Journal of Information and Data Management, 2023, 14:1, doi: 10.5753/jidm.2023.3179
This article is licensed under a Creative Commons Attribution 4.0 International License.

Improved generalization of cyclist detection on security cameras with the OpenImages Cyclists dataset

Ednilza E S Nardi [**Universidade de São Paulo** | ***ednilza@ime.usp.br***]
Bruno Padilha [**Universidade de São Paulo** | ***brunopadilha@usp.br***]
Leonardo Tadashi Kamaura [**Universidade de São Paulo** | ***ltkamaura@alumni.usp.br***]
João Eduardo Ferreira [**Universidade de São Paulo** | ***jef@ime.usp.br***]

Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, São Paulo, SP, 05508-090, Brazil.

Received: 25 February 2023 • **Published:** 20 October 2023

Abstract Most large public datasets containing cyclists for training detectors based on Deep Learning have annotations for bicycles and people, but not for cyclists. Even when it is not the case, the quality and quantity of the images are limited. To overcome these limitations, we propose the new OpenImages Cyclists dataset, built through the pre-selection of images from the OpenImages set and a new algorithm for semiautomatic generation of cyclist annotation aided by people and bicycle detectors. A cyclist detector trained with this dataset achieved identification rates up to 78% and 89% in two different sets of images obtained from security cameras at USP, Campus São Paulo - Capital.

Keywords: Deep Learning; Object Detection; Online Object Detection; Real Time Monitoring; Cyclist Detection

1 Introduction

In traffic, a good automated detection of cyclists increases the safety of the cyclist, pedestrians, and vehicles. Cyclists are vulnerable users of public roads [Li *et al.*, 2016] and, like pedestrians, are subject to risky situations, but with different speed and space occupation [Masalov *et al.*, 2019]. In addition, in the context of road monitoring by security cameras, there are specific rules for the circulation of cyclists, different from the rules for pedestrians and motor vehicles.

Automated object detection has been widely studied in several scenarios, such as automated monitoring of public roads by security cameras and autonomous driving, mainly with deep learning methods [Santhosh *et al.*, 2020]. This technique is one of the most successful in object detection, mainly due to its [Zhang *et al.*, 2021] generalization property. However, detection quality in deep learning models depends on large quantity, quality and variability of training data [Zhou *et al.*, 2017].

There is a large amount of data in publicly accessible databases to train deep learning models that recognize and locate people, bicycles, vehicles, and several other objects. For example *VOC-Pascal* [Everingham *et al.*, 2010], *COCO* [Lin *et al.*, 2014], and *Open Images* [Kuznetsova *et al.*, 2020]. In the case of a cyclist, there is little data available and the few sets that exist, such as *Tsinghua-Daimler* [Li *et al.*, 2016] and *Specialized Cyclist* [Masalov *et al.*, 2019] were collected with a camera mounted in front of a vehicle in restricted geographic regions, which recorded cyclists at an angle very different from that commonly found on security cameras (i.e. on high poles and pointing downwards). This, as demonstrated in section 4, negatively affects generalizability to security camera environments. The set *MIO-TCD* [Luo *et al.*, 2018], although coming from security cameras, contains just under two thousand images of bicycles including the driver in the same annotation. These images are of low resolution and low quality, since they contain a lot of video compression artifacts, aside from the cyclists appearing, in general, very small and with little details.

Training an object detector to be capable of accurately locating people and bicycles may not be enough for the correct identification of cyclists. A cyclist is a composite object resulting from a specific interaction of a person riding a bicycle. For example, a person standing next to a bicycle is not a cyclist, nor is a person pushing a bicycle on the sidewalk. Thus, it is necessary for the cyclist detector to learn, in addition to the characteristics of a bicycle, the characteristics of a person in the specific position of riding the bicycle, possibly wearing clothing and protective equipment suitable for this practice.

The *Open Images* dataset provides approximately 18,000 images containing cyclists, of which about a third have all of the bicycles shown in each image duly annotated, that is, bounding boxes without cutting parts of the objects and with the minimum size required. For those same images, few people were annotated or had bad bounding boxes. It is very important that most of the target objects in the same image are annotated so that a detector based on deep learning can satisfactorily learn to identify them. Alternatively, people detection can be done with models trained on the *COCO* dataset. For example, the object detector *YOLOv4* [Bochkovskiy *et al.*, 2020] provides configuration of weights pre-trained exhaustively in *COCO*, thus eliminating the need to manually annotate people in images containing cyclists from *Open Images*.

Thus, this article presents a new algorithm for the semi-automated generation of annotations of cyclists present in a

Figure 1. Example images from public datasets for object detection, with annotations.

subset of public images available on *Open Images*. Through a manual pre-selection of images containing people riding a bicycle with the bicycles well annotated, in addition to the aid of the object detector *YOLO* pre-trained in *COCO* for the automated detection of people, this algorithm associates people in a riding position with their respective bicycle to generate cyclist annotations, completely eliminating the need for manual annotations.

The main contribution of this article is OpenImages Cyclists, a new dataset based on a subset of images from *Open Images* containing annotations of cyclists. When training *YOLO* on this new dataset, we obtained significantly better results compared to when we trained this model on the other datasets that contain cyclists' annotations. Through the transfer learning technique [Zhuang *et al.*, 2020], the environment chosen to validate the results obtained was the Campus Cidade Universitária Armando Salles de Oliveira of the University of São Paulo (USP), located in the Butantã district, in São Paulo, and known as Campus São Paulo - Capital. This campus has a large real-time monitoring infrastructure [Ferreira *et al.*, 2018] and has restrictions on days and times for the practice of sport cycling. Due to its size of approximately 3,650,000 m^2, automated cyclist monitoring makes a great contribution not only to safety but also to the good coexistence of cyclists, pedestrians and cars.

