



データ可視化ツールを用いた AI開発の効率化

プラットフォーム開発部
市川 宏治

目次

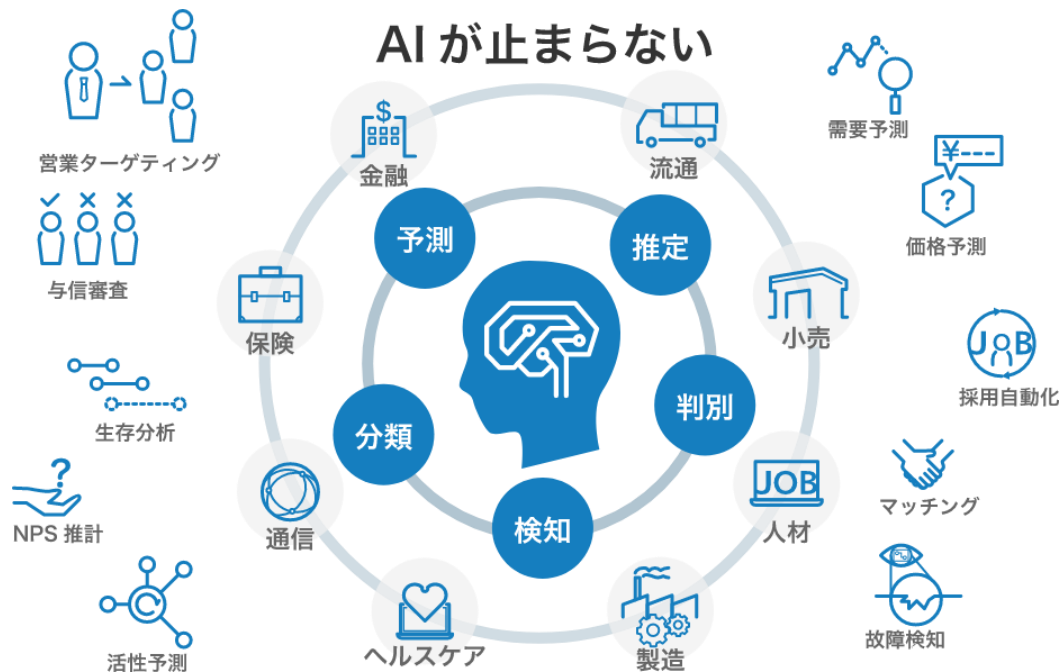
- ・自己紹介
- ・背景
- ・現状のデータ解析
- ・データ可視化ツール
- ・活用事例

自己紹介



- 名前：市川 宏治
- 所属：プラットフォーム開発部
AI・データサイエンス技術室
- 業務：自動運転認識AI評価
- 趣味：釣り、日曜大工

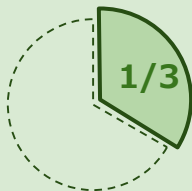
AI技術が日常生活～企業に至るまで幅広い分野で活用されている



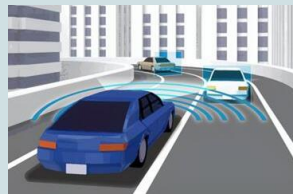
出典: (株)日立システムズ, <https://www.hitachi-systems.com/ind/ai-solution/>, 2022/10/26

