A Context-enriched Satellite Imagery Dataset and an Approach for Parking Lot Detection

Yifang Yin[1], Wenmiao Hu[1,2*], An Tran[2], Hannes Kruppa[2], Roger Zimmermann[1], See-Kiong Ng[1]
[1]Grab-NUS AI Lab, National University of Singapore
[2]GrabTaxi Holdings, Singapore
{idsyin,dcsrz,seekiong}@nus.edu.sg, {wenmiao.hu,an.tran,hannes.kruppa}@grabtaxi.com

Abstract

Automatic detection of geoinformation from satellite images has been a fundamental yet challenging problem, which aims to reduce the manual effort of human annotators in maintaining an up-to-date digital map. There are currently several high-resolution satellite imagery datasets that are publicly available. However, the associated ground-truth annotations are limited to road, building, and land use, while the annotations of other geographic objects or attributes are mostly not available. To bridge the gap, we present Grab-Pklot, the first high-resolution and context-enriched satellite imagery dataset for parking lot detection. Our dataset consists of 1344 satellite images with the ground-truth annotations of carparks in Singapore. Motivated by the observation that carparks are mostly co-appear with other geographic objects, we associate each satellite image in our dataset with the surrounding contextual information of road and building, given in the format of multi-channel images. As a side contribution, we present a fusion-based segmentation approach to demonstrate that the parking lot detection accuracy can be improved by modeling the correlations between parking lots and other geographic objects. Experiments on our dataset provide baseline results as well as new insights into the challenges and opportunities in parking lot detection from satellite images.

1. Introduction

Semantic segmentation of satellite images has been an active research area for decades [10, 26]. By using deep learning neural networks, various geographic objects and attributes can be automatically extracted from high-resolution satellite images including roads [36], buildings [13], land use [8], *etc*. The detected geoinformation can be used to fill in the missing data of a digital map. It has great importance because it significantly reduces the manual effort required from human annotators when trying to improve the completeness of a digital map. To this end, several high-resolution satellite imagery datasets such as DeepGlobe [10] and SpaceNet [29] have been made publicly available recently. However, most of the datasets focus on the detection of road and building. Ground-truth annotations of other geographic objects are seldomly available in the existing satellite imagery datasets.

In this paper we present an satellite-imagery dataset on parking lot detection, which provides important geoinformation for a map user. For example, food delivery drivers sometimes may go to places that they are unfamiliar with. Thus, it can be difficult for them to find a carpark when delivering food if such information is not available in a digital map. However, based on our investigations, a significant number of carparks are missing from the web mapping services such as OpenStreetMap. To facilitate the development of automatic parking lot detection algorithms, we present Grab-Pklot[1], the first high-resolution and context-enriched satellite imagery dataset with ground-truth carpark annotations. This dataset consists of 1344 1024×1024 satellite images in Singapore, with a ground sampling distance of 0.3 meter/pixel. As it is too time-consuming and labor-intensive to label the carparks in the satellite images from scratch, we collect the geoinformation of carpark candidates in Singapore from two public datasets, *i.e*., the OpenStreetMap and the local government public data. The raw annotations we collected this way are given as polygons or points. We then require human annotators to refine the raw annotations by removing duplicates, refining misaligned polygons, and creating new polygons around the point candidates.

In additional to the satellite images and the carpark annotations, we associate each satellite image with two contextual features extracted from OpenStreetMap based on roads and buildings. In our dataset, roads and buildings are represented by a multi-channel image where each channel is

*The corresponding author.

[1]Available upon request sent to geo.grabpklot@grabtaxi.com

Figure 1: Example of correlations between carparks and other geographic objects, *i.e.*, roads (left) and buildings (right). Original Images © 2018 Maxar Technologies Inc.

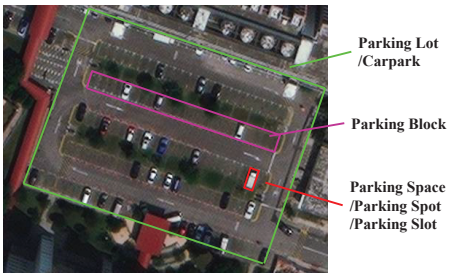

Figure 2: Illustrations of parking lot, block, and space. Original Image © 2018 Maxar Technologies Inc.

a binary mask for a certain category (*e.g.*, a service road or a residential building). This is motivated by the observation that carparks are likely to co-appear with other geographic objects as shown in Figure 1. On the left, it shows a large carpark that has service roads inside it. On the right, it shows a residential carpark that is surrounded by residential buildings. We thus believe that the parking lot detection accuracy can be improved by taking such contextual features into consideration. To this end, we present a fusion-based segmentation approach. It converts the contextual features into a 3-channel embedding, which is next added to the RGB channels of the corresponding satellite image. Thereafter, the generated feature map can be processed by any existing segmentation networks such as U-Net [22], D-LinkNet [36], and DeepLab [6]. We conducted extensive experiments on our dataset and observed a mIoU improvement of 1.80% $\sim$ 3.18% by our proposed fusion-based segmentation approach. The key contributions of this paper is summarized as follows:

- We present the first high-resolution and context-enriched satellite imagery dataset for parking lot detection, which consists of 1344 1024×1024 satellite images with carpark masks in Singapore.
- Our dataset is context-rich where each satellite image is further associated with the contextual information (*e.g.*, geometry and category) of roads and buildings extracted from OpenStreetMap.
- We present a fusion-based segmentation method and integrate it with eight state-of-the-art segmentation models. An improvement of 1.80% $\sim$ 3.18% in terms of mIoU have been observed when using the contextual features as an additional input.

