# Impact, Attention, Influence:
# Early Assessment of Autonomous Driving Datasets

Daniel Bogdoll[†‡*], Jonas Hendl[‡*], Felix Schreyer[†], Nishanth Gowda[†], Michael Färber[‡] and J. Marius Zöllner[†‡]

[†]FZI Research Center for Information Technology, Germany

bogdoll@fzi.de

[‡]Karlsruhe Institute of Technology, Germany

*Abstract*—**Autonomous Driving (AD), the area of robotics with the greatest potential impact on society, has gained a lot of momentum in the last decade. As a result of this, the number of datasets in AD has increased rapidly. Creators and users of datasets can benefit from a better understanding of developments in the field. While scientometric analysis has been conducted in other fields, it rarely revolves around datasets. Thus, the impact, attention, and influence of datasets on autonomous driving remains a rarely investigated field. In this work, we provide a scientometric analysis for over 200 datasets in AD. We perform a rigorous evaluation of relations between available metadata and citation counts based on linear regression. Subsequently, we propose an Influence Score to assess a dataset already early on without the need for a track-record of citations, which is only available with a certain delay.**

*Index Terms*—**Robotics, Autonomous Driving, Datasets, Influence, Impact, Attention, Scientometrics, Bibliometrics**

## I. INTRODUCTION

Autonomous driving technology does not only affect urban transportation [1] and delivery of goods [2], but also farming [3] or warehouse logistics [4]. With the progress of deep learning and this growing interest in AD in many fields of robotic, the number of related datasets is consistently increasing. The datasets have also increased in size and many have become increasingly specialized [5]. The most extensive collection of datasets known to us, *ad-datasets*, currently lists 231 datasets in the domain [6]. However, not all of them are being equally used in the robotics community, the distribution of their citations is heavily skewed. As part of the more impactful works, well known datasets for the core tasks perception and prediction dominate [7]–[9]. As part of the long tail, many datasets for niche research areas exist [10]–[12]. Well known datasets tend to bring many advantages with them: They enable comparison between works, have higher quality, advanced tooling, and often community knowledge and support is available. The increasing number of datasets, which are potentially interesting but lack reputation, leads to a lot of untapped potential: Many researchers are hesitant to use such datasets and stick to old, but established ones instead [6]. This is why we asked ourselves the question: Given a novel dataset without a multi-year track record of citations, is there a way to estimate its future development? Datasets with a high potential might be more appealing already in their early days.
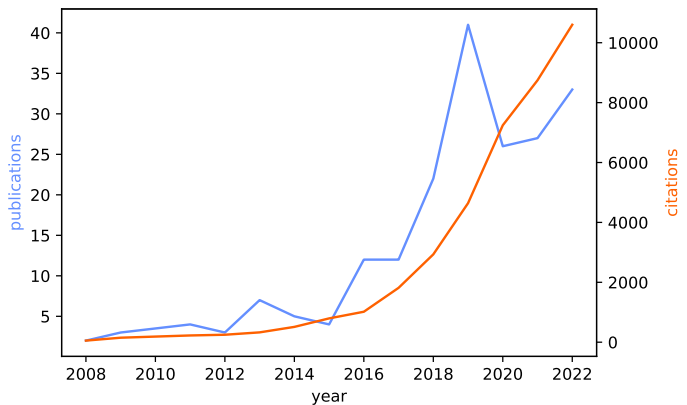
---

* These authors contributed equally



Fig. 1. Course of published datasets and citations of the accompanying publications in the domain of AD. This growing number of datasets, initially without reputation, holds a great deal of untapped potential as researchers struggle to use new datasets for their research. Datasets as listed on ad-datasets [6], citation counts from Semantic Scholar [13].

**Research Gap.** To date, citations are mostly used to assess datasets, which are not available early on. Thus, new datasets can have a hard time gaining traction, which results in untapped potential. It is yet not well understood if and how metadata of datasets relate to future impact or how they can be utilized to assess datasets early on. To the best of our knowledge, such an analysis has not yet been performed.

**Contribution.** In order to analyze the field of datasets, we first assembled the largest collection of datasets with enriched metadata available, including over 200 datasets with metadata from three different sources. We then applied linear regression to evaluate factors which relate to the future impact of datasets, measured in citations. Finally, we propose the Influence Score (IS), which is a mean to assess datasets early on without the need of a multi-year track record of citations. The IS can be used to assess any dataset at any given year, which also allows for later analysis. Our work aims to help researchers from the robotics community to better understand and assess the performance of datasets. This can lead to the design of better and thus more influential datasets as well as an actionable analysis of new datasets to assess their potential. All data used in this work is as of January 04, 2023. All code is available on GitHub.

## II. Related Work

Here, we give an introduction to scientometrics, bibliometrics, and altmetrics, followed by dataset analysis.

### A. Scientometrics, Bibliometrics, and Altmetrics

Scientometrics, Bibliometrics, and Altmetrics are highly intertwined fields that focus on the analysis of science and its processes as a whole, written works of science, and online communication of science, respectively [14].

**Scientometrics.** Ravenscroft et al. [15] examined the impact of research by comparing citation-based metrics, such as citation count or h-index [16], with altmetrics and impact other than citations, e.g., societal and economic impact. However, they found no strong relationship between the fields. Hicks et al. [17] suggest using multiple factors to portray multiple aspects.

Leydesdorff et al. [18] claim that citations are equated to impact and evaluate the relationship between impact and research quality. They found that short-term citations signify the investment in a current discourse, while long-term citations signify acceptance as reliable scientific knowledge. However, some researchers question if or to what extent citations measure scientific impact and point to issues, e.g., inconsistent reasons for citations [19]. Problems include the cumulative advantages already successful papers experience [20], self-citations, which men do more often [21], negative citations, and citing out of reasons that do not reflect actual use or relevance [22]. Valenzuela et al. [23] presented a method to identify four types of citations: "Related work, Comparison, Using the work, Extending the work" [23], which is used by Semantic Scholar to determine "Highly Influential Citations" [24]. However, it shows a high correlation with citations.

The field of trend detection analyzes large corpora of works to detect upcoming patterns [25]–[28]. Lopez Belmonte et al. [29] analyzed publications in Machine Learning and Big Data and found exponential growth of publications. They compared the popularity of keywords and the h-index.

