



mxnet* ecosystem

Naveen Swamy & Thomas Delteil

27th September 2018



Apache MXNet - History



Created by academia
(CMU and UW)



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



Accepted into
Apache Incubator
in January 2017.

Multi-language Support

R

Perl

Julia

Clojure

Python

Scala

C++

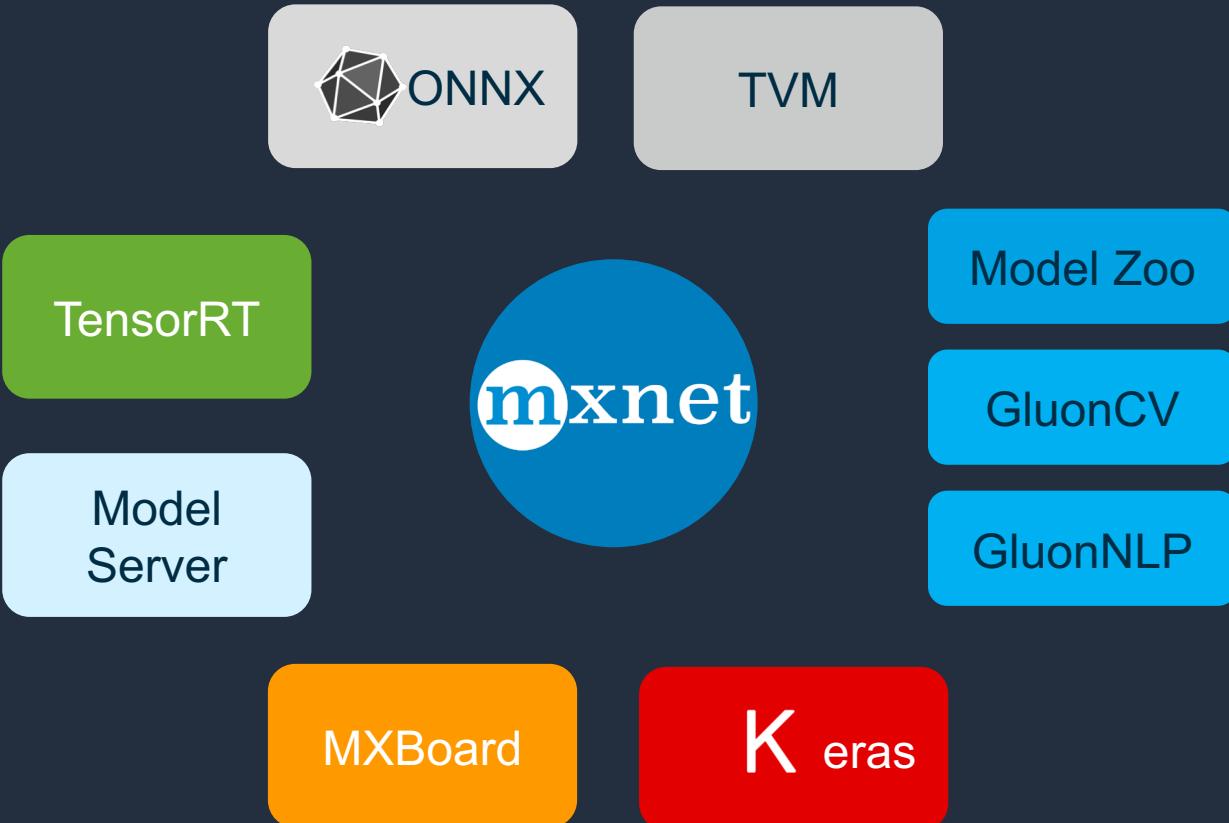
Frontend

While keeping high performance from efficient backend

Backend

C++

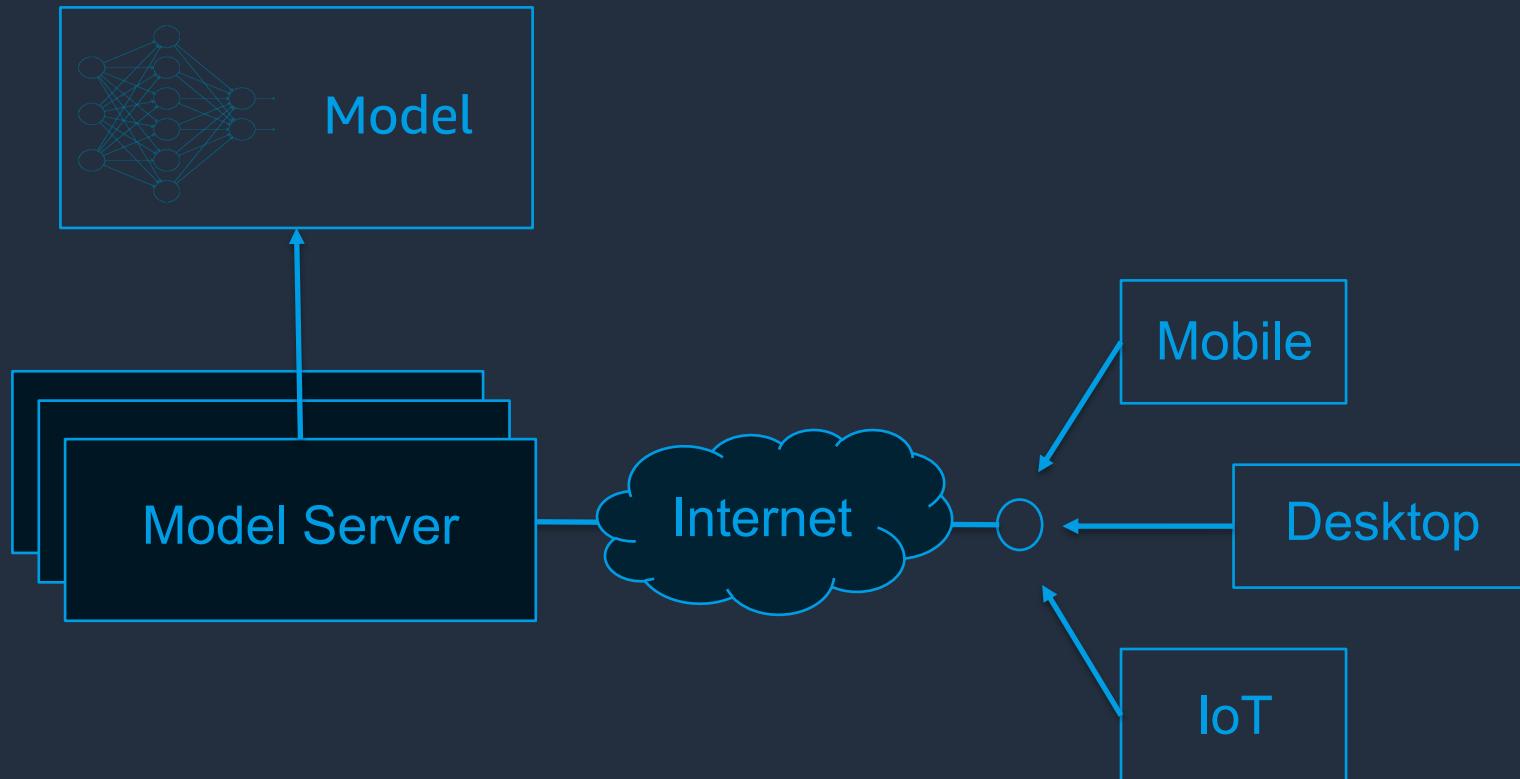
Apache MXNet Ecosystem



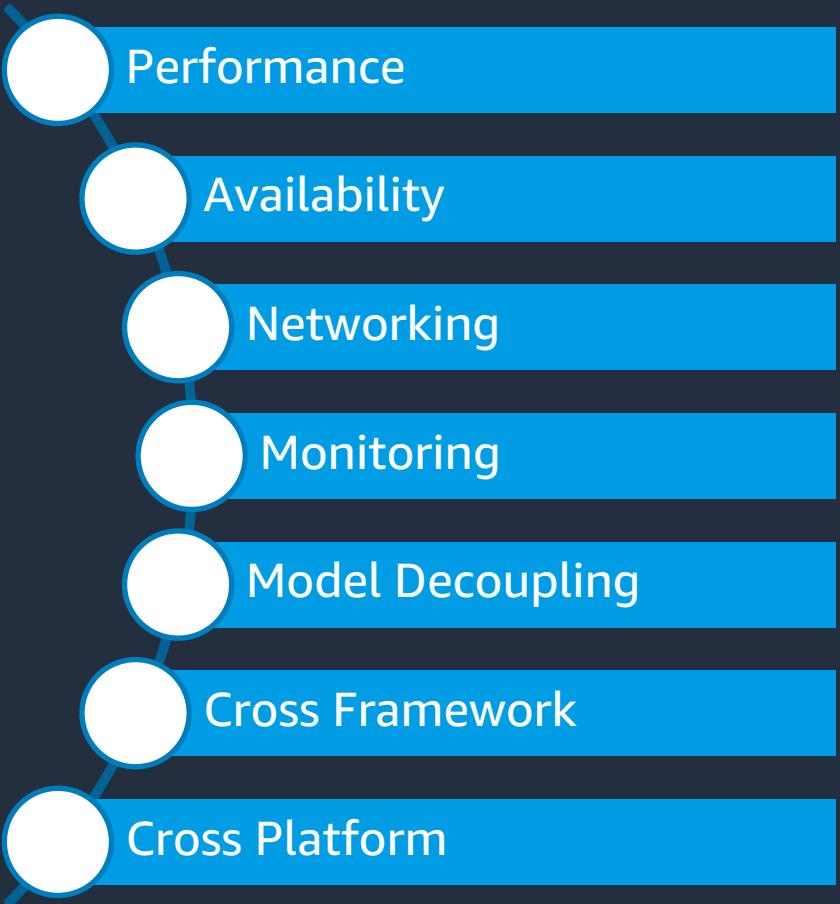


Model Server

So what does a deployed model looks like?

