EuroVis Workshop on Visual Analytics (2022)
M. Angelini and J. Bernard (Editors)

ScrutinAI: A Visual Analytics Approach for the Semantic Analysis of Deep Neural Network Predictions

Elena Haedecke[1,2], Michael Mock[2,1] and Maram Akila[2]

[1]University of Bonn, Germany
[2]Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Germany

Abstract

We present ScrutinAI, a Visual Analytics approach to exploit semantic understanding for deep neural network (DNN) predictions analysis, focusing on models for object detection and semantic segmentation. Typical fields of application for such models, e.g. autonomous driving or healthcare, have a high demand for detecting and mitigating data- and model-inherent shortcomings. Our approach aims to help analysts use their semantic understanding to identify and investigate potential weaknesses in DNN models. ScrutinAI therefore includes interactive visualizations of the model's inputs and outputs, interactive plots with linked brushing, and data filtering with textual queries on descriptive meta data. The tool fosters hypothesis driven knowledge generation which aids in understanding the model's inner reasoning. Insights gained during the analysis process mitigate the "black-box character" of the DNN and thus support model improvement and generation of a safety argumentation for AI applications. We present a case study on the investigation of DNN models for pedestrian detection from the automotive domain.

CCS Concepts

• ***Human-centered computing*** → ***Visual analytics;*** *Interactive systems and tools;* • ***Computing methodologies*** → *Neural networks; Object detection;*

1. Introduction

The advances in the field of machine learning (ML) – and especially deep neural networks (DNNs) – have led to an increasing deployment of DNN solutions in various application areas in the last decade, e.g. object detection [LAE*16] or autonomous driving [WJ22]. However, though DNNs claim several state-of-the-art (SOTA) benchmark results, to assess the reliability and trustworthiness of such models it is not sufficient to compare models only based on performance metrics, e.g. high accuracy. The issue with these common performance metrics is that they do not provide evidence that the model operates safely in its defined operational design domain, which might contain rare situations or corner cases. Furthermore, DNNs generally suffer from a lack of transparency [RBDG20]. Thus, even with a common performance metric assuring a SOTA model, the reason why a DNN model has come to a particular prediction in a specific situation remains hidden inside the "black box" [DVK17] to most extent and prevents from identifying and mitigating any weaknesses. Both combine to a crucial challenge, as the weaknesses of seeming SOTA models can neither be understood nor mitigated, which ultimately results in a lack of trust in the whole DNN application [LPK21]. To tackle this, human semantic understanding and common knowledge of the application domain has to be taken into account for the investigation of DNN predictions [AAAW22], i.e. to analyze models beyond performance indicators. This challenging task neither can be algorithmically automated nor is it feasible for a human analyst to investigate an extensive number of possible input scenarios one by one. It is therefore necessary to present the numerous, often high-dimensional and abstract data and the metrics in a human-understandable form with the possibility to select and define semantically significant dimensions and features for in-depth examination.

Motivated by these challenges, we propose our Visual Analytics (VA) approach *ScrutinAI*, offering: (1) various forms of visualization of the data to represent it in a suitable way, (2) an interactive interface that assists the analyst in the investigation process of image based DNN predictions, (3) a focus on utilizing the semantic understanding to assess potential DNN weaknesses, (4) a knowledge generation workflow, and (5) options for querying and filtering to break down the huge amount of data to comprehensible, interesting data subsets. In addition to common scene- and object-related data, e.g. the pixel-locations and classes of objects, we exploit descriptive meta data to support the generation of semantic hypotheses. This meta data describes coarser, scene-specific details on the one hand, e.g. the sun angle or the weather condition, and finer, object-specific details on the other hand, e.g. the gender of a pedestrian and the color of their clothes. As development of ScrutinAI is still work in progress, we present our interim results with a case study in the context of trustworthy AI for autonomous driving.

