to evaluate the segmentation performance on images, including (1) precision, (2) recall, (3) Jaccard/intersection-over-union (IOU), and (4) the dice coefficient /F1-score, as shown in Eqs. (2-5). In these equations, TP (true positive) represents the area of overlap between the predicted segmentation and the ground truth in the images. FP (false positive) represents the areas that belong to the correct class but that the algorithms cannot recognize, and FN (false negative) represents the areas that do not belong to the correct class, but that the algorithms incorrectly recognize them do. Using TP, FP and FN, we can calculate the evaluation metrics. Precision, also known as positive predictive value, is the fraction of the correctly classified area among the actual result area in the ground truth images. Recall, also called sensitivity, is the fraction of the correctly classified pixel area among the predicted result area in the predicted images. Next, IOU, is the fraction of the correctly classified pixel area among the union areas of the actual result areas and predicted result areas. Last, F1 is a harmonic mean that combines precision and recall score.

$$Precision = \frac{TP}{TP+FP} \qquad \text{Eq. (2)}$$

$$Recall = \frac{TP}{TP+FN} \qquad \text{Eq. (3)}$$

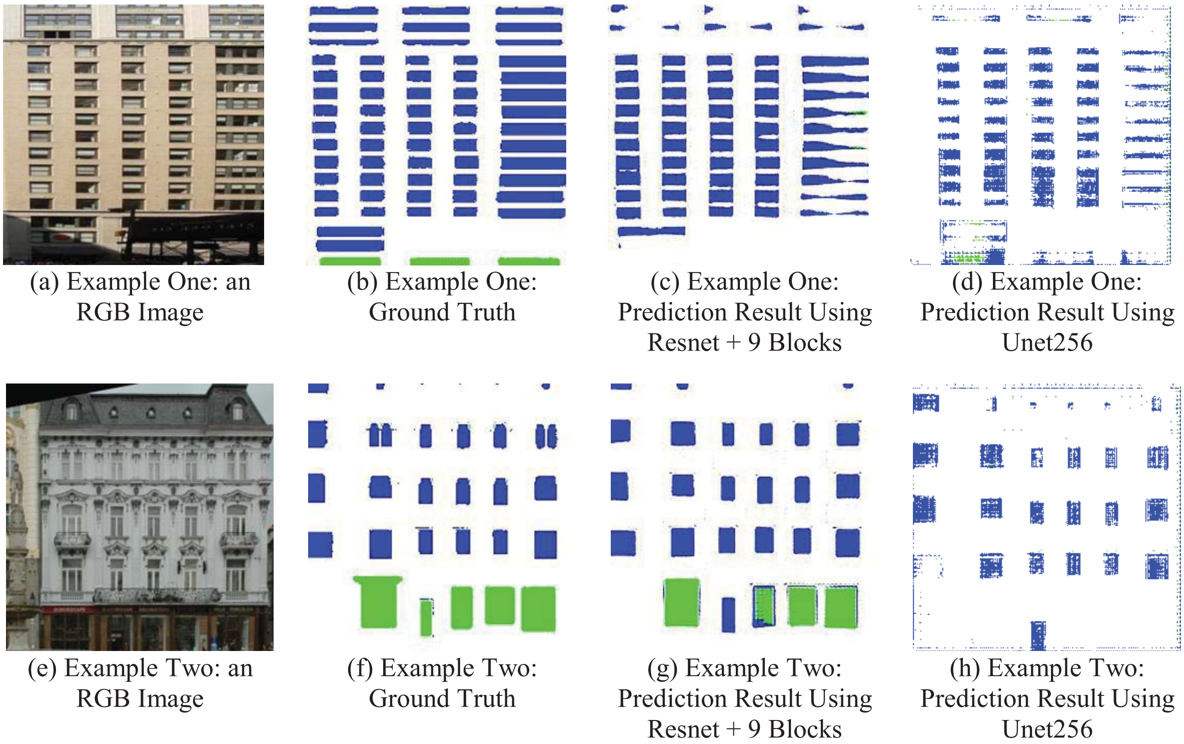
$$IOU = \frac{TP}{TP+FP+FN} \qquad \text{Eq. (4)}$$

