mxnet* ecosystem

Naveen Swamy & Thomas Delteil

27th September 2018

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

Apache MXNet - History

Created by academia
(CMU and UW)

amazon

Amazon's deep-learning
framework of choice
since November 2016.

Accepted into
Apache Incubator
in January 2017.

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

aws

Multi-language Support

R | Perl | Julia | Clojure

Python | Scala | C++

Frontend

While keeping high performance from efficient backend

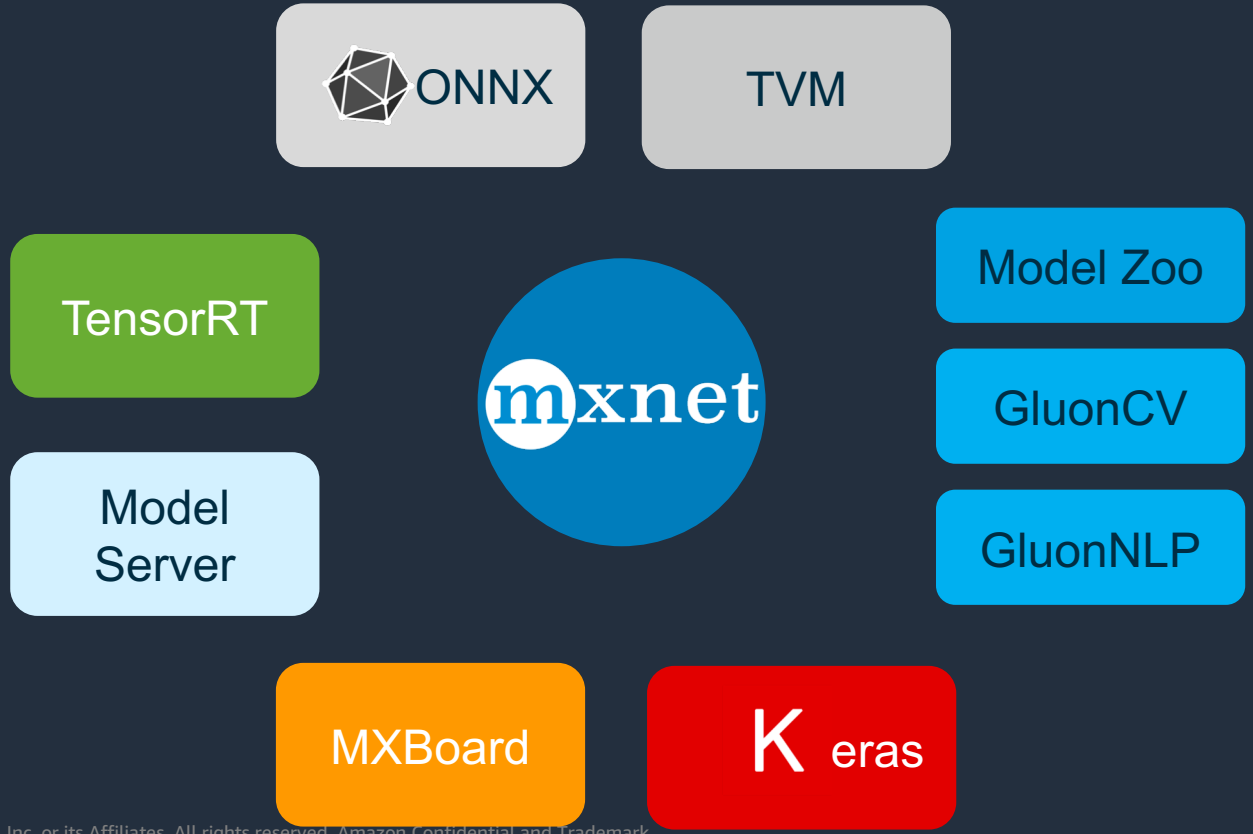
Backend

C++

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

aws

Apache MXNet Ecosystem



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

aws

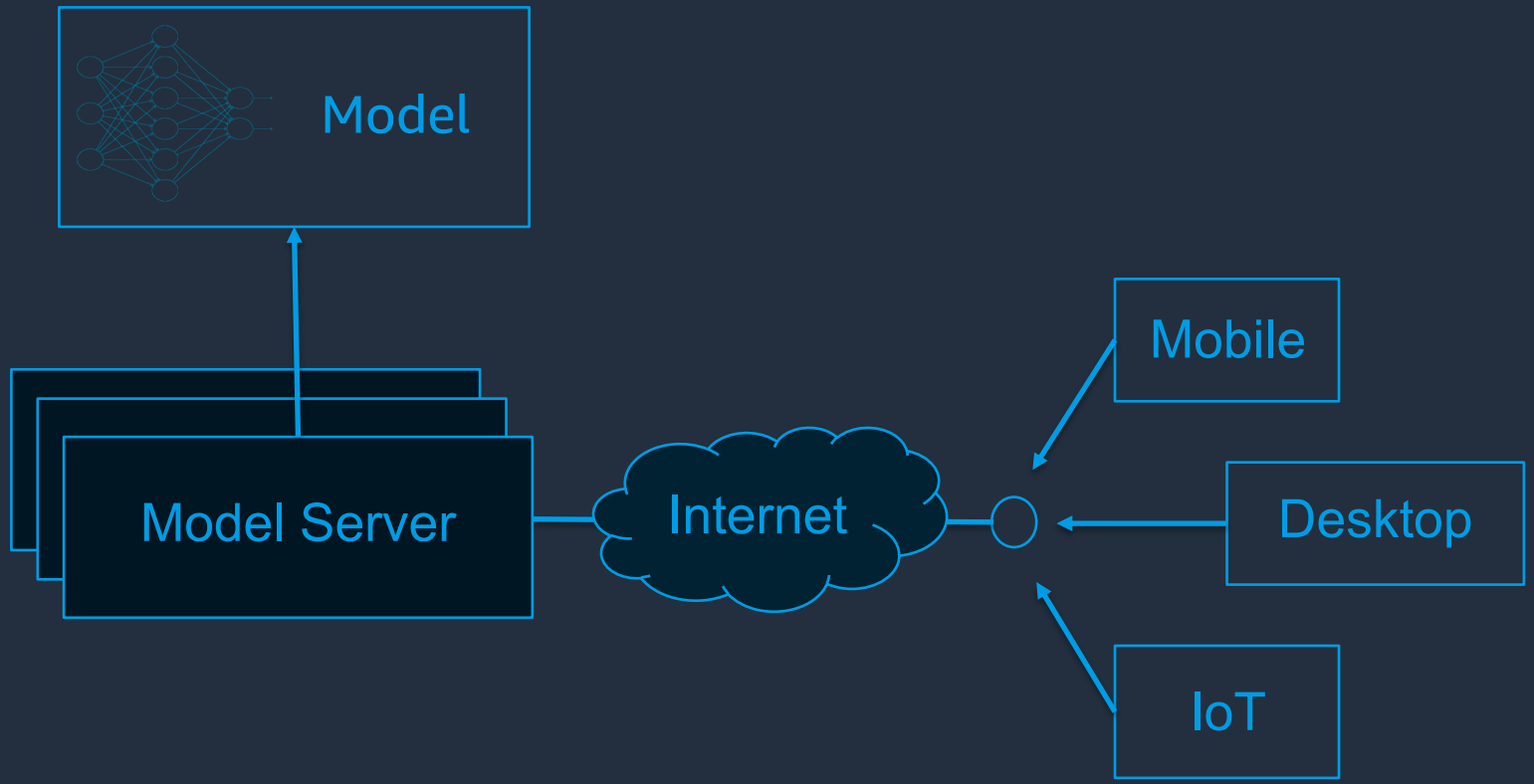
Model Server

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

aws

Credits: Hagay Lupesko

So what does a deployed model looks like?

