Geocarto International

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgei20

Automatic building outline extraction from ALS point cloud data using generative adversarial network

Gefei Kong, Hongchao Fan & Gabriele Lobaccaro

To cite this article: Gefei Kong, Hongchao Fan & Gabriele Lobaccaro (2022) Automatic building outline extraction from ALS point cloud data using generative adversarial network, Geocarto International, 37:27, 15964-15981, DOI: 10.1080/10106049.2022.2102246

To link to this article: https://doi.org/10.1080/10106049.2022.2102246

Published online: 12 Aug 2022.

GEOCARTO INTERNATIONAL
2022, VOL. 37, NO. 27, 15964–15981
https://doi.org/10.1080/10106049.2022.2102246

Automatic building outline extraction from ALS point cloud data using generative adversarial network

Gefei Kong, Hongchao Fan and Gabriele Lobaccaro

Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT

The generation of building footprints from laser scanning point clouds or remote sensing images involves three steps: segmentation, outline extraction and boundary regularization/generalization. Currently, existing approaches mainly focus on the first and third steps, while only few studies have been conducted for the second step. However, the extraction result of the building outlines directly determines the regularization performance. Therefore, high-quality building outlines are important to be delivered for the regularization. Determining parameters, such as point distance and neighborhood radius, is the primary challenge in the process of extracting building outlines. In this study, a parameter-free method is proposed by using an improved generative adversarial network (GAN). It extracts building outlines from gridded binary images with default resolution and no other input of parameters. Hence, the parameter selection problem is overcome. The experimental results on segmented point cloud datasets of building roofs reveal that our method achieves the mean intersection over union of 93.52%, the Hausdorff distance of 0.640 m and the PoLiS of 0.165 m. The comparison with α-shape method shows that our method can improve the extraction performance of concave shapes and provide a more regularized outline result. The method reduces the difficulty and complexity of the next regularization task, and contributes to the accuracy of point cloud-based building footprint generation.

ARTICLE HISTORY
Received 25 April 2022
Accepted 11 July 2022

KEYWORDS
extraction of building outlines; point cloud data; GAN; deep learning

1. Introduction

Building footprint information is one of the most important data in many geographic applications, such as mapping, 3 D building reconstruction, urban planning, and emergency response (Sugihara et al. 2012, Bittner et al. 2018, Zhu et al. 2021). Automatic footprint generation has garnered considerable attention from the research community. To date, several methods have been proposed to solve this problem. and two main data

CONTACT Hongchao Fan hongchao.fan@ntnu.no
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

sources are used: remote sensing image data and point cloud data. Many studies conducted in the last few decades (Shackelford et al. 2004, Zhu et al. 2021, Guo et al. 2022, Li et al. 2022) achieved automatic building footprint generation based on remote sensing images. However, the limitations of remote sensing images affect the practicality of these methods, such as the lack of elevation information and the presence of shadow casts. Owing to the development of Light Detection and Ranging (LiDAR) technology, point cloud data produced by the LiDAR technology can be more easily obtained. Compared with remote sensing images, point cloud data contain accurate 3 D information, and are less sensitive to occlusions and shadows (Jochem et al. 2012). Hence, point cloud data are a better choice for automatic building footprint generation and have been used by many researchers (Wang et al. 2006, Albers et al. 2016, Li 2020, Mahphood and Arefi 2022).

In general, the standard procedure for building footprint generation involves three steps for point cloud data: (1) segmentation, (2) extraction of building outlines, and (3) boundary regularization or generalization. The first segmentation step classifies points belonging to a building from the point cloud dataset. The second step involves the extraction of building boundary and generation of the preliminary footprint polygon. Finally, the third step of the boundary regularization involves the adjustment of the generated boundary and retrieval of simpler and more regular footprint polygon. In particular, the second step of the building footprint production (i.e., the extraction of building outlines) is important, as it determines the quality of input in the next regularization step and finally influences the result of building footprint generation (Awrangjeb 2016).

According to Awrangjeb (2016), methods for extracting building outlines can be divided into two types: direct and indirect. Direct methods directly extract building outlines based on points, without transformation. However, these methods are sensitive to the selection of parameters, such as neighborhood radius, and are easily affected by the noise in the point cloud data. In particular, owing to the iteration process in these methods, they are much time consuming when the parameters are not properly set, which limits the practicality of these methods. In contrast, indirect methods first transform point cloud data into binary images or triangular networks and then extract building outlines based on the transformed data. These methods are more efficient because they reduce the data volume and noise by the gridding process during the transformation. However, they have the problems of resolution selection and information loss caused by the transformation, which reduces the shape accuracy of the extracted building outlines. All these problems severely limit the application of existing methods.

In recent years, deep learning methods have been successfully applied into the field of image understanding, such as semantic segmentation (Long et al. 2015). As a new deep-learning technology, Generative Adversarial Network (GAN) has shown the potential in image inpainting, image translation and related tasks, which can achieve the information restoration of images, leading to indirect methods with less parameter selection and information loss.

To solve the parameter selection problem in existing methods and improve the shape accuracy of extracted building outlines, in this study, a new parameter-free method with an improved GAN is proposed to achieve automatic and robust outline extraction from airborne laser scanning (ALS) point cloud data. After the generation of a binary image, instead of directly extracting the building outline from the binary image using an edge detector or a boundary tracing algorithm, GAN is implemented to optimize the building binary image. The optimized image with a more accurate building shape and less noise will be applied to the next boundary tracing step in the process for the extraction of building outlines, and finally helps the vector (i.e., polygon) result of building outline extraction. A better outline extraction result will better guide the boundary regularization

in the next step of footprint generation and finally improve the performance of automatic building footprint generation. The main contributions of this study are as follows:

1. GAN is introduced into the process of building boundary extraction to optimize the building shape, repair its pore spaces, and reduce irregularities and noise in the gridded binary image. This reduces the influence of parameter selection (e.g., resolution and grid distance), ultimately avoiding reliance on parameter selection in the method for extracting building outlines and increasing its effectiveness and efficiency.
2. The network structure in the existing GAN is improved to enhance its performance in the task of building outline extraction and better solve the problem of resolution selection. The extraction results of the building outlines from the new network structure will have better shapes than the baseline structure.

This paper is organized as follows: Section 2 is a review of related research on the extraction of building outlines. Section 3 outlines the detailed workflow of the proposed method. Section 4 introduces the experimental dataset and other details. Section 5 contains the experimental results and discussions of the findings. And Section 6 discusses the conclusions of this study and the future work.

2. Related work

2.1. Direct methods

In direct methods, the convex hull algorithm (Berg et al. 2000) is one of the base. However, the traditional convex hull algorithm cannot be applied to the outline extraction of concave polygons, which are usually the outlines of buildings. Sampath and Shan (2007) used a modified convex hull algorithm proposed by Jarvis (1977) to directly extract building boundaries from point cloud data. The modified algorithm can be used to extract concave outlines. Hence, some researchers (Herve 2008, Dai et al. 2017) followed Sampath and Shan's research and used the same method to extract building outlines. However, it is difficult to set a suitable neighborhood distance parameter in this algorithm to avoid the unstable performance of the algorithm when facing uneven point distribution of the point cloud data. Wang and Shan (2009) combined convex hull estimation and the classification of boundary and non-boundary points to achieve outline extraction. Cao et al. (2017) redesigned and refined the modified convex hull algorithm to improve the performance of outline extraction. However, the two methods reduced the impact of neighborhood distance determination, but introduced new neighborhood parameters, such as the minimum number of neighbors (minPts). The problem of parameter determination remains unsolved.