Figure 2 illustrates the structure of a parking lot. Please note that the focus of this paper is to detect the location and the polygon of the parking lot. This problem has not been thoroughly studied as existing work mostly focused on the detection of the individual parking space or parking block inside a parking lot whose location is known. In previous work, "paring lot" is also termed as "carpark", and "parking space" is also termed as "parking spot" or "parking slot". Such terms are used interchangeably throughout this paper.

2. Related Work

Early parking lot detection methods mostly use wireless radio modules such as GPS and Wi-Fi to detect parking activities [3]. For example, PhonePark was proposed to detect whether a user was walking, stationary or driving based on GPS, accelerometer and Bluetooth connectivity data collected on the user's smartphones [25]. Park Here! utilized both accelerometer and gyroscope sensor to detect parking activity based on a binary classifier (*i.e.*, driving or not driving) [23]. Inspired by the great success of Convolutional Neural Networks (CNN) on image classification, machine learning based methods have been proposed to detect parking lots from surveillance or satellite imagery. Chen *et al*. presented a method to detect vacant parking spaces in a parking lot from surround-view images with the aid of pixel-level domain adaptation [5]. Seo *et al*. proposed a self-supervised method to extract the structure of parking spaces from satellite images [24]. Vadivel *et al*. proposed to localize parking spaces and vehicles in parking lots [28]. They modeled the problem as object detection from satellite images and investigated the performance of RCNN based neural network architectures. However, such methods mostly focused on the detection of parking vehicles [32] or vacant parking spaces [5, 19] rather than locating the parking lot from a global perspective.

For parking lot detection, there are only a handful of related datasets that are publicly available. For example, Tongji Parking-slot Dataset 2.0 contains surround-view images synthesized from four low-cost fisheye cameras [34]. This dataset is for parking slot detection where various parking-slot types were considered such as the vertical, the parallel, and the slant types. Do and Choi released a realistic parking slot dataset, which comprises parking slot images captured by the fish-eye cameras on vehicles with various attributes and external conditions [11]. PKLot [9],

Table 1: Comparison to existing benchmark datasets.

Dataset	Total Images	Camera View	Satellite View	Contextual Info.	Detection Target
DeepGlobe [10]	$> 10,000$	✗	✓	✗	Road, Building, Land Cover
SpaceNet [29]	$> 10,000$	✗	✓	✗	Road, Building
PKLot [9]	12,417	✓	✗	✗	Parking Space
CNRPark [2]	12,000	✓	✗	✗	Parking Space
CNRPark+EXT [1]	144,965	✓	✗	✗	Parking Space
APKLOT [15]	500	✗	✓	✗	Parking Block
Grab-Pklot, Ours	1,344	✗	✓	✓	Parking Lot

CNRPark [2], and CNRPark+EXT [1] are three benchmark datasets for visual vacancy/occupancy detection in a parking lot. These datasets comprise ground camera-view images of vacant and occupied parking spaces captured under varied weather conditions (*e.g.*, sunny, overcast or rainy) in real-world scenarios. APKLOT [15] is a dataset for parking block segmentation, which is also the most relevant to our work. This dataset contains 500 satellite view images, but the images do not have any contextual information associated with them. To our best knowledge, there is no available dataset dedicated to parking lot detection from bird-view high ground-coverage satellite imagery yet.

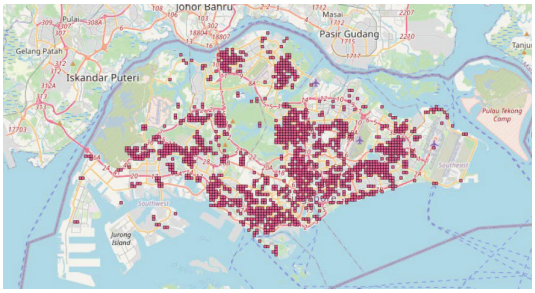

Figure 3: Dataset distribution over the Area of Interest (AoI). Pink polygons indicates the location of each sample in the dataset

3. Challenges

Parking lot detection from satellite imagery is an important, yet challenging real-world problem. Though it can be modeled as a binary semantic segmentation problem, it is different from the existing segmentation tasks in several aspects and thus posts new challenges the existing models cannot handle effectively. *First of all*, unlike the objects in scene understanding [33, 12] or medical image segmentation [22], the size of a parking lot varies significantly in different regions, depending on its capacity. Though existing models have utilized spatial pyramid and dilated convolutions to segment objects in different sizes [35, 20], such techniques may not be sufficient for the parking lot detection from satellite imagery.

