Open Set Driver Activity Recognition

Alina Roitberg   Chaoxiang Ma   Monica Haurilet   Rainer Stiefelhagen

Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology
{firstname.lastname}@kit.edu

***Abstract*— A common obstacle for applying computer vision models inside the vehicle cabin is the dynamic nature of the surrounding environment, as unforeseen situations may occur at any time. Driver monitoring has been widely researched in the context of *closed set* recognition *i.e*. under the premise that all categories are known a priori. Such restrictions represent a significant bottleneck in real-life, as the driver observation models are intended to handle the uncertainty of an *open world*.**

In this work, we aim to introduce the concept of *open sets* to the area of driver observation, where methods have been evaluated only on a static set of classes in the past. First, we formulate the problem of open set recognition for driver monitoring, where a model is intended to identify behaviors previously unseen by the classifier and present a novel *Open-Drive&Act* benchmark. We combine current closed set models with multiple strategies for novelty detection adopted from general action classification [1] in a generic open set driver behavior recognition framework. In addition to conventional approaches, we employ the prominent I3D architecture extended with modules for assessing its uncertainty via Monte-Carlo dropout. Our experiments demonstrate clear benefits of uncertainty-sensitive models, while leveraging the uncertainty of all the output neurons in a voting-like fashion leads to the best recognition results. To create an avenue for future work, we make *Open-Drive&Act* public at www.github.com/aroitberg/open-set-driver-activity-recognition.

I. Introduction

How can we deal with driver behavior that was not learned by our models during training? As we will never be able to capture and annotate all possible driver behaviors in our training data, we need to find a way to handle such *unknown* activities. While this task is vital for practical applications of driver activity recognition, previous works merely focused on optimization of the top-1 classification accuracy on a fixed set of carefully designed actions [2]–[11]. Exploring what happens if a video with a new behavior is passed to the model, has been overlooked in the past.

Driver activity recognition has a variety of applications, ranging from perceiving distraction and sending a warning to increasing comfort during autonomous driving (*e.g*. adjusting the movement dynamics if the person is drinking coffee) [11]. However, if a model developed for closed set recognition is utilized directly in an open world, it will be quickly exposed to uncertain situations. This might *e.g*. result in a high number of false positive detections which are both highly disturbing for the user and potentially dangerous.

While a large number of driver behavior recognition benchmarks have been introduced in recent years [2], [4],

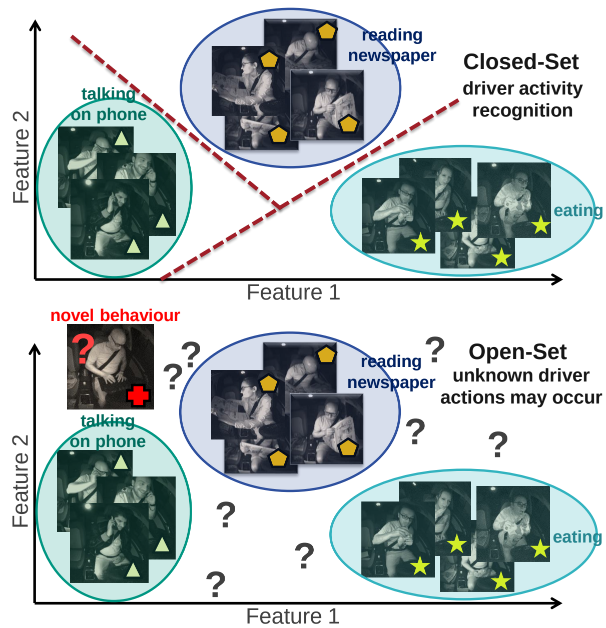

Fig. 1: Standard *closed set* driver behavior recognition benchmarks assume that the test examples only contain classes previously seen during training (top). We propose the task of *open set driver activity recognition*, where behaviors not previously seen by the model are also present (bottom).

