

Data-Efficient Double-Win Lottery Tickets from Robust Pre-training

[ICML 2022] Tianlong Chen¹, Zhenyu Zhang¹, Sijia Liu^{2,3}, Yang Zhang³, Shiyu Chang⁴, Zhangyang Wang¹

¹University of Texas at Austin, ²Michigan State University,

³MIT-IBM Watson AI Lab, ⁴University of California, Santa Barbara



The University of Texas at Austin
Electrical and Computer Engineering



IBM | **MIT**



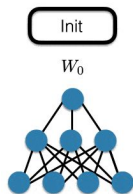
Agenda

- Backgrounds
- What is Double-Win Lottery Tickets?
- Findings
- Drawing Double-Win Lottery Tickets from Robust Pre-training
- Double-Win Tickets with Robust Pre-training Enables
Data-Efficient Transfer

Backgrounds

The Lottery Ticket Hypothesis. *A randomly-initialized, dense neural network contains a subnetwork that is initialized such that—when trained in isolation—it can match the test accuracy of the original network after training for at most the same number of iterations.*

Iterative Magnitude Pruning with Rewinding



Frankle et al., 2019
Viz: @RobertTLange

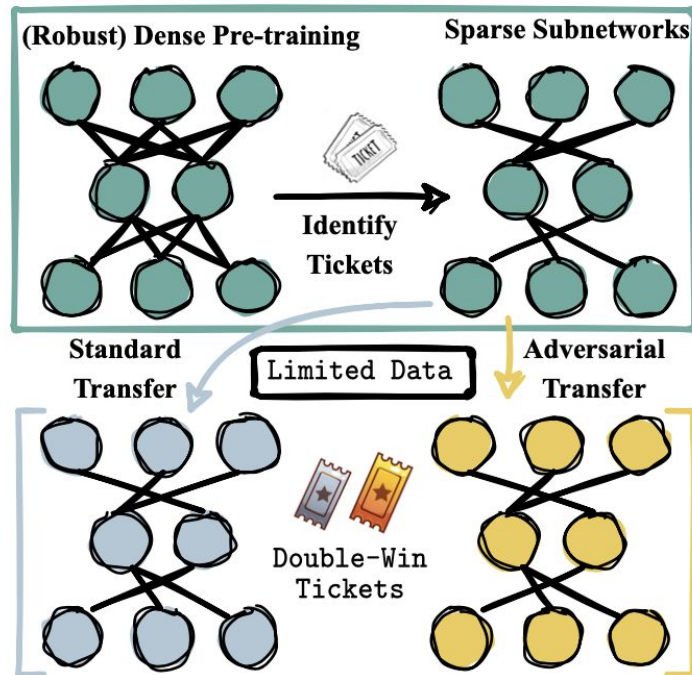
THE LOTTERY TICKET HYPOTHESIS: FINDING SPARSE, TRAINABLE NEURAL NETWORKS

Jonathan Frankle
MIT CSAIL
jfrankle@csail.mit.edu

Michael Carbin
MIT CSAIL
mcarbin@csail.mit.edu

What is Double-Win Lottery Tickets?

A located subnetwork from a pre-trained model can be independently transferred on diverse downstream tasks, to reach **BOTH** the same standard and robust generalization, under **BOTH** standard and adversarial training regimes.