Another powerful algorithm, α-shape (Edelsbrunner and Mücke 1994, Cazals et al. 2005), can also achieve direct outline extraction of a set of points. The α-shape algorithm uses a circle with radius = α to "roll" around the points and traces the order in which it hits each point to achieve the outline extraction of points. Compared to the convex hull algorithm, the α-shape algorithm can trace both convex and concave polygons. Hence, many researchers have used this algorithm to extract building outlines directly from point cloud data. In α-shape-based research, some researchers (Dorninger and Pfeifer 2008, Shahzad and Zhu 2015) set the value of α to twice the mean linear point distance. He et al. (2014) set the value of α to 1.5 times × the average point distance (spacing). Albers et al. (2016) calculated the value of α based on the point cloud density. Li et al. (2015)

further improved the α-shape algorithm using a dual threshold. However, the thresholds of α were manually defined based on experience, making these methods less adaptable to data changes. dos Santos et al. (2019) further improved the algorithm and estimated the α value based on the local point spacing rather than the global average. However, this method introduced a new parameter, neighborhood radius, to help the algorithm adapt to the density variation in the point cloud data and each building. The setting of the new parameter remains empirical.

In contrast to the above methods, Li (2020) directly extracted points at the building outline using multiple return information of the point cloud data. However, this method cannot be used for point cloud data without this information. The author had to propose a recursive convex hull algorithm to solve this problem. However, the problem of the convex hull algorithm, in which the parameter settings are difficult, ensues. In addition, this algorithm is less efficient.

2.2. Indirect methods

In indirect methods, the task of building outline extraction from point cloud data is typically simplified by transforming the 3 D point cloud data into a 2 D-image representation. Verma et al. (2006) used the oriented bounding box of each building's point cloud data to generate grids, and then marked the grids with enough points as the foreground to generate the corresponding binary image. The outline of the foreground region was regarded as the outline of the building roof. Poullis and You (2009) directly converted 3 D data to a 2 D XYZ map by subdividing and resampling the raw data, and then achieved the region boundary extraction based on the 2 D map. Zhou and Neumann (2009) divided raw point cloud data into 2 D grids and achieved the classification of points and extraction of building outlines based on grids. Awrangjeb and Fraser (2014) generated grids of point cloud data and used the Canny edge detector (Canny 1986) to extract the building roof outlines based on the grids. Mahphood and Arefi (2017) also generated grids and then extracted outlines based on the number of neighboring points. Mahphood and Arefi (2022) proposed an enlargement method to extract and regularize building outlines based on grids. However, as mentioned by Poullis and You (2009), information loss cannot be avoided during the transformation process owing to the resampling or "clustering". The selection of a proper resolution is also a problem, that has a significant impact on the accuracy of the final extraction result. Mahphood and Arefi (2022) markedly reduced the effect of parameter settings, but a proper resolution for gridding is still needed. Hence, although it is more efficient than direct methods, limited research for extracting building outlines from point cloud data has been performed based on this indirect image-based idea.

Some indirect methods are based on triangulation to avoid information loss. Awrangjeb (2016) used Delaunay triangulation to find the initial boundaries and then optimized the extraction result of the building outlines by removing the long boundary edges. However, this method loses the efficiency advantage of indirect methods as it is time consuming to achieve triangulation for many points. In addition, the threshold of the removal strategy affects the practicality of this method.

3. Methodology

3.1. Overview of our method

The workflow of the proposed method is illustrated in Figure 1. First, the point cloud data of each building are grided to create a corresponding binary image. Subsequently,

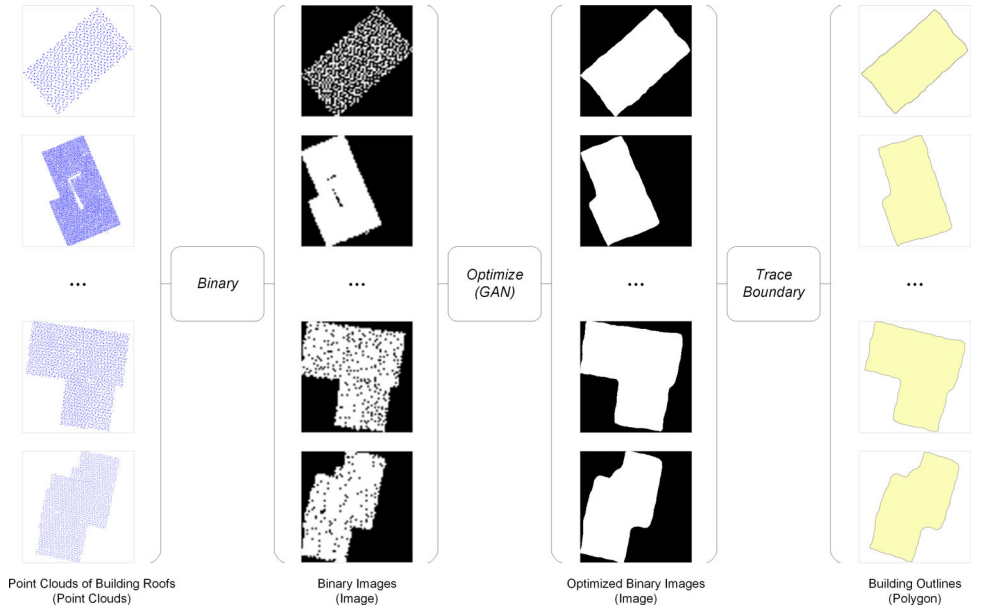

Figure 1. Workflow of our method for the extraction of building outlines.

GAN is applied to optimize the building shape and repair the pore spaces in the binary image. After image optimization, a more accurate extraction of the building outlines can be achieved by boundary tracing based on the optimized binary image.

3.2. Generation of the binary image

Unlike existing image-based indirect methods, we do not define the geographical resolution parameter, such as 2 m or 1.5 m, and then convert the point cloud data into binary images of different sizes. In our method, the resolution of the output binary images, RES_{BImg}, is set as the resolution parameter. The maximum distance in the x- or y- direction of each building is used to determine the grid size in both directions. The calculation method for the size of each grid S_g^i for each building's point cloud data PS^i is shown in Equation (1).

$$S_g^i = \frac{\max\left(\left(x_{max}^i - x_{min}^i\right),\ \left(y_{max}^i - y_{min}^i\right)\right)}{RES_{BImg}} \tag{1}$$

where S_g^i is the size of each grid for the point clouds, $P = (x, y, z)$ is the coordinate of each point, and i is the number of point cloud data of the buildings (which is equal to the number of buildings).

Each building's point cloud data PS^i are grided based on the size of each grid S_g^i. The value of each grid with points is set to 1, as the foreground of the binary image and then the binary image generation for each building is completed.

