Fraunhofer IAIS

ZERTIFIZIERTE KI Workshop „Visuell-explorative Bewertung neuronaler Netze“

Semantische Analyse neuronaler Netze mithilfe von Visual Analytics Methoden

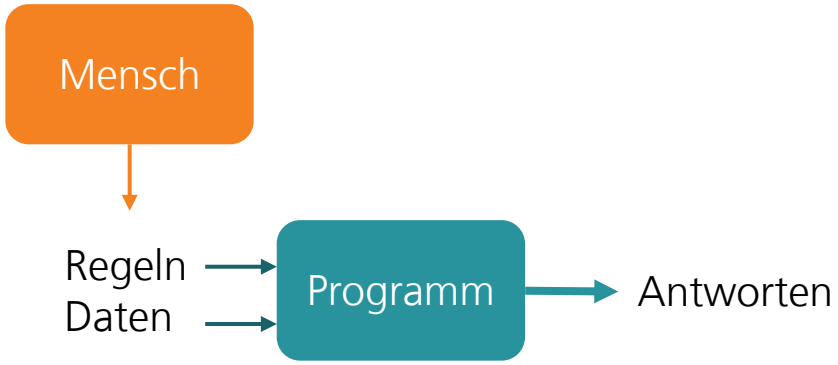
Elena Haedecke

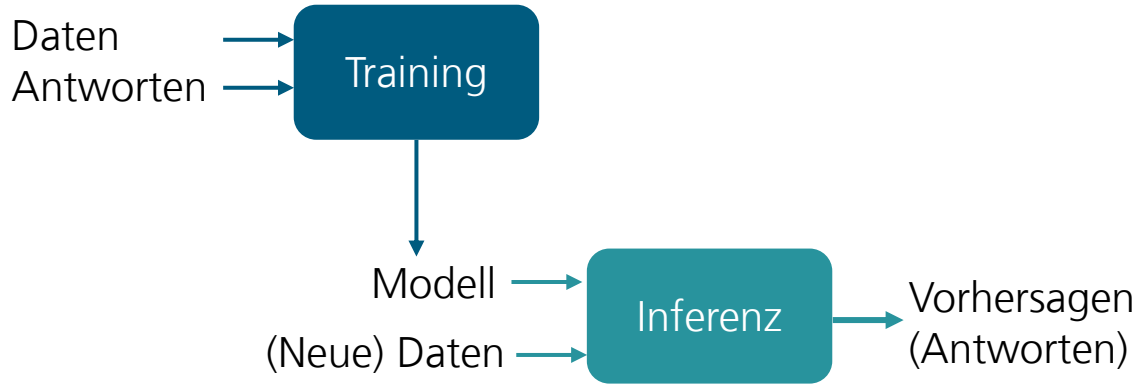
01

Warum brauchen wir Transparenz in neuronalen Netzen?

Machine Learning (ML) und Deep Neural Networks (DNN)

Unterschiede zur klassischen Programmierung





Klassische Programmierung

- Die "echte Welt" wird durch den Menschen modelliert (z.B. in Form von Regeln)
- Programm wendet das Modell/die Regeln auf die Daten an
- Ergebnis = Antworten

Machine Learning / Deep Learning

- Basis bildet ein Set aus (gelabelten) Daten und Antworten (Grundwahrheit)
  - Training des Modells
- Inferenz: Anwendung des Modells auf (neue/ungesehene) Daten
- Ergebnis = Vorhersagen (Antworten)

Machine Learning (ML) und Deep Neural Networks (DNN)

Warum ist Transparenz notwendig?

ML und DNN Systeme werden in immer mehr Bereichen eingesetzt, z.B. automatisierte Bildauswertung

Medizinische Diagnose

Inspektion von Bauteilen

Inspektion von Nahrungsmitteln

Autonome Fahrzeuge

Machine Learning (ML) und Deep Neural Networks (DNN)

Warum ist Transparenz notwendig?

