Deep Learning at Scale

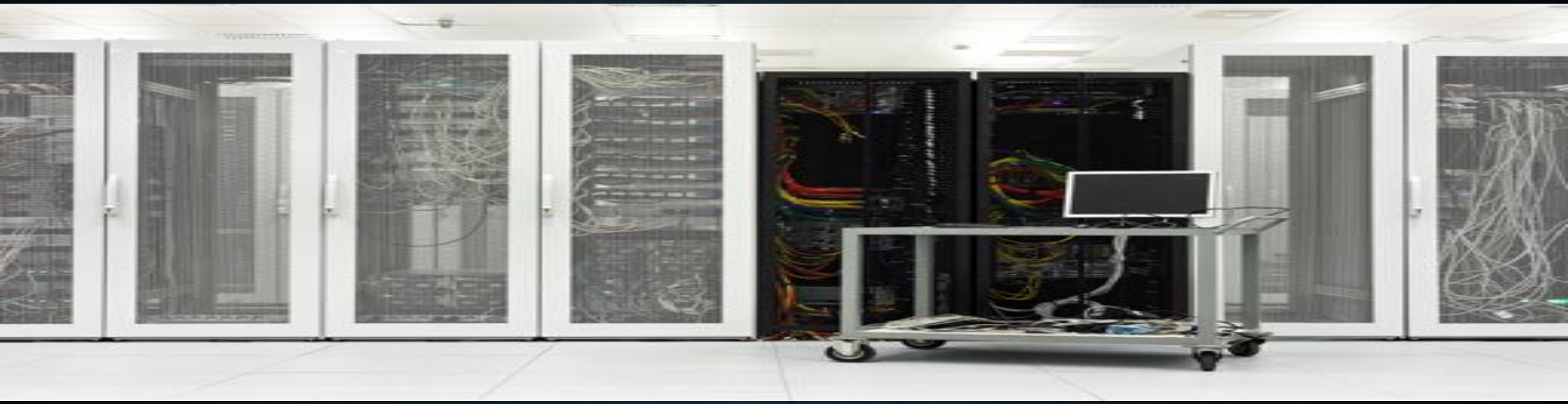
Dr. Nathalie Rauschmayr
Applied Scientist at AWS