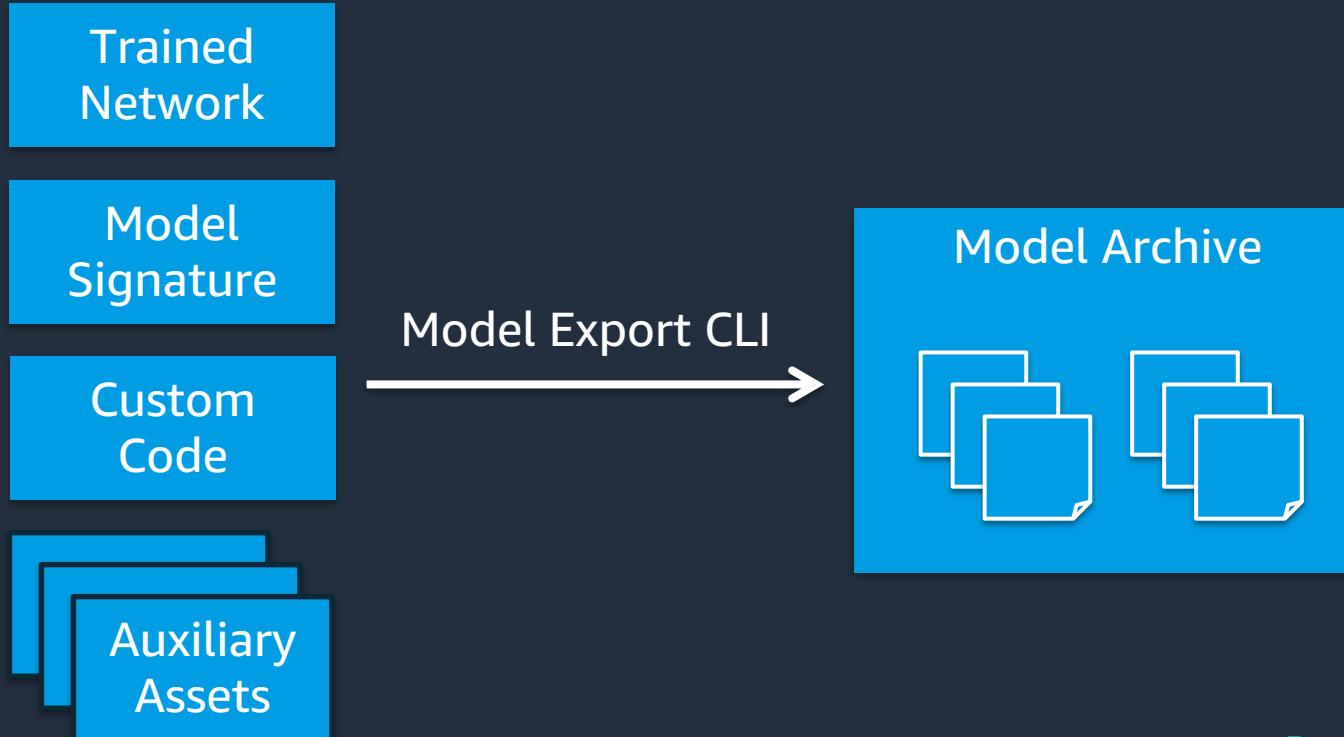


The Undifferentiated Heavy Lifting of Model Serving





Model Archive





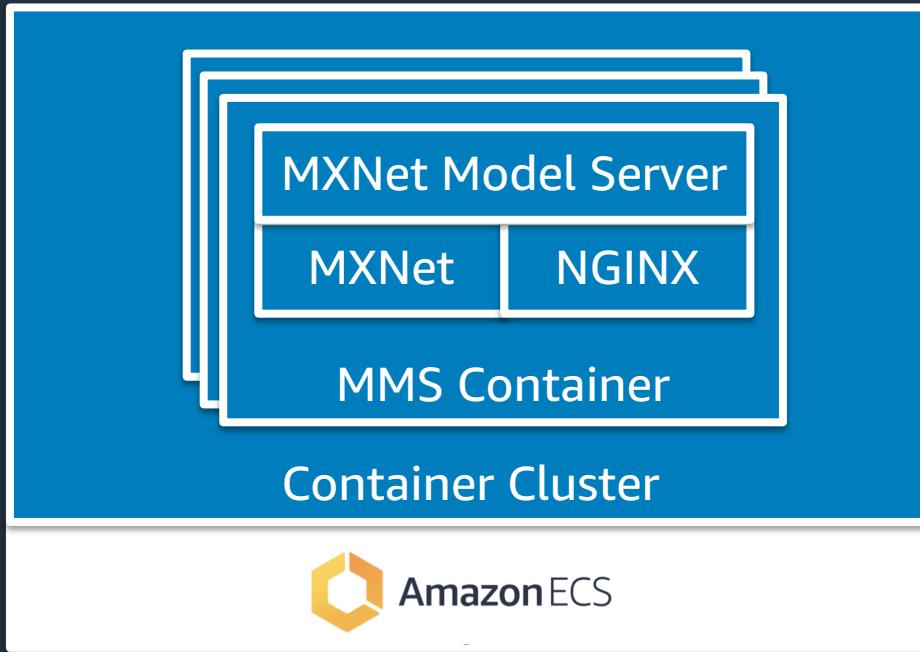
1. (bash)