ML und DNN Systeme werden in immer mehr Bereichen eingesetzt, z.B. automatisierte Bildauswertung

Medizinische Diagnose

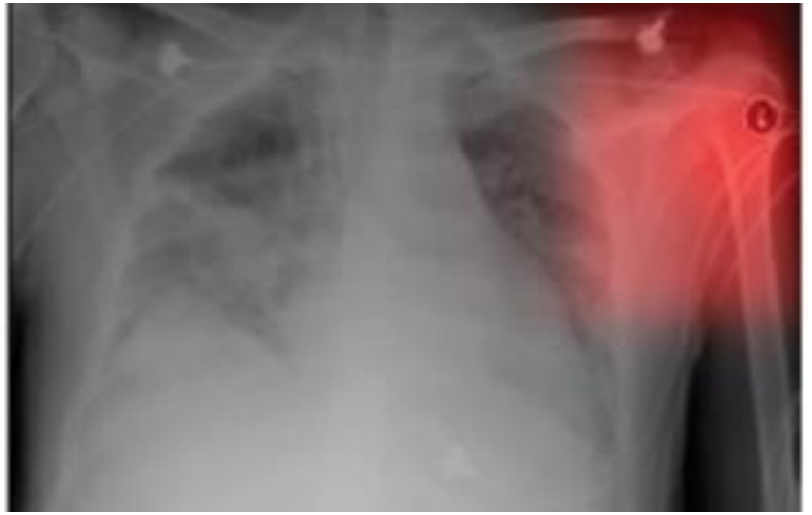
Inspektion von Bauteilen

Inspektion von Nahrungsmitteln

Autonome Fahrzeuge

Beispiel Medizinische Diagnose

- Ein System soll Lungenentzündungen in Röntgenaufnahmen erkennen
- Die Vorhersage wird fälschlicherweise auf ein Krankenhaus-Token gestützt



Quelle: Zech et al., PLOS Med (2018)

Machine Learning (ML) und Deep Neural Networks (DNN)

Warum ist Transparenz notwendig?

ML und DNN Systeme werden in immer mehr Bereichen eingesetzt, z.B. automatisierte Bildauswertung

Medizinische Diagnose

Inspektion von Bauteilen

Inspektion von Nahrungsmitteln

Autonome Fahrzeuge

Jedes technische System macht Fehler

Unerkannte Fehler bergen ein Risiko

Wir wollen Vorhersage und Entscheidungsprozess des Modells verstehen

02

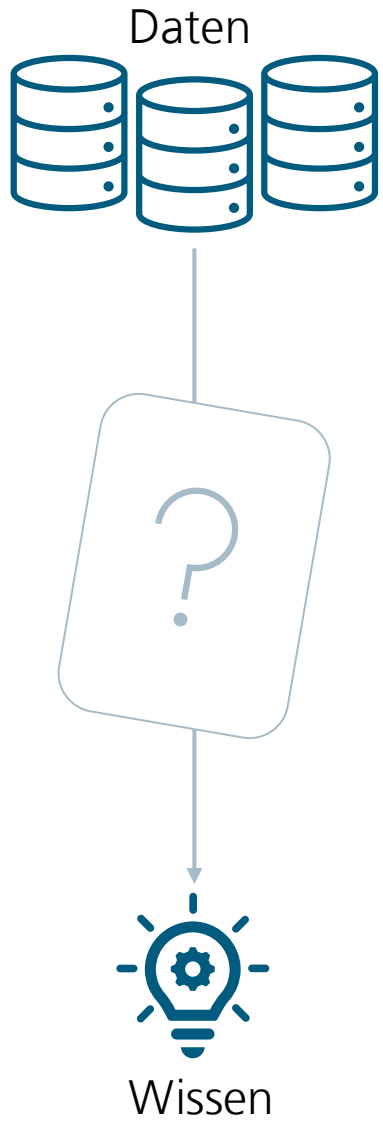
Menschenzentrierte Analyse mit Hilfe von Visual Analytics (VA)

Menschenzentrierte Analyse mit Hilfe von VA

Umwandlung von Daten in Wissen

- **Heutzutage ist das Sammeln von Rohdaten meist kein Problem**
  - Häufig sogar „Information Overload“
  - Daten sind
    - teilweise irrelevant,
    - unzureichend verarbeitet,
    - unpassend aufbereitet

