Multi-camera Vehicle Tracking and Re-identification on AI City Challenge 2019

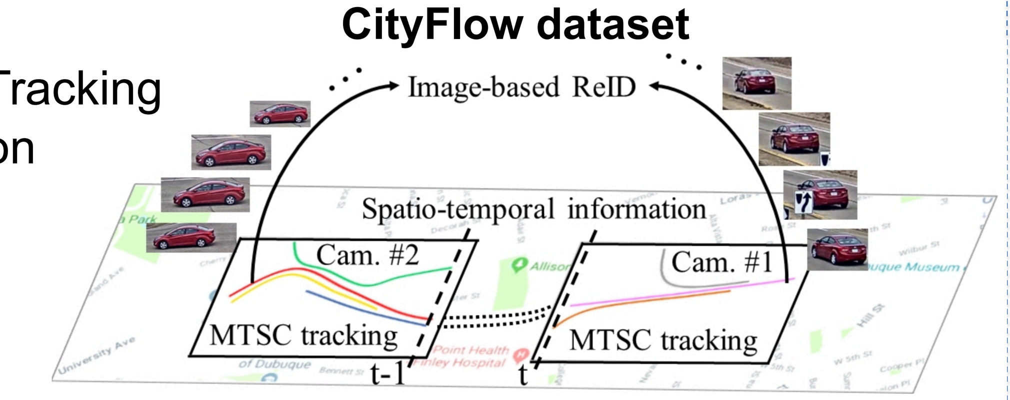
Yucheng Chen[1], Longlong Jing[2], Elahe Vahdani[2], Ling Zhang[2], Mingyi He[1], and Yingli Tian[2]
[1]Northwestern Polytechnical University [2]The City University of New York

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Introduction

1) Tracks:

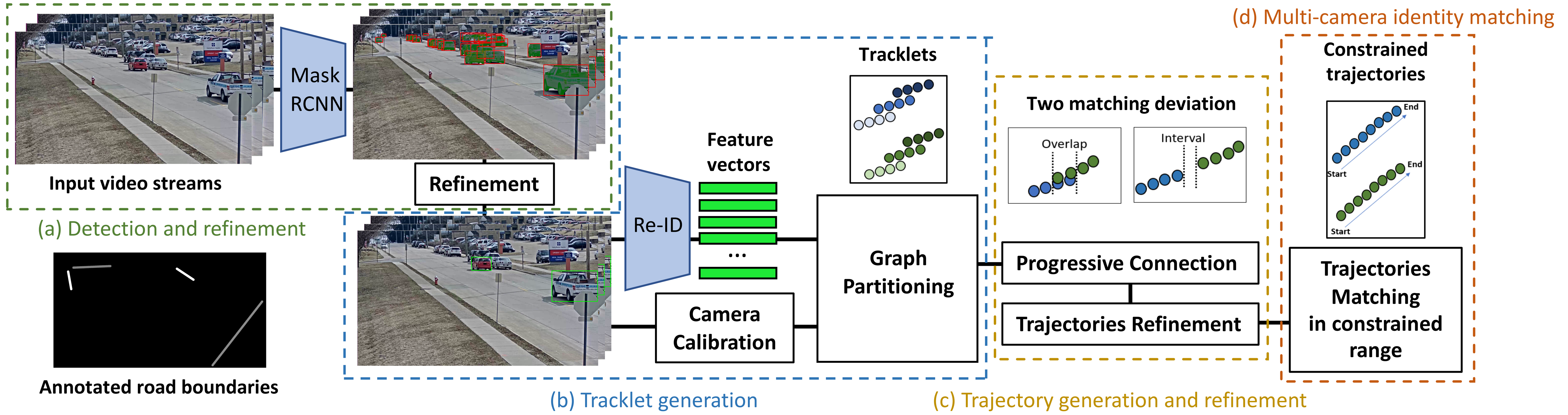
- Multi-camera Vehicle Tracking
- Vehicle Re-identification



2) Achievements:

- The proposed **enhanced multi-granularity network** with multiple branches outperforms the current state-of-the-art vehicle ReID method by **16.3%** on Veri776 dataset.
- We designed an offline pipeline for multi-camera vehicle tracking with our annotated road boundaries.
- Our algorithms are ranked the **10** and **23** in MVT and ReID tracks respectively at the NVIDIA AI City Challenge 2019.