DOI: 10.2312/eurova.20221071

2. Related Work

Incorporating human tacit or common knowledge and semantic understanding into a supportive evaluation process of DNNs is mainly related to the research field of VA. It touches also the neighboring field of e**X**plainable **A**rtificial **I**ntelligence (XAI) [LPK21], where methods for model explainability and interpretability [DVK17] are being explored, i.e. mitigating the "opaqueness" of DNN black-box models. In Cook et al. [CT05], VA is described as the "science of analytical reasoning facilitated by interactive visual interfaces", emphasizing the importance of interactive tool support. Sacha et al. [SSS*14] highlight the "structured reasoning process" of VA as it "allows analysts to find evidence in data and gain insights into the problem domain". Recent research focuses on "VA for human-centered ML", i.e. integrating the analyst into the analysis process of DNNs [AAAW22]. Using the human visual capacity, data can be presented in a suitable form of visualization to provide an overview, using different levels of abstraction to aggregate or reduce information and to detect patterns [YKSJ08, SEAKC21]. A common approach of integrating visualizations and VA into the entire ML pipeline is practiced in the field of interactive ML, as reviewed in recent surveys [SKKC19, YCY*21, ERT*17]. While being established in traditional ML areas, where models often are intrinsically interpretable, VA is a relatively new approach to facilitate and support the explainability and interpretability of DNNs [Hou21, HKPC19, CL18, GTS*18]. VA systems or tools provide exploratory workflows that often combine employing XAI methods with a targeted DNN analysis [HKPC19, SKKC19]. With DNN models relying on huge amounts of data, the difficulty is to maintain the scalability for the analyst while still offering enough flexibility for a deep dive analysis and not losing or hiding important information. Several approaches of VA systems exist, that mainly or partly follow this goal, e.g. in the fields of healthcare [XCK*20] or predictive maintenance [PMBT21], systems enabling model understanding [SSSEA20] or systems targeted at non-expert ML engineers [WPB*20]. Tools aimed at expert users often also employ some way of visualizing training progress, model architecture or neuron relations, e.g. [LWLZ17, PDD*22]. (For a more elaborate overview and classification of DNN-analysis related VA tool approaches, we refer to the works of [HKPC19, YCY*21, LWLZ17, CMJK20, KRN21].) Most existing VA tools pay little attention to the semantic context on object-level to assess potential vulnerabilities. For instance, Manifold [ZWM*19] is conceptually similar to our work, but due to its generic approach, it lacks the semantic context needed to study complex DL problems with specific granularity. Exceptions, and most related to our work, are the approaches VASS [HZS*22] and VATLD [GZL*21] that offer proven, highly specialized solutions for analysing the performance of semantic segmentation models from the automotive domain (VASS) and of traffic light detection models (VATLD), respectively. However, they strongly rely on representation learning to provide semantic information, whereas in our approach we use human-defined concepts in the form of a detailed descriptive meta data table. With these intrinsically understandable human annotations, which are more robust than learned representations, we aim to promote the generation of semantic hypotheses. Compared to other approaches, we separate semantic analysis from the study of internal technical model properties to help analysts focus on identifying model weaknesses related to semantic concepts of objects. In contrast to interactive ML, our semantic VA workflow emphasizing human semantic understanding is more focused on offline evaluation of models, which we call *scrutinizing* DNN models.

3. Problem Statement

The performance of a DNN mostly depends on a combination of different dimensions and usually cannot be explained by considering one dimension of the input alone. By relating common performance metrics to human understandable "semantic dimensions" of the input, i.e. semantic concepts defined by domain experts, methods of data visualization can support the human to identify anomalous patterns and systematic weaknesses of the DNN. While exploring several dimensions as possible influence factors, the analyst is faced with two challenges: *(1) Visual Scalability*: The sheer amount of test data hat must be examined to gain meaningful insights usually overwhelms human analysis capacity [JLC19]. *(2) Semantic Scalability*: Identifying and selecting the relevant dimensions that actually influence the network performance among all possible dimensions is far from trivial. To overcome these challenges, tools are needed that support analysts in an efficient way following the principles of VA [YKSJ08, KAF*08] and making use of explicit knowledge [FWR*17] in form of descriptive meta data. The "human-centered semantic analysis" we propose cannot be achieved with multiple separate tools, visualizations, or widgets. These must be embedded in an integral workflow to ultimately use the knowledge created during the analysis to review and improve the models and data under investigation [ALA*18]. In summary, we therefore aim to address the following two main aspects with our approach: (1) tool support through a hypothesis-driven workflow and (2) use of detailed descriptive metadata to support the investigation of DNN weaknesses based on a semantic understanding.

