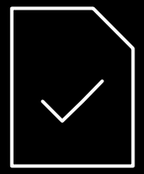
Floor 28, Tel Aviv, July 7th, 2019

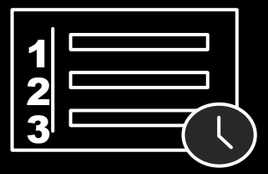
Deep Learning on Amazon SageMaker

Julien Simon
Global Evangelist, AI & Machine Learning, AWS
@julsimon

Amazon SageMaker

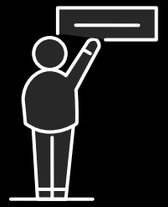


Collect and prepare training data



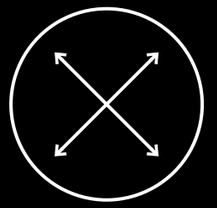
Choose and optimize your ML algorithm

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

Model options

Training code

Factorization Machines
Linear Learner
Principal Component Analysis
K-Means Clustering
XGBoost
And more

Built-in Algorithms (17)

No ML coding required
No infrastructure work required
Distributed training
Pipe mode

mxnet Chainer TensorFlow PyTorch scikit-learn

Built-in Frameworks

Bring your own code: script mode
Open source containers
No infrastructure work required
Distributed training
Pipe mode

Bring Your Own Container

Full control, run anything!
R, C++, etc.
No infrastructure work required

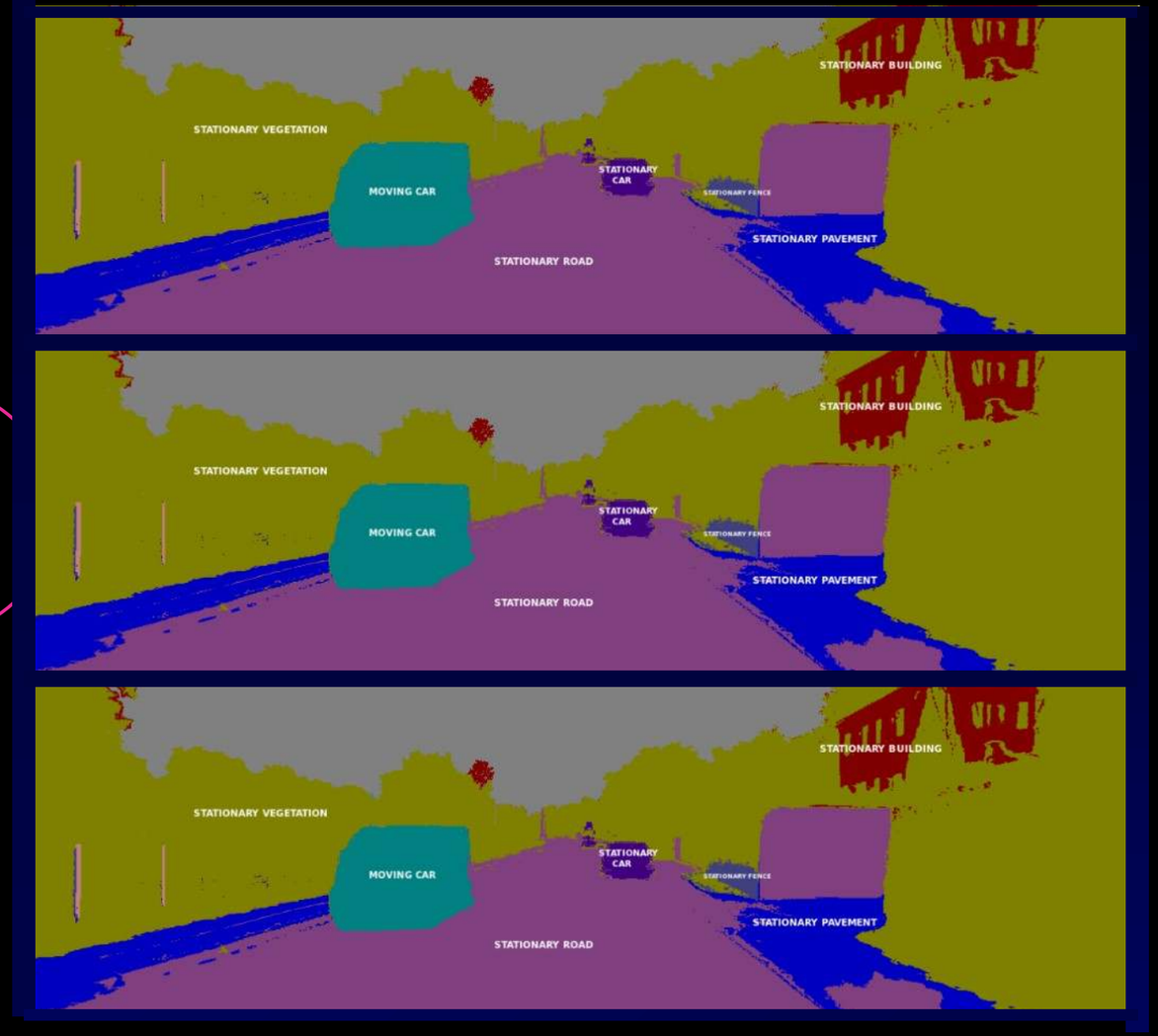
The Amazon SageMaker API

- Python SDK orchestrating all Amazon SageMaker activity
  - High-level objects for algorithm selection, training, deploying, automatic model tuning, etc.
  - Spark SDK (Python & Scala)

- AWS SDK
  - For scripting and automation
  - CLI : '*aws sagemaker*'
  - Language SDKs: boto3, etc.

