Fusion of Stereo Aerial Images and Official Surveying Data for Mapping Curbstones Using AI

Jiaojiao Tian
Remote Sensing Technology Institute
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
jiaojiao.tian@dlr.de

Xiangyu Zhuo
Remote Sensing Technology Institute
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
xiangyu.zhuo@dlr.de

Stefan Auer
Remote Sensing Technology Institute
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
stefan.auer@dlr.de

Franz Kurz
Remote Sensing Technology Institute
German Aerospace Center, DLR
Oberpfaffenhofen, Germany
franz.kurz@dlr.de

Peter Reinartz
Remote Sensing Technology Institute
German Aerospace Center (DLR)
Oberpfaffenhofen, Germany
peter.reinartz@dlr.de

***Abstract*—Semantic segmentation and object extraction from aerial images have made tremendous progress along with the evolution of deep learning neural network architectures. However, collecting high quality training data is still the bottleneck for many applications, in terms of costs and limited visibility of small objects. Conducting the training of artificial intelligence (AI) with accessible and adapted official data reduces the effort and allows to integrate independent in-situ knowledge as reference. Focusing on a prominent road element as example, this work presents a new approach for detecting curbstones from airborne stereo images with the assistance of official surveying data. To adapt the curbstone maps to the oblique view images, reference information is removed in occluded regions. The refined reference masks are fused with airborne imagery and integrated to the training of a Swin transformer segmentation model. In the end, the curbstone segments are transformed into vectors using an advanced vectorization approach. The proposed approach is tested over the city area of Brunswick, Germany.**

***Index Terms*—curbstone, deep learning, segmentation, data fusion, GIS**

I. Introduction

The task of curbstone detection is highly relevant for traffic applications [1], [2]. Curbstone is a typical companion of roads in suburban and urban regions, marking the border of roads. Curbstones are usually the connecting element between the drive-able road and slightly elevated pavement next to it. Due to the related height step, curbstones can be identified with different sensor modals in cars, e.g. optical images and radar images [3], [4]. Hence, identified curbstone lines can be used as a connecting element between observations from above, providing orientation and context, and information collected in the car, e.g. to keep vehicles on track. There are further relevant classes along roads to consider, e.g. fences or building corners, but we focus on curbstone as a prominent road element in this example.

Collecting enough reference data for training a deep learning neural network is critical in order to make the network generalize enough. The effort for manually annotating image data is costly in terms of time and money. Furthermore, the annotation of images is limited by human interpretation and spatial resolution of the images, which leads to errors. For instance, annotating curbstones from airborne imagery is a quite critical task, as in many cases, they are hardly visible. 3D street view images are necessary for the annotation, which leads to high expenses and required time for the annotation task. Fig. 1 displays two curbstone examples. Fig. 1 (a) is a regular curbstone, which appears as a separation of main road surface and the adjacent pavement (mainly sidewalks). Fig. 1 (b) is served as a base of a guardrail / fence along the road. Without the street view images, it is generally challenging to recognise the existence of curbstones.

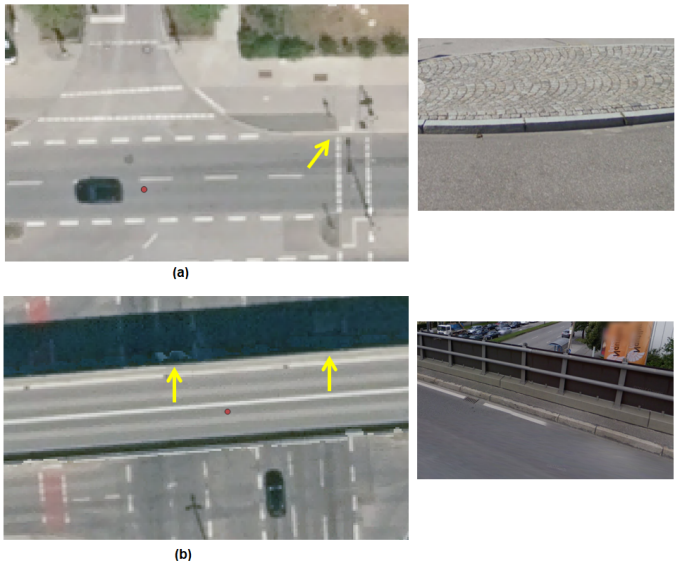

Fig. 1. Areas with required curbstone annotation (examples marked with yellow arrows). Top: curbstone next to a pedestrian way, bottom: curbstone next to the fence on a bridge.