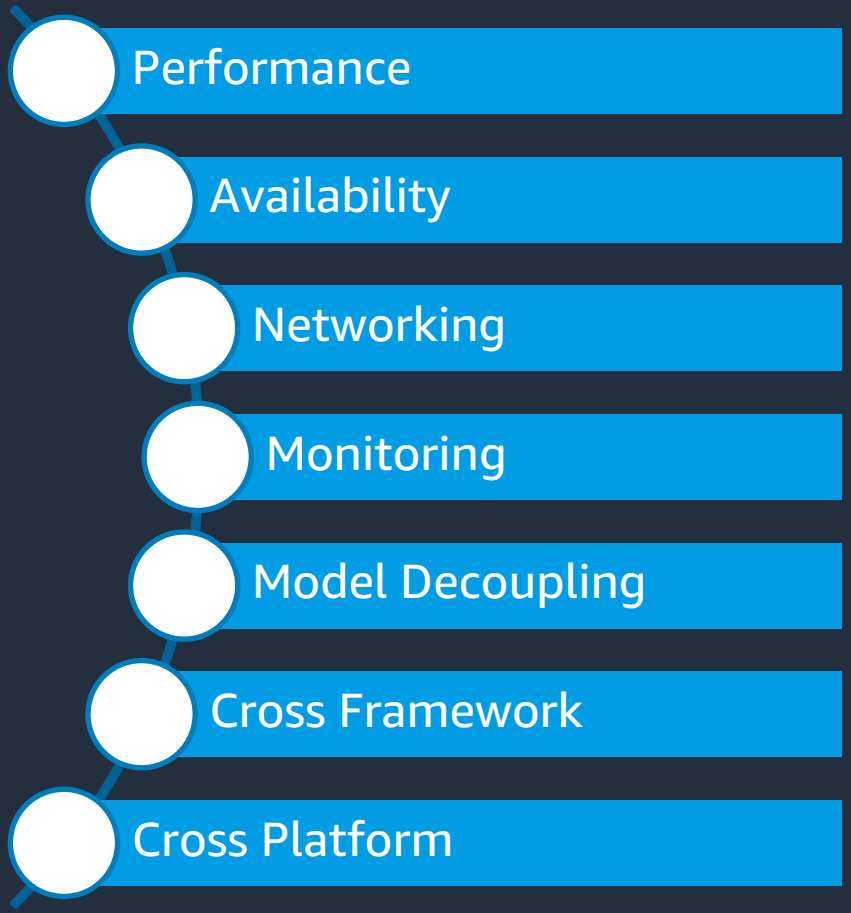


© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

aws

Credits: Hagay Lupesko

The Undifferentiated Heavy Lifting of Model Serving

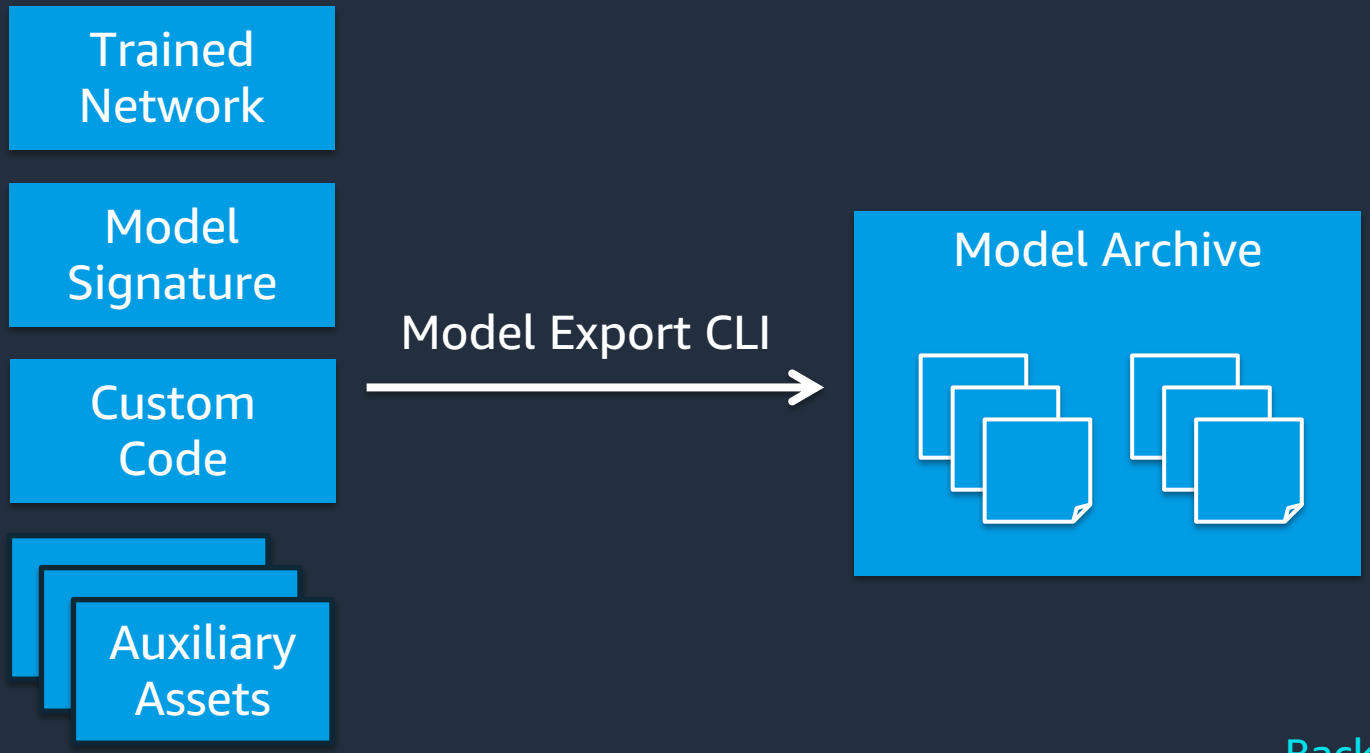


© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

aws

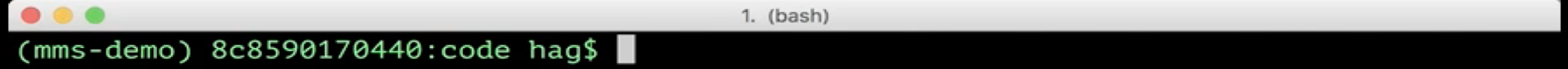
Credits: Hagay Lupesko

Model Archive



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

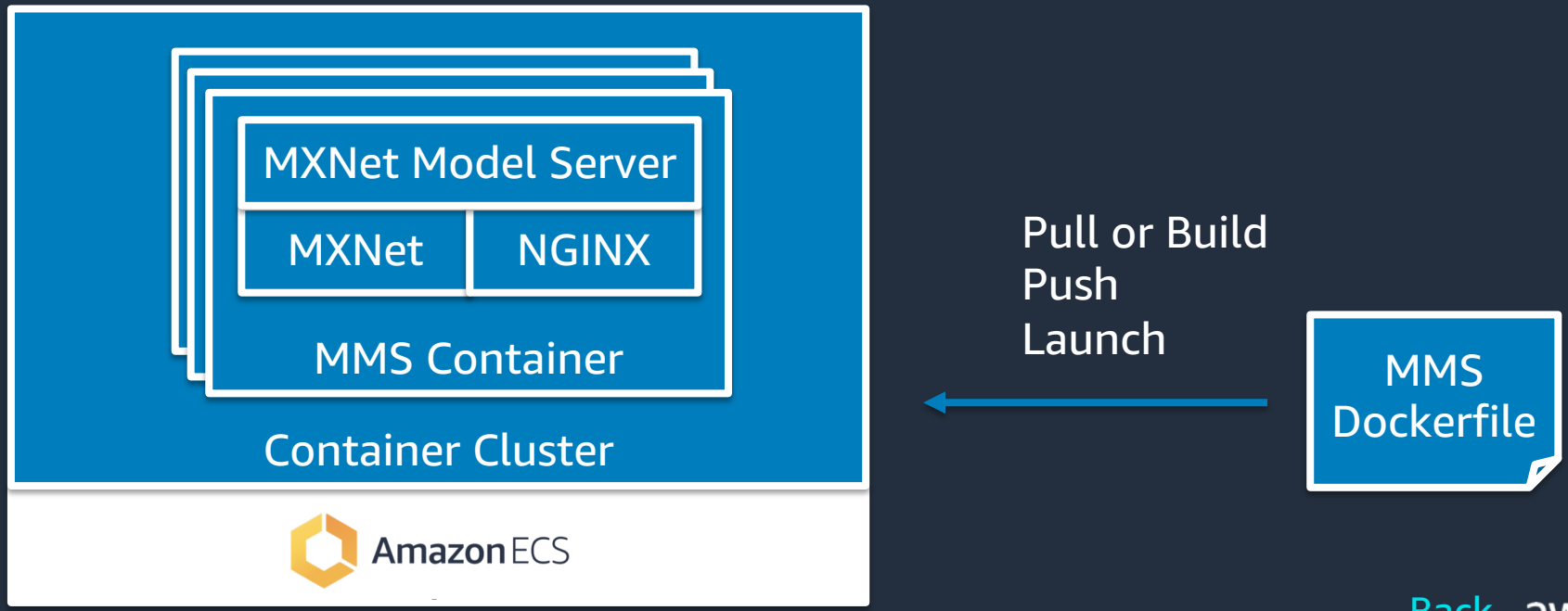
Credits: Hagay Lupesko



© 2018, Ama

Containerization

Lightweight virtualization, isolation, runs anywhere



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

Back

ONNX

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

aws

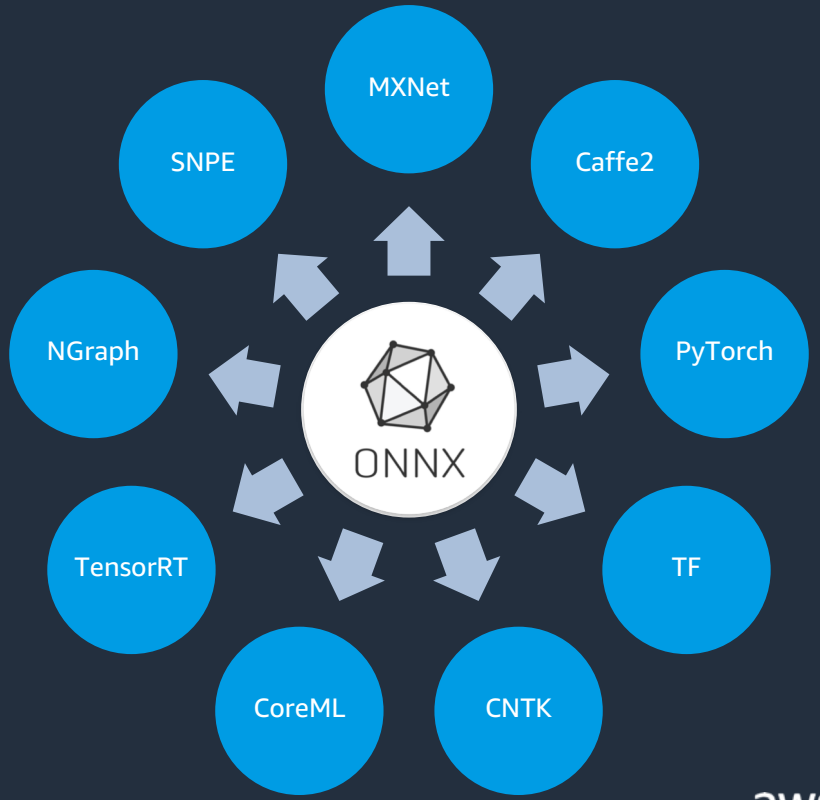
Credits: Hagay Lupesko

Open Neural Network eXchange - Overview

Many Frameworks
Many Platforms

ONNX: Common IR

- Open source
- Community driven
- Simple



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

aws

Credits: Hagay Lupesko

Import ONNX model in MXNet– Usage Example

Build and train your model with PyTorch

Load your ONNX model with MXNet

Run inference, fine tune or save as MXNet model.

```
# Import into MXNet (from MXNet 1.2)
sym, arg_params, aux_params = onnx_mxnet.import_model('model.onnx’)

# create module
mod = mx.mod.Module(symbol=sym, data_names=['input_0'], label_names=None)
mod.bind(for_training=False, data_shapes=[('input_0', input_img.shape)])
mod.set_params(arg_params=arg_params, aux_params=aux_params)
```

Credits: Hagay Lupesko

Export MXNet model to ONNX – Usage

Build and train your model in MXNet

Export trained MXNet model to ONNX format

Import in other framework like cntk, caffe2 for inference

```
# Export MXNet model to ONNX format(from MXNet 1.3)
onnx_file_path = onnx_mxnet.export_model(sym, params, [input_shape],
input_data_type, onnx_file_path)
```