[6], [8], [11], they represent a setting where the action categories in the training and test set are exactly the same (Figure 1, top). The largest of these *closed set* benchmarks is Drive&Act [11], in which all of the 34 available fine-grained activity classes are used both for evaluation and for training. However, the recent research underlines the importance of studying the behavior of such models when exposed to previously unseen classes [1], [12]–[14] and highlighting their limits when exposed to uncertainty [13], [15]. Impressive results on the conventional datasets may therefore draw an artificially idealistic picture, as the closed set constraint does not hold for *dynamic* environment of real-life applications. In this work, we propose to incorporate previously unseen behaviors in the evaluation of driver activity recognition models and expose them to *open set* conditions (Figure 1, bottom). We aim to highlight the role of uncertainty in the field of driver observation and develop models, which do not only assign one of the previously seen classes but are also able to distinguish *the known* and *the unknown* ones.

Contributions and Summary Given the dynamic nature of the environment inside the vehicle cabin, we argue that exposing activity recognition methods to previously unseen situations is crucial for applications. This work aims at bringing the notion of *open sets* to the field of driver activity recognition and has the following major contributions. (1) To counter the lack of open set benchmarks for driver behavior understanding, we introduce and publicly release the *Open-Drive&Act* testbed, which augments the original *Drive&Act* dataset with an open set setting, covering ten splits and a formalized evaluation protocol. (2) We propose a generic framework for open set activity recognition which incorporates two facets necessary for such systems: novelty detection and standard supervised classification. First, the novelty detection module computes the *newness score* of the video example and decides whether the behavior is familiar or not based on a variable threshold. Then, the behavior is either labeled as "unknown" or passed to the classification module, a neural network which classifies it with one of the known labels. (3) We evaluate multiple variants of the novelty detection module originating from the field of general activity recognition [1], which we adjust to our driver monitoring task. Different strategies for computing the *newness score* are based on conventional [16]–[18] and uncertainty-based approaches [1], [15] for outlier detection. We establish a strong open set driver behavior recognition benchmark, by systematically evaluating our framework variants on *Open-Drive&Act* in terms of binary (*i.e.* known or unknown) and multi-class accuracy. Our experiments demonstrate the effectiveness of the uncertainty-based novelty detection methods in comparison to the conventional approaches.

II. Related Work

A. Driver Activity Recognition

Vision-based activity recognition has undergone a prompt shift from machine learning approaches operating on hand-crafted features [19], [20] to end-to-end Convolutional Neural Networks (CNNs) [21], [22]. The rise of deep CNNs, however, had a rather slow effect on the field of driver behavior recognition. Presumably due to the comparably small size of available datasets [3], [8], [23] and the data-hungry nature of CNNs, most of the approaches are based on manually defined feature descriptors. The features are often computed from body pose [2], [23]–[25], eye gaze [2] or detected objects [24] and then classified with an SVM [2], random forests [3], HMM [23], [26], RNN [23], [24] or *e.g.* treated as a spatio-temporal structure passed to a graph neural network [25]. Recent emergence of large-scale benchmarks [11] and progress in deep transfer learning has lead to the first CNN-based driver activity recognition approaches surpassing the hand-crafted methods in performance [4], [10], [11]. However, all existing methods are designed and evaluated in the closed set scenario. We aim to overcome this constraint and research driver activity recognition in the realistic open set case, where the model needs to solve two tasks: identify, whether the action is known and, if so, recognize the correct known class.

B. Novelty Detection and Open Set Recognition

While open set recognition has been well-studied in some areas, such as face recognition [27], [28], the research of open set activity recognition has been very limited. Recent surveys of open set recognition [29], [30] identify the ability to distinguish between the known and unknown classes (*i.e. novelty detection*) as imperative in any recognition system. The *novelty* of captured data has been addressed through different machine learning methods with a systematic overview of approaches provided in [31], [32]. A lot of today's novelty detection research is anchored in classical machine learning *e.g.* One-class SVM [33] or Gaussian Mixture Models (GMMs) [32], which solve the problem from the probabilistic point of view. Distance-based approaches such as Local Outlier Factor are based on the idea of local density, assuming that "normal" samples have a set of k neighbors similar to the training data [34]. From the deep learning perspective, anomaly detection is usually achieved by thresholding the Softmax value of the top predicted class [35]–[37]. Recently, novelty detection has been addressed for the first time in the context of action classification [1]. The work compares classical novelty detection methods and presents a new algorithm based on Bayesian uncertainty, which we leverage for the driver monitoring task.