On the other hand, collected in-situ reference data are more complete and spatially accurate. Exchanging manual annotation with official reference data, as shown in this work, saves time and integrates knowledge that is independent from the image data to be analyzed. Besides strong advantages of efficiency, accuracy and independence, there are also drawbacks.

This paper addresses advantages and challenges of fusing optical image data and independent reference data in the context of an artificial intelligence (AI) task, with the specific goal of identifying curbstones as road elements in aerial images.

Specifically, a scenario in the context of relevant traffic applications is considered. Image analysis is focused on the detection of curbstone as road element. Image data are acquired with an airborne optical sensor, operated by the German Aerospace Center (DLR) [5]. Independent official data, documenting the position of curbstone based on terrestrial surveying, is integrated as reference in the training stage (ideal output of deep learning method). Data fusion is conducted in the training of the deep learning based segmentation model, where sensor images move in as input to be processed and known curbstone positions are compared with the output for optimizing the processing parameters of the neural network. To avoid label noise, necessary steps for appropriate preprocessing of the reference data are implemented and explained.

II. Data Description

The test region is located in Brunswick, Germany. In this test region, we have acquired aerial images with the DLR 3K camera system on 26^{th} April 2019. The aerial images from two side-looking cameras with each size 3744×5616 pixel have together a field of view of 26° along track and 80° across track. In combination with a flight height of $650m$ and focal length of $50mm$ the ground sampling distance results in around $7cm$. From altogether 1758 images covering the city area of Brunswick, 80 images from representative areas are selected and georeferenced using ground control points and onboard measured global navigation satellite system (GNSS)/inertial data. All data are feed in to a bundle adjustment which delivers precise image positions and attitudes with an expected point accuracy of better than $10cm$.

The curbstone vector data are provided by the national surveying agency in vector format as 2D polygons with attributes. One important attribute in this context is the information about the relative height of the curbstone separated in low and high. Additionally, the absolute height information for each point is derived from a digital terrain model, which is also provided from the national surveying agency. We assume further that low curbstones have the same height like the terrain whereas high curbstone have a relative height of $10 - 20cm$.

Besides the time differences, as shown in Fig. 2, the curbstones are partly hidden by building rooftops, tree branches or parked vehicle. The reference data in these regions will be handled as outliers.

Fig. 2. Reference data for curbstone (yellow) mapped on aerial image.

III. Methodology

The methodology for curbstone identification consists of three steps. The first part is the reference data preparation by using height information and a ray-tracing approach. After further dilation, the curbstone mask is employed as reference for training single class object segmentation which allows to extract curbstone pixels from the images. The third part is concerned with the vectorization of curbstone segments, resulting in curbstone polylines as vector data.

A. Curbstone Map Refinement

As illustrated in Section I, the curbstone masks are partly occluded by high-level objects such as trees in the aerial images. In addition, due to distortions in the original oblique view aerial images, curbstone reference data are often positioned at the top of building roofs. Therefore, to decide whether a curbstone pixel is visible in the aerial images, as shown in Fig. 3, we separate curbstone pixels along the height axis into high- and low-level objects. Afterward, we identify occluded parts from the oblique view airborne images.

1) High- and low-level object detection: We firstly generate the digital surface model (DSM) from the 3K stereo dataset. Image pre-orientation is performed by using open source Shuttle Radar Topography Mission (SRTM) data. GPS positions of the image projection centers are measured, precise image orientation then accomplished by bundle adjustment using automatically extracted Scale-Invariant Feature Transform (SIFT) tie points [5]. Afterward, a 3D point cloud is calculated using semi-global matching [6]–[8]. In the end, the delta surface fill algorithm [9] is performed to close areas without points with SRTM heights and generate the final DSM.