The rest of this article is organized as follows: in section 2 we present a discussion of the main related works; in section 3 we present the details of the proposed dataset as well as construction details; in section 4, we show that the proposed dataset allows the training of cyclist detectors with greater generalization capacity for the environment of USP, Campus São Paulo - Capital; finally, section 5 brings conclusions and possibilities for future works.

This article is an extension of the article (in Portuguese): "*OpenImages Cyclists*: Expandindo a Generalização na Detecção de Ciclistas em Câmeras de Segurança" presented at SBBD2022 - Full Papers [Nardi *et al.*, 2022]. Section 2 was extended as a result of a broader discussion regarding the main related works, section 3 was extended to better reflect the motivation of the process employed to tackle the challenge of monitoring cyclists by security cameras, and section 4 was extended with a new experiment to evaluate the generalization capacity for the USP environment on a larger amount of data, with a custom detector trained with the proposed dataset.

2 Related Works

2.1 Cyclist Detection

Several articles published in recent years have studied the detection of pedestrians, cyclists and motor vehicles, mainly using approaches based on computer vision, deep learning models or sensors, with regard to the evolution of autonomous vehicles and the safety of vulnerable road users. However, to the best of our knowledge, few articles have studied the detection of cyclists for the purpose of monitoring public roads.

Zou et al. provide a comprehensive analysis of research over the last 20 years in the field of object detection at [Zou *et al.*, 2023], covering topics such as milestone detectors, detection datasets, metrics, and state-of-the-art detection methods, among others.

In [Vasconcelos *et al.*, 2016b], the authors draw a parallel between computer vision and the Deep Neural Network for the detection task, applied to a pedestrian detector. In [Vasconcelos *et al.*, 2016a], they seek to improve the generalization of pedestrian detection in open scenarios through data enrichment, and this method can be applied to other detectors. [Joseph *et al.*, 2021] also studies the object detection problem in the open world and provides a solution based on contrastive clustering and energy based unknown identification.

The article [Ku *et al.*, 2019] presents a monocular 3D object detection method with the reconstruction of 3D objects from detections in a 2D scene and this method is validated in the KITTI benchmark used for autonomous driving applications, including pedestrian and cyclist classes. In [Fan *et al.*, 2022] and [Saleh *et al.*, 2017] the authors address the detection of 3D objects based on LiDAR (Light Detecting And

Figure 2. Images (duly anonymized) of Campus São Paulo - Capital (USP).

Ranging) for autonomous driving, in [Fan *et al.*, 2022] they seek to remedy the loss of information from downsampling when the ratio between object size and scene size input is significantly smaller compared to 2D detection cases such as in the case of cyclists.

Other approaches aimed at pedestrian and cyclist safety on public roads are presented in [Vial *et al.*, 2023], with the use of mobile sensors for traffic tracking applications, and in [Abadi *et al.*, 2022], [Fang and López, 2020] and [Ahmed *et al.*, 2019] who propose prediction of cyclist behavior by autonomous vehicles or driver assistance systems, using deep neural networks, which investigate movement tracking and pose estimation. In [Pool *et al.*, 2019] the authors propose a neural network that identifies contextual information for cyclist path prediction by autonomous vehicles

Researchers have studied ways to encourage greater use of roads by cyclists, aiming at green mobility and eco-driving in urban areas, as in [Dabiri *et al.*, 2022]. The increased use of public roads by cyclists makes the issue of safety for this type of user more urgent, increasing the importance of correct detection of cyclists.

2.2 Datasets

Over the last few years, several public datasets composed of thousands of images with different classes of annotated objects have been created. These sets have accelerated the evolution of deep learning algorithms for object detection. Datasets like *PASCAL Visual Object Classes (VOC)* [Everingham *et al.*, 2010], *Microsoft COCO* [Lin *et al.*, 2014], *Open Images* [Kuznetsova *et al.*, 2020; Krasin *et al.*, 2017], *MIO-TCD* [Luo *et al.*, 2018], *Tsinghua-Daimler* [Li *et al.*, 2016], and *Specialized Cyclist Detection* [Masalov *et al.*, 2019] contains images with objects annotated in everyday scenes in which these objects are presented in their natural contexts. The *VOC-PASCAL*, *COCO*, and *Open Images* datasets are general purpose datasets composed of images gathered from all over the internet, while the others are task specific, with annotations of objects of a class or a specific group of classes. The *MIO-TCD* dataset is tailored for traffic monitoring through traffic cameras, and the datasets *Tsinghua-Daimler* and *Specialized Cyclist Detection* are focused on cyclists, for application in assisted or autonomous driving scenarios using cameras in vehicles.

VOC-PASCAL has 11,540 images with 27,450 objects labeled in 20 classes, including 603 images with a bicycle, *COCO* has 328,000 images with 2.5 million objects labeled in 91 classes, around 7 thousand images with a bicycle and *Open Images* has about 9 million images with 16 million objects annotated in 600 classes, with about 18 thousand images with a bicycle. These datasets are often used in competitions to further the development of computer vision with deep learning methods, including object detection.

MIO-TCD is a dataset for vehicle classification and location, subdivided into images for classification and images for object detection, acquired by traffic surveillance cameras deployed across Canada and the United States. In *MIO-TCD-Classification* the images are clippings of scenes containing a single object of interest and in *MIO-TCD-Localization* the images are complete scenes, similar to those of the other datasets. It contains 137,743 images with 416,277 objects annotated in 12 classes, with 1,933 images with a bicycle, whose annotation also includes the person riding it, coinciding with our definition of a cyclist. Luo et al. trained and evaluated the location of vehicles with this data set, using the methods *Faster R-CNN* [Ren *et al.*, 2015], *SSD* [Liu *et al.*, 2016], *YOLO* [Redmon *et al.*, 2016; Redmon and Farhadi, 2017], method by Wang et al. [Wang *et al.*, 2017], and method by Jung et al. [Jung *et al.*, 2017].