aws

Challenges in deep learning

Data and annotations

Challenges in deep learning



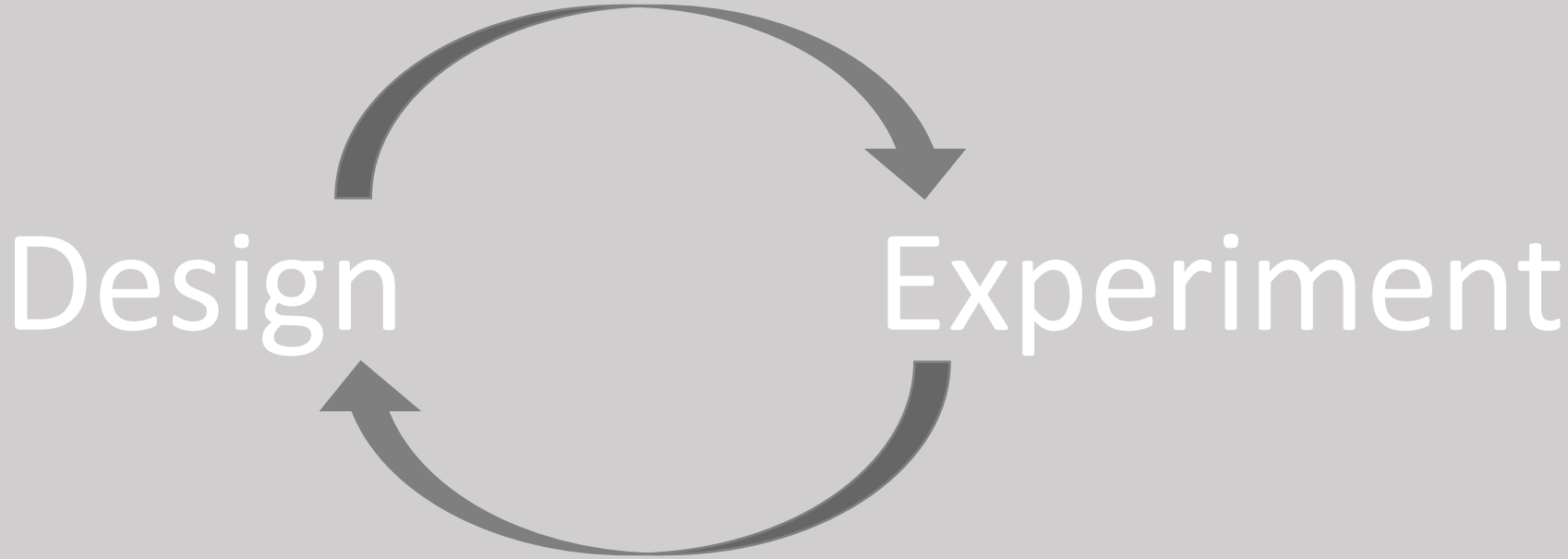
Compute and scale

Challenges in deep learning

Hyperparameter tuning

aws

Challenges in deep learning



Reproducibility

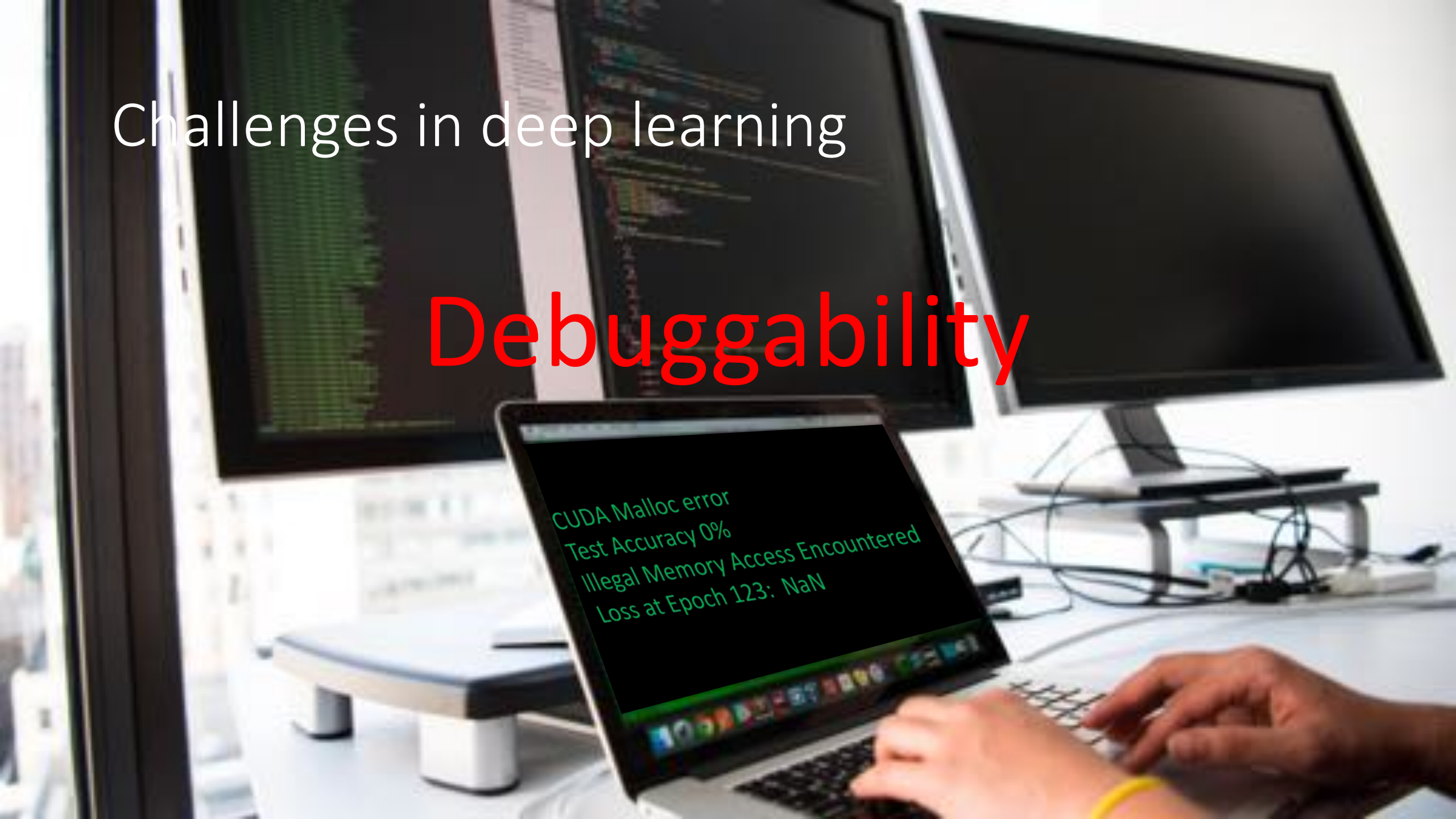
Challenges in deep learning

Image taken from https://shaoanlu.wordpress.com/2017/05/07/vihicle-detection-using-ssd-on-floybhub-udacity-self-driving-car-nano-degree/