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

Back

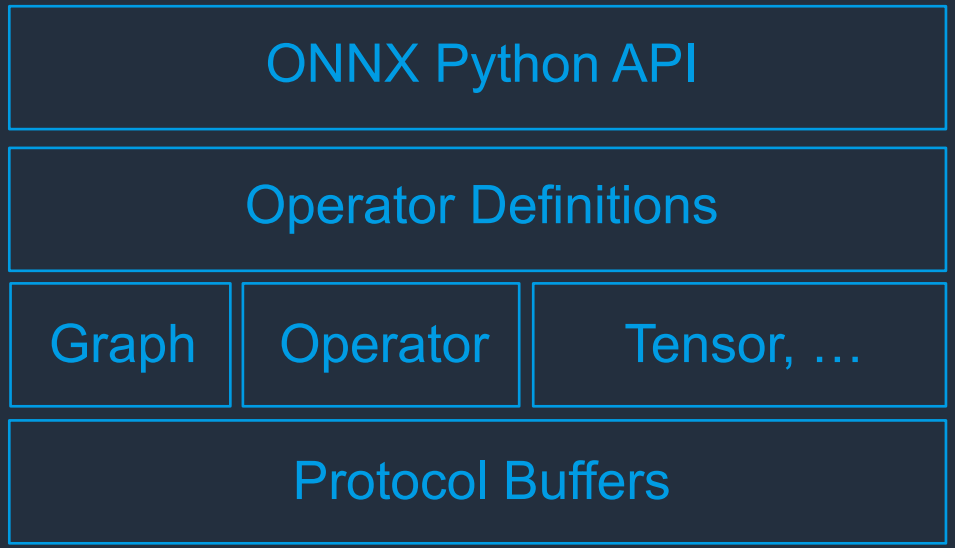
aws

Credits: Hagay Lupesko

ONNX – Internals

Protocol Buffers:

- Binary compact format
- Statically defined
- APIs for de/serialization
- Cross platform



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

aws

Credits: Hagay Lupesko

ONNX – Coverage

Framework	Export	Import
MXNet	Supported	Supported
Caffe2	Supported	Supported
PyTorch	Supported	Coming…
CNTK	Supported	Supported
Chainer	Supported (external)	N/A
TensorFlow	Supported (external)	Supported (external)
CoreML	Supported (external)	Supported (external)
SciKit-Learn	Supported (external)	N/A

Credits: Hagay Lupesko

Keras-MXNet

https://github.com/awslabs/keras-apache-mxnet

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

aws

Keras – Apache MXNet

- Deep Learning for Humans
- 2nd most popular Deep Learning framework
- Keras users leverage MXNet's great performance

```
from keras.models import Sequential
model = Sequential()
from keras.layers import Dense
model.add(Dense(units=64, activation='relu', input_dim=100))
model.add(Dense(units=10, activation='softmax'))
model.compile(loss='categorical_crossentropy',
              optimizer='sgd',
              metrics=['accuracy'])
model.fit(x_train, y_train, epochs=5, batch_size=32)
model.train_on_batch(x_batch, y_batch)
loss_and_metrics = model.evaluate(x_test, y_test, batch_size=128)
classes = model.predict(x_test, batch_size=128)
```

```
pip install mxnet-(mkl|cu92)
pip install keras-mxnet
---
~/.keras/keras.json
backend: mxnet
image_data_format: channels_first
---
```

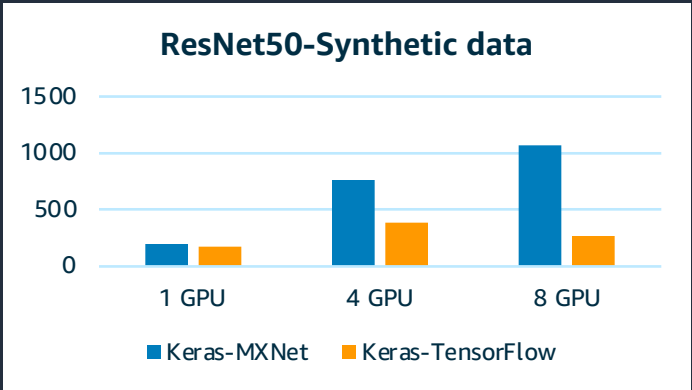
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

aws

Keras Benchmarks

Setup: https://github.com/awslabs/keras-apache-mxnet/tree/master/benchmark

	Training	Inference
Instance	P3.8x Large, P3.16x Large	C5.xLarge, C4.8xLarge
Network	ResNet50v1	ResNet50v1
Batch size	32 * Num of GPUs	32
Image size	3*256*256	3*256*256

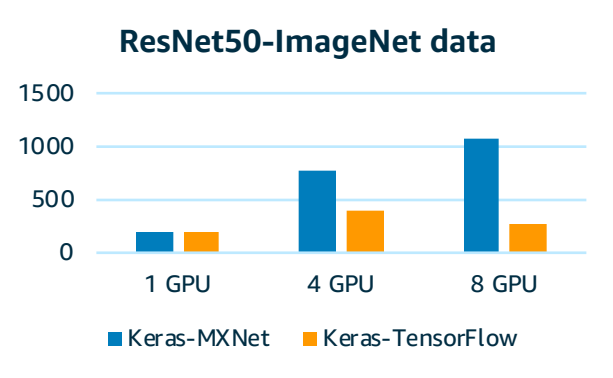


GPUs	Keras-MXNet [Image/sec]	Keras-TensorFlow [Image/sec]	Speed Up
1	194	184	1.05
4	764	393	1.94
8	1068	261	4.09

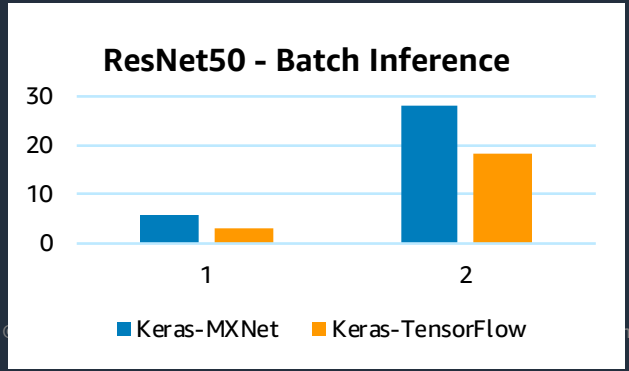
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

aws

Keras Benchmarks



GPUs	Keras-MXNet	Keras-TensorFlow	Speed Up
1	135	52	2.59
4	536	162	3.30
8	722	211	3.42



Instance	Keras-MXNet	Keras-TensorFlow	Speed Up
C5.X Large	5.79	3.27	1.782
C5.8X Large	27.9	18.2	1.53

azon Confidential and Trademark

Imperative API

Debuggable

Flexible

Scalable

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

aws

Symbolic vs Imperative

Symbolic is “define, compile, run”

```
model = Sequential()
model.add(Conv2D(32, kernel_size=(3, 3),
                 activation='relu',
                 input_shape=input_shape))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(num_classes, activation='softmax'))
```

```
model.compile(loss=keras.losses.categorical_crossentropy,
              optimizer=keras.optimizers.Adadelta(),
              metrics=['accuracy'])

model.fit(x_train, y_train,
          batch_size=batch_size,
          epochs=epochs,
          verbose=1,
          validation_data=(x_test, y_test))
```

e.g. Inception Stage

Imperative is “define-by-run”

```
net = nn.Sequential()
with net.name_scope():
    net.add(
        nn.Conv2D(channels=6, kernel_size=5, activation='relu'),
        nn.MaxPool2D(pool_size=2, strides=2),
        nn.Conv2D(channels=16, kernel_size=3, activation='relu'),
        nn.MaxPool2D(pool_size=2, strides=2),
        nn.Flatten(),
        nn.Dense(120, activation="relu"),
        nn.Dense(84, activation="relu"),
        nn.Dense(10)
    )
net.initialize(init=init.Xavier())
```