$$F1 = \frac{2TP}{2TP+FP+FN} \qquad \text{Eq. (5)}$$

3. Experiment

Thermography inspection needs a special experimental condition in which the temperature difference between the indoors and outdoors should be at least 10 °C (18 °F) (FLIR Systems, 2011). To meet this requirement, inspections need to be conducted in a hot summer or a cold winter. However, the sun radiation can cause an inaccurate façade temperature measurement and further impact the cooling energy loss audits. Therefore, thermography inspection on hot days is usually conducted in early morning or late afternoon to avoid sun radiation. However, it is still difficult to guarantee the needed temperature differences during such inspection times in the summer. Considering these facts, we conducted a heat loss inspection on a college campus and in a city area during a cold winter in Karlsruhe, Germany. In collecting data for our experiments, room temperatures were higher (the average temperature was 17 °C (63 °F) for indoor spaces when the research was conducted), and the outside ambient temperatures were lower (the outdoor temperature was -5 °C (23°F) in the early morning).

The open source dataset in which the cameras were set on the ground is annotated into 8 classes. However, we only focused on two categories (doors and windows) related to the heat loss audits for this study. As shown in Figure 5, Figure 5 (a) and (e) are two examples in the open source datasets, (b) and (f) are ground truths for these two examples, (c) and (g) are segmentation results for these two examples, and (d) and (h) are segmentation results using another algorithm.



(a) Example One: an RGB Image

(b) Example One: Ground Truth

(c) Example One: Prediction Result Using Resnet + 9 Blocks

(d) Example One: Prediction Result Using Unet256

(e) Example Two: an RGB Image

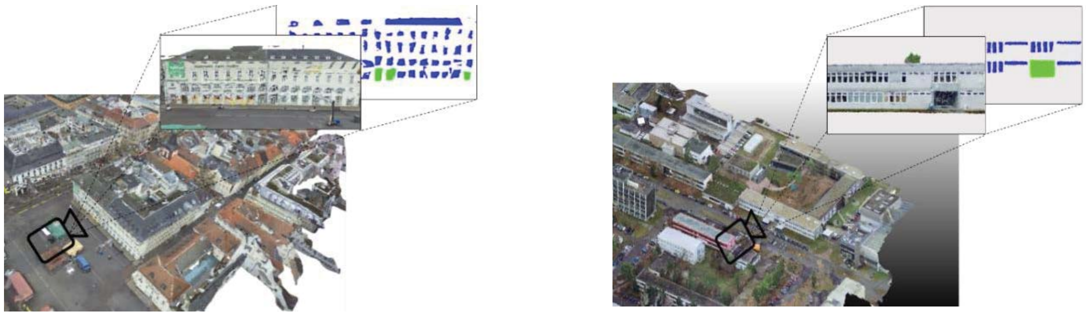
(f) Example Two: Ground Truth

(g) Example Two: Prediction Result Using Resnet + 9 Blocks

(h) Example Two: Prediction Result Using Unet256

Figure 5: Building the segmentation models

For next step, we used the two segmentation models built using "Resnet" and "Unet" to predict rendered images from the 3D point cloud models. Figure 6 (a) is an example of buildings in a city area, and Figure 6 (b) is another example for the campus buildings. A virtual camera was set in the 3D model, and a façade image with its ground truth were rendered.