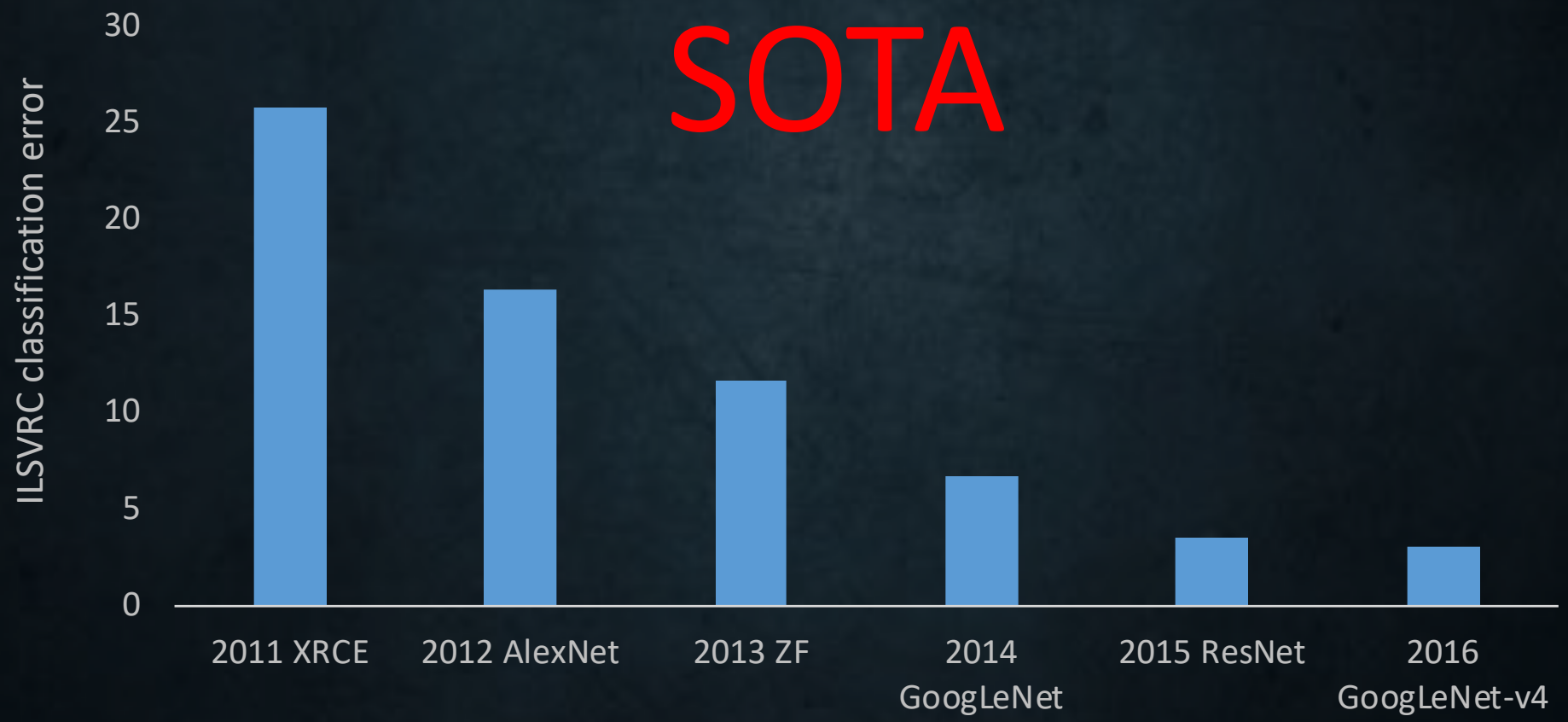
IoT and real time

aws

Challenges in deep learning



Challenges in deep learning



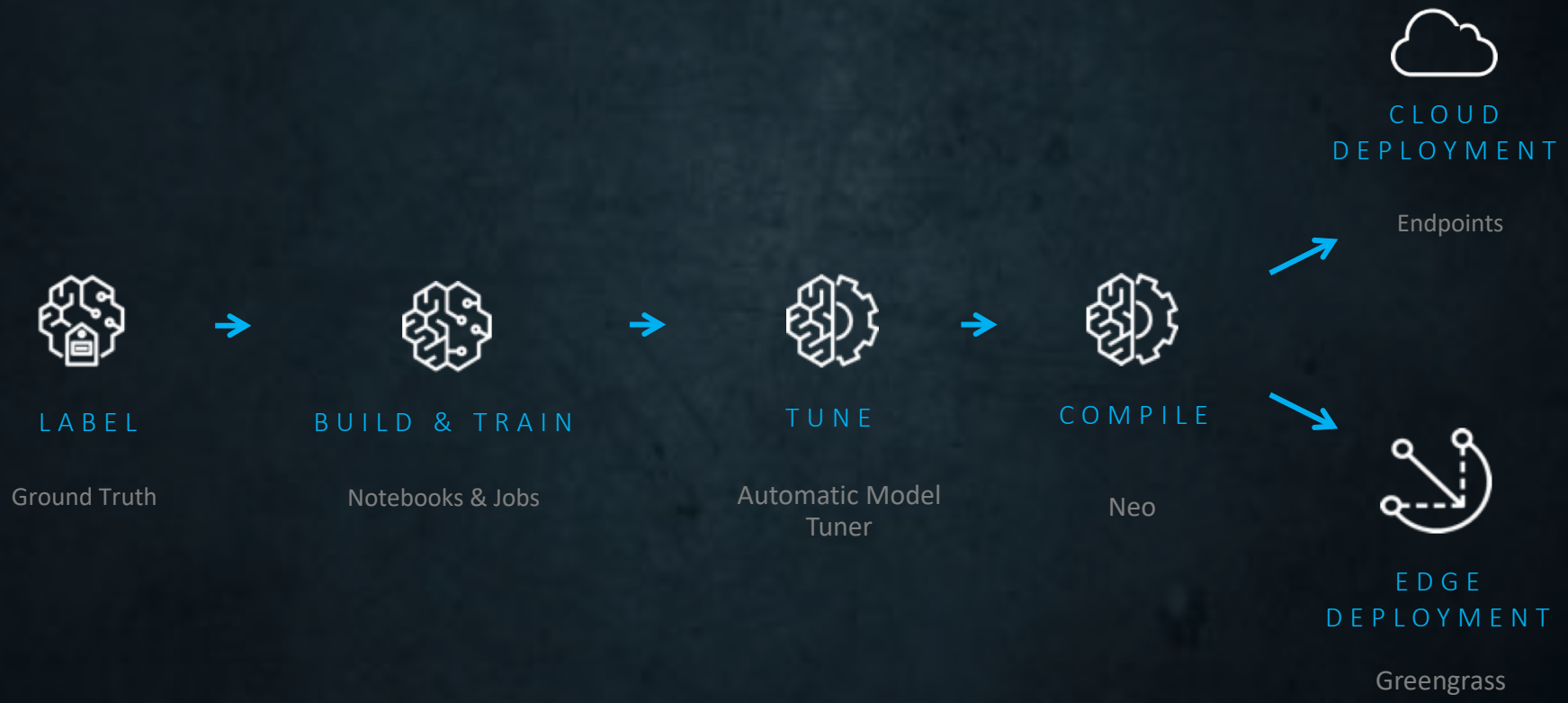
