The 2019 AI City Challenge

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

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

Abstract

The 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 2019 AI City Challenge is the third annual edition in the AI City Challenge series with significant growing attention and participation. AI City Challenge 2019 enabled 334 academic and industrial research teams from 44 countries to solve real-world problems using real city-scale traffic camera video data. The Challenge was launched with three tracks. Track 1 addressed city-scale multi-camera vehicle tracking, Track 2 addressed city-scale vehicle re-identification, and Track 3 addressed traffic anomaly detection. Each track was chosen in consultation with departments of transportation focusing on problems of greatest public value. With the largest available dataset for such tasks, and ground truth for each track, the 2019 AI City Challenge received 129 submissions from 96 individuals teams (there were 22, 84, 23 team submissions from Tracks 1, 2, and 3 respectively). Participation in this challenge has grown five-fold this year as tasks have become more relevant to traffic optimization and challenging to the computer vision community. Results observed strongly underline the value AI brings to city-scale video analysis for traffic optimization.

1. Introduction

Immense opportunity exists to make transportation systems smarter, based on sensor data from traffic, signaling systems, infrastructure, and transit. Unfortunately, progress has been limited for several reasons among them, poor data quality, missing data labels, and the lack of high-quality models that can convert the data into actionable insights. There is also a need for platforms that can handle analysis from the edge to the cloud, which will accelerate the development and deployment of these models. Two years ago, we launched the AI City Challenge [1] to address this gap. The hypothesis is that the AI City Challenge will provide a rich forum for computer vision and AI researchers to innovate and solve traffic operations challenges just like the ImageNet benchmark [30] turbocharged the domain of web-scale image analysis. Over successive years the challenge tracks of the AI City Challenge have been evolving in terms of amount of data, complexity of problems and their relevance to the real-world requirements as articulated by our partners in departments of transportation across multiple geographies.

As a result of the strong and growing participation in the two previous editions of this challenge, AI City Challenge 2017 [25] and AI City Challenge 2018 [26], the third edition continues to focus on city-scale traffic optimization. The AI City Challenge 2019 [1], organized as a workshop at CVPR 2019, focuses on the following three problems that are of great relevance in optimizing intersection as well as corridor level traffic flows and detecting and responding to incidents of interest that can help save lives, such as traffic accidents. The three tracks address the following:

- **City-scale multi-camera vehicle tracking:** This task uses a dataset we refer to as *CityFlow* [37]. This dataset has been captured in a mid-sized U.S. city using 40 cameras across multiple intersections spanning more than 2.5 km. The task is to track vehicles across multiple cameras and across multiple intersections.

- **City-scale multi-camera vehicle re-identification:** This task uses a subset of CityFlow dataset, which we dub *CityFlow-reID*. The task is to perform re-identification (ReID) of vehicles across multiple cameras based on image signatures.

- **Traffic anomaly detection:** This task uses a dataset provided by the Iowa Department of Transportation acquired from cameras overlooking major state and interstate highways. The task involves detecting anomalies, such as accidents and cars stopped on the side of the road, *etc*.

The datasets used in this challenge are, to our knowledge, the largest such labeled datasets based on real-world intersection and highway conditions. Care has also been taken in privacy preservation through the process of semi-automated redaction of faces and license plates when the objects are close to the traffic cameras.

This edition of the challenge has seen nearly a **five-fold** increase in participation. There were 334 teams with 1,004 researchers from 44 countries and regions that participated in the challenge. Of these, 96 unique teams from 22 countries and regions, including 355 individual members, submitted results to one or more tracks, totaling 129 team submissions across tracks, including 22, 84, and 23 submissions for the three tracks, respectively.

We deployed an online evaluation server (§ 4) that allows teams to submit multiple runs against each track. Each team was allowed to upload a limited number of submissions to each track to avoid excessive tweaking against the evaluation. In total, there were 1,337 successful submissions from participating teams. The challenge ran from January 2019 through May 2019. A public leader board showing the top three performing teams in each challenge track was also provided.

This paper describes the challenge setup (§ 2), challenge dataset preparation (§ 3), evaluation methodology (§ 4), and team submission results (§ 5). A discussion of developments, insights, and future trends in traffic analysis is included in § 6.