Using this new resolution parameter, the generated binary image of each building is of the same size. This enlarges buildings with small areas, which can help their outline extraction. However, if only this new resolution parameter is used, a similar problem of resolution selection in the existing image-based indirect method will still exist and the

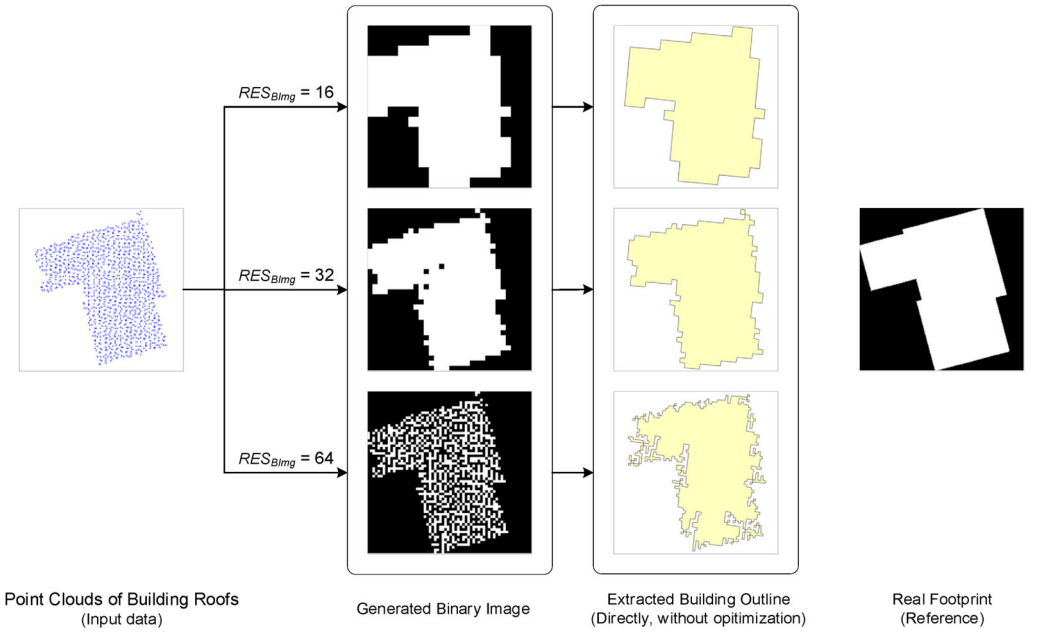

Figure 2. Sample of resolution selection problem: the direct extraction of building outlines from binary images generated by different resolution parameters.

large parameter will cause pore spaces inside the building point cloud data, as shown in the first, third and fourth buildings in Figure 1. Hence, GAN is introduced in the next subsection to solve these problems.

3.3. Optimization of the binary image using improved GAN

The gridding process reduces the data volume, and the use of binary images simplifies the boundary tracing problem. However, the resolution selection and pore spaces remain a critical challenge. An inappropriate RES_{BImg} significantly affects the final extraction performance of the building outlines, as shown in Figure 2. If RES_{BImg} is set high (e.g., $RES_{BImg} = 64$), the building shape details will be saved. However, pore spaces will be found in the binary image owing to the limited and imbalanced point density. Nonetheless, if given a low RES_{BImg} (e.g., $RES_{BImg} = 16$) to avoid the pore spaces, the boundary of the building in images will become too coarse and the outline will become inaccurate. Moreover, considering the different sizes of buildings, it is difficult and even impossible to set an appropriate resolution for all buildings. Hence, the impact of resolution selection should be avoided as much as possible.

To solve this problem, GAN is introduced. GAN uses adversarial learning, and can generate input-like outputs by imitating the input data distribution (Goodfellow et al. 2014). A GAN includes two critical submodules: a discriminator module and generator module. The discriminator is deployed to check whether the input is real or fake (i.e., generated), while the generator learns the distribution of input data and attempts to generate a realistic-looking output to fool the discriminator. By alternately training the discriminator and generator, the discriminator can better distinguish between the real input and fake input from the generator. Owing to a better discriminator, the generator gradually improves itself and generates an output that is increasingly close to the real data. In

addition to solving the problem of resolution selection, owing to the game of the discriminator and generator in GAN, the noise in the input image will also be learned and then excluded by the generator to fabricate an image close to the real distribution. Hence, GAN can also minimize noise interference in the point cloud data, which cannot be achieved by direct methods of outline extraction.

In our method, the framework *Pix2Pix* (Isola et al. 2017) is improved and used to optimize the binary image. The pipeline of *Pix2Pix* is shown in Figure 3(a). The detailed structures of the generator and discriminator of *Pix2Pix* are shown in Figure 3(b)–(d). Compared to other GANs, the original *Pix2Pix* from Isola et al. (2017) has the following characteristics:

1. *Pix2Pix* combined the input image Img_{in} and noise vector NV as the input of the generator, rather than used the noise vector as the only input. The input of the discriminator was changed from the generator output $G(Img_{in,}\ NV)$ to both the generator output and the original input image $(Img_{in},\ G(Img_{in,}\ NV))$. This design improves the control of the generator output in the network, which can generate images closer to the input image.
2. *Pix2Pix* improved its generator by using a UNet-like structure. The addition of skip connection between the corresponding layers in the encoder and decoder improves the information connection between them, and improves the generation result.
3. PatchGAN was proposed and used as the discriminator model, which classifies an image as real or fake for each patch, rather than the entire image. The high-frequency structure in the data distribution can be better learned by PatchGAN, which further improves the network performance.
4. *Pix2Pix* designed a better loss function to generate images with better edge information that are closer to the expected data distribution. The loss function is given by Equation (2):

$$\begin{aligned} & L_{Pix2Pix} = \min_G \max_D L_{GAN}(G,D) + \lambda L_1(G) \\ = \ & \min_G \max_D \mathbb{E}_{(Img_{in},\ Img_{gt})}\left[\log D(Img_{in},\ Img_{gt})\right] \\ + \ & \mathbb{E}_{(Img_{in},\ NV)}\left[\log(1 - D(Img_{in},\ G(Img_{in},\ NV)))\right] \\ + \ & \lambda \mathbb{E}_{(Img_{in},\ \ Img_{gt},\ NV)}\left[\left|\ Img_{gt} - G(Img_{in},\ NV)\right|_1\right] \end{aligned} \tag{2}$$

where Img_{gt} is the real image and λ is the weight of L_1 loss function, which can help the generator output to be more similar to the real image in structure.

Pix2Pix added the skip connection structure into its generator to enhance the information connection between the encoder and decoder, as shown in Figure 3 (b), but in the encoder of the generator, the shape information is still severely lost owing to the downsampling process. However, shape is one of the most critical pieces of information in the task of building outline extraction. To improve the shape information connection, we redesign the encoder in the generator. The improvements are described below.

