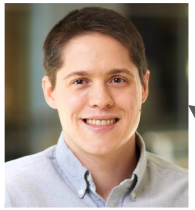
Paper and slides: https://davisjam.github.io/publications/

Analysis of Failures and Risks in Deep Learning Model Converters

A Case Study in the ONNX Ecosystem

Purvish Jajal

James C. Davis



(Joining remotely)

PURDUE UNIVERSITY | Elmore Family School of Electrical and Computer Engineering

Talk outline

- Overview of research literature: ONNX/etc.
- Our study on ONNX failures
  - Method
  - Results
  - Implications for you
- Ways you can get involved

A quick literature review

(for your reading lists)

Academic Work on ONNX

Empirical Work:

- *An Empirical Study of Challenges in Converting Deep Learning Models* [1]

Tooling Work:

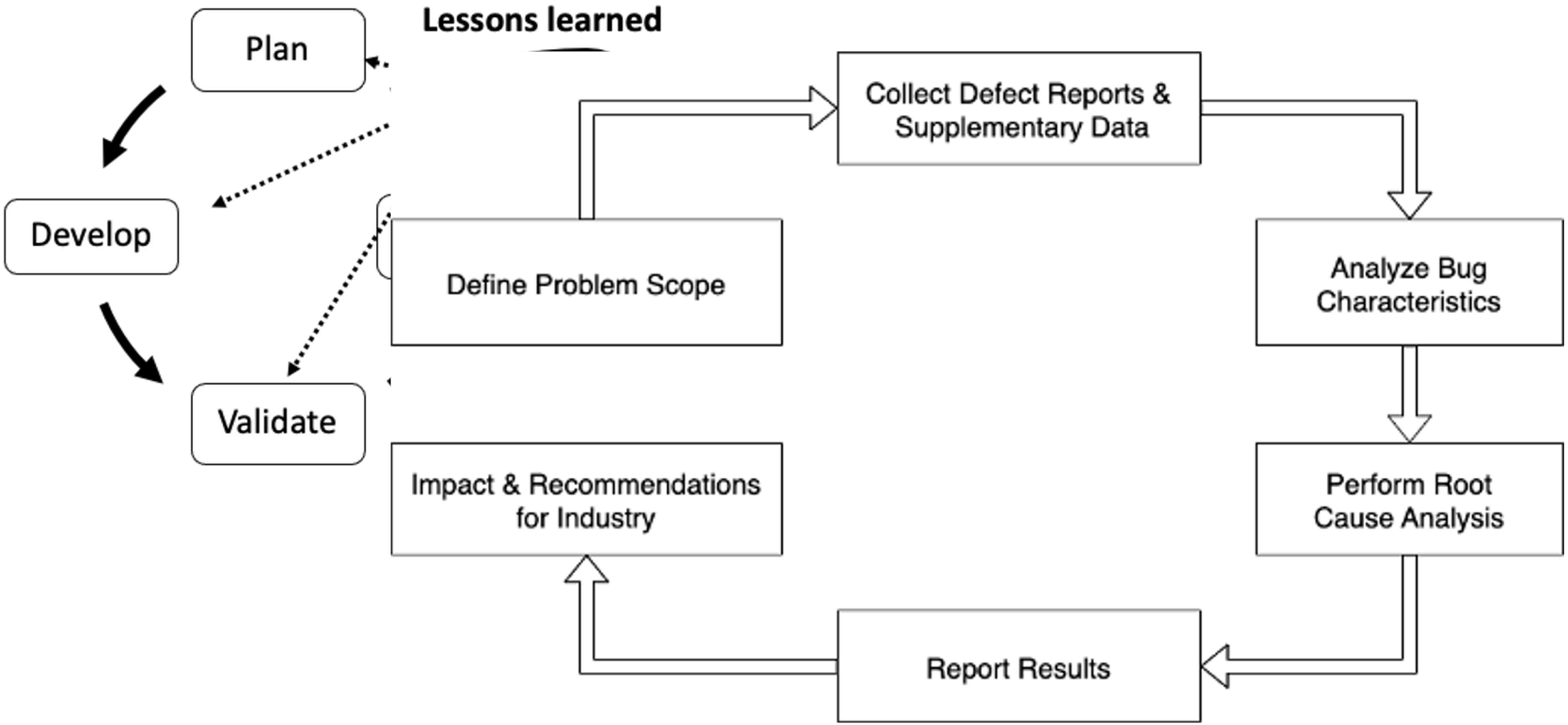
- *Sionnx: Automatic Unit Test Generator for ONNX Conformance* [2]

Selected Application Works:

- *Pre-Quantized Deep Learning Models Codified in ONNX to Enable Hardware/Software Co-Design* [3]
- *ESPnet-ONNX: Bridging a Gap Between Research and Production* [4]

Our study on ONNX failures

Background: Failure Analysis



Goal: Understanding failures of ONNX model converters.