How to tackle Deep Learning Challenges

The Amazon ML Stack

ML for everyone: different users require different tools

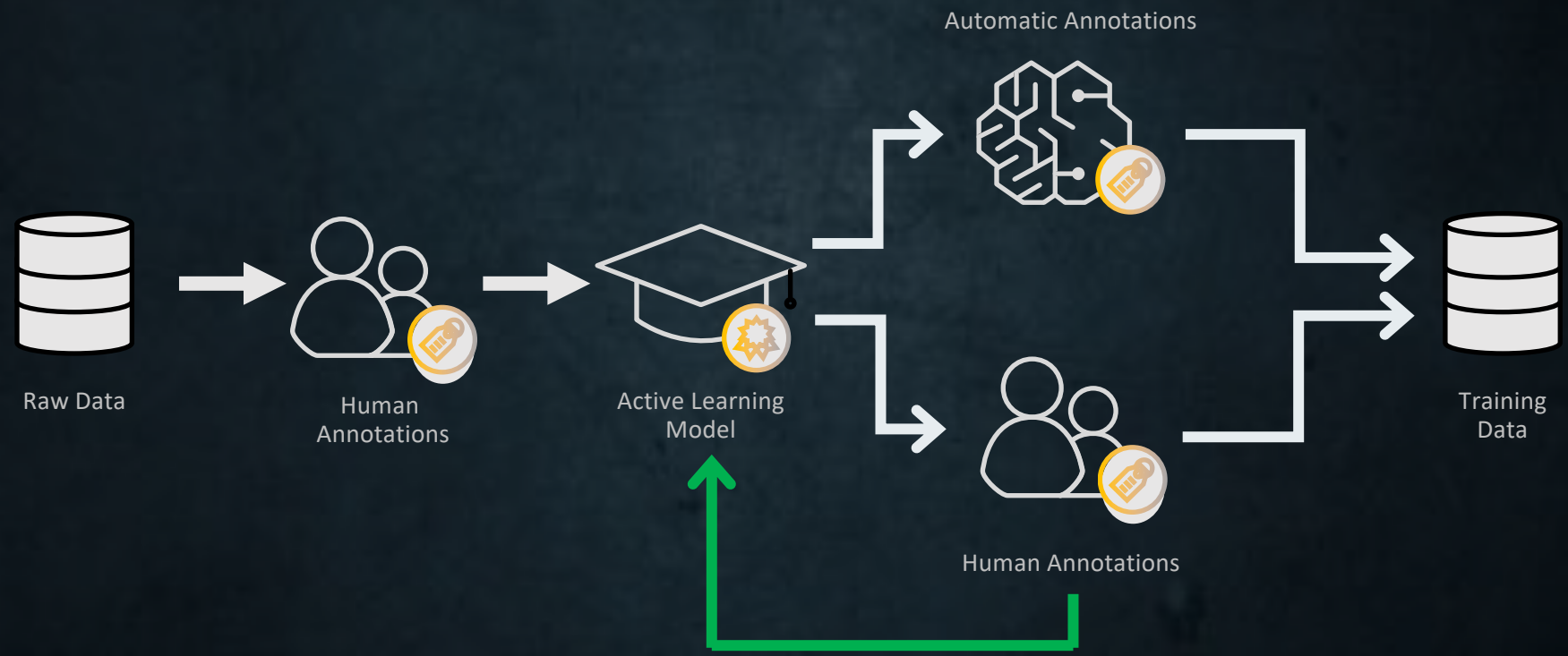
Amazon SageMaker: Build, Train, and Deploy