Tsinghua-Daimler has 30,406 images in total, with 22,161 annotated cyclists. Their images were recorded by a camera mounted on a moving vehicle in Beijing's urban traffic for about 6 hours spread over 5 days, in an area with a high concentration of cyclists and pedestrians. Xiaofei et al. evaluated detectors based on *ACF* [Dollár *et al.*, 2014], *DPM* [Felzenszwalb *et al.*, 2010] and *R-CNN* [Girshick *et al.*, 2014] with this dataset for the detection of cyclists. There were three groups of tests with the set divided into easy, moderate and difficult, depending on the level of occlusion and proportion of the size of the cyclists in relation to the image.

Specialized Cyclist Detection has 62,297 images in total, with 30 different cyclists, and about 18,200 annotated cyclists. The images of the dataset were recorded by a camera mounted on a vehicle in two different locations, with two different weather and lighting conditions. This set contains images with easy, moderate and difficult detection levels, defined by the cyclists' occlusion level, however, Masalov et al. did not evaluate any detection method with this dataset.

While the aforementioned publicly available datasets contributed significantly for advancements in cyclists detection and traffic monitoring, they fail to attain good generalization when evaluated on a large and diverse CCTV network as USP EMS. To overcome this limitation, the new *OpenImages Cyclist* dataset, presented in this article, which was built from *Open Images*, is more comprehensive for cyclist detection than public datasets specialized in cyclists. The images from *Open Images* were collected from the *Flickr* image sharing community and have different viewing angles and varied scales and dimensions [Kuznetsova *et al.*, 2020]. **Figure 1** presents some samples of images of the mentioned datasets.

3 *OpenImages Cyclist* Dataset

The new dataset *OpenImages Cyclists* was created to enable the detection of cyclists in security camera images, especially in the environment of USP, Campus São Paulo - Capital. As presented in section 4, this dataset not only allowed the training of a cyclist detector with a very good performance in the USP images (**Figure 2**) but also presented competitive results in the other sets of data containing cyclists. In this section, we present the construction details of this dataset.

3.1 Motivation

Deep learning methods tend to work well with plenty of labeled training data whose probability distribution is the same as the test data. However, in many real-world situations, there is not enough annotated data available for proper training and testing, and obtaining more data is both difficult and time-consuming to annotate. The use of transfer learning [Zhuang *et al.*, 2020] can mitigate this problem. This is a methodology that allows the transfer of knowledge acquired from a source domain to a related target domain, even with different distributions between them.

In the case studied, images of cyclists obtained by cameras installed in vehicles constitute a domain while images of cyclists obtained by security cameras constitute another domain. Both are related but have different distributions. A cyclist's appearance in images from each of these domains can be very different due to different camera angles.

Although we have being collecting hundreds of hours of surveillance footage on the USP campus, that data is not labeled. Annotating data with bounding boxes on the objects of interest is a very strenuous and prone to error task when carried out manually. Hence the need for a tool that leverages transfer learning to assist in this process.

Of the datasets with images of cyclists that were found to be publicly available, *Tsinghua-Daimler* and *Specialized Cyclist Detection* were created mainly for use in autonomous driving projects and only contain images obtained by cameras in vehicles, and *MIO-TCD* was created to aid the monitoring of various types of vehicles on highways by surveillance cameras and contains few images of interest, in which cyclists, in general, appear in reduced size or the image quality is low.

Another dataset used for training object detectors with a large amount of data is *Open Images*, which contains many images of cyclists, under the most diverse observation angles. Therefore, the distribution of data relating to cyclists in this set is closer to the distribution of data obtained by security cameras. However, in this dataset there is no cyclist annotation and many images contain either the bicycle or the rider annotated (sometimes neither). For each cyclist, there can be two annotations, bicycle and person, whose bounding boxes intersect.

There are several transfer learning approaches that can be applied to solve a variety of problems and can be interpreted from a model or data perspective. A possible data-based approach focuses on knowledge transfer by transforming or adjusting data in order to reduce the distribution difference between instances of the source domain and the target domain.

Considering the difficulty in obtaining annotated data to train cyclist detectors and the proximity between the distributions of the *Open Images* dataset and the data of interest, we created the labels for the cyclists of the subset of interest of the *Open Images* with a semiautomatic semi-automated process. We built the *OpenImages Cyclist* dataset from a selection of *Open Images* images with the new annotations.

This strategy proved to be valid since the *OpenImages Cyclist* dataset allowed the training of a deep learning model to detect cyclists in security camera images with greater precision than those obtained by training models with other datasets, as shown in section 4.

3.2 Dataset Construction

Of the almost 18,000 images with bicycles annotated by bounding boxes available in *Open Images v6* (as of 05/08/2019), about 1/3 feature sport cyclists in various scenarios [Krasin *et al.*, 2017]. The images in this set contain many bicycles that are not annotated. Also, most person annotations on these same images are inaccurate or missing.

We manually selected images containing cyclists with correctly annotated bicycles. From this selection, we obtained a set of approximately 6,000 images with cyclists and their respective bicycles annotated, without considering the person annotations, at this time. With the aid of the *YOLOv4* object detector, exhaustively pre-trained on the *COCO* dataset, we automated the creation of new annotations for the person in this obtained dataset. We then performed a visual sampling assessment (about 1200 randomly selected images) to primarily ensure that people riding bicycles were satisfactorily detected. According to [Robert *et al.*, 2022], *YOLOv4* is among the best performing detectors in terms of mean accuracy over the *COCO* dataset, which contains a large number of annotated person instances (262,465), proving to be sufficient for our experiments.

