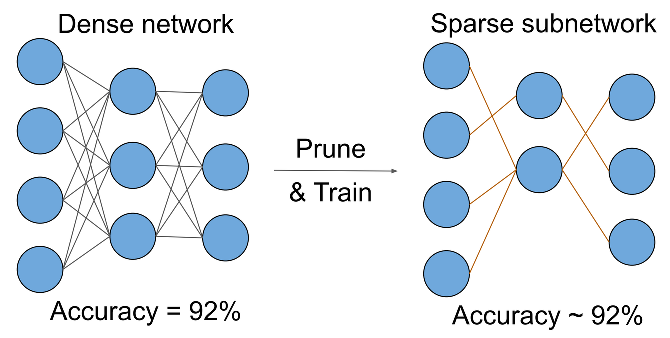
Lottery Ticket Hypothesis for Object Recognition

Sharath Girish*, Shishira R Maiya*, Kamal Gupta, Hao Chen, Larry Davis, Abhinav Shrivastava
University of Maryland, College Park

Introduction

Goal: Analyzing lottery ticket hypothesis in context of object recognition tasks.

Lottery Ticket Hypothesis (LTH): LTH states that dense randomly-initialized neural networks contain sparse subnetworks which can be trained in isolation and can match the test accuracy of the original network.

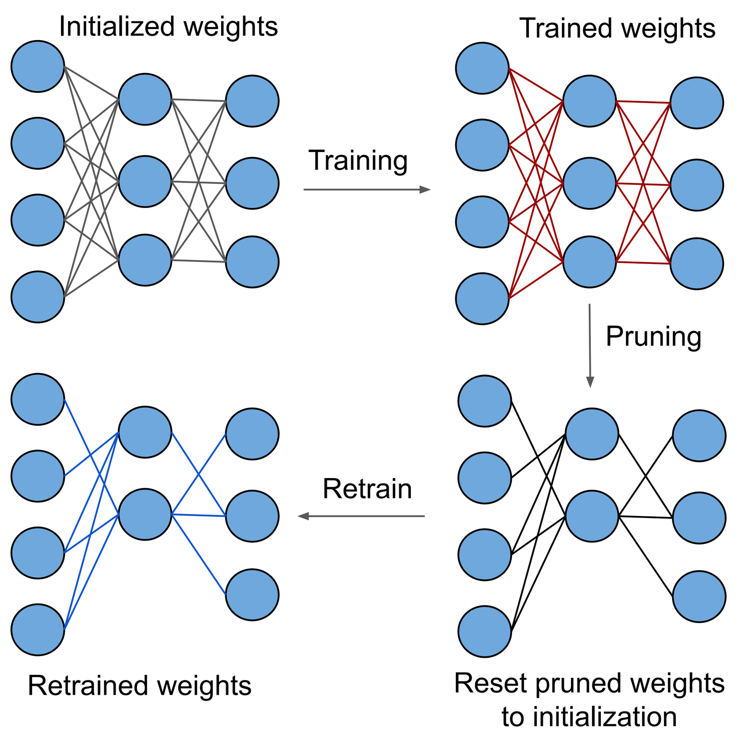


Key Contributions:

- First empirical study on implementing LTH in context of object detection, segmentation and keypoint estimation.
- Show that no "universal tickets" can be transferred to downstream tasks.
- We find winning tickets on various architectures and tasks with up to 80% sparsity.
- We investigate various properties of lottery tickets in the context of object recognition systems.

