**METEOR: A Massive Dense & Heterogeneous Behavior Dataset for Autonomous Driving**

Rohan Chandra¹, Mridul Mahajan², Rahul Kala², Rishitha Palugulla³, Chandrababu Naidu³, Alok Jain³, and Dinesh Manocha¹

Dataset page: https://gamma.umd.edu/meteor

---

**Abstract**—We present a new and complex traffic dataset, METEOR, which captures traffic patterns in unstructured scenarios in India. METEOR consists of more than 1000 one-minute video clips, over 2 million annotated frames with ego-vehicle trajectories, and more than 13 million bounding boxes for surrounding vehicles or traffic agents. METEOR is a unique dataset in terms of capturing the heterogeneity of microscopic and macroscopic traffic characteristics. Furthermore, we provide annotations for rare and interesting driving behaviors such as cut-ins, yielding, overtaking, overspeeding, zigzagging, sudden lane changing, running traffic signals, driving in the wrong lanes, taking wrong turns, lack of right-of-way rules at intersections, etc. We also present diverse traffic scenarios corresponding to rainy weather, nighttime driving, driving in rural areas with unmarked roads, and high-density traffic scenarios. We use our novel dataset to evaluate the performance of object detection and behavior prediction algorithms. We show that state-of-the-art object detectors fail in these challenging conditions and also propose a new benchmark test: action-behavior prediction with a baseline mAP score of 70.74.

---

**I. INTRODUCTION**

Recent research in learning-based techniques for robotics, computer vision, and autonomous driving has been driven by the availability of datasets and benchmarks. Several traffic datasets have been collected from different parts of the world to stimulate research in autonomous driving, driver assistants, and intelligent traffic systems. These datasets correspond to highway traffic or urban traffic and are widely used in the development and evaluation of new methods for perception [6], [10], prediction [7], [11], [12], [20], [1], behavior analysis [5], [8], [9], and navigation [28], [26].

Many initial autonomous driving datasets were motivated by computer vision or perception tasks such as object recognition, semantic segmentation or 3D scene understanding. Recently, many other datasets have been released that consist of point-cloud representations of objects captured using LiDAR, pose information, 3D track information, stereo imagery or detailed map information for applications related to 3D object recognition and motion forecasting. Many large-scale motion forecasting datasets such as Argoverse [13], and Waymo Open Motion Dataset [18], among others, have been used extensively by researchers and engineers to develop robust prediction models that can forecast vehicle trajectories. However, existing datasets do not capture the rare behaviors or heterogeneous patterns. Therefore, prediction models trained on these existing datasets are not very robust in terms of handling challenging traffic scenarios that arise in the real world.

A major challenge currently faced by these models is the heavy tail problem [18], [13]. This problem refers to the challenges of forecasting models in terms of dealing with rare and interesting instances. Such instances may include cut-ins, over-speeding, overtaking, rule-breaking, etc. Typically, there are only a handful of such instances in current datasets due to the infrequent nature of their occurrence. There are several ways that current datasets are addressing the heavy tail problem, including:

1) **Mining**: The Argoverse and Waymo datasets use a mining procedure that includes scoring each trajectory based on its “interestingness” to explicitly search for difficult and unusual scenarios [18], [13].

2) **Diversifying the taxonomy**: Train the prediction and forecasting models to identify the unknown agents at the time of testing. This approach necessitates annotating a diverse taxonomy of class labels. Argoverse and nuScenes [2] contain 15 and 23 classes, respectively.

3) **Increasing dataset size**: Another way to deal with these rare scenarios is to simply collect more data with the premise that collecting more traffic data will likely also increase the number of such scenarios in the dataset.