**Bibliometrics.** Citations can be aggregated on different levels, e.g., for the papers of one author as the h-index does, or on the journal level, like the journal impact factor (JIF), which is the two-year average ratio of citations to articles published. The JIF is ill-suited for evaluating individual papers by means of the journal it was published in [30]. This is due to the heavy skewness of the distribution of citation counts within journals [31]. The Hirsch-index, usually referred to as the h-index, combines the productivity of an author with the impact of their individual papers. Using the h-index increases robustness compared to simply counting the total number of citations, as few highly cited papers have little effect on the h-index. In addition, there have been efforts to recommend papers and citations [32], [33], predict future citation counts of papers [34]–[37] and the impact of scientists [38]. Such approaches remain challenging and are often domain-specific.

Bornmann and Marx [39] have proposed to expand the bibliometric analysis by not only considering citations but also

references. Following this idea, reference analysis has been used to identify influential references [40].

**Altmetrics.** Online interactions with papers are referred to as altmetrics and are usually available earlier than citations, which gives altmetrics an advantage over bibliometrics [41]. Bornmann and Marx [42] examined if Altmetrics can be used to predict paper quality which was measured through peer assessments and found that both tweets and readers do, with the latter having a stronger relationship. Lamb et al. [43] showed that the Altmetric Attention Score is a predictor of the citations of a paper in ecology and conservation. Zavrel et al. [44] clustered papers released at the International Conference on Machine Learning (ICML) in 2022 and calculated a score for their impact. They used Twitter mentions, citations, the authors' average h-index, and an award for outstanding papers rewarded by the conference itself. They claim to do a "simple combination of these four scores to calculate an impact score" [44] but do not reveal the formula. Färber analyzed GitHub repositories of papers, mostly from the field of AI, and found a power-law distribution of stars and forks [45]. While Haustein et al. claim that "Altmetrics measures scientific impact based on online references and activity" [46], many disagree with equating altmetrics with impact. For example, Sugimoto states that "attention is not impact" and calls online interaction with scientific works "attention" [47]. Altmetrics might reflect broader or societal impact [41].

### B. Dataset Analysis

Bogdoll et al. [5] gathered metadata of over 200 datasets in the field of autonomous driving. Similarly, Färber and Lamprecht released the data set knowledge graph, which is a collection of over 2,000 datasets with added metadata [48]. D'Ulizia et al. [49] analyzed the metadata of datasets for fake news detection. Utamachant and Anutariya [50] analyzed the datasets of Thailand's national open data portal, but relied on domain experts to assess impact. Nguyen and Weller proposed FAIRnets, a service to search for neural networks and their related datasets [51] published on GitHub. They build upon the Findable, Accessible, Interoperable, Reusable (FAIR) principles [52], which "put specific emphasis on enhancing the ability of machines to automatically find and use the data" [52]. Khan et al. [53] analyzed datasets from the Global Biodiversity Information Facility (GBIF) which publishes datasets with a DOI and indexes datasets in biodiversity. They promote data standards and the reuse of datasets [54] as well as accompanying publications, which they call "data papers", that describe a dataset thoroughly [55]. Khan et al. [53] report a strong correlation between dataset download numbers and citation counts, and suggest that downloads and citations signify a similar kind of impact. They also find correlations between altmetrics and citations. Moreover, they question whether every citation means the usage of a dataset and point to differences in citation behavior. Färber et al. proposed an approach to find methods and datasets which authors actually used when citing the related paper [56]. However, unrealistically few dataset usages were identified.
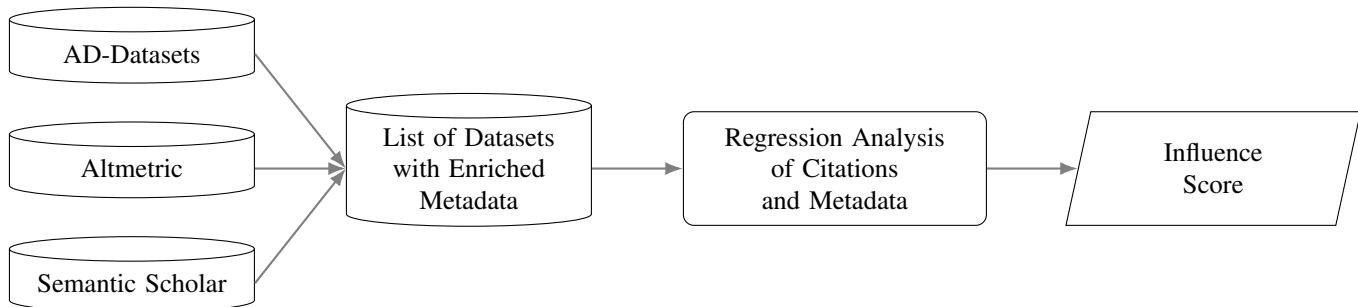
Fig. 2. Overview: First, we collect data from various sources and combine them to a single list of datasets. Based on this, we perform a regression analysis to determine which metadata correlate with future prediction counts. Based on the metadata, we compute our Influence Score (IS).

## III. REGRESSION ANALYSIS

Here, we first introduce our taxonomy of terms related to the assessment of datasets. Subsequently, we introduce our data sources. Based on these, we describe the regression analysis of citations and metadata. In Section IV, we present the resulting Influence Score. Figure 2 gives an overview over this process.

### A. Taxonomy

As became clear in Section II, no common language for specific aspects in the domain has evolved yet. Thus, we introduce a taxonomy to clearly describe different aspects with respect to the development of a dataset or paper. As general terms for this, we utilize success, progress, performance, or potential. For concrete aspects, we establish the following terms, where each one can be applied to any single paper:

**Impact:** We use the number of citations to measure the scientific impact of a paper, which is common in Scientometrics, but not without criticism [19].

**Attention:** The online reception, such as tweets and Wikipedia articles mentioning a paper, represents the attention by researchers and the public.

**Influence:** We refer to the resulting score of our proposed method, which combines a multitude of aspects, as the influence, or IS, of a dataset. We deem this term appropriate for any method that goes beyond purely impact-based assessment.

### B. Data Sources and Selection

We used three sources for our data: ad-datasets.com [6], the Semantic Scholar Academic Graph API [13], and altmetric.com [57]. Based on the DOI and arXiv-Id from ad-datasets, we automatically extracted the metadata of papers from Semantic Scholar and altmetric.com. Based on these papers, all of which describe datasets, we performed data exploration, regression, and the computation of the IS.

**AD-Datasets:** This web tool offers an overview of over 200 data sets in AD [5]. It includes a detailed breakdown of most dataset entries by 20 different meta categories, provided by the authors and the research community. This way, relations between datasets, accompanying papers and further metadata are available. The underlying data is stored in the JSON format and can be accessed accordingly. In this

work, we utilize the $n_{frames}$ and $n_{sensors}$ metadata, which indicate the size of datasets in different dimensions, which is a potential aspect of the relevance of a dataset.

