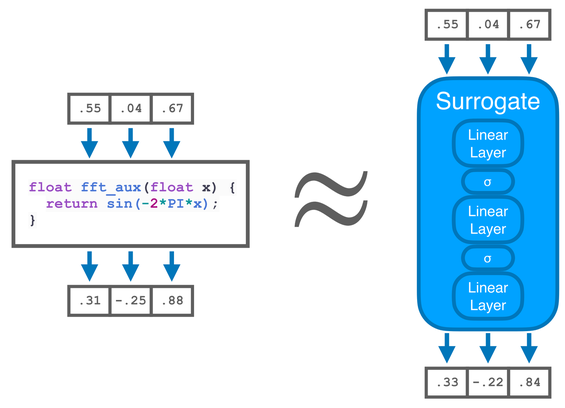
Learning to Compile Programs to Neural Networks

Logan Weber, Jesse Michel, Alex Renda, Michael Carbin

MIT

Neural Surrogates of Programs

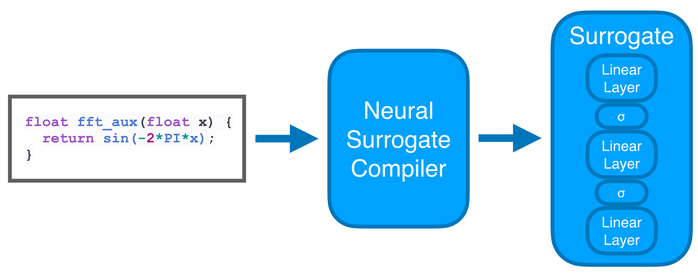
A *neural surrogate* of a program is a neural network that approximates the behavior of a program.



Researchers train these neural surrogates by collecting I/O pairs from the program.

Neural Surrogate Compilation

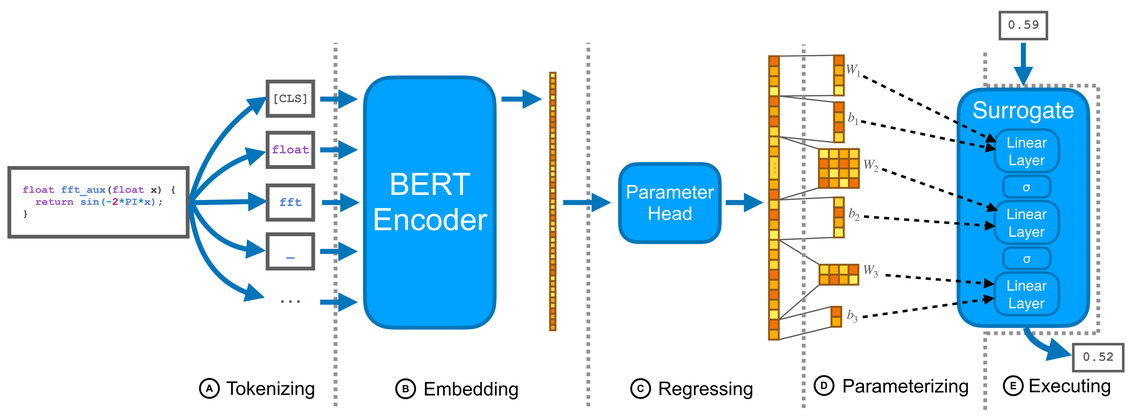
We present an alternative to standard neural surrogate development: *neural surrogate compilation*.



A neural surrogate compiler takes a program as input and produces a neural surrogate of the given program.