---

**A. Main Results**

We present a novel dataset, METEOR, corresponding to the traffic conditions in India. In many ways, the traffic patterns and surrounding environment in parts of India are different from and regarded as more challenging, than those in other parts of the world. India has a population of more than 1.35 billion and 90% of passenger and industrial transportation occurs via roads. This leads to high congestion and traffic density. Some of these roads are unmarked or unpaved. Moreover, the traffic agents moving on these roads correspond to vehicles, buses, trucks, bicycles, pedestrians, auto-rickshaws, two-wheelers such as scooters and motorcycles, etc. The traffic conditions tend to be heterogeneous in terms of microscopic variables corresponding to the size of the traffic-agents, mechanical and dynamic characteristics, lateral and longitudinal gaps between these agents, lateral distribution, and macroscopic variables corresponding to speed, density,
We introduce a new benchmark task for directly addressing the heavy tail problem called action-behavior prediction in which the goal is to predict both actions as well as rare & interesting scenarios. We provide pre-trained models with a baseline score of 70.74 mAP.

B. Applications and Benefits

We have used METEOR to evaluate the performance of some algorithms and highlight areas for future research. We report the performance of current algorithms for 2D object detection and behavior prediction on METEOR. We report accuracy results from our experiments on METEOR, which indicates 2D object detection methods may not have high accuracy on many scenarios. Our results show that even the state-of-the-art object detector [3] based on the latest transformer architecture [31] may not work well on METEOR, but yields good performance on existing datasets that mostly consist of sparse traffic. The degradation of 2D object detection methods in complex environments with high heterogeneity (up to 9 unique agents per frame) and high density (up to 40 total agents per frame) in bad weather and night-time driving is approximately $6 \times$ and $8 \times$ compared to the Argoverse and Waymo Open Motion datasets, respectively.

We also introduce a novel benchmark task on METEOR called “action-behavior prediction”. Our goal is to predict both actions as well as rare and interesting cases. We use a video activity recognition model [32] and report a strong baseline mAP score of 70.74, which is 40% more than the state-of-the-art action prediction score on the Honda Driving Dataset [23].

II. PRIOR WORK

A. Tracking and Trajectory Prediction

Datasets such as the Argoverse [13], Lyft Level 5 [22], Waymo Open Dataset [18], ApolloScape [25], nuScenes dataset [2] have been used for trajectory forecasting [7], [11], [12], [20], [1], [34], [4], [29] and tracking [6], [10].
**Figure 1: METEOR**: We summarize various characteristics of our dataset in terms of scene: traffic density, road type, lighting conditions, agents (we indicate the total count of each agent across 1250 videos), and behaviors, along with their size distribution (in GB). The total size of the current version of the dataset is around 100 GB, and it will continue to expand. Our dataset can be used to evaluate the performance of current and new methods for perception, prediction, behavior analysis, and navigation based on some or all of these characteristics. Details of the organization of our dataset are given at https://gamma.umd.edu/meteor.

METEOR has many additional characteristics with respect to these datasets in several ways. For instance, METEOR’s 2.02 million annotated frames are more than 10× the current highest number of annotated frames with respect to other dataset with high density traffic (Apolloscape). Furthermore, METEOR consists of 16 different traffic-agents that include only on-road moving entities (and not static obstacles). This is by far, the most diverse in terms of class labels. In comparison, Argoverse and nuScenes both contain 10 and 13 traffic-agents, respectively. METEOR is the first motion forecasting and behavior prediction dataset with traffic patterns from rural and urban areas that consist of unmarked roads and high-density traffic. In contrast, traffic scenarios in Argoverse, Waymo, Lyft, and nuScenes have been captured on sparse to medium density traffic with well-marked structured roads in urban areas.

