The NVIDIA AI City Challenge

Milind Naphade [1], David C. Anastasiu [2], Anuj Sharma [3], Vamsi Jagrlamudi [3] Hyeran Jeon [2], Kaikai Liu [2],
Ming-Ching Chang [4], Siwei Lyu [4], and Zeyu Gao [2]

[1]NVIDIA Coorp., Santa Clara, CA
[2]San José State University, San José, CA
[3]Iowa State University, Ames, IA
[4]University at Albany, State University of New York, NY

***Abstract*—Web image analysis has witnessed an AI renaissance. The ILSVRC benchmark has been instrumental in providing a corpus and standardized evaluation. The NVIDIA AI City Challenge is envisioned to provide similar impetus to the analysis of image and video data that helps make cities smarter and safer. In its first year, this Challenge has focused on traffic video data. While millions of traffic video cameras around the world capture data, albeit low-quality, very little automated analysis and value creation results. Lack of labeled data, and trained models that can be deployed at the edge of the city fabric, ensure that most traffic video data goes through little or no automated analysis. Real-time and batch analysis of this data can provide vital breakthroughs in real-time traffic management as well as pedestrian safety. The NVIDIA AI City Challenge brought together 29 teams from universities in 4 continents to collaboratively annotate a 125 hour data set and then compete on detection, localization and classification tasks as well as traffic and safety application analytics tasks. The result is the largest high quality annotated data set, a set of models trained using NVIDIA AI City Edge to Cloud platform and ready to be deployed at the edge solving traffic and safety problems for cities worldwide.**

***Index Terms*—Deep Learning, AI, traffic flow, pedestrian safety, video analysis, edge computing, cloud computing, GPU, mean Average Precision, Intersection over Union**

I. Introduction

Deep Learning has led an AI renaissance of sorts in recent years. Image and video analysis are among its most prominent success stories. Results of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [1] point to a dramatic improvement in object detection, localization and classification. This breakthrough has impacted many verticals from self driving vehicles in the Transportation sector to medical image analysis in the Healthcare sector. This breakthrough also has great potential for making our cities smarter [2] and safer. While there are existing corpora and benchmarks for video retrieval (*e.g.* NIST TRECVID [3], LSCOM [4]) and image classification [1], there is lack of a large scale labeled corpus of high quality traffic video data. This is also compounded by the lack of an AI platform that allows for rapid edge to cloud experimentation and deployment and a standardized evaluation of algorithm performance. To address this gap and accelerate the progress of deep learning in making cities smarter and safer we envisioned and created the NVIDIA AI City Challenge `smart-city-conference.com/AICityChallenge/` [5]. Figure 1 illustrates the life cycle of the Challenge.

Preparation for the Challenge began in May 2017 and the Challenge ended with a hackathon and workshop on Aug 5 2017 as part of the 3rd annual IEEE Smart World Congress.

We started by capturing and creating the largest video corpus of traffic video data that included high quality 1080p data by mounting new traffic cameras as well as commonly available 480p data from existing traffic cameras. Twenty-nine teams spanning four continents signed up for the Challenge and helped collaboratively annotate 125 hours of data captured at 30 frames per second. For this we extended the VIA (VGG Image Annotation Tool) [6]. 150 participants used this modified Collaborative Annotation Tool to label the data using 15 class labels identified in consultation with multiple departments of transportation in the United States. Ten hours of video data labeled with a subset of these labels, recorded in 24 Chinese cities, which was part of the UA-DETRAC benchmark [7], was also made available to participants.

After the annotation phase of the Challenge, preliminary evaluation of annotation quality was conducted and 18 teams were selected to compete in the next round of the Challenge. The selection was made based on the quality and quantity of their annotation effort and the quality of their proposal. Participants then competed in two tracks. Track 1 focused on object detection, localization and classification and used common metrics for evaluation. Track 2 was open ended and allowed participants to use Track 1 results and any other algorithms to provide solutions to common traffic flow and pedestrian safety problems that cities face.

NVIDIA provided participating teams with an edge to cloud AI infrastructure and platform. Teams used DGX servers which are equipped with 8 Tesla P100 GPUs [8] for training and Jetson TX2 [9], the fastest supercomputing edge compute device, for inferencing. The teams were also provided labeled data sets in multiple formats for the commonly used frameworks. NVIDIA provided participants with containerized frameworks including Caffe [10], Darknet [11], Tensorflow [12], MXNet [13], and Torch [14] for training various networks. Participants experimented with a variety of models by using transfer learning and extending existing models like Faster R-CNN with ResNet, SSD, YOLO9000, R-FCN, Deformable Convnets, and DeepHOG. The goals were to reduce the barrier to experimentation and time to modeling for participating teams.