一般的にAIに求められることは？

生産性の向上



社会安全性の向上

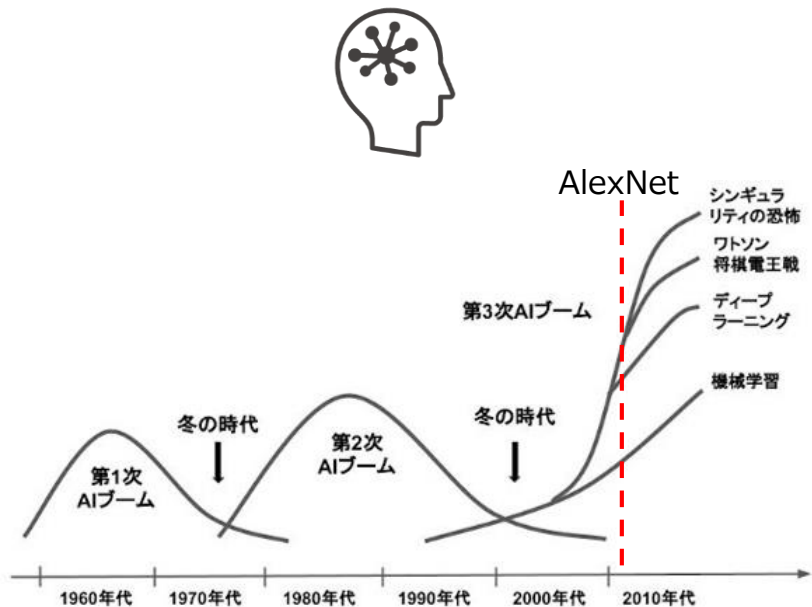


顧客満足度の向上



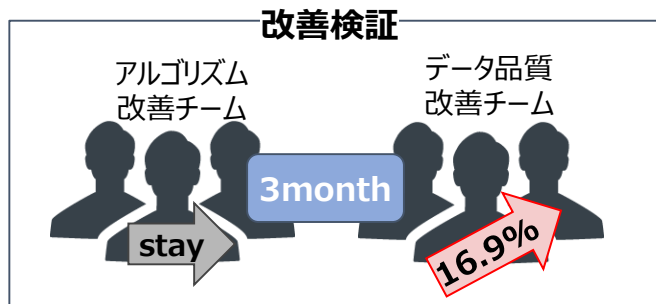
タスクに対してAIの**性能が十分**

AIの性能に影響する要因



アンドリュー・エン氏 データ中心のAI (Data-Centric AI) 2021
“モデルやアルゴリズムの構築よりも、
学習に用いるデータの品質向上を優先させるべきだ”

出典：forbesjapan, 人工知能の権威アンドリュー・エンが語る「データの質」の重要性
, <https://forbesjapan.com/articles/detail/42149/1/1/1?s=ns>

