ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

Coarse-to-Fine Multi-Class Multi-Object Tracking (COFE)

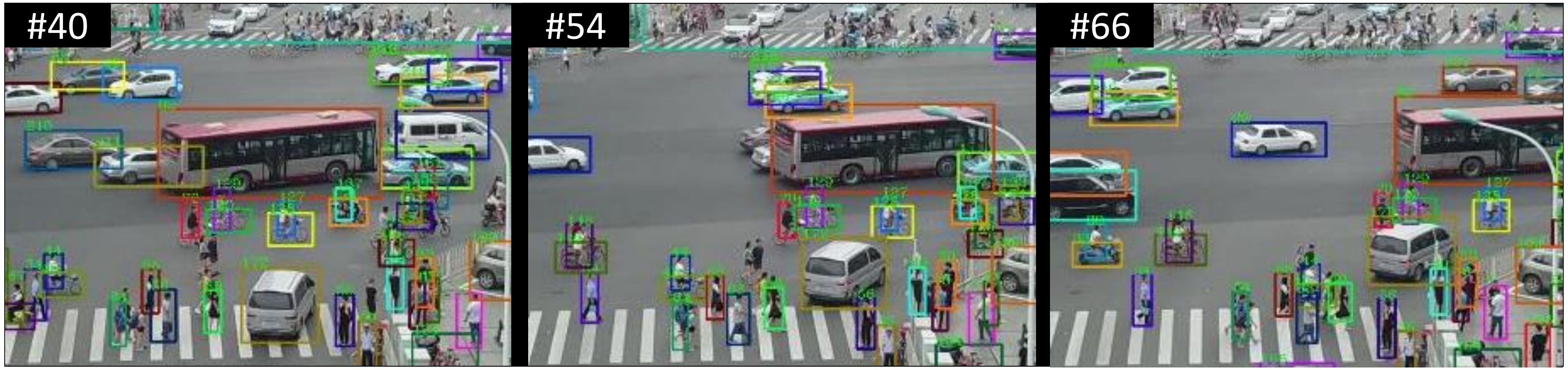
Task: Multi-Object Tracking

Yuhang He, Wentao Yu, Jie Han, Xiaopeng Hong, Xing Wei, and Yihong Gong
Xi'an Jiaotong University (XTJU), P.R.C.

1. Introduction

Goal: Track multi-class multi-object in different scenarios.

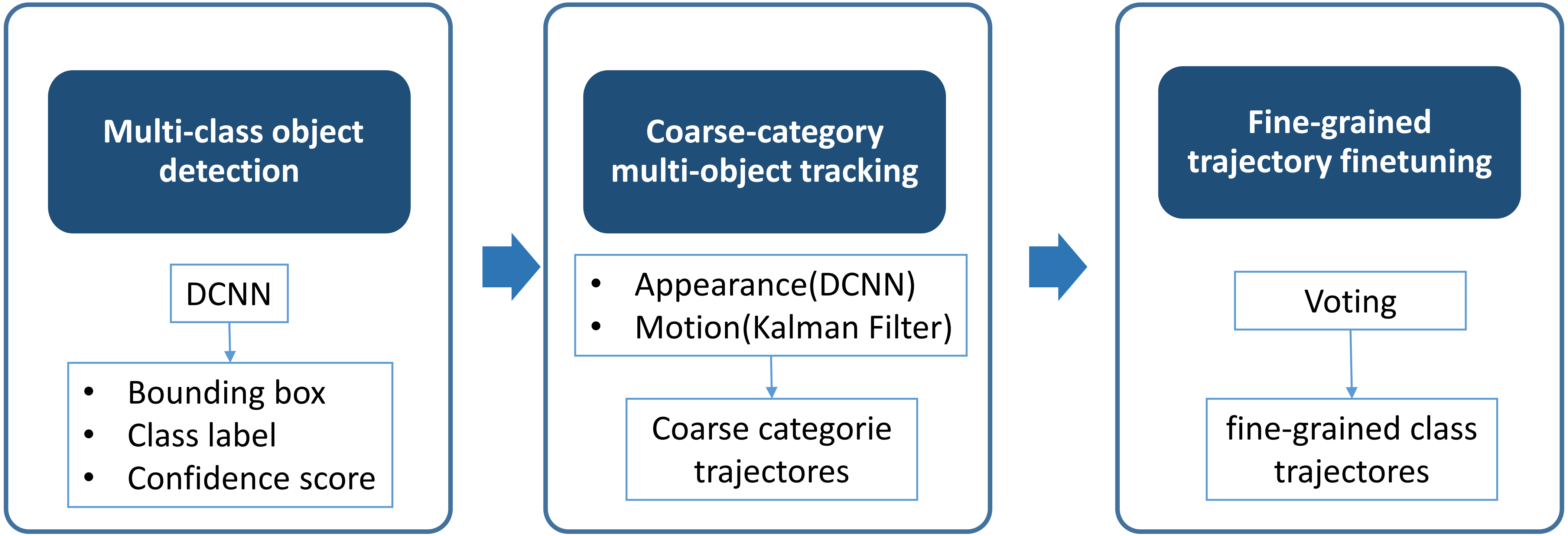
Key Insight: Interactive and mutual occluded targets are semantically discriminative in coarse categories.