4. Approach

The goal of our approach is to allow the analyst to use the insights gained during the analysis process to find evidence for the origin or absence of certain deficiencies. Improving the model during the incremental analysis can be considered as a building block to an overall argument for a trustworthy DNN model. Throughout the analysis cycle, we consider an *insight* not as the end-result of an analysis, but rather as an important part of the analysis process that, with enough evidence, might be used to generate *knowledge* [SSS*14], or otherwise to create a new hypothesis as starting point for further analysis. For the workflow (see Figure 1), we focused on efficiently supporting the human analyst on utilizing domain knowledge and concentrating on semantic aspects of the model and data. Our workflow is derived from the staged loops of Sacha et al. [SSS*14] and the "Human-Machine Interaction Ring" by Ribarsky et al. [RF16], which itself is based on the model by Sacha et al. but consolidates human and computer domain to a merged cyclic workflow. We combine the essentials of both works and extend them with a feedback loop to report back the insights and knowledge gained during the analysis process to other stakeholders, e.g. the DNN model developers. The entry point of the workflow is to visually and interactively inspect the different data that are presented in a suitable way, providing overview as well as fine-grained information,

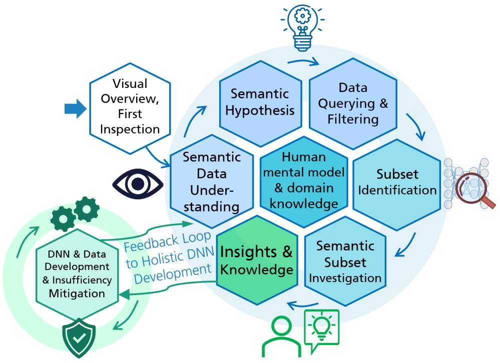

Figure 1: *The central point of the iterative workflow is the semantic understanding of the human with an emphasis on fostering the creation and evaluation of hypotheses. Insights and knowledge are reported back in a feedback loop to facilitate the improvement of the diverse DNN model and data development processes.*

to gain a *Semantic Data Understanding*. During the investigation of data, the analyst formulates *Semantic Hypotheses* based on the understanding of the data, the analyst's mental model and the interesting findings and observations. To find evidence for a certain semantic hypothesis, interesting data points need to be evaluated via *Querying & Filtering* w.r.t. the domain and tacit knowledge. Commonly, this includes relevant *Subset Identification* of close semantic distance, where a common semantic feature leads to the same prediction of the model. During the *Semantic Subset Investigation* the analyst either finds sufficient and strong evidence to verify or falsify a hypothesis, or otherwise continues with a further analysis. The *Insights* during investigation can be turned into *Knowledge* and reported back to the other stakeholders, with the possibility to request more information (e.g. more meta data, new methods, or tool widgets) that aids a further deep dive analysis. The *Feedback* can be used to mitigate DNN insufficiencies, e.g. by re-training or re-fining the DNN or incorporating new mechanisms addressing the insufficiencies, and can further contribute to a trustworthiness argumentation.

5. Scrutinizing DNNs with ScrutinAI

ScrutinAI has been programmed in Python based on open-source libraries with the goal of offering modularity and easy extensibility for DNN developers and expert users. From a comparison of different libraries providing functionalities for interactive visualizations we chose HoloViz [Hol], as it offers the best set of packages with the possibility to integrate common Python libraries and, thus, the desired modularity for our approach. The browser-based interactive application was programmed according to the Model-View-Controller software design pattern, which allows for easy integration or replacement of modules, visualizations and widgets. To support the different building blocks of the holistic DNN model development and data generation processes, ScrutinAI relies on precomputed model predictions and meta data. With *meta data* we define data in a tabular form (e.g. CSV) describing the model's input and output data in human understandable semantic dimensions. For the application of pedestrian detection in autonomous driving, examples are, e.g. its illumination or the age of the pedestrians. The tabular format gives the flexibility to exchange or expand the available meta data without much effort, i.e. using meta data suitable for each specific application and also taking a refinement of meta data during an analysis process into account. The data types for the columns are not restricted, i.e. a mix of numerical, categorical and boolean meta data is possible. Besides structured data, image data of common data types can be integrated.

