International Journal of Geo-Information

MDPI

Article

MAC-GAN: A Community Road Generation Model Combining Building Footprints and Pedestrian Trajectories

Lin Yang [1], Jing Wei [1], Zejun Zuo [1,*] and Shunping Zhou [1,2]

[1] School of Computer Science, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China
[2] National Engineering Research Center of Geographic Information System, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China
* Correspondence: zjzuo@cug.edu.cn

Abstract: Community roads are crucial to community navigation. There are automatic methods to obtain community roads using trajectories, but the sparsity and uneven density distribution of community trajectories present significant challenges in identifying community roads. To overcome these challenges, we propose a conditional generative adversarial network (MAC-GAN) supervised by pedestrian trajectories and neighborhood building footprints for road generation. MAC-GAN packs the "road trajectory–building footprint" pairs into images to characterize implicit ternary relations and sets up a multi-scale skip-connected and asymmetric convolution-based generator to incorporate such a relationship, in which the generator and discriminator mutually learn to optimize the network parameters and then derive approximate optimal results. Experiments on 37 real-world community datasets in Wuhan, China, are conducted to verify the effectiveness of the proposed model. The experimental results show that the F1 score of our model increases by 1.7–6.8%, and the IOU of our model increases by 2.2–7.5% compared with three baselines (i.e., Pix2pix, GANmapper, and DLinkGAN (configured by DLinknet)). In areas with sparse and missing trajectory data, the generated fine roads have high accuracy with the supervision of building footprints.

Keywords: community road; road extraction; deep learning; conditional generative adversarial network

Citation: Yang, L.; Wei, J.; Zuo, Z.; Zhou, S. MAC-GAN: A Community Road Generation Model Combining Building Footprints and Pedestrian Trajectories. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 181. https://doi.org/10.3390/ijgi12050181

Academic Editors: Maria Antonia Brovelli and Wolfgang Kainz

Received: 28 February 2023
Revised: 17 April 2023
Accepted: 20 April 2023
Published: 25 April 2023

1. Introduction

Community roads usually surround residential buildings, including drivable road types (e.g., mixed traffic lanes) and non-drivable road types (e.g., pedestrian paths and walkways). They are essential bridges linking people (e.g., visitors, deliverymen, etc.) and the neighborhood environment, which can help people avoid getting lost or detours during their daily commutes, and are critical to community navigation and other downstream-related applications [1].

Among multimodal information for automatic road extraction, remote sensing images and GPS trajectory are the widely accessible data sources. Existing road extraction tasks mainly focus on urban-scale roads, and the above two data sources are usually effective for this task. However, there are significant challenges when these data are used for road extraction tasks in community micro-spaces.

(i) For remote sensing data, the internal environment of residential communities is complex, and many types of elements, such as buildings, trees, and lawns, obscure the confined branch roads in the remote sensing images of the communities, thus posing a challenge for extracting fine community branch roads in community space.
(ii) For trajectory data, the trajectories in community space mainly consist of pedestrian trajectories mixed with a small number of vehicle trajectories due to the restrictions on the entry of foreign vehicles into the residential area and the common design of people–vehicle diversion. Pedestrian trajectories recorded by various mobile phone apps only exist within a limited time and space range with the characteristics of

low frequency. Firstly, such low-frequency trajectories lacking details of real paths would probably result in an ambiguous representation and undoubtedly exacerbate the difficulties of road generation. Secondly, such trajectories imprint mixed traffic flow for different agents (e.g., pedestrians and vehicles). The hybrid features in community-like spaces greatly degrade the performance of current methods. Thirdly, unlike the driving behavior of vehicles on general roads, pedestrians walk freely in the community, such as across lawns, squares, etc. This may result in trajectories with non-uniform density distribution, and random sampling may further exacerbate this problem. Existing trajectory-based methods of road extraction have limitations in dealing with such community trajectory data.

(iii) For community space, community roads are tighter and denser (i.e., adjacent roads are closer in space) than urban-scale roads. The features adopted by existing methods are insufficient for extracting the mixed staggered roads with different levels from low-frequency trajectories. In addition, the GPS drifts caused by dense tall buildings and residents induce spatial uncertainties and increase the difficulty of distinguishing adjacent roads.

Against this backdrop, we develop a MAC-GAN model to derive visually and morphologically realistic community roads from GPS tracking data and building footprint datasets. We fully consider the relationship between different features in the community-built environment, especially the relationship between building footprints and roads. On the one hand, the trajectory that records the location information of road traffic activities can reflect the geometric structure and topological relationship of roads to a certain extent. Note that in a community-like compact space, buildings and roads are the two main geographical elements closely related to people's daily commute. According to Poincare's dual characteristics, building footprints have rich road-related context information, which can compensate for the defects of density difference and incompleteness of GPS trajectory. Therefore, with the help of adjacent roads and surrounding buildings, roads in the residential scene can be inferred more precisely to supplement a more realistic community road network and enrich alternatives for community navigation.

The main contributions of this study are summarized as follows:

(1) We combine trajectory information with rich road geometry and topological features and building footprints with road contextual spatial information to enrich the research dimension of road extraction methods.

(2) We propose a generative adversarial model named MAC-GAN for community road extraction. We configure the generator MACU-Net for MAC-GAN, which has cross-perceptual field convolution blocks to enhance the attention to and perception of road space neighborhoods. It builds skip connection and adaptive attention mechanisms to fuse multi-scale features. MACU-Net captures the ternary features of the "road trajectory–building footprint" for generating roads with sparse and uneven trajectories.

(3) We explore a new geodata transformation application of GANs on a community scale to transform a coarser and accessible geospatial dataset (trajectories and building footprints) into another geospatial dataset (roads), and we verify the feasibility and effectiveness of this application to generate road data.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 elaborates on the structures of the proposed model (MAC-GAN). Section 4 presents the study dataset, baseline models, and evaluation metrics. Section 5 discusses the experimental results. Section 6 concludes the study.

2. Related Work

2.1. Road Generation Methods Based on Trajectory

Trajectory clustering methods are commonly utilized to generate roads. The turning points are treated as intersections, and the trajectory segments are clustered into a road segment [2,3]. For instance, Xie et al. [4] utilized the clustering of turning points on GPS

trajectories to identify potential intersections and then a point-by-point alignment strategy to determine the road geometry between these intersections. Zhang et al. [5] utilized a combination of K-means clustering and Gaussian modeling to recognize road centerlines and lanes. However, the scarcity of crowd-sourced trajectories prevents them from being modeled as a Gaussian distribution, resulting in poor clustering outcomes.

Raw noise data pose a challenge to road generation algorithms. Numerous scholars have proposed strategies of suppressing noise by the incremental merging of trajectories [3,6–9]. Based on the constraints of the geometric and movement characteristics (i.e., Fréchet distance, speed, and direction) between trajectory points, a graph representing the road network is incrementally generated, and the road sections with fewer trajectory data points are trimmed to remove some of the noise data [2,7,10,11]. Cao and Krumm [7] simulated physical forces to optimize GPS trajectories by pulling trajectory points closer to each other when they are in close proximity and prevent them from deviating too far from the road among the traces, which effectively suppresses the noise in the raw traces. The resulting graphics accurately display critical road connectivity and road geometry. However, it cannot make the GPS traces with significant errors and spatially scattered traces into tight bands. Both noisy residual trajectories and unevenly distributed trajectories lead to the extraction of false roads.

In addition, methods based on graph theory and image processing have been developed. Density analysis (e.g., kernel density estimation, Morse theory, improved sliding methods, etc.) is used to convert GPS trajectories into discretized images (i.e., density surfaces) [12–14], and then stable manifolds from the image are extracted for road skeletonization algorithms and topology refinement [15–17]. For example, Biagioni and Eriksson [18] suggested an incremental approach based on kernel density estimation (KDE) that involves obtaining a road sketch via gray-scale skeletonization, trimming it utilizing Viterbi map matching, and refining its topology and geometry. This image-based approach overcomes the disadvantage of insufficient GPS trajectory sampling, because it deals with trajectory points instead of GPS traces [19]. However, due to the uneven density of GPS trajectory points, it is hard to extract roads using a unified threshold.

In summary, the over-reliance on the quality of the input trajectories presents a challenge in the task of community road generation. How to deal with the nontrivial noise and unevenness of trajectories is always the key to affecting the quality of road data. It would be highly useful to develop new methods to reduce this direct dependence.

2.2. Generative Adversarial Networks and Geospatial Data Translation

Generative adversarial network (GAN) is a deep learning architecture in which generator G and discriminator D contest each other in a zero-sum game [20]. Mirza and Osindero [21] proposed a conditional generative adversarial network (CGAN), an extension of GAN, and its generators obtain more controllable outputs under the influence and control of input data.

