

CS4290 Project Presentation Group 10: Distributed Adversarial Training with JOINTSPAR

Casper Meijer Gijs Paardekooper Felix Kaubek

Delft University of Technology, The Netherlands

June 19, 2023

Contents

- ① Process
- ② Experiments
- ③ Results
- ④ Conclusion

Project Goals

- ① Reproduce Distributed Adversarial Training (DAT) framework
 - ② Reproduce JOINTSPAR paper
 - ③ Incorporate JOINTSPAR into DAT to improve training time
-
- JOINTSPAR is a novel method to reduce the computation and communication time of gradients in the distributed setting
 - Our hypothesis was that DAT could benefit from this leading to a reduced training time

Process

Challenges Faced

- Codebase of DAT incomplete, lacking many important aspects of the paper
 - No codebase for JOINTSPAR, some steps difficult to implement
 - Limited cloud credits, only one run per \$50 credit coupon possible
 - Maximum of 8 GPUs on Google Cloud
-
- Getting DAT to work with Vertex AI proved challenging and took more time than expected
 - JOINTSPAR was a more straight forward but required lots of local tests to verify its behaviour

Local Experiments

- Ran on local workstation
 - GPU: NVIDIA RTX 4090
 - CPU: AMD Ryzen 7950x
 - RAM: 64GB 4800-DDR5

Experiments

- cifarext: CIFAR-10 + 500K unlabeled Tiny Images
- Baseline cifarext without AT
- JOINTSPAR cifarext
 - Sparsity budget
 - p_{min} values

Cloud Experiments

- Ran on Google Cloud using VertexAI platform
 - 8x n1-normal-4 with 1 NVIDIA P100 each

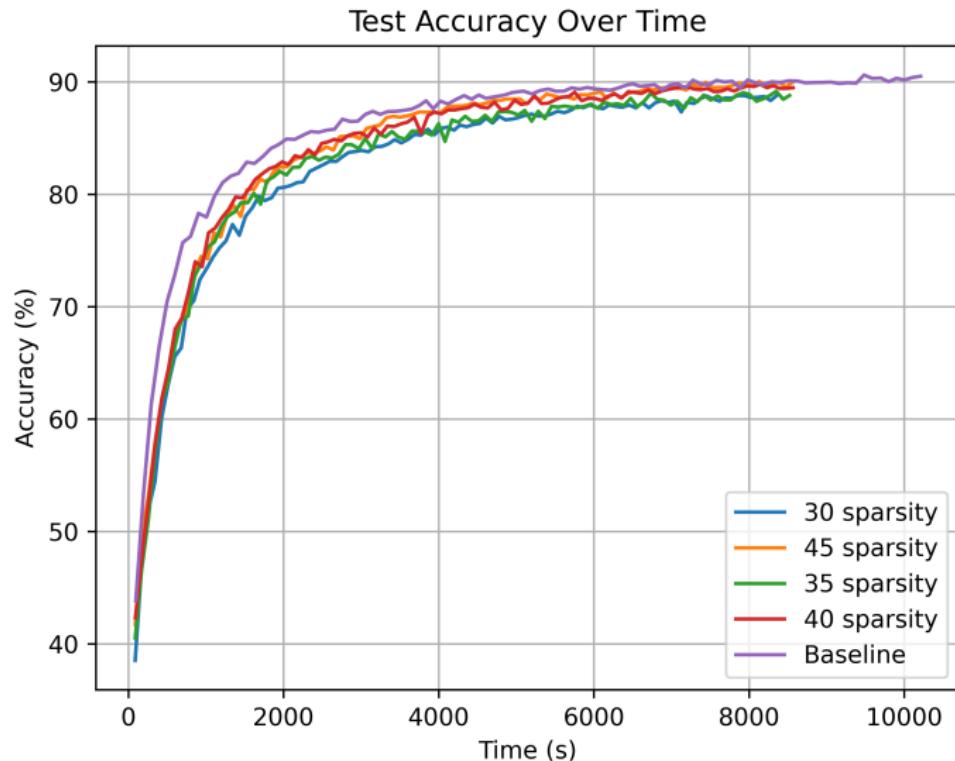
Experiments

- DAT: 4 Nodes vs 8 Nodes
- No DAT: Baseline vs JOINTSPAR
- DAT: Baseline vs JOINTSPAR