2. Challenge setup

The third edition of the AI City Challenge has been set up similarly to the second edition in terms of time-frame and rules. Participating teams were allowed to compete in one or more tracks. The dataset was made available to participants in early January 2019 and the teams had to finish their submissions by May 10, 2019. Teams were required to make their code available for verification and reproduction via GitHub. Teams were also required to make available via open-sourcing any additional labeling of the training dataset that they may have performed, which is essential to validate their performance.

Tracks 1 and 2 share the CityFlow dataset [37], which is captured from urban intersections, whereas the dataset for Track 3 is captured along state highways. To encourage advancement in traffic video analysis while simultaneously protecting privacy, the faces and license plates in the data have been appropriately redacted.

Track 1: City-Scale Multi-Camera Vehicle Tracking. Teams were asked to track vehicles across multiple cameras both at a single intersection and across multiple intersections spread out across a city. This helps traffic engineers understand journey times along entire corridors. The dataset for Track 1 is from CityFlow [37], which includes nearly 3 hours of synchronized videos collected from cameras across multiple urban intersections and urban freeways from multiple vantage points. Metadata about the collected videos, including GPS locations of cameras, camera calibration information and other derived data from videos were also made available to participating teams.

The evaluation metric used for Track 1 was IDF1 [29], which measures the ratio of correctly identified detections over the average number of ground-truth and computed detections. Compared to other measurements for multiple object tracking, *e.g.*, MOTA [5, 21], IDF1 helps resolve the ambiguity among error sources. Participating teams were provided with state-of-the-art vehicle detection and single-camera tracking baselines.

Track 2: City-Scale Multi-Camera Vehicle Reidentification. Teams were asked to perform vehicle ReID based on vehicle crops from multiple cameras placed at multiple intersections. Different from Multi-Target Multi-Camera (MTMC) tracking, image signatures of targets were extracted from trajectories to be compared directly. Spatio-temporal information was not required for solving this task. The dataset for Track 2 is a subset of CityFlow [37], which we refer to as the CityFlow-reID dataset. It contains 56,277 cropped images in total, which are split into training set, test set, and queries. To evaluate the performance of each team, we created a new evaluation metric for image-based ReID, named *rank-K mean Average Precision (mAP)*, that measures the mean of average precision for each query considering only the top K matches. K is chosen to be 100 in our evaluation.

Track 3: Traffic Anomaly Detection. The dataset for Track 3 includes more than 50 hours of video captured on state highways in Iowa. A training dataset containing a dif-

Figure 1. 5 views from one of the scenarios in the CityFlow benchmark [37] that was captured at a highway intersections.

Figure 2. 25 views from one of the scenarios in the CityFlow benchmark [37] that was captured across 7 intersections at residential area.

ferent set of 100 video clips was shared with the teams, an improvement over last year's setup. This training dataset was shared so that the teams could take advantage of a broader range of supervised or semi-supervised techniques for anomaly detection and improve detection performance. Similar to the 2018 AI City Challenge [26], 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) was considered a single anomaly. More specifically, if an anomaly occurred while another anomaly had already been in progress, the two were counted as a single anomaly. Evaluation for Track 3 was based on anomaly detection performance, measured by the F_1 score, and detection time error, measured by the root mean square error (RMSE) (§ 4.3).

3. Datasets

The datasets for the 2019 AI City Challenge came from the following sources, CityFlow [37] and Iowa DOT [26], which we further describe in this section.

3.1. The CityFlow Dataset

The CityFlow dataset has been curated specifically for Track 1 and Track 2 of the 2019 AI City Challenge. The dataset contains nearly 3 hours of video collected from 40 cameras spanning across 10 intersections in a mid-sized U.S. city. The distance between two furthest simultaneously recording cameras is 2.5 km, which is the longest among all the existing benchmarks for ReID and MTMC tracking. We thus claim CityFlow to be a city-scale benchmark and the first public benchmark to support vehicle-based MTMC tracking. Note that all human faces and vehicle license plates in CityFlow have been redacted to preserve privacy. The performance of a number of baseline methods [43, 18, 19, 38, 34] was provided to participating teams.

There are 5 scenarios in CityFlow, characterized by different scene type (highway and residential area), level of service (LOS), and number of cameras. In total, 229,680 bounding boxes of 666 vehicle identities were annotated, namely all vehicles that passed through at least 2 cameras. Resolution of each video is at least 960p, and most videos have a frame rate of 10 frames-per-second (FPS). In Track 1, the start time offset for each video in each scenario was provided, so that the participating teams could synchronize the videos for spatio-temporal analysis. Fig. 1 and Fig. 2 show the camera views from 2 of the 5 scenarios in CityFlow.

