The 2018 NVIDIA AI City Challenge

Milind Naphade[1] Ming-Ching Chang[2] Anuj Sharma[3] David C. Anastasiu[4]
Vamsi Jagarlamudi[3] Pranamesh Chakraborty[3] Tingting Huang[3] Shuo Wang[1]
Ming-Yu Liu[1] Rama Chellappa[5] Jenq-Neng Hwang[6] Siwei Lyu[2]

[1] *NVIDIA Corporation, CA, USA*
[2] *University at Albany, State University of New York, NY, USA*
[3] *Iowa State University, IA, USA*
[4] *San José State University, CA, USA*
[5] *University at Maryland, College Park, MD, USA*
[6] *University of Washington, WA, USA*

Abstract

The NVIDIA AI City Challenge has been created to accelerate intelligent video analysis that helps make cities smarter and safer. With millions of traffic video cameras acting as sensors around the world, there is a significant opportunity for real-time and batch analysis of these videos to provide actionable insights. These insights will benefit a wide variety of agencies, from traffic control to public safety. The second edition of the NVIDIA AI City Challenge, being organized as a CVPR workshop, provided a forum to more than 70 academic and industrial research teams to compete and solve real-world problems using traffic camera video data. The Challenge was launched with three tracks — speed estimation, anomaly detection, and vehicle re-identification. Each track was chosen in consultation with traffic and public safety officials based on the value of potential solutions. With the largest available dataset for such tasks, and ground truth for each track, the Challenge enabled 22 teams to evaluate their solutions. Given how complex these tasks are, the results are encouraging and reflect increased value addition year over year for the Challenge.

1. Introduction

There will be a billion cameras worldwide by 2020. The NVIDIA AI City Challenge was launched in 2017 to create datasets that would enable academic and industrial research teams around the world to advance the state-of-the-art in intelligent video analysis for a variety of real-world problems. The Challenge was inspired by the success of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [12] enabling dramatic improvement in object detection, localization and classification for web-scale images.

Despite the existence of some corpora and benchmarks for video retrieval (*e.g.*, NIST TRECVID [1]), there is a clear lack of a large scale labeled corpus of high quality video data for traffic or public safety that would reflect the scale at which such analysis needs to be executed at in order to succeed in real-world conditions. We envision intelligent video analysis to help with several city-scale problems such as traffic, public safety, crime prevention, efficient resource utilization, improving the experience of users in large public and private spaces such as malls, stadia, train stations, airports, *etc.* Traffic and transportation, for example, can benefit from actionable insights that can be derived from data captured by street cameras, where the insights can help understand traffic patterns, adaptively control signaling systems, monitor infrastructure, and detect incidents in real-time. In 2017, the inaugural NVIDIA AI City Challenge [11] focused on the analysis of traffic camera videos at several intersections. The challenge task was primarily object detection, localization and classification using the largest annotated intersection video corpus created.

To build upon the 2017 challenge, the second edition of the AI City Challenge (**AIC18**) focused on three real-world problems. We decided to take the challenge beyond bounding boxes into metrics and insights that matter to transportation and traffic agencies like the United States Department of Transportation. These agencies are in dire need of systems that can automatically analyze traffic video content. It is unrealistic for humans to eyeball all the pixels, and