Despite its relevance for practical applications, there is a lack of open set recognition research inside the vehicle cabin. We therefore introduce the first *open set* framework for driver behavior understanding, enhancing the SoA methods for standard supervised driver activity recognition with multiple strategies to quantify the *newness* of the video sample adopted from the field of general activity recognition [1]. We analyze the performance of our system on our *Open-Drive&Act* benchmark for open set driver activity recognition, which we will publicly release to the community.

III. Open Set Driver Activity Recognition

A. *Open-Drive&Act* Benchmark and Problem Formulation

To address the lack of open set benchmarks for driver observation, we introduce *Open-Drive&Act* – the first driver activity recognition testbed in which the evaluation procedure comprises both *known* and *unknown* behaviors. Thereby, we extend the *Drive&Act* dataset for closed set activity recognition inside the vehicle-cabin to the open set scenario and formalize the evaluation process to handle *unseen classes* as follows. We employ the available *Drive&Act* videos captured by a NIR-camera facing the person and all 34 annotations on the fine-grained activity level. Then, we split the dataset into 24 seen and 10 unseen categories, of which 5 unknown classes are used for validation and 5 for testing. Videos of unseen activities are not available during training, while samples of the remaining seen activities are further split into training (60%), validation (20%) and testing (20%). We randomly generate 10 splits and report the average and the standard deviation of the recognition metrics. Given a model trained on the available seen behaviors, our framework needs to handle both known and novel activities at test-time and

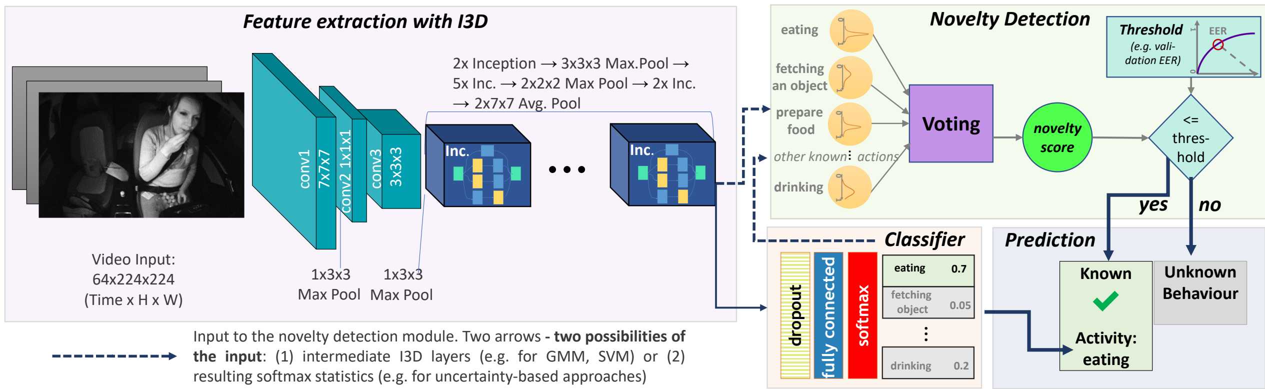

Fig. 2: Overview of the proposed neural network-based framework for open set driver activity recognition.

is examined in two different settings. (1) Novelty detection comprises binary decision whether the video depicts a known behavior evaluated as area under the ROC- and PR-curve. (2) Open set multi-class recognition aims to first reason about the novelty of the instance and then and in the case the class is known, it assigns the correct label. In the latter setting, the multi-class accuracy is used as the recognition metric, where *unknown* is treated as an additional category extending the already known classes.

B. Neural Architecture Overview

In this paper, we present a generic framework for open set driver activity recognition, incorporating both facets necessary for such systems: conventional supervised classification of previously seen activity classes and the ability to distinguish between those cases (*i.e.* novelty detection). The proposed model takes as input a video and determines whether the instance is seen or unseen by thresholding the novelty score. The threshold can be chosen *e.g.* using the Equal Error Rate (EER) of the ROC curve on the validation set. Therein, we compute an encoding of the input, by using an intermediate layer of a neural network trained for supervised classification of the *known* classes. The embedding is then passed to the novelty detection module, which assesses the input newness. Depending on the outcome, the video is either conveyed to a supervised classification module or marked as unknown. Figure 2 provides an overview of our architecture with novelty detection using uncertainty-based selective voting of the output neurons (Section III-E). While Figure 2 illustrates the best performing variant of the novelty detection module, we also adopted and benchmarked other popular novelty detection approaches, such as One-Class SVM, as described in Section III-D and Section III-E.