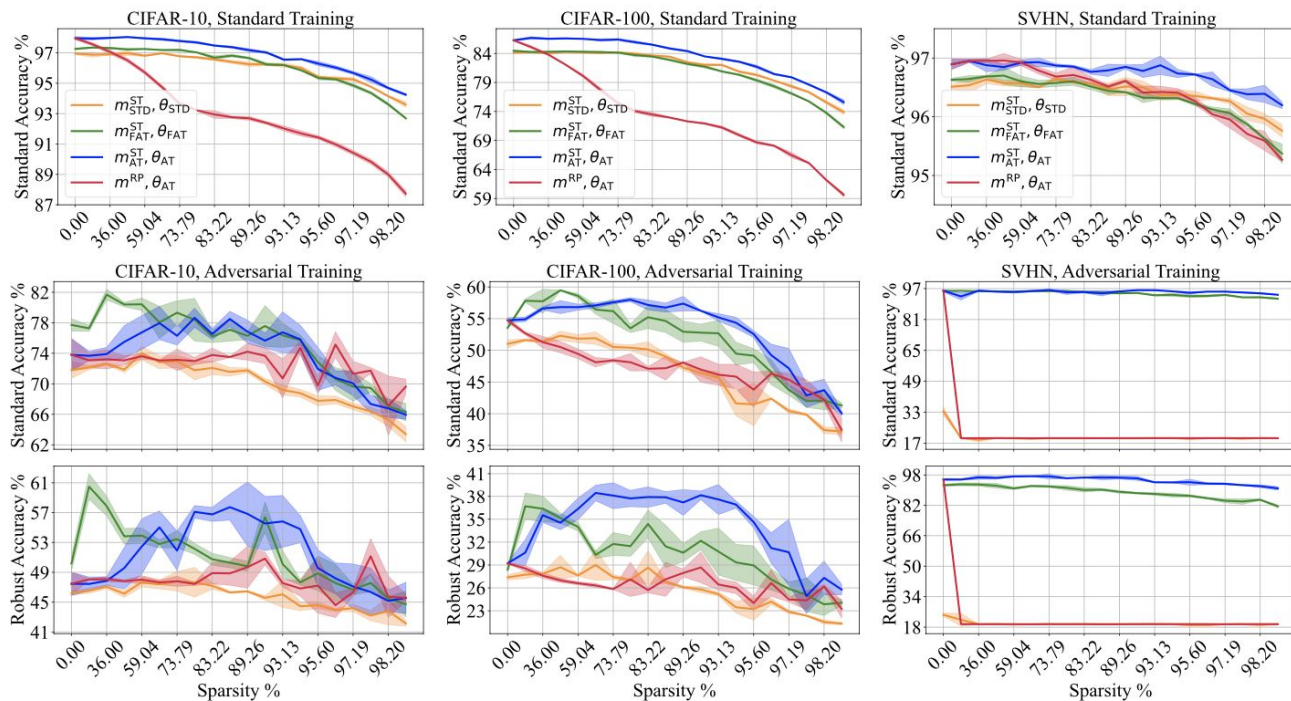


Findings

- Using IMP, we find double-win tickets broadly across diverse downstream datasets and at non-trivial sparsity levels 79.03% ~ 89.26% and 83.22% ~ 96.48% sparsity, using the fast adversarial (FAT) and adversarial (AT) pre-training. In general, subnetworks located from the AT pre-trained model have superior performance than FAT and standard pre-training.
- We further demonstrate the intriguing property of double-win tickets in the data-limited transfer settings (e.g., 10%, 1%). In this specific situation, FAT can surprisingly find higher-quality subnetworks with small sparsity while AT overtakes in a larger sparsity range.
- We show that adopting standard or adversarial training in the process of IMP makes no significant difference for the transferability of identified subnetworks on downstream tasks.