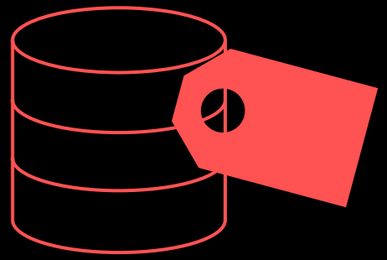
Labeling Datasets for Deep Learning

Annotating data at scale is time-consuming and expensive

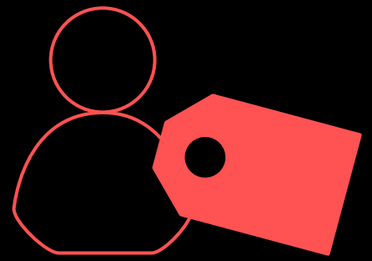


Amazon SageMaker Ground Truth

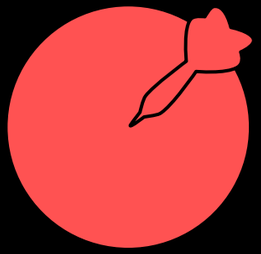
Build scalable and cost-effective labeling workflows



Quickly label training data



Easily integrate human labelers



Get accurate results

KEY FEATURES

Automatic labeling via machine learning

Ready-made and custom workflows for image bounding box, segmentation, and text

Private and public human workforce

Integrated with Deep Learning algorithms in Amazon SageMaker

Demo:

labeling images for object detection

To see how labeled images can be easily be used for training, please look at: https://github.com/awslabs/amazon-sagemaker-examples/tree/master/ground_truth_labeling_jobs

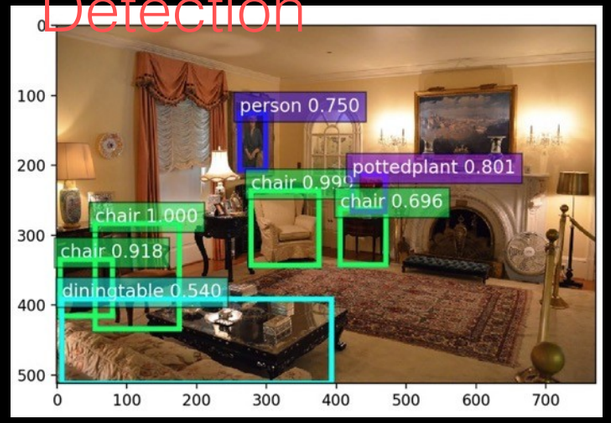
Built-in algorithms for Deep Learning

Built-in algorithms for Deep Learning

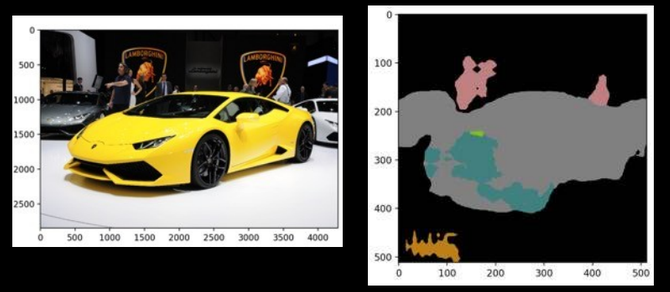
Image classification

[electric_guitar],
with probability 0.671

Object Detection



Semantic Segmentation



Time-series
(DeepAR)

Word embeddings
(BlazingText)

Machine translation
(seq2seq)

General-purpose embeddings
(Object2Vec)

Demo:

Built-in image classification with transfer learning

https://gitlab.com/juliensimon/dlnotebooks/blob/master/sagemaker/06-Image-classification-deeplens.ipynb

Demo:

Text classification with BlazingText

https://github.com/awslabs/amazon-sagemaker-examples/tree/master/introduction_to_amazon_algorithms/blazingtext_text_classification_dbpedia

Built-in frameworks for Deep Learning

Built-in frameworks: just add your code

- Built-in containers for training and prediction.
  - Available on Github, e.g. https://github.com/aws/sagemaker-tensorflow-containers
  - Build them, run them on your own machine, customize them, etc.