The goal of domain mapping in deep learning is to acquire the mapping function that transforms the source domain into the target domain [22]. Many scholars actively try to use the domain mapping capability of CGAN to discover deep-level spatial features to generate geospatial data. For example, Dong et al. [23] designed a shadow-constrained CGAN (SCGAN) to fill in the gaps of mountainous SRTM data. Zhu et al. [24] designed CEDGANs, which can learn the spatial relationship between sample data and corresponding real spatial data to generate real DEMs with local topographic structural patterns similar to real images. Many studies also predicted missing data on urban infrastructure based on urban geospatial data [25–27]. It has been reported that using the domain knowledge of urban morphology and spatial networks, building heights can be derived from the street networks and buildings data of 2D urban morphology [25]. Taking advantage of the CGAN's capability, Wu and Biljecki [28] developed a GANmapper, which transforms data from road networks into building footprints. They demonstrate the powerful ability of CGAN to identify potential patterns in intricate geospatial contexts and the feasibility

of discovering deep features based on CGAN from the intertwined features in the built environment. This inspired us to generate community roads through CGAN using building footprints and trajectory data.

The generator of a CGAN for domain mapping is usually configured as an encoder–decoder architecture [29], where its encoder captures features from external information (conditions) and its decoder utilizes these captured features to generate the ultimate output image. The design of the encoder–decoder architecture plays a key role in determining whether the CGAN will accurately perform the domain mapping task. Numerous researchers have improved the encoder–decoder structure to achieve geospatial information mining. The encoder–decoder architecture U-Net is a common network configuration scheme for road semantic segmentation. Its several skip connections between the encoder and decoder are used to improve the semantic extraction capability within the framework [30,31]. However, insufficient utilization of information flow impedes its potential. Wu and Biljecki [28] incorporated nine residual blocks [32] into the middle of an encoder–decoder architecture to increase the generator network's depth, which makes the generative model more expressive in geographical content translation tasks. However, the omission of skip connections in this structure may fail to extract the geospatial features sufficiently. Zhou et al. [33] suggested a D-LinkNet for road segmentation, which integrates skip connections and residual blocks. In addition, the dilated convolution layers are included between the encoder and decoder. This facilitates the expansion of the perceptual domain and integration of multi-scale features. However, D-LinkNet still suffers from wrong recognition and weak connectivity. The raw U-Net, resnet_9block, and D-LinkNet cannot fully extract multi-scale road features. U-Net 3+'s full-scale skip connections try to address this issue, but at the same time, a tremendous amount of computation are required [34]. Considering the complexity of community scenarios, extracting road features from building and trajectory data is a challenge in terms of capturing the spatial features implied in the deep layer as comprehensively, accurately, and with as little redundancy as possible.

3. Methodology

3.1. Framework

Inspired by the ability of CGANs to implement domain mapping to generate reliable synthetic data, we explore using CGANs to generate spatially correct synthetic road datasets using learning cues from building footprint and pedestrian trajectory datasets. We propose a MAC-GAN based on CGAN for generating roads in community scenarios using pedestrian trajectories and building footprints. Since this paper is an initial attempt to generate road spatial data using CGAN for domain mapping, the input, as the initial condition for domain mapping, largely affects the learning ability of the model, so we try to investigate the most suitable input for this task. We design the method framework as shown in Figure 1. The first step is data collection. We collect the mobile GPS trajectory data and the community center's latitude and longitude coordinate data. The second step is data pre-processing. We create a pre-processing pipeline to produce three types of input–target image datasets for the model: building footprints, pedestrian trajectories, and building overlay trajectories. The third and fourth steps are model prediction and post-processing, respectively. After obtaining a series of resultant images generated by the MAC-GAN model, we post-process the images, that is, filter and stitch these images to finally obtain an overall image of the community.

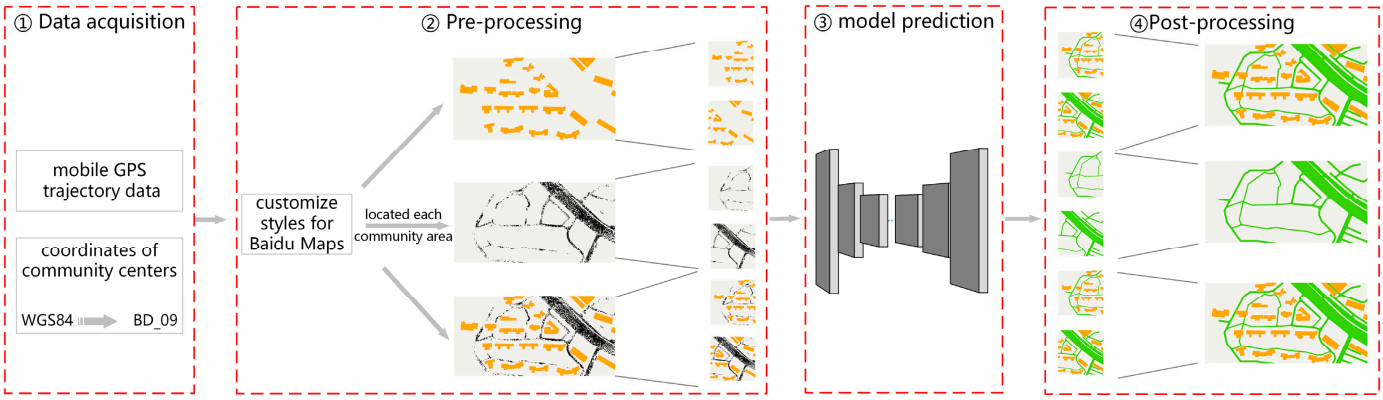


Figure 1. The framework of our method.

3.2. Pre-Processing

After collecting the mobile phone GPS trajectories and coordinates of community centers, we obtain the image dataset through the following pre-processing. To avoid the impact of other types of map elements (POI, greens, etc.) on evaluation, we customized three map styles through the Baidu Maps API: buildings only, trajectories only, and building-overlaid trajectories. Images in three map styles were used to test the optimal input–target pairs configuration. Building footprints, trajectories, and roads were represented by orange, black, and green pixels, and road levels were distinguished by the line width. Due to the small building footprints and the linear geometric shape of roads in the community, to capture the contextual features while maintaining high visual clarity, we located each community area in the custom-style Baidu map through the latitude and longitude coordinates of the community center. We grabbed a 19-level map image with a resolution of 1024×512 for each specific community area, and then divided it into two images, on average, and uniformly zoomed them to the resolution of 256×256.

3.3. Model Architecture

Figure 2 shows the structure of MAC-GAN. In order to make the generator capture accurate and sufficient spatial features of the road in the complex community environment, we configure MACU-Net as the generator (G). In each forward pass, it converts the conditional image into a road image by fully capturing the features of the conditional image (building footprint overlaid with trajectories) and fools the discriminator (D) to label the generated image as "real", while D is trained to label the generated image as "fake" and the target image as "real". The D is configured by the Markov discriminator (PatchGAN) [35]. The generated image and the target image are connected to the conditional image, and their authenticity is calculated through convolution operation. Then, the adversarial loss between the generator and the discriminator and the L1 loss are calculated. Following this, the generator and discriminator weights are updated via the backpropagation algorithm. This process repeats until the discriminator can no longer distinguish between the target and the generated image.

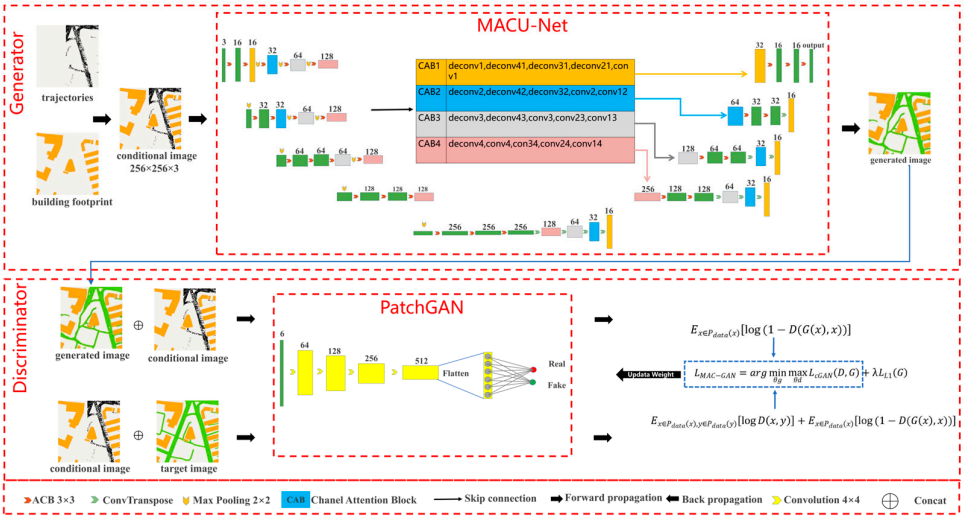


Figure 2. The structure of the MAC-GAN. In MACU-Net, features with the same size and dimensions are marked with the same color (blue, gray, orange, and red), and features with the same color will be connected and fed to the CAB module.