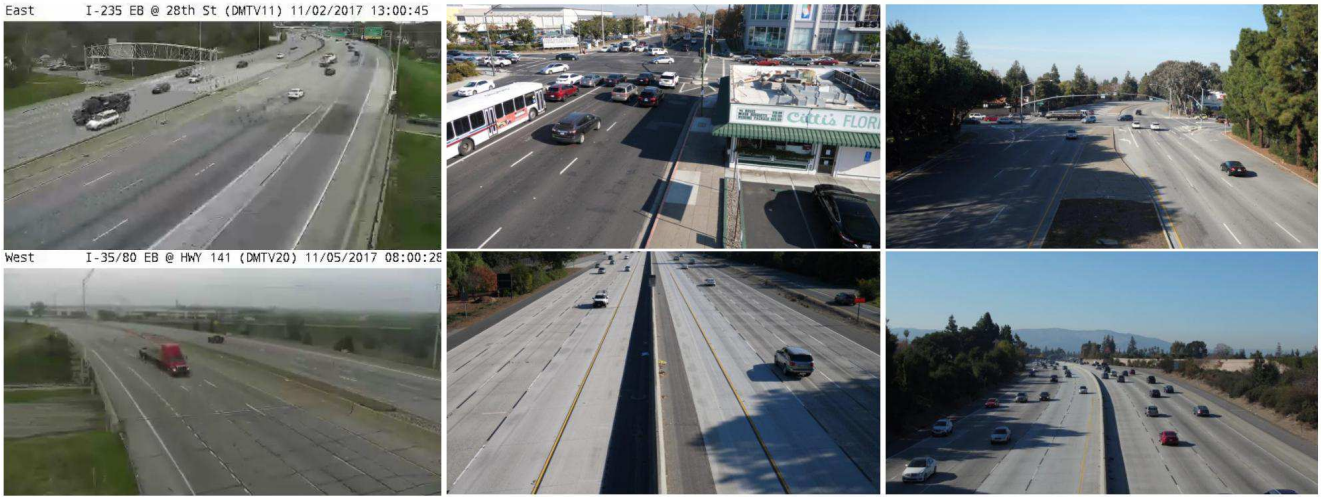

Figure 1: **AIC18 Dataset:** example data frames captured at traffic intersections and highways in Iowa and Silicon Valley.

a lack of such systems means that most of these data go unused. Upon consultation with several traffic and transportation agencies, as well as public safety organizations, we decided to focus on three tasks (§ 2):

1. Estimating traffic flow characteristics, such as the precise location and speed of each visible vehicle at any time.

2. Leveraging unsupervised approaches to detect anomalies caused by crashes, stalled vehicles, *etc.* This can be used to get the humans in the loop to pay attention to meaningful visual information in situations where timely intervention can save lives.

3. Multi-camera tracking and object re-identification in urban environments. This is very useful in traffic analysis as well as identifying and preventing crime. This also enables users to react to unfolding events as fast as possible.

We captured intersection and highway data from multiple cities and states for the Challenge (see Figure 1). The major difficulty in evaluating teams' performance on these tasks was coming up with an efficient approach to generate ground truth. On the one hand, we needed to record and measure the movement of a number of vehicles when the data was being captured. On the other hand, we wanted to avoid having to label every vehicle observed by the cameras. The approach we took was to create a fleet of control group vehicles and use their trips to generate the ground truth. On Jan 20, 2018, we shared the entire dataset with participating teams, as detailed in § 3.

More than 70 teams originally signed up to participate in the AIC18 Challenge. Submissions to the three tracks were due April 5, 2018. By then, a number of teams had dropped out, due to the complexity of the challenge and the very short period of time for completing the tasks. We received multiple submissions from 22 teams. Each team competed in one or more challenge tracks, as detailed in Table 1. An on-line evaluation system (§4) was developed and deployed that allowed teams to submit multiple runs against each track, allowing them to improve their track performance for a period of 1 week. However, teams were only allowed to see the performance of other teams when the submission period ended.

As anticipated, most teams performed well on the vehicle localization and speed estimation task, moderately well on the anomaly detection task, and really struggled but also surprised us on the vehicle re-identification task (§ 5). The re-identification task was deliberately designed to be the most challenging, especially with the large volume of video data.

The results of this year's challenge indicate that we have made progress year-over-year in moving the bar higher, going beyond mere bounding boxes around vehicles. We have also brought teams closer to real-world problems whose solutions will offer significant positive impact for traffic analytics and public safety.

2. Challenge Setup

The AIC18 Challenge allowed participants to compete in one or more of the following three tracks. Teams were required to submit their code for independent verification before being announced as winners.

Track 1 - Traffic Flow Analysis. Participating teams were asked to submit results for individual vehicle speeds in a test set containing 27 HD 1920x1080 videos, each 1-minute in length. Performance was evaluated based on the ground truth generated by a fleet of control vehicles (with accurate GPS tracking) driven during the recording. Evaluation for Track 1 was based on the detection rate of the

Table 1: The 22 participating teams and leaderboard rankings as of April 5, 2018. There are 13 teams competing in Track 1, 7 teams in Track 2, and 10 teams competing in Track 3. Top-2 teams in each track are highlighted in bold.