To obtain the absolute height of the landcover objects, we generate the normalized DSM (nDSM), representing only elevated objects in the scene, and digital terrain model (DTM) using morphological top-hat reconstruction as mentioned in [10]. With the height information from the nDSM, we category pixels in the curbstone masks into high- and low-level objects. The high-level curbstones are invisible in the images due to occlusion at trees. They are therefore treated as outliers and removed.

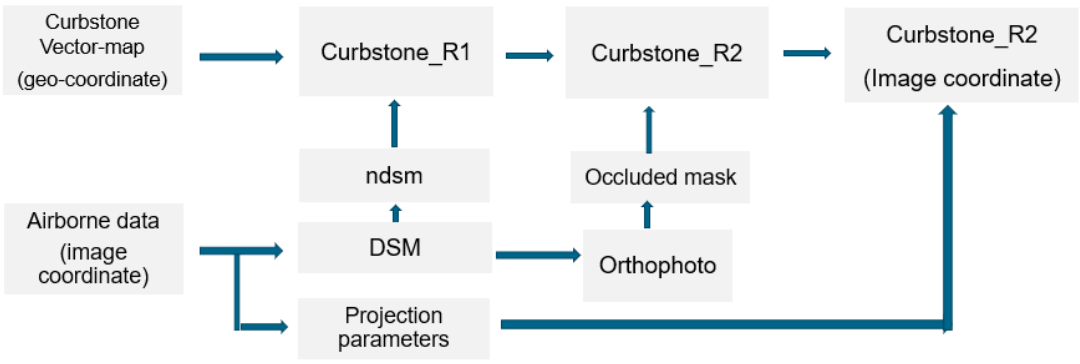

Fig. 3. Workflow of the curbstone vector data refinement.

$$Curbstone = \begin{cases} Objects_{high}, if(nDSM >= 2m) \\ Objects_{low}, if(nDSM < 2m) \end{cases}$$

2) Occluded region detection: Visibility analysis is an challenging topic in computer vision, many accelerated ray tracing approaches are proposed [11]. Fig. 4 analyzes the visibility of the objects around high buildings. As it shows, the roof of the blue building will be at the location of the bold yellow line in the oblique view image. The green line marks the occluded region where reference data are not visible from the sensor's perspective. To avoid wrong labels with impact on the training of the neural network, we generate the orthophoto (orthographic perspective from above with geometry corresponding to the reference data) of each single image in which occluded regions are be presented as no-value pixels. Fig. 5 shows an example of curbstone polygons overlaid on an ortho photo (occlusions result in white color). The red and yellow color represent the low- and high- curbstones as explained in Section II, respectively.

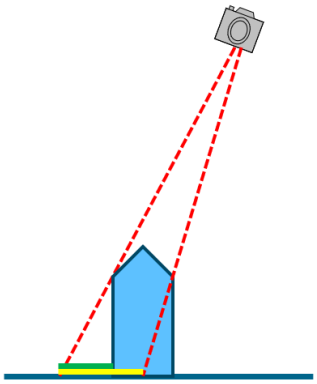

Fig. 4. Visibility analysis; occlusion happening due to building shape and sensor perspective.

B. Segmentation

To preserve the original textures in RGB images, we use the original images for curbstone detection. Therefore, the refined