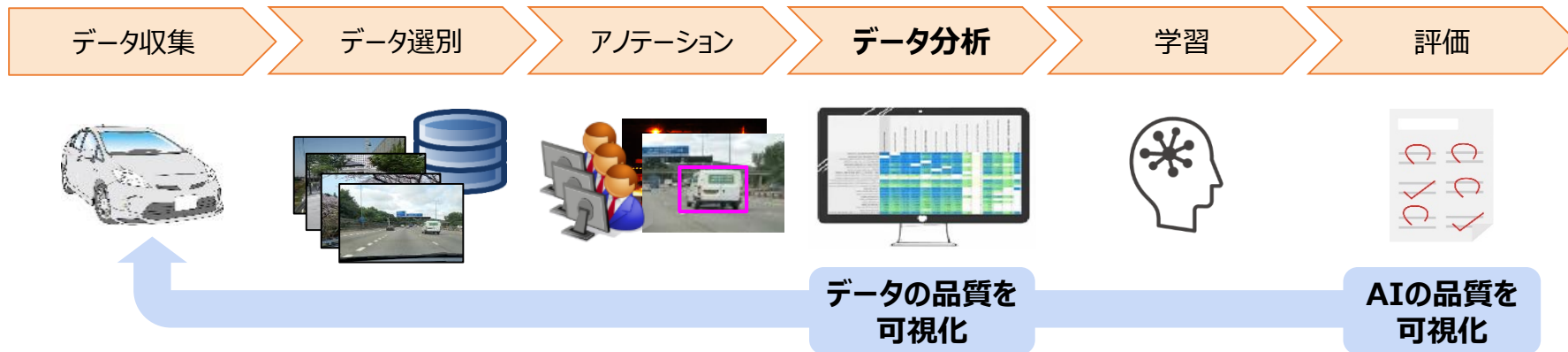


AIの性能には**データの品質**が大きく影響している

アプローチ

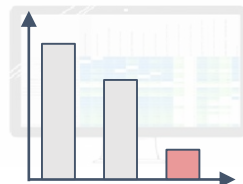
データを可視化する ≡ データの品質の可視化

一般的な物体検出モデルの開発工程



現状のデータ分析

一般的な物体検出モデルの開発工程

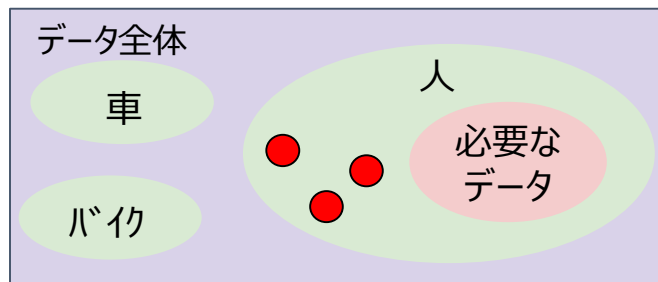


「人」のデータを増やしてみよう

車 バイク 人

オブジェクト	認識率
車	90%
バイク	80%
人	40%

「人」のデータを増やしてみよう



● 増やしたデータ

性能向上に必要なデータをFBできない可能性

なぜ？ データを詳細に分析できていない

特徴量可視化ツール

技術概要

教師データ



特徴量で分解

画像の要素：明度、色彩 ...



オブジェクト：車、自転車 ...

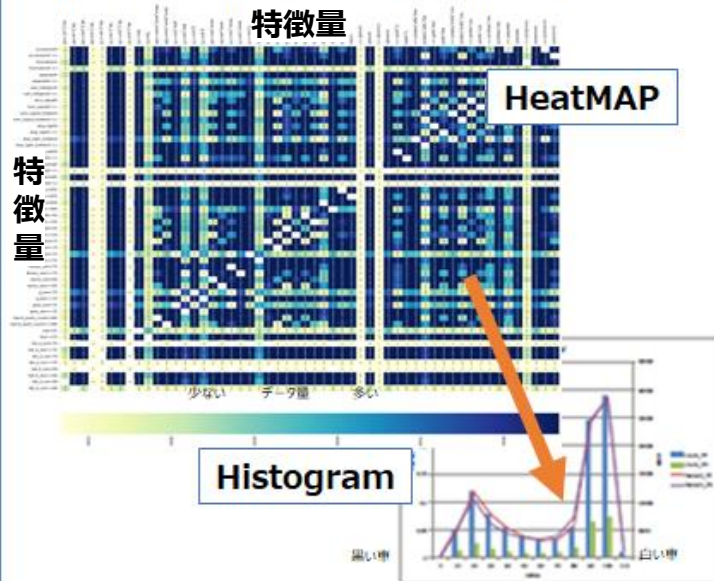


シーン：気象、道路種別 ...