C. Video Embedding and Classification

For representation learning and classification in our system, we implement a CNN based on 3D-convolution and -pooling kernels to deal with the spatial and temporal dimensions of our input. The neural network takes as input a video snippet of 64 frames with a resolution of 224×224 and learns to assign one of the *known* activity labels. Our framework is based on the Inflated 3D architecture (I3D) proposed in [21] for conventional action recognition. I3D is a version of the well-known Inception-v1 architecture extended with a temporal dimension. The network stacks 9 characteristic Inception modules: small sub-networks which use convolutions with different kernel sizes in parallel, while keeping the number of operations low by reducing the dimensions via $1 \times 1 \times 1$ convolutions. The complete I3D network consists of 27 layers with three convolution layers at the beginning and one fully-connected layer at the end, four max-pooling layers at the beginning and one average pooling layer preceding the last fc-layer and nine inception modules which themselves are two layers deep. Figure 3 visualizes the obtained I3D embedding representations for both known and unknown classes of one *Open-Drive&Act* validation split using t-SNE, depicting a clear correlation between the computed features and action semantics.

D. Classic Novelty Detection Approaches

One-Class SVM The One-class SVM introduced by Schölkopf *et al.* [17] is a widely used non-probabilistic method for novelty detection. The model learns to transform video embeddings into a feature space defined by a boundary hyperplane aiming to increase the separation margin from the origin. The novelty estimate is then quantified as the signed distance to the separating hyperplane, which is positive, if the data point is inside the boundary (*i.e.* a *known* class). We use a Radial-Basis-Function kernel and train the SVM using the intermediate I3D-embeddings as our video representation.

Gaussian Mixture Models We consider a classical generative approach for novelty detection using Gaussian Mixture Models (GMMs) [32]. We use a mixture of 24 Gaussian distributions (*i.e.* number of the known categories) and estimate model parameters using the Expectation-Maximization algorithm to fit our *known* activities, represented as the intermediate embeddings of the I3D model. We then use the estimated probability density function to quantify the novelty.

Neural Network Softmax Confidence A common way for novelty detection with neural networks is to threshold the output of the neuron with the highest activation value [18], [35], [38]. In conventional CNNs for visual recognition, the output of the last fully-connected layer is normalized using the *Softmax* function, resulting in point estimates for a fixed set of classes from which Cross-Entropy loss is computed. The resulting *Softmax* scores are often denoted as class probabilities [15], since they satisfy the properties of a probability function: they range between 0 and 1 and sum up to one. The input is assigned the class with the maximum *Softmax* score and can be directly used to quantify the data normality [18], [39]. We therefore train the I3D model to distinguish between the known activities directly through the top class Softmax score to detect unknown behaviors.

E. Deep Probabilistic Novelty Detection

When considering the *Softmax* score of the predicted class as a confidence measure, which is a common practice [18], [35], one would likely face the phenomenon of model *miscalibration* [13]. The top-1 Softmax tends to be biased towards very high values for both, correct predictions and misclassifications [13]. One way to address models confidence is to aim for the posterior *distributions* instead of single point estimates for each class. In Bayesian Neural Networks [40], this is achieved by placing a prior over models weights, so that learned parameters and therefore also the predicted values, become probability distributions. The predictive probability of the BNN is obtained by integrating over the parameter space ω (Eq. 1). Since the posterior $p(\omega|X_{train}, Y_{train})$, where X_{train}, Y_{train} denotes the training data and annotations, is intractable, it is often replaced with a variational distribution $q_\theta(\omega)$ and approximated using Monte-Carlo sampling (Eq. 2 and Eq. 3):

$$p(y|\mathbf{x}, X_{train}, Y_{train}) = \int_\omega p(y|\mathbf{x}, \omega) p(\omega|X_{train}, Y_{train})\ d\omega \quad (1)$$

$$\approx \int_\omega p(y|\mathbf{x}, \omega) q_\theta(\omega) d\omega \quad (2)$$

$$\approx \frac{1}{T} \sum_{t=1}^{T} p(\mathbf{a}|\mathbf{x}, \omega_t), \text{ with } \omega_t \sim q_\theta(\omega) \quad (3)$$

