AIRepair: A Repair Platform for Neural Networks

Xidan Song[1], Youcheng Sun[1], Mustafa A. Mustafa[1][2], Lucas C. Cordeiro[1]

[1]Department of Computer Science, University of Manchester, UK

[2]imec-COSIC, KU Leuven, Belgium

AIRepair: A brief introduction

- AIREPAIR is a tool for integrating and evaluating existing (and future) repair techniques on neural networks.

- It performs pre-processing and environment-isolating on different benchmarks to make them capable of different frameworks.

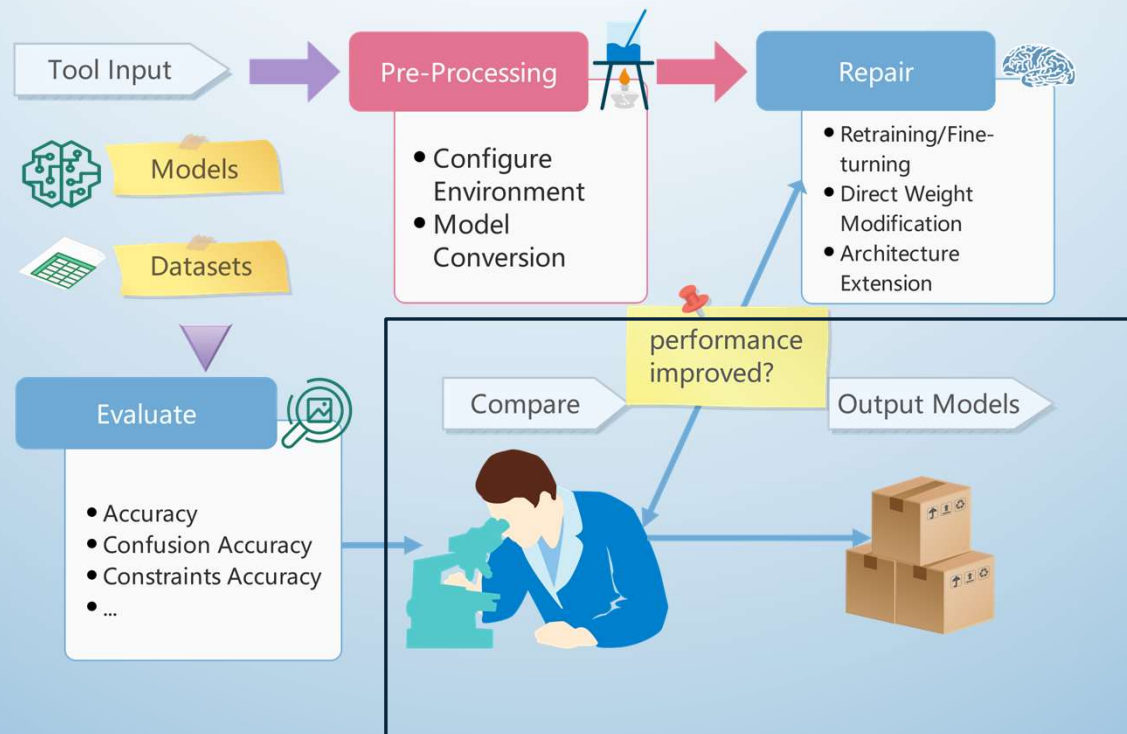- We benchmark five repair techniques on 8 types of neural network models across four datasets.