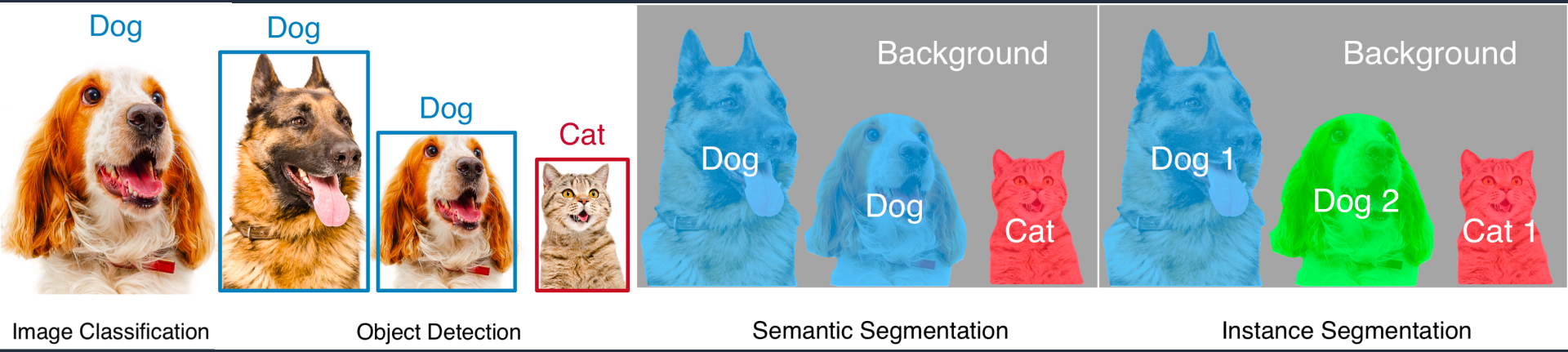
```
for epoch in range(10):
    for data, label in train_data:
        with autograd.record():
            output = net(data)
            loss = softmax_cross_entropy(output, label)
        loss.backward()
        trainer.step(batch_size)
```