Second, the visual appearance of parking lots can be quite diverse as they may have irregular shapes or get occluded by vegetation or cloud. The visual appearance of a parking lot can also shift over time due to the change of illumination or the number of cars parking inside. This challenge can degrade the performance of existing segmentation models. For example, a model may confuse an occupied parking lot with a road segment full of cars due to their similar visual appearance. Similarly, a vacant parking lot can be confused with a building rooftop or a vacant ground.

Finally, the number of public satellite images with ground-truth parking lot annotations is very limited. We can find the information of some carparks on public map data such as OpenStreetMap. However, as the public map data are mostly crowdsourced, there is no guarantee of the data quality. The annotations collected this way can be incomplete and imprecise. And it is unclear if existing models can learn effectively from such weak and noisy annotations.

4. Dataset

Geoinformation extraction from high-resolution aerial and satellite images has gained its popularity in recent years. As shown in Table 1, DeepGlobes [10] and SpaceNet [29] are two popular high-resolution satellite imagery datasets for building detection, road detection, and land cover classification. However, to the best of our knowledge, the datasets related to parking lot detection are mostly composed of camera view images only [9, 2]. Take CNRPark+EXT [1] as an example, this dataset, though consists of 144,965 images, is collected by 9 fixed cameras covering a small parking area only. APKLOT [15] is the only satellite imagery based carpark dataset, which contains 500 images with varying sizes and resolutions. To bridge the gap, we present the first high-resolution satellite imagery based carpark dataset, where each satellite image is further associated with contextual information collected from OpenStreetMap.

4.1. Dataset Overview

Our dataset contains a total of 1344 DigitalGlobe satellite images in Singapore, the distribution of which is illustrated in Figure 3. The satellite images together with

Figure 4: Dataset samples from training (left), testing (middle) and additional training (right) sets. Original Images © 2018 Maxar Technologies Inc.

the corresponding context and carpark annotations in our dataset have a size of 1024 $\times$ 1024 pixels with a ground sampling distance of 0.3 meter/pixel. To train parking lot detection models, we further divide the dataset into a training set of 527 samples, a testing set of 200 samples, and an additional training set of 617 samples. The images in the training and testing sets are labeled with manually refined high quality carpark annotations. The images in the additional training set are labeled with partial carpark annotations where missing parking blocks exist. The additional training set reflects the quality of carpark annotations in real-world crowd-sourced maps. Considering that a model trained in one country may not generalize well to another country, it is interesting to study if one can make effective use of the partially labeled images during training, to reduce the heavy cost of ground-truth labeling.

Figure 4 shows three samples in our dataset from training, testing, and additional training sets, respectively. As the size and density of carparks may vary significantly in different regions, we further group the images into three clusters based on the visible area of carparks in the satellite image and randomly divide them into the training and testing sets. The statistics are shown in Table 2. Overall speaking, our dataset is generated by three steps. First, a list of carpark candidates is extracted from public map and government data sources. Next, the initial carpark candidates are refined by experienced annotators. Satellite images with at least a partially visible carpark in the content are selected as samples in our dataset. Finally, we extract contextual information such as road network and building footprint from OpenStreetMap that correspond to each sample to provide an enriched satellite imagery dataset for parking lot detection. The details are introduced in the next section.

4.2. Dataset Generation

4.2.1 Candidate Generation

Instead of labeling the ground-truth carpark annotations from scratch, we collected a list of carpark candidates from OpenStreetMap (OSM) and Singapore's government public data. The carpark information on OSM is manually contributed by different users. As a result, the coverage of

Table 2: Statistics of the three groups divided based on the visible area of the carparks.

Group	Area (No. of pixels)	No. of samples (Train/Test)
1	$> \frac{1024 \cdot 1024}{16}$	183/79
2	$< \frac{1024 \cdot 1024}{16}$ and $> \frac{1024 \cdot 1024}{256}$	210/87
3	$< \frac{1024 \cdot 1024}{256}$	134/34
Total	> 0	527/200

Figure 5: Left: parking slots are grouped based on locations to form the polygon candidates. Right: only the points (*i.e.*, red dots) that are not covered by any polygon candidates (*i.e.*, pink polygons) are kept for further labeling. Those already been covered (*i.e.*, yellow dots) are discarded. Original Images © 2018 Maxar Technologies Inc.

carparks can be incomprehensive and imprecise. On OSM, the carpark polygons are labeled with keywords such as "Car Park", "carpark", "Parking", "garages", and "Vehicle Park" in different attributes. We therefore select OSM candidates with the corresponding polygons by filtering their keywords and relations to other geographic objects.

From the government data, we gathered information from three Singapore agencies, namely the Urban Redevelopment Authority (URA), the Housing & Development Board (HDB), and the National Parks Board (NParks). Some of the information consists of polygons of individual parking lots, and the others are given as point candidates around the actual location of the carpark. For the first type of data, parking lots are grouped based on their locations and attributes to form the polygon candidates (see Figure 5 left), which are next merged with the OSM polygon candidates to remove the duplicates. For the second type of data (*i.e.*, the point candidates), only those that are not covered by polygon candidates are kept for further labeling (see Figure 5 right) as introduced below.

