Object Detection in Curved Space for 360-Degree Camera

Kuan-Hsun Wang, Shang-Hong Lai
Department of Computer Science, National Tsing Hua University, Taiwan

Abstract

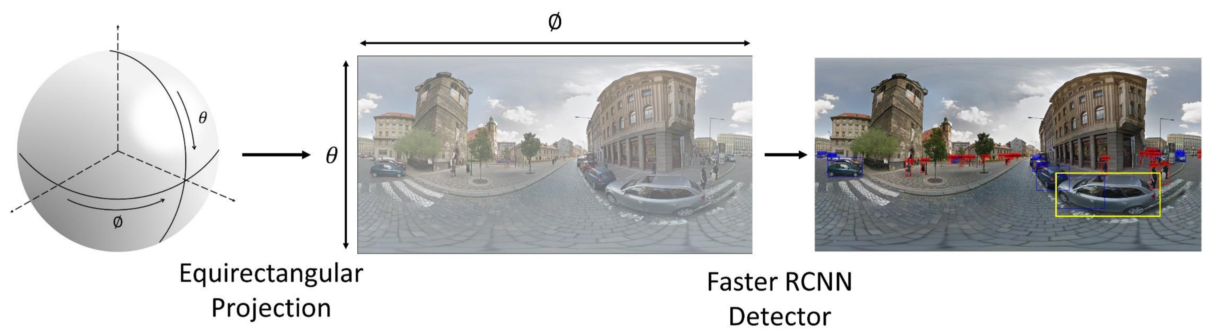
360° camera has recently become popular since it can capture the whole 360° scene. A large number of related applications have been springing up. In this paper, We propose a deep learning based object detector that can be applied directly on 360° images. The proposed detector is based on modifications of the faster RCNN model. Three modification schemes are proposed here, including (1) distortion data augmentation, (2) introducing muilti-kernel layers for improving accuracy for distorted object detection, and (3) adding position information into the model for learning spatial information. Additionally, we create two datasets, 360GoogleStreetView and 360Videos, and perform experiments on these two datasets to demonstrate that our object detector provides superior accuracy for object detection directly on 360° images.

Introduction

In this paper, we aim to develop an object detector that can directly apply on 360° images without any image preprocessing and re-projection. However, to accomplish the method, the image distortion problem should be properly handled. As we can see from the example in the following figure, some object detection benchmark models are unable to detect these distorted objects. To overcome the problem of distortion, we propose the following three methods:

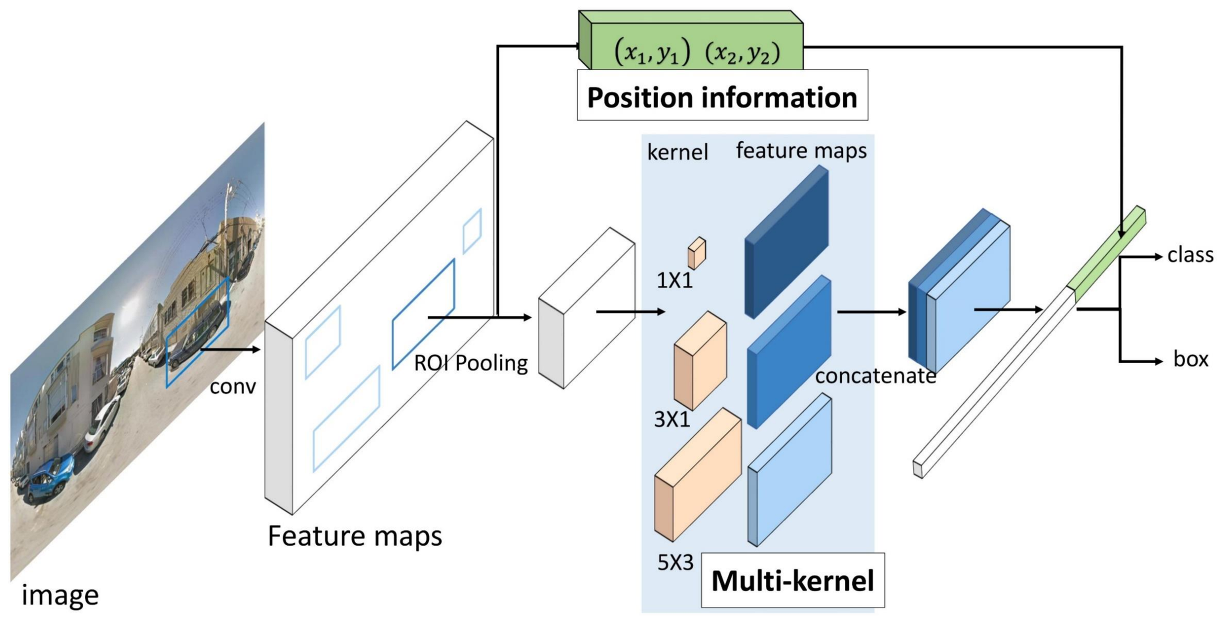
(1) distorted data augmentation — generating a variety of additional distorted data for training
(2) multi-kernel layer — applying different sizes of kernels on different regions to eliminate the distortion
(3) position information — the object position in the 360° image

In addition, we manually collect and label two datasets by ourselves to increase the number of training samples. The two datasets presented in this work are **360GoogleStreetView** and **360Videos** datasets.