A (stacked) overview of ScrutinAI is shown in Figure 2 **a)**. In (A), specific splits of a data set can be selected. The user can issue arbitrary textual queries (B) against the meta data table (C) to select interesting subsets of the data. Use case specific metrics and overviews can be included, see (D), which might depend on additional hyperparameters (e.g. confidence threshold). These can be adjusted dynamically to observe the influence of a certain parameter. As humans are much more adopted to a visual interpretation, typical elements such as histograms (E) are provided to resolve distributions along categories. The histogram offers additional grouping by a secondary attribute and automatic binning for supported numerical values. In the autonomous driving example this could be the "visible pixels" of a pedestrian, i.e. unoccluded body parts. For deeper investigations arbitrary numbers of scatter plots (F) can be created. Those allow an investigation of inter-dependency among the attributes and besides visual inspection, see (G), provide correlation coefficients, see (H). Linked brushing of the visual elements (e.g. via lasso and box select tools) allows to narrow down the focus on groups of specific interest, e.g. highly visible pedestrians with low detection performance. In addition to the meta data, (I) provides "raw" access to unstructured data, e.g. images, within the current selection. The input image in our example may be extended with an overlay of other information, e.g. segmentation masks or visualization of bounding boxes in cases of object detection. Slider widgets aid the user in adjusting the transparency of the overlayed image and "scrolling" through the current subset. Panning and zooming options allow for a "drill down" into the images, see (J). For bounding box detection, it is possible to select different boxes to be displayed (e.g. all true positives). Following the idea of linking meta-information to the visible elements, mouse hover functionalities give easily accessible information for all plots and images. For the synthetic data used this includes a unique identifier for each specific pedestrian type. This can help to detect irregularities in the performance linked to data abnormalities, e.g. if only one or few pedestrians wear a specific type of hat.

6. Case Study

In our case study, we focused on investigating the performance of an intermediate version of a modified SSD [LAE*16] model for bounding box pedestrian detection. The model was trained on a synthetic autonomous driving data set, which was developed in the project „KI Absicherung“ [KI] and is planned to be published. Following, we will elaborate one exemplary walk-through of the analysis workflow and feedback loop (see Figure 1). The meta data contains image-specific (e.g. camera information, illumination) and object-specific attributes (e.g. detection type, age of the pedestrian). In our use case, it is of high relevance to detect pedestrians in close