4.2.2 Manual Refinement and Data Selection

After obtaining a list of polygon and point carpark candidates, a manual refinement step was conducted by experienced annotators to remove non-carpark candidates, readjust existing polygon, and extend point candidates. Specifically, new polygons were created for each of the point candidate, with reference to the carpark that is visible in the

Table 3: Categories of the contextual features associated with the satellite imagery in our dataset.

Context	Category
Road	service, residential, primary, secondary tertiary, trunk, motorway, others
Building	residential, house, apartments, terrace industrial, commercial, others

corresponding satellite image. The results from refined and newly created polygons were merged to generate the geo-referenced ground truth of carparks, which contain a total of 2883 polygons. Instead of cropping the satellite imagery at the center of each polygon, we crop the entire area of interest (*i.e.*, Singapore) into non-overlapping image chips with 1024×1024 pixels and select the ones with at least a partially visible carpark in the content as candidates to form our dataset. Subsequently, there are 1484 images overlapping with the geo-referenced carpark ground truth we created. We further conduct an additional round of annotation adjustment and marginal carpark removal, resulting a dataset comprising 1344 image-mask pairs. The distribution of the dataset is shown in Figure 3.

4.2.3 Context Generation

For each of satellite images in our dataset, we additionally extract two types of contextual features from OpenStreetMap based on roads and buildings, respectively. We would like to capture not only the geometry of roads and buildings, but also their categories in our extracted contextual features [31, 30]. This is possible as there are key-value pairs in the OpenStreetMap data that indicate the category of a geographic object, *e.g.*, *highway=service* and *building=residential*. We thus group roads into 8 categories as shown in Table 3, and generate a binary mask for each of the categories. We concatenate the binary masks into an 8-channel image as the road contextual feature in our dataset. Similarly, we group buildings into 7 categories and generate a 7-channel image as the building contextual feature.

Figure 6 shows the dataset distribution over the building and road categories. Take building as an example, we compute the number of pixels that belong to the i-th category, denoted as r_i, and report the normalized $\hat{r}_i = \frac{r_i}{\sum_{i=1}^{7} r_i}$ in Figure 6. As can be seen, a majority of the pixels belong to category "others" as many buildings on OSM do not have a specific category label. The second largest category is "residential" which covers about 17.7% of building pixels. Similarly, we report the normalized $\hat{b}_i = \frac{b_i}{\sum_{i=1}^{8} b_i}$ over the 8 road categories. Category "service" covers the most road pixels, which also correlates with carparks the most. This is aligned with the OpenStreetMap Wiki [21], where the main

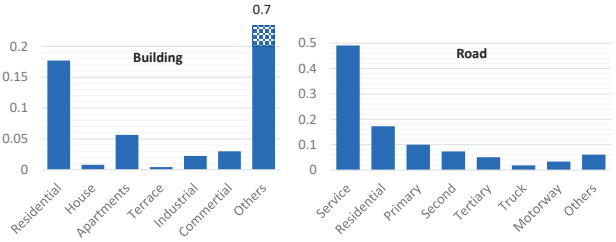

Figure 6: Distribution of categories for road and building.

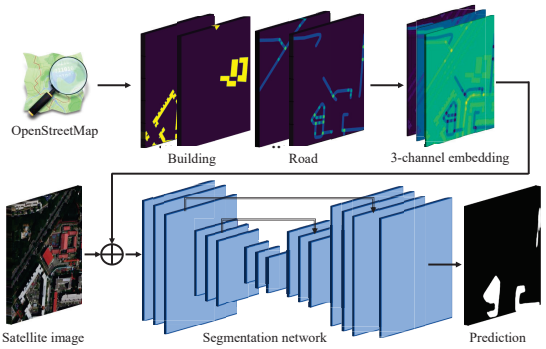

Figure 7: Network architecture of our proposed segmentation method.

ways on a parking lot that connect multiple parking aisles should be labeled with *highway=service*.

5. Approach

We model the parking lot detection as a semantic segmentation problem and present a new baseline by fusing satellite images with contextual features extracted from road network and building footprint. The advantages of utilizing contextual features as additional inputs for parking lot detection are threefold. First, the visibility of a parking lot may not always be good due to occlusions caused by trees, buildings, or heavy clouds in a satellite image. Second, the contextual features can help remove false positives that overlap with roads and buildings, thus improved segmentation results can be obtained. Third, detectors can learn from the correlations between parking lots and some of the geographic objects. For example, there is always a parking lot in the residential area for the convenience of residents.

Figure 7 illustrates the overview of our proposed method. In order to be compatible with existing segmentation models, we propose to embed the road and building contextual features into a 3-channel image, which can be easily added to the RGB channels of a satellite image. To achieve this goal, we first concatenate the channels of the road context and the building context to obtain a 15-channel image. Next, we reduce the number of channels to 3 by processing it by a 2D convolutional layer followed by Tanh acti-

Table 4: Performance comparison (mIoU (%)) of different segmentation models for parking lot detection based on satellite image with or without road and building context.