Takes away the heavy lifting of ML



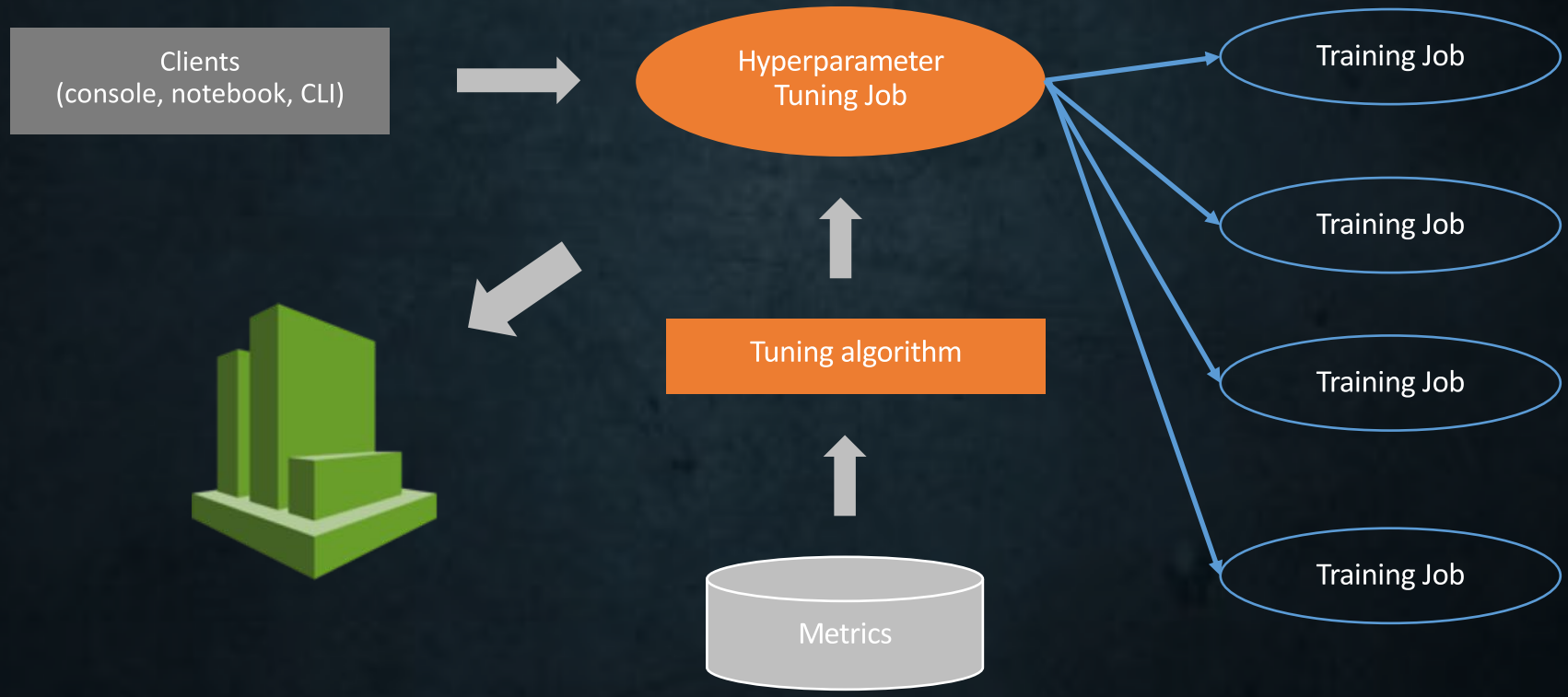
SageMaker Groundtruth

Quickly create high quality training data



SageMaker automatic model tuner

Automatically searches for the right set of hyperparameters



SageMaker Neo

Train once and run anywhere with up to 2x performance improvement

Amazon Elastic Inference

Inference amounts to 90% of production costs. Elastic inference helps to reduce those costs up to 75%