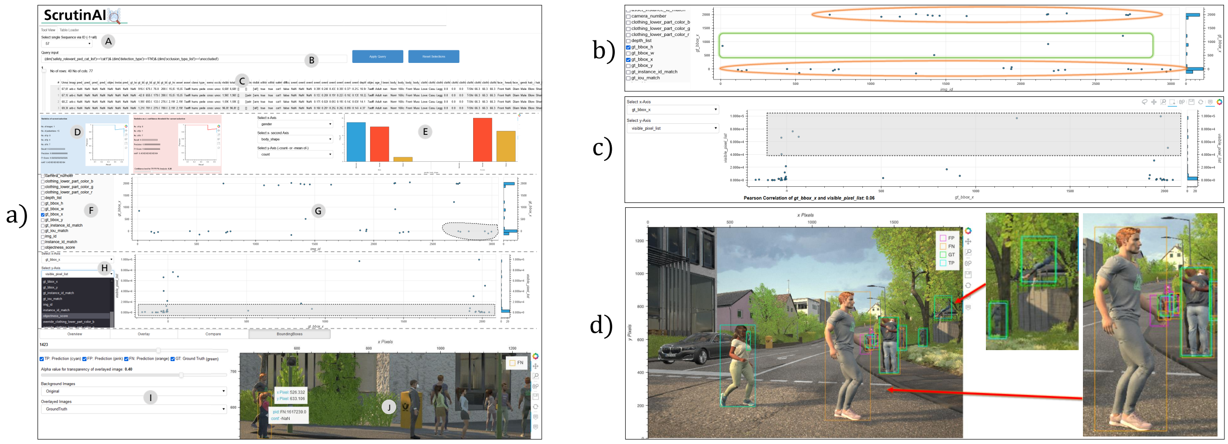

Figure 2: *ScrutinAI interface: The screenshots in a) show different interactive elements and visualizations (described in more detail in section 5) and document the findings of the case study in b) – d) (see section 6). The screenshots are truncated to give convenient overview.*

distance to the car to avoid collisions. Thus, pedestrian objects are annotated with different safety categories, with category 1 being the most critical for misdetection. We focused our investigation on those pedestrians with "safety relevant category" `cat1` that the model did not recognize, i.e. false negatives (FN). To investigate this subset of data, we filtered via a textual query for the attributes "safety relevant pedestrian == cat1" & "detection type == FN" & "occlusion type == unoccluded". The resulting subset yielded 40 pedestrians of the highest safety category that have not been recognized by the DNN model (annotated as FN). As hypothesis, we wanted to investigate the x-position of the pedestrians. The plot (see Figure 2 **b)**) revealed that only four (green rectangle) of the FNs are not located on the edges (orange circle) of the images. For the pedestrians located on the edges, a deep dive analysis of the images showed that most of them were truncated and barely visible within the image frame. From a further analysis of the scatter plot (see rectangle selection in Figure 2 **c)**) of the x-position and the attribute "visible pixels" it could additionally be observed that there are big pedestrians located that are not recognized, with one standing in the middle of the image. These data points were selected interactively with the box selection tool within the plot for a further drill-down analysis. As illustrated in Figure 2 **d)**, the model controversially shows incredible detection performance for pedestrians sitting in or hiding behind a tree, but fails to detect the big pedestrian in the middle. In summary, the analysis led to two insights: (1) there is a partially poor model performance for big pedestrians, and, (2) the `cat1` annotation of almost invisible pedestrians led to an "unfair" evaluation of model performance. Both were reported back to the data, metrics and model developers and successfully led to improvements for the next project release.

7. Discussion and Conclusion

We have presented an interactive and iterative VA workflow and tool, focused on the semantic analysis of DNNs with a feedback loop reporting the knowledge gained during the analysis to other stakeholders. Concentrating on the semantic analysis supports the analyst in identifying relevant semantic dimensions and their influence on the origins or absence of DNN weaknesses. The practical relevance was demonstrated with a case study within the domain of autonomous driving. Generally, the presented tool in section 5 can be adopted to other fields. For example, the most similar applications being domains in which models are also processing image data, e.g. DNNs used for detection of diseases in medical imaging or quality inspection in industrial assembly. The interim results showed the applicability and benefits of our approach. The knowledge generated during the exemplary analysis was reported back to the respective stakeholders for further improvement of the DNN model and training data as well as the ground truth data used in the application. As ScrutinAI is work in progress, we plan on integrating a way of tracking the insights gained during an analysis process by users, so that the final feedback report can be generated in an automated way. Possible ways to implement this are generally tracking the different interactions of the user or including interactive elements to add and modify annotations that overall result in a final report at the end of the analysis cycle. Another future direction, and currently under development, is the integration of a "query-by-example" functionality that retrieves similar images based on the content of an example image or image patch. We consider our work as a promising step towards exploiting the human knowledge for the semantic analysis of DNNs, supported by an interactive VA workflow. In addition, the knowledge gained and potentially the resulting incremental model improvements can foster further trust in specific abilities of the model.

Acknowledgments The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Energy within the project "Methoden und Maßnahmen zur Absicherung von KI basierten Wahrnehmungsfunktionen für das automatisierte Fahren (KI-Absicherung)". The development of this publication was supported by the Ministry of Economic Affairs, Innovation, Digitalization and Energy of the State of North Rhine-Westphalia as part of the flagship project ZERTIFIZIERTE KI. The authors would like to thank both consortia for the successful cooperation and Natalia Andrienko and Gennady Andrienko for inspiring discussions.