1. As shown in Figure 4, the residual block from ResNet (He et al. 2016) is used to replace the original simple downsampling encoder block. The residual block transforms the original flow of information transmission between the two layers $\{y_i \leftarrow f(y_{i-1})\}$ into the new residual connection flow $\{y_i \leftarrow g(y_{i-1}) + y_{i-1}\}$. With this change, more image details from the higher-resolution layer y_{i-1} are saved and transferred to the next layer, including the shape information. Hence, by introducing the

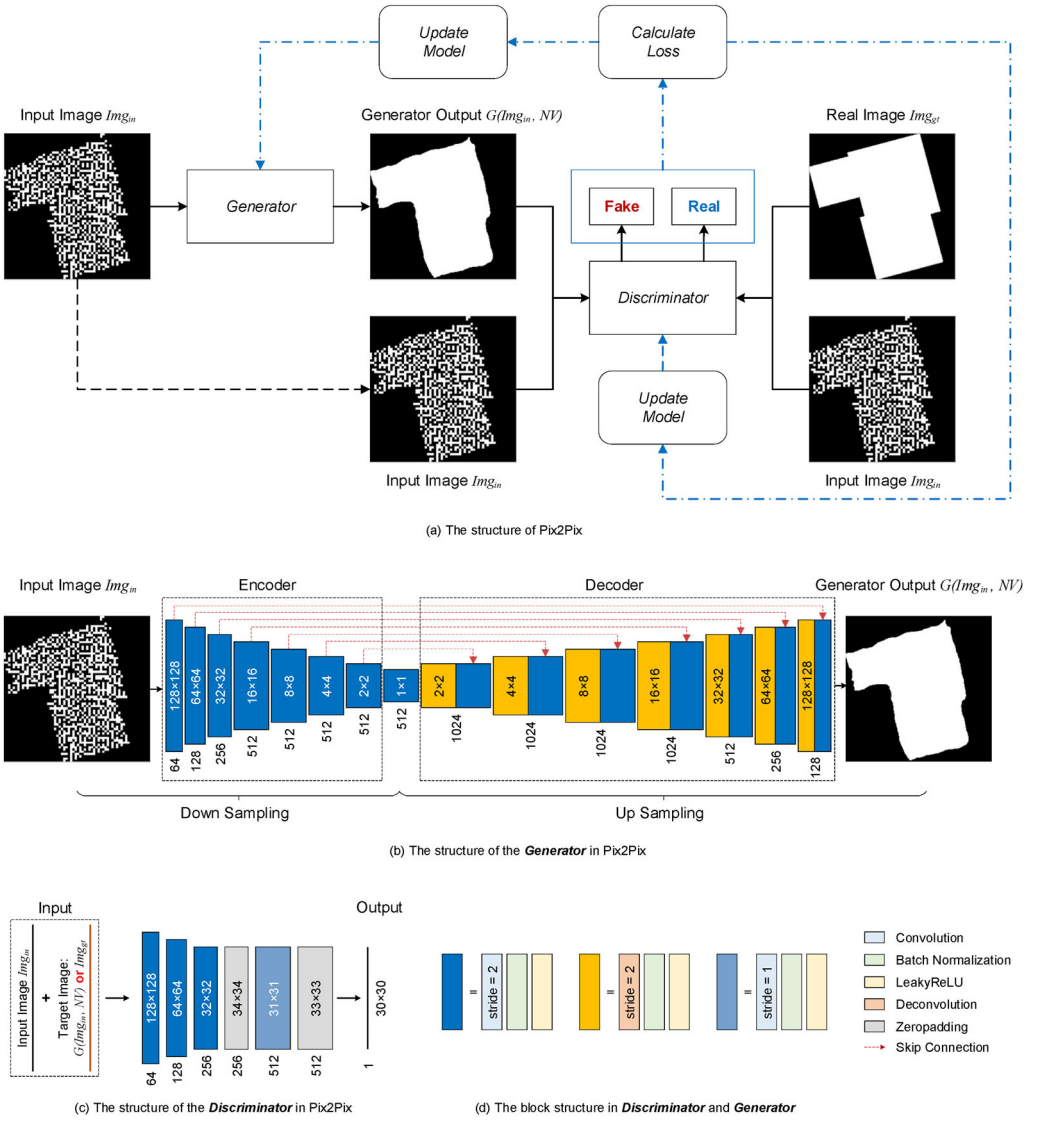

Figure 3. The pipeline of *Pix2Pix*.

residual block into the encoder, more shape information can be saved during the downsampling process, and the building shape of the generator output can be enhanced.

2. In addition to the improvement of the encoder block, the method of skip connection between the encoder and decoder in the generator is adjusted, as shown by the red dotted lines in Figure 4. Replacing the last output of each encoder block in the original generator structure for the skip connection, the new layer for the skip connection is the downsampling output in each encoder block. The new position of the skip connection will help the generator save more information from the original images, such as the original data distribution, to better guide the generator output closer to the expected distribution.

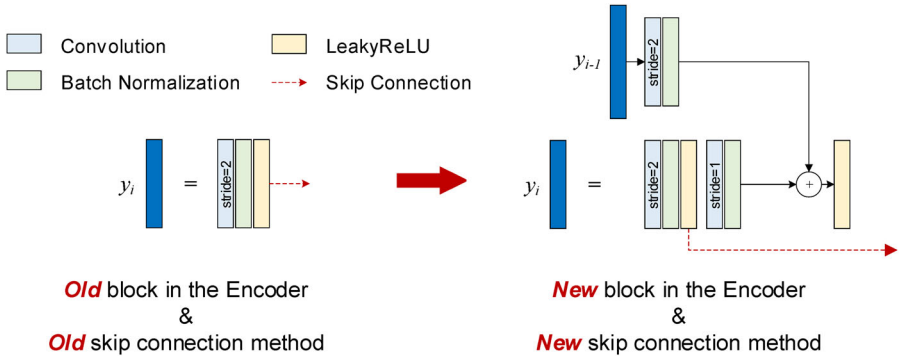

Figure 4. Improvements to the generator in *Pix2Pix*: the new block structure of the encoder in the generator and the new skip connection method.

The generator output $G(Img_{in,}\ NV)$ from the improved GAN is regarded as the optimization result. With sharper boundaries and filled pore spaces, the optimized image from GAN can better represent the real building outlines and contribute to better extraction results of building outlines.

3.4. Image to polygon to obtain the result of outline extraction

After the optimization step, the image-border-following algorithm from Suzuki and Abe (1985) is applied to trace the image building outlines O_i from the optimized binary images. Thereafter, to obtain the final building outlines, two processes should be completed: (1) transformation of the coordinate system and (2) transformation from points to polygons. In the first process, because the traced building outlines are still in the local image coordinate system, there is no geographical reference for these outlines. Hence, these outlines should be transformed from the local system to the global geographical coordinate system for following application. The second process is due to the typical use of polygon features to represent areas, such as buildings.

In the first coordinate system transformation process, the mapping between the geographical coordinates and image pixels is calculated based on the original point cloud data of each building. This process is the same as the gridding process described in section 3.2. The bounding box of each building's point cloud data is extracted and used to generate the grid $Grid_{size_g}$, where the size of the entire grid is $(size_g \times size_g)$, which is the same as the image size of the generator output, $G(Img_{in,}\ NV)$. The geographical coordinates of each pixel (i_{o_i}, j_{o_i})in the image building outline O_i is $Grid_{size_g}(i_{o_i}, j_{o_i})$. Finally, the point set of each building outline with geographical information is obtained. Thereafter, in the second process of points to polygons, each point set is sorted and reorganized into a polygon, and the output is the final extraction result of the building outline.