We employed a trajectory-level annotation strategy to efficiently annotate the MTMC tracking ground truth. First, we followed the tracking-by-detection paradigm to generate possibly noisy trajectories using the state-of-the-art in object detection [28] and single-camera tracking [38]. We then manually examined and corrected all detection and tracking errors, including misaligned bounding boxes, false negatives, false positives and identity switches. Fig. 3 shows an example of annotated targets across cameras.

Camera geometry for each scenario was provided for challenge use as well. We provided the camera *homography* matrices between the 2D image plane and the ground plane defined by GPS coordinates based on the flat-earth approximation. For camera calibration, several landmark points were manually selected in a sampled image frame from the video of each fixed camera view. The corresponding GPS coordinates in the real world were derived from Google Maps. The GPS coordinates were expressed in terms of latitude/longitude, and calibration results were stored in 64-bit double-precision floating point, which is sufficiently accurate to localize the position and movement of any street ob-

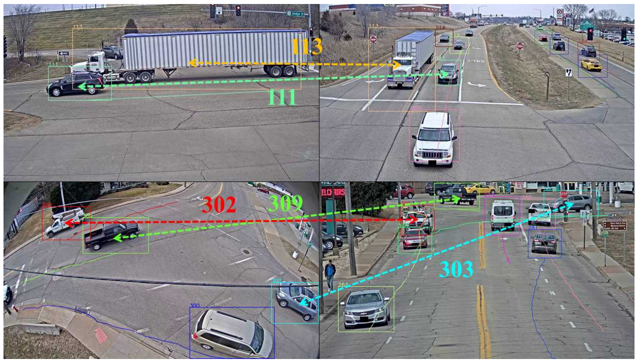

Figure 3. Demonstration of annotations on the CityFlow [37] dataset, with colored dashed lines indicating associations of object identities across camera views.

Figure 4. User interface of the Python visualization tool for qualitative evaluation on CityFlow-reID. Green and red rectangles represent true and false matches, respectively. This visualization is only available from the challenge organization, as the plotting requires ground truth.

ject. This optimization problem can be effectively solved by methods such as least median of squares, RANSAC, and the evolutionary algorithm [36, 35]. The objective cost function is the re-projection error in pixels, and the targeted homography matrix has 8 degrees of freedom.

A subset of CityFlow, namely the CityFlow-reID, was made available for the Track 2 challenge. CityFlow-reID contains 56,277 bounding boxes, where 36,935 of them from 333 vehicle identities form the training set, while the test set consists of 18,290 bounding boxes from the other 333 identities. The rest of the 1,052 images form the queries. Each vehicle is on average captured by 84.50 image signatures from 4.55 camera views. We also provided for each team a Python visualization tool that can facilitate qualitative evaluation of ReID performance; Fig 4 provides a sample screen shot.

3.2. Iowa DOT Traffic Dataset

More than 50 hours of 800×410 resolution data at 30 FPS captured by the Iowa Department of Transportation

Figure 5. Examples of detected anomalies on the Iowa DOT Traffic Dataset. Original video frames are shown on the left, and processed videos frames with detected anomalies from [39] are shown on the right.

(DOT) traffic cameras were used for the Track 3 challenge. Traffic incidences were captured in both freeways and arterial roads. The training set consists of 21 anomalies across 100 video clips, with the starting time and duration of the anomalies provided. Fig. 5 shows examples of anomaly incidences in the Iowa DOT Traffic Dataset.

4. Evaluation Methodology

Similar to the 2018 AI City Challenge [26], we encouraged teams to improve the models they learned by allowing submissions of multiple runs for each track to an **online evaluation system** that automatically measured the effectiveness of results upon submission. Teams were allowed a maximum of 5 submissions per day and a maximum number of submissions for each track (20 for Tracks 1 and 2, and 10 for Track 3). The system returned an error if the submission format was incorrect or an error was encountered during evaluation — such submissions were not counted against the team's daily or maximum submission totals. To further encourage competition between the teams, this year our evaluation system showed not only their own performance, but also the top-3 best scores on the leader board (without revealing which teams obtained those scores). To discourage excessive fine-tuning to improve performance, the results shown to the teams prior to the end of the challenge were computed on a 50% subset of the test set for each track. After the challenge submission deadline, the evaluation system revealed the full leader board with scores computed on the entire test set for each track.