Algorithm:

Iterative Magnitude Pruning for finding winning tickets

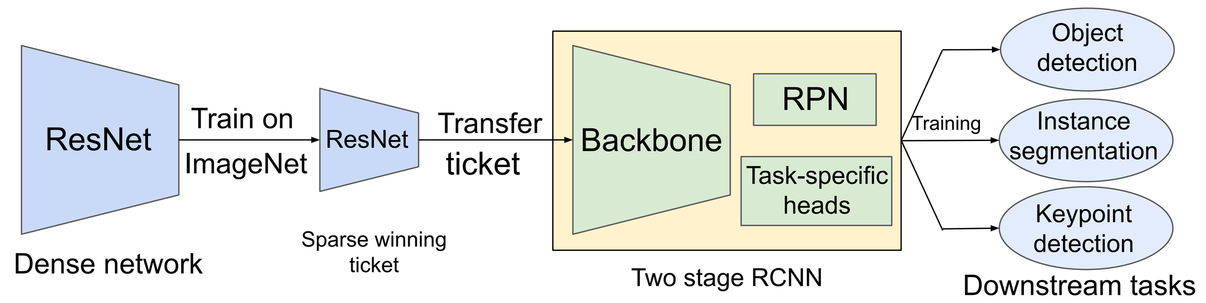


IMP: Iterative Magnitude Pruning for LTH

1. Randomly initialize network f with initial weights w_0, mask $m_0 = 1$, prune target percentage p, and T pruning rounds to achieve it.;
2. **while** $i < T$ **do**
3. Train network for N iterations $f(x; m_i \odot w_0) \rightarrow f(x; m_i \odot w_i)$;
4. Prune bottom $p^{\frac{1}{k}}\%$ of $m_i \odot w_i$ and update m_i;
5. Reset to initial weights w_0 ;
6. $i \leftarrow i + 1$ next round ;
7. **end**

ImageNet Tickets Transfer

Universal tickets: We find winning tickets in standard ImageNet models and transfer them to downstream task models.



Results:

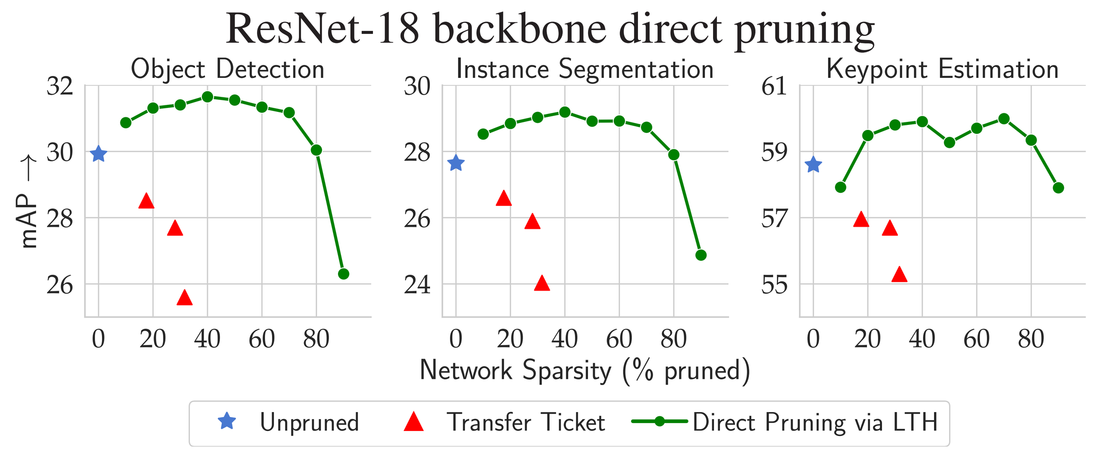
ImageNet tickets transfer: Resnet-18 backbone