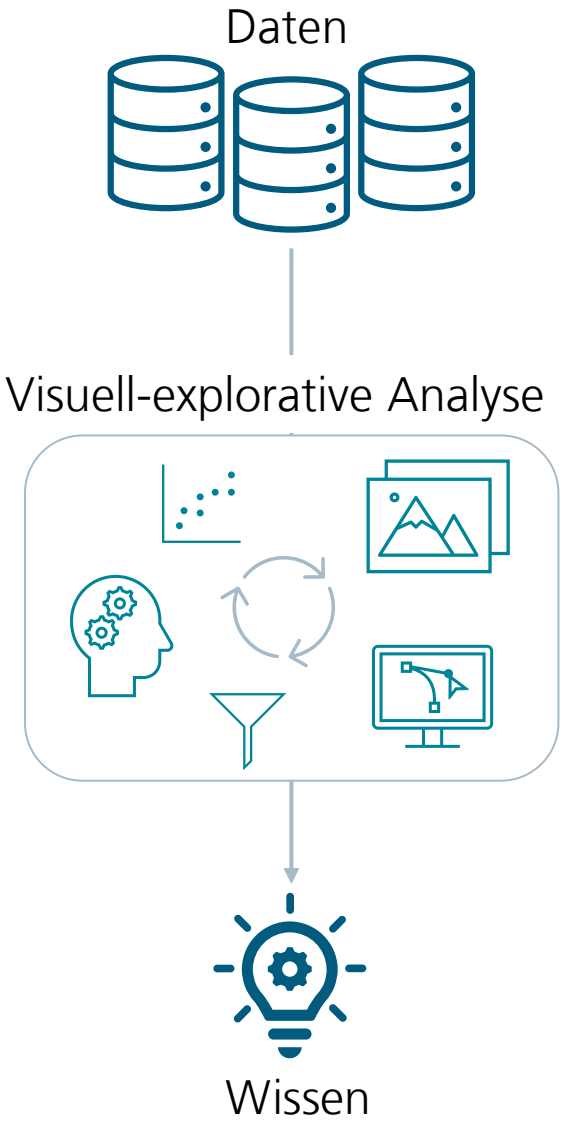
- **Die Schwierigkeit liegt vielmehr in der Umwandlung der Daten in Wissen**
  - Automatisierte Auswertungen funktionieren nur gut, wenn das Problem
    - ausreichend definiert und
    - ausreichend verstanden ist
  - Menschen allein können die Masse an Daten nicht effizient auswerten



Menschenzentrierte Analyse mit Hilfe von VA

Umwandlung von Daten in Wissen

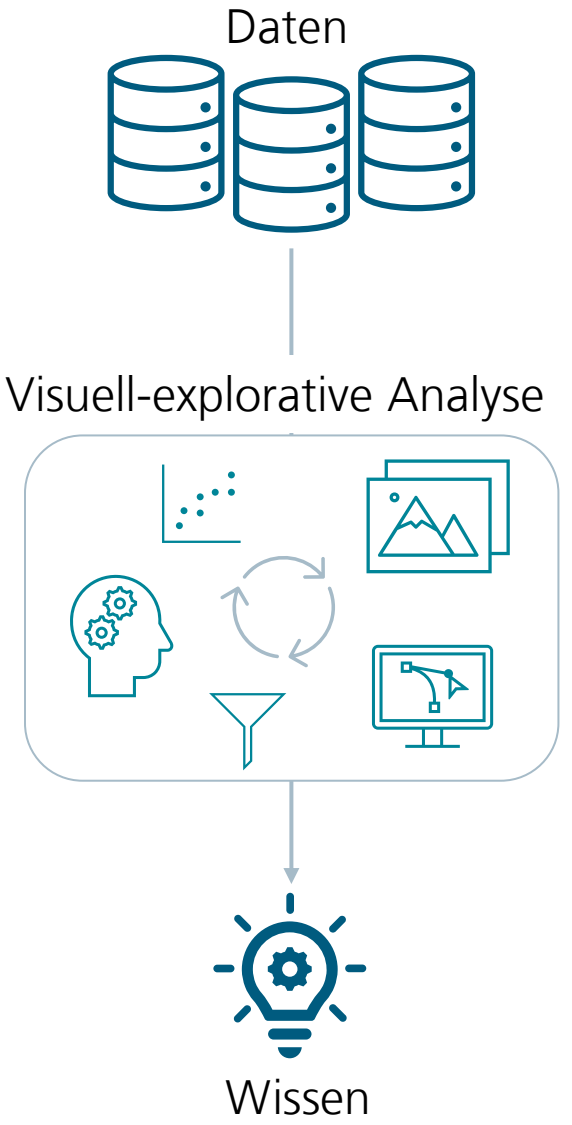
- **Visual Analytics verknüpft die Stärken von „Mensch“**
  - Visuelle Wahrnehmung und Mustererkennung
  - Semantisches Verständnis
  - Domänenwissen
- **und „Maschine“**
  - Automatisierte Verarbeitung großer Datenmengen
  - Visualisierung von Daten
  - Statistische Analysen



