The Twentieth International Conference on Autonomic and Autonomous Systems
ICAS 2024 Athens, Greece

DebiAI: Open-Source Toolkit for Data Analysis, Visualization and Evaluation in Machine Learning

Authors: Tom Mansion, Raphaël Braud, Ahmed Amrani, Sabrina Chaouche, Faouzi Adjed, Loïc Cantat.

Affiliation: Intitut de Recherche Tecnhologique SystemX, France
Speaker : Faouzi Adjed

March 12, 2024

Plan

- Context and Introduction
- State of the art
- Problematic
- Methodology and proposed tool
- Visual functionalities
- Video demonstration
- Conclusion and perspectives

Context and introduction

- **Data analysis and visualization in Machine learning (LM) Context**
  - It contributes in all steps of ML workflow
  - It enhances an iteractive ML process across various phases
- **An emergent research topic under Human-Centred Machine Learning (HCML):**
  - It creates a human interaction and contributes in ML phases optimization
  - It improves understanding of complex concepts and facilitates in-depth exploration
  - It establishes connection between concepts, datasets and models.

State of the art

- CHAMELEON [1] an interactive tool designed to attribute data iteraction, thereby enhancing model performance, data validation, and the overall quality of ML projects
- ScrutinAI [2] is a Visual Analytics tool specifically tailored for enhancing the comprehension of deep neural network (DNN) predictions
- Manifold [3] a visual analytics platform designed for comparing and debugging ML models

[1] Hohman, F., Wongsuphasawat, K., Kery, M. B., & Patel, K. (2020, April). Understanding and visualizing data iteration in machine learning. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).
[2] Haedecke, E., Mock, M., & Akila, M. (2023). ScrutinAI: A visual analytics tool supporting semantic assessments of object detection models. Computers & Graphics.
[3] Zhang, J., Wang, Y., Molino, P., Li, L., & Ebert, D. S. (2018). Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models. IEEE transactions on visualization and computer graphics, 25(1), 364-373.

State of the art

HCML tool	Advantages	Limitation
CHAMELEON [1]	Enhance experience annotation, recording and editing data Real time data analysis	It does not include model performances Non open source tool
ScrutinAI [2]	Interactive data visualization in several	It can be used only for segmentation and object detection model
Manifold [3]	Compare and debug models in terms of accuracy and it helps to enhance model performances	It uses source code and it more closer to plotly than HCML concept