Team ID	Institution	Track 1	Track 2	Track 3
4	Vietnam Nat'l Univ. [16]	6		
6	Brno Univ. Tech. [14]	8		8
10	Conduent Inc.			9
12	Columbia Univ. [9]	5	7	
15	Panasonic, HUST, NTU, CAS [20]		**1**	
18	Univ. Albany SUNY [2]	11	3	4
24	Stevens Inst. Tech.	4		
25	Peking Univ.		4	
26	San José State Univ [5]	10		
28	Peking Univ., Beijing Inst. Tech.			5
31	Hacettepe Univ.			10
37	Nat'l Taiwan Univ. [19]			**2**
39	CERTH, Maastricht Univ. [4]	13	5	
40	Iowa State Univ. [6]	9		
41	ULPGC, UNIMORE [10]			6
45	Iowa State Univ.	12		
48	Univ. Washington [15]	**1**		**1**
53	Univ. Illinois Urbana-Champaign			7
63	Beijing Univ Posts & Telecom. [17]		**2**	
65	Univ. Maryland CP [7]	7		
78	UIUC, IBM, SIT [13]	3		
79	Beihang Univ., UCAS, USC [3]	**2**	6	3

control vehicles and the *root mean square error* (RMSE) of the predicted control vehicle speeds (§ 4.1).

Track 2 - Anomaly Detection. Participating teams were asked to submit the anomalies detected in a test set containing 100 video clips, each approximately 15 minutes in length. The anomalies were either due to car crashes or stalled vehicles. Regular congestion not caused by any traffic incident was not counted as an anomaly. A multi-car event (*e.g.*, one crash followed by another crash, or a stalled car followed by someone else stopping to help) were considered a single anomaly. More specifically, if an anomaly occurred while another anomaly was already in progress, the two counted as a single anomaly. Evaluation for Track 2 was based on anomaly detection performance, measured by the F_1 score, and detection time error, measured by RMSE (§ 4.2).

Track 3 - Multi-camera Vehicle Detection and Re-identification. Participating teams were asked to identify all vehicles that were visible in the camera view at each of 4 different locations in a set of 15 videos, each 30 to 90 minutes long. Evaluation for Track 3 was based on detection accuracy and localization sensitivity for a set of ground-truth vehicles that were driven through all camera locations at least once (§ 4.3).

3. Dataset

Video data provided for this Challenge has been recorded by cameras aimed at intersections and along highways in urban areas, see Figure 1. The **AIC18 Dataset** consists of the following video data sources: [1]

- **Silicon Valley Highways and Intersection Data.** More than 15 hours of HD 1920x1080 data at 30 frames per second captured at multiple locations with synchronized recording.
- **Iowa DOT.** More than 24 hours of 800x410 resolution data at 30 frames per second captured by the Iowa Department of Transportation (DOT) traffic cameras.

To avoid having to manually label and identify each vehicle in the traffic videos, we set up a group of control vehicles. Each volunteer from the control group then drove through the various intersections and along highways based on a designed script. Each control vehicle also carried a smartphone with an application that recorded GPS information for the entire journey of the vehicle, thus providing the

[1] Participants were encouraged to use the data set available from the **SUNY Albany UA-DETRAC benchmark suite** http://detrac-db.rit.albany.edu [18] in case they needed to develop models for vehicle detection and tracking.

ground truth for speed and localization. A meticulous and manual process was used to synchronize the dataset from the journey diaries on the smartphone application and the videos in which these vehicles were observed. Ground truth files were then created for automated evaluation.

The provided re-identification dataset is much more challenging than some of the conventional re-identification datasets such as VeRi [8] in the following aspects: (1) There is a large variation in site locations, including street intersections and highways, with camera positions being separated by several miles. (2) The duration of the videos is long (up to 2 hours), spanning several days of recording. (3) The resolution of the video does not permit simple license plate recognition as the primary feature in the re-identification model. (4) The control vehicles appear in one of several cameras at least once and possibly multiple times.

The ground truth files contain the bounding box and accurate travel speed of control vehicles in all frames where they are observed. The bounding boxes of the control vehicles were manually annotated using standard tools. The travel speed of each control vehicle was recorded once per second. Instantaneous speed values for each control vehicle bounding box were assigned using interpolation and calibration.

4. Evaluation Methodology

To allow teams the most possible time to improve their results and experiment with their algorithms, we developed an **on-line evaluation system** that automatically measured the effectiveness of results for each track upon submission and stored results in a database. The system returned an error message if results were not in an acceptable format or problems were encountered when computing the performance scores for each track. Teams were allowed a maximum of 5 submissions per day for each track. After the Challenge submission deadline, teams could see a leaderboard with the best results from each team. The leaderboard was sorted in decreasing order with respect to the corresponding performance scores.

