SURVEY OF BUILT ENVIRONMENT IN THE ERA OF UAV

From Aerial Photogrammetry to Point Cloud Classification

YONGKANG LIU[1] and YI WANG[2]
[1,2]*School of Architecture, Tsinghua University.*
[1]*liuyk19@mails.tsinghua.edu.cn, 0009-0002-1984-0678.*
[2]*wangyi@tsinghua.edu.cn, 0009-0008-8130-8942. (Contact Author)*

Abstract. In order to further discover the potentials of UAV (Unmanned Aerial Vehicle) for built environment research, this article involves in drone aerial survey and its post-processing, with a special focus on point cloud classification. By operating UAV flying over villages at foot of Mount Tai, capturing images of the villages as first-hand materials, and conducting research with the help of 3D model reconstruction software, deep learning implements, GIS environment, the findings of research response the questions of the relationship between flight altitude, working efficiency, and 3D reconstruction quality, and how to utilize the deep learning tools for certain building classification. The solution to the second problem, also the most noteworthy contribution of this article, is achieved by training a customized point cloud classification model. This model can be used to identify point clouds of specific types of buildings, which is an advancement compared to the basic Automated Classification in ArcGIS Pro. The quality of point cloud recognition is also better than the latter. Potential application of this research could be reflected in the statistical work for certain types of buildings. In other words, this study plays an intermediary role between UAV-aided image gathering to further spatial statistical research.

Keywords. UAV-aided Survey, Aerial Photogrammetry, Customized Point Cloud Classification, Deep Learning.

1. Background, Related Study, and Research Question

1.1. BACKGROUND

In architectural and built environment study, it is important to investigate the field in detail, comprehensively analysing social, economic and spatial features, especially the studies related to renewal and urbanization, in which the understanding of the original built environment is emphasized. The traditional survey of built environment is typically relied on door-to-door questionnaires, or manual site measuring and drawing. These works are complex and inefficient. Thanks to UAV, also known as drone, the

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 2, 149-158.

efficiency of survey has been greatly improved. The application of UAV in architectural investigation is gradually becoming a hot topic for many scholars.

Equally important as drone technology is how the images and data it captures can be transformed into architectural and planning research. Regardless of more guidance needed for improving working efficiency for aerial photogrammetry, including UAV field operation and preliminary post-processing (includes aerotriangulation, 2D and 3D reconstruction， abbreviated as PrePP), the point cloud classification, a key step before further statistical analysis, still remains at the stage of conventional classification, decomposing 3D scene into buildings, roads and trees, etc. However, there are few related studies on further applying point cloud classification to the identification of specific building types.

1.2. LITERATURE REVIEW

Related Techniques of 2D and 3D remodelling using UAV images as raw materials are relatively mature. Scholars who had compared the performance between different UAV image processing software provided valuable experience of image processing software for building volume estimation (Ajayi, et al., 2023).

Deep learning is widespread in UAV data processing and built environment related study, no matter the 2D imagery extraction as input data for further GIS analysis in a case study of 4 neighbourhoods in Northeast Ohio (Hong, et al., 2022), or the 3D point cloud classification, for building area statistics in rural research (Zhang, 2020) and the architectural semantics study of cultural heritage (Massimiliano, et al., 2022). Classification of point cloud, distinguishing the road, building, and tree in a certain scene (Rau, et al., 2015) has become a frequently-used method.

Further studies, about algorithms and technical improvements of deep learning have made progress, with characteristics of photogrammetric data taken into consideration (Chen, et al., 2020). The approach for 3D class recognition from point cloud of deformed building elements has been proposed, with which the 3D point cloud classification is robust to deformation (Chen, et al., 2018). Super-point, the small cluster of point cloud, is treated as a basic operational unit to achieve the robustness to noise and outliers. Through a feature selection process according to a ranking of importance of point cloud features, the features with less contribution to classification task can be removed (Wang, et al., 2020). Another team's effort of ameliorating deep learning is the conversion of the laser scanner point cloud to a voxel data structure, which dramatically reduces the amount of data for processing. (Babahajiani, et al., 2015). Currently, the well-known and commonly used neural network of point cloud classification include RandLA-Net (Hu, et al., 2020), PointCNN (Li, et al., 2018) and SQN (Hu, et al., 2023). Though there are also some other famous model architectures, these three are favoured and have been adopted by ArcGIS Pro 3.2.