(mms-demo) 8c8590170440:code hag\$ █

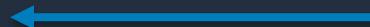


Containerization

Lightweight virtualization, isolation, runs anywhere



Pull or Build
Push
Launch





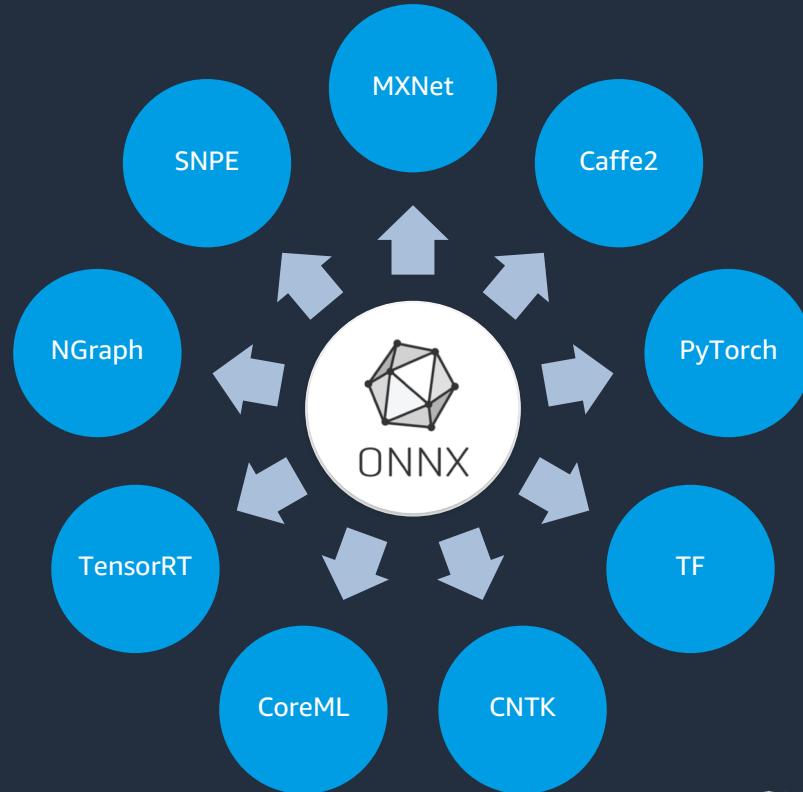
ONNX

Open Neural Network eXchange - Overview

Many Frameworks
Many Platforms

ONNX: Common IR

- Open source
- Community driven
- Simple



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)
```

Export MXNet model to ONNX – Usage

Build and train your model in MXNet

Export trained MXNet model to ONNX format

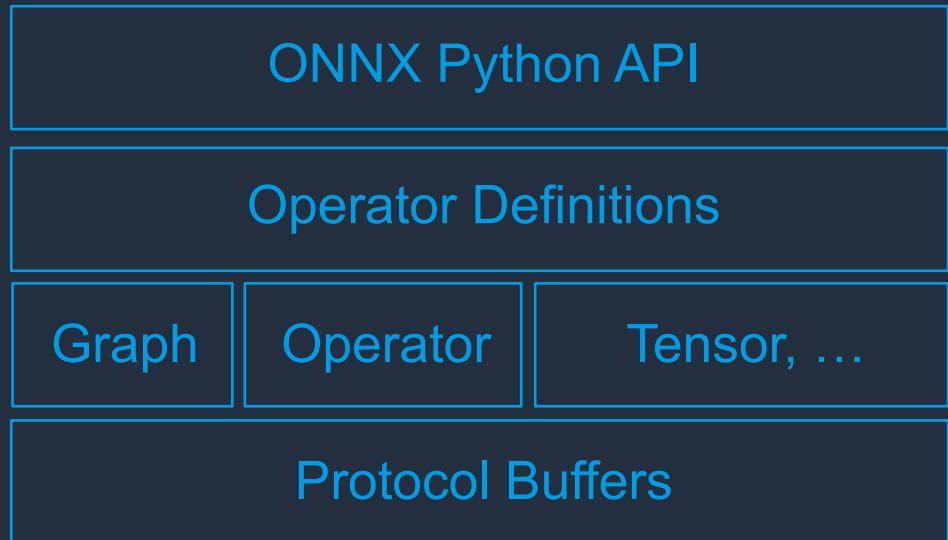
Import in other framework like cntk, caffe2 for inference