Proposed pipeline for multi-camera vehicle tracking



Results

Method	IDF1	IDP	IDR
$baseline$	0.594	0.449	0.878
$baseline + G_g$	0.605	0.459	0.890
$baseline + G_g + G_m$	0.630	0.477	0.926
$baseline + G_g + G_m + Tj_r$	0.657	0.499	0.962
$baseline + G_g + G_m + Tj_r + Tj_d$	0.755	0.647	0.907

Ablation study on CityFlow training set. (only trained on Veri776)

MVT			ReID		
Rank	Team ID	IDF	Rank	Team ID	mAP
1	21	0.7059	1	59	0.8554
2	49	0.6865	2	21	0.7917
3	12	0.6653	3	97	0.7589
10	**52**	**0.2850**	**23**	**52**	**0.4096**
22	45	0.0326	84	133	0.0003

Results and ranks on MVT and ReID tasks of AIC2019.

(a) Detection and refinement

1) Basic constraints:

Non-Maximum Suppression (NMS);
Bounding box area/height/width/aspect ratio;
The ratio of effective area (mask) …

2) Foreground-background comparison :

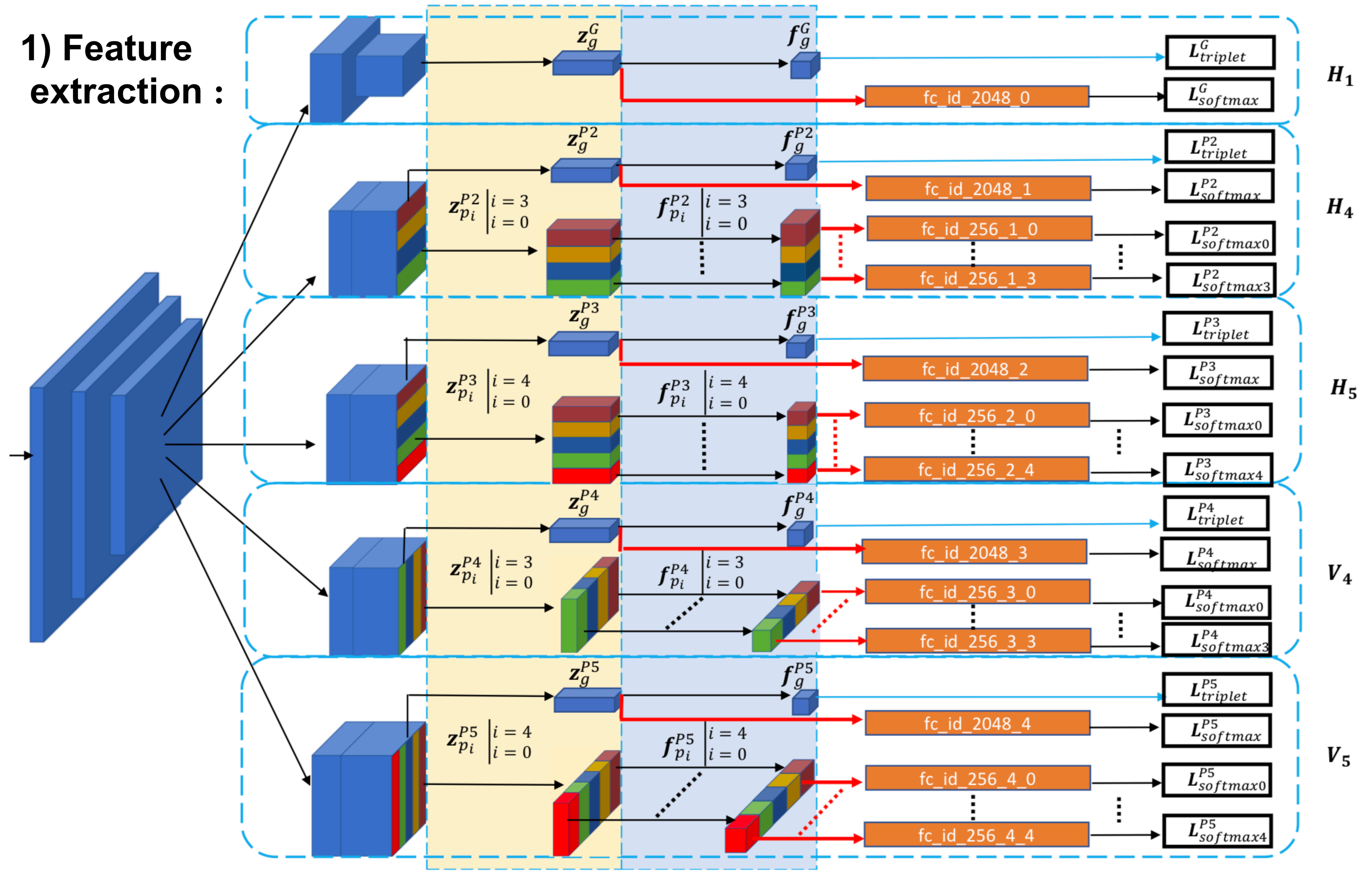
Background (averaged from non-detection area)

