AD and Collaboration for Labeling

Markus Loosen,
Lead Architect Connected Customer

Microsoft

Autonomous Driving and Labeling

- The road to L4/L5 involve collecting billions of miles of driving data from the real world and simulated world
- However the data is only useful if it has been labeled for Ground Truth
- The quality of Ground Truth impacts both the training and validation of the AI algorithms

Our footprint & references in ADAS

ACM Initiative: On April 4th 2018 the **American Center of Mobility** (ACM) in Detroit announced that Microsoft joined the Advisory Board and is the **exclusive cloud and data partner for testing and product development for automated vehicle technology**. Founding Partner of ACM are Ford, Toyota, Hyundai and more.

American Center for Mobility
CONNECTED. AUTOMATED. VALIDATED.

Based in Gaimersheim, Germany, **EFS** is the number one partner of Audi in chassis development. It examines and helps implement future-looking technologies, including automated driving. As part of its research efforts, the company used **Azure NC-series virtual machines** powered by NVIDIA Tesla P100 GPUs **to drive a deep learning AI solution** that analyzes high-resolution two-dimensional images of roads. The purpose is **to give self-driving vehicles a better understanding of those roads**. EFS proved that the concept works, and the company can now move ahead with product development.

EFS

Daimler AG, one of the world's largest manufacturers of premium cars and trucks, is driving hard to be a key player in software. To speed up software development and thus innovation, Daimler uses **Microsoft Azure DevTest Labs**. By developing in Azure, the company can onboard developers in hours versus weeks, get new ideas underway faster, and attract top talent with a state-of-the-art development environment. Peter Rothlaender, Manager of Cloud Solutions at Daimler, explains why Daimler is honking its horn about Azure DevTest Labs.

DAIMLER

Promote Open Capability
Apollo provides an **open, reliable and secure software platform** for its partners **to develop their own autonomous driving systems** through on-vehicle and hardware platforms.

apollo

cognata

ALP.Lab
Austrian Light Vehicle Proving Region for Automated Driving

e.GO:digital

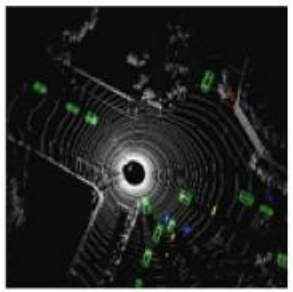
Generating Ground Truth with Labeling

Semantic Segmentation

Each Pixel of the image is assigned a category

Object Detection and Classification

Bounding box drawn around each object of interest



3D Point Cloud Labeling

Objects of interest as assigned a category in 3D LIDAR point cloud

"Ground truth" is the accuracy of the training set's classification for supervised learning techniques

- Currently done manually
- **Longer term – auto labelling**
- Partners provide
  - Results based managed service contracts
  - Trained workforce, on demand
  - Mature labeling tools

LINKER

sama SOURCE

cloudfactory

DiscoveryVest
Making an Impact in Ground Truth Discovery

INSIGHTGLOBAL

Ground Truth is one of the most critical elements of Machi ne Learning for Training and Validation

Current Labeling Approach

- **Unique definitions** per customer
- **No industry standard** for **object classes**
- **No standards** for **hierarchy of class/sub-class** defined
- **No standards** defined for **lane curvature, lane markings**
- Unique requirements per country/region
- **Different approach** per OEM even within same region
  - Road shoulder – drivable or not
  - Object visible through glass

Key Challenges

Cost

- Custom schema means manual **labeling workforce has to be retrained** for every customer
- This increases cost and time for everybody

Data Sharing

- Assume a few OMEs/Tier 1's decide to collaborate and **share data, this is not possible**
- Hinders availability of common data sets that can be used for research. Like Berkey Driving Dataset