3.3.1. Generator

Given the excellent performance of MACU-Net in solving the problem of the under-utilization of image space features [36], especially with its two core components that can meet the requirements of the practical application scenario (road generation) of this study well, the two core components of MACU-Net are integrated into our encoder–decoder structure and configured as MAC-GAN's generator. The following describes how the two core components work in the road generation task.

1. Asymmetric convolutional block (ACB). Existing neural network models typically extract features within the square window using the square convolution kernel, which is feasible for most block-shaped objects and spatial region blocks. However, the community roads in our study are narrow strips whose directions mainly extend vertically and horizontally. Using the square convolutional kernel can hardly focus on extracting linear features of the roads. In addition, the community has many road intersections, most of which are in the shape of "crossings" and "T-junctions." Due to the lack of direction sensitivity in square convolutional kernels, they cannot concentrate on extracting road information in different directions at intersections and geometric shape features of intersections. The nonlinear features of roads are generally manifested as complex geometric shapes such as curves, loops, and irregular edges. It is difficult for square convolution kernels to adequately extract nonlinear features of different scales and shapes. Moreover, the importance of features captured by square convolution kernels is heterogeneous. Specifically, the central crossover location contributes more information to feature extraction and less to the corners [36], which will cause the information extracted by the square convolution to be redundant and unrepresentative, further weakening the model's ability to extract nonlinear features. To overcome these limitations of the square convolution kernel, we choose the three-branch convolutional block (ACB) with cross-receptive fields shown in Figure 3 to extract the spatial features of the community roads. As shown in Figure 4, in ACB, the 3×3 convolution kernel is used to capture the contextual information of the road, and the 1×3 and 3×1 convolution kernels pay attention to capturing the road's linear characteristics, the intersection's geometry, and the representative linear and non-linear features at the skeleton. Thus, ACB reduces the capture of redundant information, ensures the extraction of essential road and intersection features, enhances the extraction of representative nonlinear features, and maintains sensitivity to contextual spatial features. The ACB is expressed as follows:

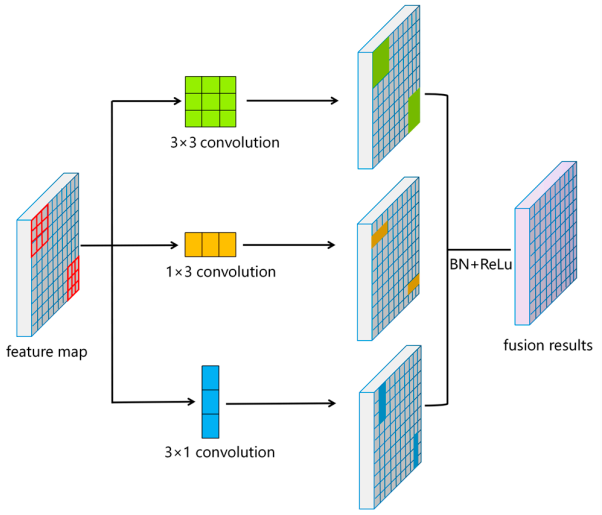
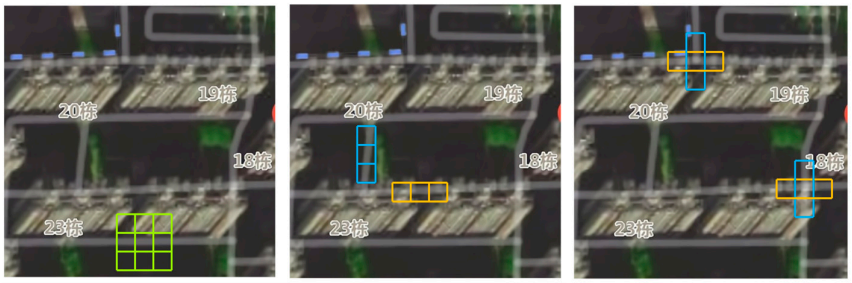


Figure 3. The structure diagram of the ACB with cross-receptive field.



(**a**) (**b**) (**c**)

Figure 4. The motivation of the asymmetric convolutional block ACB that we proposed. The three figures show the ACB convolution on the community image, and the texts in the image indicate the building units of the community. (**a**) The 3×3 convolution is used to capture the contextual information of the road. (**b**) The 1×3 and 3×1 convolutions capture features in the road's horizontal and vertical extension directions. (**c**) The 1×3 and 3×1 convolutions capture the features of road intersections.

$$\widetilde{x_i} = \mathrm{Conv}_{3\times3}(x_{i-1}) + \mathrm{Conv}_{1\times3}(x_{i-1}) + \mathrm{Conv}_{3\times1}(x_{i-1}), \tag{1}$$

$$x_i = \sigma\left(\gamma_i \frac{\widetilde{x_i} - \mathrm{E}(\widetilde{x_i})}{\sqrt{\mathrm{V}(\widetilde{x_i}) + \in_i}} + \beta_i\right), \tag{2}$$

where x_i and x_{i-1} are the output and input of ACB, respectively. The batch norm (BN) with γ and β parameters in Formula (2) is employed to improve the stability of the summation value of the three branches. The rectified linear unit $\sigma(\cdot)$ is used for nonlinear transformations. $\mathrm{V}(\cdot)$ and $\mathrm{E}(\cdot)$ denote the functions of variance and expectation. $\in$ represents a small constant.

2. Multiscale features skip connection and fusion. Considering that the insufficient information flow extraction and utilization limit the original U-Net architecture's potential, we incorporate multi-scale jump connections into the U-Net to facilitate interaction between encoders and decoders and to fully capture fine-grained road location, geometric and topology features, and coarse-grained semantic features. Figure 4 shows how X_{De}^{3} generates its feature map. The first step is the multi-scale features skip connection. Firstly, the same-level encoder layer's (i.e., X_{En}^{3}'s) feature maps are concatenated. Subsequently, the transposed convolution and ACB transmit the lower-level decoder layers' (i.e., X_{De}^{4}'s and X_{De}^{5}'s) fine-grained road geometry and topology features. Finally, the max pooling layer and ACB deliver the higher encoder

layers' (i.e., X^1_{En}'s and X^2_{En}'s) coarse-grained road semantic information. This process can be expressed as follows:

$$X^3_{De} = CAB\left(\left[ACB\left(D\left(X^k_{En}\right)_{k=1}^{2}\right), ACB\left(U\left(X^k_{De}\right)_{k=4}^{5}\right), X^3_{De}\right]\right), \quad (3)$$

X^1_{De}, X^2_{De}, X^4_{De}, and X^5_{De} are generated similarly. Their generation procedure can be uniformly formulated as follows:

$$X^i_{De} = \begin{cases} X^i_{En}, \ i = N \\ CAB\left(\left[ACB\left(D\left(X^k_{En}\right)_{k=1}^{i-1}\right), ACB\left(U\left(X^k_{De}\right)_{k=i+1}^{N}\right)\right]\right), \\ i = 1, \ldots, N-1 \end{cases} \quad (4)$$

where the channel attention block (CAB) is used to rearrange and fuse-channel related features, ACB(·) represents ACB, and U(·) and D(·) denote the up-sampling and down-sampling operations, respectively. [·] denotes concatenation. After connecting five feature maps of the same channels and resolution, we use channel attention blocks (CAB) in the second step to reduce, rearrange, and weigh the channels to emphasize essential information and filter redundant features. As shown in Figure 5, taking X^3_{De} as an example, this step consists of reducing the number of channels, compressing the spatial dimension by average pooling and max pooling, recovering the original number of channels, and generating the final output X^3_{De} through the ACB convolution layer. Similarly, the corresponding CAB generates X^4_{De}, X^2_{De}, and X^1_{De}.

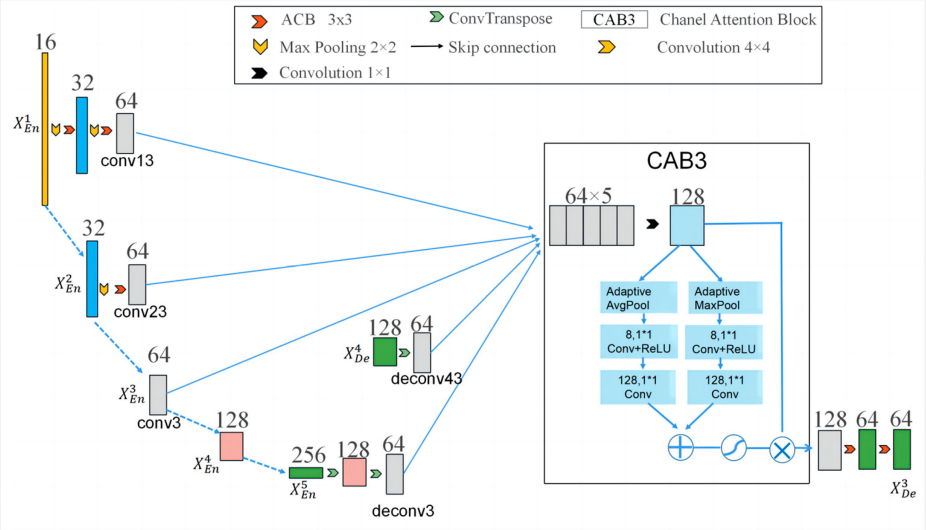


Figure 5. The construction of feature map X^3_{De}.