To generate cyclist annotations, once the annotations of bicycles and people have been obtained, it is necessary to correctly associate each rider with their respective bicycle. Algorithm 1 automatically makes this composition of annotations based on the intersection between people and bicycles, using the Intersection over Union (IOU) metric for this. This technique consists of calculating the intersection area divided by the area of the union between two bounding boxes (person and bicycle). The result is a value between zero and one, in which zero represents no intersection and one represents that one bounding box is fully enclosed by the other.

Algorithm 1, described below, starts by receiving as input a list of images along with the bicycle annotations of the preselected images and returns the bicycle annotations for these same images (line 1). For each input image, bicycle annotations are placed in the *bikes_bb* list (line 5) and the image is subjected to person detection (line 6). Then, for each bicycle, the algorithm goes through the list of people found in that image and computes the IOU for all person-bicycle pairs (lines 7-16). The largest IOU found most likely corresponds to the person riding the bicycle. Both people and bicycles for which there is no intersection with their complementary classes are discarded. (e.g. person in the background watching cyclists or an occasional bicycle without a rider). The new cyclist annotations correspond to the smallest bounding boxes resulting from the union of the annotations of each bicycle with its respective rider.

Algorithm 1 Algorithm for automated generation of cyclist annotations

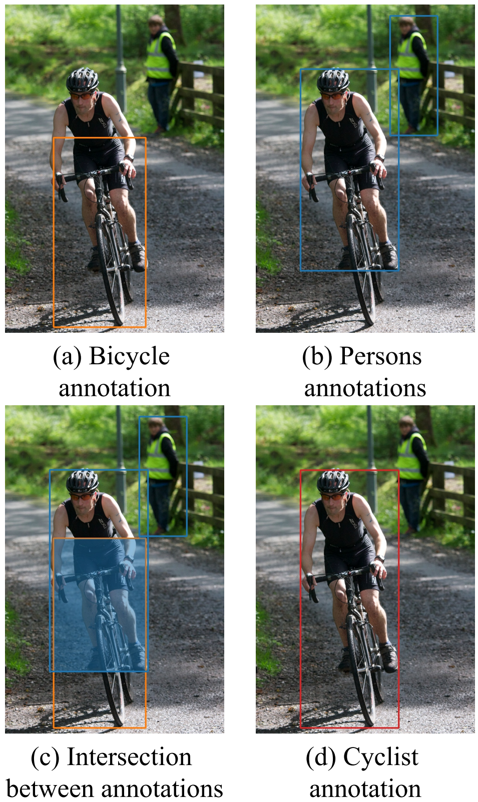
Input: List of images with annotated bicycles
Output: List of images with annotated **cyclists**

```
L ←List of images with annotated bicycles
C ← ∅                                  ▷ List of cyclist annotations by image
for i ∈ L do
    cyc_bb ← ∅
    bikes_bb ← bike_annotations(i)
    persons_bb ← detect_person(i)
    for b ∈ bikes_bb do
        max_iou ← 0
        p_tmp ← ∅
        for p ∈ persons_bb do
            iou ← calculate_iou(b, p)
            if iou > max_iou then
                max_iou ← iou
                p_tmp ← p
            end if
        end for
        if max_iou > 0 and p_tmp ≠ ∅ then
            c ← join_bb(b, p_tmp)
            cyc_bb.append(c)
        end if
    end for
    C.append(i, cyc_bb)
end for
return C
```

Figure 3 illustrates how this process works, in which the **Subfigure 3a** contains bicycle annotations from *Open Images*. **Subfigure 3b** is the result of person detection, containing the person riding the bicycle and also a spectator. The shaded region in **Subfigure 3c** corresponds to the intersection area between the two objects. Finally, in **Subfigure 3d** a new cyclist annotation is generated and the person detected in the background is discarded.

3.3 Dataset Overview

The automated generation of cyclists' annotations resulted in 5,463 images with 15,597 instances of cyclists in everyday scenes. Each image contains an average of three cyclists, most of whom have some degree of occlusion or are truncated. Only 545 cyclists appear in full view. Image resolutions are varied, for example, 1024 x 494, 852 x 768, 1024 x 1024. The relative size of cyclists is also quite varied, occupying from 0.3% to more than 90% of the image area. During the training of the cyclist detector, as presented in section 4, this data is dynamically divided between training and testing with a ratio that can vary from 80/20 to 90/10 based on the strategy of *K-Folding*. The annotations of cyclists produced in this article, as well as the indication of the images used, are available at https://data.ime.usp.br/oic.



(a) Bicycle annotation (b) Persons annotations (c) Intersection between annotations (d) Cyclist annotation

Figure 3. Composition of bicycle and person annotations to create cyclist annotation (Source Open Images). Person in the background without a bicycle (IOU = 0) is discarded in the final result (OpenImages Cyclist dataset).

4 Experimental Results

We performed some experiments to compare the new dataset with other datasets containing annotations of cyclists. Only images with annotations of cyclists were part of the experiments. We used 5463 images from *OpenImages Cyclists*, 1933 images from *MIO-TCD-Localization*, 13655 images from *Tsinghua-Daimler* and 7687 images from *Specialized Cyclist Detection*.

The object detection model chosen to evaluate *OpenImages Cyclists* was *YOLOv4*. According to [Bochkovskiy *et al.*, 2020], version 4 of *YOLO* is presented as an efficient object detection model, created from the composition of state-of-the-art methods. They compare it to other state-of-the-art object detectors using the *COCO* dataset and find that *YOLOv4* is superior to the fastest and most accurate detectors in terms of frame rate processed (FPS) and average precision (AP). Also according to [Bochkovskiy *et al.*, 2020], *YOLOv4* is twice as fast as *EfficientDet* [Tan *et al.*, 2020] with comparable performance and improves the AP and FPS of *YOLOv3* [Redmon and Farhadi, 2018] by 10% and 12%, respectively.