Menschenzentrierte Analyse mit Hilfe von VA

Fokus von Visual Analytics

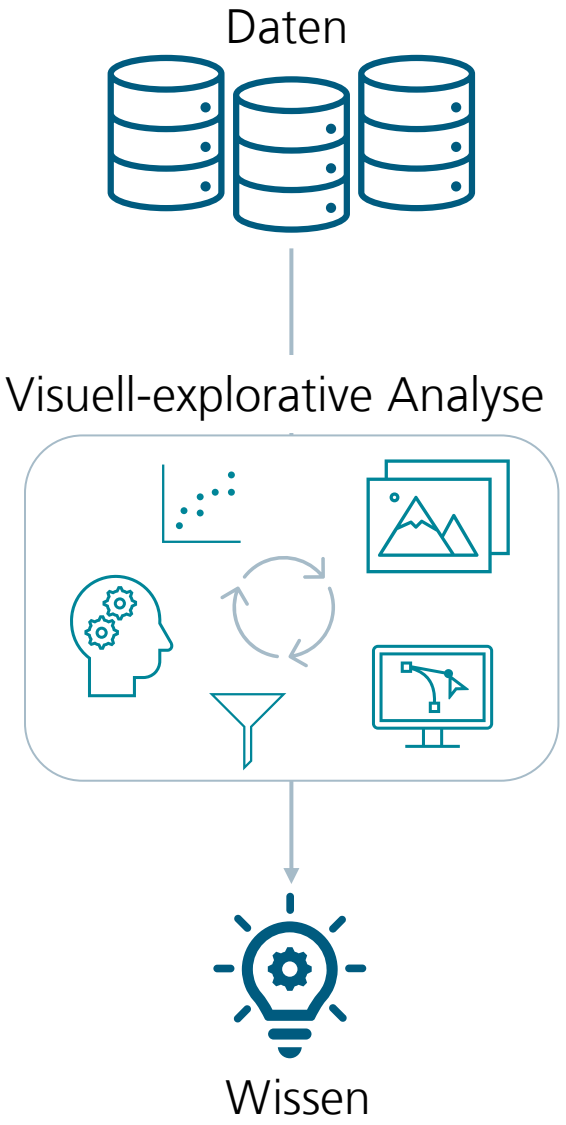
- **Tool-Support**
  - Unterstützung durch effizienten Workflow und ein
  - Interaktives Interface mit Use-Case entsprechenden Widgets
- **Visualisierung und Verknüpfung von Daten**
  - Dimensionsreduktion großer Datenmengen
  - Filtern und Durchsuchen von Daten
  - Interaktive Deep-Dive Analyse
- **Nutzbarmachen des menschlichen Wissens**
  - Hypothesen-basierte „Zoom-In & Zoom-Out“ Analyse



Menschenzentrierte Analyse mit Hilfe von VA

Fokus von Visual Analytics

- **Tool-Support**
  - Unterstützung durch effizienten Workflow und ein
  - Interaktives Interface mit Use-Case entsprechenden Widgets
- **Visualisierung und Verknüpfung von Daten**
  - Dimensionsreduktion großer Datenmengen
  - Filtern und Durchsuchen von Daten
  - Interaktive Deep-Dive Analyse
- **Nutzbarmachen des menschlichen Wissens**
  - Hypothesen-basierte „Zoom-In & Zoom-Out“ Analyse



Visual Analytics Mantra

„Analyze first, Show the Important, Zoom, Filter and analyze further, Details on demand.“

Keim et al. (2008). Visual Analytics: Definition, Process, and Challenges. https://doi.org/10.1007/978-3-540-70956-5_7

03

Visuell-explorative Analyse von neuronalen Netzen mit ScrutinAI

Haedecke et al.: „ScrutinAI: A Visual Analytics Approach for the Semantic Analysis of Deep Neural Network Predictions", EuroVis Workshop on Visual Analytics (2022)