The following is the entire workflow of the generator. In encoding, the encoder first receives a conditional image of 256 × 256 × 3; the feature maps are extracted by using two stacked 3 × 3 ACB convolution layers and a 2 × 2 maximum pooling following an ACB convolution step. With the ACB convolution layer and the max pooling layer, the fine-grained features of different sizes and dimensions contained in each layer of the encoder can be extracted, as shown by the blue, gray, and red squares in the left and middle parts of the MACU-Net, where patches of the same color represent features of the same size and dimensionality. Coarse-grained features of different sizes and dimensions contained in each decoder layer are obtained by applying a 2 × 2 transpose convolution to the feature map of each decoder layer. In decoding, we utilize multi-scale skip connections to connect road geometry, topology, context, and semantic features from the same-level encoder, lower-level encoder, and higher-level decoder layers and feed them into the channel attention block

(CAB) to rearrange and optimize. Each decoder layer will output feature maps sequentially until the final decoder layer generates images filled with road data.

3.3.2. Discriminator

The network structure of MAC-GAN's discriminator is shown in Figure 1. It is PatchGAN [35] with a 70×70 receptive field, which performs binary classification on each image patch instead of the whole, so it can better capture the local features and details of the roads. The five convolutional layers of PatchGAN process the input image into a $1 \times 1 \times 512$ tensor, which is then squashed into a one-dimensional array and transmitted to the fully connected layer. Finally, the network outputs the probability of the input image being true or false.

3.3.3. Loss Function

Designing the loss function is a critical link between the training and optimization of the MAC-GAN model. Our goal is to train a generator, G, that can learn to convert a conditional image x into the generated image G(x) similar to the target image y. At the same time, we also need to train a discriminator D to distinguish G(x) from y. For this purpose, we design the objective function $L_{MAC-GAN}$ as shown in Formula (5), which incorporates the adversarial loss of the G and D calculated by the binary cross-entropy $L_{cGAN}(D, G)$ expressed in Formula (6), and the L1 loss $L_{L1}(G)$ expressed in Formula (7). The $L_{L1}(G)$ calculated by the L1 norm denotes the difference between the generated image G(x) and the target image y for each pixel. We set it as part of the loss to constrain the image generation so that G(x) is as consistent as possible with y. λ is a parameter adjusting the importance of $L_{L1}(G)$, E(*) is the expectation of the distribution function, $P_{data}(*)$ represents the data distribution, and $||*||_1$ denotes the L1 norm.

$$L_{MAC-GAN} = L_{cGAN}(D, G) + \lambda L_{L1}(G) \tag{5}$$

$$L_{cGAN}(D, G) = E_{x \in P_{data}(x), y \in P_{data}(y)}[\log D(x, y)] + E_{x \in P_{data}(x)}[\log(1 - D(G(x), x))] \tag{6}$$

$$L_{L1}(G) = E_{x \in P_{data}(x), y \in P_{data}(y)}[|| y - G(x)||_1] \tag{7}$$

4. Study Dataset and Evaluation

4.1. Study Dataset

We randomly selected 412 communities in Wuhan, China, and marked them with green dots in Figure 6. Only four of the communities are shown in the enlarged view. Experimental community centers' latitude and longitude coordinates in the WGS-84 coordinate system were collected and converted to the same BD_09 coordinate system as Baidu Maps. GPS trajectory data used for the experiment are mobile phone GPS data collected on 8 March 2019 in the experimental community area. The sampling frequency is between 1 min and 1438 min with a mean value of 45 min and a median value of 5 min, and each record is composed of the following fields: day, time, ID, longitude, and latitude (Table 1 shows an example). Using the pre-processing data methods described in Section 3.2, we collected 824 images from these communities, which have different road layouts and building structures. In total, 750 images are captured as the training sets and 74 images are captured as the validation set for evaluating the effectiveness of the proposed model.

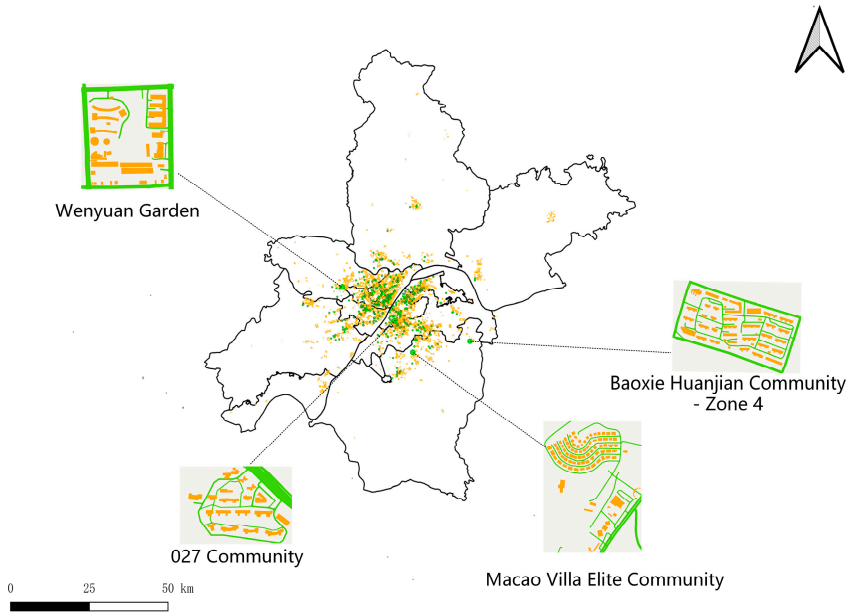


Figure 6. The study communities in Wuhan, China.

Table 1. An example of a GPS trajectory record.

Day	Time	Id	Longitude	Latitude
8 March 2019	23:31:00	7	114.310290	30.523696

4.2. Baselines and Settings

This study selects three generative models as baselines for evaluation, namely Pix2pix, GANmapper, and DLink-GAN. Pix2pix [35] is the original image-to-image CGAN that learns image mapping from input to output. GANmapper and DLink-GAN are constructed by redesigning the generator based on Pix2pix.

1. Pix2pix configures a U-Net generator whose skip connection improves the semantic extraction capability of the encoder–decoder framework. U-Net has become a standard scheme to capture nonlinear and hierarchical features of input images to reconstruct images, so we take it as one of the baseline models.
2. GANmapper's generator is configured as an encoder–decoder, which includes nine residual blocks (He et al., 2016). With the setting of the residual blocks, the loss of spatial information from down-sampling that cannot be restored by up-sampling is reduced to some extent.
3. DLink-GAN configures D-LinkNet [33] as the generator. D-LinkNet is a common encoder–decoder for road segmentation and extraction. D-LinkNet uses ResNet34 [37] to replace the encoder of U-Net, which reduces the loss of spatial information from down-sampling. In addition, the central part of its encoder and decoder uses several skip connections. The dilated convolutional layers at the center obtain a larger receptive field, which can extract and retain detailed "trajectory–building footprint–road" triple information of spatial features.

In this experiment, the minimum batch number is set to 6. The Adam optimizer (Kingma and Ba, 2014) [38] is chosen. The exponential decay rate of the first-order and second-order moment estimation is set to 0.5 and 0.999, respectively. The weight decay is set to $5 \times e^{-4}$, and the initial learning rate is set to $2 \times e^{-4}$. We use the same learning rate for the first 150 training epochs and linearly decrease it to zero during the final 150 training epochs, and the model uses the $L_{MAC-GAN}$ given by Formula (4) as the total loss, where λ is set to 100. Python3.7 and the PyTorch framework are used for experiments. The proposed model is trained on a single GPU on an Ubuntu 18.04 (64-bit) system.

4.3. Evaluation Metrics

Road extraction can be treated as a task of classifying image pixels as either road (positive) or non-road (negative). In this research task, we counted the number of generated pixels in four categories: the true example (TruePositive, TP), the negative examples (TrueNegative, TN), false positive examples (FP), and false negative examples (FN). This study selects four common indicators: pixel precision (Precision), pixel recall (Recall), Intersection-over-Union (IOU), and F1-score. They are calculated using Formulas (8)–(11).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{IOU} = \frac{\text{TP}}{\text{FP} + \text{TP} + \text{FN}} \tag{10}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \tag{11}$$

The F1-score and IOU comprehensively consider both precision and recall in evaluating the model. Their higher values indicate that the generated image overlaps more pixels with the ground truth image. However, the generated images have higher IOU scores (close to 1), which may be caused by the generator over-fitting the training data. In addition, due to the generative characteristics of GAN, although its generated roads appear to be roughly consistent with the real road in terms of overall road structure and shape, there is typically not a complete pixel-by-pixel overlap. Therefore, this paper does not rely solely on the pixels of the generated image to evaluate the road generation effect.