The CompNet Architecture

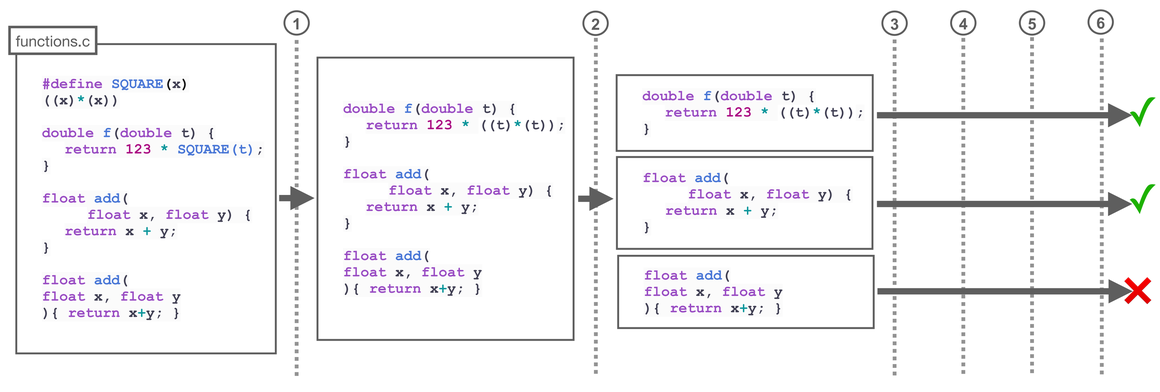
We develop a neural surrogate compiler using a *hypernetwork*—a neural network that produces neural network weights as output [1].



We call this architecture CompNet.

Datasets

To train neural surrogate compilers, we develop ExeStack, a dataset of executable programs from The Stack [2].



① Run C preprocessor
② Extract functions
③ Remove functions with pointer and numeric types
④ Collect I/O pairs
⑤ Remove nondeterministic functions
⑥ Deduplicate

We further specialize ExeStack to ExeStackCPN for CompNets and develop ParrotBenchCPN to evaluate CompNets.

ExeStackCPN ⊂ ExeStack

ParrotBenchCPN ⊂ ParrotBench [3]

Data Efficiency Improvements

We evaluate whether CompNet-initialized surrogates improve test loss over random initialization, when finetuned on I/O pairs. We compare to model-agnostic meta learning (MAML) [4] and pretrained surrogates.

ExeStackCPN

Dataset Size	CPN	MAML	PTS
0%	**84.40**×	1.42×	2.63×
0.1%	**10.43**×	0.91×	0.87×
1%	**2.90**×	0.51×	0.90×
10%	**4.12**×	1.54×	0.88×
100%	**6.67**×	1.53×	0.76×