**Altmetric:** We used the API by altmetric.com [57], which provides insight into online attention and readership. These properties are provided by the following categories:
**Attention Score:** The $aas_{curr}$ aggregates different sources into a single score [58]. It is a weighted count of different online sources. For example, the weight for a reference on Wikipedia is 3, while Twitter and Reddit mentions are both weighted with 0.25. Unfortunately, the history of this score is only provided for the most recent year.
**Attention Score after three months:** The $aas_{3m}$ is the percentile of the papers' Attention Score three months after publication. The percentile is calculated in comparison to papers that have been released at a similar time.
**Readers:** The number of people $n_{readers}$ that have saved a paper in their reference management software. Reading a paper is less significant than citing it, but the number of readers might imply interest in a paper early on. The number of readers is provided individually for multiple reference management services, which we sum into a single count for online readers. Altmetric.com cannot verify the number of readers, thus, it is not included in the attention score. This is a relevant attribute, as it decouples the metrics. However, there are no historic data available.

**Semantic Scholar:** For every accompanying paper of a dataset, we pulled data from Semantic Scholar. Sometimes, multiple datasets are described in the same paper, which will lead to the same information for those datasets. We extracted the following nested data:
- List of referenced papers, including for each a list of all citing papers and the year of citation.
- List of authors and their respective publications, including for each publication a list of citing papers and the year of citation.
- List of citing papers, including for each a list of citing papers and the year of citation.

Wherever possible, we collected associated timestamps, including the publication year $a_{pub}$ of each paper. The first

two categories, while dynamic, are directly available. We use the citations of references as a measure of the impact of references. Having impactful references might indicate that a paper is covering popular topics within AD or that the authors are knowledgeable in the field.

The performance of authors can be estimated by evaluating their paper count and how many citations they have received, which becomes only meaningful over time. As discussed earlier, not every citation means usage of a dataset. While it would have been interesting to take into account, in which section a paper has been cited, this data was not available for most papers. Based on the $n_{cit3}$ citations from the previous three years, citations of works that cited a dataset signify the value created by working with the dataset, which is why we included those.

A critical aspect of the collected data is that oftentimes, no historic information was available. Also, oftentimes, data was not available due to limitations, e.g., Altmetric is incompatible with DOIs from IEEE publications, which are common in the fields of robotics, autonomous driving, and machine learning. Similarly, Ravenscroft et al. [15] expressed concerns about Altmetric, as they were unable to find 40 % of the papers they analyzed, all from the field of computer science.

### C. Data Aggregation

To further utilize the raw data we collected, we aggregated some of it with the aim to assemble a finite list of features that describe a dataset. We aggregated some of our data sources using the concept of the h-index formula, as it is widely known, transparent, and easy to reproduce. In order to analyze smaller timespans, we deviated from the typical 5-year duration and calculated multiple 3-year indexes ourselves.

For authors, we applied the h3-index for each individual. We then aggregated the h-indices of all authors of a paper via the arithmetic mean in $aut_{\mu h3}$. Respectively, we applied the h-index formula to references and citations. For the references of a paper, the $ref_{h3}$ is calculated identically to the way it is utilized for authors. Just like an author has papers with citations, a paper has references with citations. A high h-index for references would signify that several of the referenced papers gained lots of attraction. We also applied the h3-index formula to the citations and their citations to get the h3-index of citations $cit_{h3}$, following Schubert et al. [59]. The final list of all extracted and calculated features can be found in Table I.

### D. Cluster Analysis and Regression Setup

We evaluated our computed features with respect to their ability to predict future citations. Therefor, we performed linear regression. For this, we first computed clusters of the datasets to determine a meaningful time horizon. Subsequently, we defined our regression setup.

**Cluster Analysis:** To show that there are meaningful variations between clusters, we looked at the impact of papers for up to 2 years after publication in a journal or conference
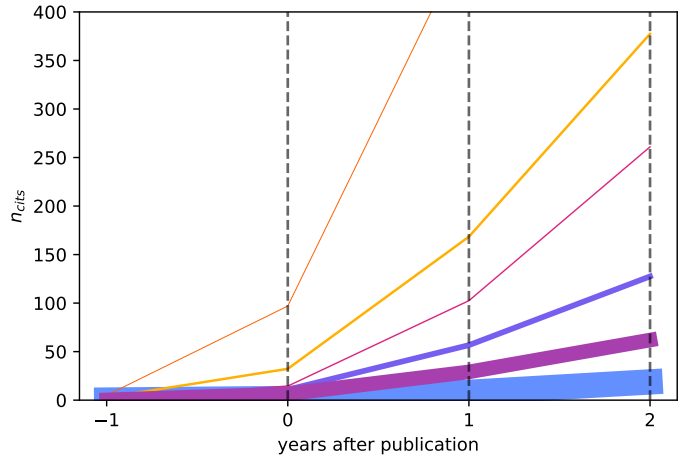


Fig. 3. Development of the number of citations for publication-clusters over a dynamic 3-year window. Papers are clustered based on similar performance. Semantic Scholar also tracks citations of pre-prints, which leads to citations prior to the publication date of the final work.

proceedings. As visualized in Fig. 3, clear clusters are visible, where line-thickness indicates cluster size. For k-means clustering, we used six clusters based on the elbow plot, which shows which additional cluster provides a non-marginal reduction of the total variation within clusters. The growth of citation counts behaves exponentially for the top performing works. A clear differentiation between all clusters becomes apparent already after one year, which we chose as the time horizon for the regression. This allowed us to include more recent papers, which would have been excluded otherwise due to their missing track record of citations.

**Regression Setup:** As independent variables, we included the features $n_{sensors}$, $a_{pub}$, $ref_{h3}$, $aut_{\mu h3}$, $n_{cit3}$, and $aas_{3m}$, as shown in Table I, in order to estimate the citation count after one year. Preliminary data exploration suggested a curvilinear relationship between $aas_{3m}$ and the number of citations. Therefore, a quadratic term was added. All predictors were standardized by subtracting the mean and dividing by the standard deviation prior to the analysis. The feature $n_{cit3}$ was log(x+1)-transformed to ensure a normal distribution of the residuals, which are the error terms of the regression.