The Fréchet inception distance (FID) [39] is a benchmark commonly used to evaluate the GAN's performance. It assesses the quality of the generative model using the Fréchet distance of the feature distribution of the generated and real images in Inception v3 [40] cyberspace. In our experiments, a lower FID score indicates that the generated community roads have more realistic shapes, sizes, distribution densities, and topologies. The FID can be represented as Formula (12):

$$\text{FID} = \left|\mu_r - \mu_g\right|^2 + \text{T}_r\left(\sum \text{r} + \sum \text{g} - 2\sqrt{\sum \text{r} \sum \text{g}}\right), \tag{12}$$

where the μ_r and μ_g refer to the averages of 2048 activation feature values obtained by inputting the real and generated images into the Inception v3 classification model, respectively, $\sum \text{r}$ and $\sum \text{g}$ are the covariance matrices for the real and generated feature vectors, and Tr represents the trace of the matrix.

5. Experimental Results and Analysis

5.1. Test for the Optimized Configurations of Input–Target Pairs

The input data serve as the critical initial condition of the generator of MAC-GAN. There are close spatial relationships between roads, trajectories, and neighboring buildings. Our method couples such relations to supervise the generation of community roads. We examine the effect of different input types on MAC-GAN's performance in predicting community roads and derive the most efficient input–target pair configuration for this research task. Figure 7 shows the three different input–target pairs adopted in this study.

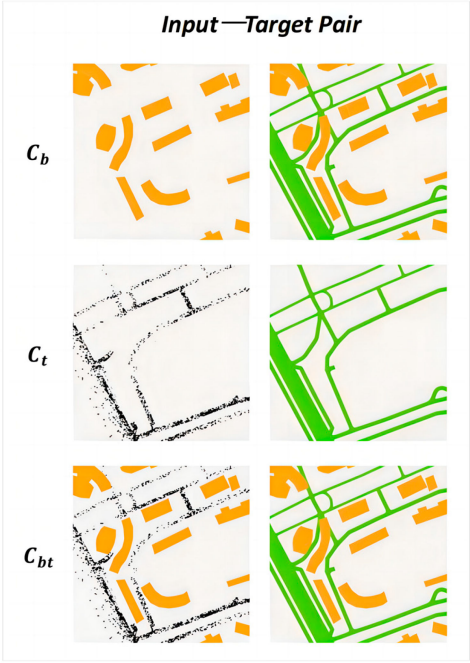


Figure 7. Examples of three different input–target pairs: (1) building footprints as input and building footprints overlaid with roads as a target (noted as *Cb*); (2) trajectories as input and roads as a target (noted as *Ct*); (3) building footprints overlaid with trajectories as input and building footprints overlaid with roads as a target (noted as *Cbt*).

We train MAC-GAN using the input–target pairs of three configurations, *Cb*, *Ct*, and *Cbt*, to obtain the corresponding models, labeled Model$_{C_b}$, Model$_{C_t}$, and Model$_{C_{bt}}$, respectively. Figure 8 shows the generated road images of these three trained models. Model$_{C_b}$ can learn the spatial structural relationship between roads and buildings and roughly predict community roads, especially the main roads of the community. However, it is difficult for Model$_{C_b}$ to accurately predict the geometric distribution and shape of refined branches in areas where buildings are sparsely distributed. As shown in the seventh row of Figure 8, the generated branches mostly appear as clumps and short discontinuous strips. This may be attributed to the weak spatial relationship between the fine branches and buildings. Specifically, it is very challenging for MAC-GAN to capture the characteristics and distribution of fine branches through the supervision of building footprint data only.

Model$_{C_t}$ generates continuous and accurately distributed roads in regions with complete trajectories. However, the generated roads are mostly wrong or missing in areas with sparse or significantly missing trajectories. There are even erroneous cases where the generated roads intersect with building footprints. This suggests that the trajectory, the directly relevant element of roads, is the essential guide for the geometric distribution and morphology of roads. In addition, the spatial information of the road neighborhood is indispensable to constraining the generation of the road's direction, distribution, shape, topology, etc. However, MAC-GAN trained with only trajectory data cannot capture this information.

Model$_{C_{bt}}$ generates roads more completely and accurately. Compared with Model$_{C_b}$, the generated roads have more regular shapes and higher integrity. Compared with Model$_{C_t}$ the generated roads' direction, shape, and topology are effectively constrained in areas with sparse and missing trajectories. This is attributed to the advantages of the two data sources of trajectory and building footprints; our model can fully combine the context features such as road geometry, topology, and building distribution of the neighborhood, thereby effectively constraining the generation of the roads.

Figure 8. Road images of nine areas (**a**–**i**) generated by Model$_{C_b}$, Model$_{C_t}$, and Model$_{C_{bt}}$.

Table 2 lists the evaluation metrics values of the results generated by Model$_{C_b}$, Model$_{C_t}$, and Model$_{C_{bt}}$. Model$_{C_{bt}}$ trained with building footprint-overlaid trajectories has the best performance. The accuracy and recall are 70.2% and 75.2%, respectively, and the FID is 92.20. Compared with Model$_{C_b}$ trained with building footprints, the F1 score and IOU are increased by 34.6% and 41.3%, respectively, and the FID is decreased by 22.09%. This indicates that the trajectory, as the directly correlated element of roads, can guide road generation to a great extent. In addition, compared with Model$_{C_t}$ only trained by trajectories, the recall of Model$_{C_{bt}}$ increases by 7.1%, and the F1 score and IOU increase by 3.1% and 3.2%, respectively. The building footprints containing rich context semantic information of roads play the role of neighborhood supervision, which makes the generated roads more complete. In terms of F1 score, IOU, and FID indicators, the performance of Model$_{C_{bt}}$ is the best. This indicates that *Cbt* is the best input–target pair configuration for road generation. Therefore, it will be chosen as the configuration of input–target pair of our model for later comparison and analysis.

Table 2. The evaluation metrics values of the results generated by Model$_{C_b}$, Model$_{C_t}$, and Model$_{C_{bt}}$.

Configuration	Model	Accuracy	Recall	F1 Score	IOU	FID
Cb	Model$_{C_b}$	0.356	0.339	0.341	0.212	114.29
Ct	Model$_{C_t}$	0.706	0.681	0.687	0.542	161.84
Cbt	Model$_{C_{bt}}$	0.702	0.752	0.718	0.574	92.20

5.2. Comparison with Baselines

This subsection compares MAC-GAN's result with three baseline models (Pix2pix, GANmapper, and DLink-GAN). All models are trained with the input–target pairs of the configuration *Cbt*. The six rows of images shown in Figure 9 are, in order, the inputs, the generated results of the four models, and the ground truths. As shown in the region marked

with red boxes in Figure 9, MAC-GAN can generate more realistic roads in complex areas with sparse or even missing trajectories compared to Pix2pix, GANmapper, and DLink-GAN.