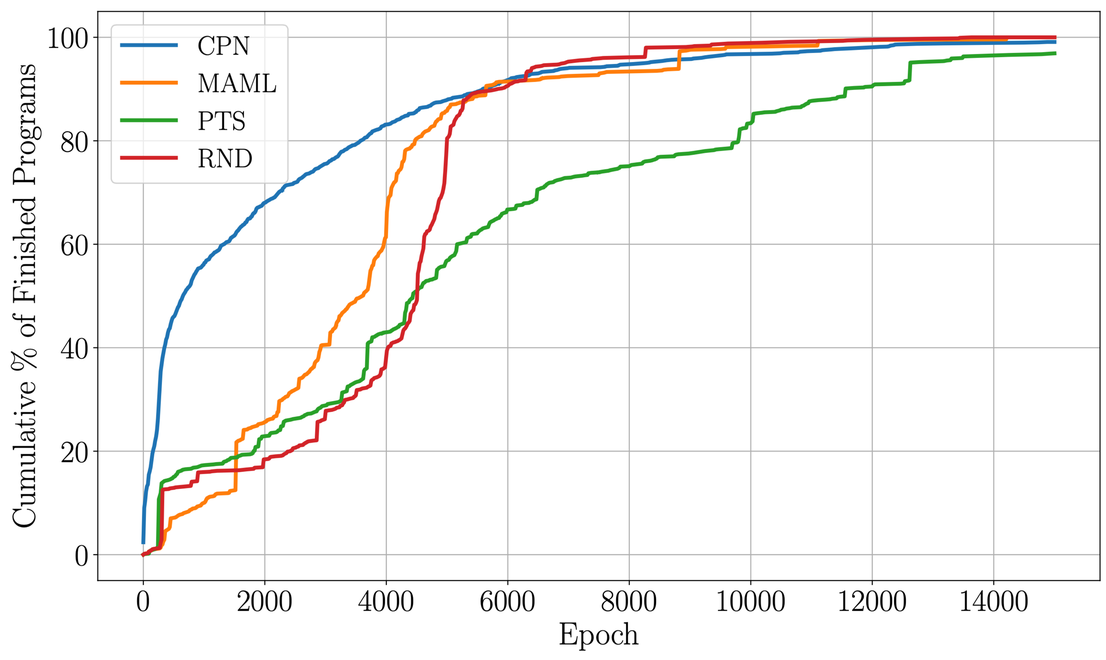
ParrotBenchCPN

% Data	CPN	MAML	PTS
0%	**1.81**×	0.90×	1.56×
0.1%	**1.98**×	0.94×	0.98×
1%	**1.77**×	0.93×	0.79×
10%	**2.38**×	1.11×	1.23×
100%	**1.68**×	0.81×	0.86×

Training Time Improvements

We evaluate whether CompNets improve training time, when finetuned on I/O pairs.