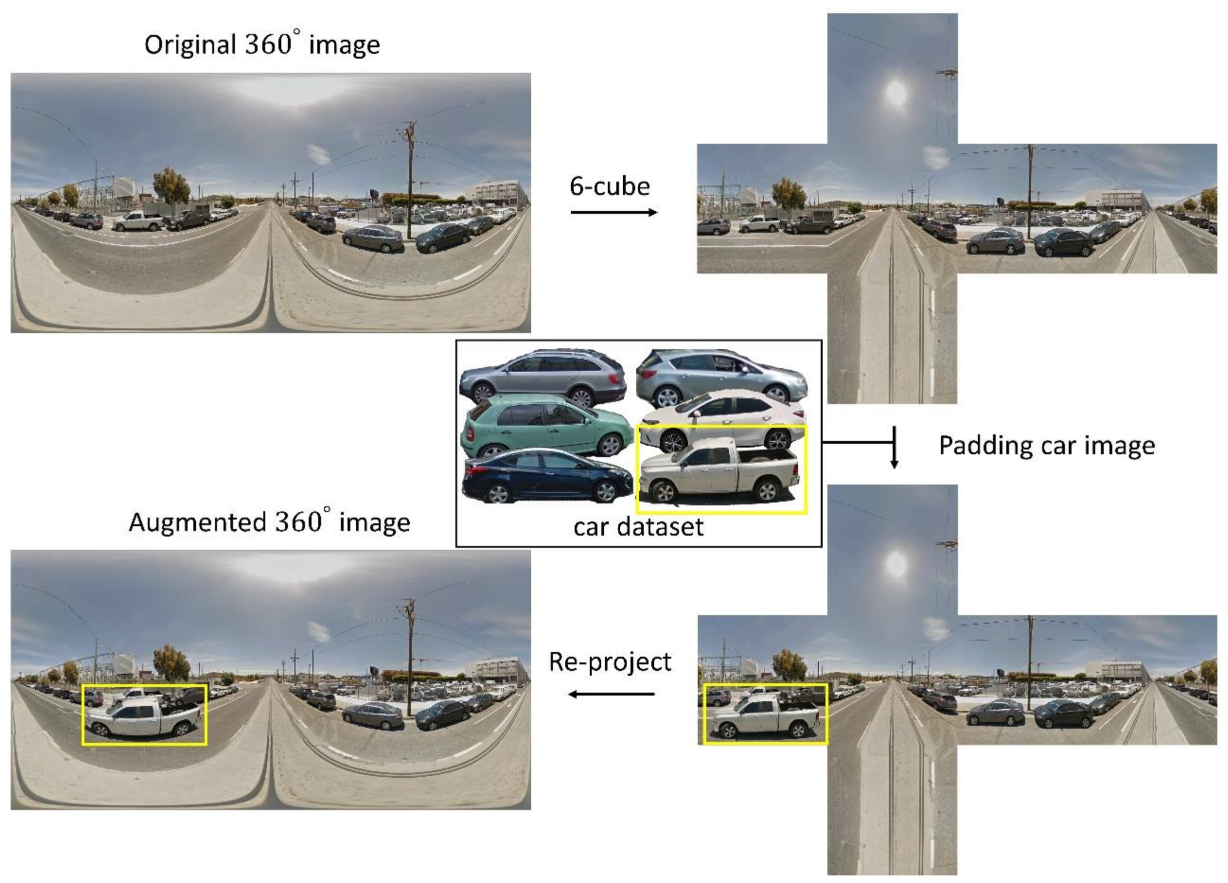
Methodology

We choose Faster R-CNN as the baseline method. The following figure is the overview of our network architecture. The input image is a distorted 360° image. We apply a multi-kernel layer after cropping the feature maps through the ROI Pooling layer. At the last fully connected layer, we even add the position information of the object region into the network.



Distorted Data Augmentation

we propose an augmentation method to increase the number of distorted training data.



Multi-kernel Layer

Since the objects located at the high latitude are stretched, the bigger kernel size should be applied to reduce the distortion.