評価・統計・可視化

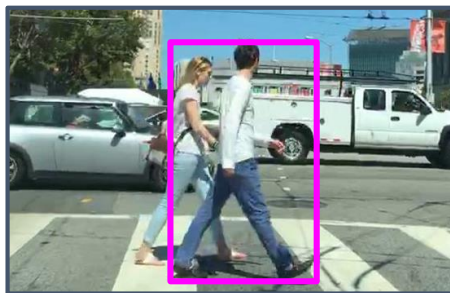
特徴量ごとに評価し、データの分布を表示



多次元の特徴量でデータを分析

特徴量可視化ツール

多次元の特徴量とは、、、

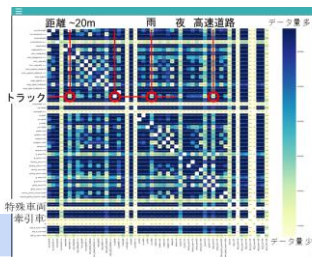


属性 × シーン = 多次元の特徴量

特徴量可視化ツール

既存のデータ解析ツールを使った場合と本ツールを使った場合の比較

一般的な物体検出モデルの開発工程



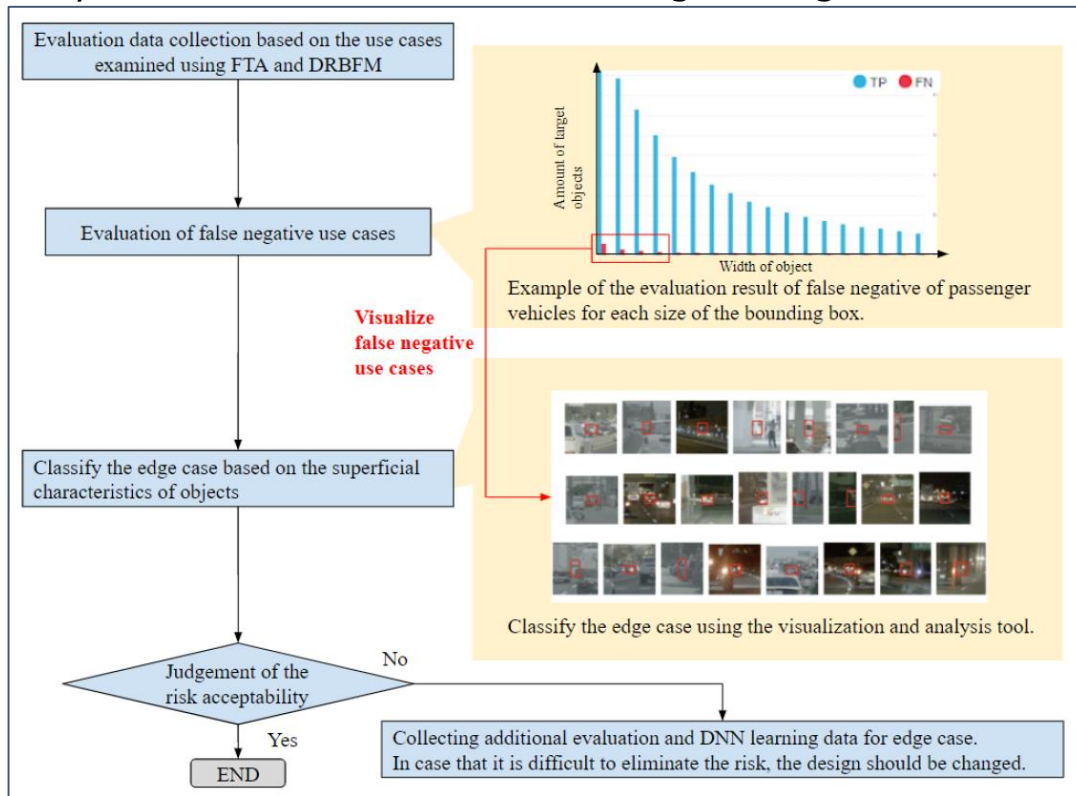
「人」のデータを増やしてみよう
昼 × 白服 × 人

「人」のデータを増やしてみよう
昼 × 交差点 × 横断中 × 人



活用事例：21年度 自動運転DNN 未検出評価プロセス

Quality Evaluation Method of DNN Image Recognition For Level 2 Partially Automated Driving