ExeStackCPN

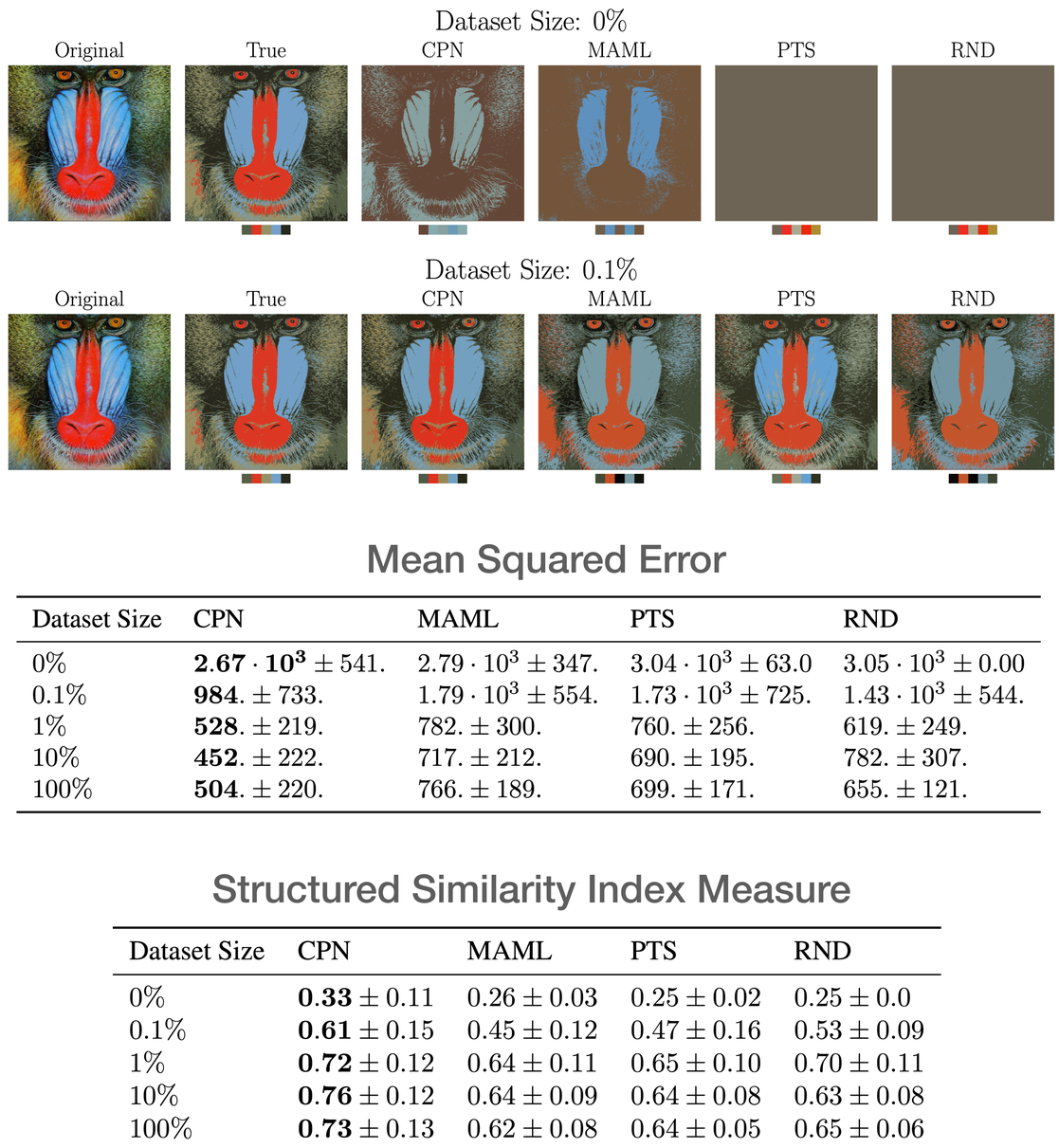


ParrotBenchCPN

Program	CPN	MAML	PTS
`fft`	**1.43**×	0.83×	0.80×
`invk2j`	0.49×	**0.65**×	0.56×
`kmeans`	**674.47**×	2.15×	86.87×
`sobel`	0.74×	**1.14**×	0.79×

End-to-End Error Improvements

We evaluate whether CompNets improve end-to-end error in a color quantization application.



Mean Squared Error

Dataset Size	CPN	MAML	PTS	RND
0%	$\mathbf{2.67 \cdot 10^3} \pm 541.$	$2.79 \cdot 10^3 \pm 347.$	$3.04 \cdot 10^3 \pm 63.0$	$3.05 \cdot 10^3 \pm 0.00$
0.1%	**984.** ± 733.	$1.79 \cdot 10^3 \pm 554.$	$1.73 \cdot 10^3 \pm 725.$	$1.43 \cdot 10^3 \pm 544.$
1%	**528.** ± 219.	782. ± 300.	760. ± 256.	619. ± 249.
10%	**452.** ± 222.	717. ± 212.	690. ± 195.	782. ± 307.
100%	**504.** ± 220.	766. ± 189.	699. ± 171.	655. ± 121.

Structured Similarity Index Measure

Dataset Size	CPN	MAML	PTS	RND
0%	**0.33** ± 0.11	0.26 ± 0.03	0.25 ± 0.02	0.25 ± 0.0
0.1%	**0.61** ± 0.15	0.45 ± 0.12	0.47 ± 0.16	0.53 ± 0.09
1%	**0.72** ± 0.12	0.64 ± 0.11	0.65 ± 0.10	0.70 ± 0.11
10%	**0.76** ± 0.12	0.64 ± 0.09	0.64 ± 0.08	0.63 ± 0.08
100%	**0.73** ± 0.13	0.62 ± 0.08	0.64 ± 0.05	0.65 ± 0.06

[1] Ha, D., Dai, A. M., and Le, Q. V. Hypernetworks. In *International Conference on Learning Representations*, 2017.

[2] Kocetkov, D., Li, R., Ben Allal, L., Li, J., Mou, C., Muñoz Ferrandis, C., Jernite, Y., Mitchell, M., Hughes, S., Wolf, T., Bahdanau, D., von Werra, L., and de Vries, H. The Stack: 3 TB of Permissively Licensed Source Code. *arXiv preprint 2211.15533*, 2022.

[3] Esmaeilzadeh, H., Sampson, A., Ceze, L., and Burger, D. Neural acceleration for general-purpose approximate pro- grams. In *International Symposium on Microarchitecture*, 2012a.

[4] Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, 2017.