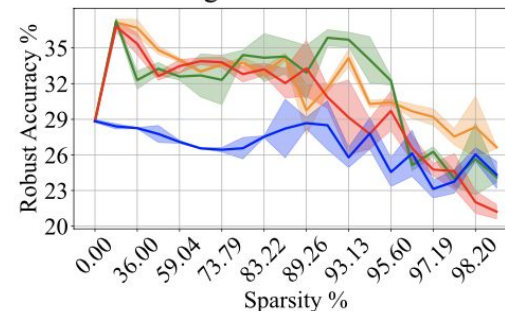
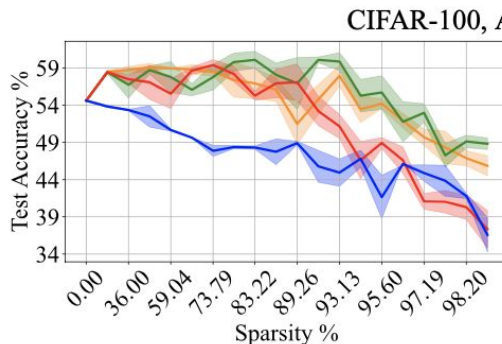
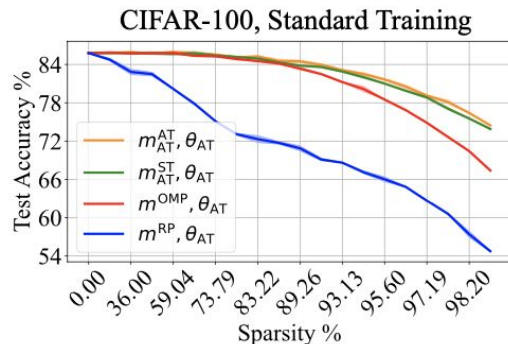
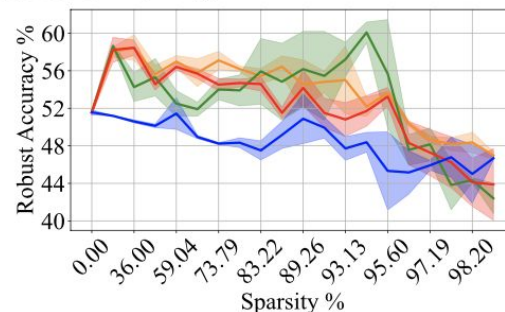
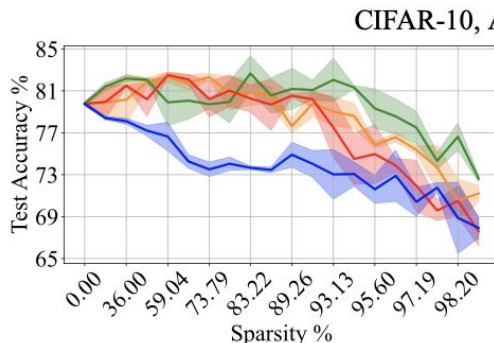
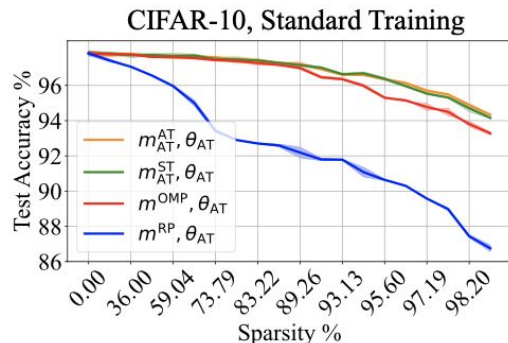
Drawing Double-Win Lottery Tickets from Robust Pre-training

[Q1] Do double-win lottery tickets exist?



Drawing Double-Win Lottery Tickets from Robust Pre-training

[Q2] Do training regimes on source domain affect the located subnetworks?