1.3. QUESTION

Regardless of many scholars having stepped into this field, further research is still need. In this paper, two issues are focused, namely, what is the relationship between flight altitude, working efficiency, and 3D reconstruction quality? And is it possible to apply point cloud classification to identifying a certain type of building?

2. Aerial Photogrammetry

2.1. TESTING OF ACCURACY AND EFFICIENCY

The tests of accuracy were conducted at 3 different flight heights: 45 meters, 90 meters and 135 meters with drone DJI M3E at villages around Mount Tai, China. The software DJI Terra was used for 3D reconstruction. Digital models constructed from drone images taken at a height of 45 meters clearly shows the building materials or the Chinese characters of the signboard. In 135 meters ones, though hard to recognize the details, it is still able to roughly judge the volume, type and texture of the building, achieving a general level of accuracy. (Figure 1)

Figure 1. The first row of images from left to right: Aerial Photogrammetry in 45m, 90m and 135m above the ground, showing building materials; The second row of images from left to right: Aerial Photogrammetry in 45m, 90m and 135m above the ground, showing store sign

Testing of efficiency includes flight efficiency tests and PrePP efficiency tests. The flight efficiency tests selected a piece of land of 100 meters by 100 meters, with the following-ground flight mode (Flight height would vary following the change of the terrain). Other flight parameters and settings include 45-degree tilt photography mode, 80% of heading overlap rate, 70% of side overlap rate, 15 m/s of maximum flight speed, and approximately 341 meters between the taking off point and the geometric centre of the surveying area. 7 times flights were conducted at the height ranging 45 meters to 180 meters relative to the ground. A computer of Nvidia Quadro P4200 and i7-8750H carried out the PrePP (includes aerotriangulation, 2D and 3D reconstruction) with software DJI Terra. It can be judged from Table 1 and Figure 2 that the duration of flight decreases at first and then increases with the raise of altitude, while the time spent on PrePP is continuously decreasing as the altitude increases. A possible explanation is, the higher the altitude, the wider the field of vision, the fewer photos needed to be taken, leading to both flight route and period shortened. However, as the flight altitude further increases, the drone needs to rise to a higher altitude for measuring, not to mention larger flight range brought by tilt photography mode, therefore inevitably provoking the curve to rise. On the other hand, since duration of PrePP is mainly judged by the number of photos, it will certainly decrease as the flight altitude rises. Based on the models' quality of 90 meters and 135 meters (Figure 1) and the fact that the total time taken is no longer shortened from the test in 112.5 meters height (Table 1), it can be concluded that for a 100 metres by 100 metres square site, if only general model quality is needed, the flight altitude of about 112.5 meters will be achieve the highest efficiency without ignoring a certain level of quality, under equipment of drone DJI M3E, a computer of Nvidia Quadro P4200 video card and i7-8750H CPU.

Table 1. Result of efficiency test at villages around Mount Tai

Flight Height	Flight Duration	Capture	PrePP Duration	Total Duration
45m	7min	60	33min	40
67.5m	6min	57	21min	27
90m	5 min	42	11min	16
112.5m	4 min	20	5min	9
135m	5 min	17	4min	9
157.5m	5 min	15	4min	9
180m	6 min	15	3min	9

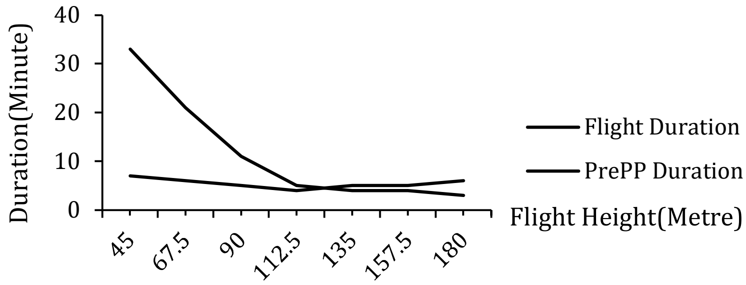


Figure 2. Flight and PrePP duration curves drawing according to Table 1

2.2. A CASE OF AERIAL PHOTOGRAMMETRY: BAIMASHI VILLAGE

Here is a case of aerial photogrammetry. Figure 3 left shows 3D model of Baimashi, a Chinese village at foot of Mount Tai, whose digital model generated by UAV-collected images and software DJI Terra. 3D model presents more information than the 2D images, a benefit for better identifying features of architecture and environment. Figure 3 right are several typical forms of built environment in Baimashi, namely, The reuse of former long buildings (①), The gradually updated cottage area (②), High quality housing built by locals themselves (③), Early built relocation dwellings of medium quality (④), The natural dominated area with sparse houses (⑤), and Recently constructed urban residences with high quality (⑥).