Zaidi et al. have done an in-depth look at the top deep learning based object detectors at [Zaidi *et al.*, 2022], providing a comprehensive review of this type of detector. They considered *YOLOv4* the state-of-the-art for real-time single-stage detectors. They evaluated the performance of the models based on the results of their articles, comparing average precision and frames per second processed in inference time.

Considering the best compromise between accuracy and speed of the object detectors evaluated by both [Bochkovskiy *et al.*, 2020] and [Zaidi *et al.*, 2022], in our experiment we

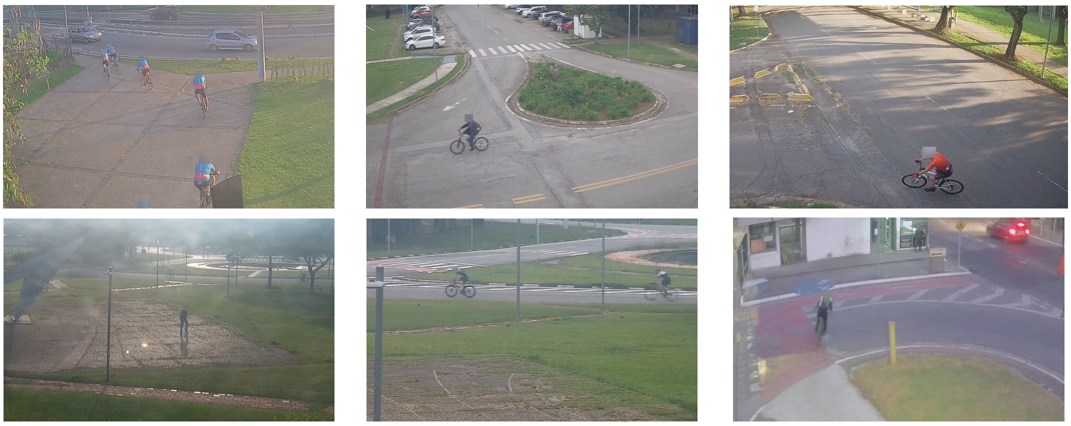


Figure 4. Images (duly anonymized) from other cameras on Campus São Paulo - Capital (USP)..

used only *YOLOv4* to evaluate and compare the datasets. One of the objectives of this article is to provide a solution that can run in real time in the USP camera monitoring environment.

For the sole purpose of testing a cyclist detector trained on the new *OpenImages Cyclists*, we created the *USP Cyclists* dataset with images obtained from the security cameras of the monitoring infrastructure of the Campus São Paulo - Capital (USP). In this first experiment set, 284 images with an average of 4 cyclists per image were manually annotated. These images, similar to those illustrated in **Figure 2**, in general, depict scenes with sport cyclists. Later, a larger set of USP images was annotated for a new experiment.

At first, *USP Cyclists*' cardinality may seem insufficient to assess the generalization ability of the detectors evaluated in section 4.3. However, these images are a sample with a uniform distribution of data collected from June 2021 to February 2022 by nine cameras positioned at different locations and angles on the USP campus. The dataset proposed in this article was used to help annotate other registered cyclists on the USP campus and create a larger dataset used in a new experiment.

4.1 Methodology

The first experiment consisted of training detector models with the standard *YOLOv4* network in each of the datasets and subsequent evaluation of the average precision (AP), considering $IOU \geq 0.50$, of each model for detecting cyclists across all datasets. The training was done using the *k-folding* technique for $k = 5$ (80% of data for training and 20% for testing) in 21,000 batches of 64 images, totaling 266 epochs. We present the results in **Table 1**.

The second experiment consisted of the evaluation of AP, considering $IOU \geq 0.50$, $Precision$, $Recall$, $F1 - score$, TP (true positive classifications), FP (false positive classifications), and FN (false negative classifications) of the detectors for detecting cyclists in *USP Cyclists* dataset. *USP Cyclists* dataset was used exclusively for evaluating the generalization of the detectors, therefore, it was not used for training any detection model. We present the results in **Table 2**.

In the third experiment, we trained a model with all images of cyclists from the sets *MIO-TCD-Localization*, *Tsinghua-Daimler*, and *Specialized Cyclist Detection* combined (23,275 images) and evaluated the detection of cyclists in *USP Cyclists* dataset, obtaining AP, with $IOU \geq 0.5$, of 65.21%, which is smaller than the AP of detection of cyclists, in this same set, of the model trained in *OpenImages Cyclists* (AP = 77.98%)

Complementarily, in the fourth experiment, also using *USP Cyclists* dataset for evaluation, the change from $IOU \geq 0.5$ to $IOU \geq 0.75$ caused a different reduction of AP in each model, depending on the dataset used for training. The model trained with *OpenImages Cyclists* reduced it in 30%, the one trained with *MIO-TCD-Localization*, in 50%, the one trained with *Tsinghua-Daimler*, in 44%, and the one trained with *Specialized Cyclist Detection*, in 48%. This experiment shows that the proposed dataset, in addition to generalizing better than the other datasets for cyclist detection, as presented in the 4.3 section, is also more accurate.

4.2 Larger Dataset Experiment

With the help of the dataset proposed in this article, we created the *Larger USP Cyclists* dataset, with 2000 images obtained from security cameras of the monitoring infrastructure of Campus São Paulo - Capital (USP) with 4084 annotations of cyclists, similar to *USP Cyclists* dataset, but with a greater number of instances. We intended to evaluate and compare the generalization capacity of the cyclist detectors described in the 4.3 section in a larger dataset with a more detailed analysis. This dataset was also used exclusively for evaluating the generalization of the detectors, not being used for training any detection model.

Using a detector trained with the standard *YOLOv4* network on the *OpenImages Cyclists* dataset, we generated annotations of cyclists for the images collected by the nine cameras positioned at different locations and angles on the USP campus, which is a source similar to that of the *USP Cyclists* dataset. used in the evaluation of the 4.3 section.