Gal *et al.* (2016) has provided a proof, that iteratively applying dropout at test-time and then computing the output statistics of the model, is a variational approximation of a BNN posterior distribution with network parameters modeled as a Gaussian Process [15]. Dropout sets the nodes to zero with a probability ρ making the network non-deterministic (therefore $q_\theta(\omega)$ follows a Bernoulli distribution for model weights to approximate the BNN posterior). This approach, often referred to as MC-Dropout, has been recently used for identifying previously unknown activity samples [1].

Novelty via Dropout Sampling Statistics We equip I3D with MC-Dropout at test-time [15] after the last average pooling layer and use its probabilistic version to quantify the normality of driver behavior [1]. Let $\mathbf{x_{emb}}$ be the embedding generated by I3D after the last average pooling layer and W, $\mathbf{b}$ be the weight matrix and bias vector of the last fully connected layer. Typically, network weights W are deterministic at test time and dropout is only active during training. Instead, we apply dropout interactively at test time,

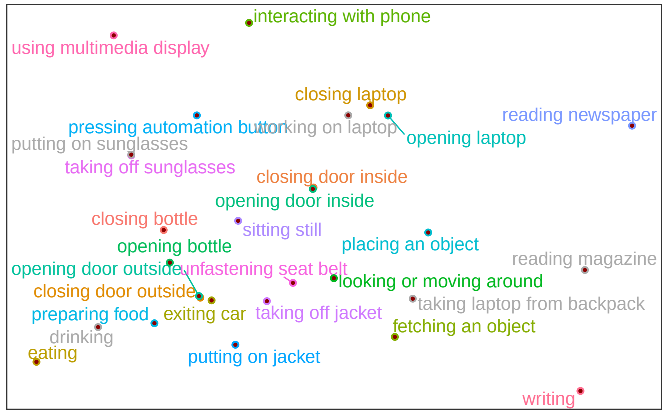

Fig. 3: T-SNE representation of the I3D video embeddings of one Open-Drive&Act validation split. Labeled dots depict the center of all samples of the corresponding category (activities *known* from training in *color*, *unknown* in *gray*).

which below is formalized by multiplying W with a diagonal matrix D, with diagonal values set to 0 with probability ρ and otherwise to 1. The probabilistic I3D no longer gives single point estimates, but now predicts *Gaussian distributions* for each known driver behavior. The mean of the computed distribution is now used to assign the known class to the data and is computed over T stochastic iterations as follows:

$$\mathbb{E}(y|\mathbf{x}) \approx \frac{1}{T} \sum_{t=1}^{T} \text{softmax}(W\ D\ \mathbf{x^t_{emb}} + \mathbf{b}) \quad (4)$$

Similarly, we compute the predicted distribution variance which is directly linked to models uncertainty:

$$U(y|\mathbf{x}) \approx \frac{1}{T-1} \sum_{t=1}^{T} [\text{softmax}(W\ D\ \mathbf{x^t_{emb}} + \mathbf{b}) - \mathbb{E}(y|\mathbf{x})]^2 \quad (5)$$

Uncertainty-based Selective Voting of Output Neurons Previous research highlights, that consolidating the uncertainty of *all* output neurons is more effective than utilizing the top-1 uncertainty alone [1]. Intuitively, we do not only consider the certainty of the model that the input indeed belongs to the predicted class, but also, how sure it is, that it *does not* belong to one of the other classes. To achieve this, we let the output neurons vote about the novelty, as in [1]. The voting is privileged to the subset of output neurons, which usually have stable uncertainty values given the predicted class. Let C be the set of action categories known during training. For each known class $c^\star \in C$, we find a subset of *informed neurons* $I_{c^\star}$, which are classes with the output neuron uncertainty usually not changing much for different examples of $c^\star$. More formally, $c_i \in I_{c^\star}$ if the variance of uncertainty of the c_i neuron calculated over the training set is below a threshold τ (we choose $\tau = 0.004$ empirically using the validation set). For a new video $\mathbf{x}$, we first find the predicted known category $c^\star$ (*e.g.* class with maximum softmax score). Then, we look up the set of informed neurons $I_{c^\star}$ of $c^\star$ (as estimated during training) and use them to compute the novelty score $\upsilon(\mathbf{x})$ as the average uncertainty of the informed neurons: $\upsilon(\mathbf{x}) = \frac{\sum_{c_j \in I_{c^\star}} U(c_j|\mathbf{x})}{|I_{c^\star}|}$.