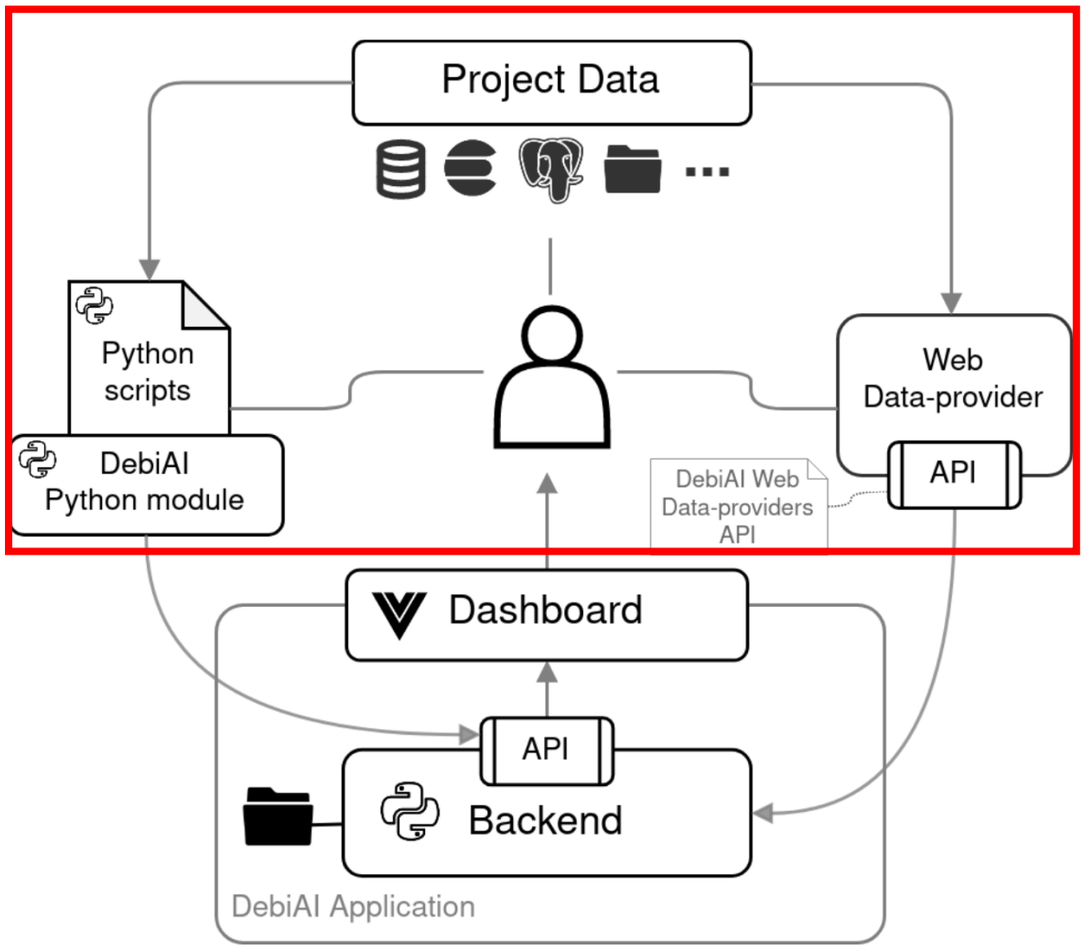
[1] Hohman, F., Wongsuphasawat, K., Kery, M. B., & Patel, K. (2020, April). Understanding and visualizing data iteration in machine learning. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).
[2] Haedecke, E., Mock, M., & Akila, M. (2023). ScrutinAI: A visual analytics tool supporting semantic assessments of object detection models. Computers & Graphics.
[3] Zhang, J., Wang, Y., Molino, P., Li, L., & Ebert, D. S. (2018). Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models. IEEE transactions on visualization and computer graphics, 25(1), 364-373.

Motivation and problematic

- Model analysis across various level of granularity (instance, subset and dataset)
- Multiplication of model functionalities such as regression classification, object detection, etc.
- Flexibility and connection of task such as filtering in the same project
- Ability to support ML pipeline in two crucial phases: pre-model and post model