未検出が起きるレアシーンを
学習データに追加したい

本ツールを用いたデータ分析

出典：27th ITS World Congress, PaperID189, Woven core, Inc Yusuke Kashiba

活用事例：21年度 自動運転DNN 未検出評価プロセス

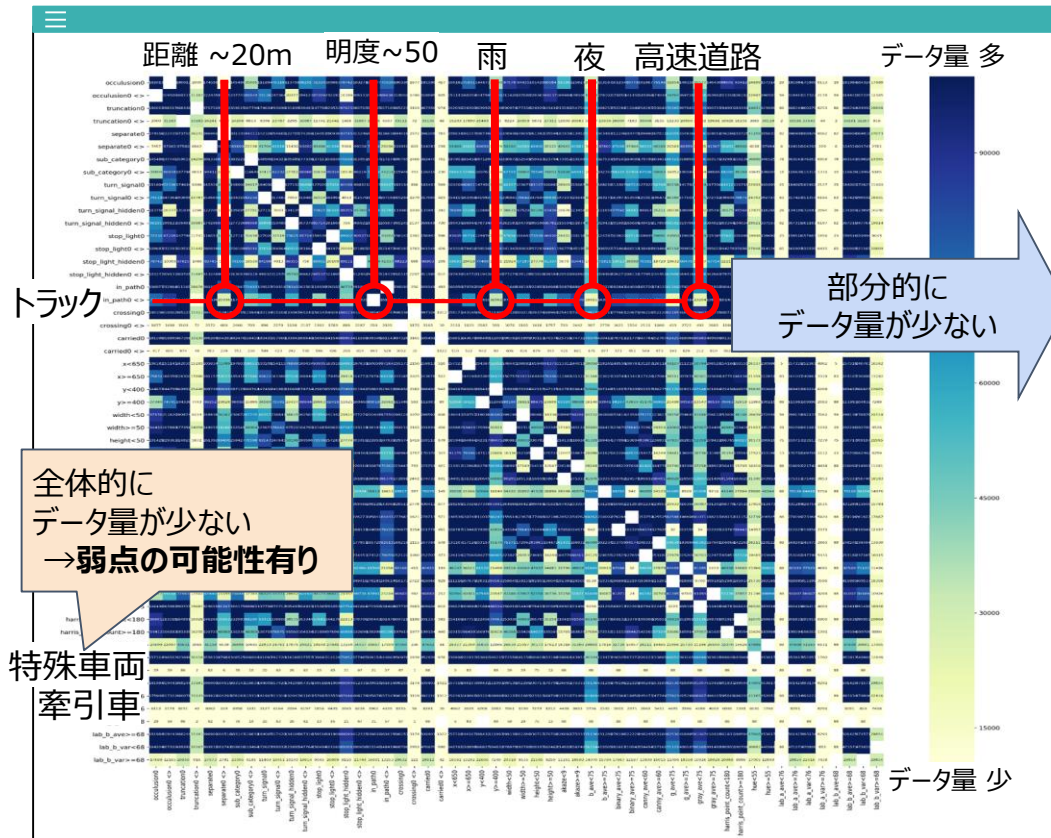
DNN の教師データを特徴量で分解・可視化

学習量評価

認識率評価

詳細分析

実データ
目視確認

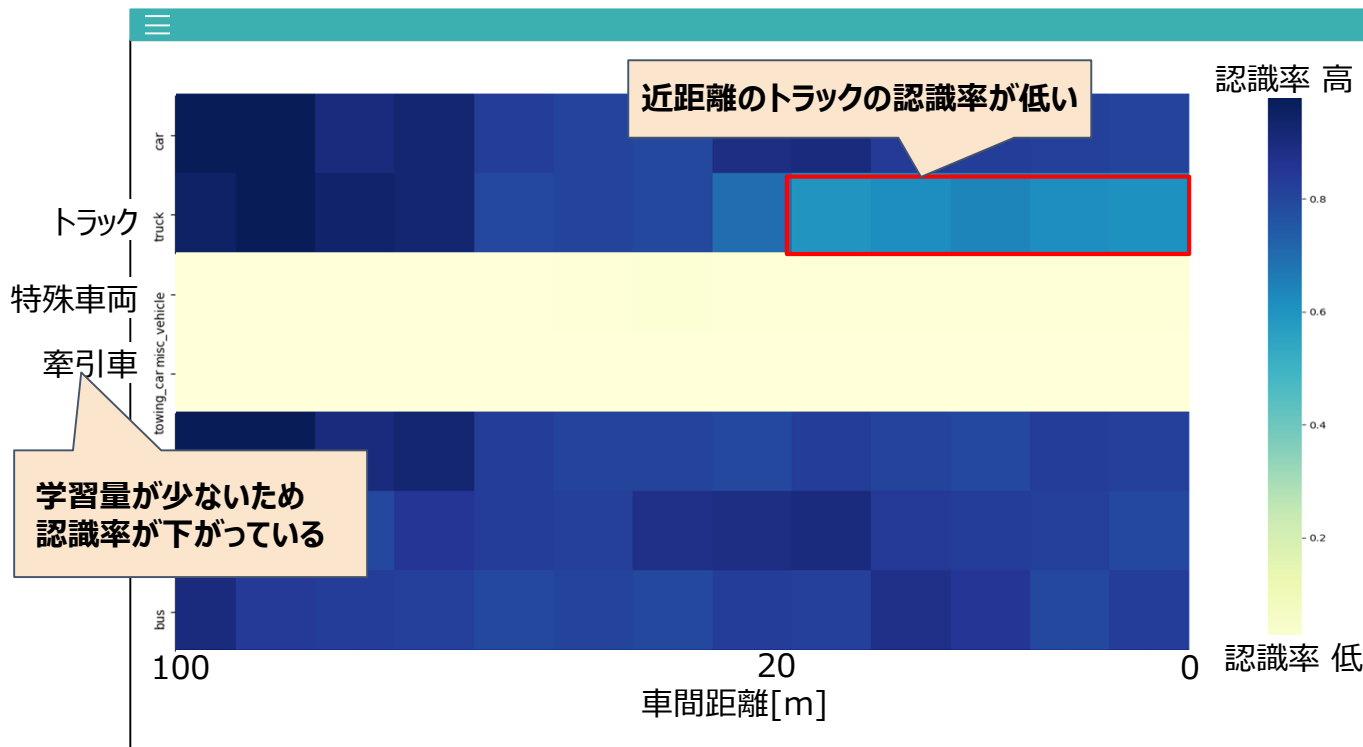


- シーン1：雨