Lower inference costs

Match capacity to demand

Available between 1 to 32 TFLOPS per accelerator

KEY FEATURES

Integrated with Amazon EC2 and Amazon SageMaker

Support for TensorFlow, Apache MXNet, and ONNX with PyTorch coming soon

Single and mixed-precision operations

aws

Deep Learning with Apache MXNet

A flexible and scalable library for deep learning

Scalable

Debuggable

Flexible

Optimized libraries

7 frontend languages

Portable

APACHE INCUBATOR

aws

Deep Learning with Apache MXNet

Choose your most preferred language








Frontend

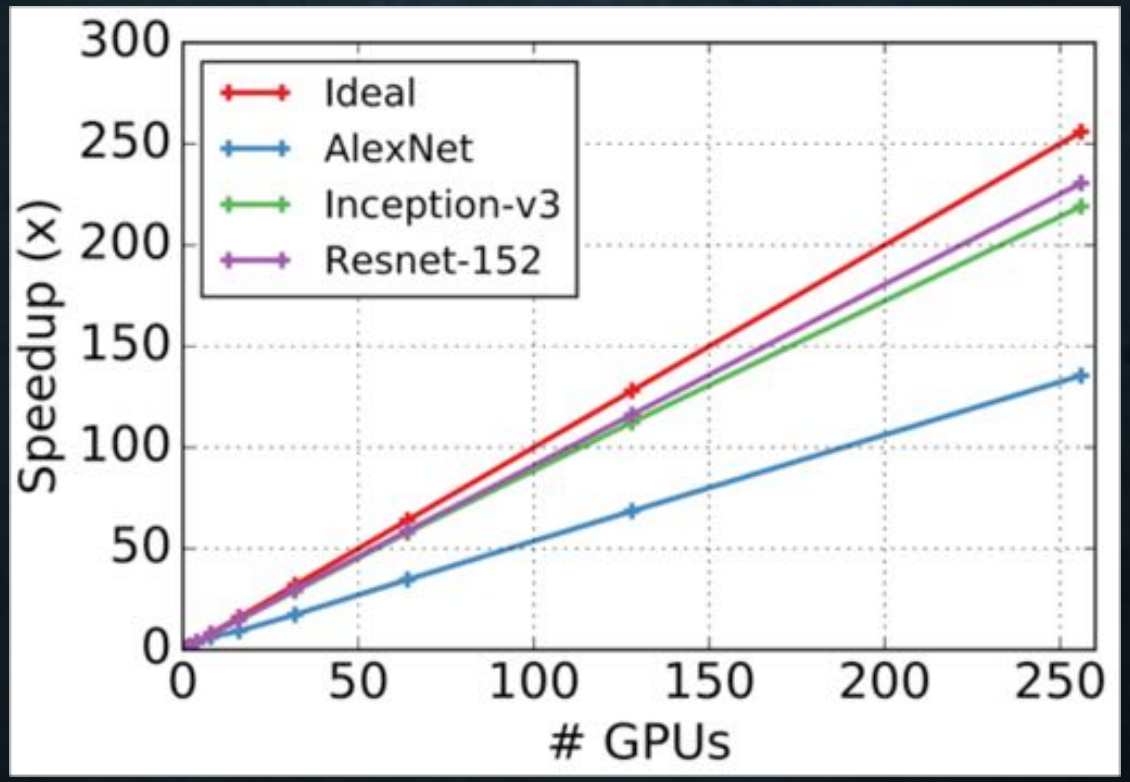
Backend

aws

Deep Learning with Apache MXNet

Nearly linear scaling across hundreds of GPUs

16 x p2.16 xlarge

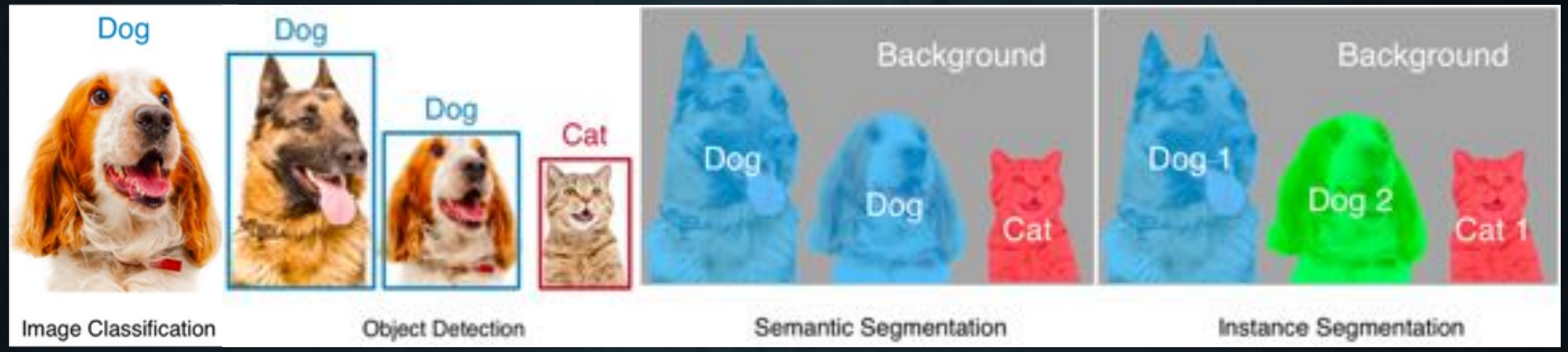