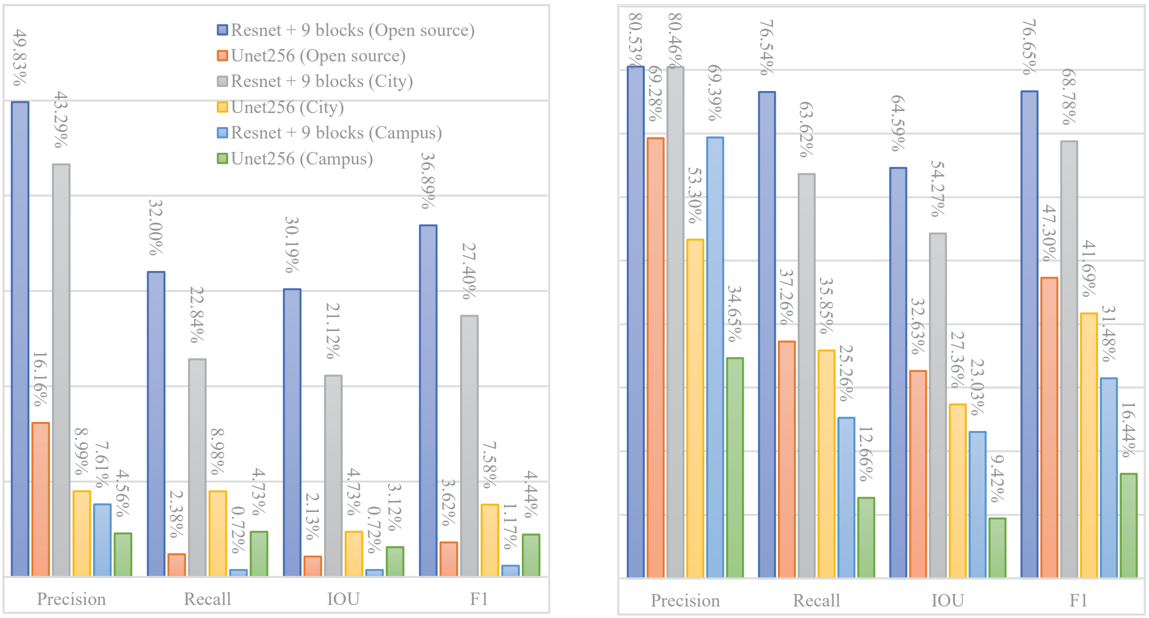
(a) Example One: an RGB Image

(b) Example Two: Ground Truth

Figure 6: Segmentation on Rendered Images

4. Results and Discussion

Based on the Eqs. (2-5), we conducted accuracy analysis of the segmentation performance for the open source datasets and performance analysis for the rendered images, as shown in Figure 7. We also used the segmentation model trained by open source datasets to predict the segmentation on rendered images, and the accuracy analyses are also shown in Figure 7.