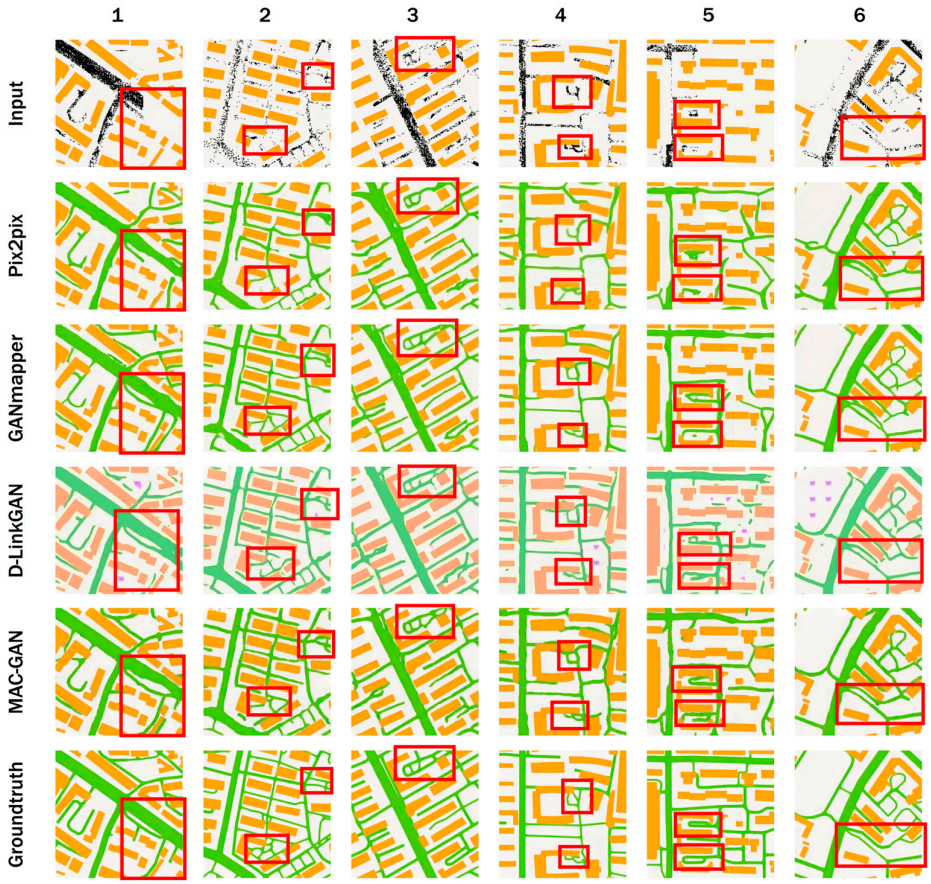


Figure 9. Results of MAC-GAN, Pix2pix, GANmapper, and DLink-GAN. Columns 1–6 in Figure 9 represent the six different study areas. In order, the six rows of images in columns 1–6 are the input, the generated results of the four models, and the ground truth. The area marked with a red box is the complex area with sparse or missing trajectories.

In the second row of Figure 9, there are some prediction errors and unsmooth road edges in the results generated by Pix2pix for the detailed roads. This is because U-Net's skip connections are only performed on feature maps of the same size, while the lack of multi-scale features as a complement makes the detailed features of the expression insufficient and incomplete. GANmapper performs better in regions with complete trajectories than regions with incomplete trajectories. For example, in scenes with complex building distribution and missing trajectories, as shown in the second row of the fifth column, it can roughly predict the roads in areas with trajectories. However, prediction errors and omissions exist in the areas with missing trajectories. The reason may be that GANmapper does not have multi-scale skip connections and feature fusion, resulting in insufficient consideration of the contribution of spatial features of neighborhood structures in road generation.

Compared with Pix2pix and GANmapper, DLink-GAN predicts road shape better but appears to have poorer topological continuity, as shown in the second and fifth columns of the fourth row. The reason may be that DLink-GAN uses dilated convolution to enlarge feature receptive fields in the central part of the network and integrates multi-scale features in the central part, which can better retain fine branch features. However, the information lost in the down-sampling module, and the dilated convolution pixels that do not participate in the computation lead to poor road topological continuity.

MAC-GAN uses an ACB with a cross-receptive field, which ensures the importance of contextual road features and effectively extracts key road features simultaneously. Notably,

it performs exceptionally well in ensuring the continuity of the road segments and intersections, as shown in the fifth and sixth columns of Figure 9. In addition, MAC-GAN replaces ordinary skip connections with multi-scale connections and a channel attention block (CAB), which can combine multi-scale information of trajectories and neighborhood buildings, thus making the generated roads more complete and have better topological continuity.

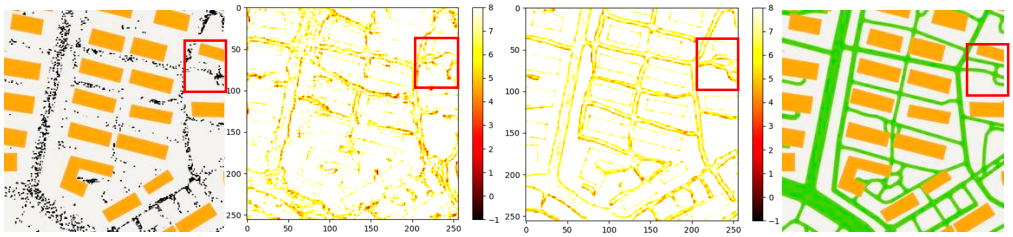
Table 3 lists five comprehensive evaluation indices for the four models. The F1 score and IOU values of MAC-GAN are better than the three baseline algorithms. MAC-GAN has an F1 score improvement of nearly 6.8% and IOU of almost 7.5% over Pix2pix. Compared with GANmapper, the F1 score and IOU gain is about 1.7% and 2.2%. Compared with the DLink-GAN, the profits of the F1 score and IOU increased by 2.4% and 3.2%, respectively. Compared with the three baselines, the profits of indicators indicate that introducing the ACB and the multi-scale connection with the channel attention block (CAB) is effective. It ensures the importance of the features in the central cross area and the multi-scale combination of the trajectory's location and neighborhood structure to road generation. Thus, MAC-GAN can produce results more consistent with the ground truth.

Table 3. Comparison of the five evaluation indices.

	MAC-GAN	**Pix2pix**	**GANmapper**	**DLink-GAN**
Accuracy	0.702	0.668	0.778	0.809
Recall	0.752	0.499	0.552	0.613
F1 score	0.718	0.650	0.701	0.694
IOU	0.574	0.499	0.552	0.542
FID	92.20	86.61	77.92	100.17

5.3. Effect of ACB Block

To verify the impact of ACB, we compare and analyze the performance of three-branch convolution blocks (ACB) and square convolutions. The ACB convolution block and square convolution are used to extract road features. Then, the feature map of the last up-sampling layer (conv9) is generated to visually compare the difference between the two kinds of blocks. As shown in Figure 10, each pixel in the heatmap represents the sum of the pixel values of all channels in the output tensor, and the color bar on the right represents the correspondence between pixel values and colors. The results indicate that the ACB module can significantly improve the feature extraction ability of the model. Specifically, even in instances where pedestrian trajectories are messy and sparse, it can still extract the linear geometry and regular boundary of the road. However, the square convolution kernel without cross-receptive fields cannot pay special attention to the linear features of the road, resulting in inaccurate extracted features. It is of interest that the ACB module outperforms the square convolution in capturing nonlinear road geometry features (e.g., curved and intersecting road regions), as shown in the red boxed area of Figure 10. This is because skeleton locations can contribute more information to feature extraction than corner points [36]. Although the 3×1 and 1×3 convolution kernels cannot fit all types of road orientation and geometric features in realistic scenes, the extra focus of the three-branch convolution block on the skeleton enables it to capture more representative nonlinear road features than square convolution.



(**a**) (**b**) (**c**) (**d**)

Figure 10. Feature visualization results of the last up-sampling layer (conv9), non-linear road features are selected in the red box. (**a**) The input image of the network model. (**b**) Feature visualization results of the conv9 layer extracted by the ACB convolution block. (**c**) Feature visualization results of the conv9 layer extracted by the square convolution. (**d**) Ground truth.

5.4. Impact of Trajectories of Different Sparseness and Missing Degrees

MAC-GAN learns basic features from the trajectory and building footprints for a given sample to predict unknown roads. However, as the degree of sparseness or absence of trajectories intensifies, the feature of the ternary spatial relationship "road trajectory–building footprint" becomes weaker, which poses a significant challenge to road prediction with sparse trajectories.

This subsection examines the impact of trajectory sparsity on model performance. We manually sample trajectories with five sparse levels and five missing levels from the real trajectory data (achieved by setting the trajectory sampling rate and the size of the sampling area). That is, the experiment is carried out by randomly adopting positions (labeled as sparse levels 1, 2, 3, 4, and 5 at the ratio of 100%, 80%, 50%, 30%, and 10%) and block sampling methods with different ratios (sampling trajectories in blocks at the ratio of 100%, 80%, 50%, 30%, and 10% of the area width, labeled as missing levels 1, 2, 3, 4, and 5). The samples, which are superimposed building footprints, are fed into our model to generate roads, which are used to test the model's performance in predicting sparse and missing trajectory data.

Figure 11 shows the results using trajectories with different sparsity. The sparser the trajectories are, the fewer trajectory features the model can extract, but the quality of road results does not degrade much under the inspiration of building footprints. For trajectories with different missing degrees, as the completeness of the trajectories decreases, the generation effect of the model decreases significantly. However, even if the integrity of the trajectory is only 10%, the model can effectively learn the spatial structure relationship between the building footprints and roads and predict the location and basic morphology of roads. This shows that at the meso-level, the spatial structure relationship of "building footprint–road" is quite robust. This relationship can be used to supervise the process of road generation in MAC-GAN and then robustly handle different sparsity and missingness of trajectory data to better achieve road generation. This provides a feasible alternative for community road prediction with sparse trajectories.

