2023

Detecting Road Intersections from Satellite Images using Convolutional Neural Networks

Fatmaelzahraa Eltaher

Luis Miralles-Pechuán

Jane Courtney

See next page for additional authors

Funder: Science Foundation Ireland

Authors
Fatmaelzahraa Eltaher, Luis Miralles-Pechuán, Jane Courtney, and Susan McKeever

Detecting Road Intersections from Satellite Images using Convolutional Neural Networks

Fatma El-zahraa El-taher
fatma.e.eltaher@mytudublin.ie
School of Computer Science, Technological University Dublin
Ireland

Luis Miralles-Pechuán
School of Computer Science, Technological University Dublin
Dublin, Ireland
luis.miralles@tudublin.ie

Jane Courtney
School of Electrical and Electronic Engineering, Technological University Dublin
Dublin, Ireland
jane.courtney@tudublin.ie

Susan Mckeever
School of Computer Science, Technological University Dublin
Dublin, Ireland
susan.mckeever@tudublin.ie

ABSTRACT

Automatic detection of road intersections is an important task in various domains such as navigation, route planning, traffic prediction, and road network extraction. Road intersections range from simple three-way T-junctions to complex large-scale junctions with many branches. The location of intersections is an important consideration for vulnerable road users such as People with Blindness or Visually Impairment (PBVI) or children. Route planning applications, however, do not give information about the location of intersections as this information is not available at scale. As a first step to solving this problem, a mechanism for automatically mapping road intersection locations is required, ideally using a globally available data source.

In this paper, we propose a deep learning framework to automatically detect the location of intersections from satellite images using convolutional neural networks. For this purpose, we labelled 7,342 Google maps images from Washington, DC, USA to create a dataset. This dataset covers a region of 58.98 km^2 and has 7,548 intersections. We then applied a recent object detection model (EfficientDet) to detect the location of intersections. Experiments based on the road network in Washington, DC, show that the accuracy of our model is within 5 meters for 88.6% of the predicted intersections. Most of our predicted centre of the intersections (approx 80%) are within 2 metres of the ground truth centre. Using hybrid images, we obtained an average recall and an average precision of 76.5% and 82.8% respectively, computed for values of Intersection Over Union (IOU) from 0.5 to 0.95, step 0.05. We have published an automation script to enable the reproduction of our dataset for other researchers.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision tasks**;

KEYWORDS

Datasets, Deep learning, Satellite images, Remote sensing images, Data acquisition, Route planning, Road intersections

ACM Reference Format:
Fatma El-zahraa El-taher, Luis Miralles-Pechuán, Jane Courtney, and Susan Mckeever. 2023. Detecting Road Intersections from Satellite Images using Convolutional Neural Networks. In *Proceedings of ACM SAC Conference (SAC'23).* ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/xx.xxx/xxx_x

1 INTRODUCTION

Predefining intersection features plays a crucial role in different domains such as route planning and road network extractions. Intersection identification is an important task to extract road networks from satellite photos, especially in complicated regions [4, 11]. Predefining intersection features increases the flexibility of route planning as it allows users to add or eliminate crossroads in accordance with their preferences. To give two examples, signalized intersections lengthen travel times and make automobiles use more fuel [1, 8]. The inclusion of simple crossroads and the avoidance of intersections promotes the safety of a person's travel, especially for People with Blindness or Visually Impairment (PBVI) [3, 6].

The majority of PBVI outdoor navigation systems use publicly available pathfinding services that generally do not provide information about the location of intersections [6]. This insight is crucial for determining the safest route and avoiding crossings for PBVI [3, 6]. Three key benefits make satellite images a viable source of information for detecting intersections. First, they capture various types of intersection structures. Second, they offer extensive coverage (such as Google Maps photos), enabling broadly-applicable detection models. Thirdly, they can be utilized to extract map data during offline procedures. Most previous works have a limited size of remote-sensing datasets. Moreover, there are no labeled satellite-images datasets indicating the location of intersections [2, 5, 6].