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

aws

GluonCV: a Deep Learning Toolkit for Computer Vision

https://gluon-cv.mxnet.io

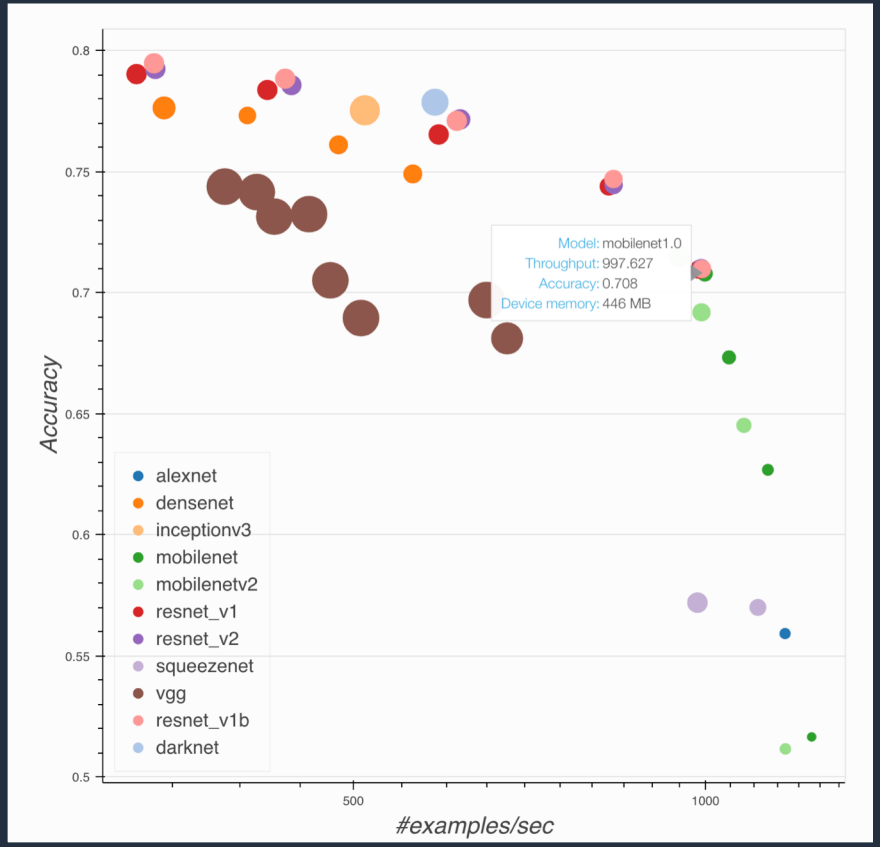


50+ Pre-trained models, with training scripts, datasets, tutorials

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

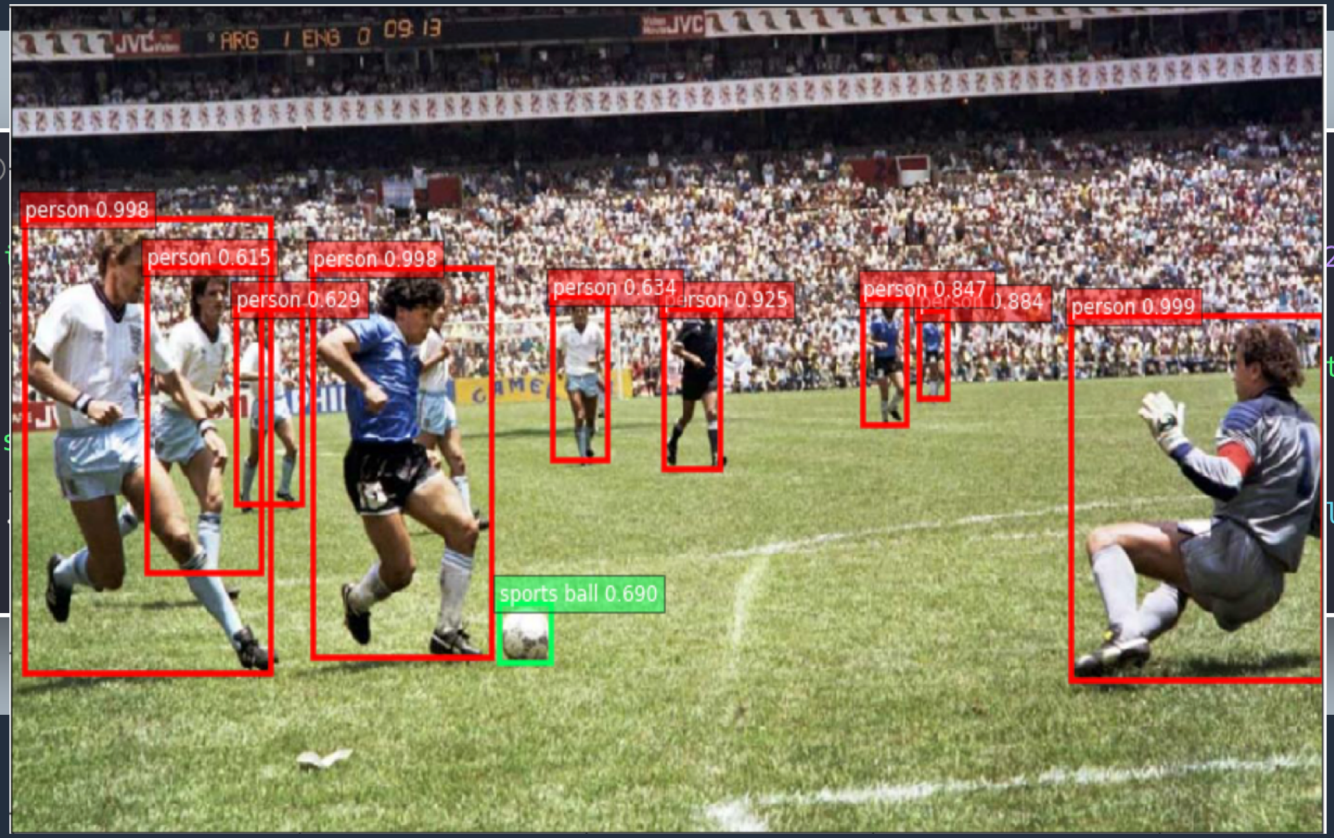
aws

GluonCV: pre-trained models, help to choose



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

GluonCV: example code



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

GluonNLP: a Deep Learning Toolkit for Natural Language Processing

https://gluon-nlp.mxnet.io

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.

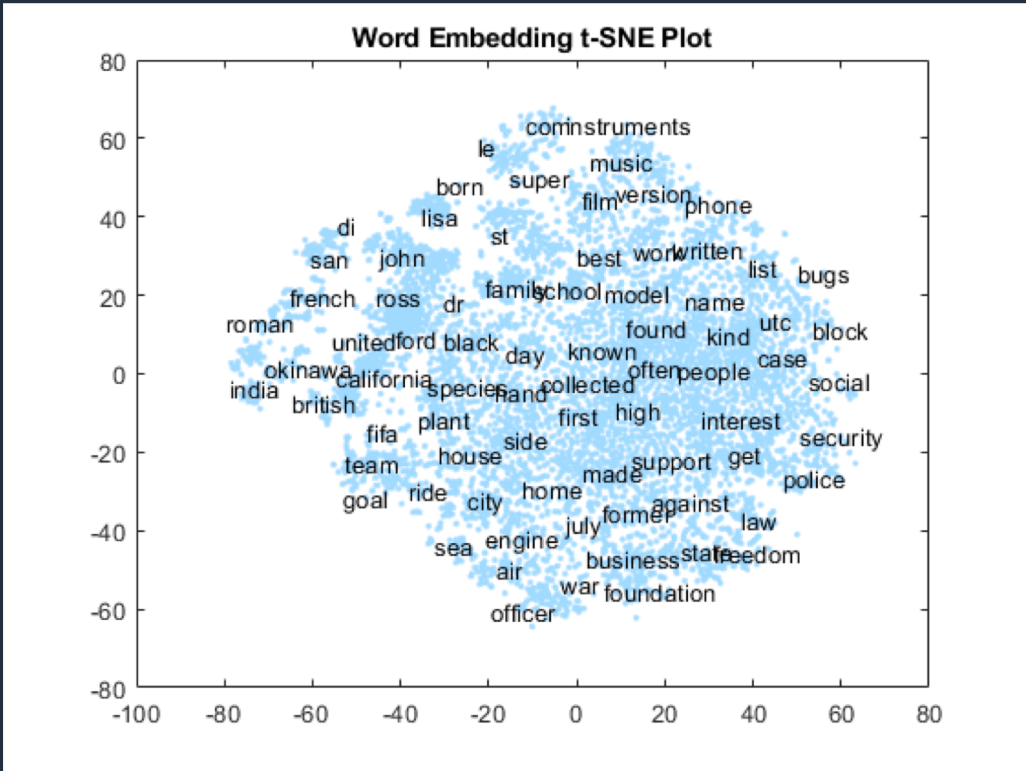
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

aws

Word embedding:

	car
dog	0
apple	0
eat	0
car	1
van	0
leaf	0
the	0
at	0

car	van
0.1	0.2
0.5	0.3
-1.1	-1.0
0.4	0.5



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

Language modeling

Trained to predict the next word ($P(w_n|w_i \dots w_{n-1})$):

- The winner of the 2018 FIFA world cup is ?

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

```
lm_model, vocab = nlp.model.get_model(name='standard_lstm_lm_200', dataset_name='wikitext-2',
pretrained=True)
data = vocab[tokenizer("This movie is considered the")]

pred, _ = lm_model(mx.nd.array(data, dtype='float32').expand_dims(axis=1))

for p in pred[-1].squeeze().topk(k=20).asnumpy():
    print(vocab.idx_to_token[int(p)])
```

This movie is considered the <?>

first	greatest	highest
most	main	final
<unk>	second	worst
only	last	sixth
same	name	third
best	largest	way
		primary

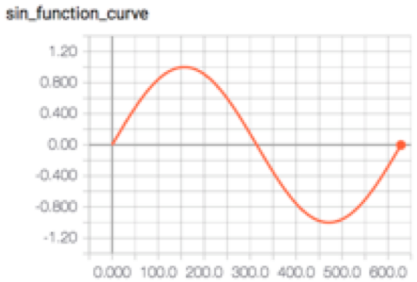
© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

