MagnifierNet: Learning Efficient Small-scale Pedestrian Detector towards Multiple Dense Regions

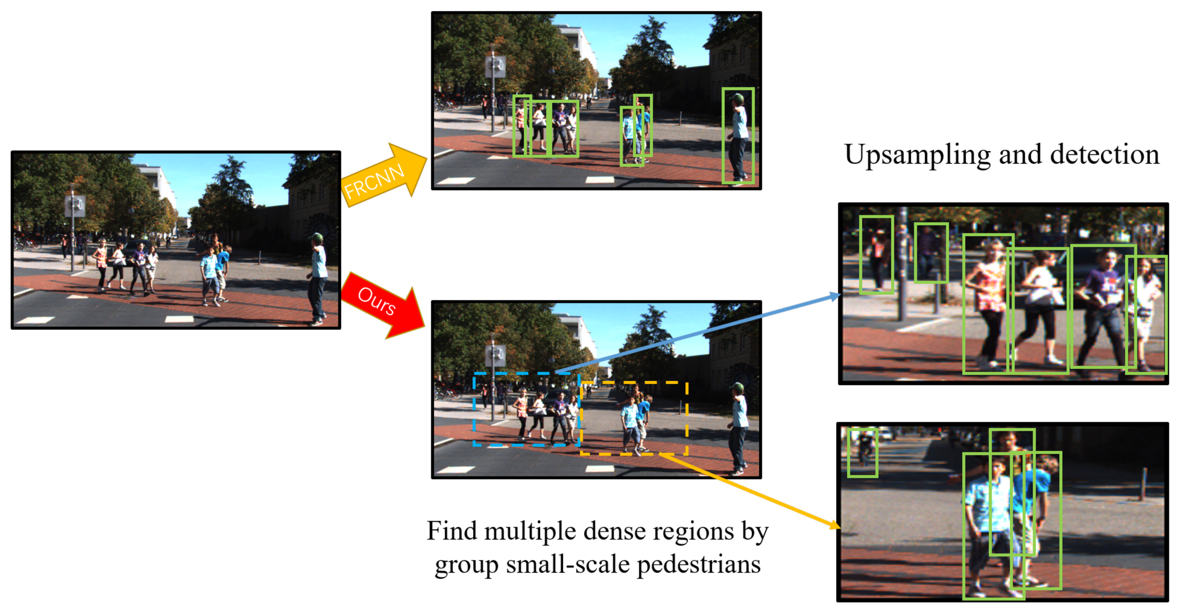
Qi Cheng   Mingqin Chen   Yingjie Wu*   Fei Chen   Shiping Lin
College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China

1. Introduction

Motivation

- Although recent deep learning object detectors such as Faster R-CNN have shown excellent performance for general object detection, they have limited success for detecting pedestrians.
- There is still a significant gap in the performance of the detection of pedestrians at different scales.
- We observed pedestrian datasets and found that pedestrians often gather together in crowded public places.

So it is considered that the detection of small-scale pedestrians is the bottleneck for improving the performance of detectors.



In this paper, we extend the existing object detection ideas and introduce a simple but effective model that is suitable for detecting small-scale pedestrians we call MagnifierNet. The main idea was inspired by the function of a magnifier.

Therefore, the MagnifierNet module implements pedestrian detection by focusing on small-scale pedestrian dense regions. During inference, these regions are up-sampled for detection as shown in the figure above.