For the regression, we were able to utilize 111 datasets, as values for all included features were available, and they had been released at least one year prior. Residuals and collinearity, the ability to linearly predict one independent variable with other independent variables, were checked. The collinearity was quantified through the variance inflation factor of each regressor which all were lower than three. We performed the Breusch-Pagan and White test for heteroskedasticity, which is the inconsistency of the variance of residuals at different levels of the dependent variable. Both tests indicated that we do not have sufficient evidence for the presence of heteroskedasticity. Still, robust standard errors were used to ensure the standard errors are calculated correctly in the presence of heteroskedasticity which at worst leads to standard errors being estimated larger.

| Feature | Description | Availability | Standardized | Log(x+1) | Influence Score | Source |
|---|---|---|---|---|---|---|
| $n_{frames}$ | Number of frames in the dataset | At publication | – | – | ✓ | AD-Datasets [6] |
| $n_{sensors}$ | Number of sensor types | At publication | ✓ | ✗ | ✓ | AD-Datasets [6] |
| $a_{pub}$ | Year of publication | At publication | ✓ | ✗ | ✗ | Semantic Scholar [13] |
| $ref_{h3}$ | 3 year h-index of references | At publication | ✓ | ✗ | ✓ | Semantic Scholar [13] |
| $aut_{\mu h3}$ | Mean 3 year h-index of authors papers | At publication | ✓ | ✗ | ✓ | Semantic Scholar [13] |
| $n_{cit3}$ | Number of citations within past 3 years | Anytime after publication | ✓ | ✓ | ✓ | Semantic Scholar [13] |
| $cit_{h3}$ | 3 year h-index of citations | >3 years after publication | – | – | ✓ | Semantic Scholar [13] |
| $aas_{curr}$ | Altmetric Attention Score | Anytime after publication | – | – | ✓ | Altmetric [57] |
| $aas_{3m}$ | Altmetric Attention Score at 3 mos | After 3 months | ✓ | ✗ | ✗ | Altmetric [57] |
| $n_{readers}$ | Number of readers | Anytime after publication | – | – | ✓ | Altmetric [57] |

We chose not to include $n_{frames}$ for the regression because numerous of the datasets did not contain this meta-information. However, we examined a model in which the feature was included, which did not lead to new findings.

*E. Regression Analysis*

With the explained regression setup, we were now interested in finding statistically significant predictor variables for the citation count at the end of the year after publication.

The $aas_{3m}$ and $aas_{3m}^2$ were positively related to the number of citations and both relationships were significant at <0.0001. Both coefficients were positive. All other features were not significantly related to the number of citations. The results are reported in Table II.

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| $ref_{h3}$ | 0.0987 | 0.103 | 0.96 | 0.337 | -0.103 | 0.3 |
| $aut_{\mu h3}$ | 0.071 | 0.086 | 0.824 | 0.41 | -0.098 | 0.24 |
| $a_{pub}$ | 0.0281 | 0.084 | 0.337 | 0.736 | -0.136 | 0.192 |
| $n_{sensors}$ | 0.1383 | 0.108 | 1.287 | 0.198 | -0.072 | 0.349 |
| $aas_{3m}$ | 0.803 | 0.116 | 6.895 | **0** | 0.575 | 1.031 |
| $aas_{3m}^2$ | 0.3375 | 0.072 | 4.72 | **0** | 0.197 | 0.478 |
| $intercept$ | 2.6402 | 0.117 | 22.48 | 0 | 2.41 | 2.87 |

Since only one feature showed a relationship with the number of citations, we do not consider a stable prediction of citations possible with the available data. In order to still perform an early evaluation of datasets, in the following we present our Influence Score (IS).

## IV. INFLUENCE SCORE

We propose the Influence Score (IS), which includes a variety of features that are available early on. These are weighted dynamically in order to receive an indication of the relative performance of any given dataset at any given time. The calculation compares each dataset with all existing ones from the domain, so that relative differences and trends are immediately recognizable.

Percentiles are used to allow relative scoring within the surrounding group of datasets. The datasets roughly follow a normal distribution in their IS scores. As shown in Table I, we utilize eight different features for the IS: $n_{frames}$, $n_{sensors}$, $ref_{h3}$, $aut_{\mu h3}$, $n_{cit3}$, $cit_{h3}$, $aas_{curr}$ and $n_{readers}$. This way, we consider more than just the citations, but do not exclude them: If early citations are already available, they become a meaningful part of the score, as the relation to other datasets of the peer group is relevant. This way, citation velocity is included. The IS is defined as follows:

$$IS(paper) = 1/n * \sum_{i=0}^{n} percentile(feature_i) \quad (1)$$

where:

$i$ = Feature Index
$n$ = Number of available features

Only features, which are available, are dynamically included in the IS. As we used percentiles of each feature to facilitate the understanding of the features, common issues are mitigated. E.g., typical feature values change over time: For example, with the growth of AD, the $n_{cit}$ value of a paper today is likely higher than a decade ago, which becomes clearly visible in Figure 1. Furthermore, commonly observed values for features might differ between different fields. This helps people who are not familiar with AD or the features to easily assess if the score a dataset achieved is high or low.

*A. Qualitative Demonstration*

To showcase the IS, we compare exemplary the development of the five most and least cited papers with a latest publication in 2019, by their IS and visualize the results in Fig. 4. It becomes clearly visible, that the two groups are easily distinguishable by their IS, but also that differences within the groups are visible.

The individual features show different pictures: For $ref_{h3}$, also papers with only a few citations can have meaningful references in their works. $n_{cit3}$ and $cit_{h3}$ only confirm what was known by our data selection, as we selected the datasets by citation count. $aut_{\mu h3}$ shows, how successful datasets can also boost personal careers, as some authors became professors and remained active in their field. $n_{sensors}$ and $n_{frames}$ show
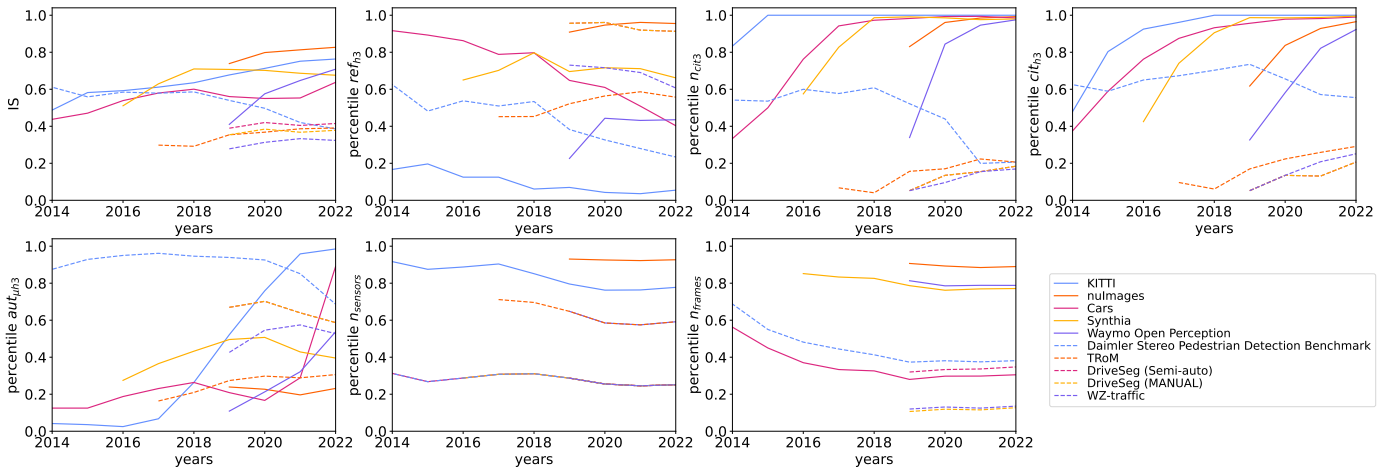
Fig. 4. Influence Score and individual features for different datasets. We show exemplary results for the five best and worst performing datasets of all time, measured by citations, with a latest release in 2019 for historical data. We also show six individual features of the IS, where historical data was available.