https://mxnet.incubator.apache.org/versions/master/tutorials/vision/large_scale_classification.html

aws

MXNet Toolkits - GluonCV

Quickly produce state of the art results with just a few lines of codes

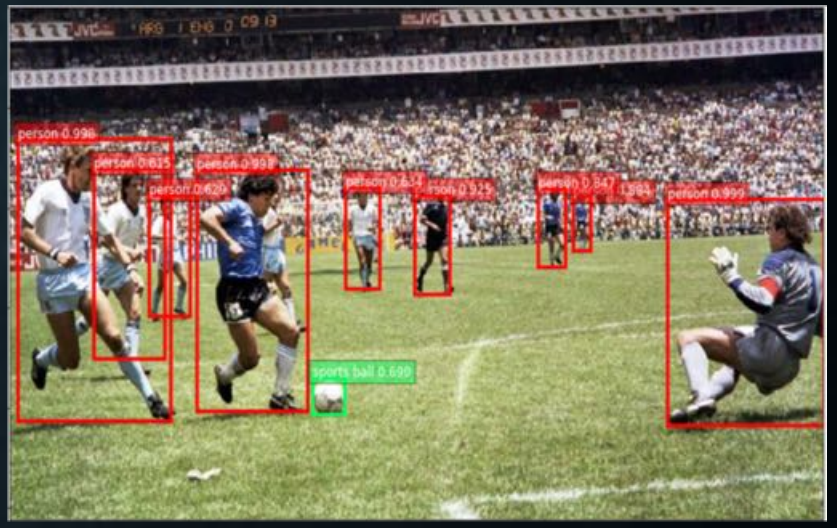


50+ Pre-trained models, with training scripts, datasets, tutorials

MXNet Toolkits - GluonCV

Quickly produce state of the art results with just a few lines of codes

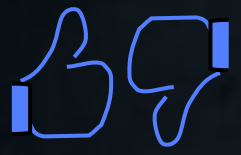
```
x, img = gcv.data.transforms.presets.ssd.load_test('support/diego.jpg', short=512)
ctx = mx.gpu()

net = gcv.model_zoo.get_model('ssd_512_resnet50_v1_coco', pretrained=True, ctx=ctx)

class_IDs, scores, bounding_boxes = net(x.as_in_context(ctx))

viz.plot_bbox(img, bounding_boxes[0], scores[0], class_IDs[0], class_names=net.classes)
```