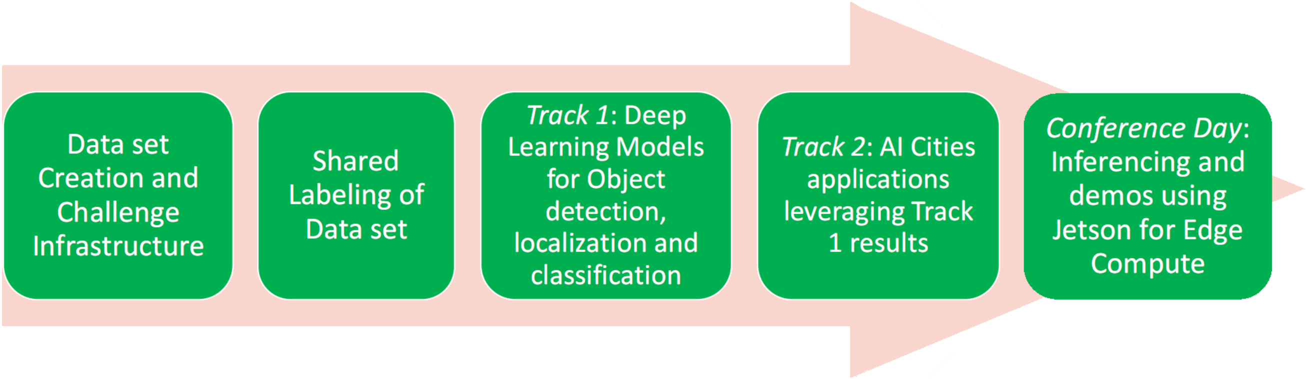

Data set Creation and Challenge Infrastructure
Shared Labeling of Data set
Track 1: Deep Learning Models for Object detection, localization and classification
Track 2: AI Cities applications leveraging Track 1 results
Conference Day: Inferencing and demos using Jetson for Edge Compute

Fig. 1: Overall structure and flow of the NVIDIA AI City Challenge.

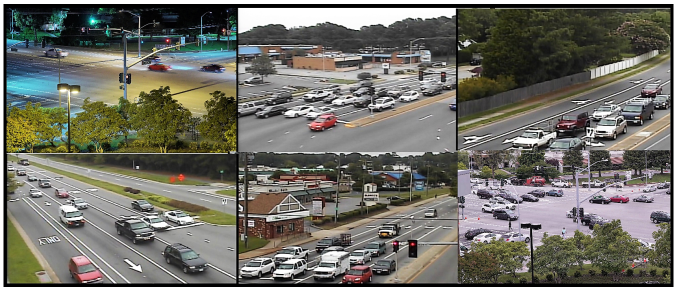


Fig. 2: A sample of images from the data captured at traffic intersections in multiple cities and states.

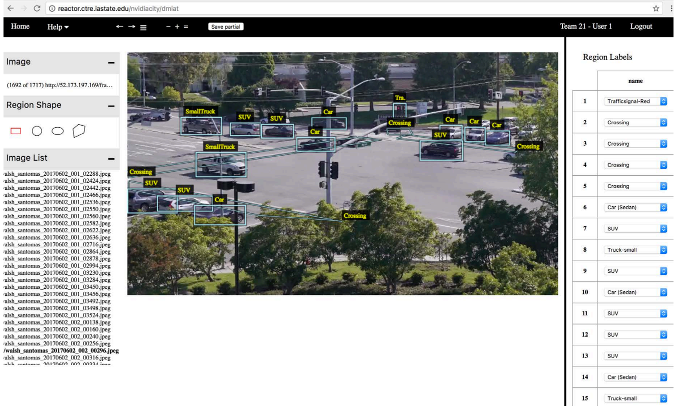
The winner of Track 1 was determined based on a composite score that combined mean Average Precision (mAP) for object detection and classification across 3 data sets that were created in the Challenge along with the localization accuracy as captured by the Intersection over Union (IoU) score for these data sets. The winner of Track 2 was determined by a panel of judges that included domain experts from NIST, GE Current, and NVIDIA. The criteria for evaluation included novelty, value, and demonstration of the innovation.

Based on the unprecedented success of this Challenge, discussions are underway to make this a recurring challenge, allow greater participation and progressively increase task diversity and complexity. These breakthroughs have a great potential for making our cities smarter and safer.

II. Dataset

Video data available for this Challenge has been recorded by cameras aimed at intersections in urban areas. Videos were recorded in diverse conditions, including daytime and nighttime conditions. The NVIDIA AI City Data Set consists of the following video data sources:

1) Silicon Valley Intersection Data - More than 70 hours of 1080p data at 30 frames per second captured from multiple vantage points.
2) Virginia Beach Intersection Data - More than 50 hours of 720x480 resolution data at 30 frames per second captured from traffic cameras.
3) Lincoln, Nebraska Data - More than 10 hours of 720x480 resolution data at 30 frames per second captured from handheld cameras.



