

PyTorch Tutorial

@shagunsodhani

Questions are welcome at all times :)

Agenda

1. **PyTorch in Action**
2. PyTorch Ecosystem
3. PyTorch for production

Questions are welcome at all times :)

Agenda

1. **PyTorch in Action**
2. PyTorch Ecosystem
3. PyTorch for production

<https://colab.research.google.com/drive/1AybRrh3G-1kwE6sFW--nQcIDPQLoLXxc#scrollTo=Hahfkrrm98Zm>

PyTorch Tutorials

New to PyTorch?

The 60 min blitz is the most common starting point and provides a broad view on how to use PyTorch. It covers the basics all the way to constructing deep neural networks.

[Start 60-min blitz >](#)

PyTorch Recipes

Bite-size, ready-to-deploy PyTorch code examples.

[Explore Recipes >](#)

PyTorch Tutorials

All	Audio	Best Practice	C++	CUDA	Frontend APIs	Getting Started
Image/Video	Interpretability	Memory Format	Model Optimization	Named Tensor		
Parallel and-Distributed-Training	Production	Quantization	Reinforcement Learning			
Tensorboard	Text	TorchScript				

Agenda

1. PyTorch in Action
- 2. PyTorch Ecosystem**
3. PyTorch for production

Questions are welcome at all times :)

PyTorch Ecosystem

- Around 40 featured projects, tools, and libraries
- Developed by researchers in academia and industry, application developers, and ML engineers.
- <https://pytorch.org/ecosystem/>

Machine Learning

skorch

skorch is a high-level library for PyTorch that provides full scikit-learn compatibility.

Machine Learning

PyTorch Lightning

PyTorch Lightning is a Keras-like ML library for PyTorch. It leaves core training and validation logic to you and automates the rest.

Machine Learning

Poutyne

Poutyne is a Keras-like framework for PyTorch and handles much of the boilerplating code needed to train neural networks.

Vision

TORCHVISION

The `torchvision` package consists of popular datasets, model architectures, and common image transformations for computer vision.

Vision

Albumentations

Fast and extensible image augmentation library for different CV tasks like classification, segmentation, object detection and pose estimation.

Vision

Kornia

Kornia is a differentiable computer vision library that consists of a set of routines and differentiable modules to solve generic CV problems.

NLP

AllenNLP

AllenNLP is an open-source research library built on PyTorch for designing and evaluating deep learning models for NLP.

Graph

DGL

Deep Graph Library (DGL) is a Python package built for easy implementation of graph neural network model family, on top of PyTorch and other frameworks.

Graph

PyTorch Geometric

PyTorch Geometric is a library for deep learning on irregular input data such as graphs, point clouds, and manifolds.

Model Interpretability

Captum

Captum (“comprehension” in Latin) is an open source, extensible library for model interpretability built on PyTorch.

Privacy Preserving ML

CrypTen

CrypTen is a framework for Privacy Preserving ML. Its goal is to make secure computing techniques accessible to ML practitioners.

PyTorch Hub

```
model = torch.hub.load('pytorch/vision', 'resnet18',  
pretrained=True)
```

Agenda

1. PyTorch in Action
2. PyTorch Ecosystem
3. **PyTorch for production**

Questions are welcome at all times :)

<https://pytorch.org/cppdocs>

C++ FRONT-END

The C++ frontend is a pure C++ interface to PyTorch that follows the design and architecture of the established Python frontend. It is intended to enable research in high performance, low latency and bare metal C++ applications.

<https://pytorch.org/cppdocs>

```
#include <torch/csrc/autograd/variable.h>
#include <torch/csrc/autograd/function.h>

torch::Tensor a = torch::ones({2, 2}, torch::requires_grad());
torch::Tensor b = torch::randn({2, 2});
auto c = a + b;
c.backward(); // a.grad() will now hold the gradient of c w.r.t. a.
```

<https://pytorch.org/docs/stable/onnx.html>

NATIVE ONNX SUPPORT

Export models in the standard ONNX (Open Neural Network Exchange) format for direct access to ONNX-compatible platforms, runtimes, visualizers, and more.

ONNX

1. Standard for exchanging ML models
2. Supports interoperability between frameworks
3. Train with framework X, deploy with framework Y
4. Supports PyTorch, TensorFlow, Keras, Scikit-Learn, mxnet,....

<https://pytorch.org/serve/>

TORCHSERVE (EXPERIMENTAL)

TorchServe is an easy to use tool for deploying PyTorch models at scale. It is cloud and environment agnostic and supports features such as multi-model serving, logging, metrics and the creation of RESTful endpoints for application integration.