To our knowledge, there are no other works in the literature estimating the location of intersections in navigation maps [6]. This paper aims to address this gap by presenting a deep learning

SAC'23, March 27- March 31, 2023, Tallinn, Estonia

ACM ISBN 978-1-4503-8713-2/22/04…$15.00
https://doi.org/xx.xxx/xxx_x

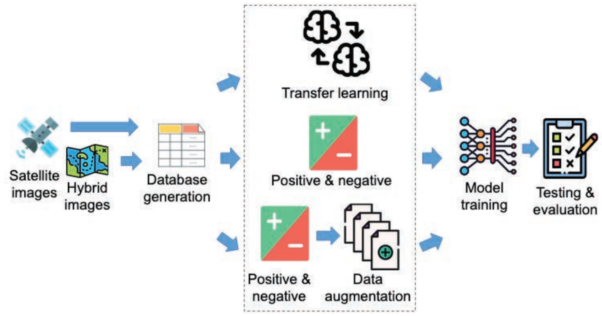

Figure 1: Pipeline of the steps involved in the proposed methodology for predicting the intersection location.

framework to automatically detect the location of intersections from satellite images. We first extended our prior dataset [7] to incorporate the location of intersections and we added a bounding box around the intersections. Then, we trained an object detection model based on convolutional neural networks (CNN). We used the resultant output bounding box from the model to calculate the location of the center of each intersection (longitude and latitude). This is a required pre-step to enabling large-scale annotation of maps with an extra layer of information about intersections.

This paper is organized as follows: Section 2 discusses the methodology. Section 3 addresses the conducted experiments and highlights the challenges. Section 4 shows concluding remarks and addresses the planned work.

2 METHODOLOGY

This section describes the complete process of the proposed automatic detection approach of the intersection's location including a description of the dataset.

2.1 Dataset generation

To curate the dataset, we first downloaded two types of images (satellite and hybrid) from Google Maps, at zoom level nº19 which clearly provides intersection details. The satellite images are unprocessed pictures where roads and intersections may be blocked by other objects such as trees and buildings. Hybrid images are satellite images that Google has labeled to highlight roads and display street names. We created a new dataset annotating the location of each intersection.

Manual annotation is time-consuming and expensive, especially in large-scale datasets. To annotate images automatically, we used two available datasets from the Washington, DC, government website [1]. The first one is the intersection points dataset which has the intersections' locations [2]. The second one is the roadway intersection approach dataset [3], which contains road segments that lead to intersections. Each road segment is defined by several points using latitude and longitude. One of these points is the centre of an intersection. We combined these two data sources to calculate the centre and the number of branches of each intersection. During the training process, we use the negative examples images that are available in the previous version of our dataset [7].

Table 1: Number of images in each of the datasets (training, validation, and testing) grouped by degree. Multiple intersections refer to images that contain more than one intersection.

Degree	Training	Validation	Testing	Total
3	2599	557	558	3714
4	2333	499	501	3333
5	56	12	13	81
6	10	2	3	15
Multiple	139	29	31	199

The annotation process for the object detection problem involves drawing a bounding box around each target object in each image of the collection. To enable the object detection model to learn the features of the target object, each bounding box should surround the intersection so that the branches and crossover are included. We assessed the ideal bounding box size before deciding on the size of the ground truth bounding box. A bounding box of 100x100 pixels only covered 76.84% of the intersections, while a box of 170x170 pixels covers 98% of them. As a result, we used a bounding box of 170x170 pixels. The total number of images in the dataset is 7,342. To develop and evaluate our model, the datasets were randomly split with 70% used for training, 15% for validation, and 15% for testing (see table 1). The data was randomly split three times, and the results presented in our paper are the average of the three splits. We made the script to download and annotate the datasets publicly available [4].