(1692 of 1717) http://52.173.197.169/fra…

/alsh_santomas_20170602_001_02288.jpeg
/alsh_santomas_20170602_001_02424.jpeg
/alsh_santomas_20170602_001_02442.jpeg
/alsh_santomas_20170602_001_02466.jpeg
/alsh_santomas_20170602_001_02536.jpeg
/alsh_santomas_20170602_001_02550.jpeg
/alsh_santomas_20170602_001_02560.jpeg
/alsh_santomas_20170602_001_02582.jpeg
/alsh_santomas_20170602_001_02622.jpeg
/alsh_santomas_20170602_001_02636.jpeg
/alsh_santomas_20170602_001_02716.jpeg
/alsh_santomas_20170602_001_02864.jpeg
/alsh_santomas_20170602_001_02878.jpeg
/alsh_santomas_20170602_001_02994.jpeg
/alsh_santomas_20170602_001_03230.jpeg
/alsh_santomas_20170602_001_03284.jpeg
/alsh_santomas_20170602_001_03450.jpeg
/alsh_santomas_20170602_001_03456.jpeg
/alsh_santomas_20170602_001_03492.jpeg
/alsh_santomas_20170602_001_03498.jpeg
/alsh_santomas_20170602_001_03524.jpeg
/alsh_santomas_20170602_002_00138.jpeg
/alsh_santomas_20170602_002_00160.jpeg
/alsh_santomas_20170602_002_00240.jpeg
/alsh_santomas_20170602_002_00256.jpeg
/walsh_santomas_20170602_002_00296.jpeg
/alsh_santomas_20170602_002_00316.jpeg

	name
1	Trafficsignal-Red
2	Crossing
3	Crossing
4	Crossing
5	Crossing
6	Car (Sedan)
7	SUV
8	Truck-small
9	SUV
10	Car (Sedan)
11	SUV
12	SUV
13	SUV
14	Car (Sedan)
15	Truck-small

Fig. 3: Image annotation tool extended from the VIA tool to provide collaborative browser based annotation capabilities.

In addition to this data, we also provided participants the option to use the labeled data set available from the SUNY Albany UA-DETRAC benchmark suite `http://detrac-db.rit.albany.edu/` [7].

A. Annotation Tool

The annotation tool [15] we developed for the AI City Challenge was extended from the VGG Image Annotator (VIA) [6], which is a single page application that provides the capability for browser based annotations. Over 250,000 keyframes were extracted from a subset of the 120 hours of video at 1 second intervals. The keyframes were shuffled randomly and an equal number of frames was assigned to each of the 29 teams. Keyframe shuffling mitigated the possibility of some videos having no annotations due to low team performance or bad annotation quality. Our annotation tool was designed to support collaborative annotation of keyframes. After users log into the application, they are presented with their assigned set of keyframes. For each keyframe, the application saves annotation information in a database in JSON format. Our tool was developed using the Express JS framework and MongoDB as the database server, and was deployed in the Microsoft Azure Cloud Services [16] computing infrastructure.

B. Class Labels

Based on the inputs from the NYC, Iowa State, Nebraska Departments of Transportation (DOT), a list 15 class labels were identified. These classes were found relevant for

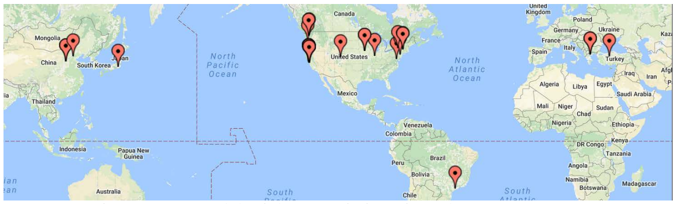


Fig. 4: Worldwide participation by 29 teams

transportation planning and operations applications. The annotators were asked to draw bounding boxes around and label objects in the identified classes, namely, *Car, SUV, Bus, Van, SmallTruck, MediumTruck, LargeTruck, Bicycle, Motorcycle, Pedestrian, GroupOfPeople, TrafficSignal-Green, TrafficSignal-Yellow, TrafficSignal-Red, and Crossing*.

An example image was provided for each class type to the annotators along with a definition. Despite these attempts, a significant amount of erroneous annotations were found for some class types, such as SmallTruck, MediumTruck, LargeTruck, Van, GroupOfPeople,, TrafficSignal-Green, TrafficSignal-Yellow, and TrafficSignal-Red. Some annotators seemed confused about the difference between tracks of different sizes. Moreover, they were not always sure when to annotate people individually and when to draw a bounding box covering all pedestrians in a frame and call them GroupOfPeople. In the case of traffic signals, certain camera angles made it impossible to differentiate colors and hence lead to confusion about the traffic signal label.

Annotators could use rectangles, ellipses, circles, and polygons to describe an object. Collaboratively, the teams contributed over 1.4M annotations in more than 153,000 keyframes. Some keyframes and annotations were removed following a quality review process. Moreover, since many of the videos were recorded at odd angles (not parallel to the road) and most popular frameworks expect rectangular bounding boxes, the "Crossing" objects lead to bounding boxes that covered many other objects and were removed from this year's dataset.