Most current datasets do not contain annotations for rare and interesting driving behaviors such as cut-ins, yielding, overtaking, overspeeding, zigzagging, sudden lane changing, running traffic signals, driving in the wrong lanes, taking wrong turns, and lack of right-of-way rules at intersections and roundabouts. In practice, motion forecasting and behavior prediction models require many such instances in the training dataset to recognize them robustly at test time. To find examples of such instances, existing datasets use a mining procedure that heuristically searches the dataset for rare and interesting examples. This is, however, an inefficient process that requires a huge amount of human labor and time. METEOR is intended to complement these datasets by explicitly providing over 2 million annotated frames that also contain many instances of such driving behaviors. Another dataset, TRAF [7] is similar to METEOR in that it also contains rural scenarios with high-density traffic in India. METEOR contains 28× the number of annotated frames as TRAF along with 2× the number of agent classes. In addition, TRAF does not have annotations for the rare and interesting behaviors that are listed in Table I.

### B. Semantic Segmentation

CityScapes [16] is a widely used dataset for several tasks, primarily semantic segmentation. It is based on urban traffic data collected from European cities with structured roads and low traffic density. In contrast, the Indian Driving Dataset (IDD) [30] is collected in India with both urban and rural areas with high-density traffic. A common aspect of both these datasets (CityScapes and IDD), however, is the relatively low annotated frame count (25K and 10K, respectively). This is probably due to the effort involved with annotating every pixel in each image. IDD also contains high-density traffic scenarios in rural areas, similar to METEOR. However, our dataset has 200× the number of annotated frames and 1.6× the number of traffic-agent classes. Similar to TRAF, the IDD does not contain the behavior data that is provided by METEOR.

### C. Action Prediction

Action prediction is closely related to behavior prediction. The former corresponds to the task of predicting turns (right,
U-turn, or left), acceleration, merging, and braking while the latter corresponds to the task of predicting driver-intrinsic behaviors such as over-speeding, overtaking, cut-ins, yielding, and rule-breaking in addition to actions. The two most prominent datasets for action prediction include the Honda Driving Dataset (HDD) [27] and the Brain4cars dataset [21]. Some of the major distinctions between M\textsc{eteor} and the HDD are the size of the dataset (approximately $10^\times$), the availability of scenes with night driving and rainy weather, and the inclusion of unstructured environments in low-density traffic. The Brain4cars dataset contains nearly as many annotated samples as M\textsc{eteor}, but it consists of low-density traffic in urban areas with well-marked roads. The annotations in prior datasets are limited to actions and do not contain the rare and interesting behaviors that we introduce in M\textsc{eteor}.

### III. M\textsc{eteor} Dataset

Our dataset is visually summarized in Figure 1. Below, we present some details of the data collection process and discuss some of the salient features and characteristics of M\textsc{eteor}.

#### A. Dataset Collection

The data was collected in and around the city of Hyderabad, India within a radius of 42 to 62 miles. Several outskirts were chosen to cover rural and unstructured roads. Our hardware capture setup consists of two wide-angle Thinkware F800 dashcams mounted on an MG Hector and Maruti Ciaz. The camera sensor has 2.3 megapixel resolution with a $140^\circ$ field of view. The video is captured in full high definition with a resolution of $1920 \times 1080$ pixels at a frame rate of 30 frames per second. The dashcam is embedded with an accurate positioning system that stores the GPS coordinates, which were processed into the world frame coordinates. The sensor synchronizes between the camera and the GPS. Recordings from the dashcam are streamed continuously and are clipped into one-minute video segments.

#### B. Dataset organization

The dataset is organized as 1250 one-minute video clips. Each clip contains static and dynamic XML files. Each static file summarizes the meta-data of the entire video clip including the behaviors, road type, scene structure etc. Each dynamic file describes frame-level information such as bounding boxes, GPS coordinates, and agent behaviors. Our dataset can be searched using helpful filters that sort the data according to the road type, traffic density, area, weather, and behaviors. We also provide many scripts to easily load the data after downloading.