Double-Win Tickets with Robust Pre-training Enables Data-Efficient Transfer

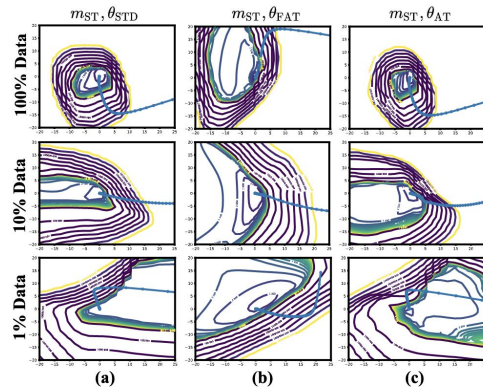
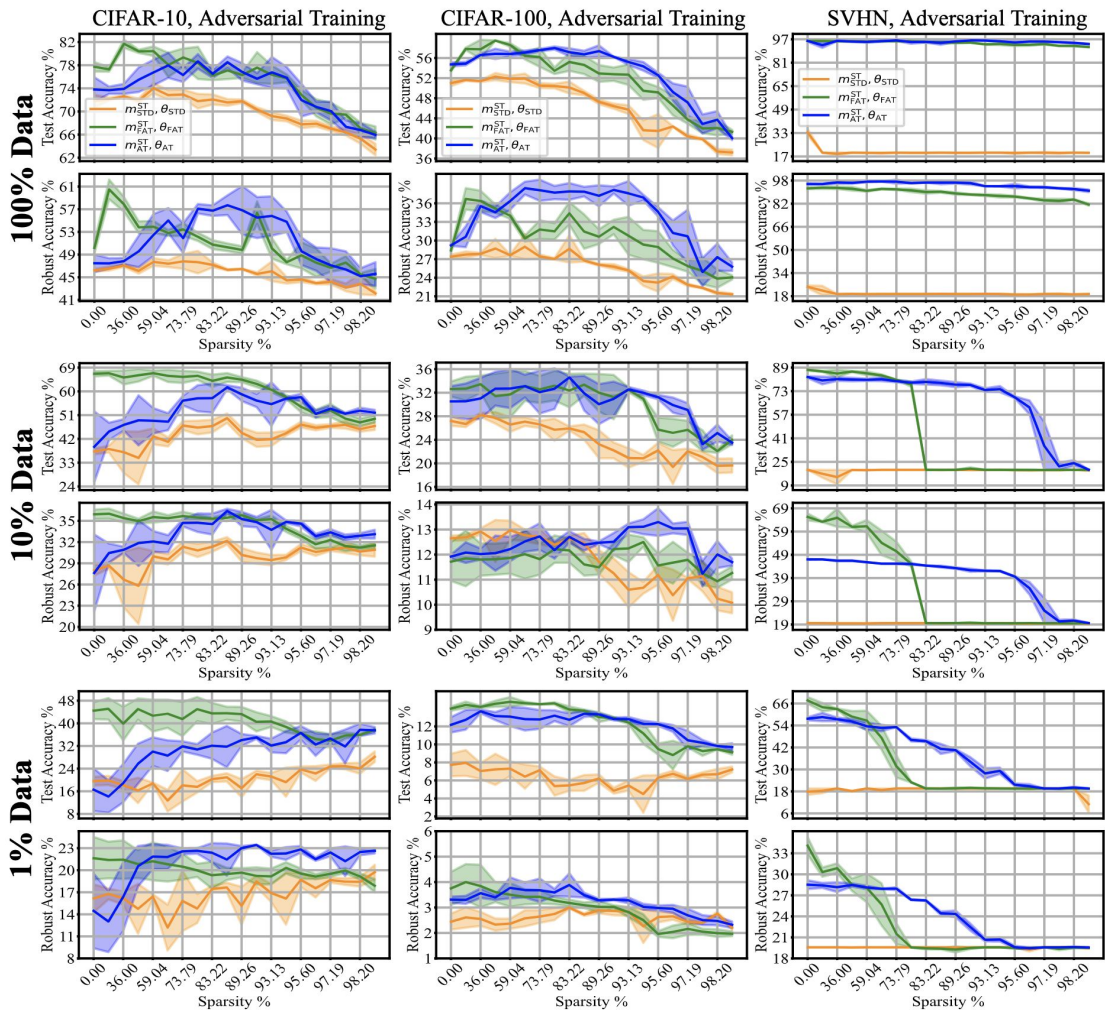


Figure 6. Visualization of loss contours and training trajectories of subnetworks located by IMP with standard re-training m^{ST} at 73.79% sparsity. Each subnetwork is adversarial trained with 100%, 10% or 1% training data on CIFAR-10. We compare three pre-training (i.e., standard θ_{STD} , fast adversarial θ_{FAT} , and adversarial pre-training θ_{AT}). The original test set is used.



Data-Efficient Double-Win Lottery Tickets from Robust Pre-training

Tianlong Chen¹, Zhenyu Zhang¹, Sijia Liu^{2,3}, Yang Zhang³, Shiyu Chang⁴, Zhangyang Wang¹

¹University of Texas at Austin, ²Michigan State University,

³MIT-IBM Watson AI Lab, ⁴University of California, Santa Barbara



➤ What is Double-Win Lottery Tickets?