ECCV'20 ONLINE 23-28 AUGUST 2020

2. Methodology

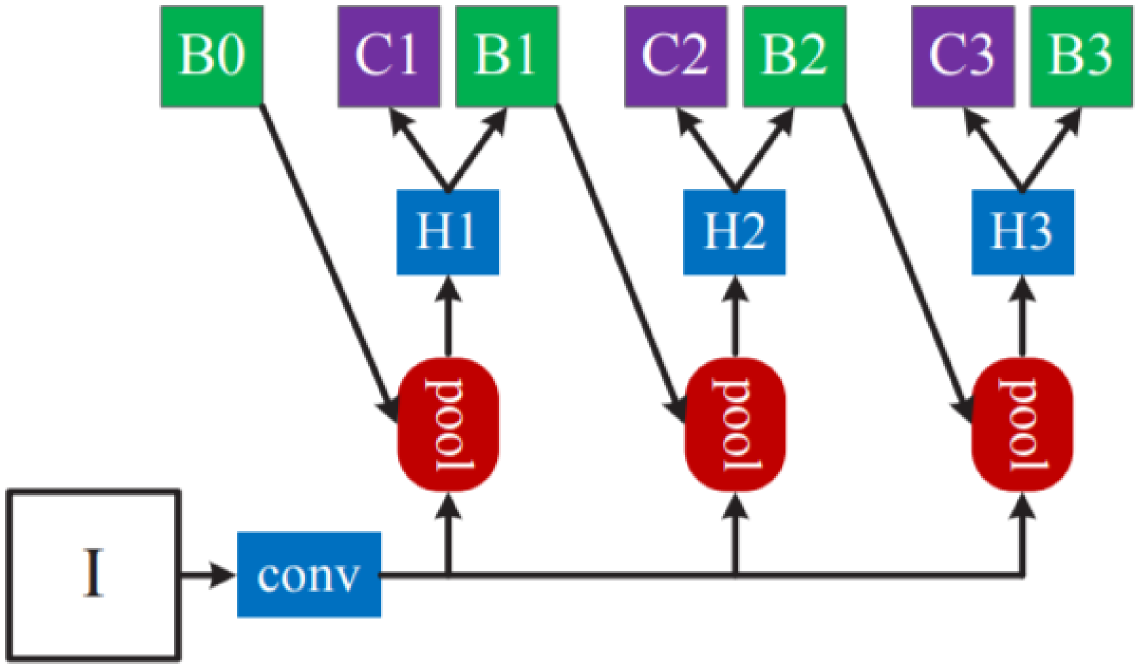
2.1 Algorithm Overview



2. Methodology

2.2 Multi-class object detection

- Baseline: Cascade R-CNN
- Backbone: resnet50
- FPN, DCN



2. Methodology

2.3 Coarse-category multi-object tracking

1. Summarize fine-grained classes into coarse category

- (Van, Bus, Car and Truck)-> Vehicle
- Pedestrian

2. Link detections into Coarse categories trajectories by appearance and motion information

- Extract appearance features by OSNet (msmt17)
- Predict motion information by Kalman Filter
- Solve data association based on Hungarian algorithm

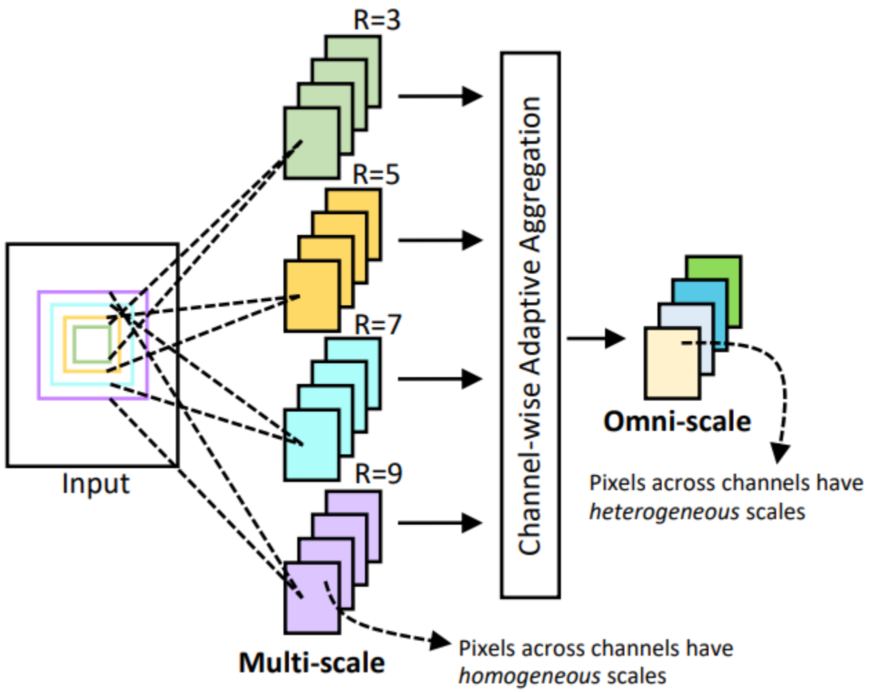


Figure 2: A schematic of the proposed building block for OSNet. R: Receptive field size.