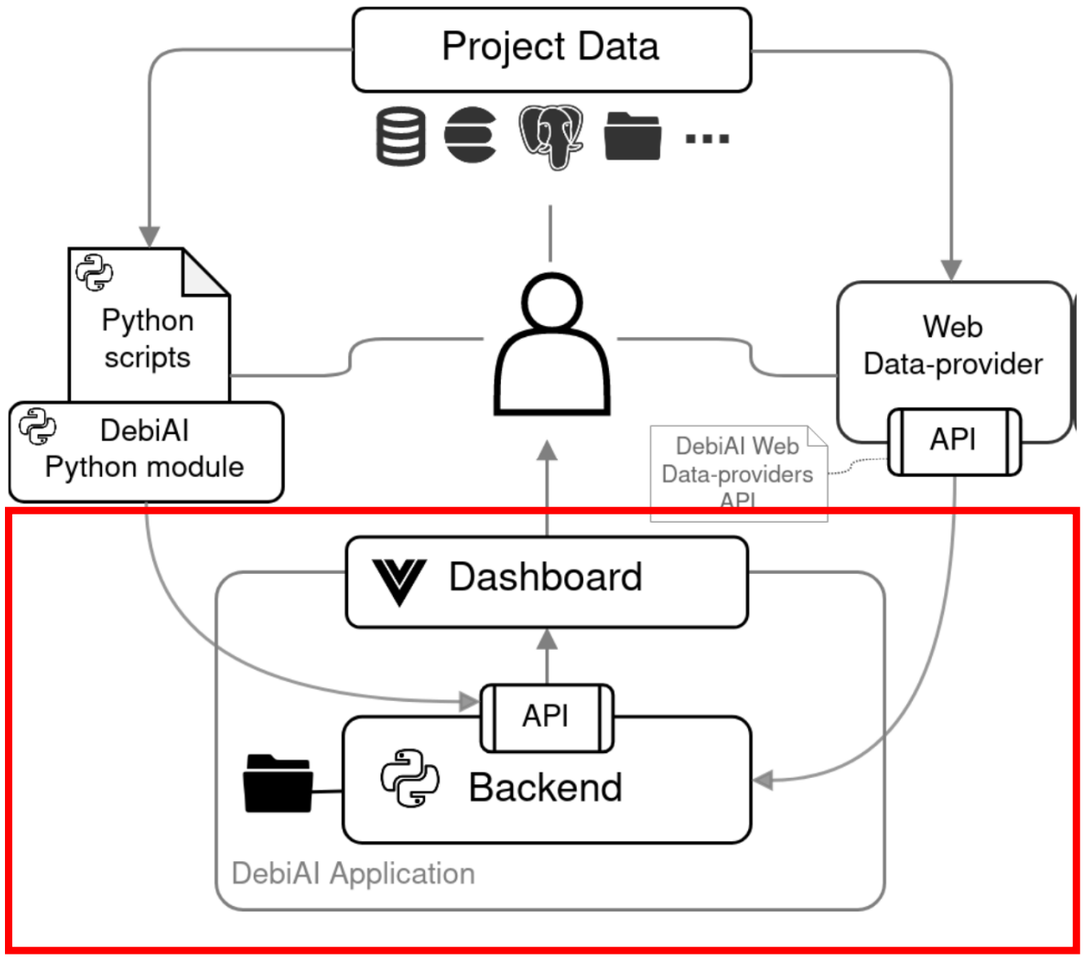
DebiAI is designed to resolve all these shortcomings and serves the data scientist during the whole ML process

DebiAI : the proposed tool



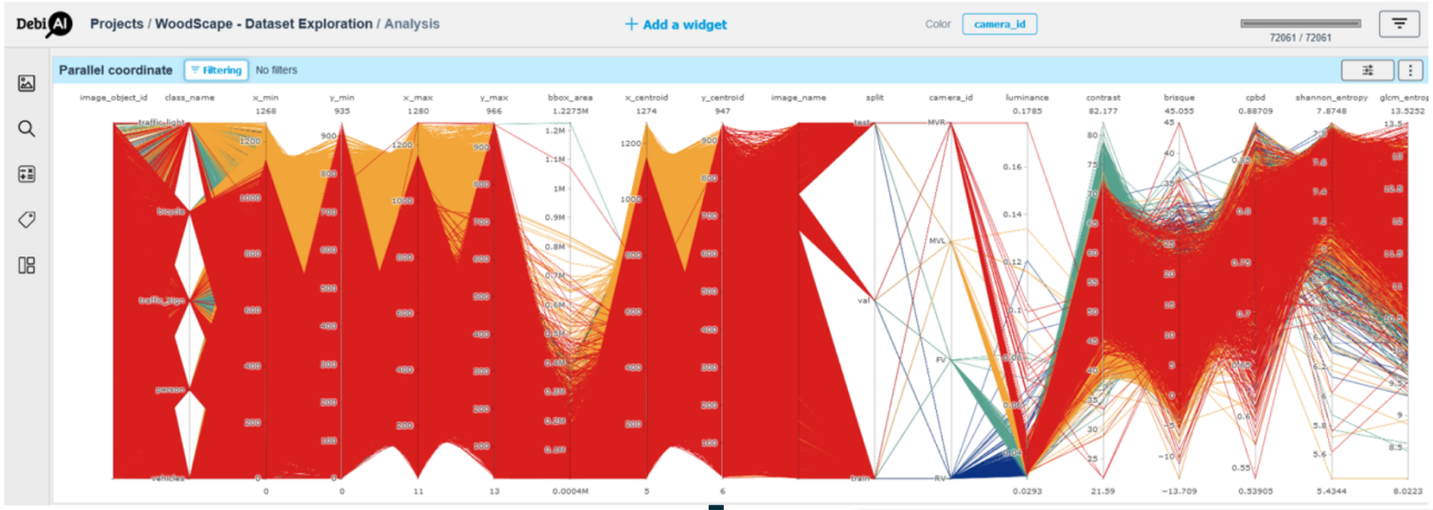
- **Project Data:**
  - This is the source of data that the user intends to analyze.
  - It may originate from a variety of sources and formats.
- **Python Scripts and DebiAI Python Module:**
  - Using the DebiAI Python module, users can adapt their existing scripts and workflows to create selections and insert data and model's results into DebiAI.
- **Web Data-Provider:**
  - These are the services created by the user's project that enable DebiAI to fetch data directly from the project's data sources.
  - A Web Data-Provider can be developed using any programming language, access data from any type of database, and be hosted on any server.