Regulations

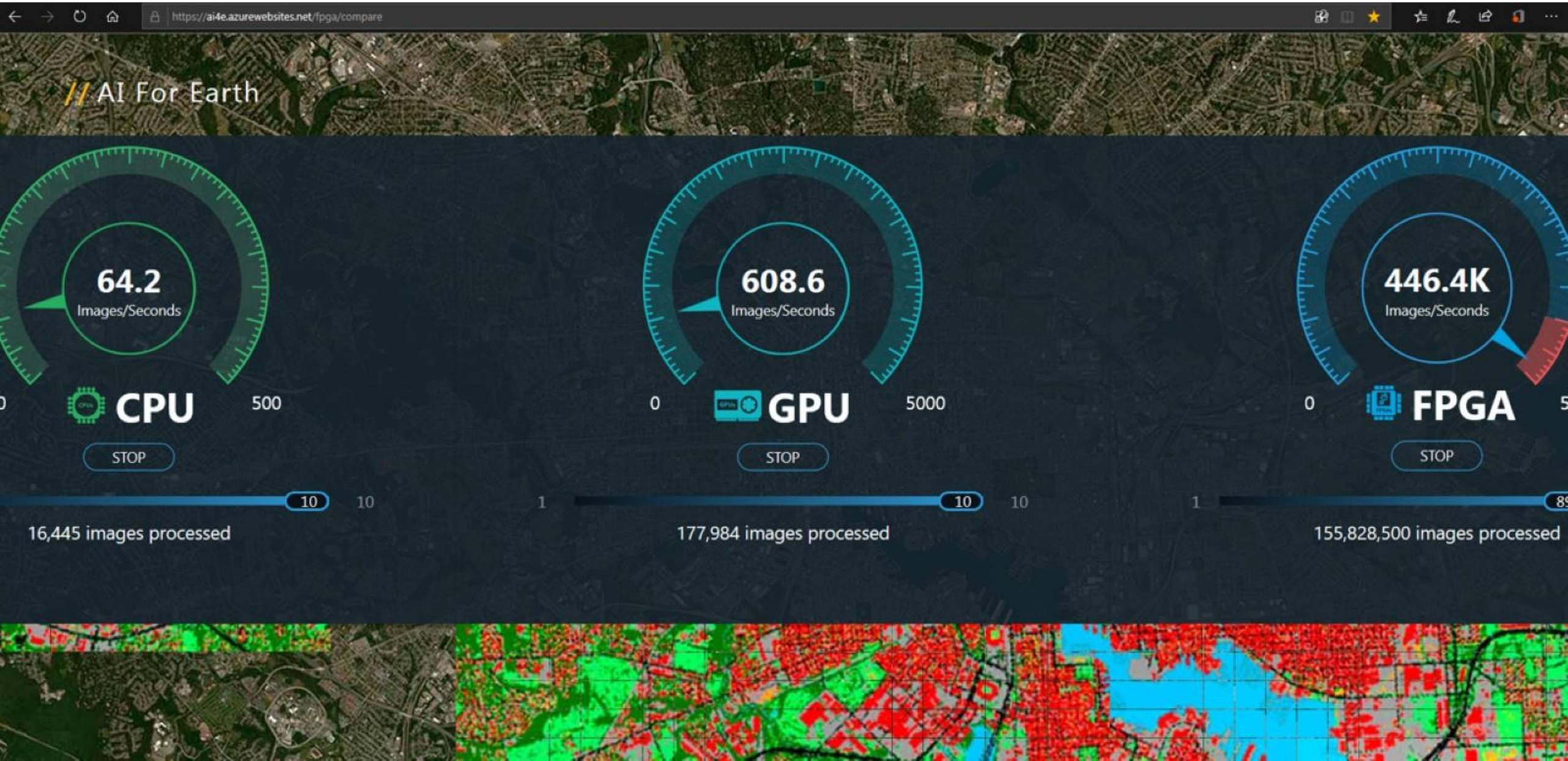
- Hypothetically - To make autonomous cars safer, **regulations could mandate OEMs to share data from incidents** with all auto makers

Addressing the Labeling Issue

There are already a few examples of open sourced labeling schemas that could be used as a starting point

1. Berkley Driving Data set –
https://github.com/ucbdrive/bdd-data/blob/master/doc/format.md
2. Apollo Data set -
http://data.apollo.auto/static/pdf/2d_obstacle_label_en.pdf

AI for Earth


https://ai4e.azurewebsites.net/fpga/compare
// AI For Earth
64.2
Images/Seconds
CPU
0
500
STOP
10
10
16,445 images processed
608.6
Images/Seconds
GPU
0
5000
STOP
1
10
10
177,984 images processed
446.4K
Images/Seconds
FPGA
0
STOP
1
155,828,500 images processed

Our Proposal

Work with OEMs and Tier 1's to define a standard schema for Labeling while also allowing for customization

Three tiers of information included in the label

1. Tier 1: System info: Acquisition equipment info, System info
    I. This information is static and does not change. It can be defined once per image folder and not needed per image
2. Tier 2: Common Labels
    I. This contains the labels for objects and obstacles defined as 2D bounding boxes, polygons, lanes etc
3. Tier 3: Custom Labels/modifier
    I. This enable any customizations needed to make the labels relevant for a region or a specific customer or to extend the labels

Thank you

Microsoft