aws

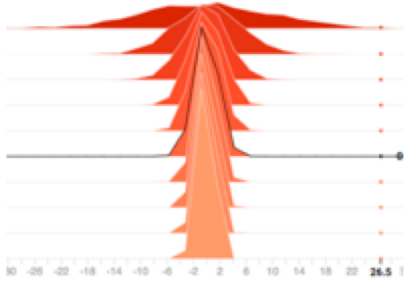
MXBoard: MXNet plugin to TensorBoard

Graph

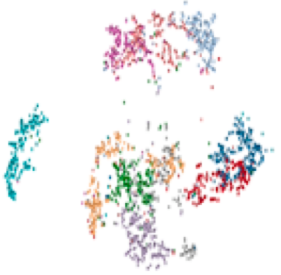
Scalar



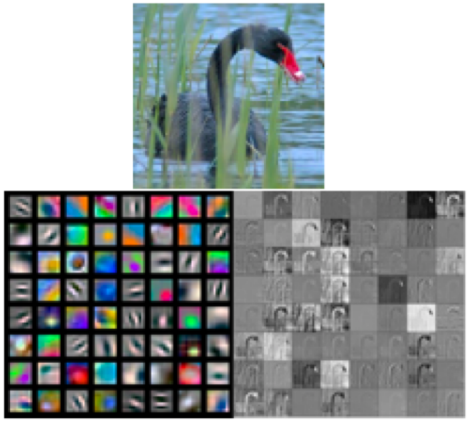
Histogram



Embedding

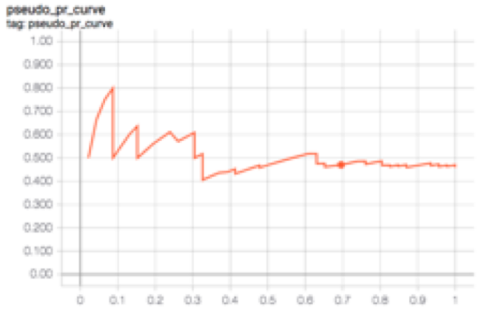


Image

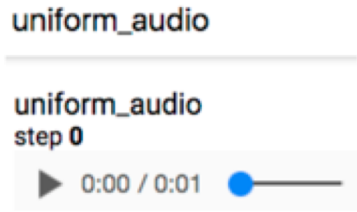


Text

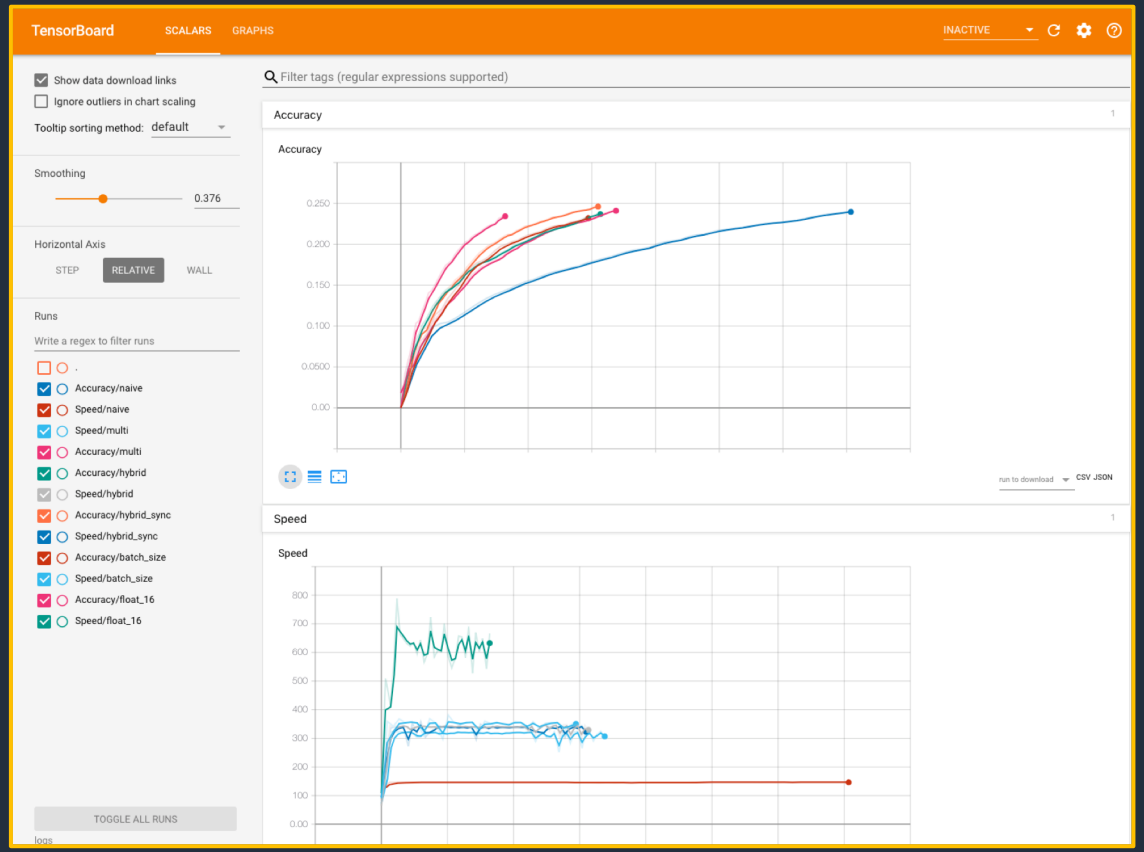
Precision-Recall Curve



Audio



MXBoard



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

Deep Learning acceleration

CUDA & CuDNN

```
pip install mxnet-cu92
```

TensorRT

```
pip install mxnet-tensorrt-cu92
```

MKL, MKLML & MKLDNN

```
e.g. pip install mxnet-mkl
```

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

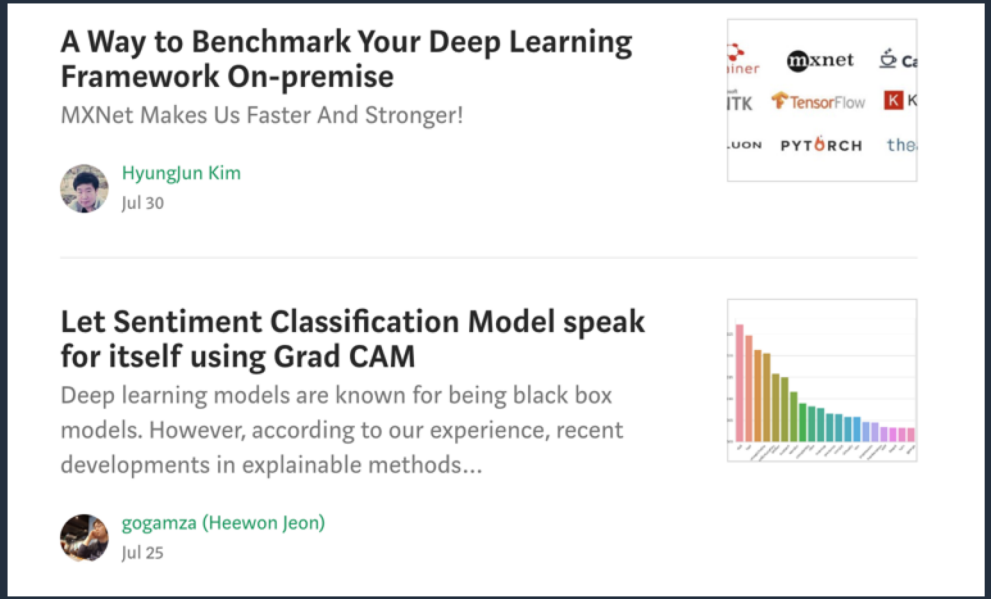
aws

Apache MXNet community

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

Keeping Up to Date

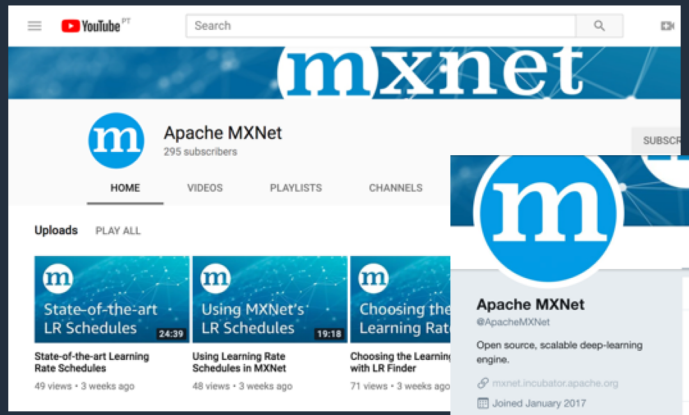
