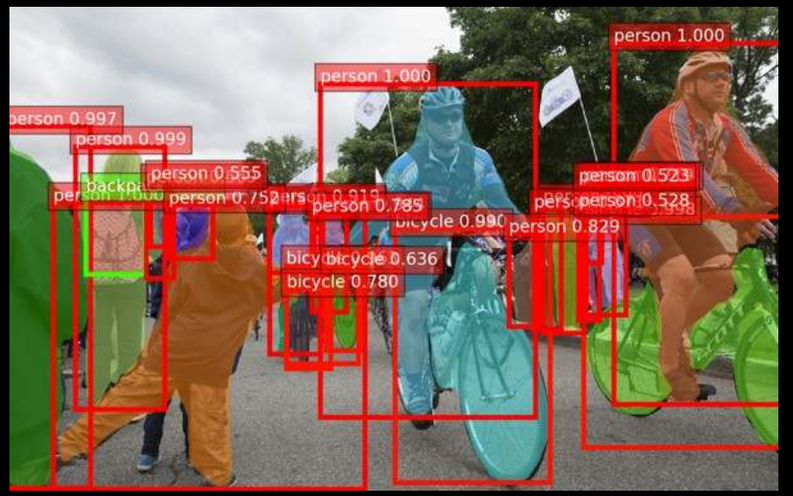
Deep Learning on AWS

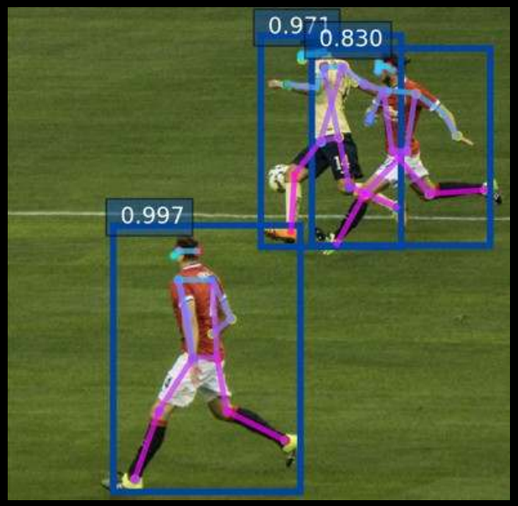
Julien Simon
Global Evangelist, AI & Machine Learning
@julsimon

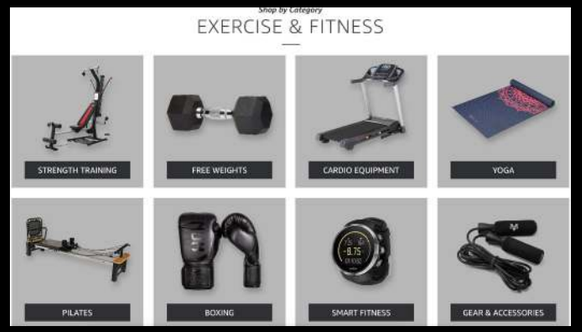
© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

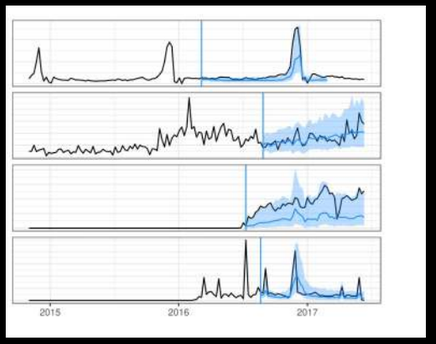
Deep Learning is changing the IT landscape

- Image and video analysis
- Natural language understanding
- Machine translation
- Speech processing
- Structured data too:
  time-series,
  personalization,
  etc.