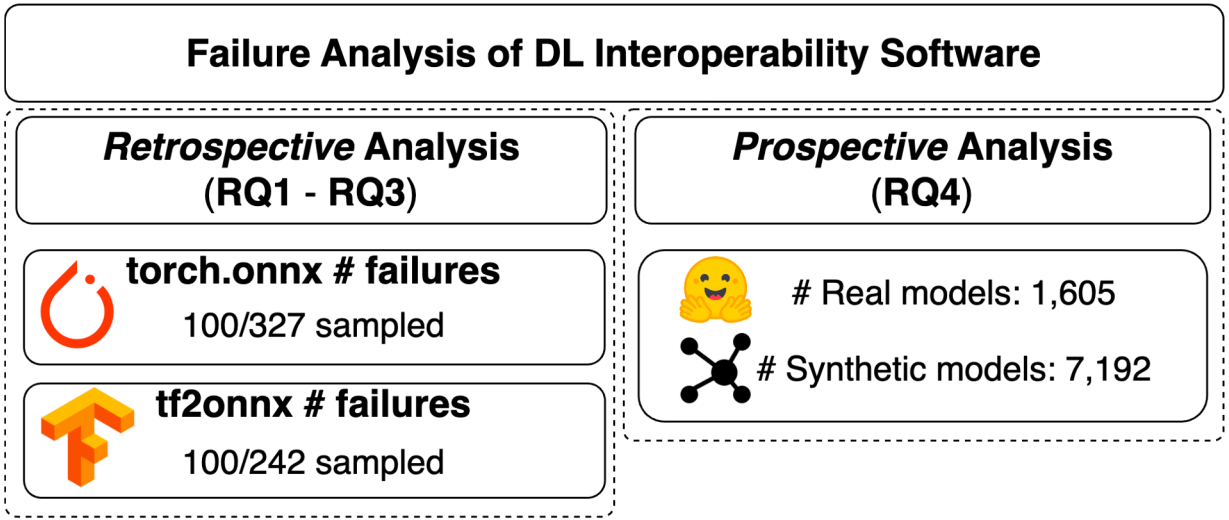
Audience:

- Product Users: To understand risks in the use of ONNX model converters
- Product Engineers: To reduce the occurrence of bugs in ONNX model converters