2.2 Approach

We consider road intersections to be distinct identifiable objects, with consistent features (centre point, three or more branches) within an image. Thus, we applied an object detection approach. In the first step, we trained the state-of-the-art object detection model (EfficientDet) [15] to detect the presence and location of intersections. To achieve the best results, we considered different decisions during the training phase (see Figure 1). Firstly, we compared two methods of training the intersection detection model; from scratch and by transfer learning methods. Secondly, we examined the effect of using images with no intersection (negative samples) in training the object detection model. Thirdly, we investigated the effect of using augmentation to extend the dataset.

3 EXPERIMENTS AND RESULTS

This section goes into the specifics of the experiments that are carried out. It also explains the rationale behind our parameter selection and the models' performance. During the experiments, we used the created dataset to train the models.

3.1 Evaluation metrics

For the performance evaluation of object detection models, a variety of measures are used. Average precision (AP) is a popular metric for evaluating the accuracy of object detectors. AP estimates the

[1] https://opendata.dc.gov/
[2] https://opendata.dc.gov/datasets/DCGIS::intersection-points/about
[3] https://opendata.dc.gov/datasets/DCGIS::roadway-intersection-approach/about
[4] https://doi.org/10.21427/g1pb-0s89

area under the curve (AUC) of the precision × recall relationship [13]. AP presents the trade-off between precision and recall.

To define the true and false detections, Intersection Over Union (IOU) is used. The IOU calculates the area of union between the predicted and the actual bounding boxes, divided by the overlapped area between them. We can categorize a detection as correct or incorrect by comparing the IOU with a specified threshold t. The detection is deemed to be accurate if $IOU > t$. If $IOU < t$, the detection is regarded as being inaccurate. The AP can be computed over different Intersection Over Union (IOU) thresholds, from 0.5 to 0.95, step 0.05, usually reported as AP@50:5:95. It also can be evaluated with only candidates over 50% or 75% of IOU, reported as AP50 and AP75 respectively. We used Average Recall (AR) as well. The AR is calculated by averaging the maximum recall values over IOUs and classes for a specified number of detections per image (1, 10 or 100). We used AP and AR as a metric for evaluating our results.

3.2 Detecting the location of the intersections

We used satellite images to figure out where the centres of the intersections are. The ground truth value for the centre location of the intersections was given in an available dataset from the Washington, DC, government website. This is an object detection problem and we used an EfficientDet model to accomplish this task.

3.2.1 Experiment I: Transfer learning versus training from scratch. Transfer learning can be generally used to improve the model's performance. MSCOCO dataset is a benchmark object detection dataset [12]. It contains natural scene images for 80 objects. The EfficientDet is pre-trained on MSCOCO dataset and their pre-training weights are used for the transfer learning method. To investigate its impact on performance, we trained the model from scratch and with a transfer learning strategy. Using a hybrid dataset, the model was trained from scratch and using transfer learning, and it achieved AP@[50:5:95] of 67.93% and 68.87% respectively. Using a satellite dataset, the model was trained from scratch and using transfer learning, and it achieved AP@[50:5:95] of 64.90% and 65.90% respectively.

The key point here is that standard datasets, such as MSCOCO, used for pre-training, are unlikely to provide a boost when training with satellite images. There is a substantial disparity between satellite images and natural scene images. Satellite images capture the roof information of geospatial objects, whereas natural scene images typically catch the profile information of the objects. Intuitively, the item detectors learned from natural scene images are difficult to transfer to satellite images [10]. Consequently, in the next experiments, the model is trained from scratch using our dataset.

3.2.2 Experiment II: Training using positive and negative images. Generally, in object detection problems, the model is trained only with images that contain the target object (positive instances). The object detection model is not trained on images that do not have the target examples (negative examples) and the background is considered a negative example.

In *positive object detector*, the model is trained exclusively with positive cases and tested over both positive and negative examples. This is important because, in practice, all the satellite images of the area of interest (city or country) will be examined by the model to locate intersections in the given area. The performance of our model was calculated using different evaluation matrices (see table 2). In the results, AP over different IOU is low. We examined the model's output to determine the nature of the errors. The fundamental error is that the model is confused between intersections and roads. This is mainly because their contexts are extremely similar. To improve the results, two different approaches are considered: the *Classifier + positive object detector* approach, and the *Positive+Negative object detector* .