DebiAI : the proposed tool



- **DebiAI Web Dashboard:**
  - This is the user interface of DebiAI.
  - It provides users with an interactive platform to manage and view their data, and is hosted and served by the DebiAI backend.
- **Backend and API:**
  - This is a Python-powered backend that not only provides an API but also serves the Web dashboard.
  - This API is employed by the dashboard for data retrieval and by the Python module for data insertion.
  - Additionally, it manages communications with the Web Data-providers and processes computational requests made by the dashboard.
- **Data storage:**
  - DebiAI uses a folder-based data store that contains data in a JSON format.
  - This data store supports the DebiAI backend by retaining projects created by the Python module and some specific dashboard elements, including layout configurations for project dashboards.

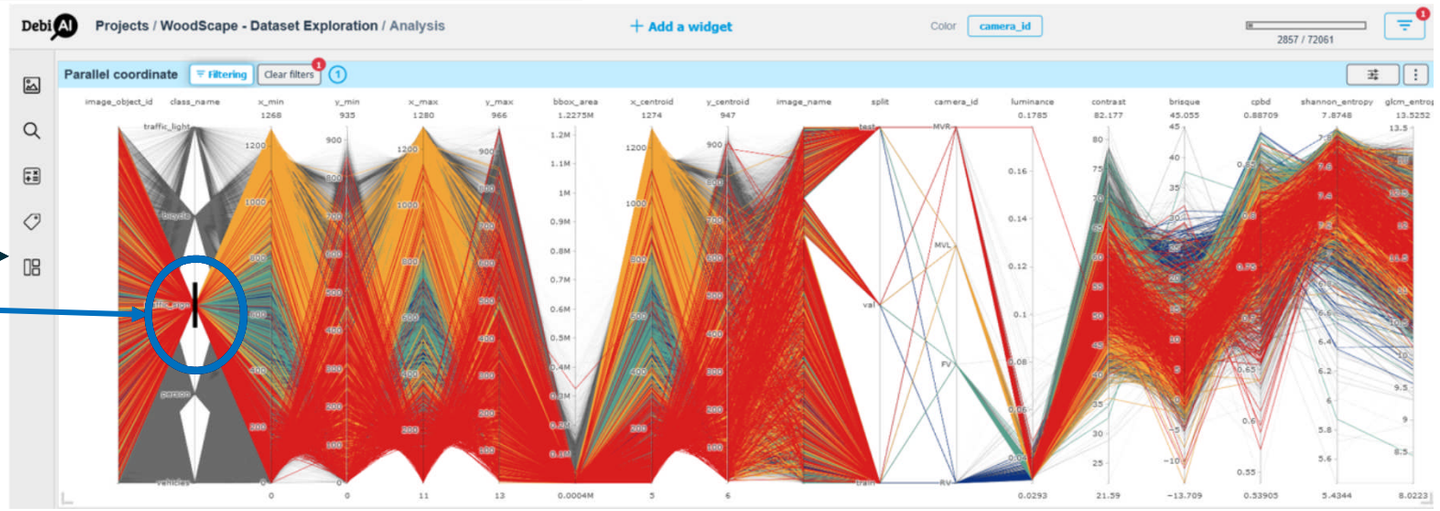
Visual functionalities: Filtering



Parallel coordinate example

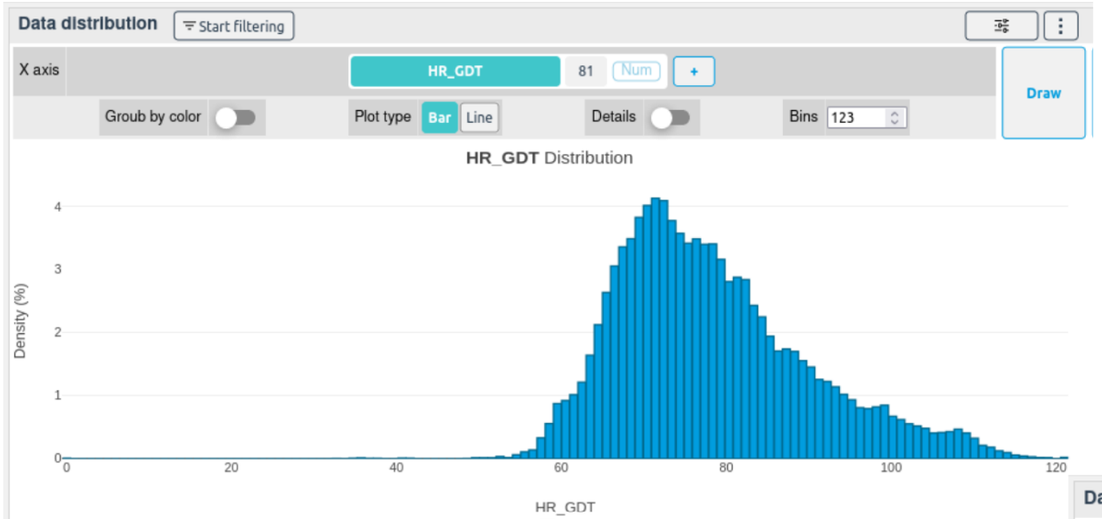
filtering

Selected dataset

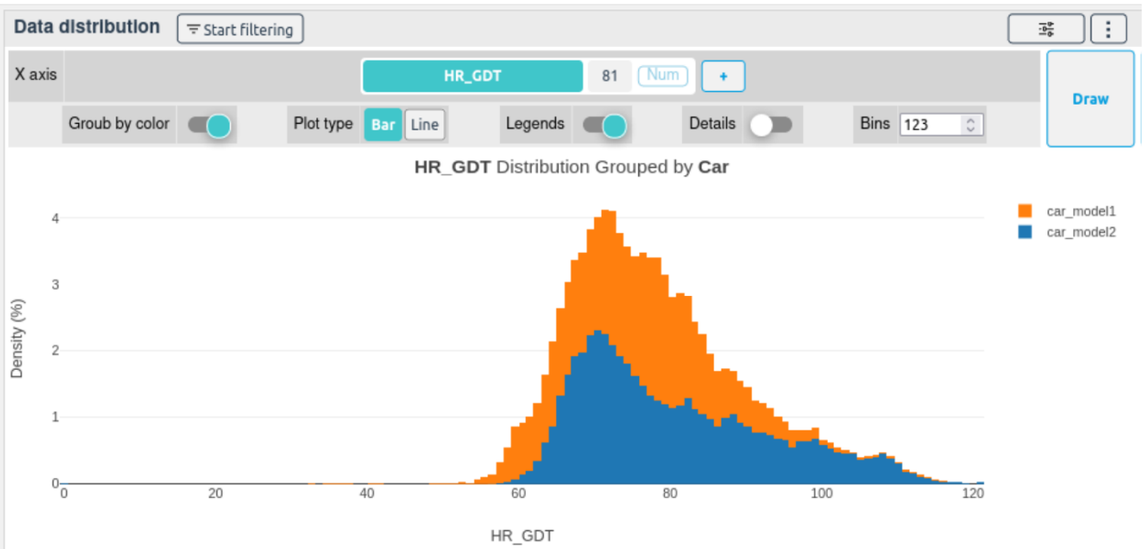


Mansion et al. *DebiAI: Open-Source Toolkit for Data Analysis Visualisation and Evaluation in Machine Learning*