Figures 12 and 13 show the quantitative evaluation metrics of the model under different degrees of sparsity and missingness of trajectories. With increasing sparsity and missingness, the F1 score and IOU decrease, and the FID is increases, but the drop and rise are small. The F1 score at each level decreases by 0.9–6.1%, IOU decreases by 1.1–6.7%, and FID increases by 0.22–9.67. This shows that the sparser the trajectories, the fewer trajectory features can be extracted. MAC-GAN can suppress the performance degradation through the learned associated feature structure, that is, "sparse trajectory–building footprint–road feature". It can be observed from Figure 13 that the fourth level of the missing trajectory degree is a threshold value. When the trajectories are missing less than the fourth level, the model's prediction performance is more stable, and the change is small. However, the model's performance decreases sharply when the missing degree of trajectories exceeds the fourth level. This indicates that for trajectories with too few features, it is difficult to supplement sufficient road features by building footprint data. However, even in the

case of severely missing trajectory features (such as the trajectory with the fourth missing level), MAC-GAN can still fully learn "trajectories–building footprints–road features" through multi-scale skip connection and asymmetric convolution with the channel attention mechanism and obtain better performance. Its F1 score reaches 51.1% and its IOU is 35.7%.

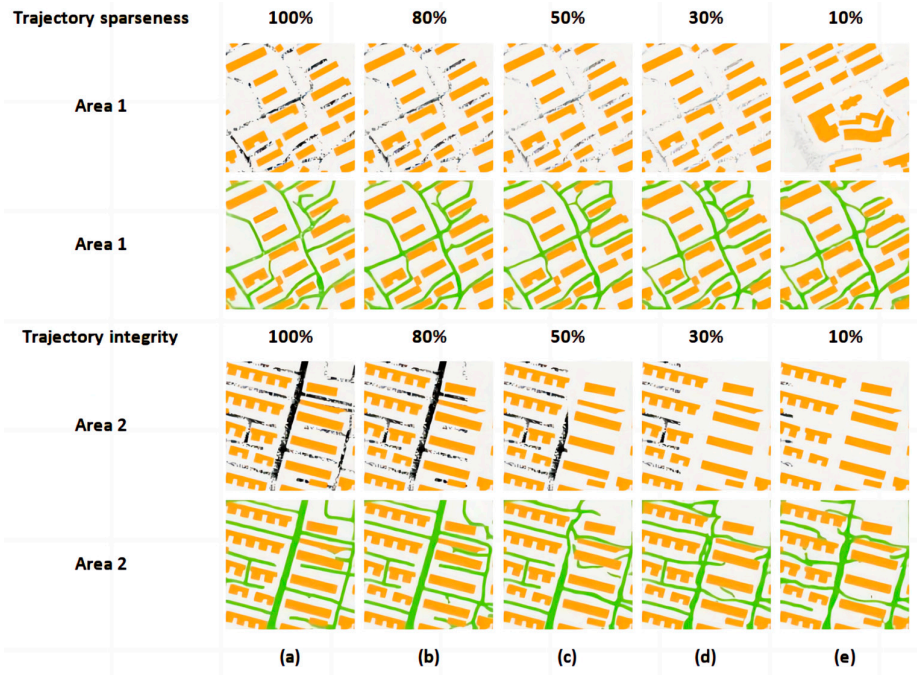


Figure 11. The results generated by our model on trajectory data with varying degrees of sparsity and missingness. The first and third rows of (**a**–**e**) show the trajectory data with the sparsity and missingness levels 1–5, and the second and fourth rows show the images generated by our model with them.

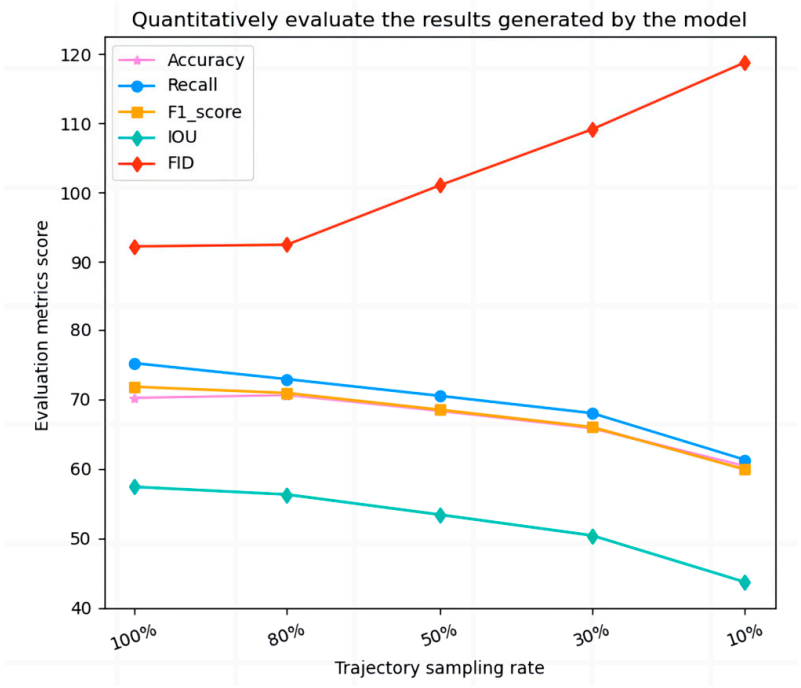


Figure 12. Evaluation metrics score of the model using the trajectory data with different sparsity.

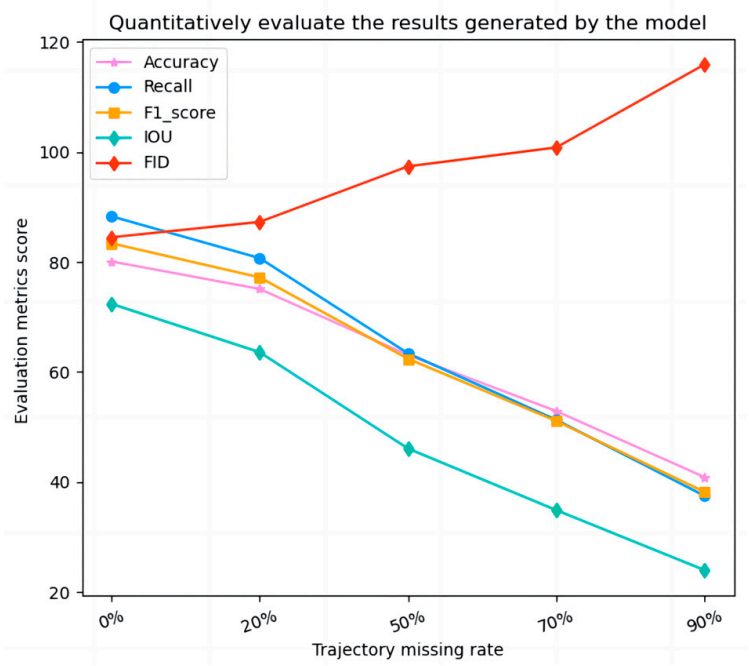


Figure 13. Evaluation metrics score of the model using the trajectory data with different missing degrees.

5.5. Loss of Model Training

Figure 14 shows the loss curve of the 300-round adversarial training process, which converges relatively quickly. The gray error curve represents the LI loss between the generated and real images. In the first 150 rounds of training, the pixel accuracy of the road image generated by the model is significantly improved, then maintains a state of slow improvement, and finally tends to stabilize. Besides the consistent downward trend, sudden spikes in the gray error curve are noticeable, which reflect the confrontation between the generator and the discriminator. The subgraph at the top right of Figure 13 shows the change in binary cross entropy adversarial loss (BCELoss) in D and G during the whole training process. It is apparent that the confusion degree of D (orange curve) tends to be the largest, and its loss is close to 0.5, and the loss in G (blue curve) continues to improve.

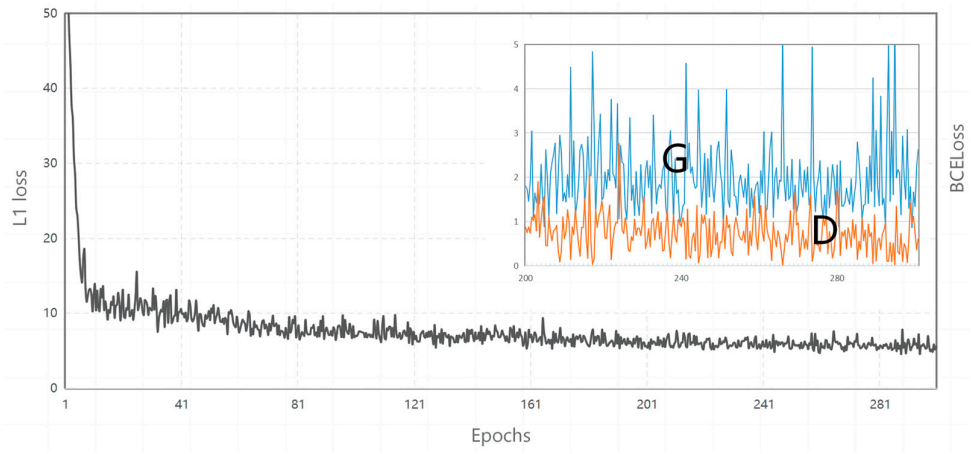


Figure 14. Loss changes in the adversarial training process of MAC-GAN. The gray curve represents the LI loss changes between the generated and real images. The blue and orange curves show the binary cross-entropy adversarial loss changes of the generator and the discriminator, respectively.