rather static results, with a trend towards larger datasets being more successful.

### B. Quantitative Demonstration

In order to show the quantitative performance of the IS, we showcase all datasets released in 2022 in a detailed overview in Table III. Such a pre-filtering process is useful in order to explore novel datasets. Here, it becomes clear that the IS captures a wide variety of different aspects of a dataset. Of particular interest is the fact that even low-performing datasets can lead in certain features. Thus, if a researcher is interested in certain aspects of a dataset, they can simply focus on the features they are interested in and omit the others, which enables less-known datasets to be discovered and used. Figure 5 shows an overview of the IS distributions.
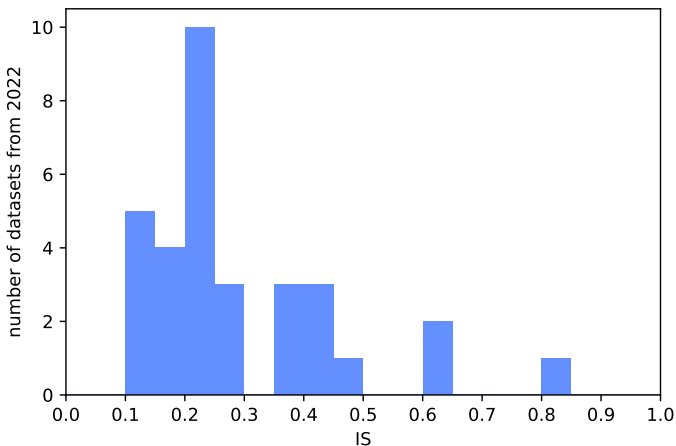


Fig. 5. Distribution of the Influence Score (IS) of all datasets from 2022.

### V. CONCLUSION

In this paper, we addressed the lack of knowledge with respect to the scientific impact, attention, and influence of datasets in robotics. Our focus was on an early assessment of datasets, given a flood of new datasets published every year. We analyzed impact measured by citations and evaluated relations of metadata and features which we extracted from multiple online sources. Our regression analysis showed no strong relation between future citations and our selected features. Subsequently, we presented our developed Influence Score (IS). This score utilizes a set of eight features to assess any dataset also early on. This is based on an analysis within the peer group of all datasets, which allows for the early detection of relative trends.

Our work contributes to a better understanding of datasets, which enables researchers to find and assess published datasets in the domain of autonomous driving without the need of waiting for a track record of citations.

**Limitations and Outlook**: For our work, we evaluated the paper accompanying the dataset assuming that the paper is a good representation of the dataset. When measuring scientific impact through citations, we think this holds because the paper is actually the cited scientific work. However, not every citation might be meaningful, positive, or indicate the usage of a dataset. Ideally, large language models could evaluate if a dataset is actually used, if cited. Khan et al. [53], who analyzed datasets in biodiversity, suggested that the correlation between the number of downloads and citations signifies that these two measures are comparable representations of impact. However, in the domain of AD, download numbers are typically not available, but this might change. As some datasets are presented in the same paper, a further decoupling of accompanying papers and the respective datasets would be helpful. We found, that the quality and availability of metadata in AD provided by the creators of datasets varies strongly. Thus, standards should be established [90]. While we focussed on dataset and paper specific features for this work, we are also interested in the venue or journal of publication, which can be considered as an additional feature in future work.

TABLE III
INFLUENCE SCORE AND FEATURES FOR DATASETS RELEASED IN 2022. SORTED BY IS, TOP 3 FEATURES BOLD.