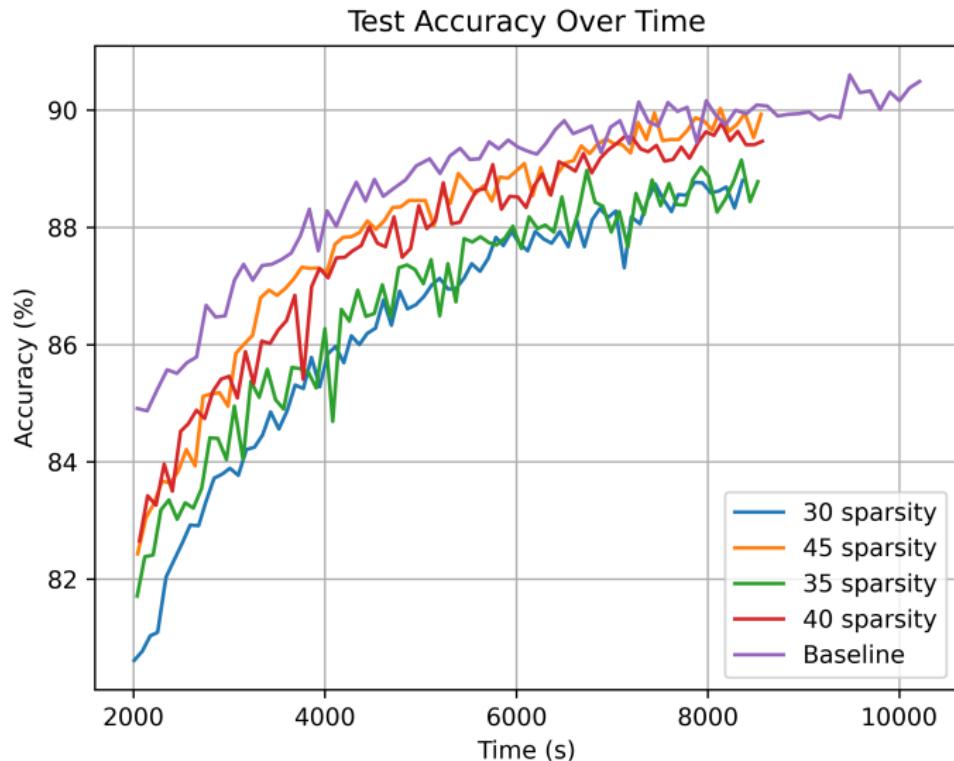
Results

- ① Local no JOINTSPAR vs JOINTSPAR with different sparsity budgets
- ② Local no JOINTSPAR vs JOINTSPAR with different p_{min} values
- ③ VertexAI: DAT 4 Machines vs 8 Machines
- ④ VertexAI: No DAT Baseline vs JOINTSPAR
- ⑤ VertexAI: DAT vs DAT with JOINTSPAR