TorchServe

1. Supports Python-based and TorchScript-based models
2. Model versioning + rollback
3. Batches inference requests
4. Dockerfile for easy deployment

TorchServe

```
torchserve --start --ncs --model-store model_store --models densenet161.mar
```

TorchServe

```
torchserve --start --ncs --model-store model_store --models densenet161.mar
```

```
curl -O https://s3.amazonaws.com/model-server/inputs/kitten.jpg  
curl http://127.0.0.1:8080/predictions/densenet161 -T kitten.jpg
```

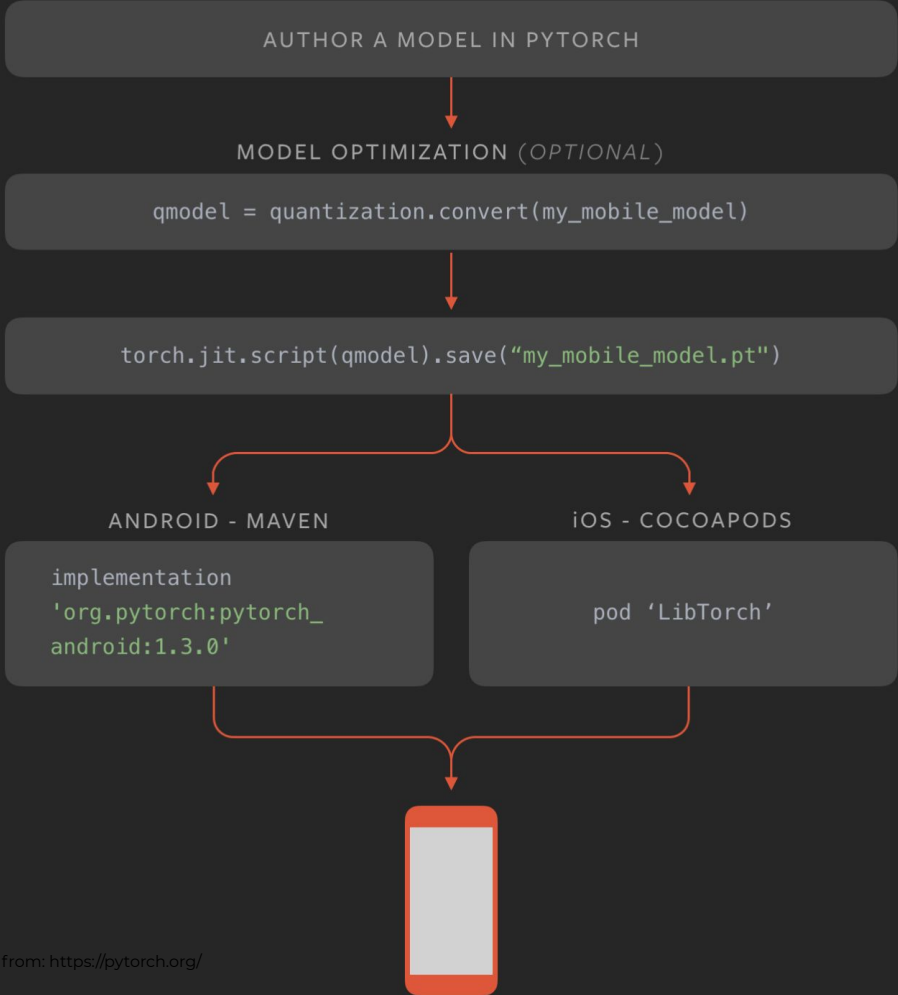
TorchServe

```
[
  {
    "tiger_cat": 0.46933549642562866
  },
  {
    "tabby": 0.4633878469467163
  },
  {
    "Egyptian_cat": 0.06456148624420166
  },
  {
    "lynx": 0.0012828214094042778
  },
  {
    "plastic_bag": 0.00023323034110944718
  }
]
```

<https://pytorch.org/mobile>

MOBILE (EXPERIMENTAL)

PyTorch supports an end-to-end workflow from Python to deployment on iOS and Android. It extends the PyTorch API to cover common preprocessing and integration tasks needed for incorporating ML in mobile applications.



TorchScript

TorchScript

1. TorchScript is an intermediate representation of a PyTorch model.
2. It can be run in a high-performance environment such as C++.

Quantization

Quantization

1. Lower precision data (int8)
2. Savings in model size, memory bandwidth, and inference time
3. PyTorch supports:
 - Dynamic Quantization
 - Post-Training Static Quantization
 - Quantization Aware Training

<https://pytorch.org/get-started/cloud-partners>

CLOUD SUPPORT

PyTorch is well supported on major cloud platforms, providing frictionless development and easy scaling through prebuilt images, large scale training on GPUs, ability to run models in a production scale environment, and more.