Figure 3. 3D model of Baimashi Village generated by UAV photos and DJI Terra

3. Point Cloud Classification: Case Study of Baimashi

3.1. STANDARD POINT CLOUD CLASSIFICATION

Point cloud of Baimashi is obtained through aerial photogrammetry, generated by software DJI Terra. Preliminarily, the standard classification using ready-made Automated Classification commands in ArcGIS Pro is tried for building identification. However, function of the standard classification command is limited according to the test in Baimashi Village case (Figure 4). The "Classify Buildings" in Automated Classification commands is able to identify buildings, but unable to further recognize a certain type of building, not to mention the uneven quality and incompleteness of point cloud recognition. Therefore, a customized point cloud classification model is worth trying, looking forward to going beyond the very basic function.

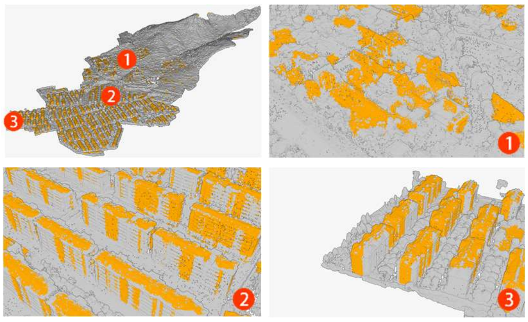


Figure 4. Automated Classification by ArcGIS Pro in the case of Baimashi Village

3.2. CUSTOMIZED POINT CLOUD CLASSIFICATION

The customized point cloud classification starts with labelling of the samples, distinguishing the expected certain buildings from others, both in training set and validating set. These labelled samples have to be transformed into trainable data format "pctd" before putting into model training tool. While the training process is completed, the newly trained model can be used for target point cloud classification (Figure 5).

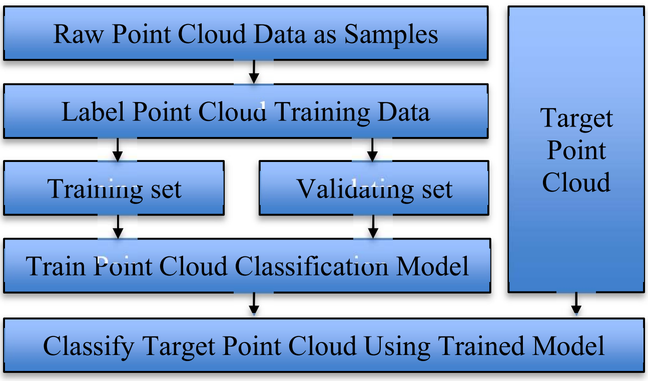


Figure 5. Steps of customized point cloud classification

Here, we take point clouds of villages at the foot of Mount Tai as the samples for discussing. RandLA-Net (Hu, et al., 2020) is adopted as model architecture. The training samples are selected from Sanhe and Matao since these two villages contain typical samples of (single layer) bungalows and (multi-story) residential buildings. Samples of T1 and V1 are from Matao, while T2 and V2 are from Sanhe (Figure 6). The labelling work usually starts from the 2D view, and then makes modifications, if necessary, on 3D view (Figure 7). Number of buildings labelled is showed in table 2. These labelled samples need to be converted to a "pctd" file subsequently.

In model training stage, F1 score is set as objective of the model optimization. The learning rate strategy is One Cycle Learning Rate: the learning rate will automatically adjust and optimize based on "1cycle" technique in FAST.AI's implementation (Smith, 2017; Smith, 2018). The training is set to halt automatically when the model is no longer improved or reaches the given limitation of 25 rounds. The curves in Figure 8 represent the training process. From the training statistics showed by Table 3, Table 4 and Table 5, it can be seen that train only one type of building at a time is better than train both, although the former is also not ideal, especially when training bungalows. Values close to zero appeared in bungalow and residence of Table 5 indicate the failure when training both of bungalow and residence at the same time, presenting somehow the mutual exclusiveness of the two building types.