```
# 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)
```

ONNX – Internals

Protocol Buffers:

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



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 Luplesko



Keras-MXNet

<https://github.com/awslabs/keras-apache-mxnet>)

Keras – Apache MXNet

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

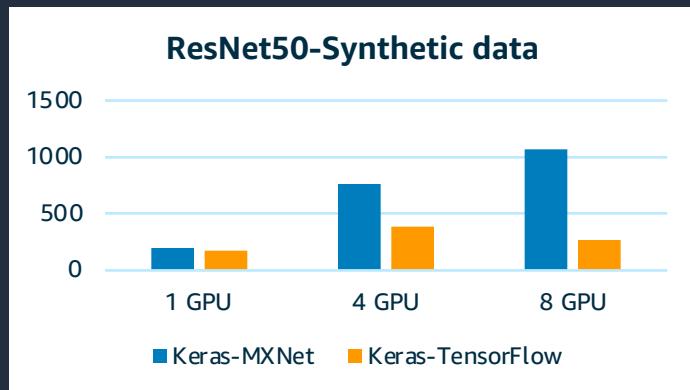
```
from keras.models import Sequential
model = Sequential()
from keras.layers import Dense
model.add(Dense(units=64, activation='relu', input_dim=100))
model.add(Dense(units=10, activation='softmax'))
model.compile(loss='categorical_crossentropy',
              optimizer='sgd',
              metrics=['accuracy'])
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
---
```

Keras Benchmarks

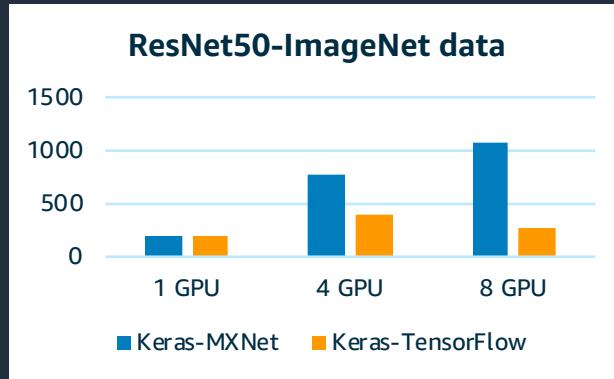
Setup: <https://github.com/awslabs/keras-apache-mxnet/tree/master/benchmark>

	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

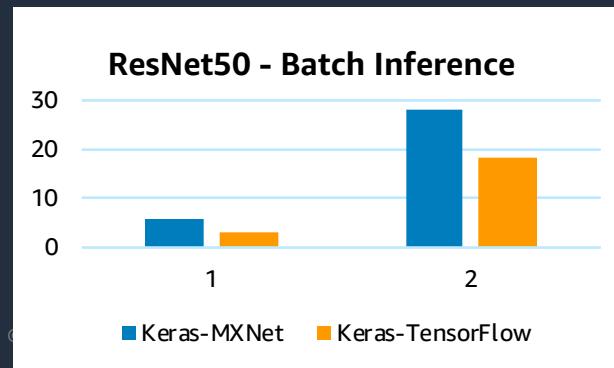


GPUs	Keras-MXNet [Image/sec]	Keras-TensorFlow [Image/sec]	Speed Up
1	194	184	1.05
4	764	393	1.94
8	1068	261	4.09

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

Imperative API



Debuggable



Flexible



Scalable

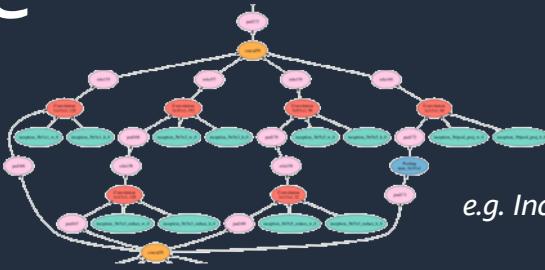
Symbolic vs Imperative

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

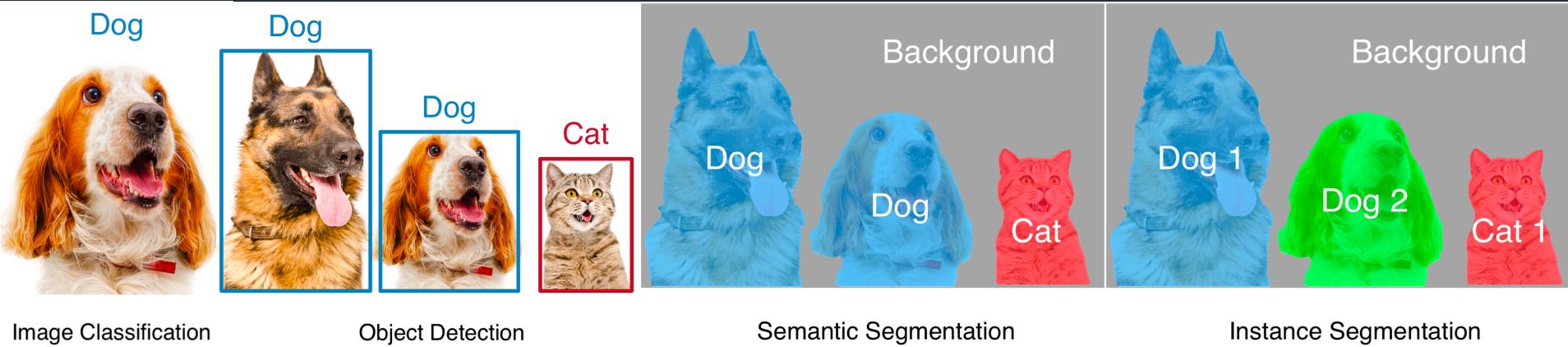
Imperative is “define-by-run”