Method: Projects and questions

Model Converters Studied:
tf2onnx** and **torch.onnx

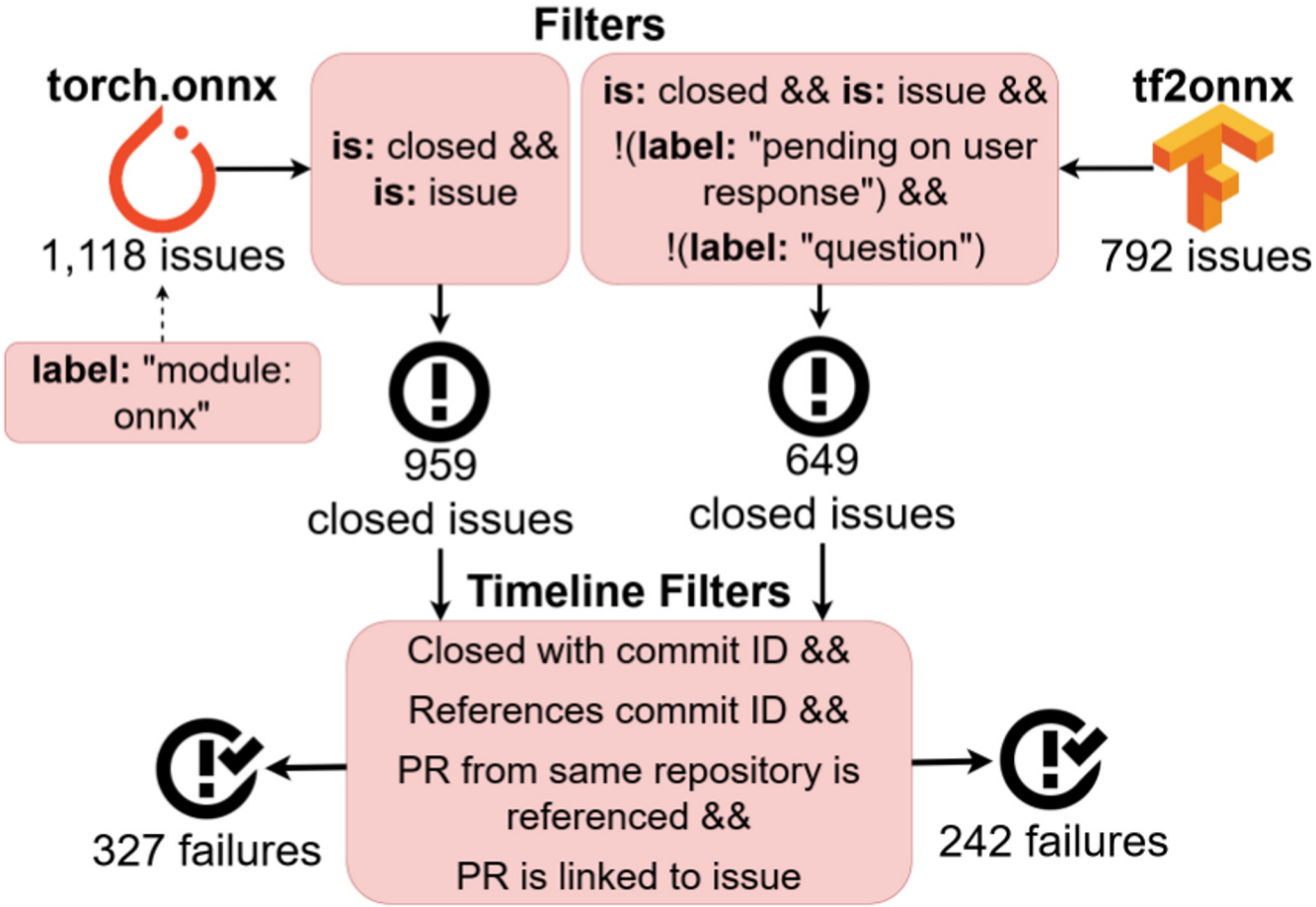
- *Retrospective Analysis*
  - *Closed GitHub Issues*
- *Systematic Testing*
  - *Real Models*
  - *Synthetic Models*



Method: Research Questions

- **RQ1:** *What are the characteristics of failures?*
- **RQ2:** *To what extent do changes in the ONNX specification correlate with model converter failures?*
- **RQ3:** *How often does interoperability software fail on real and systematically generated models?*

Method: Projects and questions



RQ1: Failure Symptoms

Common Failure Symptoms:

- Crashes
- Wrong Models

Symptom	TF	PT	Total
Crash	50	62	112 (56%)
Wrong Model	35	30	65 (33%)
Build Failure	3	2	5 (3%)
Bad Performance	2	1	3 (2%)
Hang	0	0	0 (0%)
Unreported	10	5	15 (8%)
Total	*100*	*100*	*200 (100%)*

RQ1: Failure Causes

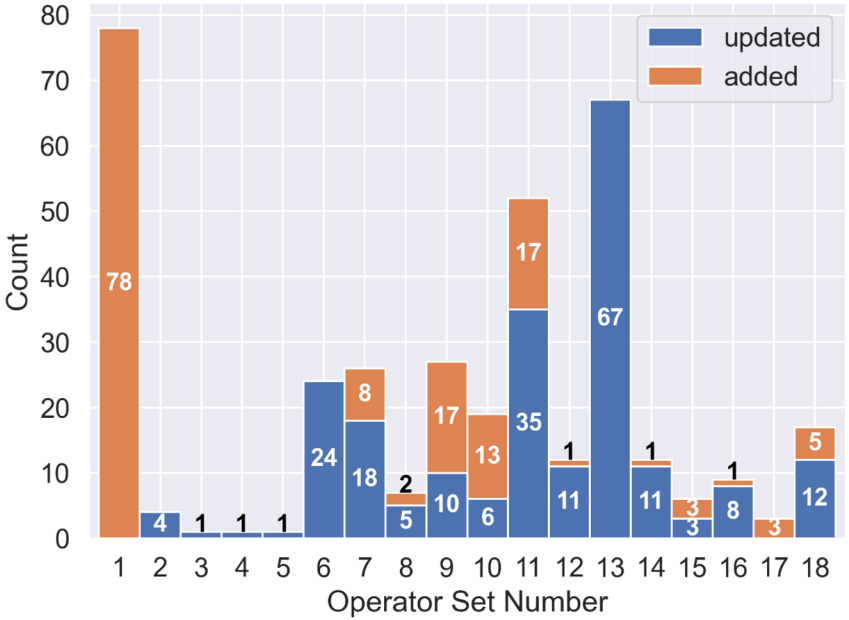
Common Causes:

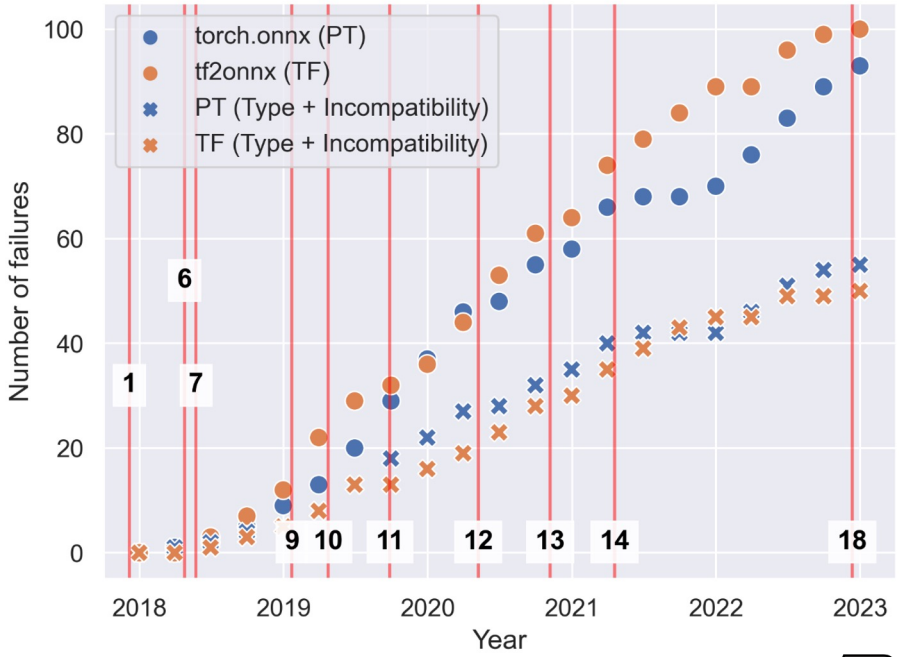
- Incompatibilities – External
- Type Problems – Node

Causes		TF	PT	Total
Incompatibility	External	**23**	**32**	**55 (28%)**
	Internal	2	0	2 (1%)
	Resource	0	0	0 (0%)
Type Problem	Node	**21**	**25**	**46 (23%)**
	Conventional	3	2	5 (3%)
	Tensor	1	2	3 (2%)
Algorithmic Error		**18**	**6**	**24 (12%)**
Shape Problem		9	12	21 (11%)
API Misuse		6	6	12 (6%)
Others		17	15	32 (16%)
Total		*100*	*100*	*200 (100%)*

RQ2: Changes sometimes break but not too often

- Weak positive correlation between changes and the number of failures (Spearman's ρ = 0.36).





RQ3: Synthetic Models

- *Real models*: Crashes and incorrect behavior → ~5% of models.
- *Synthetic models:* Reveal incorrect model behavior more often than *Real models*, ~25% (822/7,192) vs. ~1% (20/3,522).

Outcome	tf2onnx			torch.onnx		
	Real	Syn.	Syn. Con.	Real	Syn.	Syn. Con.
Start: Number of models	1,761	1,800	1,800	1,761	1,800	1,792
Unsuccessful Conversion (HF error)	456	N/A	N/A	342	N/A	N/A
Unsuccessful Conversion	65	0	0	20	190	0
Unsuccessful ORT loading	19	1,574	1,006	27	1,221	11
Incorrect Output	9	37	31	11	94	660
Successful	1,212	189	763	1,361	295	1,121

Incorrect Output is when the difference in outputs of original and converted models are $>10^{-3}$

Implications

ONNX failure modes

- Crashes are common
- Wrong models happen too – beware!
- The more unusual your model, the more likely a silent conversion failure

Testing of Converters:

- Model generation effective at inducing incorrect outputs (RQ4)
- Model generation may be a good addition to converter test suites

Ways you can get involved? Survey!

https://forms.gle/cqYqHAzNkCpaaonr8

Our paper: https://arxiv.org/abs/2303.17708

Other references

[1]: *An Empirical Study of Challenges in Converting Deep Learning Models* http://arxiv.org/abs/2206.14322

[2]: *Sionnx: Automatic Unit Test Generator for ONNX Conformance* http://arxiv.org/abs/1906.05676

[3]: *Pre-Quantized Deep Learning Models Codified in ONNX to Enable Hardware/Software Co-Design* http://arxiv.org/abs/2110.01730

[4]: *ESPnet-ONNX: Bridging a Gap Between Research and Production* https://arxiv.org/abs/2209.09756

Bonus Slides

Examples of NNSmith Synthetic Models