Visuell-explorative Analyse von neuronalen Netzen

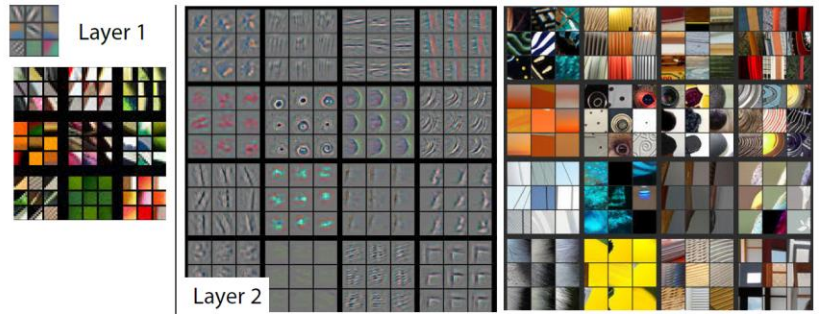
Verstehen der Black-Box

Interpretieren und Erklären der „Black-Box“ DNN Modelle und deren Vorhersagen

- Performanz des Modells ist abhängig von einer Kombination aus verschiedenen Dimensionen
  - Methoden zum Verständnis des Zusammenspiels der Eingabedimensionen werden benötigt

- Ansätze bietet das Forschungsfeld **Explainable AI (XAI)**
  - Interpretierbarkeit: **Wie** trifft das Modell generell Entscheidungen bzw. Vorhersagen?
  - Erklärung: **Warum** hat ein Modell eine spezifische Entscheidung bzw. Vorhersage getroffen?

- Aber:
  - **Evaluation durch Menschen notwendig**: Sind die Erklärungen verständlich und korrekt?
  - **Skalierbarkeitsproblem**:
    - Für welche Vorhersagen sollte ich mir die Erklärungen genauer anschauen?
    - Welche Dimensionen sind relevant?



Zeiler, Matthew D. and Rob Fergus. “Visualizing and Understanding Convolutional Networks.” ECCV (2014).

Selvaraju, Ramprasaath R. et al. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.” International Journal of Computer Vision 128 (2019): 336-359.

Visuell-explorative Analyse von neuronalen Netzen

Verstehen der Black-Box

- VA als Ergänzung zu menschenzentriertem ML und XAI Verfahren
- Analyse von Modellen kann **nicht nur** auf Performanzindikatoren gestützt werden
- **Systematische** Schwachstellen müssen **entdeckt und verstanden** werden

Visuell-explorative Analyse von neuronalen Netzen

Verstehen der Black-Box

- VA als Ergänzung zu menschenzentriertem ML und XAI Verfahren
- Analyse von Modellen kann **nicht nur** auf Performanzindikatoren gestützt werden
- **Systematische** Schwachstellen müssen **entdeckt und verstanden** werden

Zufälliger Fehler

Visuell-explorative Analyse von neuronalen Netzen

Verstehen der Black-Box

- VA als Ergänzung zu menschenzentriertem ML und XAI Verfahren
- Analyse von Modellen kann **nicht nur** auf Performanzindikatoren gestützt werden
- **Systematische** Schwachstellen müssen **entdeckt und verstanden** werden

Zufälliger Fehler

Systematischer Fehler

Visuell-explorative Analyse von neuronalen Netzen

Verstehen der Black-Box

Nutzbarmachen des menschlichen Domänenwissens und semantischen Verständnisses

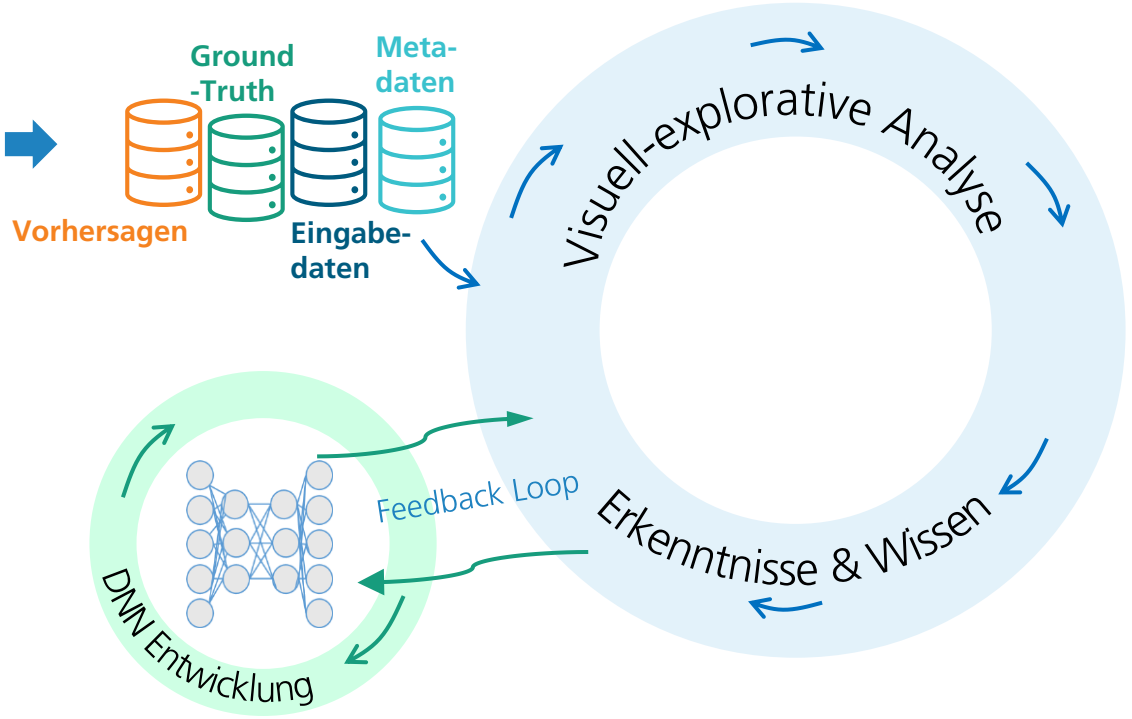
- **Menschenverständlich:**
  - Verknüpfen von „menschenverständlichen Konzepten“ (= semantische Dimensionen) mit Performanz-Metriken
- **Domänenwissen:**
  - Hypothesen-basiertes Explorieren der Daten
- **Semantik:**
  - Metadaten reichern die Eingabedaten an und fördern die semantische Analyse
- **Skalierbarkeit**:
  - Interaktives Interface und Widgets ermöglichen Untersuchung von Daten-Subsets