```
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)
```

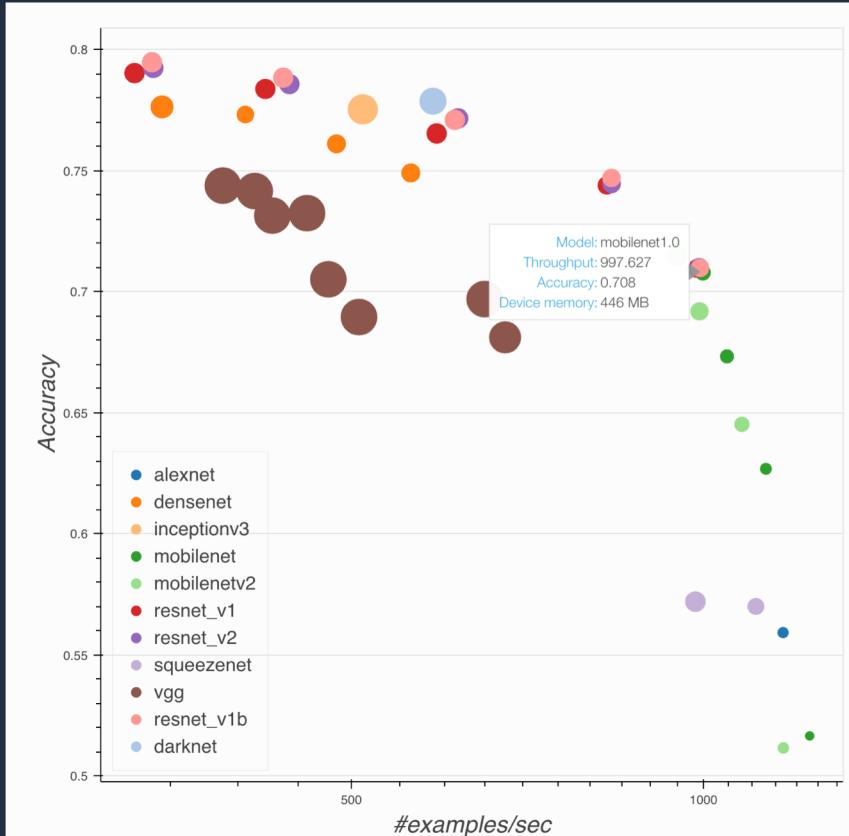
GluonCV: a Deep Learning Toolkit for Computer Vision

<https://gluon-cv.mxnet.io>

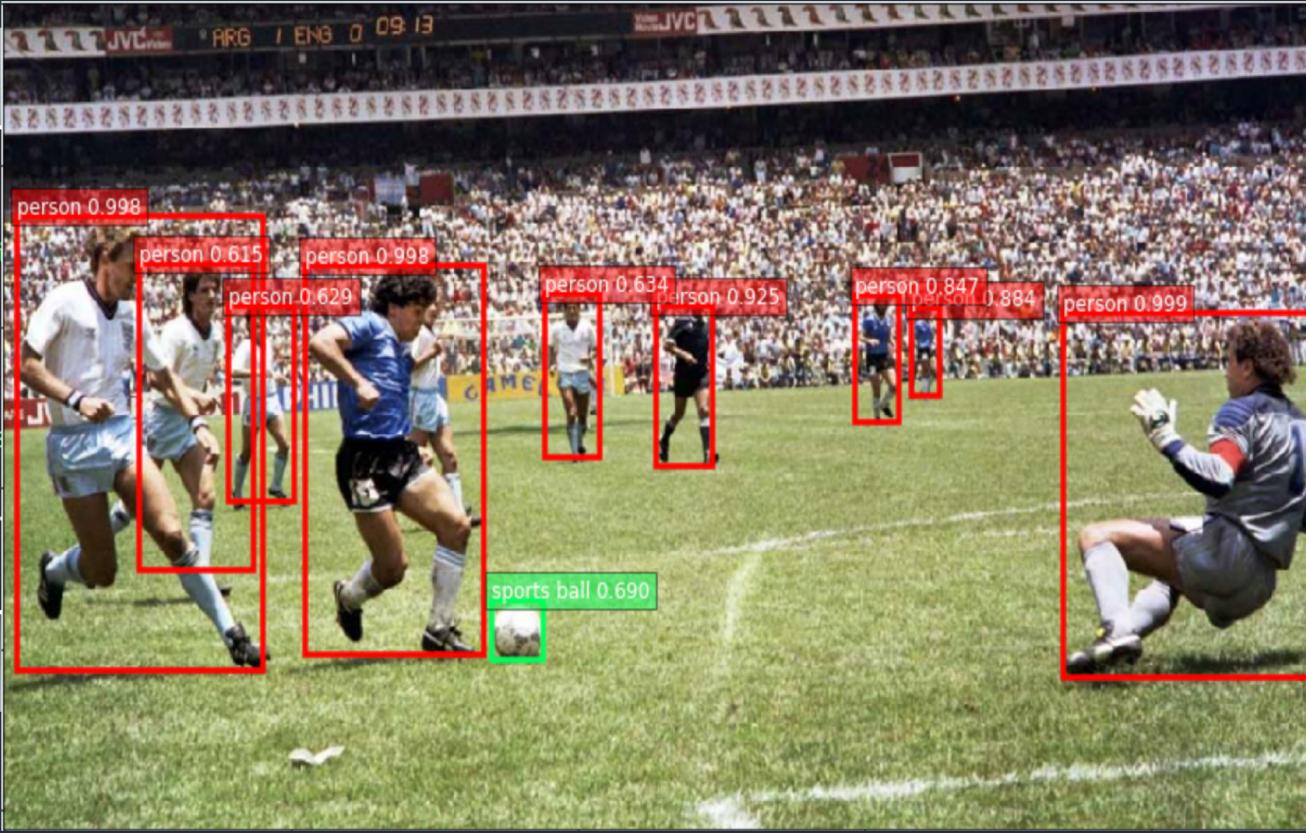


50+ Pre-trained models, with training scripts, datasets, tutorials

GluonCV: pre-trained models, help to choose



GluonCV: example code



GluonNLP: a Deep Learning Toolkit for Natural Language Processing

<https://gluon-nlp.mxnet.io>

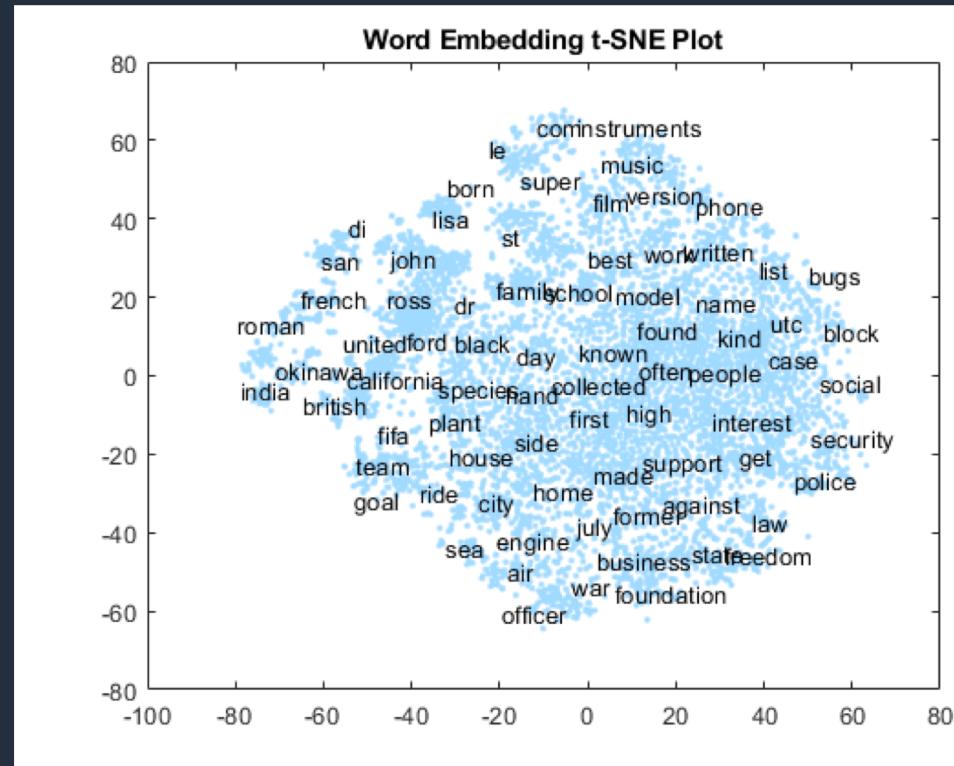
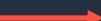


Features (as of 0.3.2)

- 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.

Word embedding:

	car	car	van
dog	0	0.1	0.2
apple	0	0.5	0.3
eat	0	-1.1	-1.0
car	1	0.4	0.5
van	0		
leaf	0		
the	0		
at	0		



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  ?

○ ○ ○