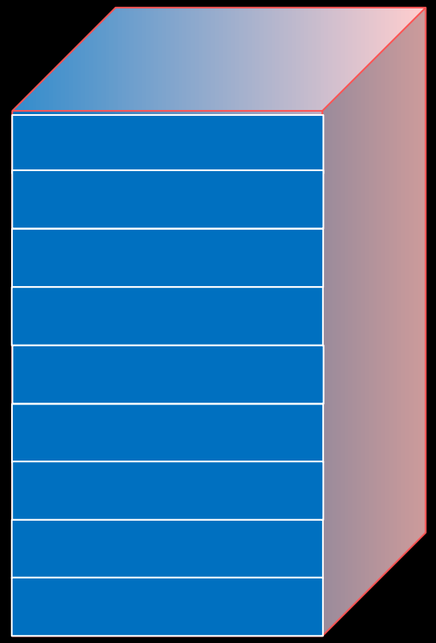






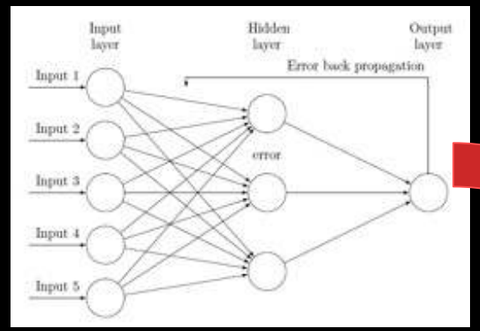
What does it take to train a Deep Learning model?

Training a neural network



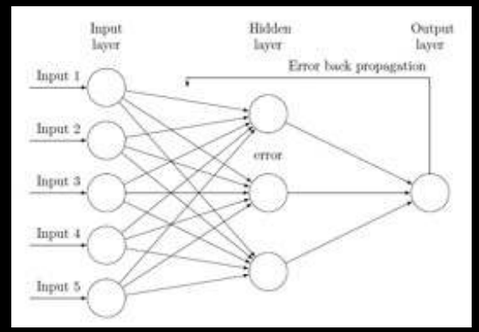
Training data set

Forward propagation



Backpropagation

Network in training



Trained network

You need to build datasets, design networks, select hyper-parameters,
train models, deploy them in production…
and of course you need to manage infrastructure.

Our mission at AWS

Put AI and machine learning
in the hands of every developer

The Amazon ML Stack: Broadest & Deepest Set of Capabilities

AI SERVICES

Vision


REKOGNITION IMAGE


REKOGNITION VIDEO



TEXTRACT

Speech


POLLY


TRANSCRIBE

Language


TRANSLATE


COMPREHEND

COMPREHEND MEDICAL

Chatbots


LEX

Forecasting



FORECAST

Recommendations



PERSONALIZE

ML SERVICES


AMAZON SAGEMAKER

BUILD

Pre-built algorithms & notebooks

Data labeling (GROUND TRUTH)

Algorithms & models (AWS MARKETPLACE)

TRAIN

One-click model training & tuning

Optimization (NEO)

Models without training data (REINFORCEMENT LEARNING)

DEPLOY

One-click deployment & hosting

ML FRAMEWORKS & INFRASTRUCTURE

Frameworks

TensorFlow

mxnet

PYTORCH

Interfaces

GLUON

Keras

Infrastructure


EC2 P3 & P3dn


EC2 C5


FPGAs


GREENGRASS



ELASTIC INFERENCE

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Trademark

aws

AI Services

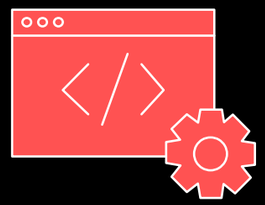
Call an API, job done!

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

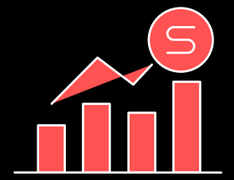
Put AI to work in minutes

AI SERVICES	Vision		Speech		Language		Chatbots	Forecasting	Recommendations
	REKOGNITION IMAGE	REKOGNITION VIDEO	TEXTRACT	POLLY	TRANSCRIBE	TRANSLATE	COMPREHEND & COMPREHEND MEDICAL	LEX	FORECAST	PERSONALIZE

Pre-trained AI services that require no ML skills



Easily add intelligence to your existing apps and workflows



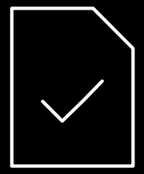
Quality and accuracy from continuously-learning APIs

Amazon SageMaker

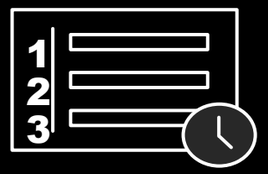
Build, train, and deploy your models at any scale on managed infrastructure