	F1	Recall	Precision
w/o refinement	0.253	0.994	0.156
w/ refinement	0.695	0.993	0.560

Detection results, without and with refinement

(b) Tracklet generation (ReID model)

1) Feature extraction :



2) Basic constraints:

- Kernighan-Lin graph partitioning
- ➢ Feature correlation
- ➢ Pixel-level motion correlation
- ➢ GPS-level motion correlation

camera calibration from CityFlow dataset

$$L_{softmax} = -\sum_{i=1}^{N} \log \frac{e^{\mathbf{W}_{y_i}^T \mathbf{f}_i}}{\sum_{k=1}^{C} e^{\mathbf{W}_k^T \mathbf{f}_i}}$$

$$L_{triplet} = -\sum_{i=1}^{P}\sum_{a=1}^{K}[\alpha + \max_{p=1\ldots K} \|\mathbf{f}_a^{(i)} - \mathbf{f}_p^{(i)}\|_2 - \min_{\substack{n=1\ldots K\\ j=1\ldots P\\ j\neq i}} \|\mathbf{f}_a^{(i)} - \mathbf{f}_n^{(j)}\|_2]_+$$

Granularity	Backbone	Resolution	MAP	Rank-1	Rank-5	MAP (RK)	Rank-1 (RK)	Rank-5 (RK)
$H1$	ResNet50	224	**66.7**	91.8	96.1	**71.4**	92.9	94.9
$H1, H2$	ResNet50	224	**74.4**	94.9	97.3	**76.9**	95.4	96.2
$H1, H2, H3$	ResNet50	224	**77.7**	95.3	97.1	**79.8**	95.5	96.5
$H1, H2, H3, H4$	ResNet50	224	**79.5**	95.5	97.6	**81.5**	96.1	97.4
$H1, H2, H3, H4$	ResNet50	256	**80.5**	96.1	98.0	**82.4**	96.4	97.4
$H1, H2, H3, H4$	SEResNext50	256	**81.9**	96.2	97.9	**83.9**	96.9	97.7

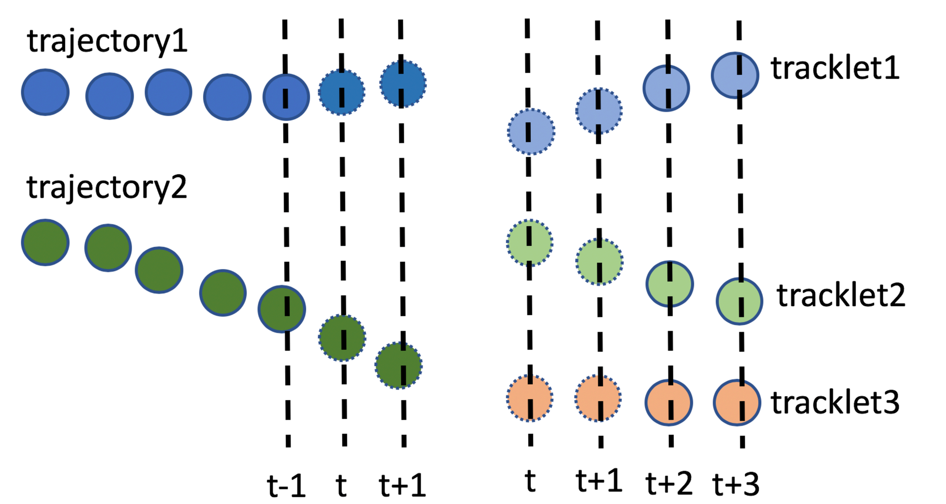
Ablation study on Veri776. "RK" refers to the Re-Ranking.