C. Participating Teams

The Challenge attracted 29 teams worldwide. Figure 4 shows the distribution of the teams. After the annotation phase, based on their written proposals and annotation effectiveness, 18 teams from 16 universities made it to the final round of the Challenge. Table I shows these institutions and IDs of teams they participated in.

TABLE I: Teams in the final Challenge round.

Institution	**Advisor**	**Team**
San José State University, USA	Liu Jeon Anastasiu	1, 25 14 21
CERTH, Greece	Kompatsiaris Tzovaras	2 3
University of Washington, USA Beijing University of Posts and Telecom., China University of Tokyo, Japan Microsoft Research, USA	Huang	4
SUNY Albany, USA GE Global Research, USA Univ. of Chinese Academy of Sciences, China	Chang Lyu	5
Iowa State University, USA	Sharma	6
Syracuse University, USA	Ozcan	7
Taiyuan University of Technology, China	Xu Zehua, Xiaofeng	10 23
University of São Paulo, Brazil	Okamoto	13
Lehigh University, USA	Chuah, Wang	16
IBM, USA UC Berkeley, USA	DesAulniers DeJana Grembek	17 17 18, 19
Southeast University, China		18
George Washington University, USA	Frick	19
University of Illinois Urbana-Champaign, USA	Shi	24

D. Resultant Corpus of Labeled Data

Thirty volunteers were used to judge the quality of annotated keyframes in an effort to clean up data before providing it to teams for training models and building smart city applications. Volunteers provided binary judgments for a random sample consisting of 1% of the keyframes assigned to each team. The annotation effort of each user in each team was checked by two volunteers and their average score was used to both identify bad quality keyframes and to choose the teams that should compete in the next phase of the competition. Based on these results, we eliminated all annotations from one user and class-specific annotations from several more.

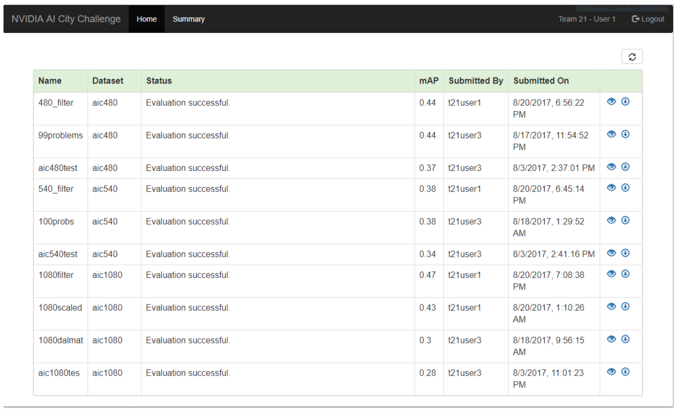
After cleaning, the annotation data was processed into three AI City (AIC) datasets, namely aic480, aic1080, and aic540. The aic480 dataset contains all videos and associated keyframes of size 720x480. Similarly, the aic1080 dataset contains all videos and associated keyframes of size 1920x1080. The aic540 dataset is a down-sampled version of the aic1080 dataset.

Each dataset was split into three sections (train, val, test). While teams were able to use the training and validation sets for several weeks before the Challenge workshop, test videos and associated extracted keyframes were only made available to teams 3 days before the submission deadline.

Each of the three AIC datasets was processed into three derived popular formats (KITTI, Pascal VOC, and DarkNet) and scripts were made available that would allow teams to create their own train/test/val split given a set of annotated keyframes in AIC format.

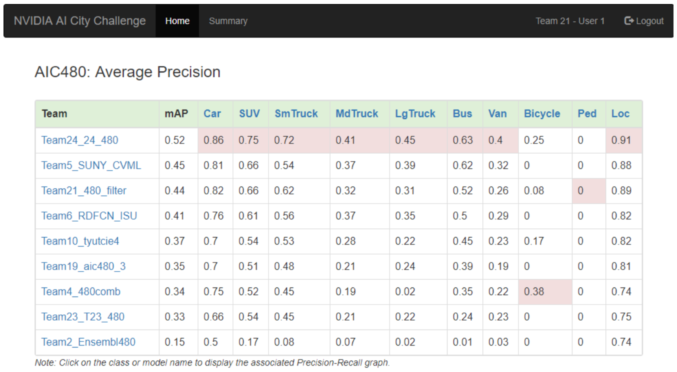
E. Track 1 Evaluation System

Track 1 results were evaluated as a function of the mean of per-class Average Precision (mAP) scores. Teams were provided with an evaluation script that computed the mAP score for a set of keyframes given their true and predicted bounding boxes and associated class designations and confidence levels.