Medium: https://medium.com/apache-mxnet



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

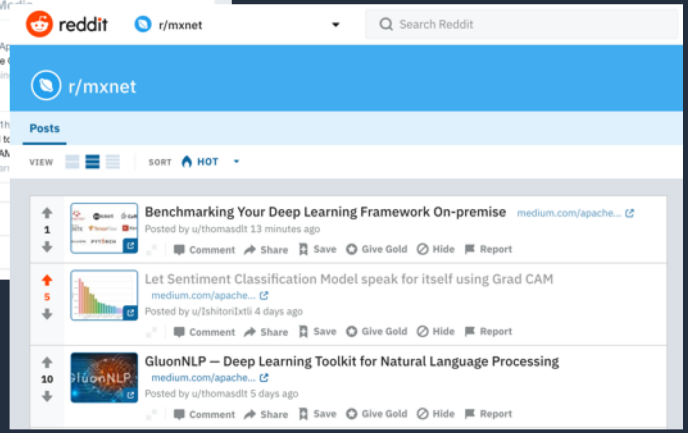
Keeping Up to Date: Social

YouTube: /apachemxnet



Twitter: @apachemxnet

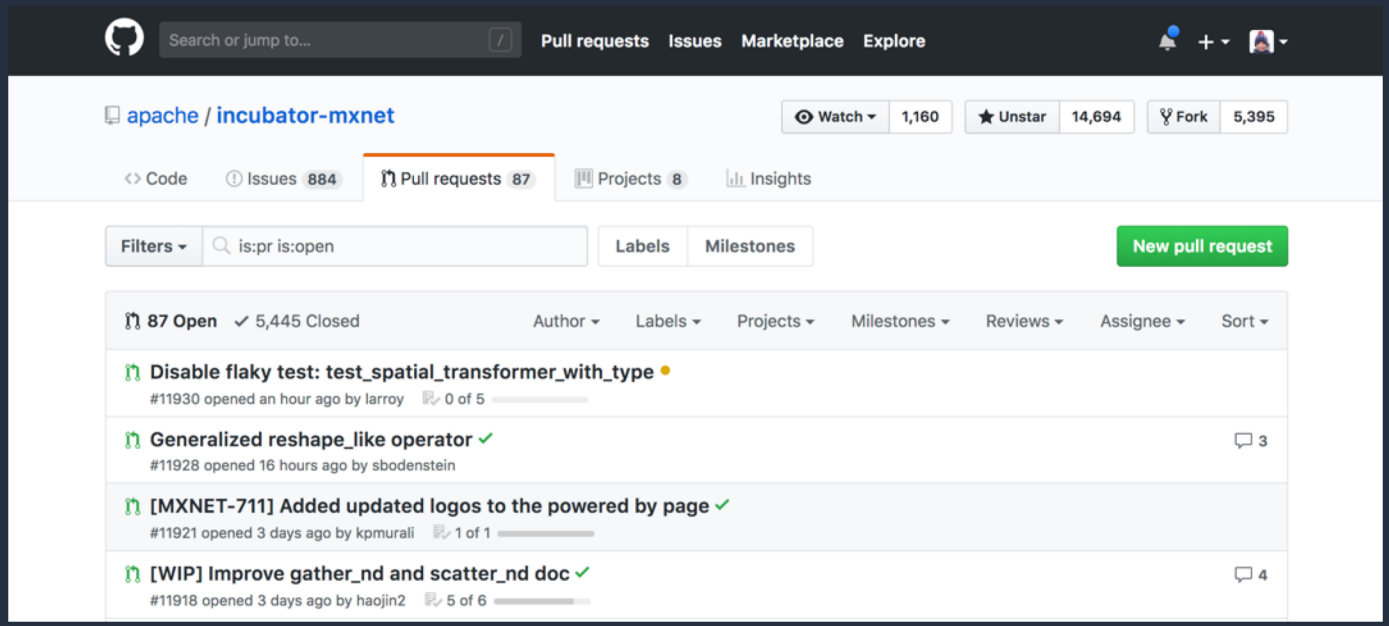
Reddit: r/mxnet



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

Community

GitHub: https://github.com/apache/incubator-mxnet

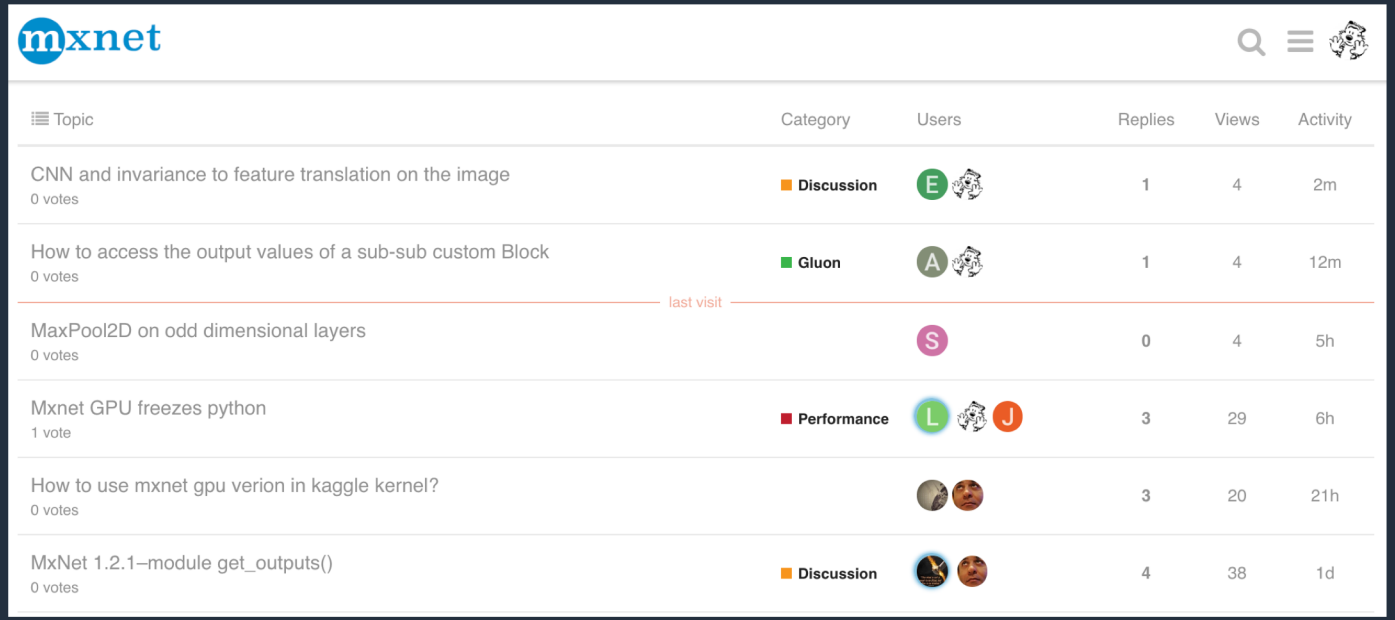


© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

aws

Community

Discuss Forum: https://discuss.mxnet.io/



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

Community

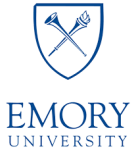
Mailing list:

dev@mxnet.apache.org

user@mxnet.apache.org

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

MXNet Customer Momentum



Thank you!

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