References

[AAAW22] Andrienko N., Andrienko G., Adilova L., Wrobel S.: Visual Analytics for Human-Centered Machine Learning. IEEE Computer Graphics and Applications 42, 1 (2022), 123–133. doi:10.1109/MCG.2021.3130314. 1, 2

[ALA*18] Andrienko N., Lammarsch T., Andrienko G., Fuchs G., Keim D., Miksch S., Rind A.: Viewing Visual Analytics as Model Building. Computer Graphics Forum 37 (2018). 2

[CL18] Choo J., Liu S.: Visual Analytics for Explainable Deep Learning. IEEE Computer Graphics and Applications 38 (2018), 84–92. 2

[CMJK20] Chatzimparmpas A., Martins R. M., Jusufi I., Kerren A.: A Survey of Surveys on the Use of Visualization for Interpreting Machine Learning Models. Information Visualization 19 (2020), 207 – 233. 2

[CT05] Cook K. A., Thomas J. J.: Illuminating the Path: The Research and Development Agenda for Visual Analytics. Tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 5 2005. URL: https://www.osti.gov/biblio/912515. 2

[DVK17] Doshi-Velez F., Kim B.: Towards A Rigorous Science of Interpretable Machine Learning. URL: https://arxiv.org/abs/1702.08608. 1, 2

[ERT*17] Endert A., Ribarsky W., Turkay C., Wong B., Nabney I., Blanco I. D., Rossi F.: The State of the Art in Integrating Machine Learning into Visual Analytics. Computer Graphics Forum 36 (2017). 2

[FWR*17] Federico P., Wagner M., Rind A., Amor-Amoros A., Miksch S., Aigner W.: The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics. 2017 IEEE Conference on Visual Analytics Science and Technology (VAST) (2017), 92–103. 2

[GTS*18] Garcia R., Telea A., Silva B. B. D., Tørresen J., Comba J.: A task-and-technique centered survey on visual analytics for deep learning model engineering. Comput. Graph. 77 (2018), 30–49. 2

[GZL*21] Gou L., Zou L., Li N., Hofmann M., Shekar A. K., Wendt A., Ren L.: VATLD: A Visual Analytics System to Assess, Understand and Improve Traffic Light Detection. IEEE Transactions on Visualization and Computer Graphics 27, 2 (2021), 261–271. doi:10.1109/TVCG.2020.3030350. 2

[HKPC19] Hohman F., Kahng M., Pienta R. S., Chau D. H.: Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers. IEEE Transactions on Visualization and Computer Graphics 25 (2019), 2674–2693. 2

[Hol] Holoviz: HoloViz. High-level tools to simplify visualization in Python. URL: https://holoviz.org/. 3

[Hou21] Houben S.: Inspect, Understand, Overcome: A Survey of Practical Methods for AI Safety. CoRR abs/2104.14235 (2021). URL: https://arxiv.org/abs/2104.14235. 2

[HZS*22] He W., Zou L., Shekar A. K., Gou L., Ren L.: Where Can We Help? A Visual Analytics Approach to Diagnosing and Improving Semantic Segmentation of Movable Objects. IEEE Transactions on Visualization and Computer Graphics 28, 1 (2022), 1040–1050. doi:10.1109/TVCG.2021.3114855. 2

[JLC19] Jiang L., Liu S., Chen C.: Recent Research Advances on Interactive Machine Learning. Journal of Visualization 22 (2019), 401–417. 2

[KAF*08] Keim D., Andrienko G., Fekete J.-D., Görg C., Kohlhammer J., Melançon G.: Visual Analytics: Definition, Process, and Challenges. In Information Visualization (2008). 2

[KI] KI Absicherung Konsortium: KI Absicherung. Safe AI for Automated Driving. URL: https://www.ki-absicherung-projekt.de/. 3