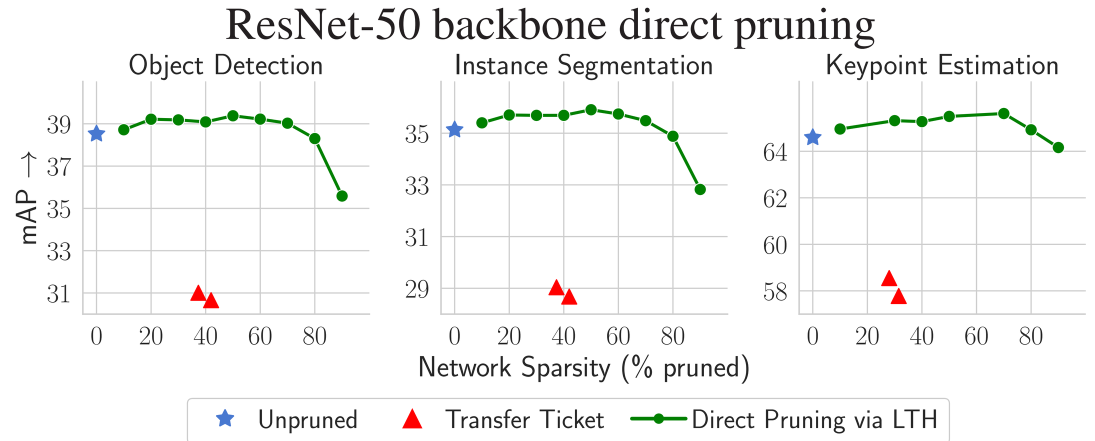
Prune %	COCO Detection		COCO segmentation		COCO Keypoint	
	Sparsity	mAP	Sparsity	mAP	Sparsity	mAP
90%	31.61%	25.59	31.61%	24.03	21.47%	55.30
80%	28.10%	27.70	28.10%	25.90	19.09%	56.70
50%	17.57%	28.52	17.57%	26.60	11.94%	56.96
0%	0%	29.91	0%	27.64	0%	58.59

- Tickets obtained from Imagenet trained backbone networks do not tranfer well to downstream tasks.
- The overall sparsity of these networks are also quite low as RPN and FC heads are not pruned.

Direct Pruning

We directly apply LTH to find task-specific tickets for 3 different tasks, using 2 backbone networks.



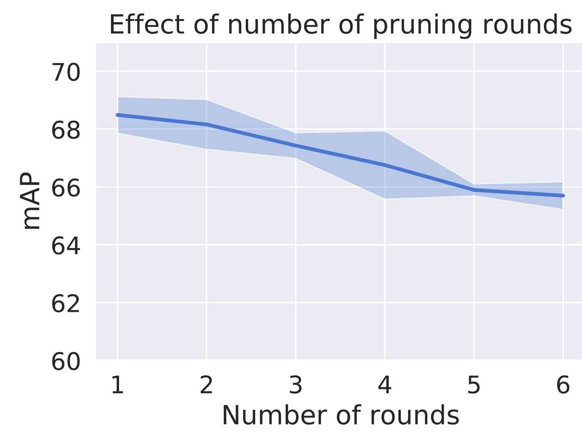


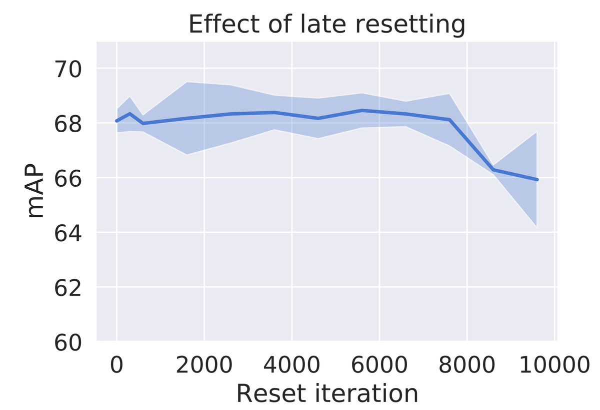
Results:

- Winning tickets with 80% sparsity for almost all tasks.
- Outperforms transfer tickets and dense baselines.