4. Experiments

4.1. Experiment dataset

The experiments are conducted on ALS point cloud data provided by the mapping authority of Trondheim Municipality. These data are collected in Trondheim, Norway in 2018, at a standard density of 12-20 points/m^2. This Trondheim 3 D point cloud dataset containing 903 building roofs with different shapes and sizes is pre-manually extracted

from the raw ALS data for the extraction of building outlines in this dataset. General building footprint (roof) types with different shapes, such as rectangle-shaped, L-shaped, T-shaped and more complex combined-shaped, are included in this dataset to ensure the generalization of our model. The building footprint dataset from the national open geographical data provided by the Norwegian Mapping Authority (FKB-Buildings Dataset 2021) is used as the ground truth dataset for network training and result evaluation. This Trondheim dataset is split into three groups at a ratio of 7:2:1 for training, validation, and testing. The three groups have 630, 180 and 93 samples, respectively.

Another benchmark dataset, the ISPRS dataset (Vaihingen) (Cramer 2010), is also considered in the experiment to more completely evaluate our method. This dataset is provided by the ISPRS Test Project on Urban Classification and 3 D Building Reconstruction, and is collected at Vaihingen, Germany in 2008, with a low point density of 4-6 points/m^2. A total of 94 building roof point clouds with ground-truth footprint shapes are included in this dataset. These building footprints also have different shapes and sizes. Following its official guidance, the dataset is split into three areas: area1, area2 and area3. The building roofs located in area2 and area3 are used for training and validation, and those in area1 are used for testing. A total of 40, 20, and 34 samples are used for training, validation, and testing, respectively.

4.2. Experiment details

The improved GAN is implemented by Tensorflow 2.6.2 (Abadi et al. 2016) and trained on a NVIDIA Tesla P100-PCIE-16GB graphics processing unit (GPU) and an Intel(R) Xeon(R) central processing unit (CPU) (2.20 GHz) supported by Kaggle (https://www.kaggle.com/). Adam (Kingma and Ba 2015) is used as the optimizer to train both the generator and discriminator in the improved GAN. network, with a learning rate of 0.0005 for the generator and 0.0003 for the discriminator, $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The parameter λ in the loss function is set to 100. The input size of the network is set to 256×256 for both the input and real images. The batch size is set to 1.

4.3. Evaluation metrics

Three metrics are used to evaluate the extraction results of building outlines: mean intersection over union (mIoU), Hausdorff distance (HD) (Huttenlocher et al. 1993), and PoLiS (Avbelj et al. 2015). The first metric is used to compare the area similarity (i.e., area completeness) between the extracted building outlines and the real building footprint polygon. The second and third metrics are used to compare the maximum shape distance and average shape similarity (distance), respectively.mIoU is calculated using Equation (3).

$$mIoU = \frac{1}{N}\sum_{n=1}^{N}\frac{(Pred_n \cap GT_n)}{(Pred_n \cup GT_n)} \tag{3}$$

where N is the number of samples, $Pred_n$ is the n extracted building outline using our method, and GT_n is the n real building footprint polygon. The intersection and union area between $Pred_n$ and GT_n are calculated based on polygons in the geographical coordinate system to ensure the reliability and referability of these metrics.

The calculation for HD and PoLiS can be found in the cited papers. HD and PoLiS are also calculated in the geographical coordinate system.

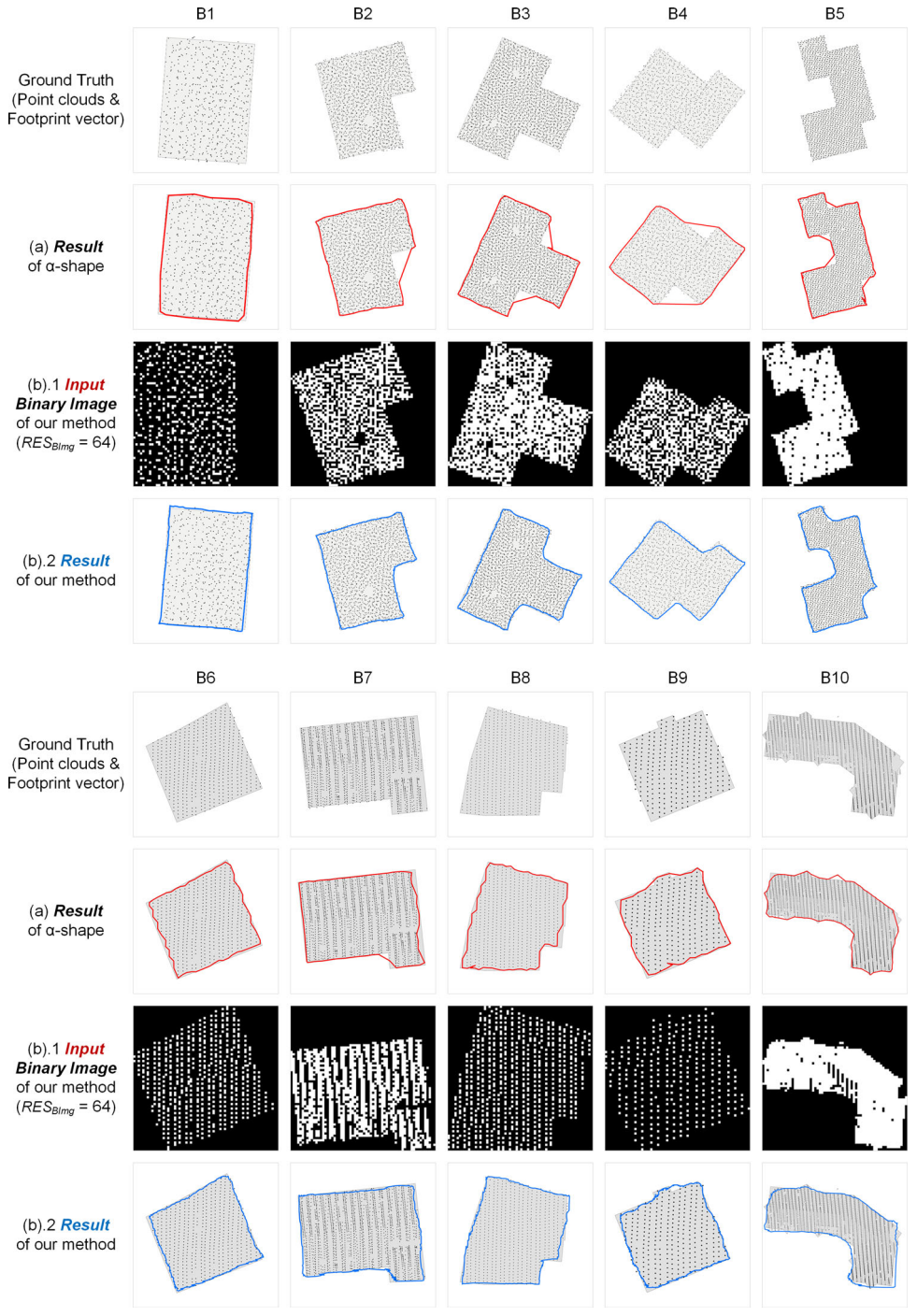

Figure 5. Results for the extraction of building outlines for different methods where B1–B5 are on the Trondheim dataset and B6–B10 are on the ISPRS dataset (images in (b).1 are shown in the image coordinate system without geographical projection).

Table 1. Quantitative result for the extraction of building outlines for different methods.