Results Local JOINTSPAR Sparsity Budgets



Results Local JOINTSPAR Sparsity Budgets



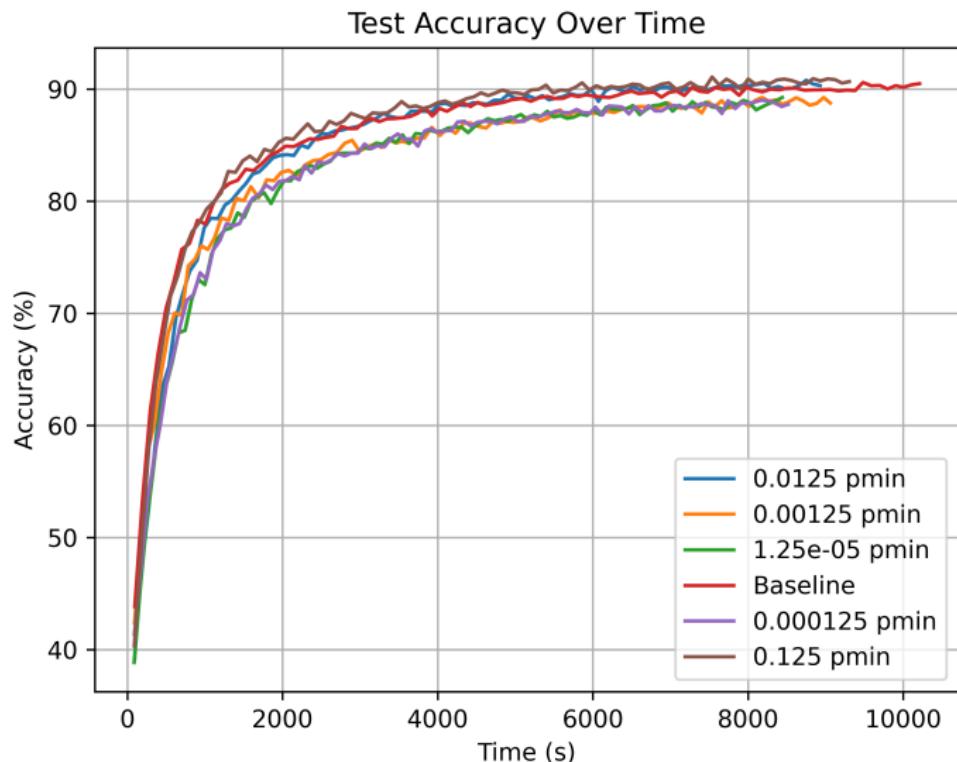
Results Local JOINTSPAR Sparsity Budgets

- Average run time of the sparsity budget runs: 8503s
- Total run time baseline: 10213 s
- Time saving factor: 1.2x

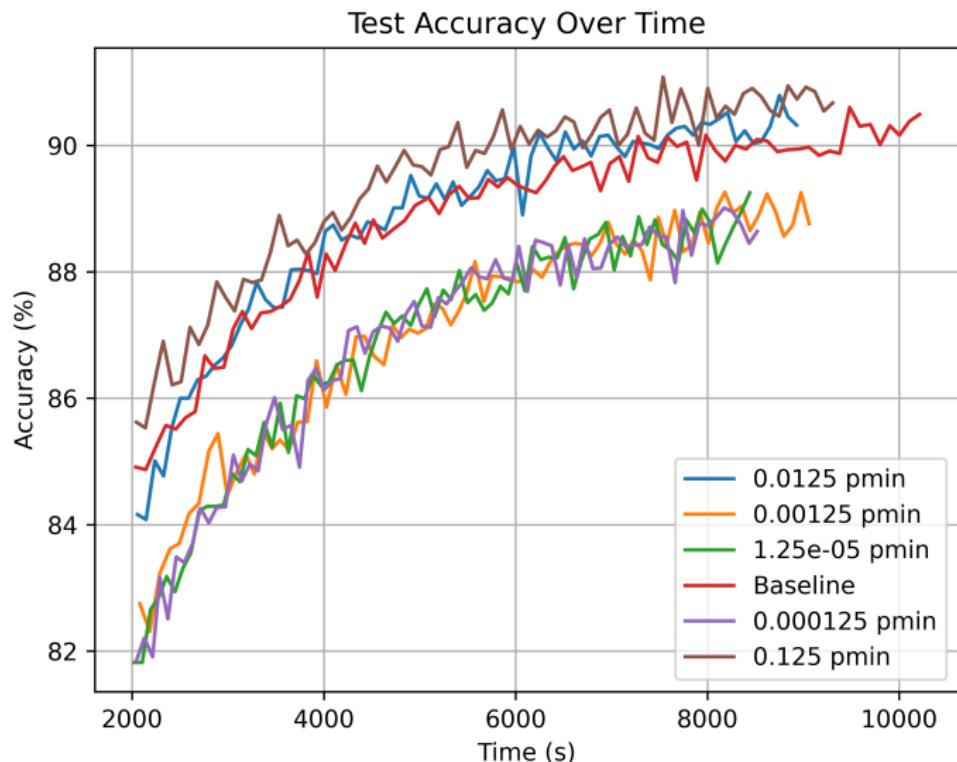
	FullPrecision	TernGrad	QSGD	ATOMO	Avg over baselines
CIFAR10	1.9x	1.4x	1.5x	1.6x	1.6x
CIFAR100	1.4x	1.3x	1.4x	1.2x	1.3x