#### C. Annotations

We provide the following annotations in our dataset:
1) Bounding boxes for every agent.
2) Agent class IDs.
3) GPS trajectory for the ego-vehicle. We also include the camera intrinsic matrix so that depth estimation techniques may be used to generate trajectories of the surrounding vehicles.
4) Environment conditions including weather, time of the day, traffic density, and heterogeneity.
are summarized below:

D. Rare and interesting behaviors

We provide a total of 17 different types of rich collection of rare and interesting cases that are unique to our dataset. They are summarized below:

- Overtaking (OT): If any of the agent of interest overtakes the other agent with sudden or aggressive movement.
- Overspeeding (OS): If the vehicle over-speeds (based on speed limits) due to any reason.
- Yield (Y): A pedestrian, bicycle, or any slow-moving agent trying to cross the road in front of another agent. If the latter slows down or stops, letting them cross the road then such behavior is labeled as yield.
- Cutting (C): If pedestrian, bicycle, or any slow-moving agent trying to cross the road is interrupted by another agent of interest cutting across them, then such behavior is labeled as cutting.

Yielding and cutting can also be re-labeled as instances of jaywalking. In a majority of these cases, one of the agents involved is a pedestrian crossing the road in the middle of traffic.

- Lane change w. lane markings (LC(m)): When agents aggressively change lanes on roads with clear lane markings.
- Lane change w/o. lane markings (LC): When agents aggressively change lanes on roads without clear lane markings.

The above two annotations can be used to identify videos in the dataset that contain roads without lane markings for relevant applications.

- Zigzagging (ZM): If any of the agent of interest undergoes a zigzag movement in the traffic, the agent behavior is classified as zigzagging.

In addition to the above driving behaviors, we also annotate traffic agents breaking traffic rules. These are particularly unique since rule breaking scenarios are rare.

- Running a traffic light (RB TL): This behavior involves passing through an intersection even though the traffic signal is red.
- Wrong Lane (RB WL): A road may not be divided for inbound and outbound traffic by a physical barrier, making it possible for the motorists to use the inbound lane for the outbound traffic and vice versa. This behavior identifies all such cases.
- Wrong Turn (RB WT): If any agent makes a turn (including U-turns) that is not allowed, the behavior is identified as a wrong turn.

E. Dataset statistics

We defer an in-depth analysis to the supplementary material and highlight a few key details in this section. Analysis of the distribution of agents and behaviors is in terms of total

| Area       | Density | Heterogeneity |
|------------|---------|---------------|
| Urban      | 76.80   | 84.22         |
| Rural      | 23.20   | 15.78         |
| Low        | 51.76   | 18.49         |
| Medium     | 29.73   | 20.61         |
| High       | 14.49   | 10.86         |

| Road type   | Time | Weather | # videos |
|-------------|------|---------|----------|
| Smooth      | 71.00| 74.25   | 1250     |
| Rough       | 29.00| 25.75   | 1250     |
| Day         | 90.38| 9.62    | 1250     |
| Night       | 25.75| 9.62    | 1250     |
| Normal      | 84.22| 15.78   | 1250     |
| Rainy       | 18.49| 10.86   | 1250     |
| Normal      | 74.25| 9.62    | 1250     |
| Rainy       | 25.75| 9.62    | 1250     |
| Smooth      | 90.38| 15.78   | 1250     |
| Rough       | 29.00| 10.86   | 1250     |
| Day         | 84.22| 15.78   | 1250     |
| Night       | 18.49| 10.86   | 1250     |

Table II: Dataset statistics: We report percentage of videos according to various aspects.
Table III: Object detection: We show the results of using the state-of-the-art object detection method, DETR [3], on autonomous driving motion forecasting and tracking datasets. Note that most methods typically report results on the mAP metric.