Method	Determi-nistic	Validation		Test	
		ROC AUC[%]	$\pm SE$	ROC AUC[%]	$\pm SE$
Conventional Methods and Baselines					
Random Chance	–	50.00	–	50.00	–
One-class SVM	✓	62.51	± 5.92	59.35	± 11.19
Gaussian Mixture Model	✓	68.84	± 6.97	65.73	± 7.09
Conventional NN Confidence	✓	82.21	± 5.66	81.05	± 4.64
Deep Models based on Bayesian Uncertainty					
Bayesian I3D – Pred. Variance (T=10)	–	82.53	± 5.40	82.28	± 4.48
Bayesian I3D – Pred. Variance (T=100)	–	83.10	± 5.47	82.69	± 4.36
Bayesian I3D – Pred. Mean (T=10)	–	83.77	± 4.71	82.88	± 3.94
Bayesian I3D – Pred. Mean (T=100)	–	84.60	± 4.50	83.60	± 3.93
Bayesian I3D – Selective Voting	–	**85.30**	± 4.06	**84.33**	± 3.85

TABLE I: Results for unknown driver behavior detection as a binary task on the Drive&Act Dataset (average and standard deviation over the ten splits). Uncertainty-based models consistently outperform standard approaches.

Method	Balanced Accuracy		Normal Accuracy	
	Acc [%]	$\pm SE$	Acc [%]	$\pm SE$
Conventional Methods and Baselines				
Random Chance	4.00	–	4.00	–
One-class SVM	24.49	±8.68	54.78	±7.90
Gaussian Mixture Model	23.61	±6.82	62.39	±5.98
Conventional NN Confidence	44.02	±5.16	74,83	±6.86
Deep Models based on Bayesian Uncertainty				
Bayes. I3D – Mean (T=10)	44.12	±5.12	76,98	±5.86
Bayes. I3D – Mean (T=100)	42.71	±5.10	77,39	±5.44
Bayes. I3D – Variance (T=10)	49.76	±7.67	75,94	±6.74
Bayes. I3D – Variance (T=100)	49.31	±9.08	75,55	±6.40
Bayes. I3D – Selective Voting	**57.55**	±9.54	**77.62**	±4.55

TABLE II: Accuracy for multi-class recognition with an unknown class (24 known classes + unknown). Normal accuracy is the recognition rate across all test samples, while balanced accuracy is the mean of the individual classes recall.

IV. Experiment Results

We evaluate the open set driver behavior recognition pipeline extensively using our testbed, equipping the I3D model with different variants of the novelty detection module. Examined methods range from standard approaches (One-Class SVM, GMM, using CNN confidence directly) to novel uncertainty-based methods (see Section III-E) and are compared with each other and a random classifier baseline.

Novelty Detection We first examine our approaches in terms of raw novelty detection *i.e.* binary decision whether the seen behavior class was present during training. We use the area under the ROC curve computed from the produced novelty scores as our evaluation metric and report the mean and standard deviation over ten splits in Table I. While all models surpass the random classifier, neural network-based approaches show clear advantages, as even using the neural network Softmax score alone outperforms a GMM by 15.32% on the test set. The recognition rate is improved by using probabilistic approaches, as all model variants based on Bayesian uncertainty surpass using conventional Softmax confidence. The predictive mean and variance denote using the MC-Dropout sampling statistics of the top-1 class confidence as in Eq. 4 and Eq. 5. While increasing the number of stochastic forward passes T from 10 to 100 improves the results in cases of both, predictive mean and variance, the advancement is small (0.41% to 0.73%) and might be omitted in favor of better computation speed. Leveraging uncertainty of all the output neurons via informed voting leads to the best recognition rates for both, training and validation set. Thereby, the informed voting strategy surpasses the raw neural network confidence by 3.28% with a total area under ROC of 84.33% on the test set.