Method	mAP	Rank-1
OIFE [21]	48.0	89.4
VAMI [27]	50.1	77.0
GSTE [2]	59.5	96.2
MoV1BS [10]	67.6	90.2
Ours	**83.9 (+16.3%)**	**96.9 (+6.7%)**

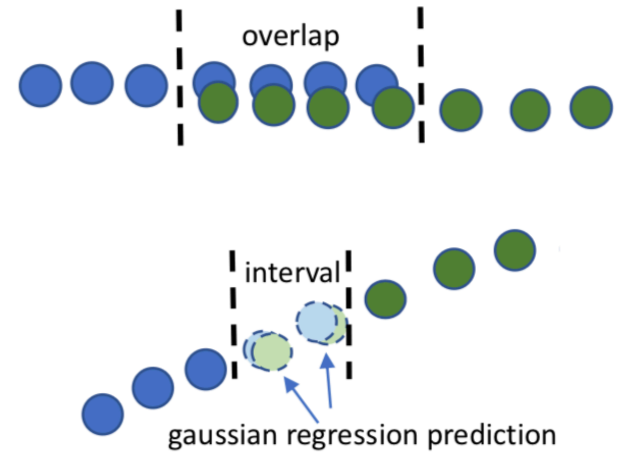
The evaluation of ReID model on Veri776

(c) Trajectory generation and refinement

1) Progressive connection:

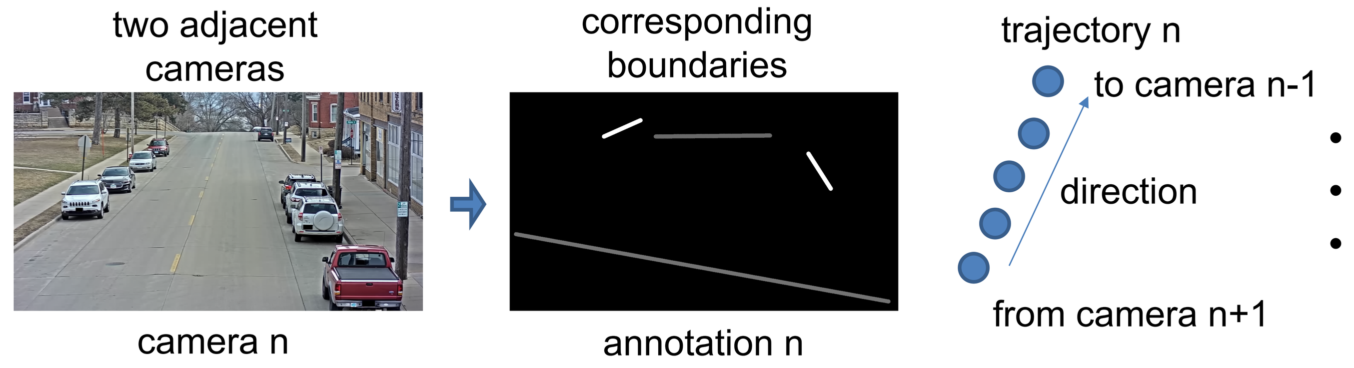


2) Trajectory refinement:



(d) Multi-camera identity matching

Trajectory matching with mean feature under three constraints



- Direction constraint
- Camera location constraint
- Road boundary constraint

Acknowledgement: This work was supported by the National Science Foundation under award number IIS-1400802 and Natural Science Foundation of China (61420106007, 61671387). Yucheng Chen's contribution was made when he was a visiting student at the City University of New York, sponsored by the Chinese Scholarship Council.