Method	Group 1			Group 2			Group 3			Overall		
	w/o cont.	w/ cont.	mIoU gain	w/o cont.	w/ cont.	mIoU gain	w/o cont.	w/ cont.	mIoU gain	w/o cont.	w/ cont.	mIoU gain
U-Net [22]	72.35	75.11	2.76	62.62	66.51	3.89	53.44	55.80	2.36	64.90	68.08	3.18
U-Net++ [37]	71.40	74.38	2.98	63.70	64.44	0.74	52.20	56.07	3.87	64.79	66.94	2.15
LinkNet [4]	72.13	74.50	2.37	63.11	66.07	2.96	51.33	52.14	0.81	64.67	67.03	2.36
D-LinkNet [36]	72.09	74.00	1.91	63.53	66.25	2.72	**56.58**	58.77	2.19	65.73	68.04	2.31
FPN [17]	72.75	75.19	2.44	63.91	66.80	2.89	55.41	56.04	0.63	65.95	68.29	2.34
PAN [18]	74.37	**75.77**	1.40	64.10	67.00	2.90	50.02	52.84	2.82	65.77	68.06	2.29
DeepLab v3 [6]	**74.62**	75.19	0.57	64.59	**67.27**	2.68	54.24	**60.62**	6.38	**66.79**	**69.27**	2.48
DeepLab v3+ [7]	73.23	75.13	1.90	**65.78**	66.14	0.36	53.64	58.88	5.24	66.66	68.46	1.80

vation. The kernel size and the stride of the convolutional layer are set to 7 and 1, respectively. The operator $\oplus$ denotes element-wise addition. Thereafter, the segmentation network in our framework can be any existing semantic segmentation models such as U-Net [22], D-LinkNet [36], DeepLabV3 [6], *etc*. The last layer outputs a 1-channel prediction map, the size of which equals to the size of the input satellite image. We adopt the Sigmoid activation to output probability scores between 0 and 1 as the parking lot detection can be seen as a binary classification problem at each pixel. For optimization, we adopt the combo loss [27]:

$$CL = \alpha \cdot BCE(y, p) + (1 - \alpha) \cdot DL(y, p) \quad (1)$$

which is a weighted sum of the binary cross-entropy loss:

$$BCE(y, p) = -(y \log(p) + (1 - y) \log(1 - p)) \quad (2)$$

and the dice loss:

$$DL(y, p) = 1 - \frac{2yp + smooth}{y + p + smooth} \quad (3)$$

The dice loss can handle the input class-imbalance problem. So the combo loss is beneficial for segmenting a small foreground from a large background, while at the same time enforcing a smooth training using the binary cross-entropy loss [27, 16]. The final prediction map is obtained based on thresholding. The predicted parking lot mask is composed of pixels whose probability scores are greater than a pre-defined threshold.

6. Experiments

We investigate the performance of eight different state-of-the-art segmentation networks on parking lot detection with or without the context information in our dataset. The eight segmentation networks include **U-Net** [22], **U-Net++** [37], **LinkNet** [4], **D-LinkNet** [36], **FPN** [17], **PAN** [18], **DeepLab v3** [6], and **DeepLab v3+** [7]. For all the models, we adopt the ResNet34 [14] as the backbone network and the combo loss [27] as the loss function. We initialize the ResNet34 with pre-trained weights on the ImageNet. And we empirically set $\alpha = 0.5$ and $smooth = 1$ in the loss function. For optimization, we train the neural networks using the Adam optimizer with a batch size of 8. The learning rate is set to 0.0002 with decays. To prevent overfiting, we applied data augmentation to the training samples including horizontal flip, vertical flip, color jittering, image shifting, scaling, and rotation.

6.1. Baseline

To investigate if contextual information is beneficial for detecting parking lots from satellite images, we train each of the segmentation models with and without the road and building context that is available in our dataset. We adopt the mean Intersection over Union (mIoU) as the evaluation metric and report the results in Table 4 with the **best result** highlighted. We train the segmentation models using the "training set" and report both the overall and the per-group results (see Table 2) on the "testing set" as described in Section 4. The results show that the detection difficulty increases while the size of the parking lots decreases regardless of the segmentation model we use. Without the road and building context, models obtained a best mIoU of 74.62%, 65.78%, and 56.58% for the three groups, respectively. With the road and building context, models obtained an improved mIoU of 75.77%, 67.72%, and 60.62%, achieving a performance gain of 1.15%, 1.94%, and 4.04% for the three groups, respectively. The results indicate that the use of road and building context is beneficial for the detection of parking lots in all sizes. Particularly, it brings significant gain in detecting small-size parking lots, which are difficult to pick up when only using satellite imagery. From our context-enriched satellite imagery dataset, models can learn rich correlations between parking lots and other geographic objects. For example, there can be service roads inside a large parking lot or residential buildings near a parking lot. Moreover, the road and building context is also

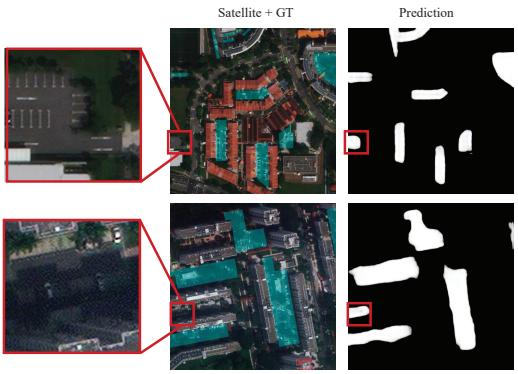

Figure 8: Two training samples with missing parking spaces. Original Images © 2018 Maxar Technologies Inc.