Open set Multi-class Recognition Next, we evaluate the open set multi-class recognition, where we use accuracy as our evaluation metric and treat *unseen* as an additional label (Table II). Since the distribution of classes in *Drive&Act* is highly unbalanced [7], we report also the balanced accuracy (mean recall of each individual class, as in [7]). Uncertainty-based selective voting outperforms other approaches in both metrics with a remarkably strong lead in balanced accuracy. This reflects, that the important question is not only *"how do we distinguish between known and unknown?"*, but also "*if we reject a known class by mistake, are we missing out an otherwise correctly predicted sample or a misclassification?*". In case of selective voting, such false positives are usually samples, which would have been incorrectly classified by I3D anyway. Their categorization as *unknown* is therefore not very damaging, and oftentimes even practical. Of course, open set recognition is a harder task and the balanced accuracy is lower than in the closed set case [7]. Still, 77.62% of the test examples are correctly classified (57.55% after balancing), which is significantly higher than the random baseline of 4% for 25 categories. While our evaluation considers all novel activities as a single *unseen* category, methods for knowledge transfer from external sources (*e.g.* via zero-shot learning [41]) would potentially allow us to distinguish different novel behaviours among each other, marking an important direction for future research.

V. CONCLUSION

Driver behavior analysis facilitates new ways of human-vehicle interaction but requires models that are reliable in a constantly changing world. In this work, we introduced the new task of open set driver activity recognition, which extends the conventional driver observation [6], [8], [11] with presence of previously unseen behaviors. We enriched the existing Drive&Act dataset with open set splits and formalized evaluation protocols in a novel *Open-Drive&Act* benchmark, in the hope of encouraging future research of driver observation in a dynamic world. We further implement a generic pipeline for open set driver activity recognition, which combines modern closed set classification methods with an additional component for computation of the novelty score. In an extensive evaluation, we examine both traditional methods and recent algorithms based on Bayesian uncertainty [1] as our novelty detection module. The experiment results reveal clear benefits of uncertainty-aware models for open set recognition, a vital step for applications of such algorithms in real-life driver monitoring systems.

REFERENCES

[1] A. Roitberg, Z. Al-Halah, and R. Stiefelhagen, "Informed Democracy: Voting-based Novelty Detection for Action Recognition," in *British Machine Vision Conference (BMVC)*, UK, 2018.

[2] E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *Transactions on intelligent transportation systems*, vol. 15, no. 6, pp. 2368–2377, 2014.

[3] L. Xu and K. Fujimura, "Real-time driver activity recognition with random forests," in *International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 2014.

[4] Y. Abouelnaga, H. M. Eraqi, and M. N. Moustafa, "Real-time distracted driver posture classification," *NeurIPS Workshop on Machine Learning for Intelligent Transportation Systems*, 2018.

[5] C. Zhao, B. Zhang, J. He, and J. Lian, "Recognition of driving postures by contourlet transform and random forests," *IET Intelligent Transport Systems*, vol. 6, no. 2, pp. 161–168, 2012.

[6] C. Yan, F. Coenen, and B. Zhang, "Driving posture recognition by convolutional neural networks," *IET Computer Vision*, 2016.

[7] S. Martin, E. Ohn-Bar, A. Tawari, and M. M. Trivedi, "Understanding head and hand activities and coordination in naturalistic driving videos," in *Intelligent Vehicles Symposium*, 2014, pp. 884–889.

[8] E. Ohn-Bar, S. Martin, A. Tawari, and M. M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *International Conference on Pattern Recognition*, 2014, pp. 660–665.

[9] S. Y. Cheng and M. M. Trivedi, "Vision-based infotainment user determination by hand recognition for driver assistance," *Transactions on intelligent transportation systems*, vol. 11, no. 3, pp. 759–764, 2010.

[10] P. Gebert, A. Roitberg, M. Haurilet, and R. Stiefelhagen, "End-to-end Prediction of Driver Intention using 3D Convolutional Neural Networks," in *Intelligent Vehicles Symposium (IV)*. IEEE, 2019.

[11] M. Martin*, A. Roitberg*, M. Haurilet, M. Horne, S. Reiß, M. Voit, and R. Stiefelhagen, "Drive&Act: A Multi-modal Dataset for Fine-grained Driver Behavior Recognition in Autonomous Vehicles," in *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, October 2019.