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker

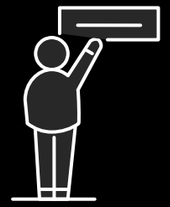


Collect and prepare training data



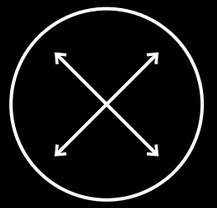
Choose and optimize your ML algorithm

Set up and manage environments for training



Train and Tune ML Models

Deploy models in production



Scale and manage the production environment

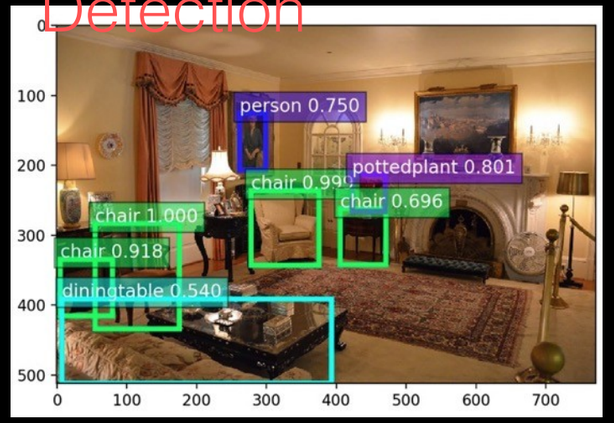
Same service and APIs from experimentation to production

Built-in algorithms for Deep Learning

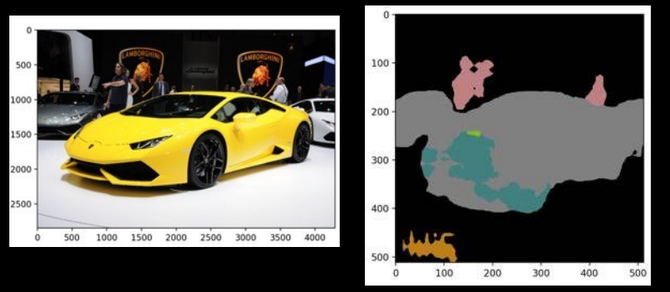
Image classification

[electric_guitar],
with probability 0.671

Object Detection



Semantic Segmentation



Time-series
(DeepAR)

Word embeddings
(Word2Vec)

Machine translation
(seq2seq)

General-purpose embeddings
(Object2Vec)

Built-in Deep Learning frameworks: just add your code

- Built-in containers for training and prediction.
  - Code available on Github, e.g. https://github.com/aws/sagemaker-tensorflow-containers
  - Build them, run them on your own machine, customize them, etc.
- Script mode: use the same code as on your laptop

No infrastructure work required: simply define instance type and instance count

Distributed training out of the box: zero setup

Pipe mode: stream infinitely large datasets directly from Amazon S3

```
tf_estimator = TensorFlow(entry_point='mnist_keras_tf.py',
                          role=role,
                          train_instance_count=1,
                          train_instance_type='ml.p3.2xlarge',
                          framework_version='1.12',
                          py_version='py3',
                          script_mode=True,
                          hyperparameters={
                              'epochs': 20,
                              'batch-size': 256,
                              'learning-rate': 0.01}
                         )
```

Frameworks & infrastructure

Total control for DIY enthusiasts and advanced users

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Deep Learning AMIs

Preconfigured environments on Amazon Linux or Ubuntu

NEW (March 27th) Deep Learning containers

Conda AMI
For developers who want pre-installed pip packages of DL frameworks in separate virtual environments.

Base AMI
For developers who want a clean slate to set up private DL engine repositories or custom builds of DL engines.

AMI with source code
For developers who want preinstalled DL frameworks and their source code in a shared Python environment.