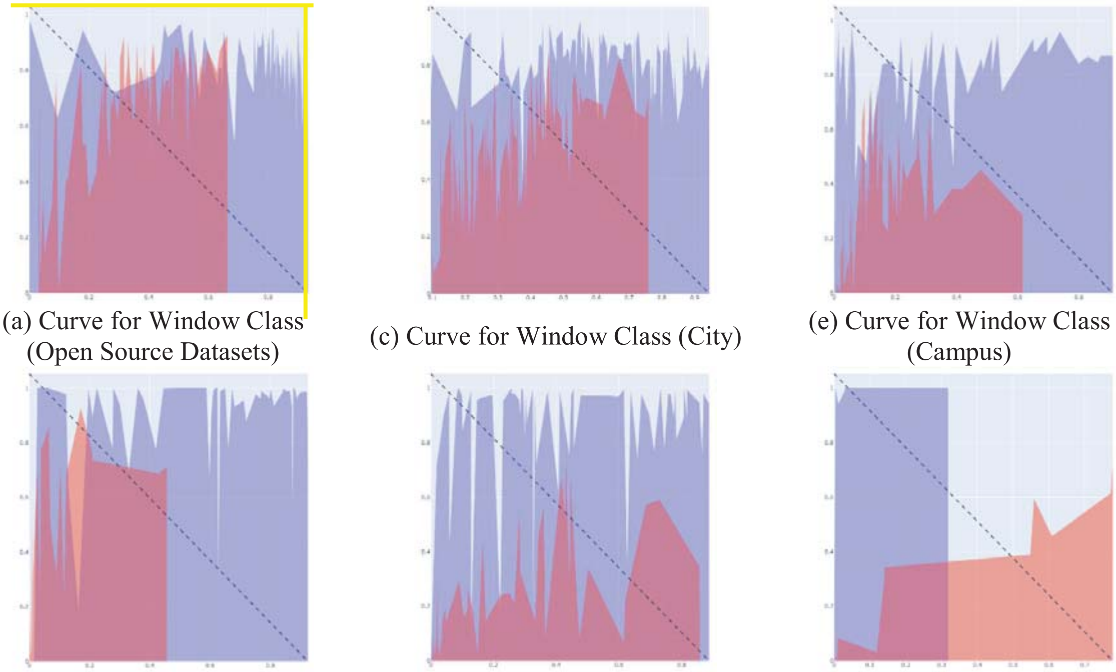
(a) Performance Analysis for Door Class

(b) Performance Analysis for Window Class

Figure 7: Segmentation Performance Analysis

We also plotted a Precision-Recall curve (PRC) as shown in Figure 8. The blue color represents "Resnet+9blocks" GAN algorithm, and red represents "Unet256" GAN algorithm. As the yellow lines shown in figure (a), the ideal test should have a PRC that passes through the upper right corner representing the 100% precision and 100% recall. In general, the closer the blue or red area is to the yellow lines, the better the performance.

There were some important findings from the results. First, as the results in Figure 7 show, "Resnet+9blocks" outperformed "Unet256" in all cases except predicting door class in rendered images from the campus datasets. Second, in general, predicting window class was more accurate than predicting door class. The blue areas are always on top of the red areas in Figure 8. This is potentially because of the unbalanced datasets. In every image in the datasets, there were more pixels belonging to window class than pixels belonging to door class. A solution needs to be found for this unbalanced dataset issue in future studies. Third, in general, our proposed approach performed better in city datasets than in campus datasets, potentially because the building styles in the open source are closer to the styles in city datasets.



(a) Curve for Window Class (Open Source Datasets)

(c) Curve for Window Class (City)

(e) Curve for Window Class (Campus)

(b) Curve for Door Class (Open Source Datasets)

(d) Curve for Door Class (City)

(f) Curve for Door Class (Campus)

Figure 8: Precision-Recall Curve

5. Conclusion and Outlook

Our results show that a 3D point cloud model can be created using aerial images and that rendered façade images for segmentation can be successfully generated by a virtual camera in the model. As the results show, the segmentation accuracy decreases from the evaluation of the segmentation performance on the open source datasets to the evaluation of the rendered images. Particularly, the performance decreases more when using the "Unet256" algorithm. Second, the accuracy of segmenting windows is higher than segmenting doors. Finally, the results show that the accuracy of semantic segmentation is higher when the approach is conducted on buildings in a city than in a university campus. In the future, there is a need to consider the unbalanced dataset issue related to the higher incidence of windows objects when compared to door objects on existing databases. Additionally, there are two options for improving the segmentation performances; one is by improving the quality of the rendered images, and the other one is by improving the segmentation algorithms.

Acknowledgement

The authors thank the CSC (China Scholarship Council), the KIT International Department, and DAAD (Deutscher Akademischer Austauschdienst, German Academic Exchange Service) for their support. Furthermore, the authors thank Marinus Vogl and Air Bavarian GmbH, his drone company, for their support with equipment and service. Last, the authors thank Leslie Ramos, Wenyun Lu, and Tianyou Dong for their time and effort in ground truth coding.

References

Bradski, G. (2000) ‘The OpenCV Library’, Dr. Dobb’s Journal of Software Tools.

Chen, M. et al. (2020) ‘Semantic Segmentation and Data Fusion of Microsoft Bing 3D Cities and Small UAV-based Photogrammetric Data’, (20220), pp.1–12. doi: arXiv:2008.09648v1.

Dino, I. G. et al. (2020) ‘Image-based construction of building energy models using computer vision’, Automation in Construction, 116(1). doi: 10.1016/j.autcon.2020.103231.

Fischer, P., Dosovitskiy, A. and Brox, T. (2015) ‘Image orientation estimation with convolutional networks’, German Conference on Pattern Recognition, pp.368–378. doi: 10.1007/978-3-319-24947-6_30.