4.1. Track 1 Evaluation

Performance evaluation in Track 1 is based on the ability to localize control vehicles and predict their speed, with speed prediction being the primary concern. As such, the score for Track 1 (S_1), for each participating team, is computed as:

$$S_1 = DR \times (1 - NRMSE^s), \tag{1}$$

where DR is the detection rate and $NRMSE^s$ is the *normalized* root mean square error of estimated speed in miles-per-hour (MPH).

The primary task of Track 1 is estimating vehicle speeds. We compute the speed estimate error as the RMSE of the ground truth vehicle speed and predicted speed for all correctly detected ground-truth vehicles. A vehicle is said to be *detected* if it was localized in at least thirty-percent of frames it appeared in. A vehicle is *localized* if at least one predicted bounding box exists with *intersection-over-union* (IOU) score of $\tau_{iou} = 0.3$ or higher relative to the annotated bounding box for the vehicle. If multiple bounding boxes with IOU $\geq \tau_{iou}$ exist, we consider only the speed estimate from the one with the highest confidence score. To obtain a normalized evaluation score, we calculate $NRMSE^s$ as the normalized vehicle speed RMSE score across all teams, which is obtained via min-max normalization given the best speed estimate scores from each team. Specifically, $NRMSE^s$ of team i is the relative speed RMSE performance of team i in comparison to all other competing teams, and is computed as

$$NRMSE^s_i = \frac{RMSE_i - RMSE_{min}}{RMSE_{max} - RMSE_{min}} \tag{2}$$

The detection rate DR is computed as the ratio of the number of detected ground truth vehicles and the total number of ground truth vehicles. We expect all control vehicles to be properly detected. The DR part of the S_1 score simply acts as a penalizing component if some control vehicles are missed.

4.2. Track 2 Evaluation

Performance evaluation in Track 2 is based on the ability of a model to detect anomalies, measured by the F_1 score, and the amount of error in detection time, measured by the RMSE of the time elapsed between the start of the anomaly and its prediction. Specifically, the Track 2 score (S_2), for each participating team, is computed as

$$S_2 = F_1 \times (1 - NRMSE^t), \tag{3}$$

where the F_1 score is the harmonic mean of the precision and recall of anomaly prediction. Precision is defined as the ratio of the anomalies correctly identified to the number of anomalies submitted. Recall is defined as the ratio of the anomalies correctly identified to the number of ground truth anomalies. For video clips containing multiple ground truth anomalies, credit is given for detecting each anomaly. Conversely, multiple false alarm submissions in a single video clip are counted as multiple false alarms. If multiple anomalies are provided within the time span of a single ground truth anomaly, we consider the one with minimum detection time error and ignore the rest.

The primary component of the score in Track 2 is the amount of time elapsed from the onset of an anomaly until its automatic detection by the model. Thus, we compute the detection time error as the RMSE of the ground truth anomaly start time and predicted anomaly start time

Table 2: **Track 1 Leaderboard.** $RMSE$ measures estimated vehicle speed in MPH. #S denotes number of submission trials.

Team	Institute	S_1	DR	$RMSE$	#S
48	UW	1.0000	100.00%	4.096	41
79	BeihangU	0.9162	100.00%	6.041	24
78	UIUC	0.8892	100.00%	6.667	22
24	Stevens IT	0.8813	100.00%	6.849	10
12	ColumbiaU	0.8331	100.00%	7.970	13
4	VietnamUN	0.7924	100.00%	8.914	13
65	UMaryland	0.7654	100.00%	9.541	8
6	BrnoUT	0.7174	81.48%	6.869	22
40	Iowa SU	0.6564	81.48%	8.609	5
26	SJSU	0.6547	100.00%	12.109	7
18	UAlbany	0.6264	85.19%	10.340	11
45	Iowa SU	0.5953	96.29%	12.957	10
39	CERTH	0.0000	88.89%	27.302	4

Table 3: **Track 2 Leaderboard.** $RMSE$ measures anomaly detection as seconds from the anomaly start. #S denotes number of submission trials.

Team	Institute	S_2	F_1	$RMSE$	#S
15	Panasonic	0.8649	0.8649	3.6152	24
63	BeijingPost	0.7853	0.8108	10.2369	19
18	UAlbany	0.4951	0.6286	48.3406	9
25	PekingU	0.2638	0.4762	97.5505	11
39	CERTH	0.0640	0.2363	157.2298	12
79	BeihangU	0.0069	0.7567	212.3274	3
12	ColumbiaU	0.0000	0.7692	214.2712	4

for all true positive predictions. We compute $NRMSE^t$ as the normalized RMSE score, showing the relative detection time error performance of the team compared to all other teams, as in Eq.(2). We expect all anomalies to be successfully detected and penalize missed detections and spurious ones through the F_1 component in the S_2 evaluation score.