Table 2: Comparison between different approaches of training for intersection detection model (the result of a hybrid dataset).

Method	AP@[50:5:95]	AP@[50]	AP@[75]	AR@[50:5:95]
Pos obj detector	67.93%	84.00%	75.90%	82.13%
Classf+pos obj detr	74.47%	92.83%	83.10%	80.57%
Pos+Neg obj detr	74.30%	92.43%	83.07%	81.33%

The *Classifier + positive object detector* approach consists a classifier model as a pre-stage to classify images into intersection and no-intersection. Then, we used the output of the classification stage as an input to the object detection model. For the classification stage, we employed the ResNet152V2 model [9], which was previously trained on a hybrid dataset. As shown in Table 2, the AP over different IOU increase with 6.5%, 8.8% and 7.2%. These results highlight that classifier can successfully classify input images into intersection and no-intersection classes.

The *Positive+Negative object detector* approach is based on training the object detection model using positive and negative images. Comparing the performance with the base approach, the AP is increased over different IOU with 6.4%, 8.43% and 7.2%. The performance of the first (Classifier + positive object detector) and second (Positive+Negative object detector) approaches are nearly identical. The second approach is more efficient in terms of training time as only the object detection model was trained. For these reasons, the second approach has been used in the rest of the paper.

Using satellite images, the model is trained using negative and positive examples (the second approach). On the hybrid images, the roads are highlighted. As the satellite images do not have this feature, they are more challenging and the results are lower. The model achieved AP@[50:5:95] of 68.20%.

3.2.3 Experiment III: The impact of using augmentation techniques. Deep learning requires large amounts of data to create generic models that make solid predictions. Augmentation is a strategy for enhancing the size and quality of the used data for training models [14]. The impact of using augmentation methods during training is examined in this paper. Various augmentation methods such as brightness and contrast, flipping, and Gaussian blur were tested with the EfficientDet model on both satellite and hybrid datasets. Horizontal flip and Gaussian blur augmentations were found to produce the best results, and these can be seen in Table 3.

The amount of improvement by augmenting varies depending on whether the dataset is satellite or hybrid. For the satellite dataset, AP@[50:5:95] and AR@[50:5:95] are increased by 3.37% and 2.47% compared when augmentation is not used. For the hybrid dataset,

Table 3: Augmentation effect on object detection model's performance.

	Satellite images			
Augmentation	AP@[50:5:95]	AP@[50]	AP@[75]	AR@[50:5:95]
No	65.43%	87.63%	73.30%	74.50%
Yes	68.80%	89.53%	76.07%	76.97%
	Hybrid images			
Augmentation	AP@[50:5:95]	AP@[50]	AP@[75]	AR@[50:5:95]
No	73.27%	92.27	82.10%	80.47%
Yes	76.50%	92.77	84.40%	82.83%

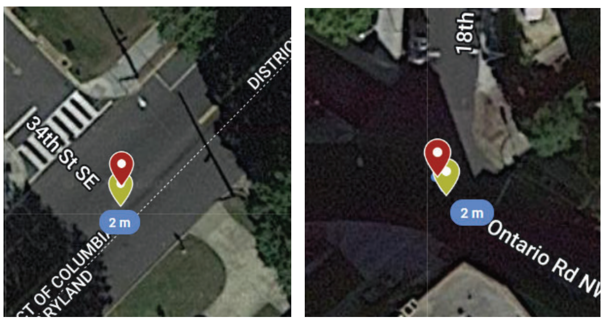

Figure 2: Examples of detected intersections within 2 meters. The green point refers to the ground truth centre location. The red point refers to the predicted centre location.

AP@[50:5:95] and AR@[50:5:95] are increased by 3.23% and 2.36%. Our approach is the first one that uses an object detector model to recognize intersections from satellite images.