Name	Dataset	Status	mAP	Submitted By	Submitted On
480_filter	aic480	Evaluation successful.	0.44	t21user1	8/20/2017, 6:56:22 PM
99problems	aic480	Evaluation successful.	0.44	t21user3	8/17/2017, 11:54:52 PM
aic480test	aic480	Evaluation successful.	0.37	t21user3	8/3/2017, 2:37:01 PM
540_filter	aic540	Evaluation successful.	0.38	t21user1	8/20/2017, 6:45:14 PM
100probs	aic540	Evaluation successful.	0.38	t21user3	8/18/2017, 1:29:52 AM
aic540test	aic540	Evaluation successful.	0.34	t21user3	8/3/2017, 2:41:16 PM
1080filter	aic1080	Evaluation successful.	0.47	t21user1	8/20/2017, 7:08:38 PM
1080scaled	aic1080	Evaluation successful.	0.43	t21user1	8/20/2017, 1:10:26 AM
1080dalmat	aic1080	Evaluation successful.	0.3	t21user3	8/18/2017, 9:56:15 AM
aic1080tes	aic1080	Evaluation successful.	0.28	t21user3	8/3/2017, 11:01:23 PM

Fig. 5: Automatic evaluation of Track 1 submissions for Team 21.



AIC480: Average Precision

Team	mAP	Car	SUV	SmTruck	MdTruck	LgTruck	Bus	Van	Bicycle	Ped	Loc
Team24_24_480	0.52	0.86	0.75	0.72	0.41	0.45	0.63	0.4	0.25	0	0.91
Team5_SUNY_CVML	0.45	0.81	0.66	0.54	0.37	0.39	0.62	0.32	0	0	0.88
Team21_480_filter	0.44	0.82	0.66	0.62	0.32	0.31	0.52	0.26	0.08	0	0.89
Team6_RDFCN_ISU	0.41	0.76	0.61	0.56	0.37	0.35	0.5	0.29	0	0	0.82
Team10_tyutcie4	0.37	0.7	0.54	0.53	0.28	0.22	0.45	0.23	0.17	0	0.82
Team19_aic480_3	0.35	0.7	0.51	0.48	0.21	0.24	0.39	0.19	0	0	0.81
Team4_480comb	0.34	0.75	0.52	0.45	0.19	0.02	0.35	0.22	0.38	0	0.74
Team23_T23_480	0.33	0.66	0.54	0.45	0.21	0.22	0.24	0.23	0	0	0.75
Team2_Ensembl480	0.15	0.5	0.17	0.08	0.07	0.02	0.01	0.03	0	0	0.74

Note: Click on the class or model name to display the associated Precision-Recall graph.

Fig. 6: Leaderboard showing best results for the aic480 dataset.

For computing the mAP score, we followed the procedure used in the Pascal VOC [17] challenge. To allow teams the most possible time to improve their results, we developed an online evaluation system that automatically measured the effectiveness of each Track 1 result upon submission and stored results in a database. Figure 5 shows Track 1 submissions for Team 21. The system returned an error if results were not in an acceptable format or other errors were encountered when computing mAP scores. Teams were allowed a maximum of 5 submissions for each dataset. After the Challenge submission deadline, teams could see a leaderboard with the best results from each team, sorted in decreasing mAP order. Figure 6 shows the leaderboard for the aic480 dataset.

III. NVIDIA AI CITY EDGE TO CLOUD PLATFORM

When designing the hardware infrastructure to host the deep learning training of the Challenge, there were several requirements from the organizing committee of IEEE Smart World Congress (SWC) and NVIDIA AI City Challenge.

- The hardware infrastructure should be able to accommodate at least 16 teams.
- Each team should be provided with access to NVIDIA GPUs to accelerate their training of deep neural networks. The compute resources should be evenly distributed to ensure equal access to the same performing hardware.

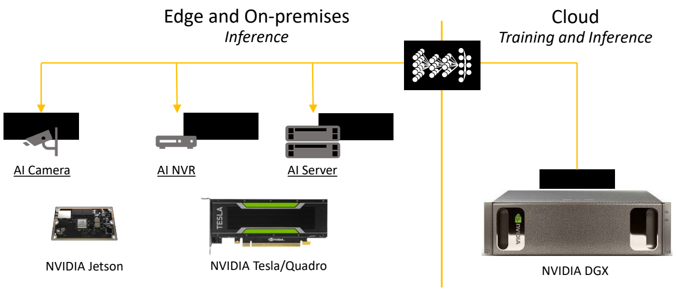


Fig. 7: The NVIDIA AI City Edge to Cloud Platform. Teams used DGX servers for training models and Jetson TX2 for inferencing at the edge.

- The annotated training dataset should be made available (read only) in a shared storage location to avoid the need for team-specific local downloads.
- Additional storage space should be dedicated for each team, to be used for scratch space storage (storing models, code, *etc.*)
- In order to provide an added layer of protection in case of any system failure, backup storage hosted on another physical system should also be provided for each team.
- The provided computing infrastructure should enable teams with all of the commonly used deep learning frameworks. This would allow teams to focus on designing their competition submissions rather than system configuration.
- Each teams resources should be isolated from others to ensure that each teams work is hidden until the end of the competition.