- JOINTSPAR paper achieves a speedup of at least 1.4x on CIFAR10

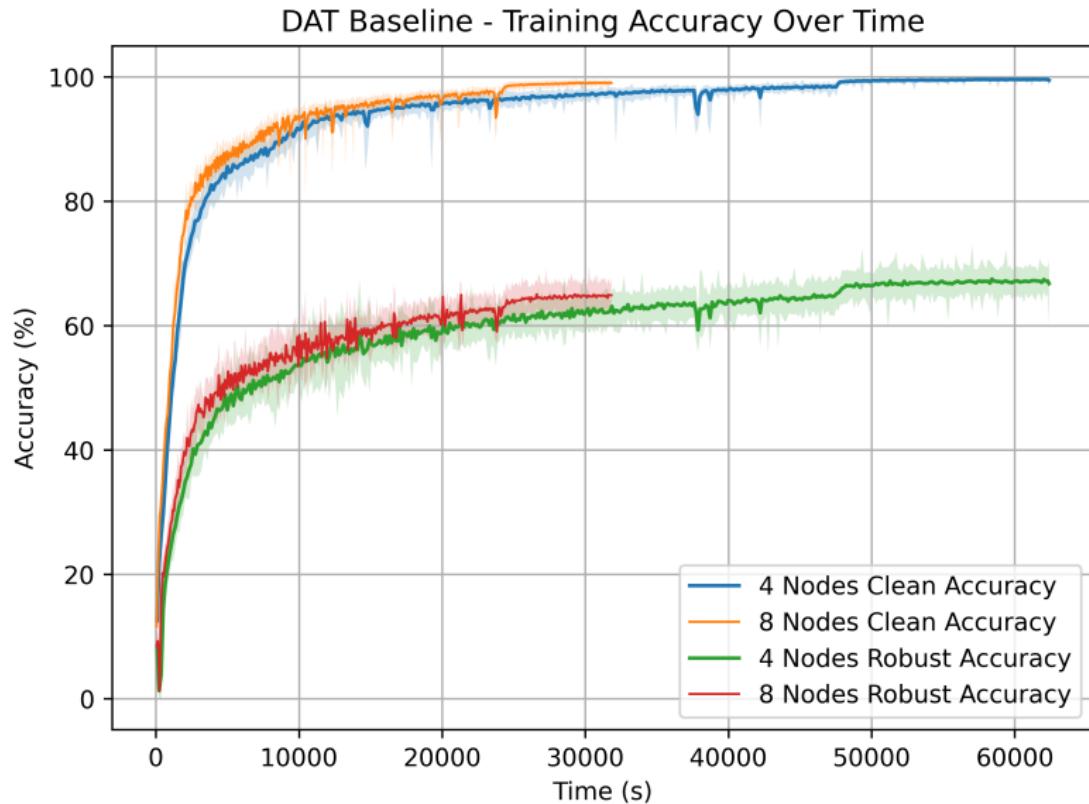
Results Local JOINTSPAR p_{min} Values