| Dataset                        | mAP | mAP50 | mAP75 | mAP5 | mAP10 | mAP15 |
|--------------------------------|-----|-------|-------|------|-------|-------|
| Kitti Object Detection [19]    | 23.00 | 42.2 | 23.7 | - | - | - |
| nuScenes [2]                   | 41.80 | - | - | - | - | - |
| Apolloscape [25]               | 48.43 | - | - | - | - | - |
| Waymo Open Motion [18]         | 65.31 | - | - | - | - | - |

| METEOR                         | 8.3 | 15.60 | 8.20 | 0.40 | 3.70 | 14.60 |

Count, uniqueness, and duration (in seconds). In terms of agents, METEOR is very dense and highly heterogeneous. For instance, in Figure 3a, we show that the total number of agents in a single frame can reach up to 40. In Figure 3d, we show that there can be up to 9 unique agents per frame. In terms of behavior, we present the distribution of behaviors across videos in Figure 3b and a distribution of each behavior’s average duration in Figure [20], [1], [7], [24], [17]. In Table II, we present the statistics for different aspects of traffic environments such as road type, traffic density, weather, and so on.

IV. Baselines

We provide the pre-trained models for object detection and behavior prediction at https://gamma.umd.edu/meteor.

A. Object detection in unstructured scenarios

We perform object detection using a DEtection TRansformer (DETR) [3] model with a ResNet-50 backbone. The model is first pre-trained on the COCO dataset and then fine-tuned on our dataset. DETR treats object detection as a direct set prediction task. In Table III, we show the average precision (AP) results of the current state-of-the-art object detection technique, DETR, on motion forecasting and tracking datasets. We highlight the fact that DETR, which succeeds on prior datasets, fails on our dataset. There are two possible reasons. First, state-of-the-art detectors, including DETR, are pre-trained on the MS COCO dataset which contains only 9 categories of commonly occurring traffic agents. This was not an issue for detectors on existing datasets since those datasets typically contain a subset of those 9 classes. METEOR, on the other hand, contains 16 agent categories that are approximately equally distributed. The second reason is due to the challenging traffic environments. In particular, DETR succeeds in daytime scenarios in our dataset, but fails in nighttime driving, in rural areas, and dense traffic. Hopefully future detectors can robustly handle such cases.

B. Action-Behavior prediction

We introduce a novel benchmark task on METEOR called “action-behavior prediction”. Our goal is to predict both actions as well as rare and interesting cases. Action-behavior prediction is distinct from action prediction (which exclusively considers actions) and behavior prediction (which includes trajectory prediction). We use a video activity recognition model [32] and report a strong baseline mAP score of 70.74 that is 40% more than the state-of-the-art action prediction score on the Honda Driving Dataset [23]. The reason for this improvement is that the behaviors and actions in METEOR are between 1 – 3 seconds in duration, which at 30 frames per second corresponds to 24 – 72 frames. The HDD dataset, on the other hand, uses only 20 frames as input, which by comparison has relatively less information to learn from.

We use a Temporal Segment Network (TSN) [32] with a ResNet-101 backbone to predict behaviors and rare instances from an input RGB video. The model is first pre-trained on the Multi-Moments in Time dataset and then refined on our dataset. TSN employs a sparse temporal sampling strategy which first divides the input video into several segments of the same size. Next, a few frames are randomly sampled for each segment. Unlike dense sampling, this not only lowers the computation cost but also helps in capturing long-range temporal structures. The network then makes predictions for each segment over the behavior labels. A consensus of these segment-wise predictions is obtained using segment-wise aggregation of class scores. We present the overall mAP results in Table IV. We provide a strong baseline that will serve as a backbone for future research in behavior and rare instance prediction in unstructured traffic environments.

Table IV: Behavior detection results: We use a video activity prediction model [32] to perform behavior prediction on our dataset and compare the mAP with a behavior prediction on the HDD.

| Dataset                        | mAP |
|--------------------------------|-----|
| Honda Driving Dataset [27]     | 51.10 |
| METEOR                         | 70.74 |

V. Conclusion

We present a new dataset, METEOR, for autonomous driving applications in dense, heterogeneous, and unstructured traffic scenarios. Our dataset consists of 17 different behaviors and rare instances such as cut-ins, yielding, jaywalking, overtaking, and overspeeding. Additionally, METEOR has 16 unique traffic-entities. METEOR is diverse in terms of bad weather scenarios, nighttime driving, rural areas with unmarked lanes, and high-density traffic. Overall, METEOR is the largest and most diverse autonomous driving dataset to-date collected in unstructured traffic.