```
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))

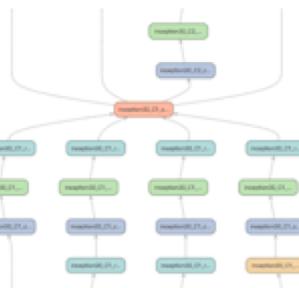
for p in pred[-1].squeeze().topk(k=20).asnumpy():
    print(vocab.idx_to_token[int(p)])
```

This movie is considered the <?>

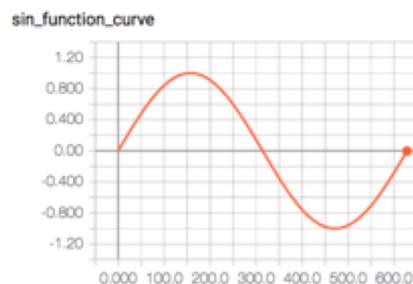
first	greatest	highest
most	main	final
<unk>	second	worst
only	last	sixth
same	name	third
best	largest	way
		primary

MXBoard: MXNet plugin to TensorBoard

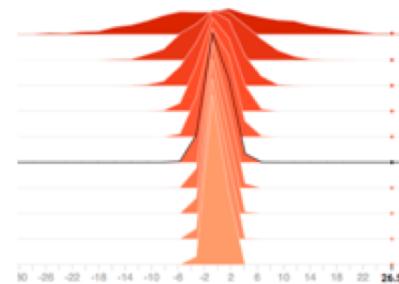
Graph



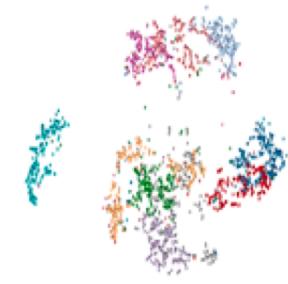
Scalar



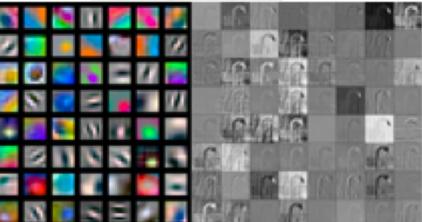
Histogram



Embedding



Image



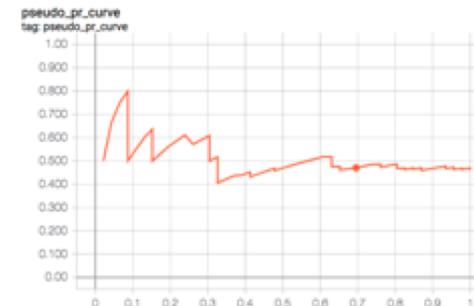
Text

markdown_table
tag: markdown_table

step 0

Hello	MXNet,
This	is
so	awesome!

Precision-Recall Curve



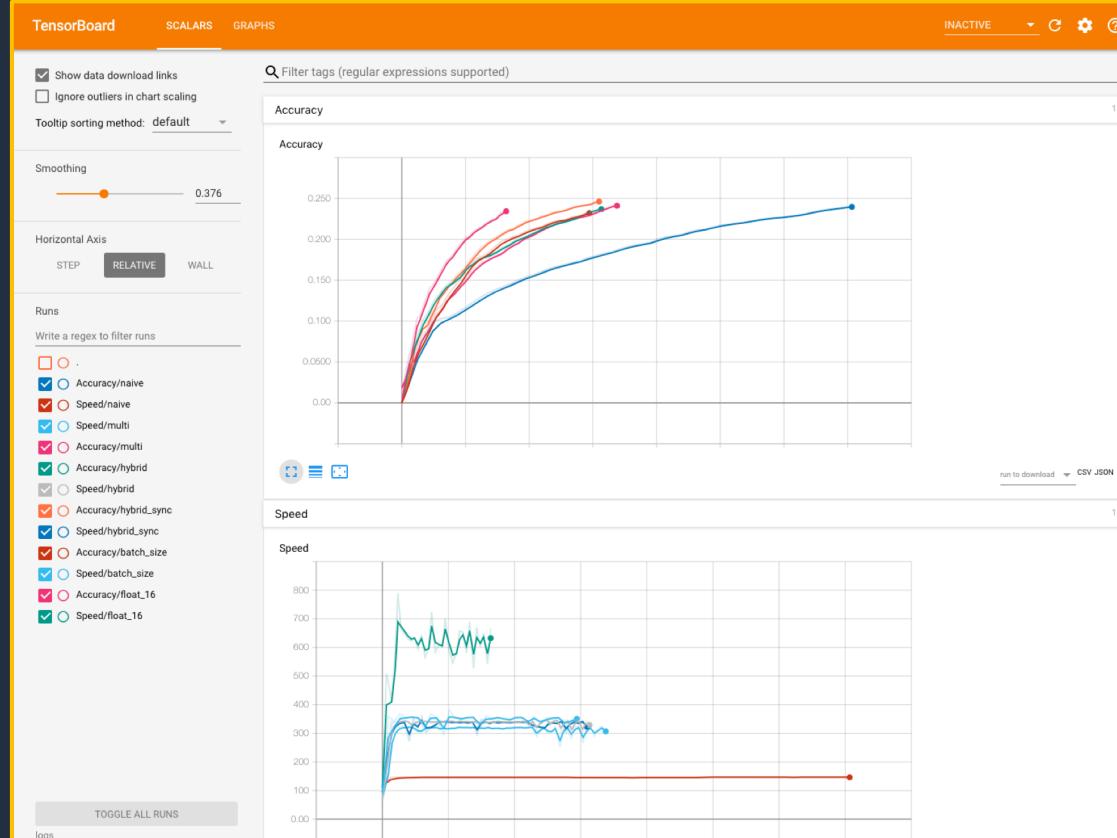
Audio

uniform_audio

uniform_audio
step 0

▶ 0:00 / 0:01

MXBoard



Deep Learning acceleration



CUDA & CuDNN