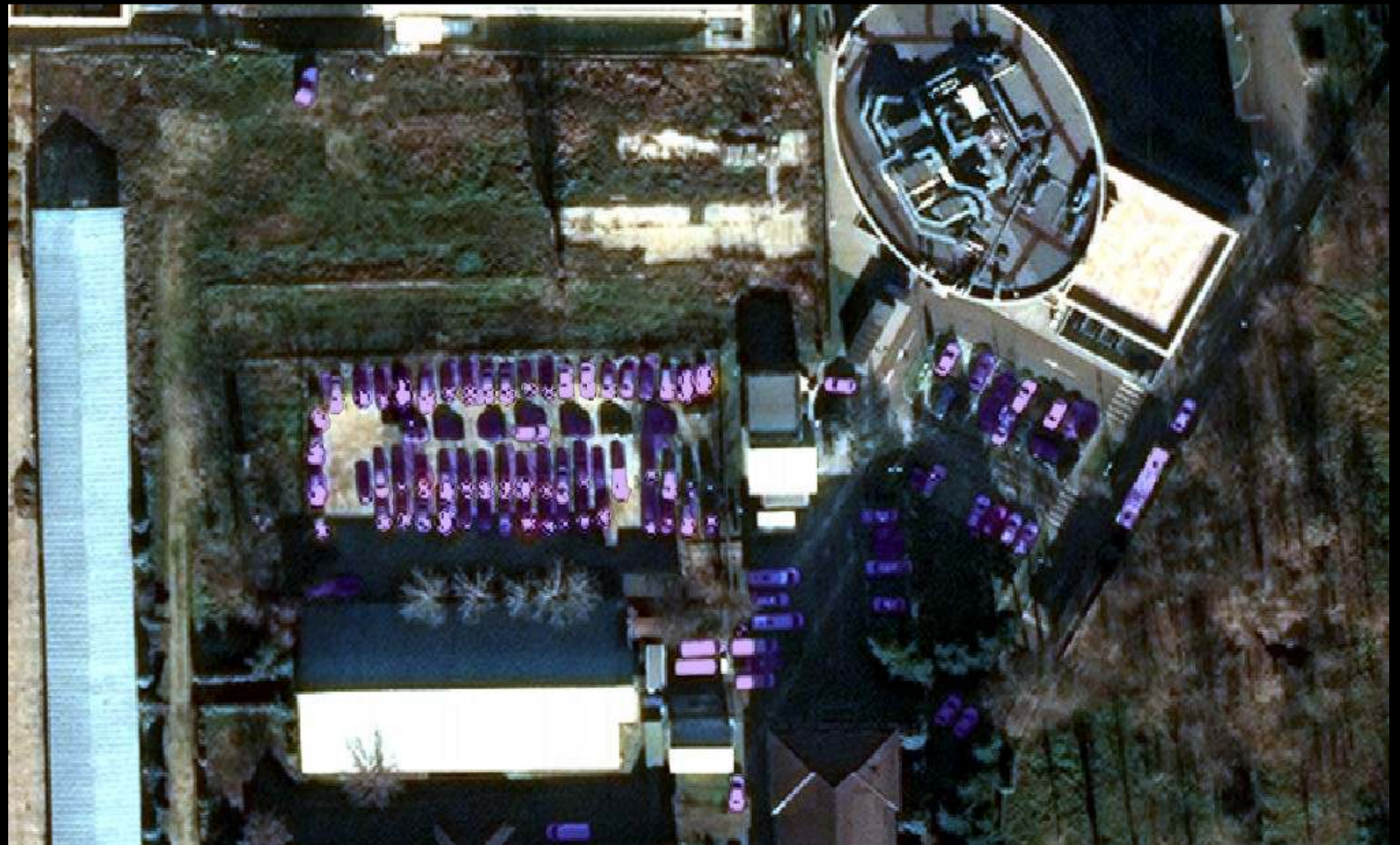
AWS: The platform of choice to run TensorFlow

85% of all TensorFlow workloads in the cloud runs on AWS

Source: Nucleus Research, November 2018

Satellite image analysis with Tensorflow

https://www.earthcube.eu/



Optimizing Tensorflow for Amazon EC2 instances

C5 instances (Intel Skylake)

Training ResNet-50 with the ImageNet dataset using our optimized build of Tensorflow 1.11 on a c5.18xlarge instance type is 11x faster than training on the stock binaries.

P3 instances (NVIDIA V100)

