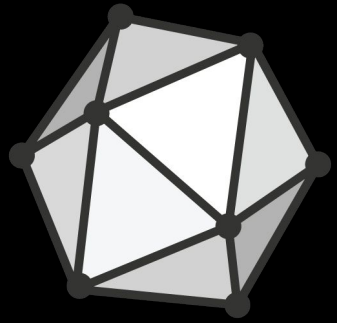
Shape Inference Improvements

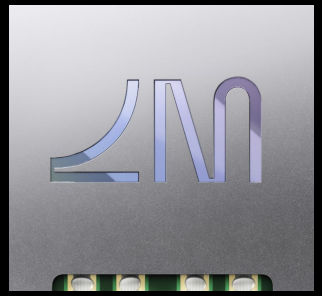
Presenter: Calvin McCarter
Lightmatter

How we use ONNX

Fine-tune weights



Analyze coverage & optimize performance



Run at the speed of light

Presenter: Calvin McCarter
Lightmatter

How we use ONNX shape inference

ONNX shape inference is helpful for our ONNX analyzer, and for our compiler (based on Glow).

Presenter: Calvin McCarter
Lightmatter

Shape Inference Improvements

1. Continued progress with data propagation

   MLPerf Resnet34-SSD: image -> shape -> cast -> slice -> concat -> cast -> reshape
2. Shape inference where rank is unknown but certain axes are known

   MLPerf RNN-T Joint model: Add{unknown, [29,]} -> [unknown 29,]
3. Arithmetic expressions containing variables
4. (?) A documented API for users to manually update `graph.value_info`

Shape inference might feel tedious to work on, but it's really helpful.
Thank you!

Presenter: Calvin McCarter
Lightmatter