- シーン2：夜
- シーン3：高速道路
- 属性1：トラック
- 属性2：車間距離 ~20m
- 属性3：明度 ~50

弱点の可能性有り

活用事例：21年度 自動運転DNN 未検出評価プロセス

DNN の認識結果を特徴量を基に可視化



活用事例：21年度 自動運転DNN 未検出評価プロセス

認識結果を多次元の特徴量でフィルタリング



① 教師データ量可視化

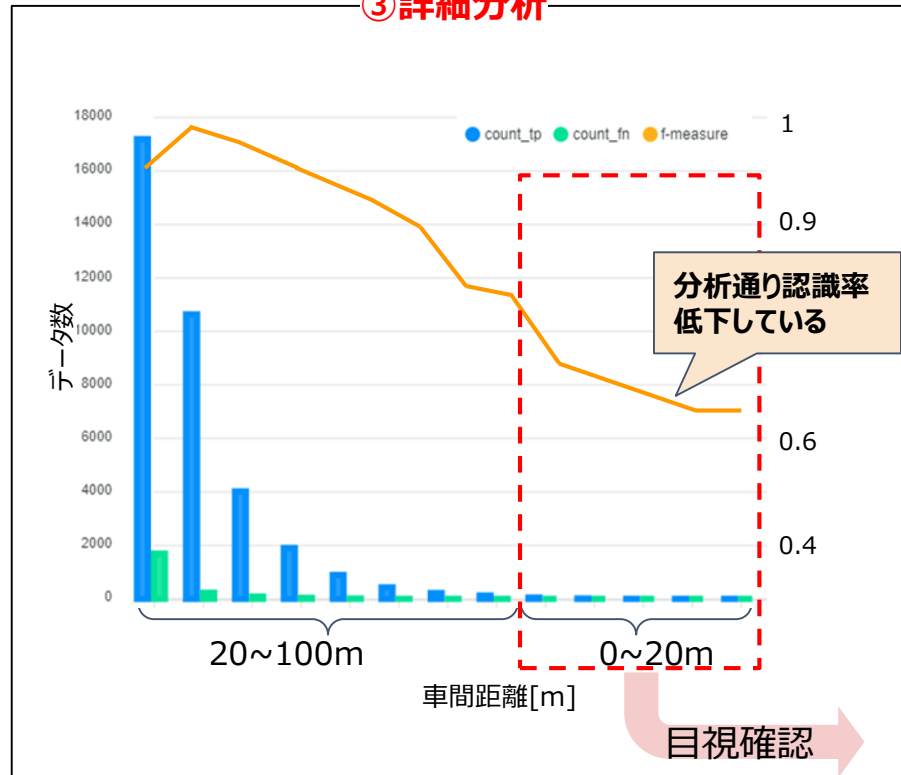
シーン1：雨
シーン2：夜
シーン3：高速道路
属性1：トラック
属性2：車間距離 ~20m
属性3：明度 ~50