```
pip install mxnet-cu92
```

TensorRT

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

MKL, MKLML & MKLDNN

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



Apache MXNet community

Keeping Up to Date

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



A Way to Benchmark Your Deep Learning Framework On-premise

MXNet Makes Us Faster And Stronger!



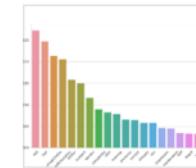
HyungJun Kim

Jul 30



Let Sentiment Classification Model speak for itself using Grad CAM

Deep learning models are known for being black box models. However, according to our experience, recent developments in explainable methods...



gogamza (Heewon Jeon)

Jul 25

Keeping Up to Date: Social

YouTube: /apachemxnet



The screenshot shows the Apache MXNet YouTube channel interface. It features a large blue header with the 'mxnet' logo. Below the header, there's a navigation bar with 'HOME', 'VIDEOS', and 'CHANNELS'. Under 'VIDEOS', there are three thumbnail cards: 'State-of-the-art LR Schedules' (24:39), 'Using Learning Rate Schedules in MXNet' (19:18), and 'Choosing the Learning Rate with LR Finder' (71 views). Each thumbnail includes a small 'm' icon and a timestamp.

Twitter: @apachemxnet



The screenshot shows the Apache MXNet Twitter profile. It displays basic user statistics: Tweets 227, Following 129, Followers 1,362, Likes 172, Lists 1, and Moments 0. The bio is: Open source, scalable deep-learning engine. There is also a link to mxnet.incubator.apache.org and a note about joining January 2017.

Reddit: r/mxnet



The screenshot shows the r/mxnet Reddit community. It lists several posts: 1. 'Benchmarking Your Deep Learning Framework On-premise' by u/thomasdit (13 minutes ago). 2. 'Let Sentiment Classification Model speak for itself using Grad CAM' by u/ishitorix (4 days ago). 3. 'GluonNLP – Deep Learning Toolkit for Natural Language Processing' by u/thomasdit (5 days ago).

Community

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



Screenshot of the GitHub repository page for apache/incubator-mxnet. The page shows 87 open pull requests, 1,160 stars, and 5,395 forks. The repository has 8 projects and 8 milestones. A search bar at the top includes the filter "is:pr is:open". Below the search bar, there are buttons for "Filters", "Labels", "Milestones", and a green "New pull request" button. The main area displays a list of pull requests:

- Disable flaky test: test_spatial_transformer_with_type** (PR #11930) - 0 of 5 comments
- Generalized reshape_like operator** (PR #11928) - 3 comments
- [MXNET-711] Added updated logos to the powered by page** (PR #11921) - 1 of 1 comment
- [WIP] Improve gather_nd and scatter_nd doc** (PR #11918) - 5 of 6 comments

Community

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



Topic	Category	Users	Replies	Views	Activity
CNN and invariance to feature translation on the image 0 votes	Discussion	E 🐻	1	4	2m
How to access the output values of a sub-sub custom Block 0 votes	Gluon	A 🐻	1	4	12m
MaxPool2D on odd dimensional layers 0 votes		S 🐻	0	4	5h
Mxnet GPU freezes python 1 vote	Performance	L 🐻 J 🐻	3	29	6h
How to use mxnet gpu verion in kaggle kernel? 0 votes		U 🐻 V 🐻	3	20	21h
MxNet 1.2.1–module get_outputs() 0 votes	Discussion	X 🐻 Y 🐻	4	38	1d

Community

Mailing list:

dev@mxnet.apache.org

user@mxnet.apache.org

MXNet Customer Momentum



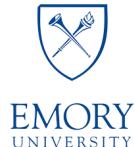
Software Platform Lab
Seoul National University



BOREALIS AI
RBC Institute for Research



Carnegie
Mellon
University



MediaNET LAB
Media Network Laboratory
KAIST
Korea Advanced Institute of
Science and Technology
School of
Electrical Engineering



Thank you!