The object detection evaluation metrics do not provide sufficient information to evaluate the performance of our method. In this case study, the centre of the intersection is more critical than the bounding box. Therefore, we calculated the difference between the actual centre of intersections and the predicted centre of the intersections. We counted intersection detections within 5 and 10 meters distances from ground truth as true positives. Empirically, the accuracy is 88.6%, within 5 meters and 89.73%, within 10 meters. Most of our detections (approx. 80%) are within 2m of the ground truth (see fig. 2).

4 CONCLUSION AND FUTURE WORK

A global solution to mapping road intersections would benefit route planning, traffic prediction, and road extraction application. Specifically, route planning applications do not provide users with information about intersection locations when embarking on a journey. Satellite images are a potentially global source for enabling the extraction of this information. We present a deep learning framework to detect the location of intersections. EfficientDet can detect the location of the intersection from satellite images with AP@[50:5:95] equals 68.80% and AR@[50:5:95] equals 76.97%. The performance is raised at hybrid images to equal 76.50% AP@[50:5:95] and equal 82.83% AR@[50:5:95].

Our model has been tested with images from the suburbs, it would be interesting in the future to test it in more dense areas of big cities such as New York or London. And also test it over areas covered with trees or with narrow paths. Additionally, one of the limitations of our methodology is that it finds it difficult to detect multiple intersections in one single image. To overcome this problem we would need to train the model with more images of this kind. For future work, we plan to extend our dataset to collect more images of other geo-areas with different road network styles and terrain.

ACKNOWLEDGMENTS

This publication has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 18/CRT/6222.

REFERENCES

[1] Behrang Asadi and Ardalan Vahidi. 2010. Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time. *IEEE transactions on control systems technology* 19, 3 (2010), 707–714.
[2] Jianghua Cheng, Tong Liu, Yueyong Zhou, and Yanye Xiong. 2019. Road Junction Identification in High Resolution Urban SAR Images Based on SVM. In *Proceedings of the International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. Springer, 597–606.
[3] Achituv Cohen and Sagi Dalyot. 2021. Route planning for blind pedestrians using OpenStreetMap. *Environment and Planning B: Urban Analytics and City Science* 48, 6 (2021), 1511–1526.
[4] Dragos Costea, Alina Marcu, Emil Slusanschi, and Marius Leordeanu. 2017. Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2100–2109.
[5] Jiguang Dai, Yang Wang, Wantong Li, and Yuqiang Zuo. 2020. Automatic Method for Extraction of Complex Road Intersection Points from High-resolution Remote Sensing Images Based on Fuzzy Inference. *IEEE Access* 8 (2020), 39212–39224.
[6] Fatma El-Zahraa El-Taher, Ayman Taha, Jane Courtney, and Susan Mckeever. 2021. A systematic review of urban navigation systems for visually impaired people. *Sensors* 21, 9 (2021), 3103.
[7] Fatma El-zahraa El-taher, Ayman Taha, Jane Courtney, and Susan Mckeever. 2022. Using Satellite Images Datasets for Road Intersection Detection in Route Planning. *International Journal of Computer and Systems Engineering* 16, 10 (2022), 411–418.
[8] Shan Fang, Lan Yang, Tianqi Wang, and Shoucai Jing. 2020. Trajectory planning method for mixed vehicles considering traffic stability and fuel consumption at the signalized intersection. *Journal of Advanced Transportation* 2020 (2020).
[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Identity mappings in deep residual networks. In *Proceedings of European conference on computer vision*. Springer, 630–645.
[10] Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. 2020. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing* 159 (2020), 296–307.
[11] P Li, Y Li, J Feng, Z Ma, and X Li. 2020. Automatic Detection and Recognition of Road Intersections for Road Extraction from Imagery. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 43 (2020), 113–117.
[12] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.
[13] Rafael Padilla, Sergio L Netto, and Eduardo AB Da Silva. 2020. A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)*. IEEE, 237–242.
[14] Connor Shorten and Taghi M Khoshgoftaar. 2019. A survey on image data augmentation for deep learning. *Journal of big data* 6, 1 (2019), 1–48.
[15] Mingxing Tan, Ruoming Pang, and Quoc V Le. 2020. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10781–10790.