helpful for removing the false positives, *i.e.*, the mistakenly detected pixels due to the similar visual appearances to the pixels inside a parking lot. In terms of the presence of missing and misaligned annotations, we observe that models can actually learn from the weak and imprecise carpark masks when sufficient training samples with reasonably good annotations are available. Figure 8 illustrates two training samples in our dataset. The middle column shows the satellite images together with the carpark annotations available in our dataset, while the right column shows the predicted carpark masks outputted by a segmentation model after converging. As can be seen, though the annotations for some carparks are missing (*e.g.*, the region in the red bounding box) from the ground truth, our segmentation model is still able to detect such carparks successfully after learning from the whole training dataset.

Among the eight models, DeepLab v3 obtained the best overall mIoU, followed by DeepLab v3+ and FPN. Some models favor the detection of large parking lots such as PAN, while others favor the detection of small parking lots such as D-LinkNet. DeepLab v3 calculated the final prediction mask based on bilinear interpolation. DeepLab v3+ introduced a decoder module, which combines high-level and low-level features in order to obtain more accurate segmentation boundary. However, DeepLab v3 outperformed DeepLab v3+ on our dataset. One possible reason can be that the carpark annotations of the training samples in our dataset are crowdsourced and thus have small misalignment and a few missing parking spaces. Moreover, as the road width information is mostly not available on OSM, we approximately set the road width to 10 meters in our dataset. Both the annotation noise and the input approximation can cause performance degradation of DeepLab v3+ when learning the segmentation boundary. To summarize, DeepLab v3 obtained the best overall mIoU of 69.27% and 66.79% with and without the context information, respectively. The road and building context is beneficial for addressing the challenges on parking lot detection from satellite images. By taking it as an additional input, performance gain has been observed among all segmentation models.

Table 5: Performance comparison (mIoU (%)) of U-Net on parking lot detection using different input channels.

Input Channel			Group 1	Group 2	Group 3	Overall
Sat.	Road	Buil.				
3	0	0	72.35	62.62	53.44	64.90
3	1	0	74.99	66.07	51.95	67.19
3	8	0	75.11	65.90	50.74	66.96
3	0	1	70.78	62.20	52.77	64.00
3	0	7	73.38	65.09	53.05	66.32
3	1	1	73.11	67.03	50.40	66.61
3	8	7	75.11	66.51	55.80	68.08

6.2. Contextual Features

This experiment studies the use of the contextual features in our segmentation framework. We report the detection results obtained by fusing satellite image with only one of the contextual features and study the impact of the road/building context on parking lot detection. Recall that the road context in our dataset is represented by an 8-channel image where a binary mask is generated for each road type including "service", "residential", "primary", "secondary", *etc*. To study the impact of the road type on parking lot detection, we remove the type information by using only a 1-channel image (*i.e.*, a binary mask for all types) to represent the road context. Similarly, we also generate a binary mask for all types of buildings and use it as the building context. We perform experiments on U-Net, and report the results in Table 5.

From the results we observe that the road type does not have much impact on parking lot detection as competitive results have been obtained by fusing with either the 1-channel road context or the 8-channel road context. On the other hand, the building type turns out to be an important and indispensable feature for parking lot detection. When fusing the satellite image with the 1-channel building context, we only observed an overall mIoU of 64%. Then the mIoU got significantly improved to 66.32% when fusing with the 7-channel building context. One possible reason is that parking lots only correlate with certain types of roads and buildings such as the service roads and residential buildings. While category "service" is dominant for roads, category "residential" covers much fewer pixels than category "others" for building. Subsequently, it becomes difficult for the segmentation models to learn the correlations when the building type information is not available due to the noise introduced by the pixels belonging to "others". As can be seen, by fusing the satellite image with the road context only or the building context only, we achieved an over-

