



ONNX

OPERATORS SIG

BREAKOUT SESSION

Emad Barsoum (Microsoft)
Michal Karzynski (Intel)



AGENDA

- Operators SIG
- ONNX 1.7 Release
- Dynamic shapes decision
- Discussion: Coding style
- Discussion: PR and Issues

GOAL

Keep Up

Keep up with the latest progress in AI

Quality

Improve the quality of ONNX Operators

Clarity

Reduce ambiguity

Size

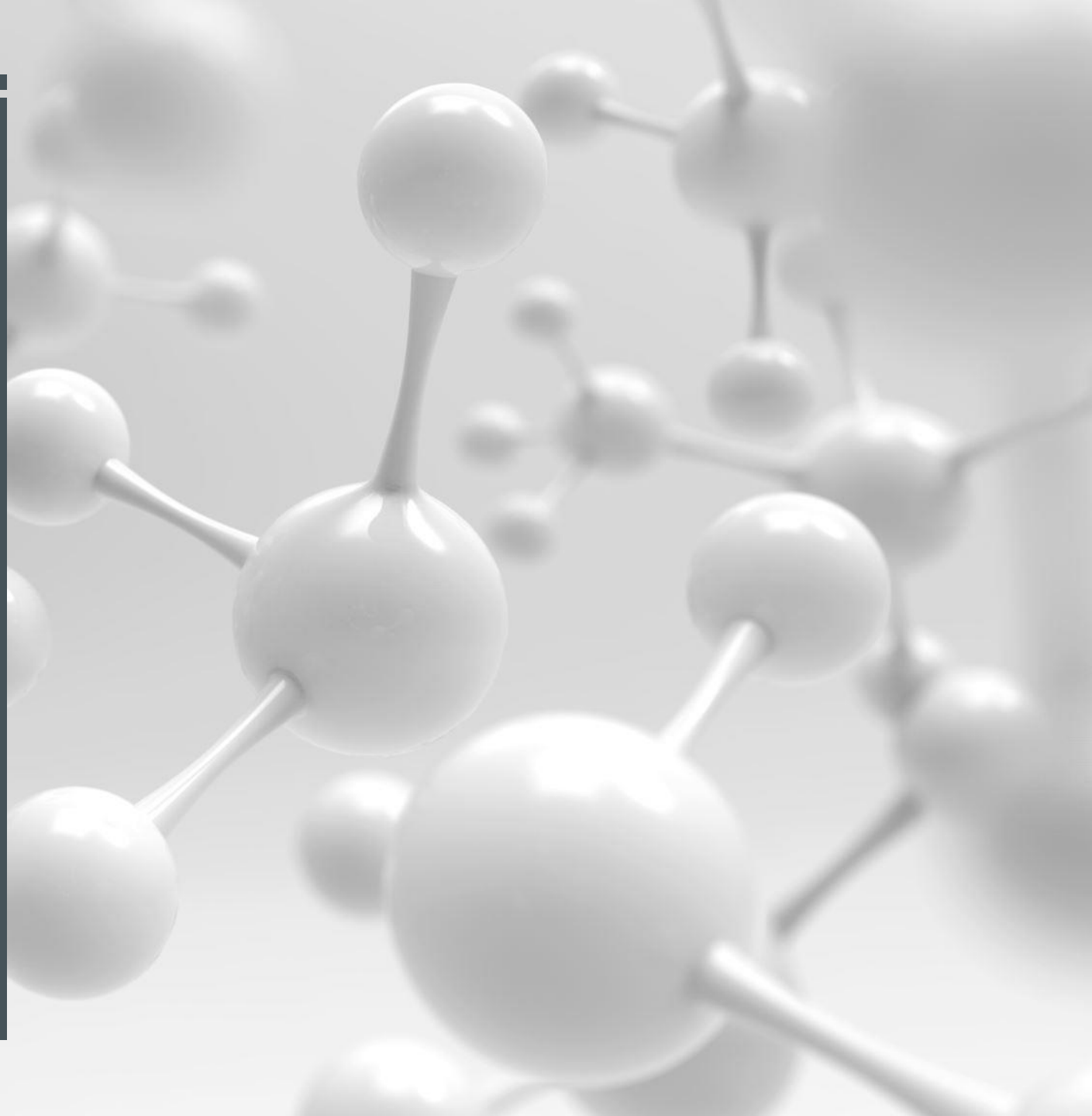
Avoid bloating ONNX spec

PRs and Issues

Keep up with PRs and operator issues

PARTICIPANTS

- Akinlawon Solomon (Qualcomm)
- Darren Crews (Intel)
- Dilip Sequeira (NVidia)
- Ganesan Ramalingam (Microsoft)
- Itay Hubara (Habana)
- Jianhao Zhang (JD)
- Ke Zhang (Alibaba)
- Leonid Goldgeisser (Habana)
- Milan Oljaca (Qualcomm)
- Ofer Rosenberg (Qualcomm)
- Rajeev K Nalawadi (Intel)
- Scott Cyphers (Intel)
- Shlomo Raikin (Habana)
- Simon Long (GraphCore)
- Spandan Tiwari (Microsoft)
- Wei-Sheng Chin (Microsoft)
- Weiming Zhao (Alibaba)
- Liqun Fu (Microsoft)