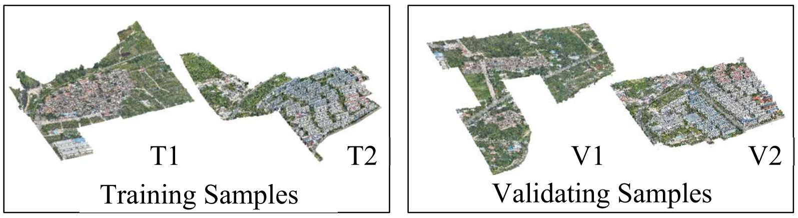


Figure 6. Samples were selected from Matao and Sanhe around Mount Tai

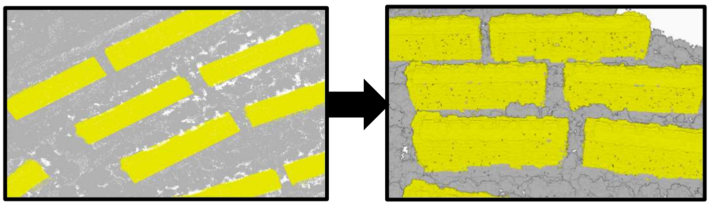
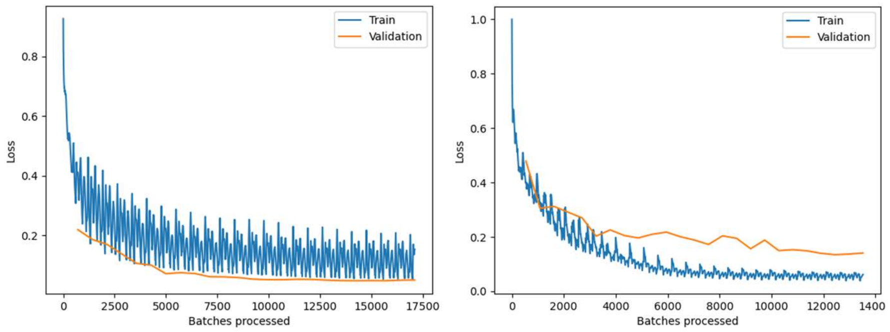


Figure 7. Samples show labelling process, firstly label on 2D view, then fixed on 3D model

Table 2. Labelling class and number

CLASS	DESCRIPTION	TRAINING	VALIDTAING
Residence	Multi-story building for relocated villagers or newly settled citizens.	51	40
Bungalow	The local dwelling of the villagers, mostly only one floor.	313	170



(a) Individually train the bungalow *(b) Individually train the residence*

Figure 8. Training curve generated by integrated module in ArcGIS PRO

Table 3. Individually train the bungalow, labelled samples of residence are set as background

EPOCH	CLASS	PRECISION	RECALL	F1 SCORE
0	Background	0.999775	0.97894	0.988984
0	Bungalow	7.77E-05	0.034653	0.000154
…	…	…	…	…
23	Background	0.996445	0.987446	0.991885
23	Bungalow	0.193959	0.418373	0.24676

Table 4. Individually train the residence, labelled samples of bungalow are set as background

EPOCH	CLASS	PRECISION	RECALL	F1_SCORE
0	Background	0.802683	0.967173	0.872591
0	residence	0.303915	0.225495	0.245674
…	…	…	…	…
24	Background	0.963999	0.989032	0.976031
24	Residence	0.346165	0.317636	0.328948

Table 5. Train both of bungalow and residence at the same time

EPOCH	CLASS	PRECISION	RECALL	F1_SCORE
0	Background	0.997164	0.87867	0.927582
0	Bungalow	0.00205	0.054263	0.003858
0	Residence	0.00012	0.238652	0.000162
…	…	…	…	…
8	Background	0.990581	0.879291	0.925084
8	Bungalow	0.036478	0.115565	0.043238
8	Residence	1.04E-05	0.123799	2.09E-05