- Local mode: train and predict on your local machine

- Script mode: use the same code as on your local machine

AWS: the platform of choice to run Tensorflow

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

Source: Nucleus Research, November 2018

Optimizing Tensorflow on AWS

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

Demo:

Image classification with Keras/Tensorflow and Keras/MXNet

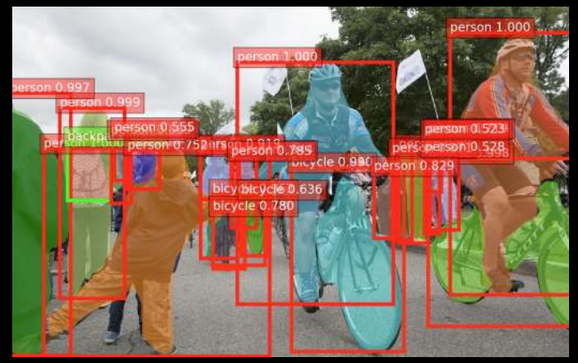
https://gitlab.com/juliensimon/dlnotebooks/tree/master/keras/05-keras-blog-post

Bonus: a quick look at Gluon

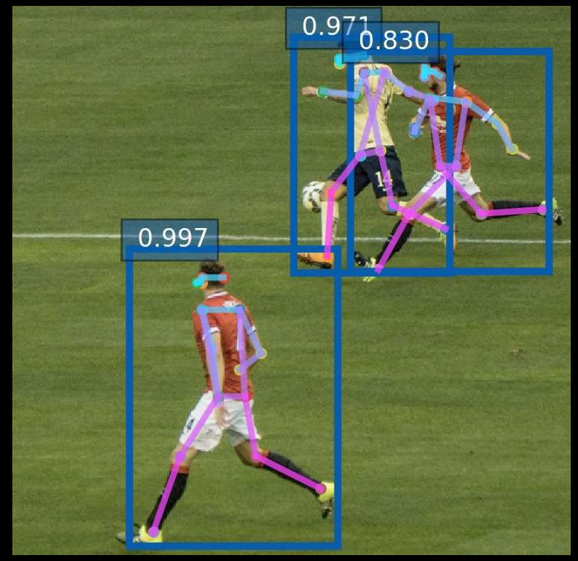
GluonCV

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

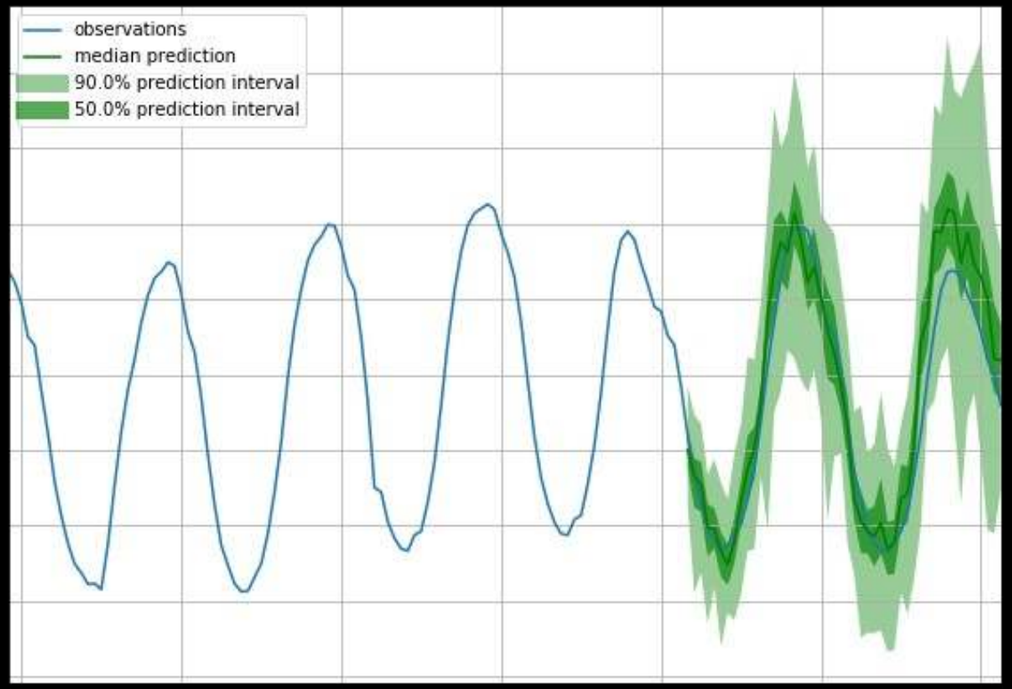
- State-of-the-art deep learning tools for natural language processing
  - Pre-trained models and embeddings
  - Training and fine-tuning scripts
  - Prototype products, validate new ideas and learn NLP

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

GluonTS

https://gluon-ts.mxnet.io/
https://github.com/awslabs/gluon-ts

- State-of-the-art deep learning tools for time-series forecasting
- Real and artificial datasets
- Loading and iterating over time series datasets
- Models ready to be trained
- Building blocks to define your own models



Demo: GluonCV

https://gitlab.com/juliensimon/dlnotebooks/tree/master/gluoncv

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/aws/sagemaker-spark
https://github.com/awslabs/amazon-sagemaker-examples

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

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

Merci!

Julien Simon
Global Evangelist, AI & Machine Learning, AWS
@julsimon