Fraunhofer IAIS

Visuell-explorative Analyse von neuronalen Netzen

ScrutinAI

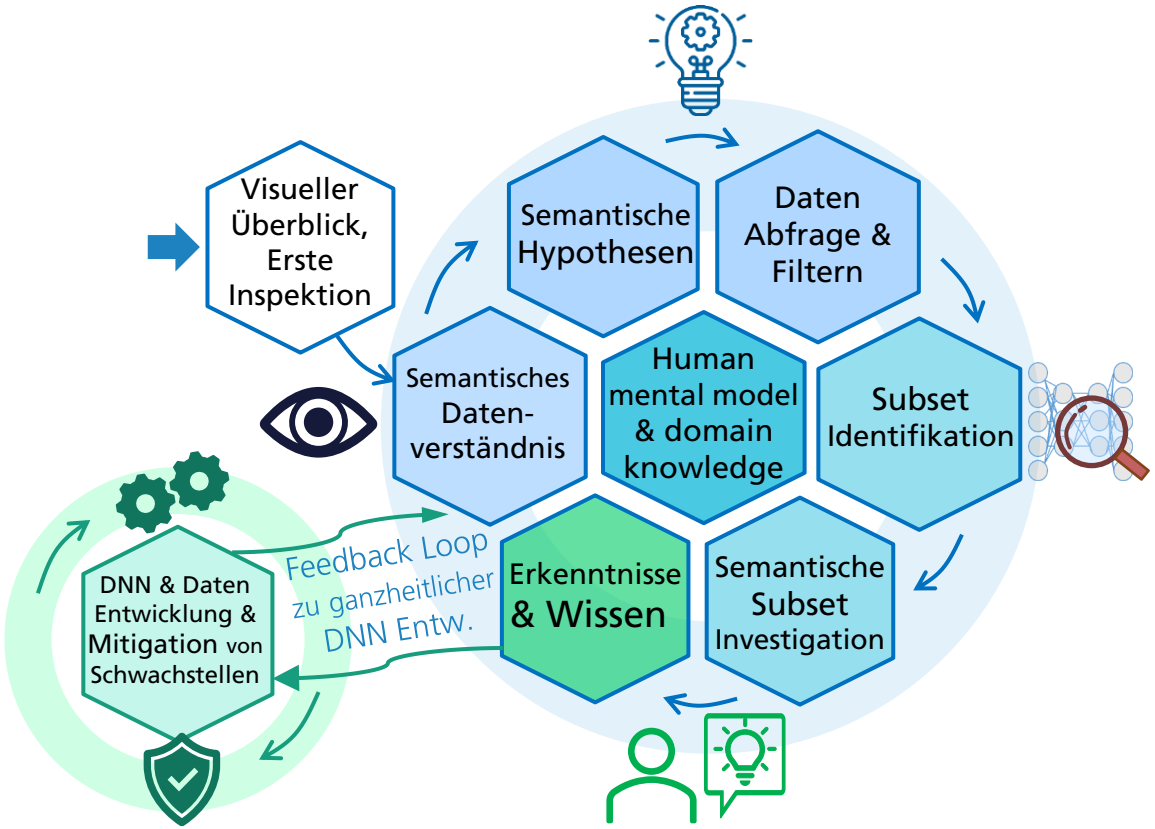
Grobkonzept des Workflows



Visuell-explorative Analyse von neuronalen Netzen

ScrutinAI

Feinkonzept des Workflows



Visuell-explorative Analyse von neuronalen Netzen

ScrutinAI: Feature Übersicht

Metadaten & Meta Informationen

Verständliche Darstellung

Daten Visualisierung

Interaktivität & Verknüpfung

Semantische Abfragen

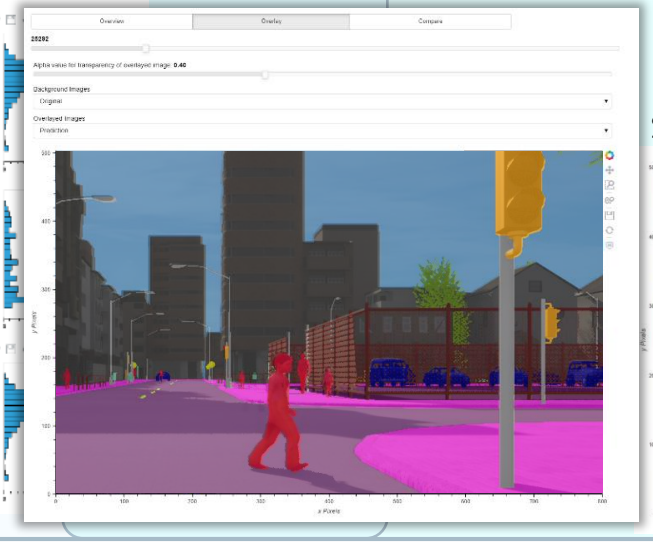
