Object Detection with Self-Supervised Scene Adaptation

Zekun Zhang[1], Minh Hoai[1,2]
[1]Stony Brook University, [2]VinAI Artificial Intelligence Application and Research

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

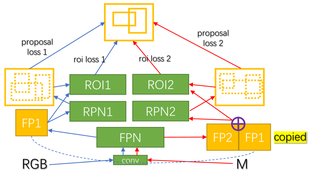
Motivation & Challenges

- train once, adapt to any scene
- require no manual annotation
- mitigate domain shift issue

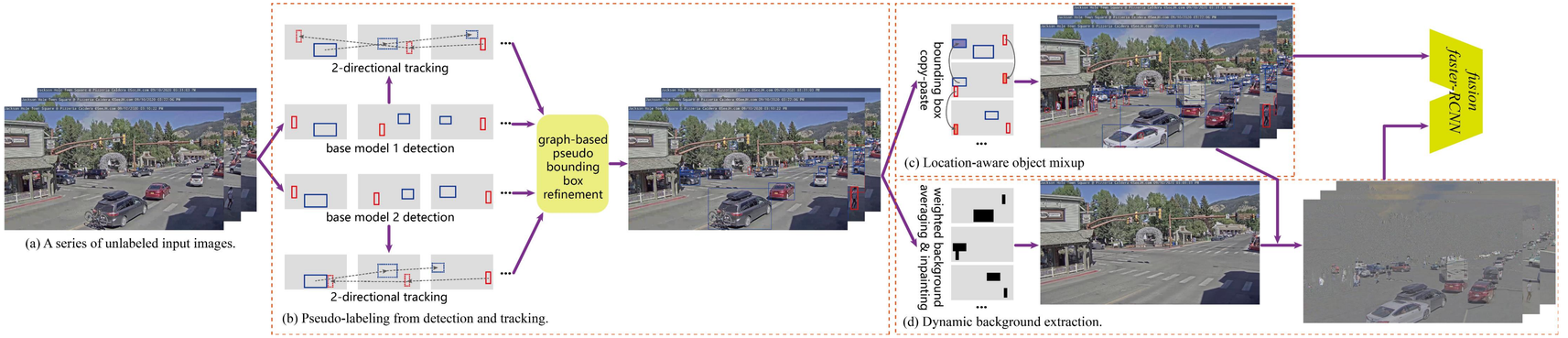
Scenes100 Dataset

- first video object detection dataset focused on scene adaptation
- large-scale and diverse

Framework Architecture & Data Flow



mid-fusion faster-RCNN



(a) A series of unlabeled input images.

(b) Pseudo-labeling from detection and tracking.

(c) Location-aware object mixup

(d) Dynamic background extraction.

Quantitative Results & Ablation Study

Method	mAP-gain
ST (RoyChowdhury *et al.*, CVPR 2019)	+0.80
STAC (Sohn *et al.*, arXiv 2020)	-1.26
AT (Li *et al.*, CVPR 2022)	-0.75
H^2FA (Xu *et al.*, CVPR 2022)	-3.10
TIA (Zhao & Wang, CVPR 2022)	-0.32
LODS (Li *et al.*, CVPR 2022)	+0.45
Proposed (Zhang & Hoai, CVPR 2023)	**+3.76**

Mixup	Fusion	mAP-gain
✗	✗	+0.95
Location-aware	✗	**+1.72**
Random	✗	+1.22
✗	early	+1.85
✗	mid	**+3.40**
✗	late	+3.34
Location-aware	early	+2.25
Location-aware	mid	**+3.76**
Location-aware	late	+3.66

Code & Dataset

Poster # THU-PM-094