[ALMA-AI]
[WednesdAIs]
Real-world industrial data.
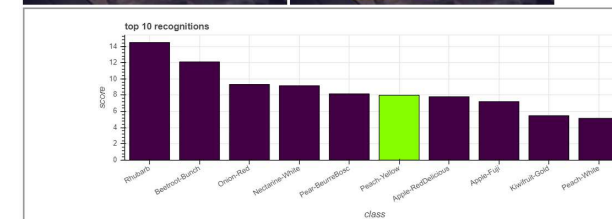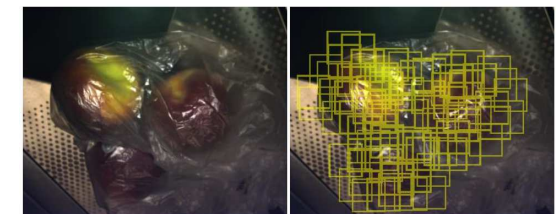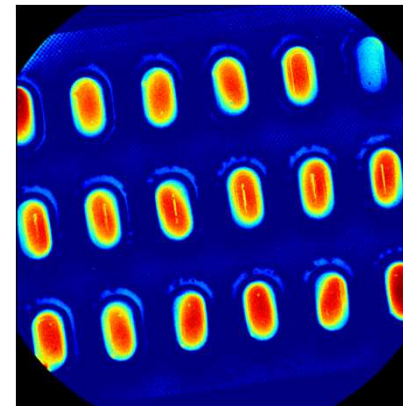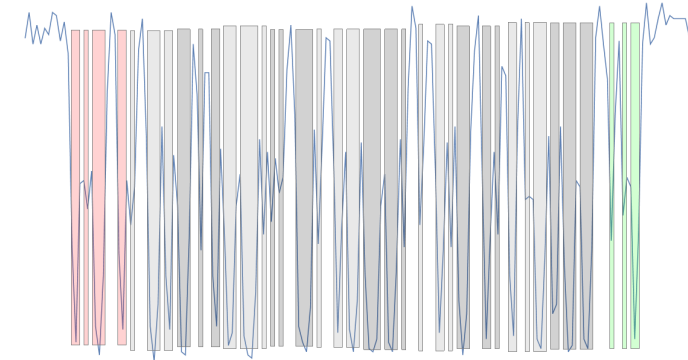
27 MAY 2020

Stefano Santi,   Datalogic USA

Some of «Data»Logic Applications

In Datalogic "Data" is available and used for

- 1D Barcodes
- 2D Barcodes
- Industrial and T&L OCR
- Hazardous Material Detection
- Quality inspection
- Produce Recognition
- General Item Identification
- etc.

«Data» Properties

"Data" is a means to obtain a "good" model, i.e. the end goal is a "good" model

"Data" Expenses may be related to
- Data Collection
- Data Annotation
- Data Anonymization
- Data Generation
- Data Storage
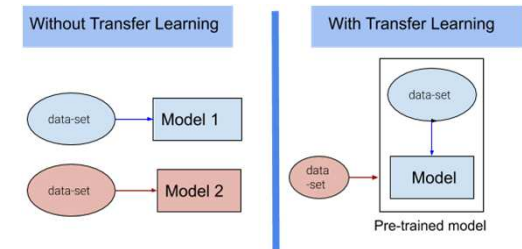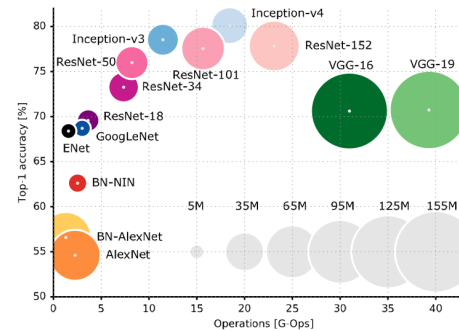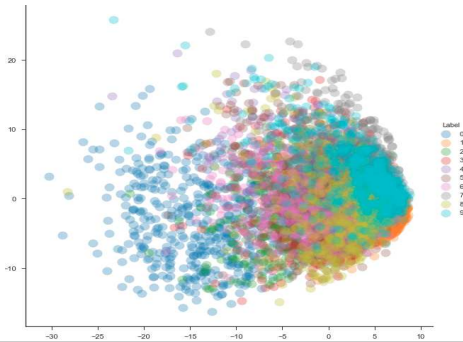- Licensing Fees
- etc.

"Data" Availability may be subject to
- Contract w/ Customer
- Data Protection Regulations: personal and sensitive data

"Data" Type may be:
- "Real": Most Expensive and possibly Not Available
- "Augmented": Cheap but depends on "Real" and must be Representative
- Synthetic: Cheap but needs Representative Generative Model

«Data» Size

- No "Rule of Thumb" here (I apologize for that...)
- We do not want to use more data than strictly necessary to find a "good enough model"
- Minimum Size depends on:

"Data Distribution" ...unknown ("dog chasing its tail" problem):
- visualize data
- Try different models
- Simple Models First!

"Model Complexity"...very related to "Data Distribution" but the choice of Models is very limited while the "Data Distribution" can be very different depending on the Applications

Availability of "Pretrained Models"
- Transfer Learning