Due to the poor effect of the previous training, improving the training results is the next key work. This round of improvement is merely aimed at residence (multi-story, same as above) identification, and the main improvement measures are to increase samples. In the new training, samples are selected from Shangyu, Aiwa and Sanhe. The numbers of labelled residence in training set and validating set are increased to 82 and 88, respectively (Table 6). After retraining, the F1 score of model has been improved from 0.33 below to 0.42 above (Table 7). Figure 9 shows the classification result of Baimashi residence using the newly trained model. It can be seen that training a point cloud recognition model specially focused on multi-layer residential building is possible. It is also worth mentioning that the quality of point cloud classification showed in partially enlarged image of Figure 9 is much higher than the standard categorization available in ArcGIS Pro in Figure 5.

The reason for the poor performance of individually training of bungalows could be, at least, the negative influence of cluttered environment in 3D model, bringing difficulty to label the bungalows with high quality. However, deep learning network for 2D image data such as U-Net architecture and ResNet (Ronneberger, et al., 2015; He, et al., 2015) have become very popular, potentially serving as a supplement in case of which 3D point cloud data is hard to label and train.

Table 6. Labelling number of residential buildings in the second rounds training

CLASS	TRAINING	VALIDTAING
Residence	82	88

Table 7. Improved individually training of residence

EPOCH	CLASS	PRECISION	RECALL	F1_SCORE
13	Background	0.991473	0.879944	0.926526
13	Residence	0.329407	0.720755	0.424353

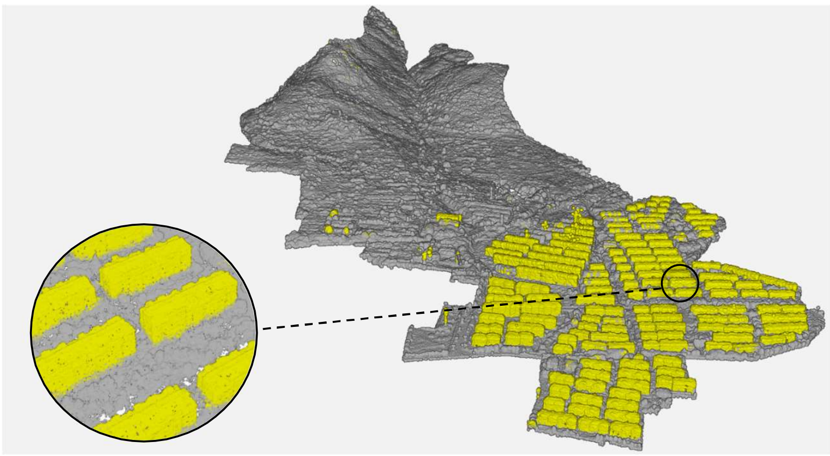


Figure 9. Classification of Baimashi residence using the improved model

3.3. IMPLICATIONS FOR FURTHER BUILT ENVIRONMENT STUDY

Identifying a specific type of building through training a customized point cloud classification model goes one step further than the Automated Classification in ArcGIS Pro. Compared to 2D images, 3D point cloud data reflects morphological features of a building, letting it to be more accurate than 2D image deep learning. The exploration is valuable for statistics work of specific types of buildings. For example, in GIS environment, by combining the feature of extracting 2D building contour based on point cloud classification results and subsequently extracting geometric centre from 2D contour, each target building can be abstracted as a point, and the density feature of the point can be regarded as one of the features of the built environment, serving as an indicator in quantitative research. However, limited by the length of this article, these further explorations will be reflected in future research.

4. Conclusion

For aerial photogrammetry of 100 metres by 100 metres square site, if only general model quality is needed, the flight altitude of about 112.5 meters will achieve the highest efficiency with equipment of drone DJI M3E, a computer of Nvidia Quadro P4200 video card and i7-8750H CPU, without neglecting a certain level of quality. Certainly, the most noteworthy contribution of the paper is the customized point cloud classification. To train a customized point cloud classification model for identifying a certain type of building, as research has proven, is workable. This is a functional advance compared with Automated Classification in ArcGIS Pro, which can only roughly identify non-specific type of buildings. While the next problem is, each training can only target at one specific type of building. The attempt to train a model for simultaneously classifying two or more kinds of buildings is temporarily unsuccessful. This issue will be studied in the future.