2. Methodology

2.3 Coarse-category multi-object tracking

2. Link detections into Coarse categories trajectories by appearance and motion information

Listing 1 Matching Cascade

Input: Track indices $\mathcal{T} = \{1, \ldots, N\}$, Detection indices $\mathcal{D} = \{1, \ldots, M\}$, Maximum age $A_{\max}$

Compute cost matrix $\boldsymbol{C} = [c_{i,j}]$ using Eq. 5
Compute gate matrix $\boldsymbol{B} = [b_{i,j}]$ using Eq. 6
Initialize set of matches $\mathcal{M} \leftarrow \emptyset$
Initialize set of unmatched detections $\mathcal{U} \leftarrow \mathcal{D}$
for $n \in \{1, \ldots, A_{\max}\}$ **do**
    Select tracks by age $\mathcal{T}_n \leftarrow \{i \in \mathcal{T} \mid a_i = n\}$
    $[x_{i,j}] \leftarrow \text{min\_cost\_matching}(\boldsymbol{C}, \mathcal{T}_n, \mathcal{U})$
    $\mathcal{M} \leftarrow \mathcal{M} \cup \{(i,j) \mid b_{i,j} \cdot x_{i,j} > 0\}$
    $\mathcal{U} \leftarrow \mathcal{U} \setminus \{j \mid \sum_i b_{i,j} \cdot x_{i,j} > 0\}$
end for
return $\mathcal{M}, \mathcal{U}$

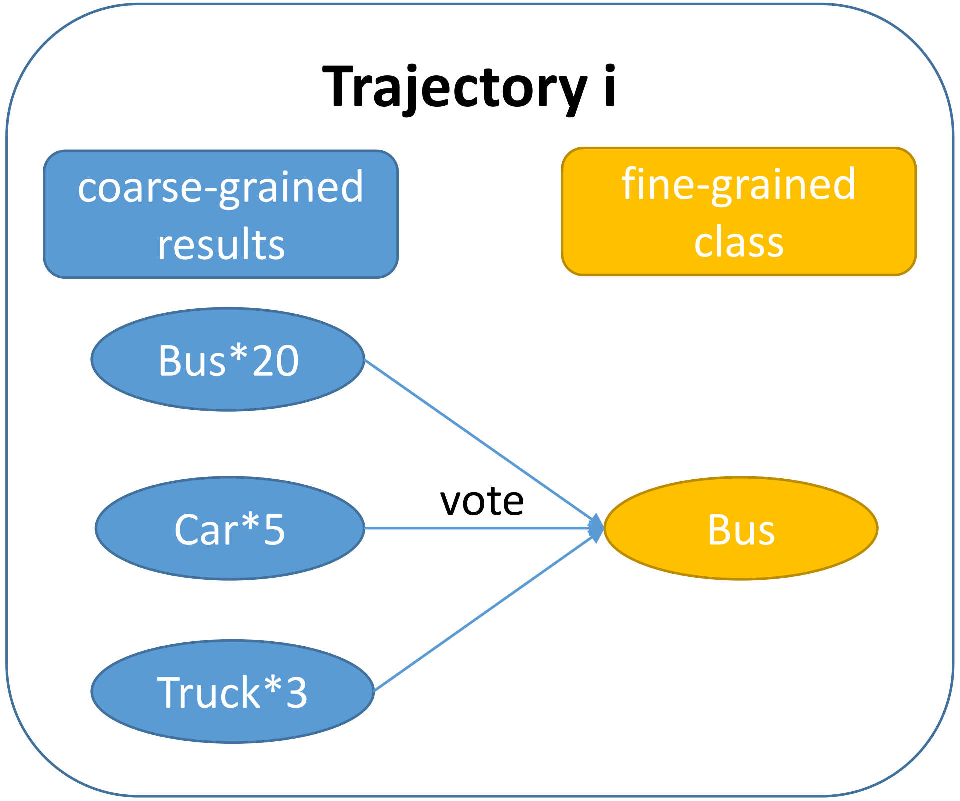
2. Methodology

2.4 Fine-grained trajectory finetuning

1. finetune fine-grained class label by voting

2. refine the tracking results by post processing

- Trajectory padding, smoothing



3. Experimental Results

rank	team	mAP
1	**visDrone-MOT**	**61.88**
2	DeepBlueAI-MOT	57.65
3	Daniel Stadler	50.80
4	Julyyyyy	42.11
5	Zhizhao Duan	42.10
6	Li Peng	38.05
7	shengwen li	35.76

The proposed method COFE achieves the best result and significantly outperforms state-of-the-art methods by a large margin

Thanks !