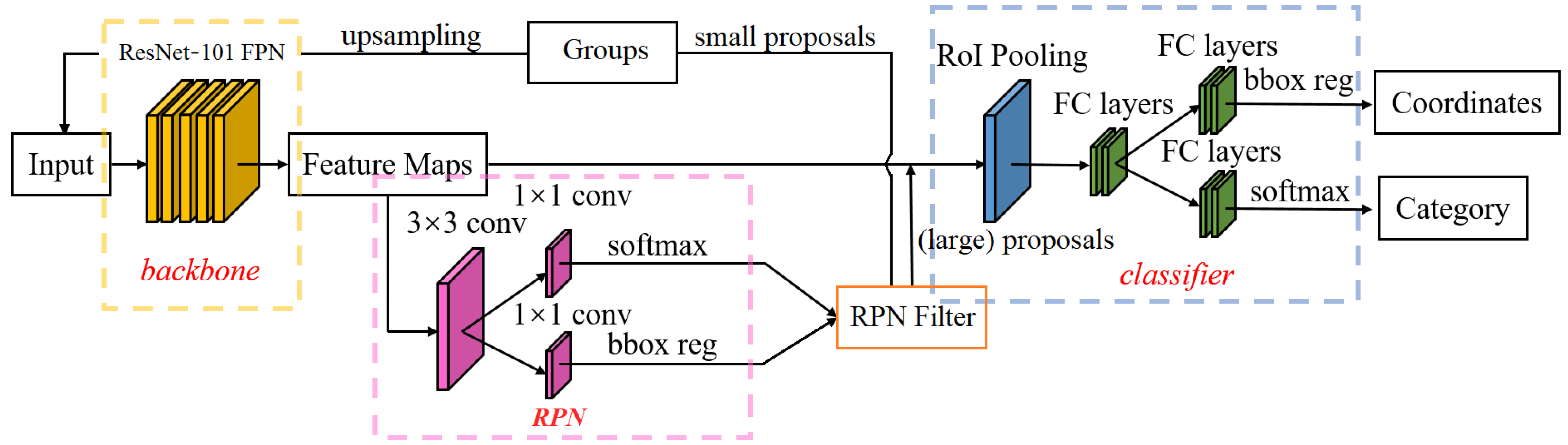
Challenges

- Finding multiple dense regions of small-scale pedestrians in the image.

Main Contribution

- We propose a pedestrian detector which has better performance and is particularly effective for small-scale pedestrians. Since this method was inspired by the magnifier, we named it MagnifierNet.
- To find the multiple dense regions of small-scale pedestrians in the image, we propose a small object grouping algorithm based on sweep-line.
- We introduce a new data augmentation strategy suitable for pedestrian datasets based on the grouping algorithm.
- MagnifierNet achieves the best detection performance on the CityPersons benchmark without using any external data and improves the performance of smallscale pedestrians significantly. In addition, we also achieve competitive performance on the KITTI dataset.

2. Model architecture



The difference from the Faster R-CNN model is that the output of RPN has an additional branch to find small-scale pedestrian dense regions, which works in the inference stage

3. MagnifierNet

Area threshold

We believe that the area distribution of the small-scale pedestrian bounding box should be similar to the area distribution of the large-scale pedestrian bounding box. The size of the group box can be determined by:

$$s_t = \frac{\overline{L_t}}{\overline{S_t}} \tag{1}$$

$$gw_t \cdot gh_t \cdot s_t = w \cdot h \tag{2}$$

$$gw_t = \frac{w}{\sqrt{s_t}} \qquad gh_t = \frac{h}{\sqrt{s_t}} \tag{3}$$

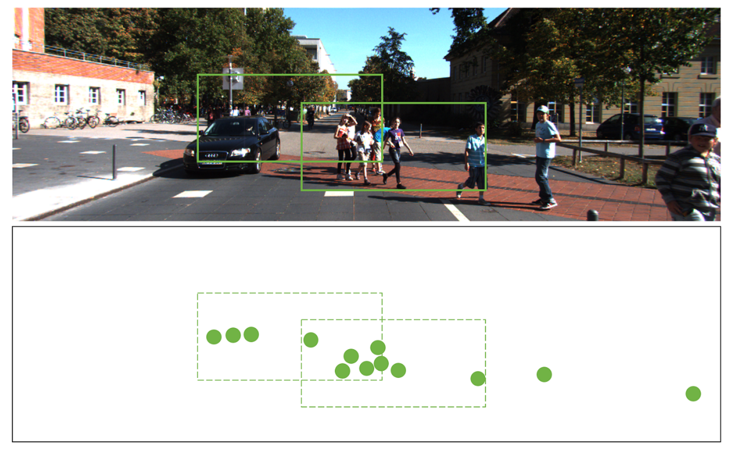
The calculation formula of $\hat{t}$ is as follows:

