PRF-Ped: Multi-scale Pedestrian Detector with Prior-based Receptive Field



Yuzhi Tan^{1,2} Hongxun Yao¹ Haoran Li¹ Xiusheng Lu³ Haozhe Xie¹

¹Harbin Institute of Technology ²Advanced Institute of Information Technology, Peking University ³HiSilicon

Introduction

Proposed Method

Pedestrian detection is required in many real-world scenarios, including autonomous driving, and intelligent robots, and intelligent monitoring.

Existing issue

Vast scale variation.

Characteristic of pedestrian target

The pedestrian usually has a specific aspect ratio.

How to solve it?

Focusing on multi-scale pedestrian feature extraction and aggregation.

Our advantage

The proposed detector can obtain better pedestrian feature representation with a wider range of scales for prediction, while less affected by the surrounding complex background.

Contributions

- We present a bidirectional feature enhancement module (BFEM), which enhances the semantic information of low-level features and enriches the localization information of high-level features.
- We propose a prior-based receptive field block (PRFB), which takes the shape prior of the pedestrian into account, guiding the network to focus on the pedestrian.
- Experimental results demonstrate that the proposed method outperforms the state-of-theart methods on the CityPersons and Caltech datasets.



Prior-based Receptive Field Block (PRFB)





In the upper picture, a square receptive field produced by the general convolutional neural network covers more background, which could affect the pedestrian feature extraction under a complex environment. We design a prior-based receptive field block (PRFB), whose receptive field is close to the pedestrian area, to extract more representative features before aggregation.

Experimental Results

The pedestrian detection results evaluated with MR^{-2} on the CityPersons dataset

Methods	Backbone	Reasonable	Bare	Partial	Heavy	Large	Medium	Small	Test Time
FRCNN [1]	VGG-16	15.4	-	-	-	7.9	7.2	25.6	-
FRCNN+Seg [1]	VGG-16	14.8	-	-	-	8.0	6.7	22.6	-
OR-CNN [2]	VGG-16	12.8	6.7	15.3	55.7	-	-	-	-
RepLoss [3]	Resnet-50	13.2	7.6	16.8	56.9	-	-	-	-
TLL [4]	Resnet-50	15.5	10.0	17.2	53.6	-	-	-	-
TLL+MRF [4]	Resnet-50	14.4	9.2	15.9	52.0	-	-	-	-
ALFNet [5]	Resnet-50	12.0	8.4	11.4	51.9	6.6	5.7	19.0	0.27s / image
MGAN [6]	VGG-16	11.5	-	-	51.7	-	-	-	-
CSP [7]	Resnet-50	11.0	7.3	10.4	49.3	6.5	3.7	16.0	0.33s / image
PRF-Ped (Ours)	Resnet-50	9.7	6.5	8.8	47.3	5.8	3.9	12.9	0.16s / image

The pedestrian detection results evaluated with MR^{-2} on the Caltech dataset



The ablation study of PRF-Ped on the CityPersons dataset

Method	Concatenate	ASFF	BFEM-concat	BFEM-add	BFEM	PRFB	$MR^{-2}(\%)$	#Parameters	Test Time
CSP	\checkmark						11.0	40.0MB	0.33s / image
		\checkmark					10.7	28.1MB	0.11s / image
		\checkmark	\checkmark				10.7	35.1MB	0.14s / image
Ours		\checkmark		\checkmark			10.4	31.6MB	0.13s / image
		\checkmark			\checkmark		10.1	33.4MB	0.14s / image
		\checkmark			\checkmark	\checkmark	9.7	34.6MB	0.16s / image

Visualization results



References

[1] S. Zhang, R. Benenson, and B. Schiele, "Citypersons: A diverse dataset for pedestrian detection," in CVPR, 2017.

[2] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Occlusion-aware r-cnn: detecting pedestrians in a crowd," in ECCV, 2018.

[3] X. Wang, T. Xiao, Y. Jiang, S. Shao, J. Sun, and C. Shen, "Repulsion loss: Detecting pedestrians in a crowd," in CVPR, 2018.

[4] T. Song, L. Sun, D. Xie, H. Sun, and S. Pu, "Small-scale pedestrian detection based on somatic topology localization and temporal feature aggregation," in ECCV, 2018.

[5] W. Liu, S. Liao, W. Hu, X. Liang, and X. Chen, "Learning efficient single-stage pedestrian detectors by asymptotic localization fitting," in ECCV, 2018.

[6] Y. Pang, J. Xie, M. H. Khan, R. M. Anwer, F. S. Khan, and L. Shao, "Mask-guided attention network for occluded pedestrian detection," in ICCV, 2019.

[7] W. Liu, S. Liao, W. Ren, W. Hu, and Y. Yu, "High-level semantic feature detection: A new perspective for pedestrian detection," in CVPR, 2019.