From these annotated images, we obtained a random sample with uniform distribution, whose automatic generated annotations were manually verified and corrected. This sample constituted the *Larger USP Cyclists* dataset, whose images are similar to those illustrated in **Figure 2** and **Figure 4** and portray both sports and non-sports cyclists, the former being the majority group.

The experiment consisted of the evaluation of AP, con-

Table 1. Comparison between detector performance (AP)

	OIC	MIO	Daimler	Specialized
YOLO_{OIC}	**86.37%**	**99.72%**	63.19%	82.33%
YOLO_{MIO}	35.34%	92.84%	10.85%	59.39%
$\text{YOLO}_{\text{Daimler}}$	63.02%	27.41%	**73.59%**	81.08%
$\text{YOLO}_{\text{Specialized}}$	76.93%	36.35%	48.49%	**96.01%**

Table 2. Capability to generalize the detectors to *USP Cyclists* dataset

	YOLO_{OIC}	YOLO_{MIO}	$\text{YOLO}_{\text{Daimler}}$	$\text{YOLO}_{\text{Specialized}}$
AP	**77.98%**	52.82%	29.36%	40.06%
Precision	**0.93**	0.86	0.81	0.85
Recall	**0.71**	0.48	0.21	0.34
F1-score	**0.80**	0.62	0.33	0.48
TP	**494**	336	146	236
FP	39	53	**35**	41
FN	**203**	361	551	461

sidering $IOU \geq 0.50$, $Precision$, $Recall$, $F1-score$, TP, FP, and FN of the detectors for detecting cyclists in *Larger USP Cyclists* dataset. We present the results in **Table 3**.

Figure 2 and **Figure 4** illustrates camera positions found in the USP Campus environment, from which the data for the experiments were obtained.

4.3 Results analysis

For identification in **Table 1**, **Table 2** and **Table 3**, we name the detectors trained with images originating from the datasets *OpenImages Cyclist*, *MIO-TCD-Localization*, *Tsinghua- Daimler* and *Specialized Cyclist Detection*, respectively, as YOLO_{OIC}, YOLO_{MIO}, $\text{YOLO}_{\text{Daimler}}$ and $\text{YOLO}_{\text{Specialized}}$. For the same data sets, we identify their respective test sets by *OIC*, *MIO*, *Daimler* and *Specialized*.

Table 1 presents the comparison between the performance of the detectors. In the rows are the detectors and in the columns are the test datasets. Each cell shows AP of the detector indicated in the respective row for detecting cyclists in the test set images indicated in the respective column. The main diagonal represents the cases in which the test set and the training set come from the same dataset and contains, in general, the largest values of AP, per column. This is expected, as the images from both sets have the same probability distribution. The *MIO* exception is possibly due to the relatively small number of cyclists in the YOLO_{MIO} training data.

We also observe from **Table 1**, that the YOLO_{OIC} detector has better AP for cyclist detection in all test sets, except for the diagonal cases. This result corroborates the initial hypothesis that, with the use of good quality images, with good variability and good annotations, it is possible to train more accurate object detectors and, in the case of the USP images, resulting in a more generalizable model.

Table 2 presents the generalization capability of the detectors for the USP domain. In the columns are the detectors and in the rows are the metrics regarding the performance of each detector to find out cyclists in the *USP Cyclists* dataset. Greater AP and $F1 - score$ of the YOLO_{OIC} detector for detecting cyclists in *USP Cyclists* dataset, reinforces the better generalization capacity of the detector trained with the *OpenImages Cyclist* dataset, already indicated by **Table 1**. In addition, the good performance of the YOLO_{OIC} detector in *USP Cyclists* allows its application in monitoring the roads of Campus São Paulo - Capital of USP for the detection of professional cyclists.

The lower generalization capacity of the other detectors can possibly be explained by the small number of cyclists in the YOLO_{MIO} training data, by the little variability of the scenes in the $\text{YOLO}_{\text{Daimler}}$ and $\text{YOLO}_{\text{Specialized}}$ training data, being a little larger in the latter, and due to the difference in observation angles between the images of *MIO* and *USP Cyclists* in relation to *Daimler* and *Specialized*

Table 3 also presents the generalization capability of the detectors for the USP domain, similar to **Table 2**, but considering a larger set of images. In the columns are the detectors and in the rows are the metrics regarding the performance of each detector for identifying cyclists in the images of *Larger USP Cyclists* dataset.

Comparing **Table 3** with **Table 2**, the values of AP, recall and, consequently, F1-score of all detectors were higher with the *Larger USP Cyclists* dataset than with the *USP Cyclists* dataset. This is due to adjustments in the settings of some cameras to improve the overall quality of images (e.g. reduced the amount of motion blur in low light conditions, slightly improved contrast and sharpness), and a larger number of images from the *Larger USP Cyclists* dataset. An example of the improvement achieved is shown in **Figure 5**.

This adjustment may have made the images more suitable for the detector trained with the *Specialized Cyclist Detection* dataset, which would explain the greater difference for the $\text{YOLO}_{\text{Specialized}}$ in the two tables.

The highest values of the metrics AP and $F1 - score$, for the YOLO_{OIC} detector in relation to the other detectors, found in **Table 3**, as well as **Table 2**, confirm its better generalization capacity, already indicated by **Table 2**, and its usefulness in monitoring cyclists on the streets of Campus São Paulo - Capital of USP.