$$\hat{t} = \arg\min_{t} KL\left(P_t(\xi) \| Q_t(\xi)\right) \tag{4}$$

Grouping algorithm

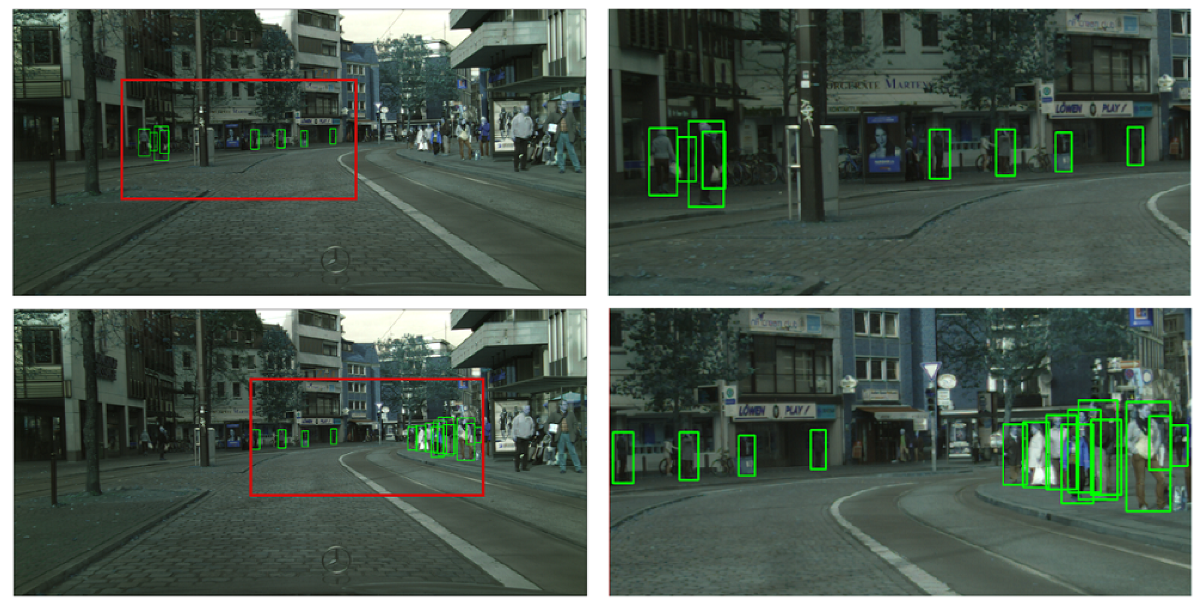
We propose the sweep-line based grouping algorithm to find multiple dense regions. The transfer equation for dynamic programming is:

$$dp\left(x_i\right) = \min_{x_i - x_j < gw} \{dp\left(x_j\right) + \text{count}\left(x_j, x_i\right)\} \tag{5}$$



After finding small-scale pedestrian dense regions, we cropped these regions from the image, and upsampled them.

Data augmentation



After finding small-scale pedestrian dense regions, we cropped these regions from the image, and upsampled them.

4. Experiments

To demonstrate the effectiveness of the proposed method, we evaluate two of the largest pedestrian detection benchmark, i.e., KITTI and CityPersons.

CityPersons

Methods	R	H	P	B	S	M	L
FRCNN	15.4	-	-	-	25.6	7.2	7.9
FRCNN+Seg	14.8	-	-	-	22.6	6.7	8.0
OR-CNN	12.8	55.7	15.3	**6.7**	-	-	-
RepLoss	13.2	56.9	16.8	7.6	-	-	-
TLL+MRF	14.4	52.0	15.9	9.2	-	-	-
ALFNet	12.0	51.9	11.4	8.4	19.0	5.7	6.6
CSP	11.0	49.3	10.4	7.3	16.0	**3.7**	**6.5**
Baseline	18.7	52.5	19.1	13.1	32.6	12.1	8.4
+DA	15.3	48.5	12.6	12.0	18.2	7.2	9.3
+GA	15.6	49.7	15.5	11.0	22.9	9.4	8.0
+DA & GA	**10.8**	**42.2**	**10.1**	7.4	**12.6**	5.5	7.7

KITTI

Method	E	M	H
FRCNN	78.35	65.91	61.19
MS-CNN	83.70	73.62	68.28
RPN+BF	75.58	61.29	56.08
F-ConvNet	83.63	72.91	67.18
VMVS	81.11	70.89	67.23
MonoPSR	85.60	68.56	63.34
SubCNN	83.17	71.34	66.36
Aston-EAS	85.12	74.52	69.35
MHN	85.80	74.60	68.94
CLA	–	73.96	–
Baseline	84.15	68.78	63.60
+DA	85.51	73.75	68.54
+GA	86.21	72.67	67.55
+DA & GA	**86.95**	**74.95**	**69.50**

To our best knowledge, our MagnifierNet detector achieves the best small-scale pedestrian detection performance on CityPersons benchmark without any external data, and also achieve competitive performance on the KITTI dataset.

5. Conclusion

In this work, it is considered that the detection of small-scale pedestrians is the bottleneck for improving the performance of pedestrian detectors.

To address this, we propose MagnifierNet, a simple but effective detector which is focused on smallscale pedestrian dense regions. We introduce a sweep-line based grouping algorithm to find multiple dense regions. With the help of the effective data augmentation strategy, MagnifierNet brings significant improvements in detecting small-scale pedestrians.

The experimental results show that it outperforms other detection methods on KITTI and CityPersons datasets.

6. Acknowledgment

This work is supported by the National Natural Science Foundation of China (61771141) and the Natural Science Foundation of Fujian Province, China (2018J01799).