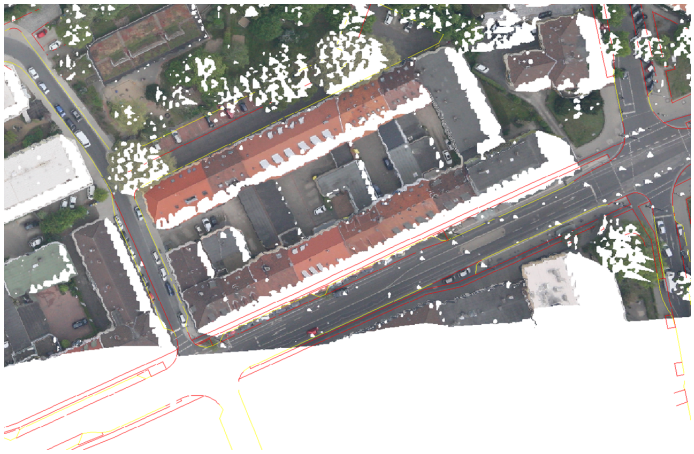

Fig. 5. Example of the curbstone polygons overlaid on generated ortho image. Occlusions due to sensor perspective lead to pixels with white color. Reference data in that area are to be excluded from the training procedure.

curbstone masks are reprojected to each single image by using the projection parameters derived from the bundle adjustment. We dilate the polylines to a width of five pixels in order to mitigate the impact of co-registration errors between the aerial images and the curbstone reference masks.

As representative for state-of-the-art AI-based segmentation methods, the Swin transformer [12] is used with self-attention mechanisms as one of its key elements. As a new vision transformer, Swin Transformer serves as a general-purpose backbone for computer vision. It can produce a hierarchical feature representation and has linear computational complexity with respect to input image size. Furthermore, Swin Transformer achieves the state-of-the-art performance on many semantic segmentation datasets, significantly surpassing previous best methods [13].

C. Vectorization

The vectorization step takes segmentation result as input and converts pixel segments (raster data) of identified curbstones into polygons (vector data). The conversion process is divided into three stages:

1) A skeleton is extracted from the raster representation of curbstones, resulting in a chain of pixels.
2) As not all pixels are necessary to define representative polygons, key points (nodes) and edges are detected from the pixel chain to represent the shape of curbstones.
3) A graph is built to connect the detected nodes and edges.

IV. Experiment

A. Experiment setup

Our experiments are conducted with the PyTorch deep learning framework. We adopt Swin transformer from the mmsegmentation framework [14] as our codebase and follow the default augmentation. As a highly unbalanced segmentation, we take the Dice loss [15] function to train the segmentation model, and set the weight between curbstone and background to (20:1). In our experiment, we use 70 images for training and 10 images for testing. To this end, the refined curbstone reference data is reprojected to the image geometry and resampled to the same spatial resolution.

B. Results

Comparing the original curbstone reference data with the aerial images, many polylines locate on building roof and tree crowns. These lines are detected as outliers and removed for refinement. Fig.6 shows one example of the original curbstone maps for a scene in Brunswick, and the refined curbstone masks which are used to train the Swin transformer.

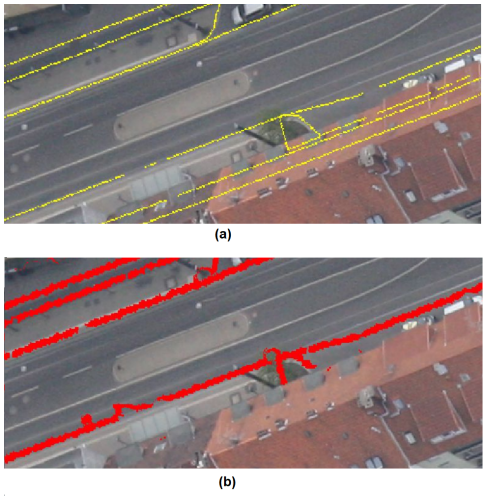

Fig. 6. Comparison of the Curbstone masks before (a) and after (b) the refinement step.

Fig. 7 presents the predicted curbstone probabilities for two test images from the Brunswick data set. In Fig. 7 (a) and (c), the probability maps are highlighted with yellow color and overlaid on the original images. To visually analyze the quality of the prediction result, the corrected reference maps for curbstones of the same regions are displayed in Fig 7 (b) and (d). We show two categories of curbstones with yellow and read colors in the reference data, which present high- and low-curbstones as described Section II. This visual comparison shows that the segmentation results and reference data correspond to each other in terms of position and direction.