A. *DGX based virtualized environment for training models*

NVIDIA provided teams with access to two DGX-1 deep learning systems (16 P100 GPUs) hosted in its Santa Clara lab. The resources were partitioned among the 16 teams by virtualizing each of the GPUs into a VM using KVM and PCI passthrough. We chose this approach versus utilizing containers directly on the host since using only containers would have allowed users to negatively impact each others work and potentially have root access to the host system. Using VMs provided a layer of isolation between the teams, as teams were only provided "non-`sudo`" access to their assigned VM instead of the underlying DGX-1 system. Each VM used Ubuntu 14.04 as the operating system, and came provisioned with the GPU driver, CUDA toolkit, Docker and nvidia-docker.

To enable teams with deep learning frameworks, we leveraged the DGX-1 nvidia-docker container technology. These containers include fully configured and tested installations of deep learning frameworks that are highly optimized to run on DGX-1. Due to high demand, we also built and made a container for the Darknet framework available in addition to the set of DGX-1 containers. Although some of the teams were new to Docker, these containers simplified the process for teams to get started training their models and made it easier for them to use more than one framework at a time. Adequate storage and backup was provided for teams to run multiple experiment with data augmentation, and train several models using multiple frameworks.

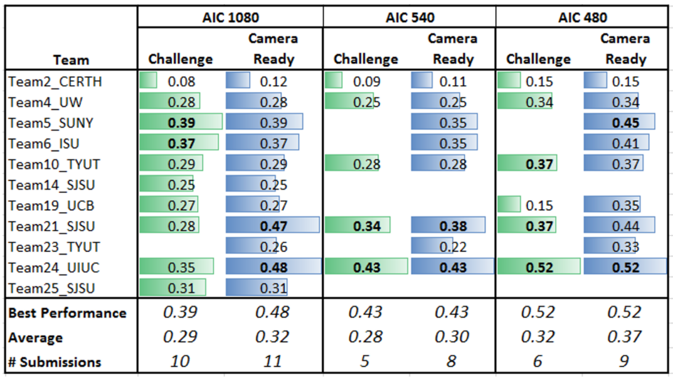
B. *Jetson TX2 for inferencing*

Each team was provided an NVIDIA Jetson TX2 to deploy models they trained on their allocated NVIDIA DGX instance. Jetson TX2 is NVIDIA's second-generation CUDA-capable edge device [9]. Like its predecessor, TX1 [9], TX2 runs Linux using a quad-core ARM CPU. It is equipped with an NVIDIA Pascal GPU [18], which contains specialized architecture for AI applications. TX2 has 8 GB of LPDDR4 RAM and supports PCIe2.0 and various peripherals such as UART, GPIOs, HDMI, USB 3.0 and 2.0, Ethernet, and 802.11ac WLAN. NVIDIA provides the required drivers and CUDA toolkits for TX2 via the JetPack SDK [19]. For this Challenge, all teams installed JetPack 3.1.

At the Challenge workshop, all participating teams demonstrated their designs using a TX2. Various DNN frameworks such as DarkNet, Caffe, and MXNet were installed on TX2. Teams downloaded trained models from their assigned DGX server to the TX2. Traffic object localization and classification was demonstrated via an HDMI monitor connected to the TX2, while running the trained model on the installed DNN frameworks. For testing processing of video inputs, one of the testing video clips was provided to the teams before the start of the workshop. For the teams that used frameworks that did not support video inputs, keyframes were provided for the same test video clip.

IV. CHALLENGE TRACK 1 EXPERIMENTS AND RESULTS

Figure 8 shows Track 1 results at the Challenge stage (August 04, 2017) and at the camera ready stage (August 20, 2017). All teams worked individually till the Challenge deadline. University of Illinois Urbana-Champaign team (UIUC, Team 24) was announced to be the winner of Track 1 Challenge. They had the best performance at the Challenge and also the best performance across all data sets for the camera ready deadline of August 20. They also had overall best performance across all the data sets for the camera ready deadline. The techniques used by each team were presented in the NVIDIA AI City Workshop held on August 05, 2017. At this workshop, the participating teams interacted with each other and exchanged notes to further improve their models. The main goal of the workshop was to catalyze innovation by bringing together multiple interdisciplinary teams to learn from each other and produce results exceeding individual performances. After the cross-pollination of ideas, the teams were given additional 15 days to improve on their work and submit 5 more models by August 20, 2017. In just additional 15 days, we could see the impact of the workshop as more than half the teams improved their scores. As the next stage for continuing innovation, the code developed by each team has been shared on GitHub [20]. The details of techniques used by the top teams are being published in IEEE SWC 2017 conference proceeding papers. Given the benchmark dataset and base models, we hope to see a significant improvement in the state of art for using cameras as traffic sensors.



