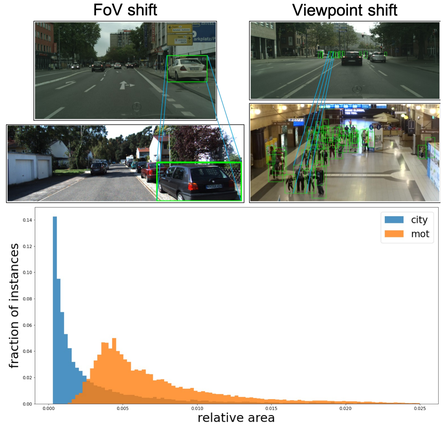
Learning Transformations To Reduce the Geometric Shift in Object Detection

Vidit, Martin Engilberge, Mathieu Salzmann
CVLab, EPFL

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction

- Domain gap caused due to geometric shifts
- Underlying transformations of these shifts can be unknown
- **Cause**: Field of View (FoV) or Viewpoint changes
- **Effect**: Apparent change in bounding boxes sizes



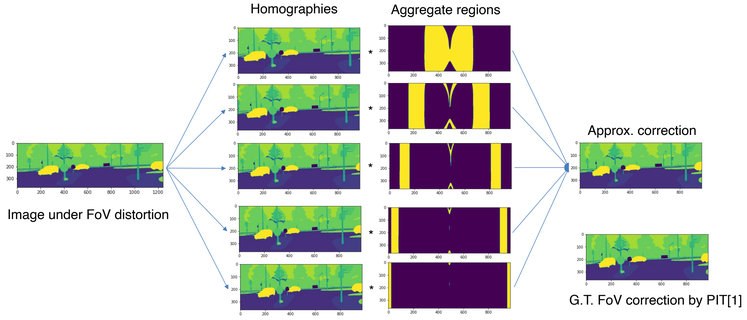
Change in Bbox distribution between two domains

Contribution

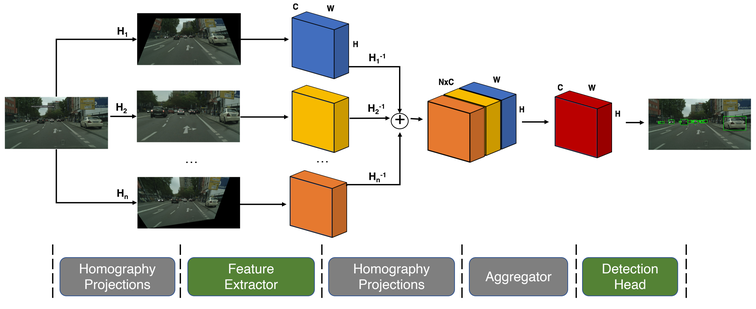
- Learning transformations to bridge the domain gap
- Our method is general as it does not require the knowledge of ground-truth transformations
- Our self-training based method achieve state-of-the-art results

Method

- A set of homographies can approximate any geometric shift

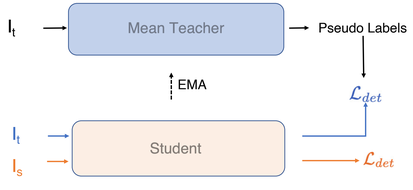


- It is hard to know exact homographies and regions to aggregate for all shifts



Our architecture

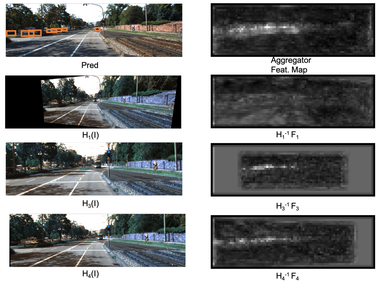
- Our proposed method learns both of these components with MeanTeacher training.



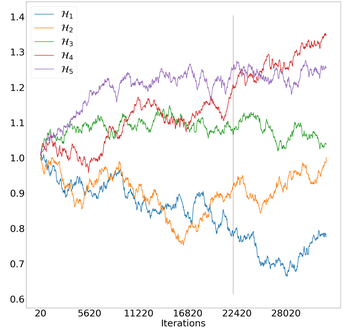
MeanTeacher training

Results

Aggregator learns to combine different car regions



Homographies become diverse during training



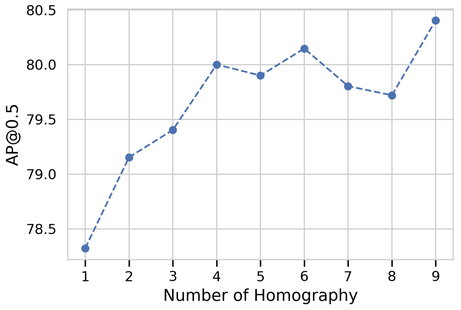
FoV Adaptation

Method	Car AP@0.5
FR [3]	76.1
AT [2]	77.2
FR+PIT	77.6
MT	78.3
MT+PIT [1]	79.7
Ours	**80.4** $\pm$ 0.15

Viewpoint Adaptation

Method	Pedestrian AP@0.5
FR [3]	43.7
AT [2]	63.5
MT	64.7
Ours	**65.3** $\pm$ 0.37

Performance improves with more homographies



Conclusion

- Geometric shifts are fairly common but exact transformations can be unknown
- Few homographies are sufficient to close the gap between domains
- Our method generalizes for shifts like Viewpoint for which ground-truth transformations is unknown

References:

[1] Qiqi Gu, et. al. Pit: Position-invariant transform for cross-fov domain adaptation, ICCV'21

[2] Yu-Jhe Li, et. al. Cross-domain adaptive teacher for object detection, CVPR'22

[3] Shaoqing Ren, et. al. Faster r-cnn: towards real-time object detection with region proposal networks. TPAMI'16