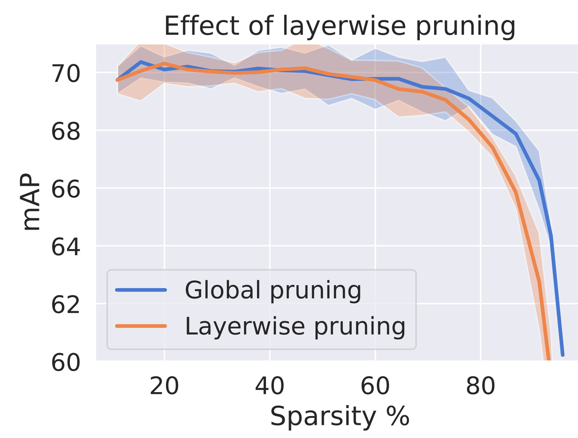
Training Lottery Tickets

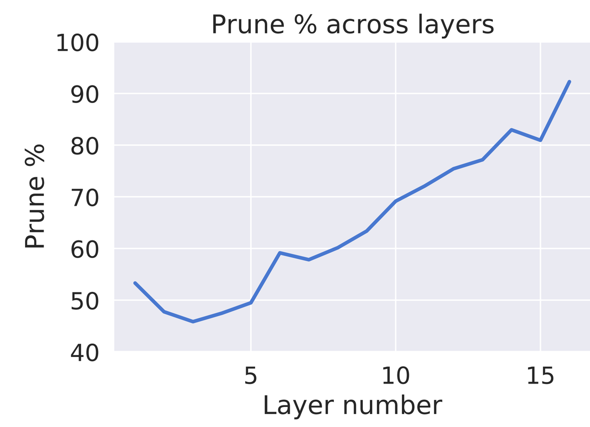
We provide insights into training lottery tickets for the obejct detection task. All the experiments in this section are performed by training a Faster-RCNN with ResNet-18 backbone on the VOC dataset.



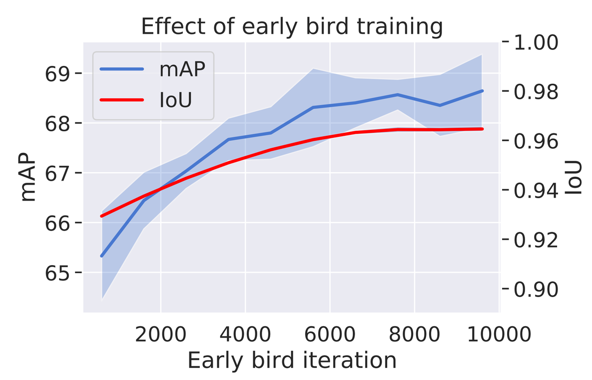


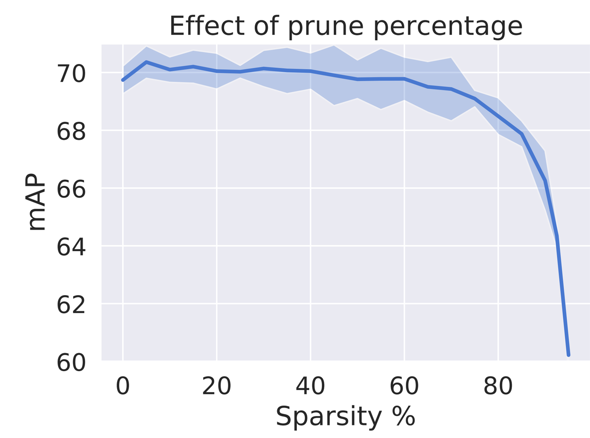
- Performance decrease with more pruning.
- This is in contrast with LTH for classification.
- Late resetting has little to no effect on mAP.
- Likely due to non-random ImageNet initialization.





- Global pruning does better than layerwise pruning for higher levels of sparsity.
- In layerwise pruning, deeper layers are pruned at a larger rate, hurting the performance.