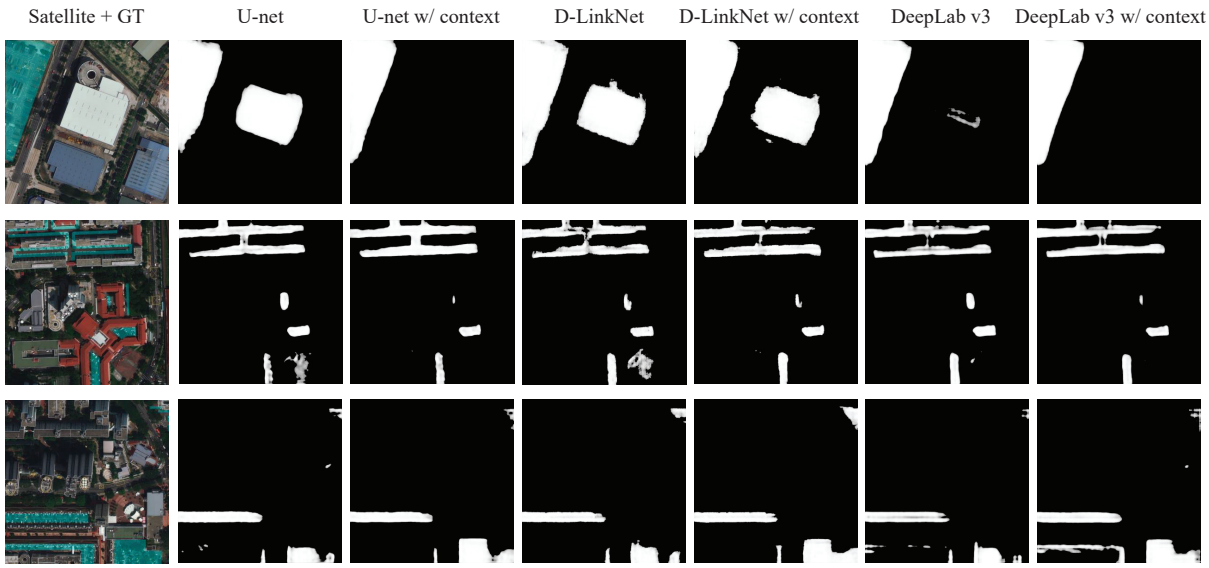

Figure 9: Parking lot detection results of different segmentation models trained with or without the context of road network and building footprint. Original Images © 2018 Maxar Technologies Inc.

all mIoU of 66.96% and 66.32%, respectively. This result indicates that the road context tends to be more important than the building context for parking lot detection. By fusing the satellite image with both contextual features with or without the type information, we achieved an overall mIoU of 68.08% and 66.61%, respectively, which verifies the importance of the type information for parking lot detection.

6.3. Visualization

Figure 9 visualizes the parking lot detection results of different segmentation models trained with or without the road and building contextual features. In our experiments, DeepLab v3 outperformed U-Net and D-LinkNet. For instance, in the first example, U-Net and D-LinkNet mistakenly recognized the rooftop of a building as the parking lot. In the second example, U-Net and D-LinkNet again mistakenly recognized a vacant ground as the parking lot. DeepLab v3 performed much better where only a small region of the rooftop/ground was wrongly detected. These issues can be addressed by taking the road and building context as an additional input as shown in the third, fifth, and seventh columns. In the third example, there is a narrow parking lot at the bottom left corner that is hard to be detected based on the satellite image only, but can be successfully recognized with the facilitation of the contextual features. To summarize, by fusing satellite image with road and building contextual features, false positives (*i.e.*, mistakenly detected building rooftop and vacant ground) can be effectively removed. Moreover, the detection rate of the hard instances, which are difficult to be recognized due to irregular shape, tiny size, vegetation occlusion, *etc*., increases by modeling the correlations between different geographic objects. Thus, the visualization results further verifies the effectiveness of our proposed approach.

7. Conclusion and Future Work

We present a high-resolution satellite imagery dataset with high quality carpark annotations. For each satellite image, we additionally collect the contextual information of roads and buildings from OSM, represented by a multi-channel image that captures not only their geometry, but also their categories. We show through experiments that it is beneficial to take the context as an additional input to address the challenges of parking lot detection. In the future, we plan to develop semi-supervised parking lot detection method to leverage the large number of partially labeled satellite images. We will also continue refining the carpark annotations in the additional training set.

8. Acknowledgment

This work was funded by the Grab-NUS AI Lab, a joint collaboration between GrabTaxi Holdings Pte. Ltd. and National University of Singapore, and the Industrial Postgraduate Program (Grant: S18-1198-IPP-II) funded by the Economic Development Board of Singapore, and in part supported by Singapore Ministry of Education Academic Research Fund Tier 2 under MOE's official grant number MOE2018-T2-1-103. We would also like to thank Maxar Technologies Inc. for allowing the original satellite images to be released as part of Grab-Pklot dataset for non-profit academic research purposes.

References

[1] Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Carlo Meghini, and Claudio Vairo. Deep Learning for Decentralized Parking Lot Occupancy Detection. *Expert Systems with Applications*, 72:327–334, 2017.

[2] Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, and Claudio Vairo. Car Parking Occupancy Detection using Smart Camera Networks and Deep Learning. In *IEEE Symposium on Computers and Communication*, pages 1212–1217, 2016.

[3] Pietro Edoardo Carnelli, Joy Yeh, Mahesh Sooriyabandara, and Aftab Khan. Parkus: A Novel Vehicle Parking Detection System. In *Twenty-Ninth IAAI Conference*, 2017.

[4] Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In *IEEE Visual Communications and Image Processing*, pages 1–4, 2017.

[5] J. Chen, L. Zhang, Y. Shen, Y. Ma, S. Zhao, and Y. Zhou. A Study of Parking-Slot Detection with the Aid of Pixel-Level Domain Adaptation. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2020.

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017.

[7] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *European Conference on Computer Vision*, pages 801–818, 2018.

[8] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional Map of the World. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6172–6180, 2018.

[9] Paulo RL De Almeida, Luiz S Oliveira, Alceu S Britto Jr, Eunelson J Silva Jr, and Alessandro L Koerich. PKLot–A Robust Dataset for Parking Lot Classification. *Expert Systems with Applications*, 42(11):4937–4949, 2015.

[10] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 172–181, 2018.

[11] H. Do and J. Y. Choi. Context-Based Parking Slot Detection with a Realistic Dataset. *IEEE Access*, 8:171551–171559, 2020.