Visual functionalities: Grouped by



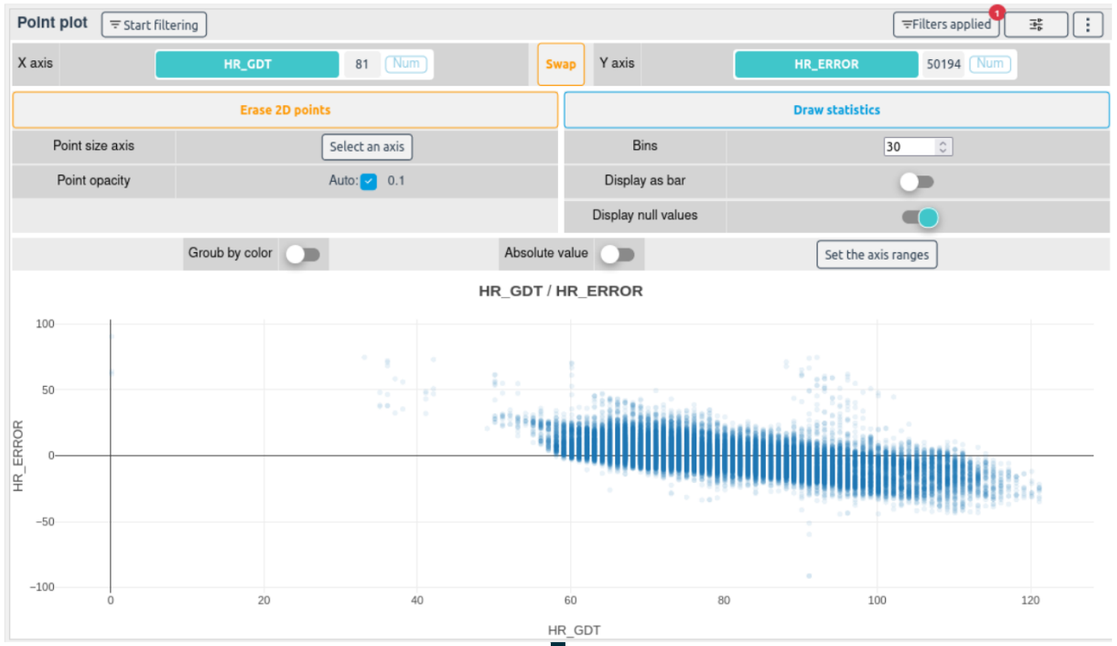
Grouped by



Data distribution example

Mansion et al. *DebiAI: Open-Source Toolkit for Data Analysis Visualisation and Evaluation in Machine Learning*