4.3. Track 3 Evaluation

Performance evaluation in Track 3 is based on tracking accuracy and localization sensitivity for a set of ground-truth vehicles that were driven through 4 camera locations at least once. Specifically, the Track 3 score (S_3), for each participating team, is computed as

$$S_3 = \frac{TDR + PR}{2}, \tag{4}$$

where TDR is the track detection rate and PR is the localization precision.

TDR is computed as the ratio of correctly identified ground-truth vehicle tracks and the total number of ground-truth vehicle tracks. A vehicle track is correctly identified if the vehicle has been localized (IOU $\geq \tau_{iou}$) and associated with the same object ID in at least thirty-percent of the frames containing the ground-truth vehicle in a given video. PR is the localization precision, which is calculated as the ratio of correctly localized bounding boxes and the total number of predicted boxes across all videos. Since both detection and tracking of the vehicles in question and precise localization are important in Track 3, the S_3 score is simply computed as the average of TDR and PR.

Table 4: **Track 3 Leaderboard.** #S denotes number of submission trials.

Team	Institute	S_3	TDR	PR	#S
48	UW	0.7106	0.4286	0.9925	22
37	NTaiwanU	0.2861	0.5714	0.0007	20
79	BeihangU	0.0785	0.1429	0.0142	17
18	UAlbany	0.0074	0.0000	0.0147	22
28	PekingU	0.0026	0.0000	0.0052	6
41	UNIMORE	0.0024	0.0000	0.0049	6
53	UIUC	0.0002	0.0000	0.0004	1
6	BrnoUT	0.0001	0.0000	0.0001	11
10	Conduent	0.0000	0.0000	0.0000	1
31	HacettepeU	0.0000	0.0000	0.0000	1

5. Submission Results

Table 1 summarizes all participating teams and their challenge results. Out of all **79** registered teams (56 for Track 1, 53 for Track 2, 61 for Track 3), **22** teams submitted results (13 for Track 1, 7 for Track 2, 10 for Track 3).

5.1. Track 1 Challenge Summary

Most Track 1 methods adopt the tracking-by-detection paradigm followed by inference of real-world speeds from pixel distance increments. Thanks to the rapid advancement of deep neural networks (DNN), teams were able to obtain very good detection results using some of the latest DNNs, including YOLO2 (Team 48), DenseNet (Team 79), Mask R-CNN (Teams 78, 12, 65), and Faster R-CNN (Teams 4, 6, 40, 18, 39). The Mask R-CNN model, in particular, was able to detect and localize small vehicles with excellent precision. As seen in Table 2, several leading teams were able to obtain a DR of 100%, which highlights the strong capabilities of the latest deep learning methods.

For tracking, the IOU score between detector and tracker boxes was often used for vehicle ID assignment (Teams 4, 18, 26). Tracking methods varied considerably among teams, relying on, *e.g.*, clustering-based association (Team 48), graph optimization (Team 79), medianflow (Team 78), and Kalman filtering (Teams 12, 65, 18). Both on-line and off-line tracking methods were used.

The mapping from tracked vehicle pixels to real-world

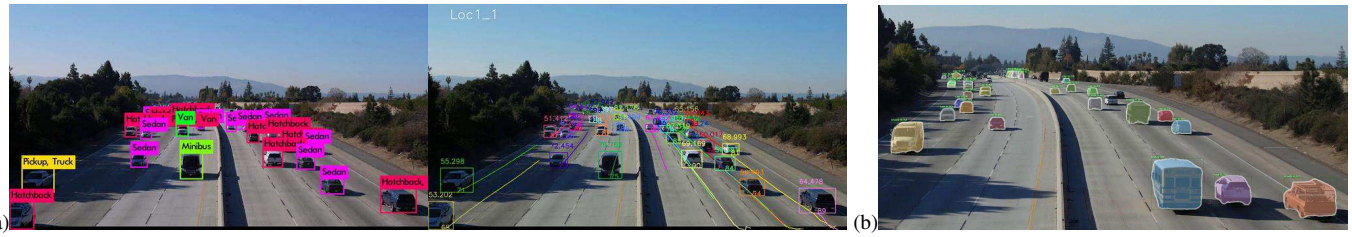

Figure 2: **Highlights of Track 1 methods.** (a) **Team 48:** left: detected vehicles with car types, right: vehicle trajectories with estimated speed in MPH. (b) **Team 78:** Vehicle detection using Mask R-CNN.