[12] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual Attention Network for Scene Segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019.

[13] Nitin L Gavankar and Sanjay Kumar Ghosh. Automatic Building Footprint Extraction from High-resolution Satellite Image using Mathematical Morphology. *European Journal of Remote Sensing*, 51(1):182–193, 2018.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[15] Nisim Hurst-Tarrab, Leonardo Chang, Miguel Gonzalez-Mendoza, and Neil Hernandez-Gress. Robust Parking Block Segmentation from a Surveillance Camera Perspective. *Applied Sciences*, 10(15), 2020.

[16] Shruti Jadon. A Survey of Loss Functions for Semantic Segmentation. In *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, pages 1–7, 2020.

[17] Alexander Kirillov, Kaiming He, Ross Girshick, and Piotr Dollár. A Unified Architecture for Instance and Semantic Segmentation. `http://presentations.cocodataset.org/COCO17-Stuff-FAIR.pdf`, 2017.

[18] Hanchao Li, Pengfei Xiong, Jie An, and Lingxue Wang. Pyramid Attention Network for Semantic Segmentation. *arXiv preprint arXiv:1805.10180*, 2018.

[19] Wei Li, Libo Cao, Lingbo Yan, Chaohui Li, Xiexing Feng, and Peijie Zhao. Vacant Parking Slot Detection in the Around View Image Based on Deep Learning. *Sensors*, 20(7), 2020.

[20] Sachin Mehta, Mohammad Rastegari, Anat Caspi, Linda Shapiro, and Hannaneh Hajishirzi. Espnet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation. In *the European Conference on Computer Vision*, pages 552–568, 2018.

[21] OpenStreetMap Wiki. https://wiki.openstreetmap.org/.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 234–241, 2015.

[23] R. Salpietro, L. Bedogni, M. Di Felice, and L. Bononi. Park Here! A Smart Parking System based on Smartphones' Embedded Sensors and Short Range Communication Technologies. In *IEEE 2nd World Forum on Internet of Things*, pages 18–23, 2015.

[24] Young-Woo Seo, Nathan Ratliff, and Chris Urmson. Self-supervised Aerial Images Analysis for Extracting Parking Lot Structure. In *International Joint Conference on Artificial Intelligence*, 2009.

[25] L. Stenneth, O. Wolfson, B. Xu, and P. S. Yu. PhonePark: Street Parking Using Mobile Phones. In *IEEE International Conference on Mobile Data Management*, pages 278–279, 2012.

[26] Tao Sun, Zonglin Di, Pengyu Che, Chun Liu, and Yin Wang. Leveraging Crowdsourced GPS Data for Road Extraction from Aerial Imagery. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7509–7518, 2019.

[27] Saeid Asgari Taghanaki, Yefeng Zheng, S Kevin Zhou, Bogdan Georgescu, Puneet Sharma, Daguang Xu, Dorin Comaniciu, and Ghassan Hamarneh. Combo Loss: Handling Input and Output Imbalance in Multi-organ Segmentation. *Computerized Medical Imaging and Graphics*, 75:24–33, 2019.

[28] Murugesan Vadivel, SelvaKumar Murugan, Suriyadeepan Ramamoorthy, Vaidheeswaran Archana, and Malaikannan Sankarasubbu. Detecting Parking Spaces in a Parcel using Satellite Images. *arXiv preprint arXiv:1909.05624*, 2019.

[29] Adam Van Etten, Dave Lindenbaum, and Todd M. Bacastow. SpaceNet: A Remote Sensing Dataset and Challenge Series. *arXiv preprint arXiv:1807.01232*, 2018.

[30] Yifang Yin, An Tran, Ying Zhang, Wenmiao Hu, Guanfeng Wang, Jagannadan Varadarajan, Roger Zimmermann, and See-Kiong Ng. Multimodal Fusion of Satellite Images and Crowdsourced GPS Traces for Robust Road Attribute Detection. In *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2021.

[31] Yifang Yin, Jagannadan Varadarajan, Guanfeng Wang, Xueou Wang, Dhruva Sahrawat, Roger Zimmermann, and See-Kiong Ng. A Multi-Task Learning Framework for Road Attribute Updating via Joint Analysis of Map Data and GPS Traces. In *The Web Conference*, pages 2662–2668, 2020.

[32] Sebastian Zambanini, Ana-Maria Loghin, Norbert Pfeifer, Elena Màrmol Soley, and Robert Sablatnig. Detection of Parking Cars in Stereo Satellite Images. *Remote Sensing*, 12(13):2170, 2020.

[33] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. Context Encoding for Semantic Segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7151–7160, 2018.

[34] Lin Zhang, Junhao Huang, Xiyuan Li, and Lu Xiong. Vision-based Parking-slot Detection: A DCNN-based Approach and a Large-scale Benchmark Dataset. *IEEE Transactions on Image Processing*, 27(11):5350–5364, 2018.

[35] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid Scene Parsing Network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2881–2890, 2017.

[36] Lichen Zhou, Chuang Zhang, and Ming Wu. D-linknet: Linknet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 182–186, 2018.

[37] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, 2018.