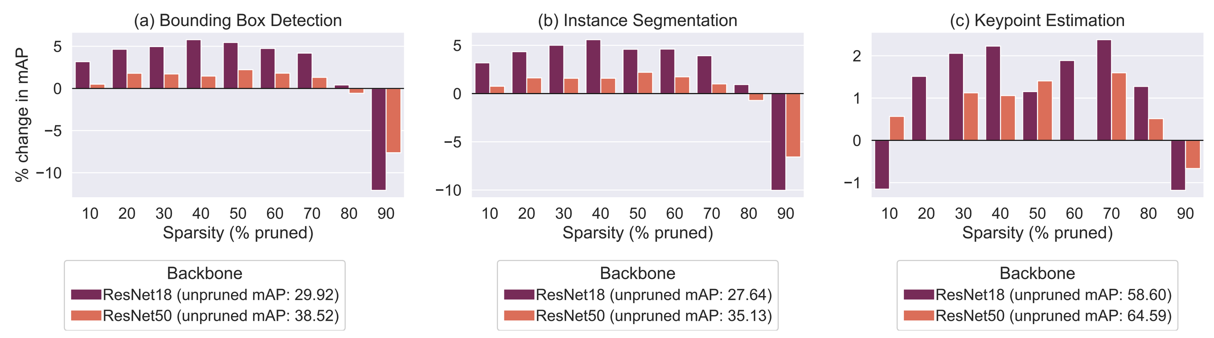


- We find pruning masks within 50% of the training iterations which yield little drop in performance.
- We observe that the performance of the model steeply decreases after the 80% sparsity mark.

Properties of Winning Tickets

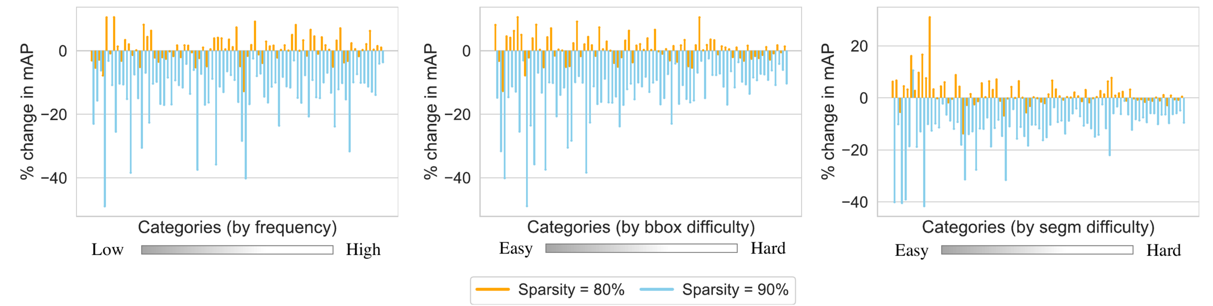
We perform experiments to rigorously analyze the behaviour of winning lottery tickets.

Effect of backbone architecture:



The ResNet-18 network with fewer number of parameters shows higher gains or drops in mAP depending upon the sparsity level.

Do tickets behave differently on easy *vs.* hard categories?



As we increase sparsity beyond the 80% threshold, we observe that the smaller and easier (easy bounding boxes) objects are worst hit, registering a steep performance decline across all tasks.

Do winning tickets transfer across downstream tasks?

Target task	Source task	Network sparsity	mAP	AP50
Det	Det/Seg	78.4%	30.04	49.40
	Keypoint	50.11%	23.94	41.08
Seg	Det/Seg	78.4%	27.90	46.68
	Keypoint	50.11%	23.02	39.01
Keypoint	Det/Seg	76.98%	58.31	81.53
	Keypoint	79.4%	59.34	82.36

We observe that detection/segmentation tickets transfer fairly well to the keypoint estimation task. However, the reverse is not true because keypoint only uses the 'human' class which does not transfer to the full COCO dataset.

Conclusion

- We provide rigorous empirical analysis of LTH for multiple object recognition tasks.
- We observe that ImageNet tickets don't transfer to downstream tasks even at lower levels of sparsity.
- With direct pruning of the entire network, we find winning tickets with upto 80% sparsity for different network architectures and tasks.
- Typical LTH methods for classification, such as late resetting, iterative pruning, do not necessarily work for object recognition tasks.
- Please visit project website for more information: `http://lth-recognition.github.io/`