Tensorflow scaling efficiency with 256 GPUs

65% Stock version

90% AWS-optimized version

Apache MXNet: Deep Learning for enterprise developers

Start with off-the-shelf models

- Gluon CV and Gluon NLP
- ONNX compatibility

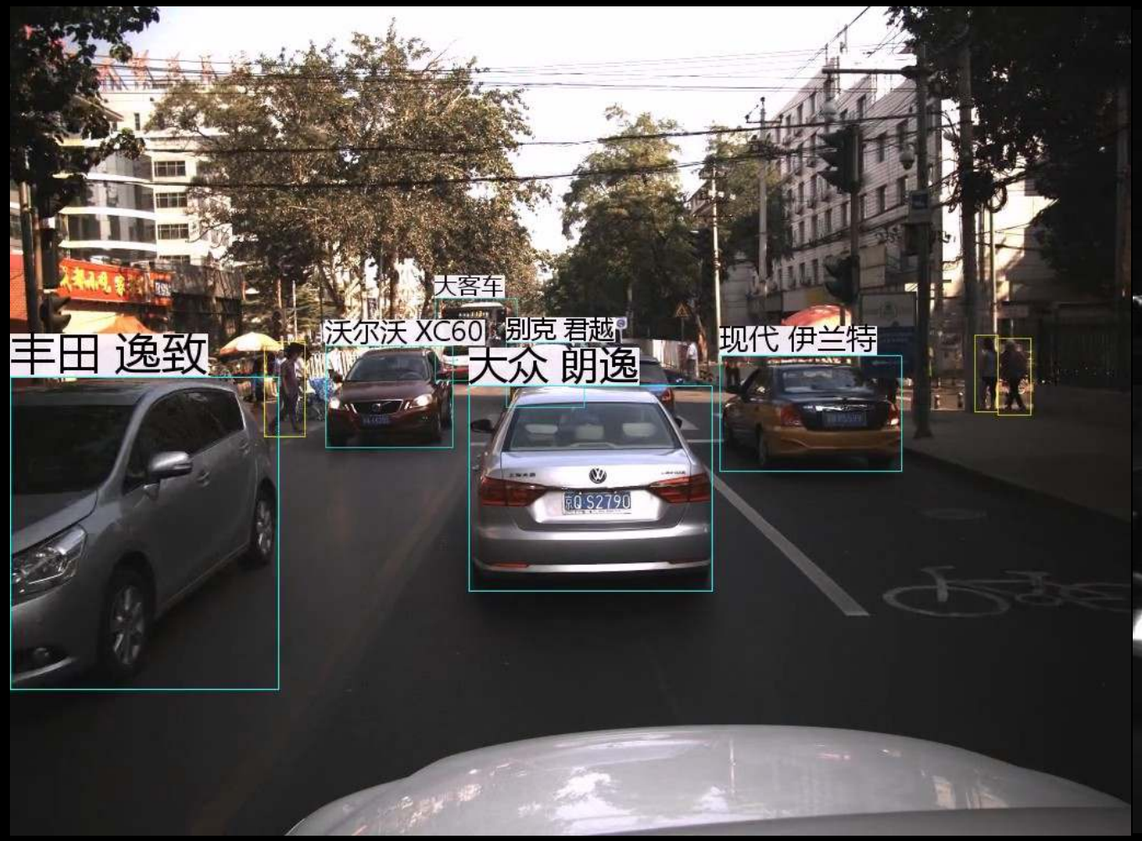
Fast and scalable training

- Keras-MXNet up to 2x faster than Keras-TensorFlow
- Near-linear scalability up to 256 GPUs
- Dynamic training