Team	AIC 1080		AIC 540		AIC 480	
	Challenge	Camera Ready	Challenge	Camera Ready	Challenge	Camera Ready
Team2_CERTH	0.08	0.12	0.09	0.11	0.15	0.15
Team4_UW	0.28	0.28	0.25	0.25	0.34	0.34
Team5_SUNY	**0.39**	0.39		0.35		**0.45**
Team6_ISU	**0.37**	0.37		0.35		0.41
Team10_TYUT	0.29	0.29	0.28	0.28	**0.37**	0.37
Team14_SJSU	0.25	0.25				
Team19_UCB	0.27	0.27			0.15	0.35
Team21_SJSU	0.28	**0.47**	**0.34**	**0.38**	**0.37**	0.44
Team23_TYUT		0.26		0.22		0.33
Team24_UIUC	0.35	**0.48**	**0.43**	**0.43**	**0.52**	**0.52**
Team25_SJSU	0.31	0.31				
Best Performance	*0.39*	*0.48*	*0.43*	*0.43*	*0.52*	*0.52*
Average	*0.29*	*0.32*	*0.28*	*0.30*	*0.32*	*0.37*
# Submissions	*10*	*11*	*5*	*8*	*6*	*9*

Fig. 8: Track 1 teams performances. The top two results for each dataset are highlighted in bold font.

TABLE II: Track 2 teams and topics.

Team	Topic
Team1_SJSU	Video annotation tool, vehicle counting and intersection traffic patterns
Team3_CERTH	Vehicle tracking
Team4_UW	Vehicle tracking, segmentation, counting, re-acquisition with 3D modeling
Team5_SUNY	Vehicle tracking, counting, traffic analysis
Team13_SaoPaulo	Simulation of emergency response time
Team18_UCB	Traffic light timing simulation based on intersection data analysis
Team21_SJSU	Vehicle counting

V. CHALLENGE TRACK 2 EXPERIMENTS AND RESULTS

Track 2 of the NVIDIA AI City Challenge was open ended. Table II indicates how various teams approached Track 2. A number of teams focused on object tracking and vehicle counting. A few teams used the data for simulation purposes coming up with interesting results. To assist teams with Track 2, a baseline tracking algorithm was run [21] and results of the tracking were shared for 5 video clips.

A panel of judges evaluated Track 2 submissions using a combination of innovation novelty, value to real-world problems and demonstration of the system at the Challenge. The three judge panel unanimously chose the submission from Team 4 (University of Washington, Seattle) as the winner of Track 2 for their work on multiple-kernel based vehicle tracking using 3D deformable model and camera self-calibration. Figure 9 shows example results of this method. The judge panel also selected Team 5 (SUNY Albany) for an "honorary mention" award. Figure 10a shows example vehicle tracking results shown from an aerial top-down view presented by Team 5. Traffic analysis can be performed straightforwardly on this representation. Figure 10(b) shows results of vehicle speed and motion type estimation, where the motion type is classified into 4 categories (going straight, left turn, right turn, or stopped).

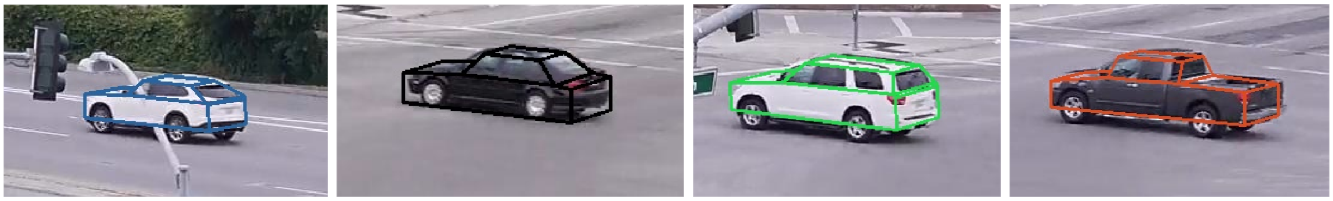


Fig. 9: **Team 4 (University of Washington, Seattle):** Multiple-kernel based vehicle tracking, segmentation and re-acquisition along with 3D model fitting.