② 認識率可視化



③ 詳細分析



活用事例：21年度 自動運転DNN 未検出評価プロセス

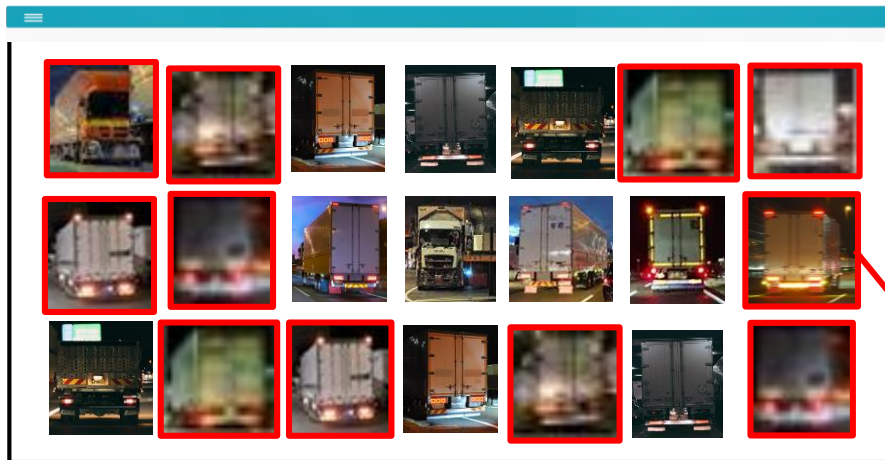
DNNが正確に認識できなかった箇所を画像で確認

学習量評価

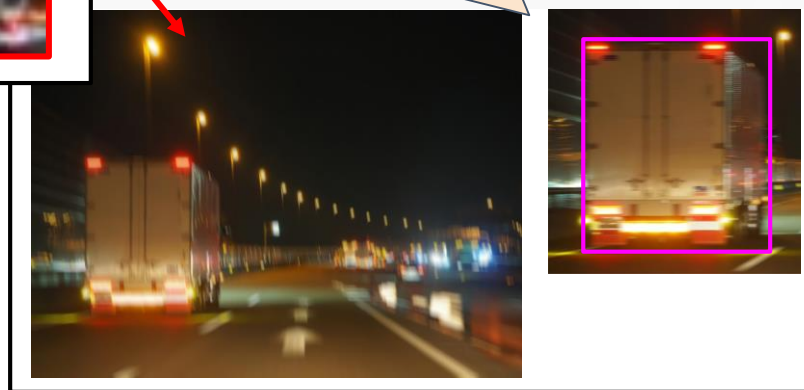
認識率評価

詳細分析

実データ
目視確認



ブラー(ボケ)も認識率低下に
影響している可能性有り



活用事例：21年度 自動運転DNN 未検出評価プロセス

DNNが正確に認識できなかった箇所を画像で確認

学習量評価

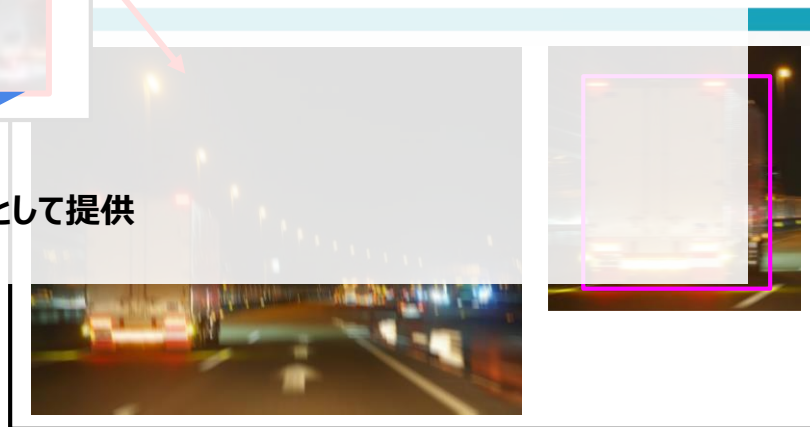
認識率評価

詳細分析

実データ
目視確認



対象データをDNNの学習データとして提供

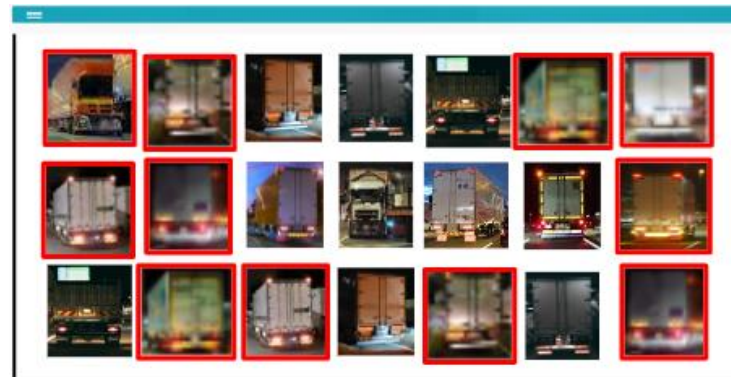