Dataset	Method	mIoU(%)	HD(m)	PoLiS(m)
Trondheim	α-shape	92.80	0.858	0.187
	Our method	**93.52**	**0.640**	**0.165**
ISPRS	α-shape	83.10	1.825	**0.413**
	Our method	**85.18**	**1.796**	0.429

5. Results and discussions

5.1. Comparison with the existing method

To evaluate the effectiveness of our method on datasets with different point densities, a classic method for the extraction of building outlines, α-shape, is compared with our method. Based on related research (Dorninger and Pfeifer 2008, Shahzad and Zhu 2015, dos Santos et al. 2018, 2019), the different α values for each building are provided, which are set to twice the average point distance density of each building's point cloud data.

A qualitative comparison of the results is presented in Figure 5. The buildings of different shapes are shown in this figure. The results of the α-shape algorithm are shown in Figure 5(a) and those of the proposed method are shown in Figure 5(b). The black points in Figure 5 represent the building points from the point cloud dataset, the gray polygons represent the ground truth building footprint polygons, and the lines with different colors represent the results of the building outlines extracted by different methods. Based on the qualitative results, our method is effective and more accurately retains the shape details of the buildings. Both the α-shape algorithm and our method perform well on a simple convex polygon (B1 and B6). However, the α-shape algorithm cannot trace the concave part of the building polygon well, as shown in B2–B5 and B7–B10, while our method can handle building polygons with both convex and concave parts. Our method can also extract building outlines with fewer noise points than the α-shape algorithm as shown by the upper boundaries of B1 and B8. Moreover, compared with the "zig-zag" building outlines extracted by α-shape, our method can directly extract more smooth outlines, as shown in B8. These results demonstrate that our method can contribute to the accuracy of the final building footprint. Overall, the visualization result further indicates that our method can accept a binary image with pore spaces, which is still not addressed by the newest related research (Mahphood and Arefi 2022). The adaptation of pore spaces in a binary image allows our method to retain more details of building shapes and thus extract outlines with higher shape accuracy.

The quantitative evaluation results are presented in Table 1. On the Trondheim dataset and comparing the α-shape, the metric mIoU in our method increases 0.72% and HD increases by 0.22 m. For the strictest metric – PoLiS, our method could result in an increase of 0.02 m. On the ISPRS dataset, the metrics mIoU and HD in our method increase by 2.1% and 0.03 m, respectively, with PoLiS decreasing by 0.02 m. According to the quantitative result, compared with the α-shape algorithm, the introduction of GAN is effective, and our method obtains better results for both datasets. Only the PoLiS of our method on the ISPRS dataset is slightly lower than the α-shape result. The increase in mIoU indicates that our method can extract building outlines more completely and close to the regularized footprint, and the increase in HD and PoLiS indicates that our method better retains the shape details. Although our method performs similar to the α-shape in shape accuracy on the ISPRS dataset, our method better guarantees the completeness of the extracted building outlines and provides a more regularized result.

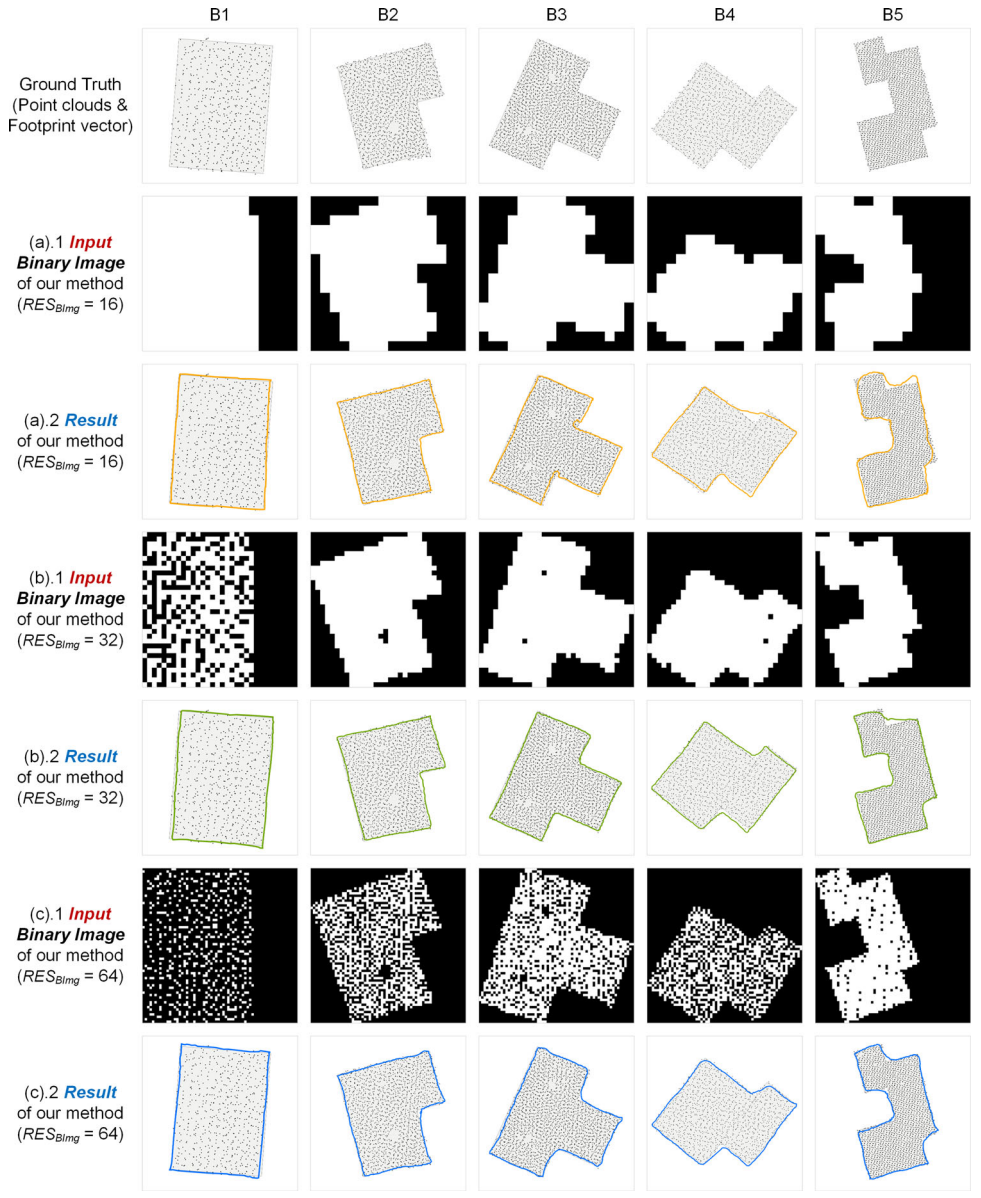

Figure 6. Results for the extraction of building outlines for different resolutions on the Trondheim dataset (point density: 12-20 points/m^2).

The point density and ground truth quality of the ISPRS dataset are markedly lower than those of the Trondheim dataset, which negatively affects the accuracy and information volume of the binary image. Hence, the decrease in shape accuracy on the ISPRS dataset is predictable. Examples of B6–B10 in Figure 5 highlight such case. Our method obtains a slightly lower average shape similarity than the α-shape on this dataset. However, the building outlines extracted using our method is markedly smoother than those extracted using the α-shape. In addition, less noise is presented in the results of the proposed method.

Table 2. Quantitative results for the extraction of building outlines for different resolutions on the Trondheim dataset (point density: 12-20 points/m^2).