Moving one step further towards mapping of curbstone areas with vector data, the red polylines in Fig. 8 (a) represent the automatic vectorization result along a main road of Brunswick. For evaluation, we compare it to the reference curbstone map (blue color in Fig. 8 (b)). It can be seen that most of the detected polylines correspond to the reference in terms of position, whereas some curbstone vectors show deviations in terms of direction.

V. Discussion and Conclusion

Neural networks dominate segmentation and object extraction tasks in many applications and research fields [16]–[18]. However, the performance largely depends on the quality and the amount of training data, which is expensive and challenging to generate [13]. Therefore, integration of openly available official data (with known origin and quality) into automatic image analysis is one solution to reduce the effort and benefit from independent observations. Representative for this direction, we explore the possibility of the fusion of airborne stereo imagery and official surveying data for the automatic segmentation of curbstones along roads. Experimental results show that the proposed workflow is able to solve the objective and can be useful for mapping of large regions.

Integrating existing even open-source datasets saves time and reduces costs. However it also requires pre-processing to reduce label noise and fit the geometry of the data to be analyzed, as was exemplified in this paper.

The curbstone segmentation and vectorization can be further improved in various ways. Better performance can be expected with advanced loss functions which are more sensitive to detecting edges. In addition, contextual information can be more considered to improve the direction of identified curbstone polylines.

References

[1] S. Hegemann, S. Lüke, and C. Nilles, "Curbstone detection as one part of urban driver assistance," *ATZ worldwide*, vol. 115, no. 11, pp. 34–39, 2013.

[2] L. Zhou and G. Vosselman, "Mapping curbstones in airborne and mobile laser scanning data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 18, pp. 293–304, 2012.

[3] S. E. C. Goga and S. Nedevschi, "Fusing semantic labeled camera images and 3d lidar data for the detection of urban curbs," in *2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 301–308, IEEE, 2018.

[4] C. Chai, T. Liu, J. Liu, X. Cao, Y. Hu, and Q. Li, "Surrogate evaluation model for assessing lane detection reliability of automated vehicles in complex road environments," *Transportation Research Record*, p. 03611981231152260, 2023.

[5] F. Kurz, S. Tuermer, O. Meynberg, D. Rosenbaum, H. Runge, P. Reinartz, and J. Leitloff, "Low-cost optical camera systems for real-time mapping applications," *Photogrammetrie-Fernerkundung-Geoinformation*, pp. 159–176, 2012.

[6] J. Tian, P. Reinartz, P. d'Angelo, and M. Ehlers, "Region-based automatic building and forest change detection on cartosat-1 stereo imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 79, pp. 226–239, 2013.