In **Table 2** the $Precision$ metric was better for YOLO_{OIC}, although it was greater than 80% for all detectors. This same metric was similar across detectors in **Table 3**, except for

Table 3. Capability to generalize detectors to *Larger USP Cyclists* dataset

	YOLO_{OIC}	YOLO_{MIO}	$\text{YOLO}_{\text{Daimler}}$	$\text{YOLO}_{\text{Specialized}}$
AP	**89.01%**	67.07%	38.61%	80.70%
Precision	0.89	0.90	**0.92**	0.79
Recall	**0.88**	0.60	0.26	0.78
F1-score	**0.89**	0.72	0.41	0.79
TP	**3594**	2453	1062	3187
FP	441	273	**93**	846
FN	**490**	1631	3022	897

old image new image

Figure 5. Images (duly anonymized) from the same USP camera exemplifying the improvement achieved by adjustments to its settings.

$\text{YOLO}_{\text{Specialized}}$, because despite having a high rate of correct positive classifications, it had the highest rate of incorrect positive classifications. Considering that the power of human attention is limited, the rate of 89% of correct positive classifications of YOLO_{OIC} detector in the identification of cyclists is of great aid for monitoring by security cameras.

The $Recall$ metric was quite different between detectors in both tables, indicating that the success in the positive classifications of YOLO_{OIC} detector was considerably higher than that of the other detectors.

A high value of $Recall$ is important in the case of road monitoring, as a false negative, that is, the non-identification of a cyclist may pose a greater security risk than a false positive, that is, an object that is not a cyclist to be identified as such.

The results of detector YOLO_{OIC}, which was trained on *OpenImages Cyclists* dataset, in detecting cyclists in datasets *USP Cyclists* and *Larger USP Cyclists* indicate that dataset *OpenImages Cyclists* allows generalization in the detection of cyclists.

The new dataset *OpenImages Cyclists*, thanks to its construction process, presents greater diversity in the images than the other datasets, with variation in the observation angle, size of the cyclist in relation to the image, position, color, background, number of cyclists. In addition, its images are of good quality because, in general, *Open Images* has images with a great diversity of scenes and of superior quality than other public datasets in terms of resolution, sharpness and lighting, as the community *Flickr* allows for this image quality [MacAskill, 2018]. Thus, it is expected that models trained on this dataset will result in more accurate detectors with greater generalization capacity than those trained on other datasets.

5 Conclusion and Future Works

The new *OpenImages Cyclists* dataset substantially improved the cyclist detection precision in USP security camera images, a potentially replicable result in similar monitoring environments. The experiment with the larger test set confirms this result. The variability of the training data, expressed by the variation in camera positioning, lighting, and image quality, favors the generalization of detection.

An object detector based on Deep Learning and trained on this dataset certainly contributes to increasing the safety of cyclists, who often need to share roads with cars and pedestrians. The good performance of the *YOLOv4* detector, trained on *OpenImages Cyclists*, when evaluated on the *Tsinghua-Daimler* and *Specialized Cyclist Detection* datasets is a strong indication that our dataset could also be applied in the context of autonomous driving.

The next steps after this article include: 1) the detection of a cyclist squad, for which the dataset *OpenImages Cyclists* will be of fundamental importance; 2) the creation of a new annotation for classes of objects composed of other objects, such as motorcyclists, using the same process that was used to create the annotation of cyclists.

Acknowledgements

We thank the funding agencies that fund our research activities: FAPESP and CNPq.

Funding

This research was funded by FAPESP no. 2020/06950-4 (Center for Research and Development on Live Knowledge) and CNPq no. 308820/2021-5.

Authors' Contributions

EN, BP and JF contributed to the conception of this study. LTK performed the experiments. EN is the main contributor and writer of this manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no conflicting interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in

- OpenImages Cyclist: `https://data.ime.usp.br/oic`.
- Open Images v6: `https://storage.googleapis.com/openimages/web/download.html`

- MIO-TCD: `https://tcd.miovision.com/challenge/dataset.html`
- Tsinghua-Daimler: `http://www.gavrila.net/Datasets/Daimler_Pedestrian_Benchmark_D/Tsinghua-Daimler_Cyclist_Detec/tsinghua-daimler_cyclist_detec.html`
- Specialized Cyclist Detection: `https://www.mrt.kit.edu/software/datasets.html`.

References

Abadi, A. D., Gu, Y., Goncharenko, I., and Kamijo, S. (2022). Detection of cyclists' crossing intentions for autonomous vehicles. In *2022 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6. DOI: 10.1109/ICCE53296.2022.9730559.

Ahmed, S., Huda, M. N., Rajbhandari, S., Saha, C., Elshaw, M., and Kanarachos, S. (2019). Pedestrian and cyclist detection and intent estimation for autonomous vehicles: A survey. *Applied Sciences*, 9(11):2335. DOI: 10.3390/app9112335.

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *ArXiv*, abs/2004.10934. DOI: 10.48550/arXiv.2004.10934.

Dabiri, A., Hegyi, A., and Hoogendoorn, S. (2022). Optimized speed trajectories for cyclists, based on personal preferences and traffic light information-a stochastic dynamic programming approach. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):777–793. DOI: 10.1109/TITS.2020.3014448.

Dollár, P., Appel, R., Belongie, S., and Perona, P. (2014). Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545. DOI: 10.1109/TPAMI.2014.2300479.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338. DOI: 10.1007/s11263-009-0275-4.

Fan, L., Pang, Z., Zhang, T., Wang, Y.-X., Zhao, H., Wang, F., Wang, N., and Zhang, Z. (2022). Embracing single stride 3d object detector with sparse transformer. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8448–8458. DOI: 10.1109/CVPR52688.2022.00827.

Fang, Z. and López, A. M. (2020). Intention recognition of pedestrians and cyclists by 2d pose estimation. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4773–4783. DOI: 10.1109/TITS.2019.2946642.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645. DOI: 10.1109/TPAMI.2009.167.