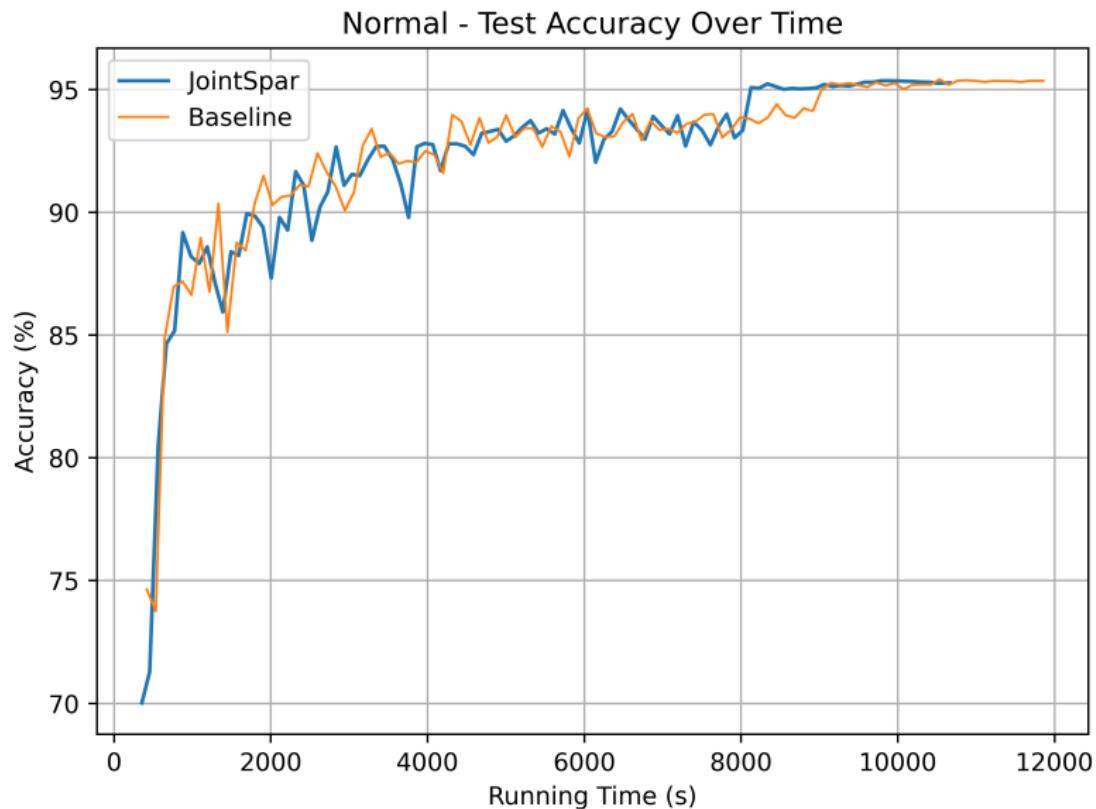
Results Local JOINTSPAR p_{min} Values



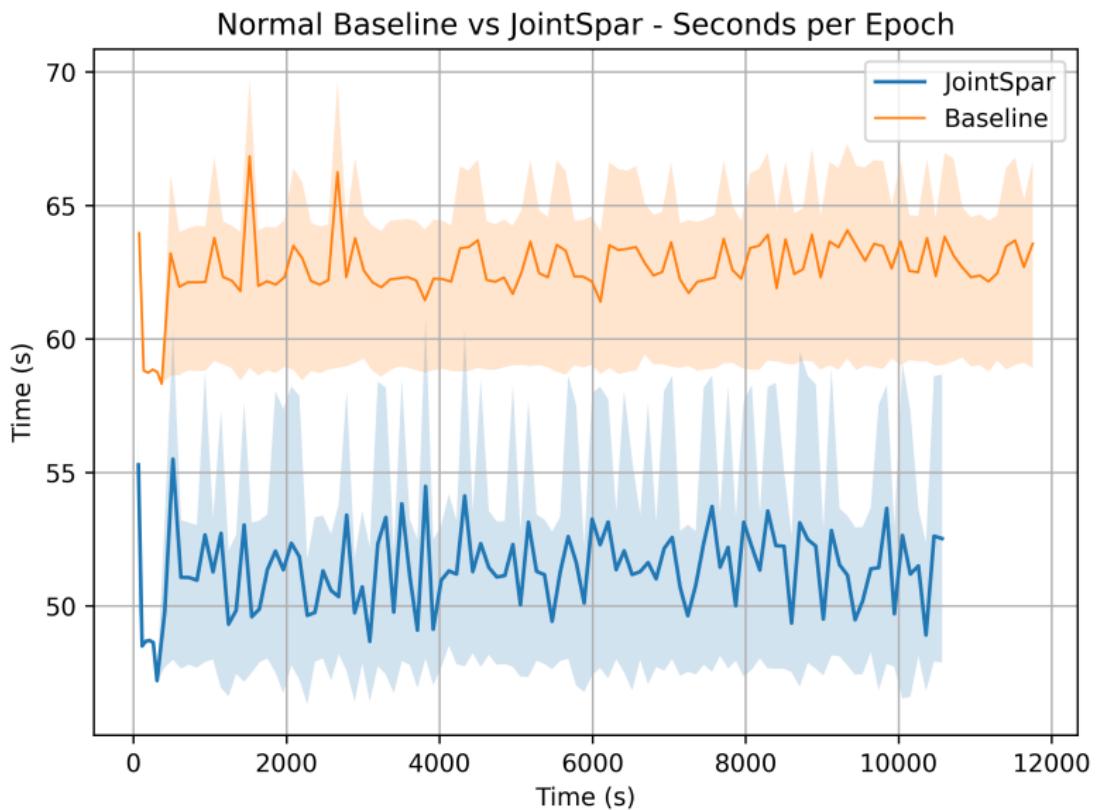
Results VertexAI DAT 4 Machines vs 8 Machines



