Welcome to the PyTorch developer's wiki!

Please read our [[best practices|Where or how should I add documentation]] if you're interested in adding a page or making edits

User docs

- Release notes
- PyTorch Versions
- Public API definition and documentation

Onboarding

New to PyTorch? Don't know where to start? - [[Core Onboarding|Core Frontend Onboarding]]

Developer docs

- Developer FAQ
- [[Where should I add documentation?|Where or how should I add documentation]]
- PyTorch Data Flow and Interface Diagram
- Multiprocessing Technical Notes
- Software Architecture for c10
- PyTorch JIT IR format (slightly out of date now)
- TH to ATen porting guide
- Writing Python in C++ (a manifesto)
- Introducing Quantized Tensor
- Life of a Tensor
- How to use `TensorIterator`
- Running and writing tests
- Writing memory format aware operators
- Guide for adding type annotations to PyTorch
- The torch.fft module in PyTorch 1.7
- PyTorch-ONNX exporter

Notes

- Automatic Mixed Precision package
- Automatic Mixed Precision examples
- Autograd mechanics
- Broadcasting semantics
- CPU threading and TorchScript inference
- CUDA semantics
- Frequently Asked Questions
- Extending PyTorch
- Features for large-scale deployments

- Multiprocessing best practices
- Reproducibility
- Serialization semantics
- Windows FAQ
- Python Language Reference Coverage
- Complex Numbers
- Android
- iOS
- How-to: Writing PyTorch & Caffe2 Operators
- CUDA IPC Refcounting implementation explained
- Autograd
- Code Coverage Tool for Pytorch
- How to write tests using FileCheck
- PyTorch Release Scripts
- Serialized operator test framework
- Observers
- Snapdragon NPE Support
- Using TensorBoard in ifbpy

Named Tensors

- Named Tensors
- Named Tensors operator coverage

Quantization

- Introduction to Quantization
- Quantization Operation coverage
- Implementing native quantized ops
- Extend PyTorch Quantization to Custom Backends

JIT/TorchScript

- JIT Technical Overview
- Current workflow
- Static Runtime
- TorchScript serialization
- PyTorch Fuser
- Implementation reference for the CUDA PyTorch JIT Fuser
- TorchScript
- TorchScript Language Reference
- TorchScript Unsupported Pytorch Constructs

Distributed

- Distributed RPC Framework
- Distributed Autograd Design

- Remote Reference Protocol
- Distributed Data Parallel
- Distributed communication package
- Contributing to PyTorch Distributed

C++

- PyTorch with C++
- The C++ Frontend
- PyTorch C++ API
- Tensor basics
- Tensor Creation API
- Tensor Indexing API
- MaybeOwned<Tensor>
- Installing C++ Distributions of PyTorch
- Torch Library API
- libtorch
- C++ / Python API parity tracker
- TensorExpr C++ Tests
- JIT C++ Tests
- C++ Frontend Tests
- FAQ
- Best Practices to Edit and Compile Pytorch Source Code On Window

Benchmarks

- Distributed Data Parallel Benchmark
- Fast RNN benchmarks
- PyTorch/Caffe2 Operator Micro-benchmarks
- **torch_function** micro-benchmarks
- Benchmarking tool for the autograd AP
- Modular Benchmarking Components

DataLoader

- DataPipe
- DataPipe test requirements

Workflow docs

- Continuous Integration
- Bot commands
- Code review values
- Lint as you type
- Pull request review etiquette
- Debugging with SSH on Github Actions
- Using hud.pytorch.org

Community

- Code of Conduct
- Contributing
- PyTorch Contribution Guide
- PyTorch Governance

Archived

- Breaking Changes from Variable and Tensor merge (from 0.4 release)
- Tensor API changes for Caffe2 developers (from 1.0 release, plus some stuff on master)
- Autograd and Fork

Caffe2

- Caffe2
- Building Caffe2
- Doxygen Notes
- Docker & Caffe2
- Caffe2 implementation of Open Neural Network Exchange (ONNX)
- Caffe2 ONNX op coverage
- nomnigraph
- Caffe2 & TensorRT integration
- Playground for Caffe2 Models
- How to run FakeLowP vs Glow tests
- Using ONNX and ATen to export models from PyTorch to Caffe2
- An ATen operator for Caffe2