4.1. Track 1 Evaluation

The primary task of Track 1 was identifying and tracking vehicles that can be seen in at least two of the 40 cameras in the CityFlow dataset. As such, we adopted the IDF1 score [29] from the MOTChallenge [4, 21] to rank the performance of each team. IDF1 measures the ratio of correctly identified detections over the average number of ground-truth and computed detections. In the multi-camera setting, the score is computed in a video made up of the concatenated videos from all cameras. The ground truth consists of the bounding boxes of multi-camera vehicles labeled with a consistent global ID. Our evaluation system automatically removes identities that appear only within a single camera. A high IDF1 score is obtained when the correct multi-camera vehicles were discovered, accurately tracked within each video, and labeled with a consistent ID across all videos in the dataset. For each submission, the evaluation server also computes several other performance measures, including ID match precision (IDP), ID match recall (IDR), and detection precision and recall. While these scores were shared with the teams for their own submissions, they were not used in the overall team ranking and were not displayed in the leader board.

4.2. Track 2 Evaluation

Performance evaluation in Track 2 is based on the ability of a model to find the correct matches in the test set for each of the query vehicles. Given the large size of CityFlow-reID, we propose the rank-K mAP measure, which computes the mean of the average precision (the area under the Precision-Recall curve) over all the queries when considering only the top-K results for each query. We requested up to $K = 100$ results for each query. In addition to the rank-K mAP results, our evaluation server also computes the rank-K Cumulative Matching Characteristics (CMC) scores for $K \in \{1, 5, 10, 15, 20, 30, 100\}$, which are popular metrics for person ReID evaluation. Similar to Track 1, while these scores were shared with the teams for their own submissions, they were not used in the overall team ranking and were not displayed in the leader board.

4.3. Track 3 Evaluation

Performance evaluation in Track 3 is 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 3 score (S_3), for each participating team, is computed as

$$S_3 = F_1 \times (1 - NRMSE^t), \tag{1}$$

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 predictions 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 only consider the one with minimum detection time error and ignore the rest. We expect all anomalies to be successfully detected and penalize missed detection and spurious ones through the F_1 component in the S_3 evaluation score.

Table 1. Summary of Track 1 leader board.

Rank	Team ID	Team name (and paper)	IDF1
1	21	U. Washington IPL [12]	0.7059
2	49	DiDi Global [20]	0.6865
3	12	BUPT Traffic Brain [10]	0.6653
5	97	Australian National U. [11]	0.6519
6	59	Baidu ZeroOne [33]	0.5987
9	104	Shanghai Tech. U. [40]	0.3369
10	52	CUNY NPU [9]	0.2850
17	79	NCCU-UAlbany [7]	0.1634
18	64	GRAPH@FIT [32]	0.0664
19	43	U. Autonoma de Madrid [23]	0.0566

Another primary component of the score in Track 3 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 between the ground-truth anomaly start time and predicted start time for all true positives. To obtain a normalized evaluation score, we calculate $NRMSE^t$ as the normalized detection time RMSE using min-max normalization between 0 and 300 frames (for videos of 30 FPS, this corresponds to 10 seconds), which represents a reasonable range of RMSE values for the anomaly detection task. Specifically, $NRMSE^t$ of team i is computed as

$$NRMSE^t_i = \frac{\min(RMSE_i, 300)}{300}. \tag{2}$$

5. Challenge Results

Tables 1, 2, and 3 summarize the leader boards for Track 1 (city-scale multi-camera vehicle tracking), Track 2 (city-scale multi-camera vehicle ReID), and Track 3 (traffic anomaly detection) challenges, respectively.

5.1. Summary for Track 1 Challenge

The best performing team (Team 21 U. Washington IPL [12]) combined single-camera tracking, deep-learning-based ReID and camera link models for inter-camera tracking into one system. In addition, effective spatio-temporal association was proven useful in improving tracking performance, as in the case of the second and third best performing teams (Team 49 DiDi Global [20] and Team 12 BUPT

Table 2. Summary of Track 2 leader board.