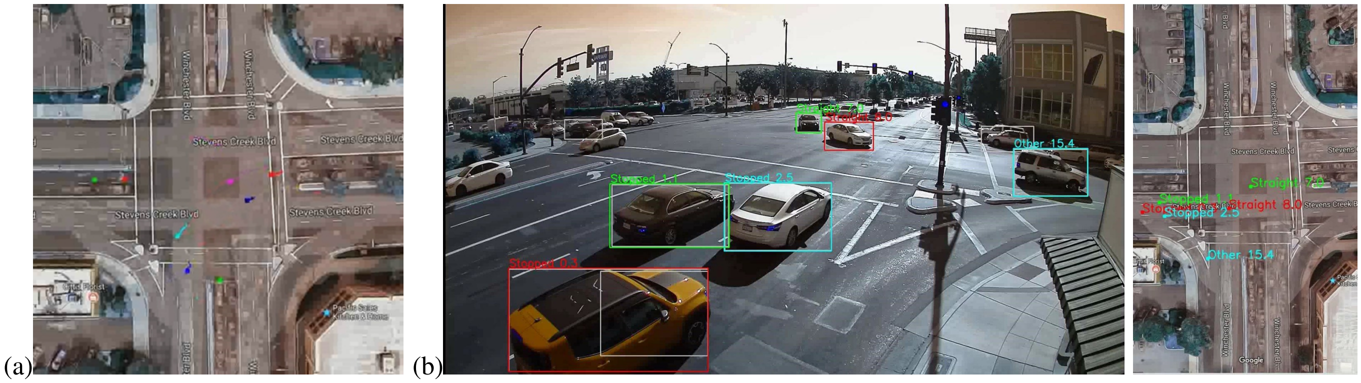


Fig. 10: **Team 5 (SUNY Albany):** Vehicle tracking and traffic analysis results on the Stevens-Winchester-1 video. (a) A top-down visualization of the traffic scenario on Google Map. (b) Visualization of traffic analysis including vehicle speed in MPH and motion status in both the original video and the corresponding top-down views.

VI. Observations and Future Work

There is tremendous potential in recent advances in deep learning to make our cities smarter and safer. The NVIDIA AI City Challenge demonstrated this specifically by bringing together worldwide research teams, an edge to cloud AI platform and a labeled traffic intersection video data set to push the boundaries of automated traffic analysis for traffic flow and pedestrian safety. The huge response in terms of participation and the experiments conducted by the teams within a very short period of time have validated the hypothesis that such a Challenge can dramatically accelerate creation and adoption of valuable AI technology for transportation agencies and also created demand for extending this Challenge globally.

Acknowledgment: A large team at NVIDIA, SJSU and ISU made this Challenge execution possible within such a short period of time. We would like to specifically acknowledge Murali Gopalakrishna, Arman Toorians, Abid Itagi, Douglas Holt, Prethvi Kashinkunti, Michael Knox, Michael Young and Phi Lawrence at NVIDIA. Acknowledgements are also due to the 150 volunteering students from the 29 participating teams for collaboratively annotating the data set. Finally, we thank judges John Garofolo of NIST, Maulin Patel of GE Current and Farzin Aghdasi of NVIDIA.

References

[1] 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," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[2] Milind R. Naphade, Guruduth Banavar, Colin Harrison, Jurij Paraszczak, and Robert Morris, "Smarter cities and their innovation challenges," *IEEE Computer*, vol. 44, pp. 32–39, 06 2011.

[3] George Awad, Asad Butt, Jonathan Fiscus, Martial Michel, David Joy, Wessel Kraaij, Alan F. Smeaton, Georges Qunot, Maria Eskevich, Roeland Ordelman, Gareth J. F. Jones, and Benoit Huet, "TRECVID 2017: Evaluating ad-hoc and instance video search, events detection, video description and hyperlinking," in *Proceedings of TRECVID 2017*. NIST, USA, 2017.

[4] Milind Naphade, John R. Smith, and Jelena Tesic, "Large-scale concept ontology for multimedia," *IEEE MultiMedia*, vol. 13, no. 3, pp. 86–91, July-September 2006.

[5] "NVIDIA AI City Challenge," http://smart-city-conference.com/AICityChallenge/.

[6] "VGG Image Annotator," https://gitlab.com/vgg/via/.

[7] Longyin Wen, Dawei Du, Zhaowei Cai, Zhen Lei, Ming-Ching Chang, Honggang Qi, Jongwoo Lim, Ming-Hsuan Yang, and Siwei Lyu, "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *arXiv CoRR*, vol. abs/1511.04136, 2015.

[8] "NVIDIA Tesla P100," http://www.nvidia.com/object/tesla-p100.html.

[9] "NVIDIA Jetson TX2," http://www.nvidia.com/object/embedded-systems-dev-kits-modules.html.

[10] "Caffe," http://caffe.berkeleyvision.org/.

[11] Joseph Redmon, "Darknet: Open source neural networks in C," http://pjreddie.com/darknet/, 2013–2016.

[12] "Tensorflow," https://www.tensorflow.org/.

[13] "MXNET, a flexible and efficient library for deep learning," https://mxnet.incubator.apache.org/.

[14] "Torch, a scientific computing framework for LuaJIT," http://torch.ch/.

[15] "Image Annotation Tool source code," https://github.com/jvkrishna/Image-Annotation-Tool.

[16] "Microsoft Azure," https://azure.microsoft.com/en-us/.

[17] "The PASCAL Visual Object Class Challenges," http://host.robots.ox.ac.uk/pascal/VOC/.

[18] "NVIDIA Pascal GPU," https://developer.nvidia.com/pascal.

[19] "NVIDIA JetPack SDK," https://developer.nvidia.com/embedded/jetpack.

[20] "NVIDIA AI City Challenge code repository," https://github.com/NVIDIAAICITYCHALLENGE.

[21] Longyin Wen, Wenbo Li, Junjie Yan, Zhen Lei, Dong Yi, and Stan Z Li, "Multiple target tracking based on undirected hierarchical relation hypergraph," in *CVPR*, 2014, pp. 1282–1289.