FLIR Systems (2011) An informative guide for the use of thermal imaging cameras for inspecting buildings, solar panels and windmills. Thermal Image Guidebook for Building and Renewable Energy Applications Content. Available at: http://www.flirmedia.com/MMC/THG/Brochures/T820325/T820325_EN.pdf (Accessed: 13 June 2020).

Goodfellow, I. et al. (2014) ‘Generative Adversarial Nets’, Advances in neural information processing systems, pp.2672–2680. doi: 10.1109/ICCVW.2019.00369.

Hou, Y. et al. (2019) ‘Factors affecting the performance of 3D thermal mapping for energy audits in a district by using infrared thermography (IRT) mounted on unmanned aircraft systems (UAS)’, in Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019, pp.266–273. doi: https://doi.org/10.22260/ISARC2019/0036.

Hou, Y. et al. (2021) ‘Automation in Construction Fusing tie points’ RGB and thermal information for mapping large areas based on aerial images : A study of fusion performance under different flight configurations and experimental conditions’, Automation in Construction. Elsevier B.V., 124. doi: 10.1016/j.autcon.2021.103554.

Hou, Y., Volk, R. and Soibelman, L. (2021) ‘A Novel Building Temperature Simulation Approach Driven by Expanding Semantic Segmentation Training Datasets with Synthetic Aerial Thermal Images’, Energies, 14(2). doi: https://doi.org/10.3390/en14020353.

Mathias, M., Martinović, A. and Van Gool, L. (2016) ‘ATLAS: A Three-Layered Approach to Facade Parsing’, International Journal of Computer Vision, 118(1), pp.22–48. doi: 10.1007/s11263-015-0868-z.

Matl, M., Mahler, J. and Goldberg, K. (2017) ‘An algorithm for transferring parallel-jaw grasps between 3D mesh subsegments’, IEEE International Conference on Automation Science and Engineering, 2017-Augus, pp.756–763. doi: 10.1109/COASE.2017.8256195.

Mayer, Z. et al. (2021) ‘Thermal Bridges on Building Rooftops - Hyperspectral (RGB + Thermal + Height) drone images of Karlsruhe, Germany, with thermal bridge annotations (Version 0.1.0)’, in Zenodo. Zenodo. doi: http://doi.org/10.5281/zenodo.4767772.

Pedregosa, F. et al. (2019) ‘Generating the blood exposome database using a comprehensive text mining and database fusion approach’, Environmental Health Perspectives, 127(9), pp.2825–2830. doi: 10.1289/EHP4713.

Shahandashti, S. M. et al. (2010) ‘Data-Fusion Approaches and Applications for Construction Engineering’, Journal of Construction Engineering and Management, 137(10), pp.863–869. doi: 10.1061/(ASCE)CO.1943-7862.0000287.

Shi, Z. and Ergan, S. (2020) ‘Towards Point Cloud and Model-Based Urban Façade Inspection: Challenges in the Urban Façade Inspection Process’, Construction Research Congress 2020, p. 385. doi: https://doi.org/10.1061/9780784482872.042.

Simon, L. et al. (2011) ‘Random Exploration of the Procedural Space for Single-View’, International Journal of Computer Vision, 93(2), pp.253–271. doi: https://doi.org/10.1007/s11263-010-0370-6.

Teboul, O. (2008) Ecole Centrale Paris Facades Database. doi: http://vision.mas.ecp.fr/Personnel/teboul/data.php.

Verykokou, S. et al. (2018) ‘A Photogrammetry-Based Structure From Motion Algorithm Using Robust Iterative Boundle Adjustment Techniques’, ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, IV(October), pp.1–2. doi: https://doi.org/10.5194/isprs-annals-IV-4-W6-73-2018.

Yang, M. Der, Su, T. C. and Lin, H. Y. (2018) ‘Fusion of infrared thermal image and visible image for 3D thermal model reconstruction using smartphone sensors’, Sensors (Switzerland), 18(7). doi: 10.3390/s18072003.