AIRepair Architecture

AIRepair Architecture
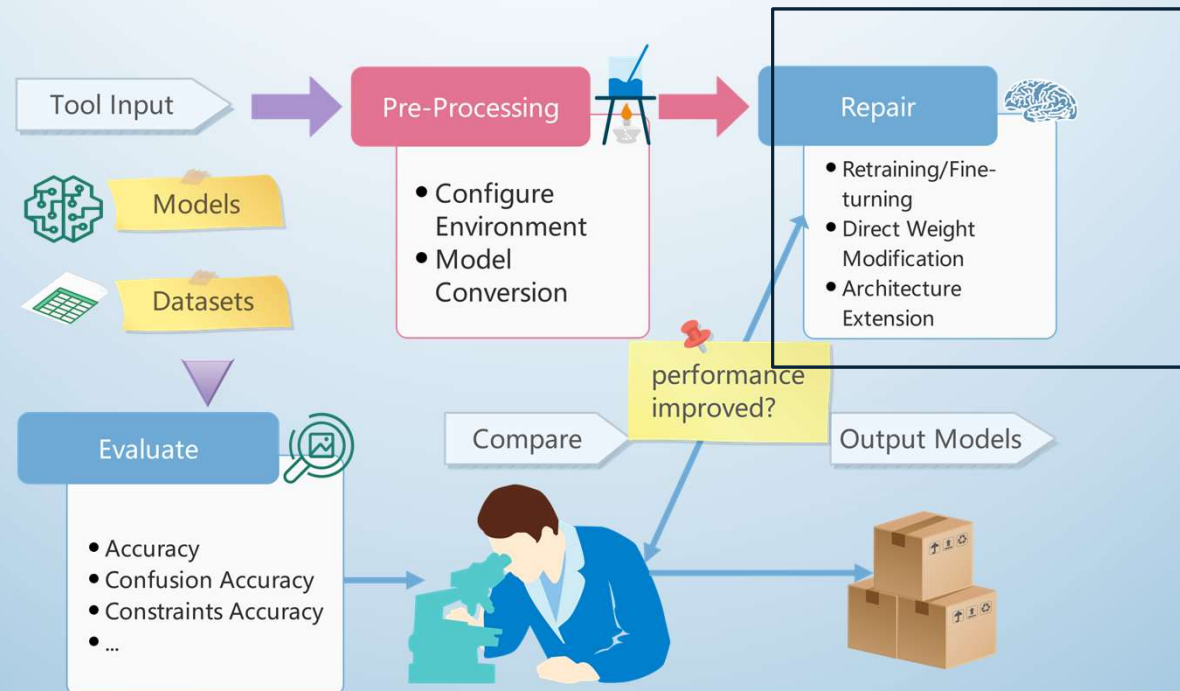
AIRepair Architecture
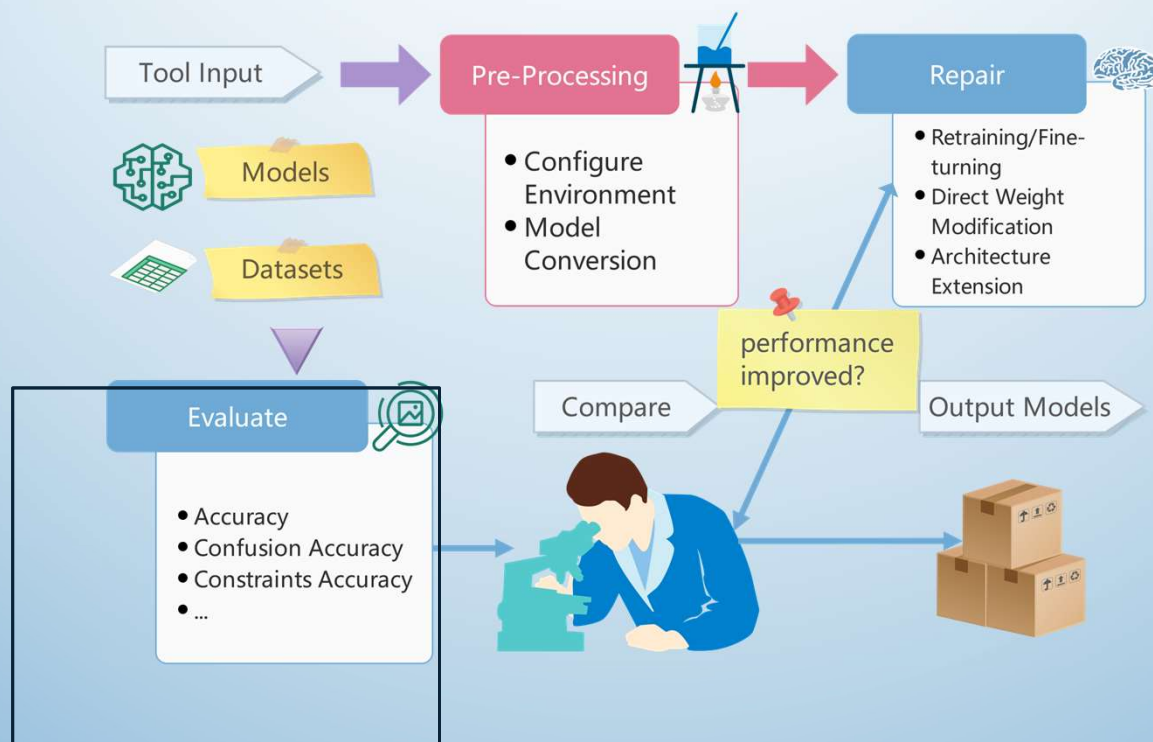
AIRepair Architecture

