



ONNX

Converter SIG Updates

Kevin Chen (NVIDIA)

Outline

Convertor updates

- Frontend converters (Framework → ONNX)
 - PyTorch, Tensorflow, sklearn
- Backend convertors (ONNX → Runtime Framework)
 - TensorRT, Tensorflow

Roadmap

Get involved!

PyTorch → ONNX

Latest release: PyTorch 1.12 + opset 16

- Supports exporting nn.Module calls as ONNX local functions
- New op support: tanhshrink, hardshrink, softshrink, __xor__, isfinite, log10, diagonal, cross, linalg_cross, cdist, pairwise_distance, bucketize, pixel_unshuffle, embedding_renorm, aminmax, amax, amin, grid_sample
- Added ONNX BFloat16 and Optional type support
- Added support for exporting quantized models, and exporting models with Apex O2
- Added support for ATen fallback for non-Caffe2 backends
- Supported PyTorch ATen operators are listed on [documentation online](#)
- Stay tuned for "PyTorch Converter for ONNX" talk by Bowen Bao at 1:35pm PST

Tensorflow → ONNX

Latest release: tf2onnx 1.11.1 + opset 16, up to TF 2.8

- Default opset export version updated to 13
- New op support: TFL_BATCH_MATMUL, TFL_MATMUL, TensorScatterAdd , Rint
- QDQ model export improvements
- Improvements in support for TFLite

sklearn → ONNX

Latest release: sklearn-onnx 1.11.2 + opset 15

- New op support: sgd_oneclass_svm,
- Bug fixes for ops: Local Outlier Factor for n_neighbors, Simple Imputer, Isolation Forest and WOE-Transformer
- Improved library import performance

ONNX → TensorRT

Latest release: onnx-tensorrt 8.4 + opset 17

- New op support: Shrink, Xor
- Improved support for: Argmax, Argmin, Range, Celu, GEMM, Einsum
- Improved support for dynamically shaped models
- General performance improvements

ONNX → Tensorflow

Latest release: v1.10 + opset 15, up to TF 2.8

- New op support: All opset 15 operators are at least partially supported
- Added support for ONNX optional datatype

Roadmap

Improved documentation across convertors - done!

Minimal opset support - WIP

- Formal proposal to deprecate opset version < 9 in progress

Improved community-driven tooling - WIP

- Fuzz-testing, shape / type inference, quantization tools, constant-folding

Improve export of PyTorch models with LazyTensor
and Autograd functions

Get Involved!

- Feedback? Join us on Slack in the #onnx-converters channel
- Subscribe to [ONNX Converters SIG mailing list](#)
- Open up issues and partake in discussions on Github