Ferreira, J. E., Antônio Visintin, J., Okamoto, J., Cesar Bernardes, M., Paterlini, A., Roque, A. C., and Ramalho Miguel, M. (2018). Integrating the university of são paulo security mobile app to the electronic monitoring system. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1377–1386. IEEE. DOI: 10.1109/BigData.2018.8622069.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 580–587. IEEE. DOI: 10.1109/CVPR.2014.81.

Joseph, K. J., Khan, S., Khan, F. S., and Balasubramanian, V. N. (2021). Towards open world object detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5826–5836. DOI: 10.1109/CVPR46437.2021.00577.

Jung, H., Choi, M.-K., Jung, J., Lee, J.-H., Kwon, S., and Jung, W. Y. (2017). Resnet-based vehicle classification and localization in traffic surveillance systems. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 934–940. IEEE. DOI: 10.1109/CVPRW.2017.129.

Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Veit, A., *et al*. (2017). Openimages: A public dataset for large-scale multi-label and multi-class image classification. https://github.com/openimages.

Ku, J., Pon, A. D., and Waslander, S. L. (2019). Monocular 3d object detection leveraging accurate proposals and shape reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2019.01214.

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., *et al*. (2020). The open images dataset v4. *International Journal of Computer Vision*, 128(7):1956–1981. DOI: 10.1007/s11263-020-01316-z.

Li, X., Flohr, F., Yang, Y., Xiong, H., Braun, M., Pan, S., Li, K., and Gavrila, D. M. (2016). A new benchmark for vision-based cyclist detection. In *2016 IEEE Intelligent Vehicles Symposium (IV)*, pages 1028–1033. IEEE. DOI: 10.1109/IVS.2016.7535515.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer. DOI: 10.1007/978-3-319-10602-1_{4}8.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European Conference on Computer Vision*, pages 21–37. Springer. DOI: 10.1007/978-3-319-46448-0_{2}.

Luo, Z., Branchaud-Charron, F., Lemaire, C., Konrad, J., Li, S., Mishra, A., Achkar, A., Eichel, J., and Jodoin, P.-M. (2018). Mio-tcd: A new benchmark dataset for vehicle classification and localization. *IEEE Transactions on Image Processing*, 27(10):5129–5141. DOI: 10.1109/TIP.2018.2848705.

MacAskill, D. (2018). Putting your best photo forward: Flickr updates. https://blog.flickr.net/.

Masalov, A., Matrenin, P., Ota, J., Wirth, F., Stiller, C., Corbet, H., and Lee, E. (2019). Specialized cyclist detection dataset: Challenging real-world computer vision dataset for cyclist detection using a monocular rgb camera. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 114–118. IEEE. DOI:

10.1109/IVS.2019.8813814.

Nardi, E., Padilha, B., Kamaura, L. T., and Ferreira, J. E. (2022). Openimages cyclists: Expandindo a generalização na detecção de ciclistas em câmeras de segurança. In *Anais do XXXVII Simpósio Brasileiro de Bancos de Dados*, pages 229–240. SBC. DOI: 10.5753/sbbd.2022.224626.

Pool, E. A. I., Kooij, J. F. P., and Gavrila, D. M. (2019). Context-based cyclist path prediction using recurrent neural networks. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 824–830. DOI: 10.1109/IVS.2019.8813889.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788. IEEE. DOI: 10.1109/CVPR.2016.91.

Redmon, J. and Farhadi, A. (2017). Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6517–6525. IEEE. DOI: 10.1109/CVPR.2017.690.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *ArXiv*, abs/1804.02767. DOI: 10.48550/arXiv.1804.02767.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, volume 28, pages 91–99. Curran Associates, Inc.

Robert, Ross, Marcin, Elvis, Guillem, Andrew, and Thomas (2022). Papers with code. https://paperswithcode.com/sota/object-detection-on-coco. Accessed on May 20, 2022.

Saleh, K., Hossny, M., Hossny, A., and Nahavandi, S. (2017). Cyclist detection in lidar scans using faster r-cnn and synthetic depth images. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. DOI: 10.1109/ITSC.2017.8317599.

Santhosh, K. K., Dogra, D. P., and Roy, P. P. (2020). Anomaly detection in road traffic using visual surveillance: A survey. *ACM Comput. Surv.*, 53(6). DOI: 10.1145/3417989.

Tan, M., Pang, R., and Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10778–10787. IEEE. DOI: 10.1109/CVPR42600.2020.01079.

Vasconcelos, C. N., Paes, A., and Montenegro, A. (2016a). Towards deep learning invariant pedestrian detection by data enrichment. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 837–841. DOI: 10.1109/ICMLA.2016.0150.

Vasconcelos, C. N., Vargag, A. C. G., Paes, A., and Montenegro, A. (2016b). Pedestrian detection using convolutional neural networks. In *Proceedings of XII Workshop de Visão Computacional, 2016*, pages 289–294.

Vial, A., Hendeby, G., Daamen, W., van Arem, B., and Hoogendoorn, S. (2023). Framework for network-constrained tracking of cyclists and pedestrians. *IEEE Transactions on Intelligent Transportation Systems*, 24(3):3282–3296. DOI: 10.1109/TITS.2022.3225467.

Wang, T., He, X., Su, S., and Guan, Y. (2017). Efficient scene layout aware object detection for traffic surveillance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 926–933. IEEE. DOI: 10.1109/CVPRW.2017.128.

Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., and Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126:103514. DOI: 10.1016/j.dsp.2022.103514.

Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Commun. ACM*, 64(3):107–115. DOI: 10.1145/3446776.

Zhou, X., Gong, W., Fu, W., and Du, F. (2017). Application of deep learning in object detection. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pages 631–634. IEEE. DOI: 10.1109/ICIS.2017.7960069.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76. DOI: 10.1109/JPROC.2020.3004555.

Zou, Z., Chen, K., Shi, Z., Guo, Y., and Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276. DOI: 10.1109/JPROC.2023.3238524.