

Training libraries

Inference libs

Python client

C++ client ...

C API

Distributed master

Dataflow executor

Const Var MatMul Conv2D ReLU Queue ...

Kernel implementations

RPC RDMA ...

Networking layer

CPU GPU ...

Device layer