Dataset	RES_{BImg}	mIoU(%)	HD(m)	PoLiS(m)
Trondheim	16	92.14	0.703	0.194
	32	93.21	**0.596**	0.171
	64	**93.52**	0.640	**0.165**

Table 3. Quantitative results for the extraction of building outlines for different resolutions on the ISPRS dataset (point density: 4-6 points/m^2).

Dataset	RES_{BImg}	mIoU(%)	HD(m)	PoLiS(m)
ISPRS	16	83.55	1.851	0.480
	32	84.43	1.823	0.444
	64	**85.18**	**1.796**	**0.429**

Table 4. Results of the ablation study.

Structure	RES_{BImg}	Changes		mIoU(%)	HD(m)	PoLiS(m)
		(a)	(b)			
Pix2Pix	16			91.67	0.796	0.218
	32			92.67	0.655	0.185
	64			93.48	0.676	0.189
Pix2Pix–(a)	16	√		91.89	0.727	0.209
	32	√		92.76	0.642	0.180
	64	√		93.31	**0.564**	**0.165**
Our improved Pix2Pix (Our improved GAN)	16	√	√	92.14	0.703	0.194
	32	√	√	93.21	0.596	0.171
	64	√	√	**93.52**	0.640	**0.165**

5.2. Analysis of resolution parameter choosing

The proposed method is parameter-free. It means that, in comparison with existing methods, the parameter setting is not very important, and our method does not actually need to set parameter when applied in a practical project. In our method and many other indirect methods, this critical parameter is the resolution for gridding point clouds into binary images, referred to as RES_{BImg} in this study. To investigate whether our method avoids the impact on resolution selection, RES_{BImg} is set to three different values: 16, 32 and 64. The Trondheim dataset is mainly used for the experiment. In this dataset, the chosen values correspond to three different gridding levels:

1. 16: few details of building shape but almost no pore spaces in the binary image (can be accepted by existing methods), as shown in Figure 6 (a).1;
2. 32: some pore spaces but better shape details in the binary image (cannot be accepted by existing methods), as shown in Figure 6 (b).1;
3. 64: more shape details but many pore spaces in the binary image (cannot be accepted by existing methods), as shown in Figure 6 (c).1.

The qualitative and quantitative experimental results are presented in Figure 6 and Table 2. As reported in the quantitative experiment result, from RES_{BImg}=64 to 32 with the 2^2-fold information loss as shown in Figure 6 (c).1–(b).1, the mIoU only decreases by 0.32%, the PoLiS only decreases by 0.01 m, and the HD increases by 0.04 m. From RES_{BImg}=32 to 16 with the 2^2-fold information loss as shown in Figure 6 (b).1–(a).1, the mIoU decreases by 1.07% and the PoLiS decreases 0.02 m, and the HD only decreases by 0.11 m. Considering the degree of information loss between two resolution levels, our

method performs quite stably. Moreover, compared with existing indirect methods, pore spaces in the binary image are allowed in our method and the resolution parameter does not need to be carefully adjusted. These indicate that our method improves and balances the performance of the extraction of building outlines under different resolution parameters. The parameter setting in our method is not as important as that in existing methods. The qualitative visualization results also support this conclusion.

Furthermore, our method is tested on the ISPRS dataset using different resolution parameters RES_{BImg}. The ISPRS dataset has a lower point density than the Trondheim dataset. The experimental results are listed in Table 3. As shown in Table 3, our method also obtains the best result at RES_{BImg}=64 for the ISPRS dataset. This result indicates that the default parameter setting (64) can adapt to datasets with different point densities.

Overall, in our method, setting $\mathrm{RES_{BImg}}$ is not as important as setting this parameter in existing related methods. In fact, our method does not require parameter tuning in practical project as the best results can be achieved using the default parameter (RES_{BImg}=64).

5.3. Ablation study

In our method, two changes are introduced to improve the performance of *Pix2Pix* in the task of building outline extraction. The ablation study is performed to evaluate the effectiveness of these changes. Table 4 shows the results of the ablation study, where change (a) is the residual block and change (b) is the new skip connection position. As there is no additional convolutional layer for the new skip connection position without the addition of the residual block, the *Pix2Pix*–(b) (the *Pix2Pix* network with only the new skip connection position) does not need to be considered in the ablation study.

As shown in Table 4, the two changes proposed in our method are effective. Our improved *Pix2Pix* structure with two changes generally performs better than the original *Pix2Pix* structure and the *Pix2Pix* structure with change (a) in the task of extracting building outlines. Although the HD of our improved *Pix2Pix* structure at $RES_{BImg} = 64$ is 0.08 m less than the HD of the *Pix2Pix* structure with the residual block alone, all metrics of our improved *Pix2Pix* structure perform better at $RES_{BImg} = 16$ and 32. Such finding indicates that our improved *Pix2Pix* structure is more stable and robust when facing different resolution parameters, and better solves the resolution selection problem in image-based indirect methods. Hence, the improved *Pix2Pix* structure with two changes is a better choice for extracting building outlines.

6. Conclusion and future work

The extraction of building outlines is a critical step in the generation of building footprint. However, traditional methods have the parameter selection problem. To solve the problem of existing methods and reduce the difficulty of the next regularization step to improve the accuracy of building footprints, a new method for the extraction of building outlines is proposed in this study. This method achieves an automatic and more accurate extraction of building outlines from segmented point cloud data of building roofs. By introducing GAN after the binary image generation of point cloud data, the information loss and pore spaces in the transformation from the point cloud data to the image are recovered, thereby improving the performance of extracting the building outline. The residual block and new skip connection position are introduced in our GAN structure to improve its performance for this specific task. The proposed method is trained and

evaluated on two segmented point cloud datasets with different point densities. The quantitative experimental result of our method includes the mIoU of 93.52%, the HD of 0.640 m and the PoLiS of 0.165 m. The qualitative and quantitative experimental results demonstrate that our method with an improved GAN not only effectively and efficiently solves the problem of parameter selection in the existing methods, but also improves the shape accuracy of the extracted building outline, which provides a better basis for the building footprint generation.

The extracted outline polygons retain many details of the building shape and should be further regularized and generalized to generate the final building footprint polygons; this is the boundary regularization work in the next step of the building footprint generation. Hence, in the future, we will attempt to find an automatic and parameter-free method for boundary regularization and combine this method with that proposed in this study, to achieve full automation of building footprint generation. This research will also be combined with research on roof plane segmentation and façade parsing to further achieve automatic 3 D building reconstruction in LoD2 and LoD3 from raw geographic data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work in this paper is supported by NTNU Digital project (project No. 81771593) and NFR project HELIOS (project No. 324243).

ORCID

Hongchao Fan http://orcid.org/0000-0002-0051-7451
Gabriele Lobaccaro http://orcid.org/0000-0003-1603-3520

Data availability statement

The authors confirm that the data and source codes used in this study can be available upon request.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, et al. 2016. TensorFlow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). Presented at the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), p. 265–283.

Albers B, Kada M, Wichmann A. 2016. Automatic extraction and regularization of building outlines from airborne LiDAR point clouds. ISPRS - Int Archiv Photogrammetr Remote Sens Spatial Inform Sci XLI-B3:555–560.

Avbelj J, Müller R, Bamler R. 2015. A metric for polygon comparison and building extraction evaluation. IEEE Geosci Remote Sensing Lett. 12(1):170–174.