Visual functionalities: Statistical Analysis

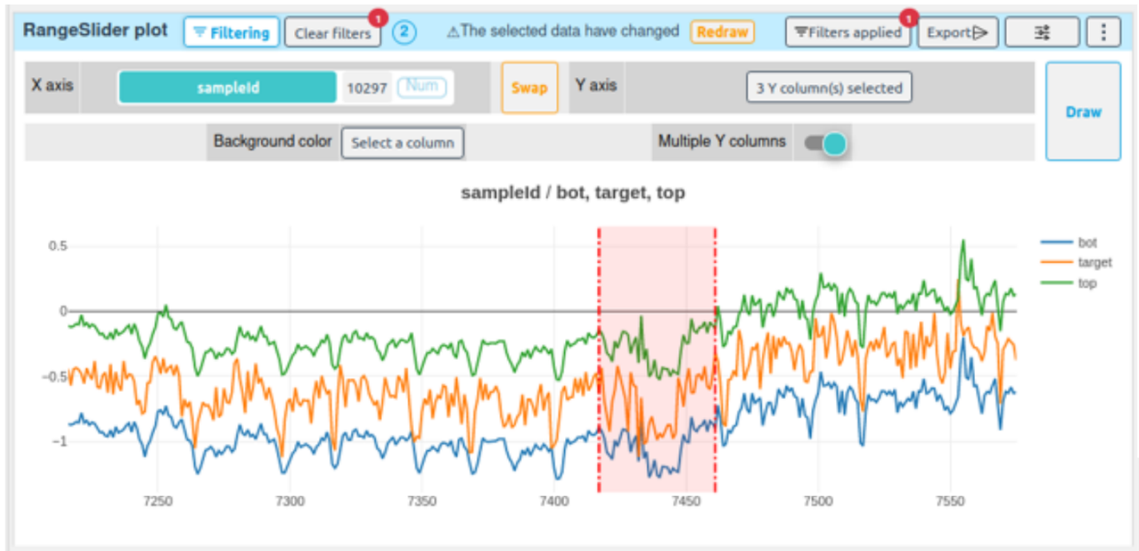


2D cloud points example

Statistical analysis

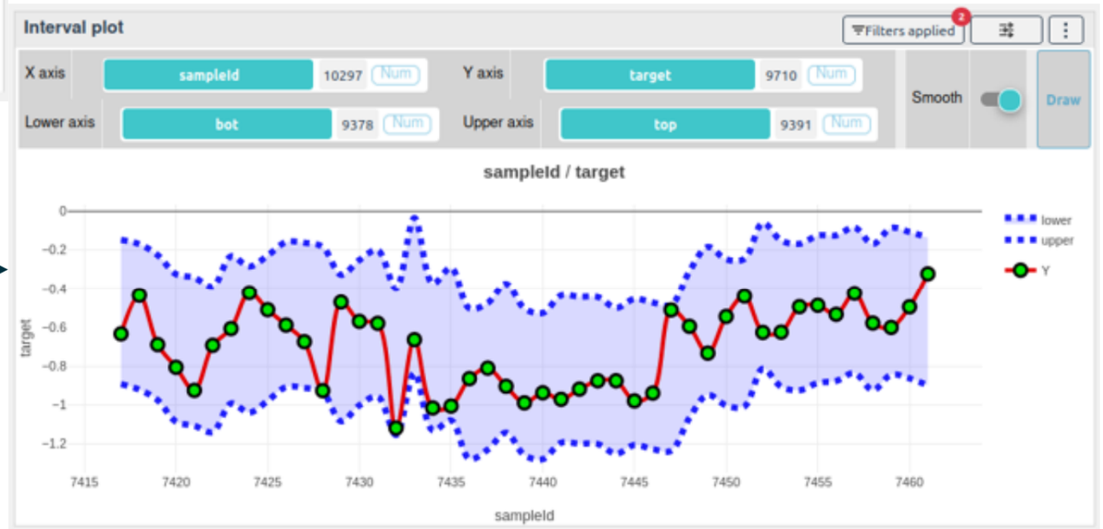
Mansion et al. *DebiAI: Open-Source Toolkit for Data Analysis Visualisation and Evaluation in Machine Learning*