[12] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proceedings of International Conference on Learning Representations*, 2017.

[13] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *Conference on Computer Vision and Pattern Recognition*, 2015.

[14] P. Panareda Busto and J. Gall, "Open set domain adaptation," in *International Conference on Computer Vision*. IEEE, 2017.

[15] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016, pp. 1050–1059.

[16] F. Zorriassatine, A. Al-Habaibeh, R. Parkin, M. Jackson, and J. Coy, "Novelty detection for practical pattern recognition in condition monitoring of multivariate processes: a case study," *The International Journal of Advanced Manufacturing Technology*, vol. 25, 2005.

[17] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[18] C. Richter and N. Roy, "Safe visual navigation via deep learning and novelty detection," in *Robotics: Science and Systems Conf.*, 2017.

[19] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *IEEE International conference on computer vision*, 2013.

[20] A. Roitberg, A. Perzylo, N. Somani, M. Giuliani, M. Rickert, and A. Knoll, "Human activity recognition in the context of industrial human-robot interaction," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2014.

[21] J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017.

[22] M. Asadi-Aghbolaghi, A. Clapes, M. Bellantonio, H. J. Escalante, V. Ponce-López, X. Baró, I. Guyon, S. Kasaei, and S. Escalera, "A survey on deep learning based approaches for action and gesture recognition in image sequences," in *International Conference on Automatic Face & Gesture Recognition*. IEEE, 2017.

[23] A. Jain, H. S. Koppula, S. Soh, B. Raghavan, A. Singh, and A. Saxena, "Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture," *arXiv preprint arXiv:1601.00740*, 2016.

[24] P. Weyers, D. Schiebener, and A. Kummert, "Action and object interaction recognition for driver activity classification," in *2019 IEEE Intelligent Transportation Systems Conference*. IEEE, 2019.

[25] P. Li, M. Lu, Z. Zhang, D. Shan, and Y. Yang, "A novel spatial-temporal graph for skeleton-based driver action recognition," in *2019 IEEE Intelligent Transportation Systems Conference*. IEEE, 2019.

[26] N. Akai, T. Hirayama, L. Y. Morales, Y. Akagi, H. Liu, and H. Murase, "Driving behavior modeling based on hidden markov models with driver's eye-gaze measurement and ego-vehicle localization," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 949–956.

[27] M. Gunther, S. Cruz, E. M. Rudd, and T. E. Boult, "Toward open-set face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 71–80.

[28] R. Vareto, S. Silva, F. Costa, and W. R. Schwartz, "Towards open-set face recognition using hashing functions," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 634–641.

[29] T. Boult, A. Dhamija, M. Gunther, J. Henrydoss, and W. J. Scheirer, "Learning and the unknown: Surveying steps toward open world recognition," in *AAAI Conference on Artificial Intelligence*, 2019.

[30] C. Geng, S.-j. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *arXiv preprint arXiv:1811.08581*, 2018.

[31] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[32] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, 2014.

[33] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in *Advances in neural information processing systems*, 2000, pp. 582–588.

[34] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000.

[35] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proceedings of International Conference on Learning Representations*, 2017.

[36] M. Augusteijn and B. Folkert, "Neural network classification and novelty detection," *International Journal of Remote Sensing*, 2002.

[37] M. Markou and S. Singh, "Novelty detection: A review - Part 2:: Neural network based approaches," pp. 2499–2521, 2003.

[38] ——, "A neural network-based novelty detector for image sequence analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1664–1677, 2006.

[39] S. Ramos, S. Gehrig, P. Pinggera, U. Franke, and C. Rother, "Detecting unexpected obstacles for self-driving cars: Fusing deep learning and geometric modeling," in *Intelligent Vehicles Symposium*. IEEE, 2017.

[40] R. M. Neal, *Bayesian learning for neural networks*. Springer Science & Business Media, 2012, vol. 118.

[41] A. Roitberg, M. Martinez, M. Haurilet, and R. Stiefelhagen, "Towards a fair evaluation of zero-shot action recognition using external data," in *Proceedings of the ECCV Workshop on Shortcomings of Vision and Language, Springer*, 2018.