Easy deployment

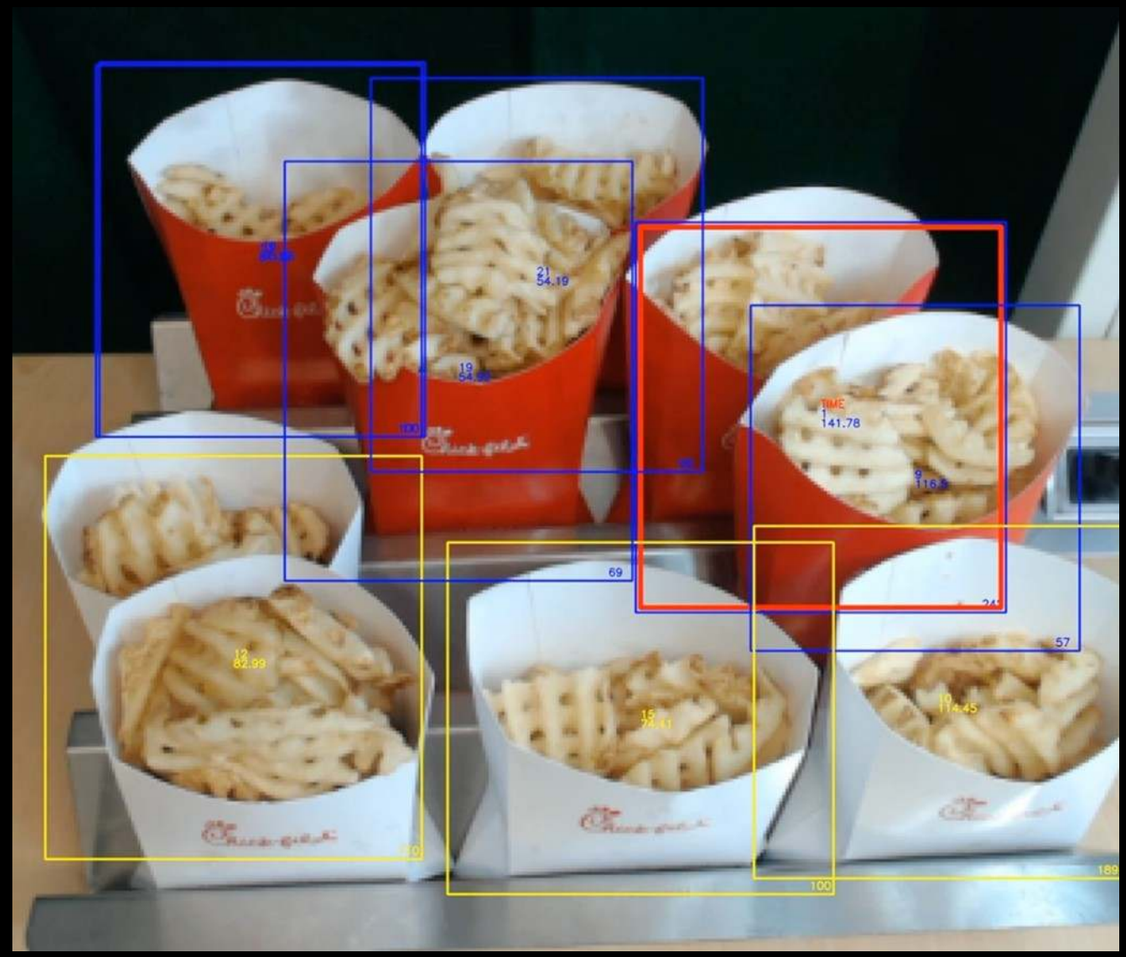
- Java/Scala APIs
- Model Server

Autonomous driving with Apache MXNet



Keeping fries fresh with Apache MXNet

https://www.youtube.com/watch?v=dKcyAjCtXqc



Track waffle fry freshness

Identify fries that have exceeded the hold time

Solution based on Deep Learning

Computer vision model for object detection and tracking

Model built on Apache MXNet

A team of enterprise developers with no ML expertise

GluonCV

https://gluon-cv.mxnet.io
https://github.com/dmlc/gluon-cv

- State-of-the-art deep learning tools for computer vision
  - Pre-trained models
  - Training and fine-tuning scripts
  - Prototype products, validate new ideas and learn computer vision

- Image classification: 50+ models
- Object detection: Faster RCNN, SSD, Yolo-v3
- Semantic segmentation: FCN, PSP, DeepLab v3
- Instance segmentation: Mask RCNN
- Pose estimation: Simple Pose
- Person re-identification (Market1501 dataset)
- GANs: Wasserstein GAN, Super Resolution GAN, CycleGAN

GluonNLP

https://gluon-nlp.mxnet.iohttps://github.com/dmlc/gluon-nlp

- State-of-the-art deep learning tools for natural language processing
  - Pre-trained models and embeddings
  - Training and fine-tuning scripts
  - Prototype products, validate new ideas and learn NLP

- Word embeddings: Word2Vec, FastText, GloVE, BERT
- Machine translation: GNMT, Transformer
- Sentiment analysis: TextCNN
- Text classification: FastText
- Language models
- Text generation
- Natural language inference
- Parsing

Getting started

http://aws.amazon.com/free

https://ml.aws
https://aws.amazon.com/sagemaker
https://github.com/aws/sagemaker-python-sdk
https://github.com/awslabs/amazon-sagemaker-examples
https://aws.amazon.com/machine-learning/amis/

https://medium.com/@julsimon
https://gitlab.com/juliensimon/dlnotebooks

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Julien Simon
Global Evangelist, AI and Machine Learning

@julsimon

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.