Rank	Team ID	Team name (and paper)	rank-K mAP
1	59	Baidu ZeroOne [33]	0.8554
2	21	U. Washington IPL [14]	0.7917
3	97	Australian National U. [24]	0.7589
4	4	U. Tech. Sydney [42]	0.7560
5	12	BUPT Traffic Brain [10]	0.7302
8	5	U. Maryland RC [16]	0.6078
13	27	INRIA STARS [8]	0.5344
18	24	National Taiwan U. [22]	0.4998
19	40	Huawei AI Brandits [2]	0.4631
23	52	CUNY-NPU [9]	0.4096
25	113	VNU HCMUS [27]	0.4008
36	26	SYSU ISENET [13]	0.3503
45	64	GRAPH@FIT [32]	0.3157
50	79	NCCU-UAlbany [7]	0.2965
51	63	Queen Mary U. London [15]	0.2928
54	46	Siemens Bangalore [17]	0.2766
60	43	U. Autonoma de Madrid [23]	0.2505

Traffic Brain [10], respectively). Also, the data association graph has been a reliable model for this task (Team 97 Australian National U. [11] ranked 5th, Team 104 Shanghai Tech. U. [40] ranked 9th, and Team 52 CUNY NPU [9] ranked 10th). However, the majority of the MTMC tracking methods followed the tracking-by-detection paradigm as in the previous years, with adaptations to multi-camera views.

5.2. Summary for Track 2 Challenge

The best performing team was Team 59 Baidu ZeroOne [33], using a method based on extracting visual features from convolutional neural networks (CNNs), and leveraging semantic features from traveling direction and vehicle type classification. The utilization of vehicles semantic attributes has also been exploited by Team 21 U. Washington IPL [14], which was the runner-up in this track. Though such attributes are highly useful for ReID, they require additional annotation and thus were not widely used by other teams. In addition, some leading teams also used pre-trained models to extract vehicle pose, which can infer the orientation information [14, 16]. Since the trajectory-level information was provided in CityFlow-reID, several teams leveraged temporal attention/pooling in their methods [33, 14, 24]. We also noticed that many teams [14, 16, 13, 17] post-processed their results by re-ranking to further improve the ReID performance. Moreover, teams were allowed to incorporate additional data from other public benchmarks in training [42, 10, 16, 22], but the domain gap between various datasets cannot be neglected, which was addressed by Team 24 National Taiwan U. [22]. Finally, most of the other methods [8, 2, 9, 7, 15, 17, 23] relied on feature embedding schemes and distance metric learning to push the limits of appearance-based ReID.

Table 3. Summary of Track 3 leader board.

Rank	Team ID	Team name (and paper)	S_3
1	12	BUPT Traffic Brain [3]	0.9534
2	21	U. Washington IPL [39]	0.9362
6	79	NCCU-UAlbany [7]	0.6997
7	48	BUPT MCPRL [41]	0.6585
8	113	VNU HCMUS [27]	0.6129
12	65	Trivandrum CET CV [31]	0.3636
16	61	MNIT Vision Intelligence [6]	0.2641
17	5	U. Maryland RC [16]	0.2207

5.3. Summary for Track 3 Challenge

Most methods for Track 3 used foreground segmentation to reduce the search (Teams 12 BUPT Traffic Brain [3], Team 79 NCCU-UALbany [7], and Team 48 BUPT MCPRL [41]). The best performance was achieved by Team 12 BUPT Traffic Brain [3] using a spatio-temporal anomaly matrix. The runner-up (Team 21 U. Washington IPL [39]), on the other hand, proposed a novel two-stage framework based on anomaly candidate identification and starting time estimation. The solution of the third place team (Team 79 NCCU-UALbany [7]) was based on the second-place winning method from the 2018 AI City Challenge [26], with refined event recognition of stalled vehicles and back-tracking to accurately locate event occurrence.

6. Discussion

The success of the 2017 and 2018 AI City Challenges made us confident that many of the state-of-the-art technologies developed in the computer vision community over the years are now ready for deployment to solve real-world traffic and public safety problems. The third edition of the AI City Challenge has attracted five-fold the number of teams and growing attention from the research community. This year's AI City Challenge strives to bring the computer vision community closer to real-world stakeholders, focusing on more challenging city-scale multi-camera tracking of vehicles, vehicle ReID and anomaly detection. Based on this year's team submissions, we would like to highlight several observations.

Compared to previous versions of the multi-camera vehicle tracking challenge, this year we included multiple camera views that do not have a large overlap. The camera views came from sites that physically span multiple intersections. The introduction of such city-scale camera network further complicates MTMC tracking, a problem that has already been difficult due to the large number of existing vehicle models, occlusion, and changes in lighting conditions. We started to see the problems of MTMC tracking and vehicle ReID being jointly solved with appropriate visual features and spatio-temporal association to achieve high performance. Indeed, the best performing team this year (Team 21, U. Washington IPL [12]) combines single-camera tracking, CNN features, and camera link models in their solution. However, most of the algorithms are still far from being able to be used in real-time processing.