coordinates is required for speed estimation. Both semi-automatic camera calibration methods, *e.g.*, relying on estimating the *vanishing lines, points* and scale (Teams 48, 78, 6), or fully manual methods (Team 18) were used. Several teams did not rely on calculating the camera projection matrix, and instead estimated the planar homography to compute the projective image warping using image landmarks (Teams 12, 4, 65, 40) and then estimated the scale. Robust speed estimation was then obtained using smoothing. Several teams made assumptions that vehicles on the highway traveled at constant speed (Team 48, 4, 26, 18), or followed a known speed limit (Team 26).

Table 2 summarizes the leaderboard as of Apr 5, 2018 for the Track 1 challenge. Team 48 was at the top on the board. Their YOLO2 vehicle detector was trained based on manual labeling of 4500 frames of Track 1 videos into 8 categories. Tracking was performed based on a large set of dedicated designed loss functions, considering tracklet smoothness, appearance, velocity, and time interval. The explicit clustering of the tracklet assignment, merge, split, and switch also improved the overall vehicle detection score. Figure 2 shows visual highlights of results from the top teams in the Track 1 challenge.

5.2. Track 2 Challenge Summary

Test videos in Track 2 are real-world traffic videos recorded in a wide range of viewpoints, weather, and road conditions. These issues make it difficult to design general-purpose anomaly detection methods. Thus, most successful approaches are based on traffic motion flow analysis (*e.g.*, using optical flow) rather than trying to detect and track individual vehicles. In fact, stalled vehicles mostly occur on the side of the road. Thus, a region-of-interest (ROI) can be estimated after traffic lanes are delineated. Teams 15, 18, and 79 performed lane finding based on optical flow and background analysis to obtain a refined anomaly ROI. An event classifier was then applied on the ROI to detect stalled vehicles across large time windows, using, *e.g.*, ResNet (Team 15), VGGNet (Team 63), feature histograms (Team 18), or SVM (Team 39). Both the top-2 teams (Teams 15, 63) used Faster R-CNN to detect stalled vehicles.

Table 3 shows the leaderboard for the Track 2 challenge. Team 15 claimed the top position by using a dual-mode

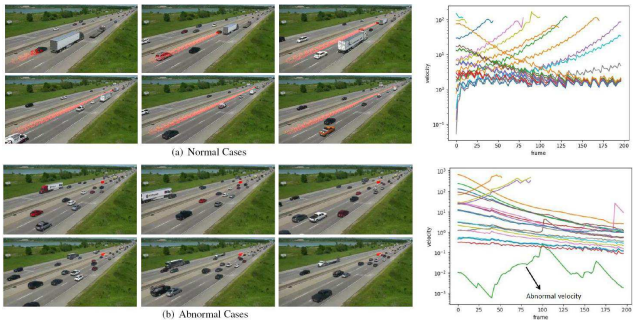

Figure 3: **Highlights of Track 2 methods. Team 15:** images on the left show trajectories in several frames at distinct times. Curves on the right show estimated vehicle velocity. The stalled vehicle can be easily identified.

(static and dynamic) analysis method that integrates background modeling, vehicle detection and segmentation using Mask R-CNN, followed by outlier filtering. Figure 3 shows highlights of this method.

5.3. Track 3 Challenge Summary

The Track 3 challenge is extremely difficult, since the vehicle to be (re-)identified can appear anywhere, anytime, and possibly multiple times in the videos across the 4 sites. Thus, although a naïve method could rely on brute-force pairwise comparisons, it would take too long to execute. The solutions proposed by the teams are much more effective. They compared vehicles based on tracking, relying on the whole space-time tracklets (Team 48, 37), or based on vehicle images, focusing on one image from each tracklet (*e.g.*, Team 18 selected the image enclosed by the largest vehicle bounding box in each tracklet). Various deep features are extracted for pairwise re-identification matching, *e.g.*, using a fusion of loss functions considering vehicle types, appearance, and other similarities (Team 48), or based on the triplet loss (Team 18). Candidate vehicle tuples across the 4 test sites can be nominated by repeatedly applying and extending these pairwise matches. Note that while person re-identification is actively researched, much fewer vehicle re-identification datasets are available for training the deep vehicle features. The "VeRi" dataset[2] was used for training by both Teams 37 and 18.