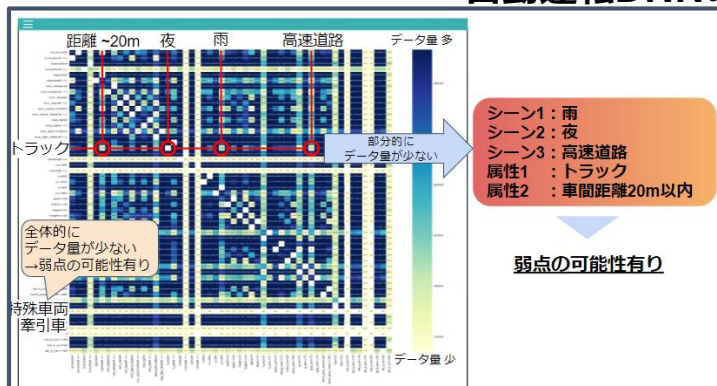


まとめ

近年、Data-Centricな考え方が提唱されているが、
現状のデータ分析では必要なデータをフィードバックできない可能性がある。

→**多次元の特徴量を基にデータ量を可視化することで効率的に高性能なAI開発を実現**

自動運転DNNの弱点シーンを分析



本ツールは自動運転用の画像認識AIだけでなく、一般的な画像認識AIでも適応できます。

使用希望の方は下記までご連絡ください（**研究目的での使用はフリー**）※商用利用は個別に相談

TTDC プラットフォーム開発部 AI・データサイエンス技術室 mail : mlops@ml.toyota-td.jp Twitter : [@TTDC keisoku](https://twitter.com/TTDC_keisoku)