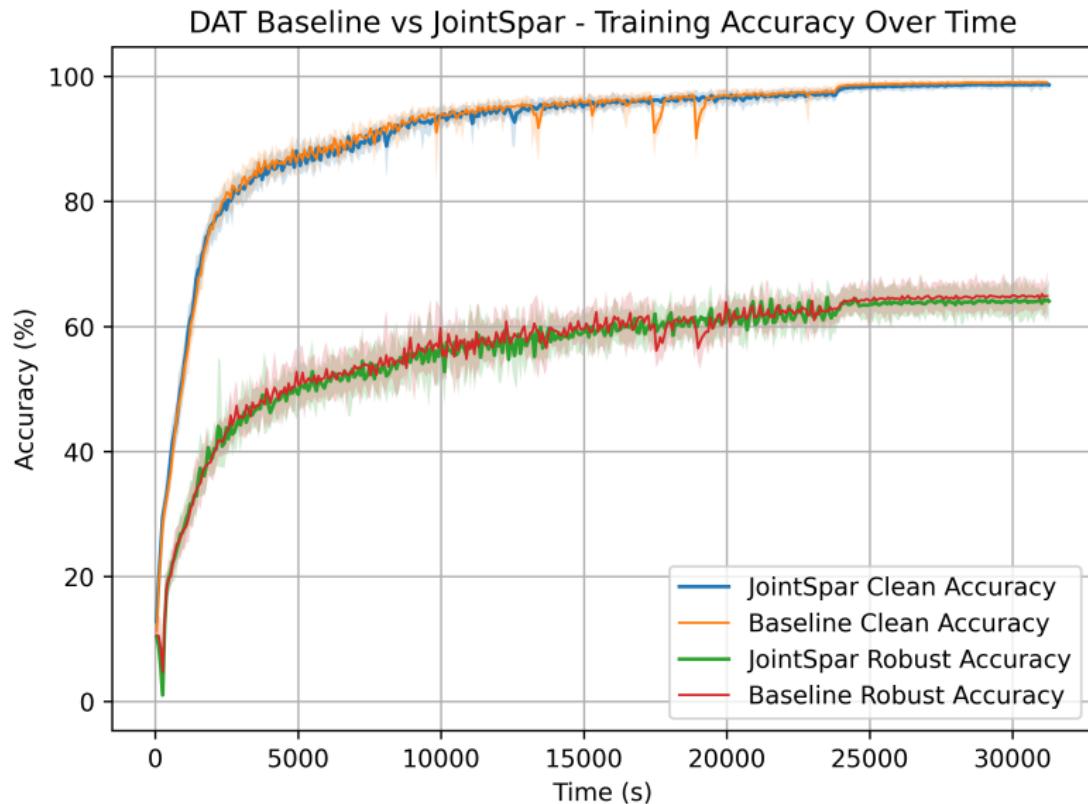
Results VertexAI Baseline vs JOINTSPAR



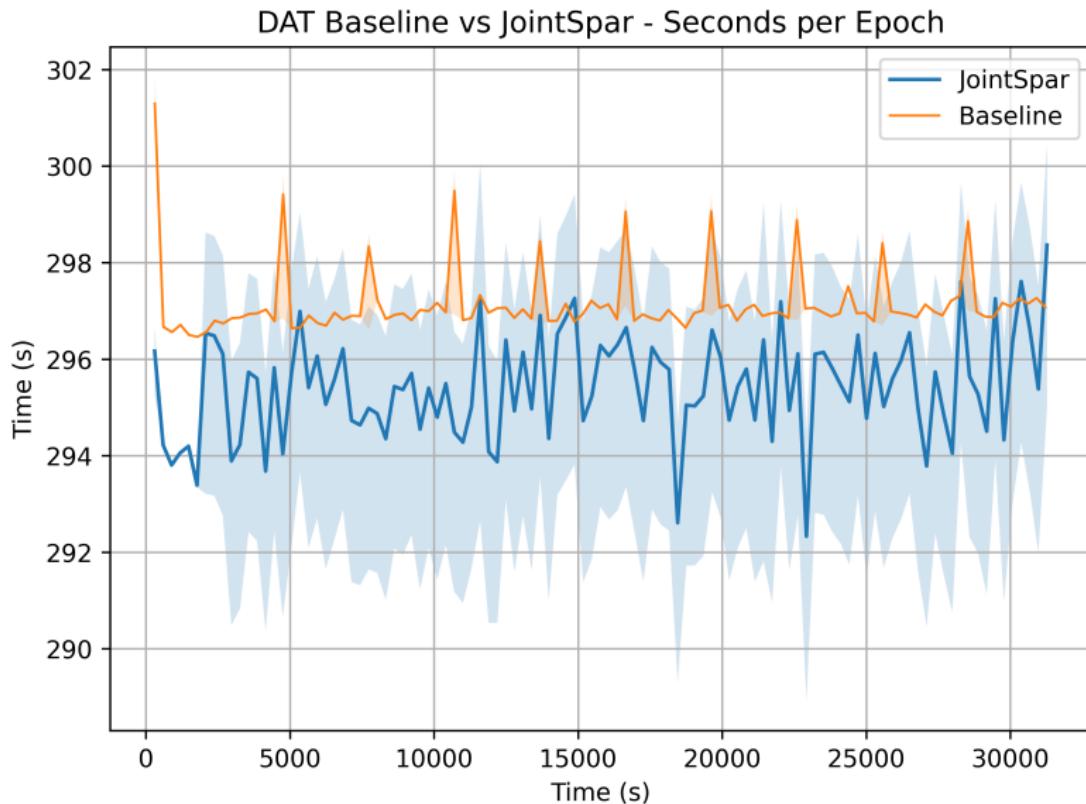
Results VertexAI Baseline vs JOINTSPAR



Results VertexAI DAT vs DAT JOINTSPAR



Results VertexAI DAT vs DAT JOINTSPAR



Conclusion

- Successfully reproduced and implemented:
 - Distributed Adversarial Training (DAT) framework
 - JOINTSPAR algorithm
- JOINTSPAR gave a significant speedup on normal training
- JOINTSPAR gives no significant training time decrease for DAT
 - Biggest portion of DAT training time spend on the attack step
- JOINTSPAR paper uses 2x as many nodes and 8x as many GPUs
 - More nodes and more GPUs lead to a more noticeable speedup