Track 2 vehicle ReID is a new task for AI City Challenge this year. Person-based ReID has been developing rapidly in recent years and advanced the research in distance metric learning. However, vehicle-based problems are more challenging, due to the high intra-class variability caused by the dependence of shapes on viewing angles, and high inter-class similarity, as vehicle models produced by different manufacturers look visually alike. Since viewpoint and orientation information is crucial for feature extraction and comparison, the two leading teams (Team 59 Baidu ZeroOne [33] and Team 21 U. Washington IPL [14]) respectively introduced direction classification and vehicle orientation feature descriptors (extracted from keypoints) in their formulation. Other popular strategies adopted by leading teams are trajectory-level association, post-processing by re-ranking, and inclusion of training data from other benchmarks. To handle the domain gap across datasets, Team 24 National Taiwan U. [22] proposed supervised joint domain learning, which can handle misaligned feature distribution between domains. We believe such techniques will become a trend, as the availability of vehicle data scales rapidly nowadays. The generation of synthetic data is getting mature, which may benefit ReID on real data if their domain gap can be minimized. Last but not least, the room for improvement on ReID based on appearance features only is limited, because of the inter-class similarity of vehicle models. That is why the problem of Track 1, which provides spatio-temporal cues, is more realistic.

The third track of the 2019 AI City Challenge is on unconstrained traffic anomaly detection, which continues to be a challenging problem, not only because of confounding factors, such as video quality, illumination and environmental conditions, but more importantly, atypical events are relatively rare and complex in nature. In the past years, this has been an area with less prominent progress in comparison with the other two tracks, but we have seen significant improvement in performance this year. The two best performing teams (Team 12 BUPT Traffic Brain [3] and Team 21 U. Washington IPL [39]) achieve very high scores with a leading margin around $\sim 30\%$ over other teams. Interestingly, most algorithms in this track are not solely based on end-to-end trained deep neural networks. This may be attributed to the fact that there are only few large-scale datasets available for traffic anomaly detection, which are necessary for training deep-learning networks. The availability of a training set in the challenge this year, albeit small, still makes it possible for teams to fine-tune their configuration parameters before experimenting on the test set.

7. Conclusion

The 2019 AI City Challenge is the first city-scale multi-camera challenge in terms of amount of synchronized video data, number of cameras used, geographical area covered and the diversity in camera vantage and orientation. It evaluates participants on extremely challenging tasks of ReID and MTMC tracking, as well as anomaly detection. The challenge has seen significant worldwide participation with 96 teams submitting results to three tracks. The performance of the top teams in each track gains improvements over provided baselines, which shows good progress achieved to solve these complex problems. The challenge and the datasets appear to be achieving their objectives of attracting computer vision and AI researchers to push the boundaries of the state-of-the-art on problems that matter and under real-world conditions.

8. Acknowledgement

The datasets of the 2019 AI City Challenge would not have been possible without significant contributions from an urban traffic agency in the United States and the Iowa Department of Transportation. This challenge was also made possible by significant help from NVIDIA Corporation, which curated the datasets for public dissemination.

References

[1] The AI City Challenge. `https://www.aicitychallenge.org/`.

[2] Abner Ayala-Acevedo, Akash Devgun, Sadri Zahir, and Sid Askary. Vehicle re-identification: Pushing the limits of re-identification. In *Proc. CVPR Workshops*, 2019.

[3] Shuai Bai, Zhiqun He, Yu Lei, Wei Wu, Chengkai Zhu, and Ming Sun. Traffic anomaly detection via perspective map based on spatial-temporal information matrix. In *Proc. CVPR Workshops*, 2019.

[4] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: The clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008(1):246309, May 2008.

[5] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: The CLEAR MOT metrics. *Image and Video Processing*, 2018.

[6] Kuldeep Marotirao Biradar, Ayushi Gupta, Murari Mandal, and Santosh Kumar Vipparthi. Challenges in time-stamp aware anomaly detection in traffic videos. In *Proc. CVPR Workshops*, 2019.