A located subnetwork from a pre-trained model can be independently transferred on diverse downstream tasks, to reach **BOTH** the same standard and robust generalization, under **BOTH** standard and adversarial training regimes.

➤ Insightful Findings

- ❑ Double-win tickets broadly exist across diverse downstream datasets and at non-trivial sparsity levels 79.03% ~ 89.26% and 83.22% ~ 96.48% sparsity, using the fast adversarial (FAT) and adversarial (AT) pre-training. In general, subnetworks located from the AT pre-trained model have superior performance than FAT and standard pre-training.
- ❑ In the data-limited transfer settings (e.g., 10%, 1%), FAT can surprisingly find higher-quality subnetworks with small sparsity while AT overtakes in a larger sparsity range.
- ❑ Adopting standard or adversarial training in the process of IMP makes no significant difference for the transferability of identified subnetworks on downstream tasks.

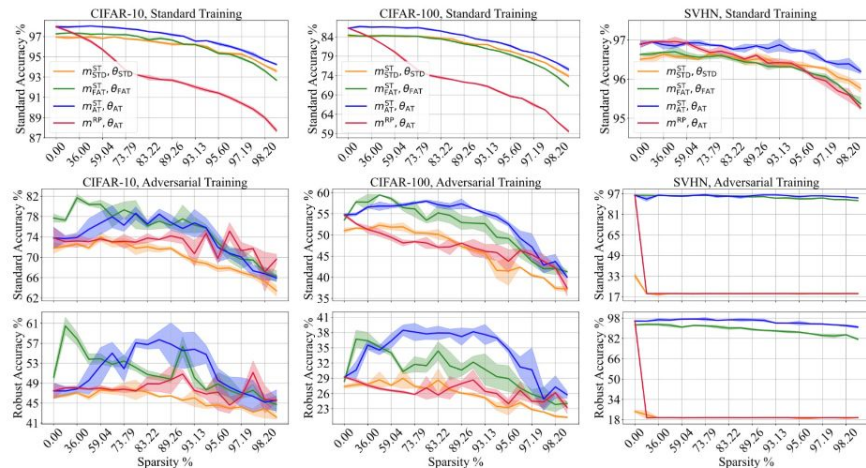
➤ The Existence of Double-Win Lottery Tickets

[Q1] Do Double-Win Lottery Tickets exist?

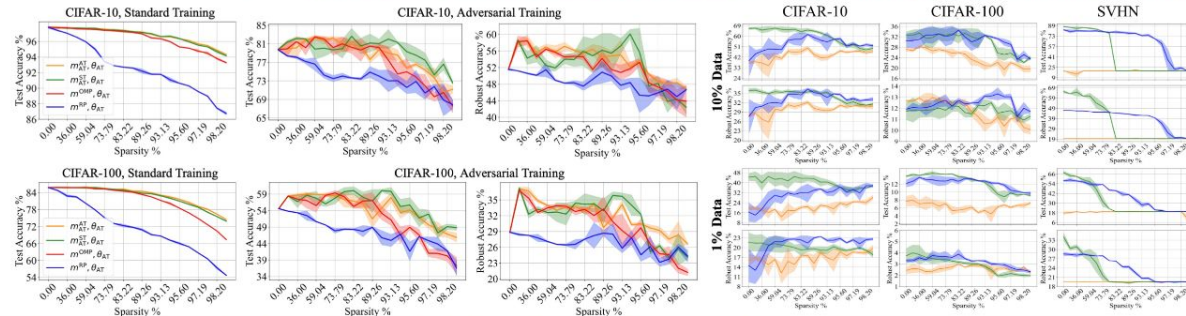
[Q2] Do training regimes on source domain affect the located subnetworks?

➤ References

- ❖ Frankle, J. et al. "The lottery ticket hypothesis: Finding sparse, trainable neural networks".
- ❖ Chen, T. et al. "The lottery ticket hypothesis for pre-trained BERT networks".
- ❖ Chen, T. et al. "The lottery ticket hypothesis for supervised and self-supervised pre-training in computer vision models".



➤ Double-Win Tickets with Robust Pre-training Enables Data-Efficient Transfer



Q&A