COMMUNICATION

- Gitter channel: <https://gitter.im/onnx/operators>
- Discussions on GitHub PRs and issues
- Meetings announcement are on Gitter
- Docs and meeting notes are in onnx/sigs
<https://github.com/onnx/sigs/tree/master/operators>
- Procedure for adding or updating op
 - <https://github.com/onnx/onnx/blob/master/docs/AddNewOp.md>

ONNX 1.7 RELEASE

Loss

- SoftmaxCrossEntropy
- NegativeLogLikelihood
- MeanSquareDistance

Operators

- Einsum
- GreaterOrEqual
- LessOrEqual
- UnfoldToDepth
- Inverse
- Gradient
- GraphCall
- Celu

Quantization

- MaxPool
- ReduceMin
- ReduceMax

Optimizer

- Momentum SGD
- Adagrad

GET INVOLVED: SUBMIT AND REVIEW PRS

Filters

☐ Clear current search query, filters, and sorts

☐ ⓘ 101 Open ✓ 70 Closed Author ▾ Label ▾

☐ ⓘ Add new "Diag" operator into ONNX **operator**
#2512 opened on Dec 17, 2019 by thangdt

☐ ⓘ Really need onnx to support DCN (deformal conv) operator.... **operator**
#2485 opened on Dec 2, 2019 by lancasterJie

☐ ⓘ Resize use Sizes but Scales is single **operator**
#2451 opened on Nov 12, 2019 by leroyzhao

Filters

☐ Clear current search query, filters, and sorts

☐ ⓘ 22 Open ✓ 55 Closed Author ▾ Label ▾ Projects ▾

☐ ⓘ Feature/fft ifft ✗ **operator**
#2625 opened on Feb 26 by jeremycochoy • Draft

☐ ⓘ Fix axis optional for CumSum • **operator**
#2611 opened on Feb 18 by pluradj • Changes requested

☐ ⓘ Update Reshape op: Remove special meaning of 0 dims ✗ **operator**
#2553 opened on Jan 15 by jignparm • Review required

PR REVIEW

- PRs should be marked with the **Operator** label
 - <https://github.com/onnx/onnx/pulls?q=is:pr+is:open+label:operator>
- Ops Contributors Group should review the PRs according to guidelines
- Mature PRs can be discussed during bi-weekly sync
- Final approval by member of **SIG-operators-approvers** group

GITHUB DISCUSSION ISSUES

- Many open discussions marked with **Operator** label
- Ops Contributors Group should be active in discussions and encourage submission of PRs
- We should decide which discussions can be closed
- Still looking for best way to triage this large number of open issues



PROPOSAL

- Feel free to propose any improvements, such as:
 - Better testing, validation and coverage of ONNX operators.
 - Better documentation generation.
 - More operators.
 - A lot of existing manual steps need automation.
- For any big proposal, you will be invited in the SIG meeting to present it.

DYNAMIC SHAPES IN ACCELERATED DEVICES

- Dynamic shapes pose a problem to accelerators, which cannot know the memory size to allocate for a tensor.
- The issue applies to three different cases:
 - inferable ops (Shape, Slice),
 - non-inferable ops (NonZeros, Unique),
 - dynamic inputs (to the model).
- Should we add a flag/option to enable vendors to replace a dynamic shape with a statically defined shape?

DYNAMIC SHAPES IN ACCELERATED DEVICES

- Spec versus runtime implementation.
- Having a max size hint, should be part of spec or runtime?
 - Who decide the max size? Scientist who created the model? Based on what?
 - Shouldn't the size depend on the target hardware and the required accuracy?
 - If it is a hint, then it can be ignored. So should it be part of the spec?
- Even that some of the OP, can output variable size output. The underlying implementation can have a fixed size.
 - It could clip or error out, if the output exceeded the max size.

DISCUSSION: CODING STYLE

- We have a coding convention for operator:
 - <https://github.com/onnx/onnx/blob/master/docs/OpConventions>.
 - How many know about it?
- Currently, we aren't following the convention.
- Should we enforce it? Part of the CI?



DISCUSSION: PR AND ISSUE

- How long should we keep PR or issue open?
- If the original author didn't follow up in X weeks, should we close the issue?
 - What good value of X?
- If a PR or issue is opened, then how long should it take us to review it?
 - We need help, we can't keep up with all the PR and issues.
- How to improve the process?
- How to entice more people to participate?

THANKS FOR COMING!!!

Operator SIG resources

- Gitter channel <https://gitter.im/onnx/operators>
- Documents and artifacts:
<https://github.com/onnx/sigs/tree/master/operators>