| | IS | $n_{cit3}$ | $cit_{h3}$ | $ref_{h3}$ | $aut_{\mu h3}$ | $n_{frames}$ | $n_{sensors}$ | $aas_{curr}$ | $n_{readers}$ |
|---|---|---|---|---|---|---|---|---|---|
| Waymo Block-NeRF [60] | 0.82 | **0.7** | **0.53** | **0.95** | 0.87 | – | – | **1.0** | **0.89** |
| SHIFT [61] | 0.62 | 0.25 | 0.2 | **0.99** | **0.88** | **0.94** | **0.77** | 0.65 | **0.42** |
| Street Hazards [62] | 0.62 | **0.67** | **0.58** | 0.83 | **0.92** | 0.16 | 0.25 | **0.73** | **0.45** |
| KITTI-360-APS [63] | 0.48 | 0.23 | 0.06 | 0.68 | 0.67 | 0.56 | 0.25 | **0.97** | 0.21 |
| ScribbleKITTI [64] | 0.45 | 0.2 | 0.15 | 0.83 | **0.97** | 0.32 | 0.25 | 0.63 | 0.02 |
| BDD100K-APS [63] | 0.42 | 0.23 | 0.06 | 0.68 | 0.67 | 0.1 | 0.25 | **0.97** | 0.21 |
| Ithaca365 [65] | 0.4 | 0.15 | 0.15 | 0.68 | 0.22 | **0.82** | **0.58** | 0.53 | 0.26 |
| CODA [66] | 0.39 | 0.2 | 0.15 | 0.75 | 0.36 | – | 0.25 | 0.53 | 0.33 |
| Rope3D [67] | 0.37 | 0.15 | 0.15 | 0.83 | 0.42 | 0.53 | **0.58** | 0.25 | 0.25 |
| Comma2k19 LD [68] | 0.37 | 0.12 | 0.15 | 0.71 | 0.56 | – | – | 0.64 | 0.02 |
| RoadSaW [69] | 0.29 | 0.09 | 0.06 | 0.27 | 0.19 | **0.83** | 0.25 | – | – |
| K-Radar [70] | 0.26 | 0.09 | 0.06 | 0.47 | 0.11 | 0.44 | **0.93** | 0.4 | 0.26 |
| CARLA-WildLife [71] | 0.26 | 0.03 | 0.06 | **0.92** | 0.12 | – | 0.25 | 0.34 | 0.08 |
| AugKITTI [72] | 0.25 | 0.03 | 0.06 | 0.88 | 0.16 | – | – | 0.25 | 0.1 |
| WildDash 2 [73] | 0.24 | 0.17 | 0.2 | 0.52 | 0.08 | – | 0.25 | – | – |
| MONA [74] | 0.23 | 0.03 | 0.06 | 0.36 | 0.48 | – | 0.25 | – | – |
| Street Obstacle Sequences [71] | 0.23 | 0.03 | 0.06 | **0.92** | 0.12 | 0.07 | 0.25 | 0.34 | 0.08 |
| HDBD [75] | 0.22 | 0.03 | 0.06 | 0.16 | 0.64 | – | **0.58** | – | – |
| GLARE [76] | 0.22 | 0.03 | 0.06 | 0.47 | 0.29 | – | – | 0.42 | 0.06 |
| Boreas [77] | 0.22 | **0.29** | **0.25** | 0.14 | 0.2 | – | – | – | – |
| Autonomous Platform Inertial [78] | 0.21 | 0.2 | **0.25** | 0.36 | 0.03 | – | – | – | – |
| aiMotive [79] | 0.2 | 0.03 | 0.06 | 0.33 | 0.04 | 0.41 | **0.93** | 0.44 | 0.13 |
| CarlaScenes [80] | 0.2 | 0.03 | 0.06 | 0.52 | 0.2 | – | **0.77** | – | – |
| LUMPI [81] | 0.19 | 0.09 | 0.06 | 0.02 | 0.07 | 0.74 | **0.58** | – | – |
| A9 [82] | 0.19 | 0.25 | **0.25** | 0.14 | 0.1 | – | **0.58** | 0.29 | 0.12 |
| Amodal Cityscapes [83] | 0.19 | 0.09 | 0.06 | 0.27 | 0.43 | 0.11 | 0.25 | 0.29 | 0.08 |
| R-U-MAAD [84] | 0.16 | 0.03 | 0.06 | 0.33 | 0.35 | – | 0.25 | 0.15 | 0.06 |
| TJ4DRadSet [85] | 0.15 | 0.12 | 0.15 | 0.14 | 0.05 | – | – | 0.42 | 0.02 |
| OpenMPD [86] | 0.14 | 0.18 | 0.15 | 0.02 | 0.07 | 0.27 | **0.58** | – | – |
| I see you [87] | 0.12 | 0.03 | 0.06 | 0.09 | 0.01 | – | – | 0.44 | 0.08 |
| SceNDD [88] | 0.11 | 0.09 | 0.06 | 0.16 | 0.19 | – | – | 0.15 | 0.02 |
| exiD [89] | 0.11 | 0.12 | 0.06 | 0.02 | 0.24 | – | 0.25 | – | – |

## REFERENCES

[1] Waymo, "Waymo One," https://waymo.com/waymo-one/, 2022, accessed: 2022-12-14.

[2] Waabi, "Introducing the Waabi Driver," https://waabi.ai/introducing-the-waabi-driver/, 2022, accessed: 2022-12-14.

[3] World Economic Forum, "3 Ways Autonomous Farming is Driving a New Era of Agriculture," https://www.weforum.org/agenda/2022/01/autonomous-farming-tractors-agriculture/, 2022, accessed: 2022-12-14.

[4] Amazon News, "Meet Amazon's First Fully Autonomous Mobile Robot," https://www.youtube.com/watch?v=AmmEbYkYfHY, 2022, accessed: 2022-12-14.

[5] D. Bogdoll, F. Schreyer, and J. M. Zöllner, "AD-Datasets: A Meta-Collection of Data Sets for Autonomous Driving," in *International Conference on Vehicle Technology and Intelligent Transport Systems*, 2022.

[6] D. Bogdoll, "ad-datasets," https://ad-datasets.com/, 2022, accessed: 2022-09-02.

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research*, 2013.

[8] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in Perception for Autonomous Driving: Waymo Open Dataset," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[9] S. Qiao, Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen, "VIP-DeepLab: Learning Visual Perception With Depth-Aware Video Panoptic Segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[10] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection using deep neural networks with images captured through a smartphone," *arXiv:1801.09454*, 2018.

[11] F. Codevilla, E. Santana, A. M. Lopez, and A. Gaidon, "Exploring the Limitations of Behavior Cloning for Autonomous Driving," in *IEEE/CVF International Conference on Computer Vision*, 2019.

[12] A. Lehner, S. Gasperini, A. Marcos-Ramiro, M. Schmidt, M.-A. N. Mahani, N. Navab, B. Busam, and F. Tombari, "3D-VField: Adversarial Augmentation of Point Clouds for Domain Generalization in 3D Object Detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[13] Allen Institute for AI, "Semantic Scholar Academic Graph API," https://www.semanticscholar.org/product/api, accessed: 2022-09-17.

[14] P. Chellappandi and C. S. Vijayakumar, "Bibliometrics, Scientometrics, Webometrics/Cybermetrics, Informetrics and Altmetrics – An Emerging

Field in Library and Information Science Research," *Shanlax International Journal of Education*, 2018.

[15] J. Ravenscroft, M. Liakata, A. Clare, and D. Duma, "Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements," *PLOS ONE*, 2017.

[16] J. E. Hirsch, "An index to quantify an individual's scientific research output," *National Academy of Sciences*, 2005.

[17] D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, and I. Rafols, "Bibliometrics: The Leiden Manifesto for research metrics," *Nature*, 2015.

[18] L. Leydesdorff, L. Bornmann, J. A. Comins, and S. Milojević, "Citations: Indicators of Quality? The Impact Fallacy," *Frontiers in Research Metrics and Analytics*, 2016.

[19] L. Bornmann and H.-D. Daniel, "What do citation counts measure? A review of studies on citing behavior," *Journal of Documentation*, 2008.

[20] D. D. S. Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American Society for Information Science*, 1976.

[21] D. S. Chawla, "Men cite themselves more than women do," *Nature*, 2016.

[22] J. Mingers and L. Leydesdorff, "A review of theory and practice in scientometrics," *European Journal of Operational Research*, 2015.

[23] M. Valenzuela, V. A. Ha, and O. Etzioni, "Identifying meaningful citations," in *AAAI Workshop: Scholarly Big Data*, 2015.

[24] Semantic Scholar, "What are highly influential citations?" https://www.semanticscholar.org/faq#influential-citations, 2022, accessed: 2022-12-18.

