PyTorch JIT

This folder contains (most of) the C++ code for the PyTorch JIT, a language and compiler stack for executing PyTorch models portably and efficiently. To learn more about the JIT from a user perspective, please consult our reference documentation and tutorials.

A brief summary of the source tree: - OVERVIEW.md: High-level technical overview of the JIT. - frontend/: Taking PyTorch modules in Python and translating them into the JIT IR. - ir/: Core IR abstractions. - runtime/: Interpreter, graph execution, and JIT operators. - codegen/: Generating efficient, hardware-specific code for JIT subgraphs. - serialization/: Saving and loading modules. - api/: Any user-facing C++ or Python interfaces. - python/: Binding stuff into Python or accessing information from the Python environment. - testing/: Utilities and helpers for testing. - mobile/: Mobile-specific implementations of runtime components. - passes/: IR-to-IR passes, generally for optimization and lowering. - generated/: This folder is generated by the PyTorch build, and contains bindings for native PyTorch operators into the JIT.

Refer to each folder for more in-depth documentation.

Other relevant parts of the codebase not contained here: - aten/src/ATen/core: contains JIT code re-used by other elements of the runtime system (eager, mobile, etc.)