[2] https://github.com/VehicleReId/VeRidataset

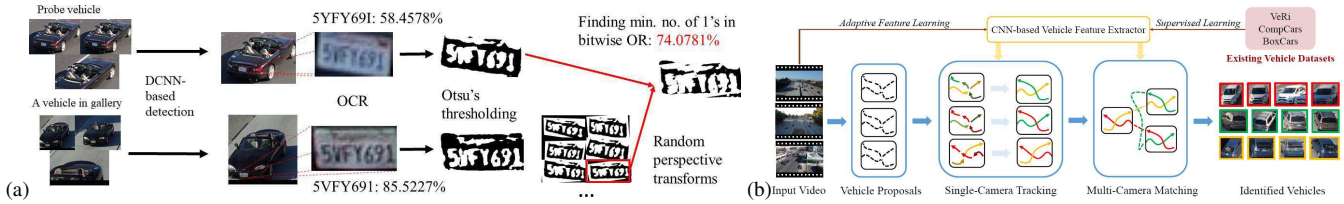

Figure 4: **Highlights of Track 3 methods.** (a) **Team 48:** the use of license plate matching requires accurate 3D vehicle modeling, however can greatly improve re-identification. (b) **Team 37:** re-identification is based on matching space-time track features.

Table 4 shows the leaderboard for the Track 3 challenge. Team 48 claimed the leading spot on the board. Their method incorporated a wide range of loss estimations (*e.g.*, 1024-dimensional DCNN features based on the GoogLeNet model trained on CompCar) with a large ensemble of rules (appearance, license plate recognition, travel time loss, *etc.*) that lead to successful re-identification of about half of the control vehicles. As seen in Table 4, the re-identification precision of Team 48 was also remarkable. Team 37 performed vehicle detection using the Facebook Detectron model configured similarly to the ResNet101 model, based on a feature pyramid network. Deep space-time features were extracted from the tracks using the ResNet50 model trained on various datasets. Re-identification was performed using multi-task learning, followed by matching across multiple cameras. Figure 4 shows highlights of approaches from the top two teams.

6. Discussion

Based on the results of the 2017 AI City Challenge, we hypothesized that the community was ready for taking on higher-level use cases. The 2018 Challenge focused on the use cases of speed estimation, anomaly detection and vehicle re-identification. The Challenge succeeded in bringing the computer vision community closer to leveraging traffic video analysis for real-world traffic and public safety problems. These real-world problems are not specifically addressed in most of the existing efforts that we are aware of. Based on the teams' submissions, we make the following observations.

Multiple object tracking (following the tracking-by-detection paradigm) is yet to mature for problems such as occlusions. Light-weight, on-line tracking methods are preferred, however, most current leading methods are complex and off-line (such as the leading method used by Team 48). Furthermore, the joint problems of multi-camera multi-object tracking and vehicle re-identification can be resolved together, if appropriate deep features can be leveraged. Camera calibration or image warping (which requires a pixel-to-world mapping) is a well-studied topic. Practical methods such as auto-calibration (Teams 48, 78, 26), planar homography (Teams 79, 12), affine warping (Teams 65, 40), or even a full site calibration (Team 18) all require manual initialization, in order to estimate scale from knowledge of world measures. Vehicle speed estimation seems to be a feasible problem in general, but there is still a wide spread in $RMSE$ across the various submissions. We expect vision-based methods to be nearly as accurate as other speed estimation techniques such as RADAR-based estimation, and be used pervasively in the near future.

Traffic anomaly detection is a difficult problem especially when no constraining assumptions can be made about the video quality, illumination and environmental conditions. Some anomalies are easier to describe, whereas others may be more complex. Training sets for such anomalies are also rare. The results we saw this year were therefore promising given the level of difficulty. We hope to extend this to several other anomalies such as wrong-lane driving, illegal turns and traffic light violation, *etc.* We anticipate that AI will become a pervasive detection and alerting tool to the human operators in command and control centers as the performance on this track improves.

Finally, vehicle re-identification at city-scale is the most challenging of the three tracks. To the best of our knowledge, this challenge is the first to attempt this task at such a large scale, across many hours of videos and multiple sites. We note that machine generated re-identification is indeed very different from human approaches, which generally start from determining vehicles types, makes, model, colors, *etc.* Fine-grained vehicle model recognition continues to be an open problem. This challenge will be considered successful when AI becomes a force multiplier in public safety cases such as Amber Alert. New frontiers of visual AI technology can perhaps soon allow public safety officials to shorten time in forensic investigation through hundreds of hours of videos across a city.