Our dataset has some limitations. While METEOR contains bounding box information for the surrounding agents, we currently do not provide trajectory information from a fixed reference frame. One would have to use depth estimation techniques to extract such trajectories. Furthermore, our dataset does not contain HD maps, which are used in many applications. For future work, we hope that our dataset can benefit in terms of design and evaluation of new motion forecasting and action-behavior prediction algorithms in dense and heterogeneous traffic. Finally, we hope to include semantic segmentation capability as part of METEOR by providing pixel labels for each object.
REFERENCES

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 961–971, 2016.

[2] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liang, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. Nuscenes: A multimodal dataset for autonomous driving. arXiv preprint arXiv:1903.11027, 2019.

[3] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In European Conference on Computer Vision, pages 213–229. Springer, 2020.

[4] Rohan Chandra, Aniket Bera, and Dinesh Manocha. Stylepredict: Machine theory of mind for human driver behavior from trajectories. arXiv preprint arXiv:2011.04816, 2020.

[5] Rohan Chandra, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. Densespeds: Pedestrian tracking in dense crowds using front-rvo and sparse features. In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 468–475. IEEE, 2019.

[6] Rohan Chandra, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. Traffic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8483–8492, 2019.

[7] Rohan Chandra, Uttaran Bhattacharya, Trisha Mittal, Aniket Bera, and Dinesh Manocha. Cmetric: A driving behavior measure using centrality functions. arXiv preprint arXiv:2003.04424, 2020.

[8] Rohan Chandra, Uttaran Bhattacharya, Trisha Mittal, Xiaoyu Li, Aniket Bera, and Dinesh Manocha. Graphrqi: Classifying driver behaviors using graph spectrums. arXiv preprint arXiv:1910.00049, 2019.

[9] Rohan Chandra, Uttaran Bhattacharya, Tanmay Randhavane, Aniket Bera, and Dinesh Manocha. Roadtrack: Tracking road agents in dense and heterogeneous environments. 2019.

[10] Rohan Chandra, Trinuji Guan, Srujan Panuganti, Trisha Mittal, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms. IEEE Robotics and Automation Letters, 5(3):4882–4890, 2020.

[11] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Aymery Constant and Emmanuel Lagarde. Protecting vulnerable road and behavior of road-agents using spectral clustering in graph-lstms. Biomedicine and Pharmacotherapy, 125:109877, 2020.

[12] Ashesh Jain, Liting Sun, Di Wang, Haojie Shi, Aubrey Clausse, Maximilian Naumann, Julius Kummerle, Hendrik Konigshof, Christoph Stiller, Arnaud de la Fortelle, et al. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. arXiv preprint arXiv:1910.03083, 2019.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008, 2017.

[14] Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaogang Tang, and Lu Van Gool. Temporal segment networks: Towards good practices for deep action recognition, 2016.

[15] Wei Zhan, Liting Sun, Di Wang, Haojie Shi, Aubrey Clausse, Maximilien Naumann, Julius Kummerle, Hendrik Konigshof, Christoph Stiller, Arnaud de la Fortelle, et al. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. arXiv preprint arXiv:1910.03083, 2019.

[16] Hang Zhao, Jiayi Gao, Tian Lan, Chen Sun, Benjamin Sapp, Balakrishnan Varadarajan, Yue Shen, Yi Shen, Yuning Chai, Cordelia Schmid, et al. Tnt: Target-driven trajectory prediction. arXiv preprint arXiv:2008.08294, 2020.

[17] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In Conference on Computer Vision and Pattern Recognition (CVPR), 2012.