Awrangjeb M. 2016. Using point cloud data to identify, trace, and regularize the outlines of buildings. Int J Remote Sens. 37(3):551–579.

Awrangjeb M, Fraser CS. 2014. An automatic and threshold-free performance evaluation system for building extraction techniques from airborne LIDAR data. IEEE J Sel Top Appl Earth Observations Remote Sens. 7(10):4184–4198.

Berg MT, de Kreveld MV, Overmars M, Schwarzkopf O. 2000. Computational geometry: Algorithms and applications. Berlin, Heidelberg: Springer Science & Business Media.
Bittner K, Adam F, Cui S, Körner M, Reinartz P. 2018. Building footprint extraction from VHR remote sensing images combined with normalized DSMs using fused fully convolutional networks. IEEE J Sel Top Appl Earth Observations Remote Sens. 11(8):2615–2629.
Canny J. 1986. A computational approach to edge detection. IEEE Transact Pattern Anal Machine Intell, PAMI. 8(6):679–698.
Cao R, Zhang Y, Liu X, Zhao Z. 2017. 3D building roof reconstruction from airborne LiDAR point clouds: a framework based on a spatial database. Int J Geograph Inform Sci. 31(7):1359–1380.
Cazals F, Giesen J, Pauly M, Zomorodian A. 2005. Conformal alpha shapes. In: Proceedings Eurographics/IEEE VGTC Symposium Point-Based Graphics, 2005. p. 55–61.
Cramer M. 2010. The DGPF-test on digital airborne camera evaluation overview and test design. PFG. 2010(2):73–82.
Dai Y, Gong J, Li Y, Feng Q. 2017. Building segmentation and outline extraction from UAV image-derived point clouds by a line growing algorithm. Int J Digital Earth. 10(11):1077–1097.
Dorninger P, Pfeifer N. 2008. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. Sensors (Basel). 8(11): 7323–7343.
dos Santos RC, Galo M, Carrilho AC. 2018. Building boundary extraction from LIDAR data using a local estimated parameter for alpha shape algorithm. Int Arch Photogramm Remote Sens Spatial Inf Sci. XLII-1:127–132.
dos Santos RC, Galo M, Carrilho AC. 2019. Extraction of building roof boundaries from LiDAR data using an adaptive alpha-shape algorithm. IEEE Geosci Remote Sensing Lett. 16(8):1289–1293.
Edelsbrunner H, Mücke EP. 1994. Three-dimensional alpha shapes. ACM Trans Graph. 13(1):43–72.
FKB-Buildings Dataset. 2021. [online], 2021. [accessed 2022 Mar 5]. https://kartkatalog.geonorge.no/metadata/fkb-bygning/8b4304ea-4fb0-479c-a24d-fa225e2c6e97.
Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. 2014. Generative adversarial nets. In: Advances in neural information processing systems. Montreal, Canada: Curran Associates, Inc.
Guo H, Du B, Zhang L, Su X. 2022. A coarse-to-fine boundary refinement network for building footprint extraction from remote sensing imagery. ISPRS J Photogramm Remote Sens. 183:240–252.
He Y, Zhang C, Fraser CS. 2014. An energy minimization approach to automated extraction of regular building footprints from airborne LiDAR data. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Presented at the ISPRS Technical Commission III Symposium (Volume II-3); p. 5–7. September 2014, Zurich, Switzerland, Copernicus GmbH, 65–72.
He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 770–778.
Herve L. 2008. Outlining buildings using airborne laser scanner data [master thesis]. Netherlands: ITC.
Huttenlocher DP, Klanderman GA, Rucklidge WJ. 1993. Comparing images using the Hausdorff distance. IEEE Trans Pattern Anal Machine Intell. 15(9):850–863.
Isola P, Zhu J-Y, Zhou T, Efros AA. 2017. Image-to-image translation with conditional adversarial networks. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 1125–1134.
Jarvis R. 1977. Computing the shape hull of points in the plane. Proceedings of the IEEE Computing Society Conference on Pattern Recognition and Image Processing. p. 231–241.
Jochem A, Höfle B, Wichmann V, Rutzinger M, Zipf A. 2012. Area-wide roof plane segmentation in airborne LiDAR point clouds. Comput Environ Urban Syst. 36(1):54–64.
Kingma DP, Ba J. 2015. Adam: A method for stochastic optimization. In: Bengio Y, LeCun Y, editors. 3rd International Conference on Learning Representations, ICLR 2015. San Diego, CA.
Li Q, Mou L, Hua Y, Shi Y, Zhu XX. 2022. Building footprint generation through convolutional neural networks with attraction field representation. IEEE Trans Geosci Remote Sens. 60:1–17.
Li X. 2020. Automatic building footprint generation from airborne lidar point cloud [Doctoral diss]. The University of Texas at Dallas.
Li Y, Gao G, Cao B, Zhong L, Liu Y. 2015. Building boundaries extaction from point clouds using dual-threshold Alpha Shapes. In: 2015 23rd International Conference on Geoinformatics. Presented at the 2015 23rd International Conference on Geoinformatics, 1–4.

Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440.

Mahphood A, Arefi H. 2017. A data driven method for flat roof building reconstruction from lidar point clouds. Int Archiv Photogram Remote Sens Spatial Inform Sci XLII-4/W4:167–172.

Mahphood A, Arefi H. 2022. Grid-based building outline extraction from ready-made building points. Autom Constr. 139:104321.

Poullis C, You S. 2009. Automatic reconstruction of cities from remote sensor data. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2775–2782.

Sampath A, Shan J. 2007. Building boundary tracing and regularization from airborne LIDAR point clouds. Photogramm Eng Remote Sens. 73(7):805–812.

Shackelford AK, Davis CH, Wang X. 2004. Automated 2-D building footprint extraction from high-resolution satellite multispectral imagery. In: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, 1996–1999.

Shahzad M, Zhu XX. 2015. Reconstruction of building footprints using spaceborne tomosar point clouds. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci. II-3/W5II-3/W5:385–392.

Sugihara K, Zhou X, Murase T. 2012. Knowledge-based system for automatic 3D building generation from building footprint. In: Watanabe T, Watada J, Takahashi N, Howlett RJ, Jain LC, editors. Intelligent interactive multimedia: Systems and services. Berlin, Heidelberg: Springer, p. 363–373.

Suzuki S, Abe K. 1985. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Process. 30(1):32–46.

Verma V, Kumar R, Hsu S. 2006. 3D building detection and modeling from aerial LIDAR data. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Presented at the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2213–2220.

Wang O, Lodha SK, Helmbold DP. 2006. A Bayesian approach to building footprint extraction from aerial LIDAR data. In Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06). Presented at the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), p. 192–199.

Wang J, Shan J. 2009. Segmentation of LIDAR point clouds for building extraction. In: American Society for Photogramm. Remote Sens. Annual Conference. Presented at the American Society for Photogramm. Remote Sens. Annual Conference, Baltimore, MD, 1–13.

Zhou Q-Y, Neumann U. 2009. A streaming framework for seamless building reconstruction from large-scale aerial LiDAR data. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2759–2766.

Zhu Q, Liao C, Hu H, Mei X, Li H. 2021. MAP-Net: Multiple attending path neural network for building footprint extraction from remote sensed imagery. IEEE Trans Geosci Remote Sens. 59(7):6169–6181.