Ähnlichkeits-suche

Aggregation von Methoden, Metriken und Metadaten

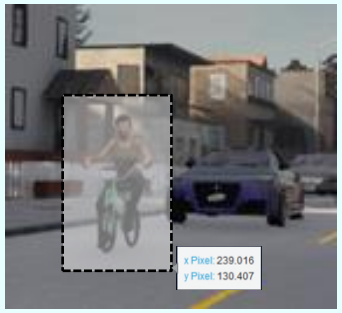
Query input

(dim('RecError')>=0.058) & (dim('inODD')=='False') & (dim('camera_type')=='arb')

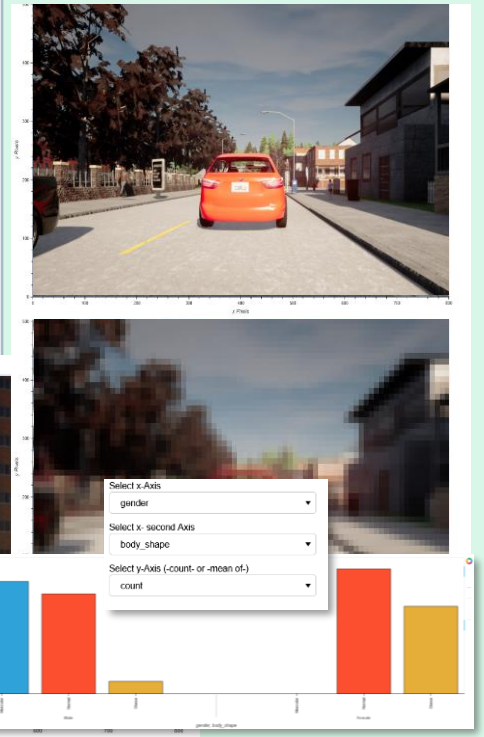
#	img_id	Filename	FußgängerIn	Auto	Ampel	Verkehrsschi	road	sidewalk
0	6,888	/data/share/k	0.0	NaN	NaN	NaN	NaN	0.907102
1	8,888	/data/share/k	0.174917	0.0	NaN	NaN	0.0	0.987142
2	8,892	/data/share/k	0.263533	0.0	NaN	NaN	0.972678	0.979261