5.6. Limitations

Our model uses a combination of trajectories and building footprints around the road as context information to supervise road generation. During the process, a unified map zoom level and a specific image size were designated to capture road context to assist in inferring roads. The density of road contextual information (e.g., building footprints, trajectories) varies across communities, which may affect road generation. For areas with detailed trajectories, the model will generate more convincing results. Figure 15 shows typical cases of prediction failures. For the first row, the sparse trajectories in this community provide too little information, and the feature of community branches and building footprints is insufficient, resulting in poor results. For the second line, according to the structural duality of main roads and building footprints of the community, our model predicts main community roads well, but the branches only moderately. The region lacks trajectories for road extraction. For the third row, our model fails to capture effective contextual information to support road prediction in areas with missing trajectories and sparse building footprints. Therefore, the performance of our model depends on the complementarity information of trajectories and building footprints and the size of the total amount of feature information.

Figure 15. Examples of model prediction failures. (**a**–**c**) represent three community areas.

6. Conclusions

Existing road extraction methods based on multimodal information have limitations in identifying community roads from sparse trajectory data, resulting in low accuracy and integrity of community roads. To solve this problem, we propose a generative adversarial network (MAC-GAN) for the road generation task supervised by pedestrian trajectories and neighborhood building footprints. MAC-GAN is configured with a MACU-Net with asymmetric convolution blocks and multi-scale skip connections with channel attention. MACU-Net replaces standard convolution layers with asymmetric convolution blocks to enhance the network's feature representation and extraction capabilities, and utilizes multi-scale skip connection combination and the channel attention mechanism to adaptively fuse the "road trajectory–building footprint" ternary spatial features contained in low-level and high-level feature maps to supervise road generation. The combined image closely related to road information serves as a constraint that controls the generator's output, which is expected to ease the heavy dependence of the generated road data quality on the trajectory data quality alone at the community level.

Numerical experiments conducted on a large number of residential study areas in Wuhan confirmed the feasibility of our method for generating fine-scale community road data with different sparse and missing degrees and showed its advantages over baseline methods. To improve the application value of the method, the multi-level semantic constraints (pixel-level, object-level, feature-level) of road neighborhoods will be further explored to guide optimal road generation.

Author Contributions: Conceptualization, Lin Yang and Jing Wei; methodology, Lin Yang and Zejun Zuo; software, Jing Wei; validation, Lin Yang and Jing Wei; formal analysis, Zejun Zuo; investigation, Lin Yang; re-sources, Lin Yang and Zejun Zuo; data curation, Jing Wei; writing—original draft preparation, Lin Yang and Jing Wei; writing—review and editing, Lin Yang, Jing Wei and Zejun Zuo; visualization, Lin Yang; supervision, Zejun Zuo; project administration, Shunping Zhou and Lin Yang; funding acquisition, Lin Yang and Shunping Zhou. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant No. 42071383).

Data Availability Statement: The data presented in this study are available from the author upon reasonable request.

Acknowledgments: The authors thank the editors and anonymous reviewers for their insightful comments and constructive suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bettencourt, L.M.J. The origins of scaling in cities. *Science* **2013**, *340*, 1438–1441. [CrossRef]
2. Karagiorgou, S.; Pfoser, D. On vehicle tracking data-based road network generation. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, New York, NY, USA, 6–9 November 2012; pp. 89–98.
3. Liu, X.; Zhu, Y.; Wang, Y.; Forman, G.; Ni, L.M.; Fang, Y.; Li, M.J.H.L. *Road Recognition Using Coarse-Grained Vehicular Traces*; Hewlett-Packard Development Company: Palo Alto, CA, USA, 2012.
4. Xie, X.; Wong, K.B.-Y.; Aghajan, H.; Veelaert, P.; Philips, W. Inferring directed road networks from GPS traces by track alignment. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2446–2471. [CrossRef]
5. Zhang, L.; Thiemann, F.; Sester, M. Integration of GPS traces with road map. In Proceedings of the Third International Workshop on Computational Transportation Science, New York, NY, USA, 2 November 2010; pp. 17–22.
6. Bruntrup, R.; Edelkamp, S.; Jabbar, S.; Scholz, B. Incremental map generation with GPS traces. In Proceedings of the 2005 IEEE Intelligent Transportation Systems, Vienna, Austria, 16 September 2005; pp. 574–579.
7. Cao, L.; Krumm, J. From GPS traces to a routable road map. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, NY, USA, 4–6 November 2009; pp. 3–12.
8. Quddus, M.A.; Ochieng, W.Y.; Noland, R.B. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transp. Res. Part C Emerg. Technol.* **2007**, *15*, 312–328. [CrossRef]
9. Ahmed, M.; Wenk, C. Constructing street networks from GPS trajectories. In Proceedings of the Algorithms—ESA 2012: 20th Annual European Symposium, Ljubljana, Slovenia, 10–12 September 2012; pp. 60–71.
10. Edelkamp, S.; Schrödl, S. Route planning and map inference with global positioning traces. In *Computer Science in Perspective*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 128–151.
11. Worrall, S.; Nebot, E. Automated process for generating digitised maps through GPS data compression. In Proceedings of the Australasian Conference on Robotics and Automation, Brisbane, Australia, 10–12 December 2007.
12. Wang, J.; Rui, X.; Song, X.; Tan, X.; Wang, C.; Raghavan, V. A novel approach for generating routable road maps from vehicle GPS traces. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 69–91. [CrossRef]
13. Guo, Y.; Bardera, A.; Fort, M.; Silveira, R.I. A scalable method to construct compact road networks from GPS trajectories. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1309–1345. [CrossRef]
14. Davies, J.J.; Beresford, A.R.; Hopper, A. Scalable, distributed, real-time map generation. *IEEE Pervasive Comput.* **2006**, *5*, 47–54. [CrossRef]
15. Biagioni, J.; Eriksson, J. Inferring road maps from global positioning system traces: Survey and comparative evaluation. *Transp. Res. Rec.* **2012**, *2291*, 61–71. [CrossRef]
16. Yang, X.; Tang, L.; Ren, C.; Chen, Y.; Xie, Z.; Li, Q. Pedestrian network generation based on crowdsourced tracking data. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 1051–1074. [CrossRef]
17. Shi, W.; Shen, S.; Liu, Y. Automatic generation of road network map from massive GPS, vehicle trajectories. In Proceedings of the 2009 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA, 4–7 October 2009; pp. 1–6.
18. Biagioni, J.; Eriksson, J. Map inference in the face of noise and disparity. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, New York, NY, USA, 6–9 November 2012; pp. 79–88.
19. Li, Y.; Xiang, L.; Zhang, C.; Wu, H. Fusing taxi trajectories and RS images to build road map via DCNN. *IEEE Access* **2019**, *7*, 161487–161498. [CrossRef]
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
21. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.

22. Song, J.; Li, J.; Chen, H.; Wu, J. MapGen-GAN: A fast translator for remote sensing image to map via unsupervised adversarial learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2341–2357. [CrossRef]
23. Dong, G.; Huang, W.; Smith, W.A.; Ren, P. A shadow constrained conditional generative adversarial net for SRTM data restoration. *Remote Sens. Environ.* **2020**, *237*, 111602. [CrossRef]
24. Zhu, D.; Cheng, X.; Zhang, F.; Yao, X.; Gao, Y.; Liu, Y. Spatial interpolation using conditional generative adversarial neural net-works. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 735–758. [CrossRef]
25. Milojevic-Dupont, N.; Hans, N.; Kaack, L.H.; Zumwald, M.; Andrieux, F.; de Barros Soares, D.; Lohrey, S.; Pichler, P.-P.; Creutzig, F. Learning from urban form to predict building heights. *PLoS ONE* **2020**, *15*, e0242010. [CrossRef] [PubMed]
26. Mocnik, F.-B. Benford's law and geographical information—The example of OpenStreetMap. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1746–1772. [CrossRef]
27. Majic, I.; Naghizade, E.; Winter, S.; Tomko, M. There is no way! Ternary qualitative spatial reasoning for error detection in map data. *Trans. GIS* **2021**, *25*, 2048–2073. [CrossRef]
28. Wu, A.N.; Biljecki, F.J. GANmapper: Geographical data translation. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 1–29. [CrossRef]
29. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef]
30. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
31. Li, R.; Zheng, S.; Duan, C.; Su, J.; Zhang, C. Multistage attention ResU-Net for semantic segmentation of fine-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
33. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 182–186.
34. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.-W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059.
35. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
36. Li, R.; Duan, C.; Zheng, S. MACU-Net Semantic Segmentation from High-Resolution Remote Sensing Images. *arXiv* **2020**, arXiv:2007.13083.
37. Chaurasia, A.; Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
39. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
40. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.