Position Information

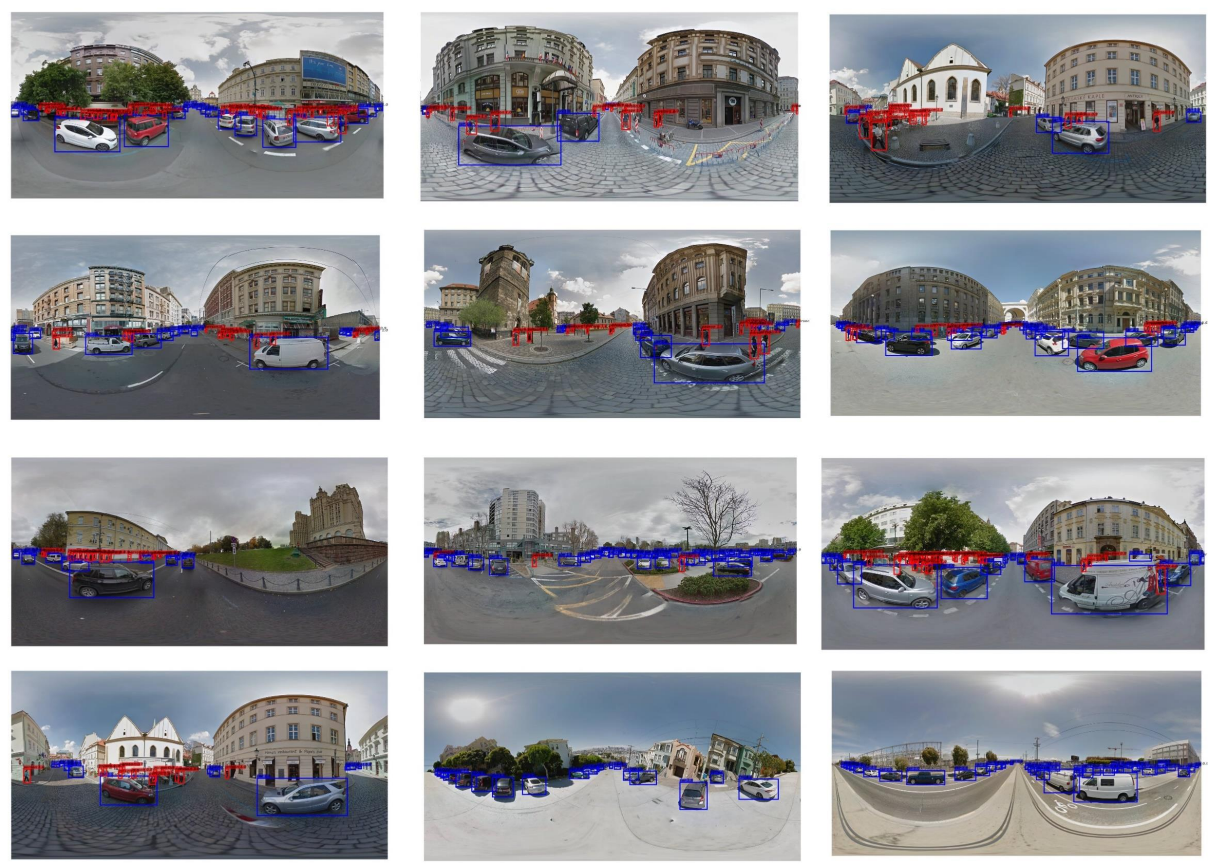
The position information is composed of 6 components, including bounding box coordinates aThe definition of position information is as follows: nd the width and height of the bounding box.

$$P^i = \left(\frac{x_1^i}{W}, \frac{y_1^i}{H}, \frac{x_2^i}{W}, \frac{y_2^i}{H}, \frac{w^i}{W}, \frac{h^i}{H}\right)$$

Experimental Results

We perform experiments on the 360GoogleStreetView dataset and compare our detector with the state-of-the-art object detectors. The result is reported in the following table. All the methods, including Faster RCNN and Deformable CNN, are trained on both 360GoogleStreetView and 360Videos.

Method	mAP@0.5/0.7	D-car/D-person	A_D@0.5/0.7
Faster RCNN	72.88 / 56.24	88.06 / 79.94	87.01 / 75.87
D-CNN	69.15 / 52.28	69.97 / 79.12	89.69 / 80.00
Ours (Position)	73.46 / 56.38	88.95 / 80.53	88.87 / 76.70
Ours (MK)	76.78 / 56.67	89.45 / 80.41	89.48 / 80.00
Ours (Aug)	76.13 / 56.01	90.37 / 79.96	89.89 / 77.53
Ours (All)	76.38 / 56.76	90.60 / 80.01	90.30 / 81.03



Method	mAP@0.5/0.7	D-car	D-person
Faster RCNN	86.05 / 76.25	85.53 / 77.30	86.56 / 75.20
D-CNN	87.76 / 75.30	89.85 / 76.45	85.67 / 74.15
Ours	91.14 / 76.80	95.23 / 78.75	87.04 / 74.84

Conclusion

We propose a modified Faster RCNN model for object detection directly on 360° images. The distorted data augmentation method generates additional distorted objects for training. We propose to include a multi-kernel layer that incorporates different kernel sizes to alleviate distortion effect. In addition, we include object position information into the network to obtain better prediction. Furthermore, we created two datasets for performance evaluation of object detection on 360° images, and demonstrated the proposed object detector provided superior object detection accuracy compared to the state-of-the-art object detectors.

Acknowledgements

The authors would like to thank Qualcomm Technologies Inc. for supporting this research work