[25] M.-H. Le, T.-B. Ho, and Y. Nakamori, "Detecting Emerging Trends from Scientific Corpora," *International Journal of Knowledge and Systems Sciences*, 2005.

[26] A. A. Salatino, "Early Detection and Forecasting of Research Trends," in *Doctoral Consortium Co-located with the International Semantic Web Conference*, 2015.

[27] M. Farber and A. Jatowt, "Finding Temporal Trends of Scientific Concepts," in *International Workshop on Bibliometric-enhanced Information Retrieval*, 2019.

[28] M. Farber, C. Nishioka, and A. Jatowt, "ScholarSight: Visualizing Temporal Trends of Scientific Concepts," in *ACM/IEEE Joint Conference on Digital Libraries*, 2019.

[29] J. López Belmonte, A. Segura-Robles, A.-J. Moreno-Guerrero, and M. E. Parra-González, "Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science," *Symmetry*, 2020.

[30] F. M. Paulus, N. Cruz, and S. Krach, "The Impact Factor Fallacy," *Frontiers in Psychology*, 2018.

[31] E. Callaway, "Beat it, impact factor! Publishing elite turns against controversial metric," *Nature*, 2016.

[32] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: a literature survey," *International Journal on Digital Libraries*, 2016.

[33] M. Färber, A. Thiemann, and A. Jatowt, "To cite, or not to cite? detecting citation contexts in text," in *European Conference on Information Retrieval*, 2018.

[34] L. Fu and C. Aliferis, "Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature," *Scientometrics*, 2010.

[35] N. Pobiedina and R. Ichise, "Citation count prediction as a link prediction problem," *Applied Intelligence*, 2016.

[36] A. Ma, Y. Liu, X. Xu, and T. Dong, "A deep-learning based citation count prediction model with paper metadata semantic features," *Scientometrics*, 2021.

[37] M. Li, J. Xu, B. Ge, J. Liu, J. Jiang, and Q. Zhao, "A Deep Learning Methodology for Citation Count Prediction with Large-scale Biblio-Features," in *2019 IEEE International Conference on Systems, Man and Cybernetics*, 2019.

[38] E. Bütün and M. Kaya, "Predicting Citation Count of Scientists as a Link Prediction Problem," *IEEE Transactions on Cybernetics*, 2020.

[39] L. Bornmann and W. Marx, "The proposal of a broadening of perspective in evaluative bibliometrics by complementing the times cited with a cited reference analysis," *Journal of Informetrics*, 2013.

[40] A. W. K. Yeung, M. G. Georgieva, A. G. Atanasov, and N. T. Tzvetkov, "Monoamine Oxidases (MAOs) as Privileged Molecular Targets in Neuroscience: Research Literature Analysis," *Frontiers in Molecular Neuroscience*, 2019.

[41] L. Bornmann, "Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics," *Journal of Informetrics*, 2014.

[42] L. Bornmann and R. Haunschild, "Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data," *PLOS ONE*, 2018.

[43] C. T. Lamb, S. L. Gilbert, and A. T. Ford, "Tweet success? Scientific communication correlates with increased citations in Ecology and Conservation," *PeerJ*, 2018.

[44] J. Zavrel, "Can AI help us understand ICML 2022?" https://www.zeta-alpha.com/post/can-ai-help-us-understand-icml-2022, 2022, accessed: 2022-08-18.

[45] M. Färber, "Analyzing the github repositories of research papers," in *ACM/IEEE Joint Conference on Digital Libraries*, 2020.

[46] S. Haustein, I. Peters, J. Bar-Ilan, J. Priem, H. Shema, and J. Terliesner, "Coverage and adoption of altmetrics sources in the bibliometric community," *Scientometrics*, 2014.

[47] C. Sugimoto, ""Attention is not Impact" and Other Challenges for Altmetrics," https://www.wiley.com/network/researchers/promoting-your-article/attention-is-not-impact-and-other-challenges-for-altmetrics, 2015, accessed: 2022-07-04.

[48] M. Färber and D. Lamprecht, "The data set knowledge graph: Creating a linked open data source for data sets," *Quantitative Science Studies*, 2021.

[49] A. D'Ulizia, M. C. Caschera, F. Ferri, and P. Grifoni, "Fake news detection: A survey of evaluation datasets," *PeerJ Computer Science*, 2021.

[50] P. Utamachant and C. Anutariya, "An Analysis of High-Value Datasets: A Case Study of Thailand's Open Government Data," in *International Joint Conference on Computer Science and Software Engineering*, 2018.

[51] A. Nguyen and T. Weller, "FAIRnets Search - A Prototype Search Service to Find Neural Networks," in *International Conference on Semantic Systems Posters&Demos Track*, 2019.

[52] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The fair guiding principles for scientific data management and stewardship," *Scientific Data*, 2016.

[53] N. Khan, M. Thelwall, and K. Kousha, "Measuring the impact of biodiversity datasets: Data reuse, citations and altmetrics," *Scientometrics*, 2021.

[54] Global Biodiversity Information Facility, "What is gbif?" https://www.gbif.org/what-is-gbif, 2022, accessed: 2022-09-19.

[55] ——, "Data papers," https://www.gbif.org/data-papers, 2022, accessed: 2022-09-19.

[56] M. Färber, A. Albers, and F. Schüber, "Identifying used methods and datasets in scientific publications," in *AAAI Workshop: Scientific Document Understanding*, 2021.

[57] altmetric.com, "Altmetric," https://www.altmetric.com/, accessed: 2022-09-17.

[58] Williams, "The Altmetric score is now the Altmetric Attention Score," 2016.

[59] A. Schubert, "Using the h-index for assessing single publications," *Scientometrics*, 2008.

[60] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. P. Srinivasan, J. T. Barron, and H. Kretzschmar, "Block-NeRF: Scalable Large Scene Neural View Synthesis," *arXiv:2202.05263*, 2022.

[61] T. Sun, M. Segu, J. Postels, Y. Wang, L. Van Gool, B. Schiele, F. Tombari, and F. Yu, "SHIFT: A Synthetic Driving Dataset for Continuous Multi-Task Domain Adaptation," *arXiv:2206.08367*, 2022.

[62] D. Hendrycks, S. Basart, M. Mazeika, A. Zou, J. Kwon, M. Mostajabi, J. Steinhardt, and D. Song, "Scaling Out-of-Distribution Detection for Real-World Settings," *arXiv:1911.11132*, 2022.

[63] R. Mohan and A. Valada, "Amodal Panoptic Segmentation," *arXiv:2202.11542*, 2022.