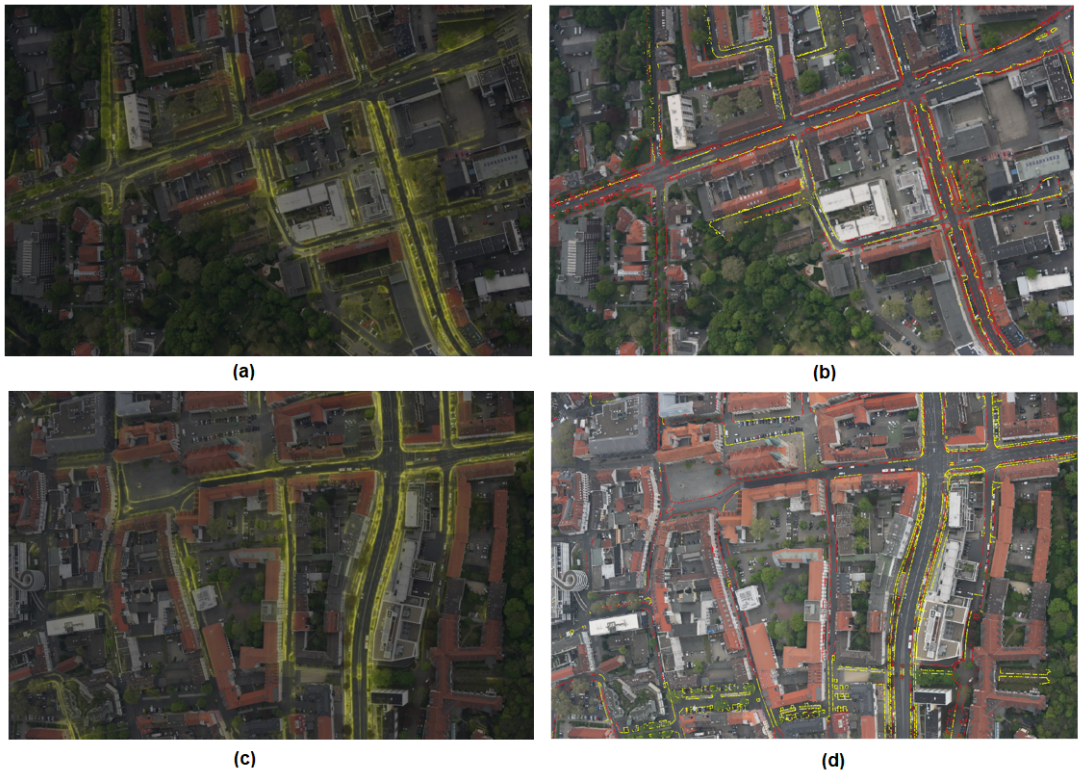

Fig. 7. Curbstone segmentation result for two scenes in Brunswick: Images (a) and (c) show the predicted probability, images (b) and (d) present the refined training data derived from official surveying data. Background: aerial images used for segmentation.

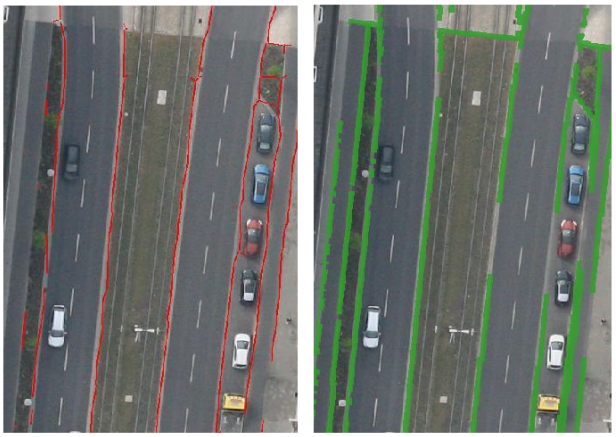

Fig. 8. A subset of vectorization result (a) compares to the reference curbstone map (b) overlaid on the original RGB image.

[7] P. d'Angelo, "Improving semi-global matching: cost aggregation and confidence measure," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41, pp. 299–304, 2016.

[8] C. Kempf, J. Tian, F. Kurz, P. D'Angelo, T. Schneider, and P. Reinartz, "Oblique view individual tree crown delineation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 99, p. 102314, 2021.

[9] G. Grohman, G. Kroenung, and J. Strebeck, "Filling srtm voids: The delta surface fill method," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 3, pp. 213–216, 2006.

[10] R. Qin, J. Tian, and P. Reinartz, "Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images," *International Journal of Remote Sensing*, vol. 37, no. 15, pp. 3455–3476, 2016.

[11] A. Marsetič, "Robust automatic generation of true orthoimages from very high-resolution panchromatic satellite imagery based on image incidence angle for occlusion detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3733–3749, 2021.

[12] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021.

[13] H. Li, J. Tian, Y. Xie, C. Li, and P. Reinartz, "Performance evaluation of fusion techniques for cross-domain building rooftop segmentation," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 501–508, 2022.

[14] M. Contributors, "MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark." https://github.com/open-mmlab/mmsegmentation, 2020.

[15] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, pp. 1–7, IEEE, 2020.

[16] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–35, 2019.

[17] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.

[18] Y. Xie, J. Tian, and X. X. Zhu, "A co-learning method to utilize optical images and photogrammetric point clouds for building extraction," *International Journal of Applied Earth Observation and Geoinformation*, vol. 116, p. 103165, 2023.