Visual functionalities: Model Analysis



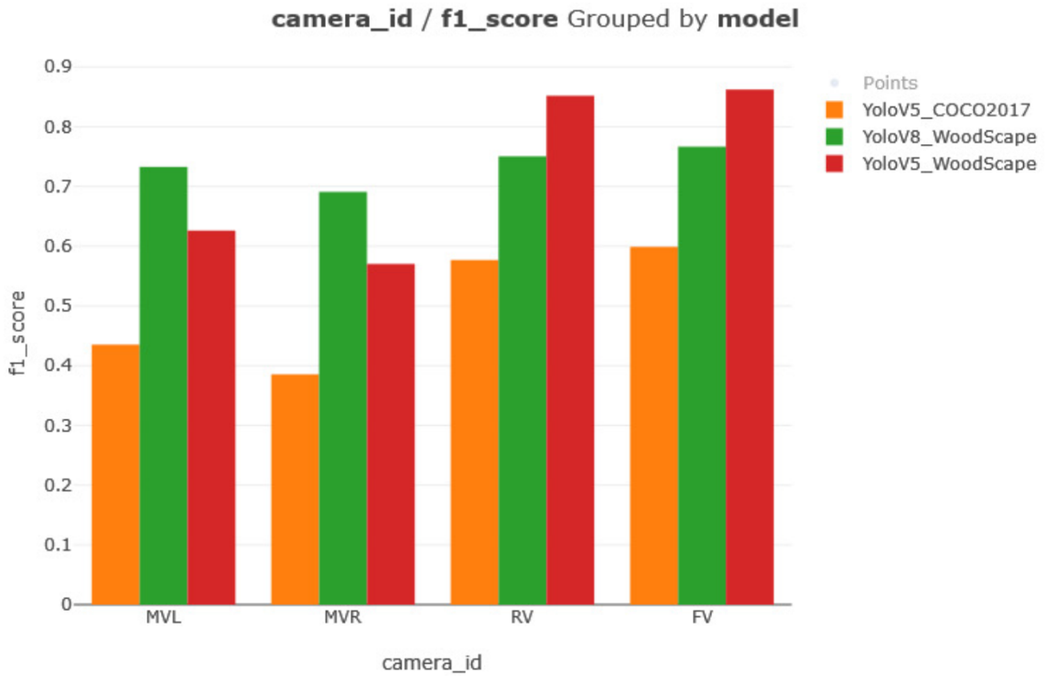
Time series example

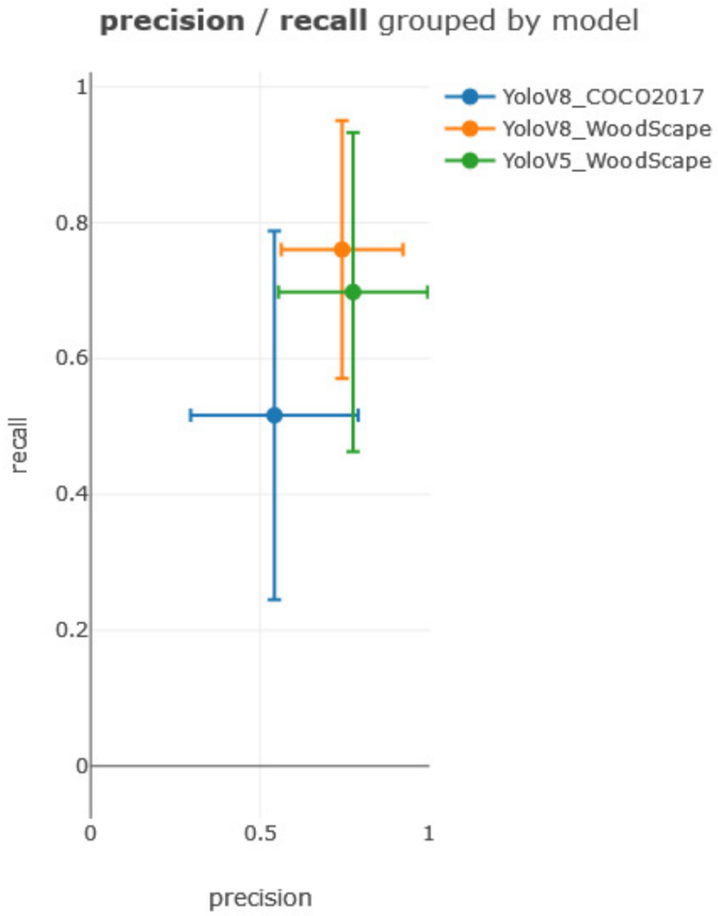
Model analysis



Mansion et al. *DebiAI: Open-Source Toolkit for Data Analysis Visualisation and Evaluation in Machine Learning*

Visual functionalities: Model Analysis





Models performance comparison

Mansion et al. *DebiAI: Open-Source Toolkit for Data Analysis Visualisation and Evaluation in Machine Learning*

Interactive video demonstration

Conclusions and perspectives

- **DebiAI is versatile web-based visual analytics tools that:**
  - Enhances data preparation and quality assessment
  - Integrates model result analysis and comparison
  - Can be adapted to various uses cases

- **Perspectives and ongoing developments are based on**
  - AlgoHub for data quality assessment
  - Mastering Model quality assessment and traceability
  - Data versioning system to ensure traceability of datasets and models

DebiAI

Confiance.ai

Systemx
INSTITUT DE RECHERCHE TECHNOLOGIQUE

IARIA

https://debiai.irt-systemx.fr/

www.confiance.ai
contact@irt-systemx.fr