[KRN21] Kaluarachchi T., Reis A., Nanayakkara S.: A Review of Recent Deep Learning Approaches in Human-Centered Machine Learning. Sensors (Basel, Switzerland) 21 (2021). 2

[LAE*16] Liu W., Anguelov D., Erhan D., Szegedy C., Reed S. E., Fu C.-Y., Berg A. C.: SSD: Single Shot MultiBox Detector. In ECCV (2016). 1, 3

[LPK21] Linardatos P., Papastefanopoulos V., Kotsiantis S.: Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy 23 (2021). 1, 2

[LWLZ17] Liu S., Wang X., Liu M., Zhu J.: Towards Better Analysis of Machine Learning Models: A Visual Analytics Perspective. Vis. Informatics 1 (2017), 48–56. 2

[PDD*22] Park H., Das N., Duggal R., Wright A. P., Shaikh O., Hohman F., Chau D. H.: Neurocartography: Scalable automatic visual summarization of concepts in deep neural networks. IEEE Transactions on Visualization and Computer Graphics 28 (2022), 813–823. 2

[PMBT21] Paraschos S., Mollas I., Bassiliades N., Tsoumakas G.: VisioRed: A Visualisation Tool for Interpretable Predictive Maintenance. In IJCAI (2021). 2

[RBDG20] Roscher R., Bohn B., Duarte M. F., Garcke J.: Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access 8 (2020), 42200–42216. doi:10.1109/ACCESS.2020.2976199. 1

[RF16] Ribarsky W., Fisher B. D.: The Human-Computer System: Towards an Operational Model for Problem Solving. 2016 49th Hawaii International Conference on System Sciences (HICSS) (2016), 1446–1455. 2

[SEAKC21] Streeb D., El-Assady M., Keim D. A., Chen M.: Why Visualize? Arguments for Visual Support in Decision Making. IEEE Computer Graphics and Applications 41, 2 (2021), 17–22. doi:10.1109/MCG.2021.3055971. 2

[SKKC19] Sacha D., Kraus M., Keim D., Chen M.: VIS4ML: An Ontology for Visual Analytics Assisted Machine Learning. IEEE Transactions on Visualization and Computer Graphics 25 (2019), 385–395. 2

[SSS*14] Sacha D., Stoffel A., Stoffel F., Kwon B. C., Ellis G. P., Keim D.: Knowledge Generation Model for Visual Analytics. IEEE Transactions on Visualization and Computer Graphics 20 (2014), 1604–1613. 2

[SSSEA20] Spinner T., Schlegel U., Schäfer H., El-Assady M.: explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning. IEEE Transactions on Visualization and Computer Graphics 26 (2020), 1064–1074. 2

[WJ22] Wen L.-H., Jo K.-H.: Deep Learning-based perception systems for autonomous driving: A comprehensive survey. Neurocomputing 489 (2022), 255–270. 1

[WPB*20] Wexler J., Pushkarna M., Bolukbasi T., Wattenberg M., Viégas F. B., Wilson J.: The What-If Tool: Interactive Probing of Machine Learning Models. IEEE Transactions on Visualization and Computer Graphics 26 (2020), 56–65. 2

[XCK*20] Xie Y., Chen M., Kao D., Gao G., Chen X. A.: CheXplain: Enabling Physicians to Explore and Understand Data-Driven, AI-Enabled Medical Imaging Analysis. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (2020). 2

[YCY*21] Yuan J., Chen C., Yang W., Liu M., Xia J., Liu S.: A Survey of Visual Analytics Techniques for Machine Learning. Comput. Vis. Media 7 (2021), 3–36. 2

[YKSJ08] Yi J. S., Kang Y., Stasko J., Jacko J.: Understanding and Characterizing Insights: How Do People Gain Insights Using Information Visualization? In BELIV '08 (2008). 2

[ZWM*19] Zhang J., Wang Y., Molino P., Li L., Ebert D. S.: Manifold: A Model-Agnostic Framework for Interpretation and Diagnosis of Machine Learning Models. IEEE Transactions on Visualization and Computer Graphics 25 (2019), 364–373. 2