<https://pytorch.org/get-started/cloud-partners>



Alibaba Cloud



Amazon Web Services



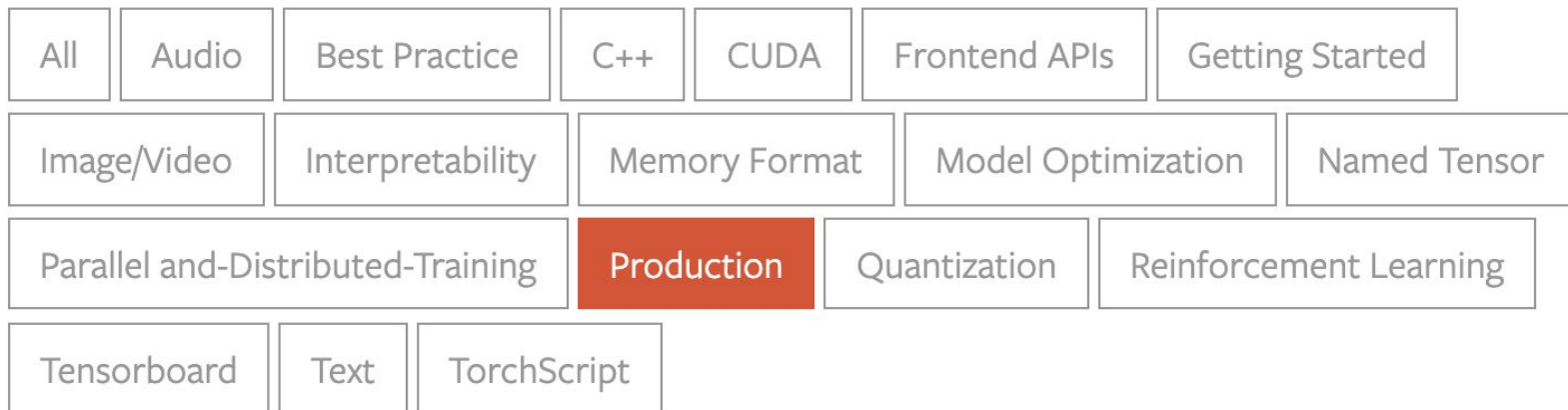
Google Cloud Platform



Microsoft Azure



<https://pytorch.org/tutorials/>



Community

pytorch / pytorch

Watch 1.5k Unstar 39.5k Fork 10.2k

Code Issues 4,469 Pull requests 1,550 Actions Projects 8 Wiki Security Insights

Branch: master

Go to file

Add file

Clone

xcheng16 authored and facebook-github-bot committed 0b375... 27,599 commits 3,805 branches 39 tags

.circleci	Revert D220688657: [pytorch][PR] Remove global CMAKE_INSTALL_R...	6 hours ago
.ctags.d	Add a .ctags.d/ toplevel directory (#18827)	15 months ago
.github	Stop running target bot on ci-all (#40186)	4 hours ago
.jenkins	Revert D22076711: [Reland #3] Include AT_PARALLEL_OPENMP/AT_P...	2 hours ago
android	[android][test_app] cleanup (#40136)	8 hours ago
aten	Revert D22076711: [Reland #3] Include AT_PARALLEL_OPENMP/AT_P...	2 hours ago
benchmarks	observer bench: add CUDA (#39360)	12 days ago
binaries	Add optimization blacklist as second arg to optimizeForMobile metho...	34 minutes ago
c10	[pytorch] fix CUDA_KERNEL_ASSERT macro for android build (#40151)	2 hours ago
caffe2	Revert D22076711: [Reland #3] Include AT_PARALLEL_OPENMP/AT_P...	2 hours ago
cmake	Revert D220688657: [pytorch][PR] Remove global CMAKE_INSTALL_R...	6 hours ago
docker	Add sccache support for hcc and hip-clang in ROCm (#38451)	last month
docs	Adding torch.futures to API docs (#40051)	1 hour ago
ios	[iOS] Disable NNPACK on iOS builds (#39868)	17 hours ago
modules	Remove `Caffe2_MAIN_LIBS` (#38408)	last month
scripts	[iOS] Disable NNPACK on iOS builds (#39868)	17 hours ago
src/modules	'Re-secure with internal repository' (#12652)	? years ago

About

Tensors and Dynamic neural networks in Python with strong GPU acceleration

pytorch.org

neural-network autograd gpu numpy deep-learning tensor python machine-learning

Readme

View license

Latest release

Stable C++ Frontend, Distributed RPC framework, and more. New experimental higher-level autograd API, Channels Last memory format, and more.
on Apr 21

+ 38 releases

Used by 5,000+

32,381

Community



all categories ▾

Latest

New (98)

Unread (81)

Top

Categories

+ New Topic

Topic		Replies	Views	Activity
Does number of gradient accumulation steps affect model's performance? • ■ autograd		0	3	9m
How to solve this error?		3	258	12m
A simple extension of nn.Sequential • ■ vision		2	11	19m
'aten::slow_conv_transpose2d' not support in 'QuantizedCPUtensorID' • ■ quantization		2	25	27m
Learning - Why does .unsqueeze(-1) magically work? •		3	22	28m
Dose static quantization support CUDA? • ■ quantization		1	14	28m
<input checked="" type="checkbox"/> Segmentation fault (core dumped). when I was using CUDA		7	48	1h
[libtorch] build failed with opencv-4.2.0 using cmake •		0	5	1h
A little code problem •		0	7	1h
Is there a way to insert data augmentation into the training data based on the validation data? •		1	5	1h
Memory difference depending on whether the tensor was creating on gpu or pushed to gpu? Strange		2	21	2h

is-number-of-gradient-accumulation-steps-affect-models-performance/85859

Agenda

1. PyTorch in Action
2. PyTorch Ecosystem
3. PyTorch for production

Questions are welcome at all times :)

Thank you

@shagunsodhani