7. Conclusion

The 2018 NVIDIA AI City Challenge (AIC18) evaluated the application of state-of-the-art computer vision and deep learning technologies to real-world traffic analysis problems, providing insights into understanding the opportunities and gaps that need to be overcome for the pervasive use of AI in traffic and public safety solutions. We hope that future versions of this challenge continue to push the envelope of computer vision and deep learning in these solutions.

8. Acknowledgement

The AIC 2018 Challenge would not have been possible without significant help from several volunteers at NVIDIA Corporation who helped create the ground truth for the various challenge tracks including the control fleet operations. Sushma Nagaraj of San José State University created the bounding box ground truth based on the travel diaries of control fleet volunteers.

References

[1] G. Awad, A. Butt, J. Fiscus, M. Michel, D. Joy, W. Kraaij, A. F. Smeaton, G. Qunot, M. Eskevich, R. Ordelman, G. J. F. Jones, and B. Huet. TRECVID 2017: Evaluating ad-hoc and instance video search, events detection, video description and hyperlinking. In *Proc. TRECVID 2017*. NIST, USA, 2017.

[2] M.-C. Chang, Y. Wei, N. Song, and S. Lyu. Video analytics in smart transportation for the AIC'18 challenge. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[3] W. Feng, D. Ji, Y. Wang, S. Chang, H. Ren, and W. Gan. Challenges on large scale surveillance video analysis. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[4] P. Giannakeris, V. Kaltsa, K. Avgerinakis, A. Briassouli, S. Vrochidis, and I. Kompatsiaris. Speed estimation and abnormality detection from surveillance cameras. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[5] S. Hua, M. Kapoor, and D. C. Anastasiu. Vehicle tracking and speed estimation from traffic videos. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[6] T. Huang. Traffic speed estimation from surveillance video data. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[7] A. Kumar, P. Khorramshahi, W.-A. Lin, P. Dhar, J.-C. Chen, and R. Chellappa. A semi-automatic 2D solution for vehicle speed estimation from monocular videos. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[8] X. Liu, W. Liu, T. Mei, and H. Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *ECCV*, pages 869–884. Springer, 2016.

[9] T. Mao, W. Zhang, H. He, Y. Lin, V. Kale, A. Stein, and Z. Kostic. Aic2018 report: Traffic surveillance research. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[10] P. A. Marín-Reyes, A. Palazzi, L. Bergamini, S. Calderara, J. Lorenzo-Navarro, and R. Cucchiara. Aic2018 aimagelab. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[11] M. Naphade, D. C. Anastasiu, A. Sharma, V. Jagrlamudi, H. Jeon, K. Liu, M.-C. Chang, S. Lyu, and Z. Gao. The NVIDIA AI City Challenge. In *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing&Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2017.

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 115(3):211–252, 2015.

[13] H. Shi, Z. Wang, Y. Zhang, X. Wang, and T. Huang. Geometry-aware traffic flow analysis by detection and tracking. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[14] J. Sochor, J. Špaňhel, R. Juŕanek, P. Dobeš, and A. Herout. Graph@fit submission to the nvidia ai city challenge 2018. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[15] Z. Tang, G. Wang, H. Xiao, A. Zheng, and J.-N. Hwang. Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[16] M.-T. Tran, T. Dinh-Duy, T.-D. Truong, V. Ton-That, T.-N. Do, Q.-A. Luong, T.-A. Nguyen, V.-T. Nguyen, , and M. N. Do. Traffic flow analysis with multiple adaptive vehicle detectors and landmark-based scanlines. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[17] J. Wei, J. Zhao, and Y. Zhao. Unsupervised anomaly detection for traffic surveillance. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[18] L. Wen, D. Du, Z. Cai, Z. Lei, M. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *arXiv CoRR*, abs/1511.04136, 2015.

[19] C.-W. Wu, C.-T. Liu, C.-E. Jiang, W.-C. Tu, and S.-Y. Chien. Vehicle re-identification with the space-time prior. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.

[20] Y. Xu, X. Ouyang, Y. Cheng, S. Yu, L. Xiong, S. Pranata, S. Shen, and J. Xing. Dual-mode vehicle motion pattern learning for high performance road traffic anomaly detection. In *CVPR Workshop (CVPRW) on the AI City Challenge*, 2018.