aws

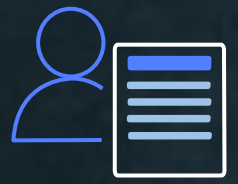
MXNet Toolkits - GluonNLP

Quickly produce state of the art results with just a few lines of codes

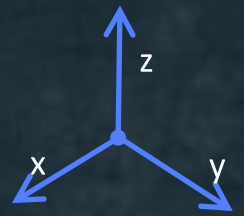


Sentiment Analysis

Text Generation



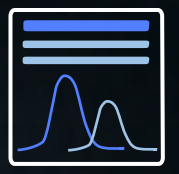
Named Entity Recognition



Representation Learning

Machine Translation

Question Answering



Language Modeling

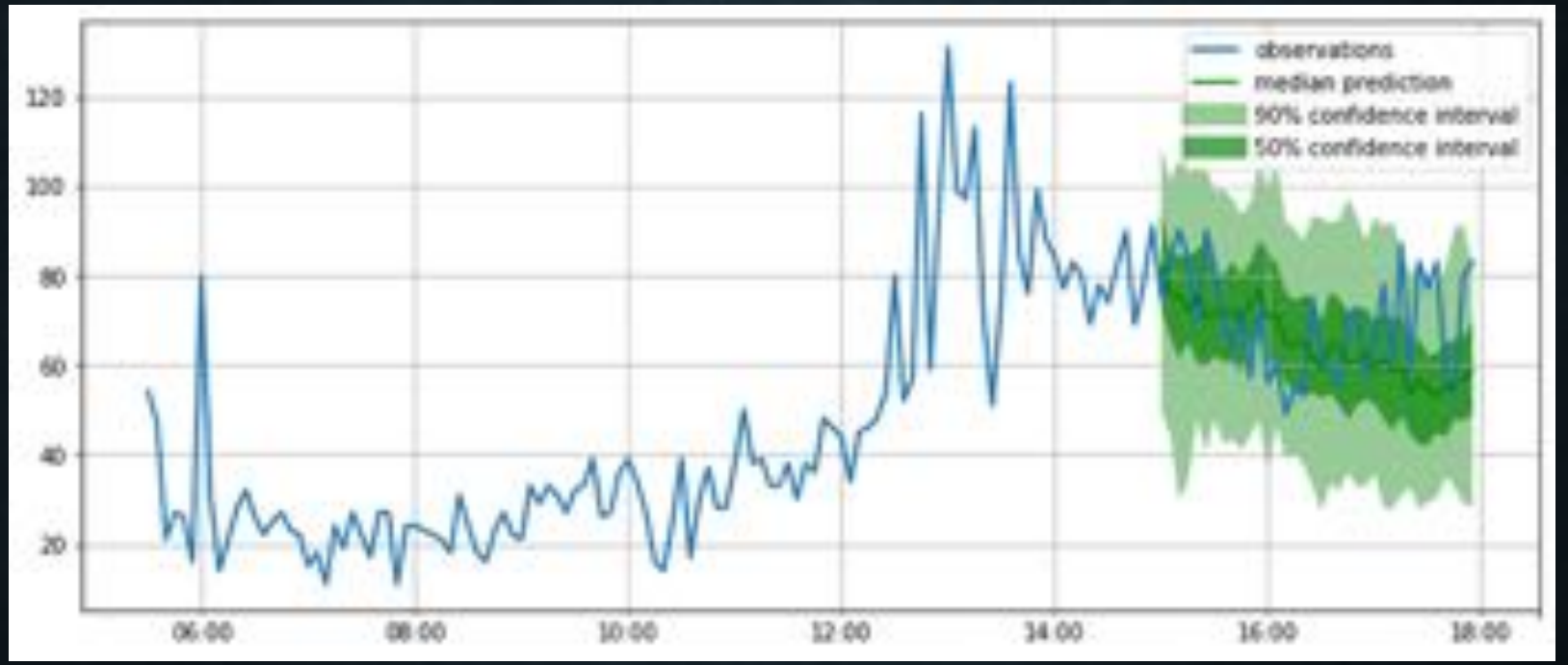
Features (as of 0.3.2)

- Pre-trained models: over 300 word-embedding
- 5 language models
- Neural Machine Translation (Google NMT, Transformer)
- Flexible data pipeline tools and many public datasets.
- NLP examples such as sentiment analysis.

aws

MXNet Toolkits - GluonTS

Quickly produce state of the art results with just a few lines of codes



aws

Go and build!

Amazon SageMaker: https://aws.amazon.com/sagemaker/

MXNet https://mxnet.apache.org/

Gluon https://gluon.mxnet.io/

GluonCV https://gluon-cv.mxnet.io/

GluonNLP https://gluon-nlp.mxnet.io/

Gluon-TS https://gluon-ts.mxnet.io/

Deep learning book http://www.d2l.ai/