Example Usage

```
python AIRepair.py [-h] [--all]
[--net_arch NETARCH] [--dataset DATASET]
[--pretrained PATH_AND_FILENAME]
[--depth DEPTH][--method METHOD] [--auto]
[--additional_param PARAM]
[--input_logs INPUT_LOGS]
[--testonly]

python AIRepair.py --method apricot deeprepair dl2 --
pretrained cifar10_resnet34_baseline.pt --dataset cifar10 -
-net_arch resnet --depth 34
```

Available at: https://github.com/theyoucheng/AIRepair

AIRepair Results

Datasets		CIFAR-10			CIFAR-100			MNIST	Fashion-MNIST
Models		ResNet18	ResNet34	ResNet50	ResNet18	ResNet34	ResNet50	MNIST	Fashion-MNIST
Baselines	Acc.	92.05%	91.34%	94.42%	46.84%	44.16%	47.36%	99.45%	92.20%
	Const.	90.51%	90.27%	90.66%	86.62%	85.95%	85.21%	99.96%	100%
Apricot	Acc.	-2.65%	-0.38%	-3.4%	+9.02%	+13.74%	+11.15%	+0.06%	+0.61%
DeepRepair	Acc.	+0.5%	-1.27%	-4.14%	+10.91%	+21.42%	+20.32%	+0.17%	+0.47%
	Const.	-9.46%	-8.82%	-12.77%	-37.62%	-34.95%	-29.71%	-0.43%	-4.10%
DL2	Acc.	-2.16%	+0.23%	-1.95%	+0.87%	+1.17%	-1.16%	+0.08%	+0.28%
	Const.	+9.3%	+9.61%	+5.4%	-0.49%	-0.89%	-0.4%	+2.55%	+6.27%

We ran experiments on a machine with Ubuntu 18.04.6 LTS OS Intel(R) Xeon(R) Gold 5217 CPU @ 3.00GHz and two Nvidia Quadro RTX 6000 GPUs.

The models in the experiments are available at: https://zenodo.org/record/7627801#%23.Y-X6g3bP3tU

References: Fischer, Marc, et al. "DL2: training and querying neural networks with logic." *International Conference on Machine Learning*. PMLR, 2019. Yu, Bing, et al. "Deeprepair: Style-guided repairing for deep neural networks in the real-world operational environment." IEEE Transactions on Reliability 71.4 (2021): 1401-1416. Zhang, Hao, and W. K. Chan. "Apricot: A weight-adaptation approach to fixing deep learning models." 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2019.

Conclusion and Future works

- We present AIREPAIR, a comprehensive platform for repairing neural networks.

- We give the results of five existing neural network repair tools integrated into AIREPAIR.

- In the future, we will support and test more neural network repair methods, propose a unified interface for developers to test and benchmark their repair methods

Thank you!