[7] Ming-Ching Chang, Jiayi Wei, Zheng-An Zhu, Yan-Ming Chen, Chan-Shuo Hu, Ming-Xiu Jiang, and Chen-Kuo Chiang. AI City Challenge 2019 – City-scale video analytics for smart transportation. In *Proc. CVPR Workshops*, 2019.

[8] Hao Chen, Benoit Lagadec, and Francois Bremond. Partition and reunion: A two-branch neural network for vehicle re-identification. In *Proc. CVPR Workshops*, 2019.

[9] Yucheng Chen, Longlong Jing, Elahe Vahdani, Ling Zhang, YingLi Tian, and Mingyi He. Multi-camera vehicle tracking and re-identification on AI City Challenge 2019. In *Proc. CVPR Workshops*, 2019.

[10] Zhiqun He, Yu Lei, Shuai Bai, and Wei Wu. Multi-camera vehicle tracking with powerful visual features and spatial-temporal cue. In *Proc. CVPR Workshops*, 2019.

[11] Yunzhong Hou, Heming Du, and Liang Zheng. Think small, deliver big: A locality aware city-scale multi-camera vehicle tracking system. In *Proc. CVPR Workshops*, 2019.

[12] Hung-Min Hsu, Tsung-Wei Huang, Gaoang Wang, Jiarui Cai, Zhichao Lei, and Jenq-Neng Hwang. Multi-camera tracking of vehicles based on deep features Re-ID and trajectory-based camera link modelsodels. In *Proc. CVPR Workshops*, 2019.

[13] Runhui Huang, Peixiang Huang, Jianjie Huang, Rushi Yangchen, and Zongyao He. Deep feature fusion with multiple granularities for vehicle re-identification. In *Proc. CVPR Workshops*, 2019.

[14] Tsung-Wei Huang, Jiarui Cai, Hao Yang, Hung-Min Hsu, and Jenq-Neng Hwang. Multi-view vehicle re-identification using temporal attention model and metadata re-ranking. In *Proc. CVPR Workshops*, 2019.

[15] Aytac Kanaci, Minxian Li, Shaogang Gong, and Georgia Rajamanoharan. Multi-task mutual learning for vehicle re-identification. In *Proc. CVPR Workshops*, 2019.

[16] Pirazh Khorramshahi, Neehar Peri, Amit Kumar, Anshul Shah, and Rama Chellappa. Attention driven vehicle re-identification and unsupervised anomaly detection for traffic understanding. In *Proc. CVPR Workshops*, 2019.

[17] Varghese Alex Kollerathu, Adithya Shankar, Sheetal Reddy, Chandan Yeshwanth, Akhil Purshottam, and Vinay Sudhakaran. Comparative study on various losses for vehicle re-identification. In *Proc. CVPR Workshops*, 2019.

[18] Ratnesh Kumar, Edwin Weill, Farzin Aghdasi, and Parthsarathy Sriram. Vehicle re-identification: An efficient baseline using triplet embedding. In *Proc. IJCNN*, 2019.

[19] Young-Gun Lee, Zheng Tang, and Jenq-Neng Hwang. Online-learning-based human tracking across non-overlapping cameras. *T-CSVT*, 28(10):2870–2883, 2018.

[20] Peilun Li, Guozhen Li, Zhangxi Yan, Youzeng Li, Meiqi Lu, Pengfei Xu, Yang Gu, and Bing Bai. Spatio-temporal consistency and hierarchical matching for multi-target multi-camera vehicle tracking. In *Proc. CVPR Workshops*, 2019.

[21] Yuan Li, Chang Huang, and Ram Nevatia. Learning to associate: Hybrid boosted multi-target tracker for crowded scene. In *Proc. CVPR*, pages 2953–2960, 2009.

[22] Chih-Ting Liu, Man-Yu Lee, Chih-Wei Wu, Yao Ting Hsu, Tsai-Shien Chen, Bo-Ying Chen, and Shao-Yi Chien. Supervised joint domain learning for vehicle re-identification. In *Proc. CVPR Workshops*, 2019.

[23] Elena Luna, Juan Carlos San Miguel, and José M. Martínez. Clustering-based multi target multi camera tracking. In *Proc. CVPR Workshops*, 2019.

[24] Kai Lv, Weijian Deng, Yunzhong Hou, Heming Du, Hao Sheng, Jianbin Jiao, and Liang Zheng. Vehicle re-identification with the location and time stamp. In *Proc. CVPR Workshops*, 2019.