Example Image / Patch



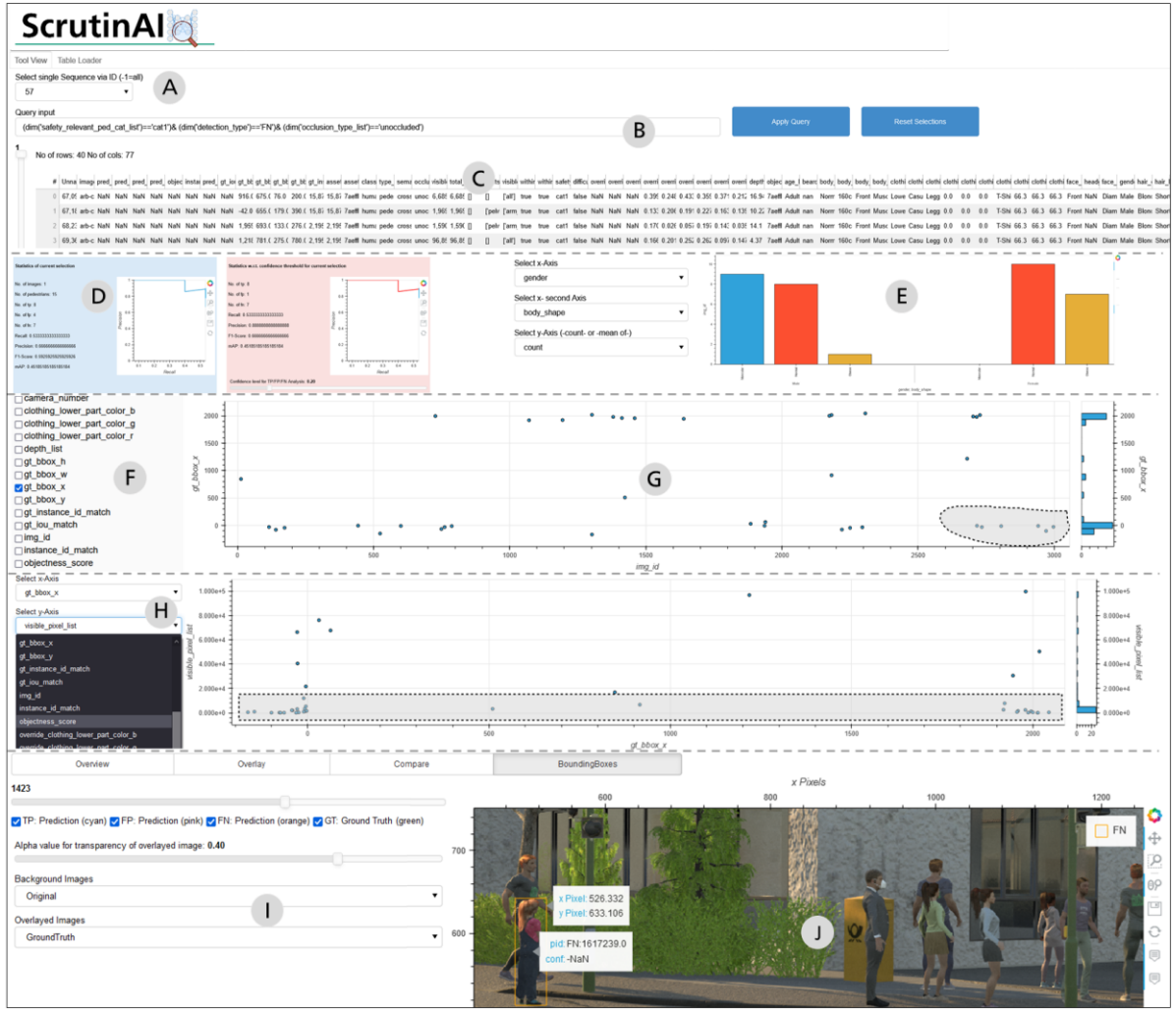
Similar Image



Visuell-explorative Analyse von neuronalen Netzen

ScrutinAI: Feature Übersicht

Live Demo!



A: Sequence Selection

B: Textual Query

C: Metadata table

D: Performance Statistics

E: Semantic Histogram

F: Plot selection

G: Visual Pattern Exploration

H: Correlation Plot

I: Image Overlay

J: Image Prediction Investigation

04

Fragen?

Kontakt

Elena Haedecke
Team Absicherung und Zertifizierung
elena.haedecke@iais.fraunhofer.de

Fraunhofer-Institut für intelligente
Analyse- und Informationssysteme IAIS
Schloss Birlinghoven 1
53757 Sankt Augustin

www.iais.fraunhofer.de