Acknowledgement and Illustrations

Thanks to Tai'an Natural Resources and Planning Bureau, and the Statistics Bureau of Shandong Province, for providing geographic or administrative boundaries of Baimashi and some other villages, which have facilitated this study. Except for the contributions mentioned as acknowledgement above, all images and tables in this paper are drawn by the authors.

References

Ajayi, O. G., Ogundele, B. S., & Aleji, G. A. (2023). Performance evaluation of different selected UAV image processing software on building volume estimation. Advances in Geodesy & Geoinformation, 72(1), 1–17. https://doi.org/10.24425/agg.2023.144591.

Babahajiani, P., Fan, L., Kamarainen, J., & Gabbouj, M. (2015). Automated super-voxel based features classification of urban environments by integrating 3D point cloud and image content. In 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia. (pp. 372-377). https://doi.org/10.1109/ICSIPA.2015.7412219.

Chen, M., Feng, A., McAlinden, R, &Soibelman, L. (2020). Photogrammetric point cloud segmentation and object information extraction for creating virtual environments and simulations. Journal of Management in Engineering. 36(2), 04019046.

https://doi.org/10.1061/(ASCE)ME.1943-5479.0000737.
Chen, J., Cho, Y. K., &Ueda, J. (2018). Sampled-point network for classification of deformed building element point clouds. In 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia. (pp. 2164-2169). https://doi.org/10.1109/ICRA.2018.8461095.
Hong, X., Sheridan, S. & Li, D. (2022). Mapping built environments from UAV imagery: a tutorial on mixed methods of deep learning and GIS. Computational Urban Science. 2(1), 12. https://doi.org/10.1007/s43762-022-00039-w.
Hu, Q., Yang, B., Fang, G., Guo, Y., Leonardis, A., Trigoni, N., &Markham, A. (2023). SQN: Weakly-supervised semantic segmentation of large-scale 3D point clouds. arXiv:2104.04891 [cs.CV]. https://doi.org/10.48550/arXiv.2104.04891.
Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., &Markham, A. (2020). RandLA-Net: Efficient semantic segmentation of large-scale point clouds. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, (pp. 11105-11114). https://doi.org/10.1109/CVPR42600.2020.01112.
Kaiming He, Xiangyu Zhang, Shaoqing Ren, &Jian Sun. (2015). Deep residual learning for image recognition. arXiv:1512.03385 [cs.CV]. https://doi.org/10.48550/arXiv.1512.03385.
Li, Y., Bu, R., Sun, M., Wu, W., Di, X., &Chen, B. (2018). PointCNN: Convolution on X-transformed points. In Advances in Neural Information Processing Systems 31. https://proceedings.neurips.cc/paper_files/paper/2018/file/f5f8590cd58a54e94377e6ae2eded4d9-Paper.pdf.
Massimiliano, P., Vincenzo S. A., Domenica, C., &Daniele, S. (2022). Data for 3D reconstruction and point cloud classification using machine learning in cultural heritage environment. Data in Brief, 42, 108250, ISSN 2352-3409. https://doi.org/10.1016/j.dib.2022.108250.
Rau, J. Y., Jhan, J. P., &Hsu, Y. C. (2015). Analysis of oblique aerial images for land cover and point cloud classification in an urban environment. In IEEE Transactions on Geoscience and Remote Sensing, 53(3), 1304-1319. https://doi.org/10.1109/TGRS.2014.2337658.
Ronneberger, O., Fischer, P., &Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351, (pp. 234–241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28.
Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay. arXiv:1803.09820 [cs.LG]. https://doi.org/10.48550/arXiv.1803.09820.
Smith, L. N. (2017). Cyclical learning rates for training neural networks. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA. (pp. 464-472). https://doi.org/10.1109/WACV.2017.58.
Wang, Y., Jiang, T., Yu, M., Tao, S., Sun, J., &Liu, S. (2020). Semantic-based building extraction from LiDAR point clouds using contexts and optimization in complex environment. Sensors. 20(12), 3386. https://doi.org/10.3390/s20123386.
Zhang, X. (2020). Village-level homestead and building floor area estimates based on UAV imagery and U-Net algorithm. In ISPRS International Journal of Geo-Information. 9(6), 403. https://doi.org/10.3390/ijgi9060403.