[25] Milind Naphade, David C. Anastasiu, Anuj Sharma, Vamsi Jagrlamudi, Hyeran Jeon, Kaikai Liu, Ming-Ching Chang, Siwei Lyu, and Zeyu 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.

[26] Milind Naphade, Ming-Ching Chang, Anuj Sharma, David C. Anastasiu, Vamsi Jagarlamudi, Pranamesh Chakraborty, Tingting Huang, Shuo Wang, Ming-Yu Liu, Rama Chellappa, Jenq-Neng Hwang, and Siwei Lyu. The 2018 NVIDIA AI City Challenge. In *Proc. CVPR Workshops*, pages 53–60, 2018.

[27] Khac-Tuan Nguyen, Trung-Hieu Hoang, Trung-Nghia Le, Minh-Triet Tran, Ngoc-Minh Bui, Trong-Le Do, Viet-Khoa Vo-Ho, Quoc-An Luong, Mai-Khiem Tran, Thanh-An Nguyen, Thanh-Dat Truong, Vinh-Tiep Nguyen, and Minh Do. Vehicle re-identification with learned representation and spatial verification and abnormality detection with multi-adaptive vehicle detectors for traffic video analysis. In *Proc. CVPR Workshops*, 2019.

[28] Joseph Redmon and Ali Farhadi. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.

[29] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *Proc. ECCV*, pages 17–35, 2016.

[30] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 115(3):211–252, 2015.

[31] Linu Shine, Anitha Edison, and Jiji C.V. A comparative study of Faster R-CNN models for anomaly detection in 2019 AI City Challenge. In *Proc. CVPR Workshops*, 2019.

[32] Jakub Špaňhel, Vojtěch Bartl, Roman Juránek, and Adam Herout. Vehicle re-identifiation and multi-camera tracking in challenging city-scale environment. In *Proc. CVPR Workshops*, 2019.

[33] Xiao Tan, Zhigang Wang, Minyue Jiang, Xipeng Yang, Jian Wang, Yuan Gao, Xiangbo Su, Xiaoqing Ye, Yuchen Yuan, Dongliang He, Shilei Wen, and Errui Ding. Multi-camera vehicle tracking and re-identification based on visual and spatial-temporal features. In *Proc. CVPR Workshops*, 2019.

[34] Zheng Tang and Jenq-Neng Hwang. MOANA: An online learned adaptive appearance model for robust multiple object tracking in 3D. *IEEE Access*, 7(1):31934–31945, 2019.

[35] Zheng Tang, Yen-Shuo Lin, Kuan-Hui Lee, Jenq-Neng Hwang, and Jen-Hui Chuang. ESTHER: Joint camera self-calibration and automatic radial distortion correction from tracking of walking humans. *IEEE Access*, 7(1):10754–10766, 2019.

[36] Zheng Tang, Yen-Shuo Lin, Kuan-Hui Lee, Jenq-Neng Hwang, Jen-Hui Chuang, and Zhijun Fang. Camera self-calibration from tracking of moving persons. In *Proc. ICPR*, pages 260–265, 2016.

[37] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David C. Anastasiu, and Jenq-Neng Hwang. CityFlow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. 2019.

[38] Zheng Tang, Gaoang Wang, Hao Xiao, Aotian Zheng, and Jenq-Neng Hwang. Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features. In *Proc. CVPR Workshops*, pages 108–115, 2018.

[39] Gaoang Wang, Xinyu Yuan, Aotian Zhang, Hung-Min Hsu, and Jenq-Neng Hwang. Anomaly candidate identification and starting time estimation of vehicles from traffic videos. In *Proc. CVPR Workshops*, 2019.

[40] Minye Wu, Guli Zhang, Ning Bi, Ling Xie, Yuanquan Hu, Shenghua Gao, and Zhiru Shi. Multiview vehicle tracking by graph matching model. In *Proc. CVPR Workshops*, 2019.

[41] Jianfei Zhao, Zitong Yi, Siyang Pan, Yanyun Zhao, and Bojin Zhuang. Unsupervised traffic anomaly detection using trajectories. In *Proc. CVPR Workshops*, 2019.

[42] Zhedong Zheng, Tao Ruan, Yunchao Wei, and Yi Yang. VehicleNet: Learning robust feature representation for vehicle re-identification. In *Proc. CVPR Workshops*, 2019.

[43] Kaiyang Zhou. deep-person-reid. https://github.com/KaiyangZhou/deep-person-reid.