[64] O. Unal, D. Dai, and L. Van Gool, "Scribble-Supervised LiDAR Semantic Segmentation," *arXiv:2203.08537*, 2022.

[65] C. A. Diaz-Ruiz, Y. Xia, Y. You, J. Nino, J. Chen, J. Monica, X. Chen, K. Luo, Y. Wang, M. Emond, W.-L. Chao, B. Hariharan, K. Q. Weinberger, and M. Campbell, "Ithaca365: Dataset and Driving Perception under Repeated and Challenging Weather Conditions," *arXiv:2208.01166*, 2022.

[66] K. Li, K. Chen, H. Wang, L. Hong, C. Ye, J. Han, Y. Chen, W. Zhang, C. Xu, D.-Y. Yeung, X. Liang, Z. Li, and H. Xu, "CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving," *arXiv:2203.07724*, 2022.

[67] X. Ye, M. Shu, H. Li, Y. Shi, Y. Li, G. Wang, X. Tan, and E. Ding, "Rope3D: TheRoadside Perception Dataset for Autonomous Driving and Monocular 3D Object Detection Task," *arXiv:2203.13608*, 2022.

[68] T. Sato and Q. A. Chen, "Towards Driving-Oriented Metric for Lane Detection Models," *arXiv:2203.16851*, 2022.

[69] K. Cordes, C. Reinders, P. Hindricks, J. Lammers, B. Rosenhahn, and H. Broszio, "RoadSaW: A Large-Scale Dataset for Camera-Based Road Surface and Wetness Estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022.

[70] D.-H. Paek, S.-H. Kong, and K. T. Wijaya, "K-Radar: 4D Radar Object Detection for Autonomous Driving in Various Weather Conditions," *arXiv:2206.08171*, 2022.

[71] K. Maag, R. Chan, S. Uhlemeyer, K. Kowol, and H. Gottschalk, "Two Video Data Sets for Tracking and Retrieval of Out of Distribution Objects," *arXiv:2210.02074*, 2022.

[72] Y. Pan, F. Xie, and H. Zhao, "Understanding the Challenges When 3D Semantic Segmentation Faces Class Imbalanced and OOD Data," *arXiv:2203.00214*, 2022.

[73] O. Zendel, M. Schorghuber, B. Rainer, M. Murschitz, and C. Beleznai, "Unifying Panoptic Segmentation for Autonomous Driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[74] L. Gressenbuch, K. Esterle, T. Kessler, and M. Althoff, "MONA: The Munich Motion Dataset of Natural Driving," in *IEEE International Conference on Intelligent Transportation Systems*, 2022.

[75] Y. Qiu, C. Busso, T. Misu, and K. Akash, "Incorporating Gaze Behavior Using Joint Embedding With Scene Context for Driver Takeover Detection," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022.

[76] N. Gray, M. Moraes, J. Bian, A. Tian, A. Wang, H. Xiong, and Z. Guo, "GLARE: A Dataset for Traffic Sign Detection in Sun Glare," *arXiv:2209.08716*, 2022.

[77] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. K. Leung, A. P. Schoellig, and T. D. Barfoot, "Boreas: A Multi-Season Autonomous Driving Dataset," *arXiv:2203.10168*, 2022.

[78] A. Shurin, A. Saraev, M. Yona, Y. Gutnik, S. Faber, A. Etzion, and I. Klein, "The Autonomous Platforms Inertial Dataset," *IEEE Access*, vol. 10, 2022.

[79] T. Matuszka, I. Barton, Á. Butykai, P. Hajas, D. Kiss, D. Kovács, S. Kunsági-Máté, P. Lengyel, G. Németh, L. Pető, D. Ribli, D. Szeghy, S. Vajna, and B. Varga, "aiMotive Dataset: A Multimodal Dataset for Robust Autonomous Driving with Long-Range Perception," *arXiv:2211.09445*, 2022.

[80] A. Kloukiniotis, A. Papandreou, C. Anagnostopoulos, A. Lalos, P. Kapsalas, D.-V. Nguyen, and K. Moustakas, "CarlaScenes: A synthetic dataset for odometry in autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022.

[81] S. Busch, C. Koetsier, J. Axmann, and C. Brenner, "LUMPI: The Leibniz University Multi-Perspective Intersection Dataset," in *IEEE Intelligent Vehicles Symposium*, 2022.

[82] C. Creß, W. Zimmer, L. Strand, V. Lakshminarasimhan, M. Fortkord, S. Dai, and A. Knoll, "A9-Dataset: Multi-Sensor Infrastructure-Based Dataset for Mobility Research," *arXiv:2204.06527*, 2022.

[83] J. Breitenstein and T. Fingscheidt, "Amodal Cityscapes: A New Dataset, its Generation, and an Amodal Semantic Segmentation Challenge Baseline," *arXiv:2206.00527*, 2022.

[84] J. Wiederer, J. Schmidt, U. Kressel, K. Dietmayer, and V. Belagiannis, "A Benchmark for Unsupervised Anomaly Detection in Multi-Agent Trajectories," *arXiv:2209.01838*, 2022.

[85] L. Zheng, Z. Ma, X. Zhu, B. Tan, S. Li, K. Long, W. Sun, S. Chen, L. Zhang, M. Wan, L. Huang, and J. Bai, "TJ4DRadSet: A 4D Radar Dataset for Autonomous Driving," in *IEEE International Conference on Intelligent Transportation Systems*, 2022.

[86] X. Zhang, Z. Li, Y. Gong, D. Jin, J. Li, L. Wang, Y. Zhu, and H. Liu, "OpenMPD: An Open Multimodal Perception Dataset for Autonomous Driving," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, 2022.

[87] H. Quispe, J. Sumire, P. Condori, E. Alvarez, and H. Vera, "I see you: A Vehicle-Pedestrian Interaction Dataset from Traffic Surveillance Cameras," *arXiv:2211.09342*, 2022.

[88] A. Prabu, N. Ranjan, L. Li, R. Tian, S. Chien, Y. Chen, and R. Sherony, "SceNDD: A Scenario-based Naturalistic Driving Dataset," in *IEEE International Conference on Intelligent Transportation Systems*, 2022.

[89] T. Moers, L. Vater, R. Krajewski, J. Bock, A. Zlocki, and L. Eckstein, "The exiD Dataset: A Real-World Trajectory Dataset of Highly Interactive Highway Scenarios in Germany," in *IEEE Intelligent Vehicles Symposium*, 2022.

[90] Google Search Central, "Dataset (dataset, datacatalog, datadownload) structured data," https://developers.google.com/search/docs/appearance/structured-data/dataset, accessed: 2023-01-05.