One-shot Network Pruning at Initialization with Discriminative Image Patches

Yinan YANG, Yu WANG, Ying JI, Heng QI, Jien KATO

Ritsumeikan Univ., Hitotsubashi Univ., Nagoya Univ., Dalian Univ. of Technology

Introduction

- One-shot Network Pruning at Initialization (OPaI) is an effective method to decrease network pruning costs. Recently, there is a growing belief that data is unnecessary in OPaI.
- However, extensive experiments reveal that OPaI is data-dependent in two representative OPaI methods
- We propose two novel methods, *Discriminative One-shot Network Pruning (DOP)* and *Super Stitching*, to prune the network by high-level visual discriminative image patches.

Related Work

- **One-cut Network Pruning at Initialization (OPaI)**
  - Samples a randomize mini-batch in training data.
  - Mask unimportant parameters with 0, fine-tuning pruned network.
- **SNIP** computes the connection sensitivity

$$s(\theta_j) = \left| \frac{\partial L(T^b; \theta_j \odot m_j)}{\partial m} \right|_{m=1} \left| = \right| \frac{\partial L(T^b; \theta_j)}{\partial \theta_j} \odot \theta_j \Bigg| . \qquad (1)$$

- **GraSP** uses the Hessian H to preserve the gradient flow

$$s(\theta_j) = -\theta_j (H \frac{\partial L(T^b; \theta_j)}{\partial \theta_j})_j . \qquad (2)$$

- **Is OPaI data-independent?**
  - A much-debated question is whether data-independent in OPaI. Recent research demonstrate in sanity check approaches, that the data using in pruning step is unimportant.

Proposed Method

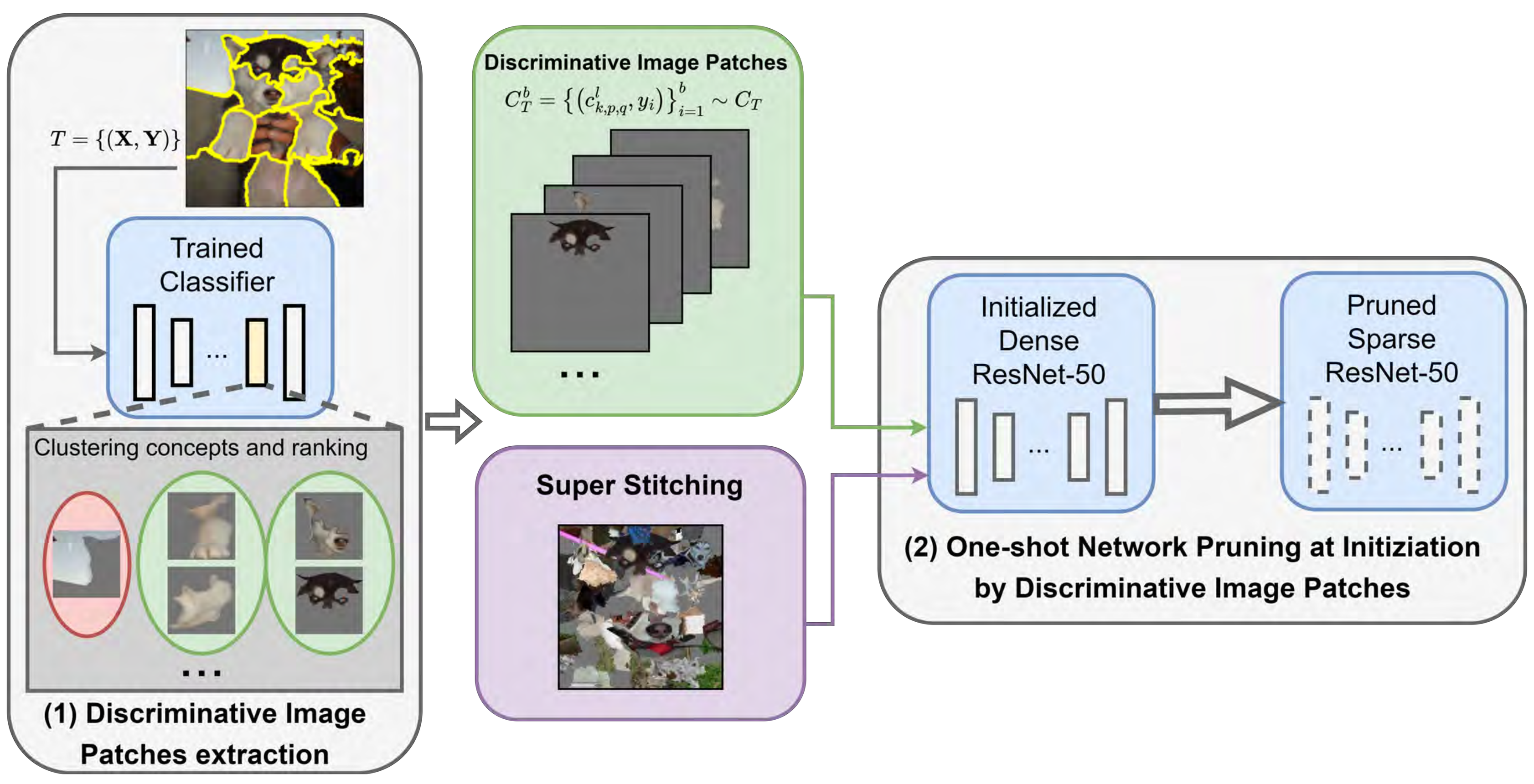

Figure: Overview of the Discriminative One-shot Network Pruning (DOP) and Super Stitching. (1) Cluster segments in trained network's activation space, extract Discriminative Image Patches. The green is meaningful in network prediction, and the red is meaningless. (2) Using Discriminative Image Patches or Super Stitching to prune unimportant parameters by an specific OPaI algorithm.

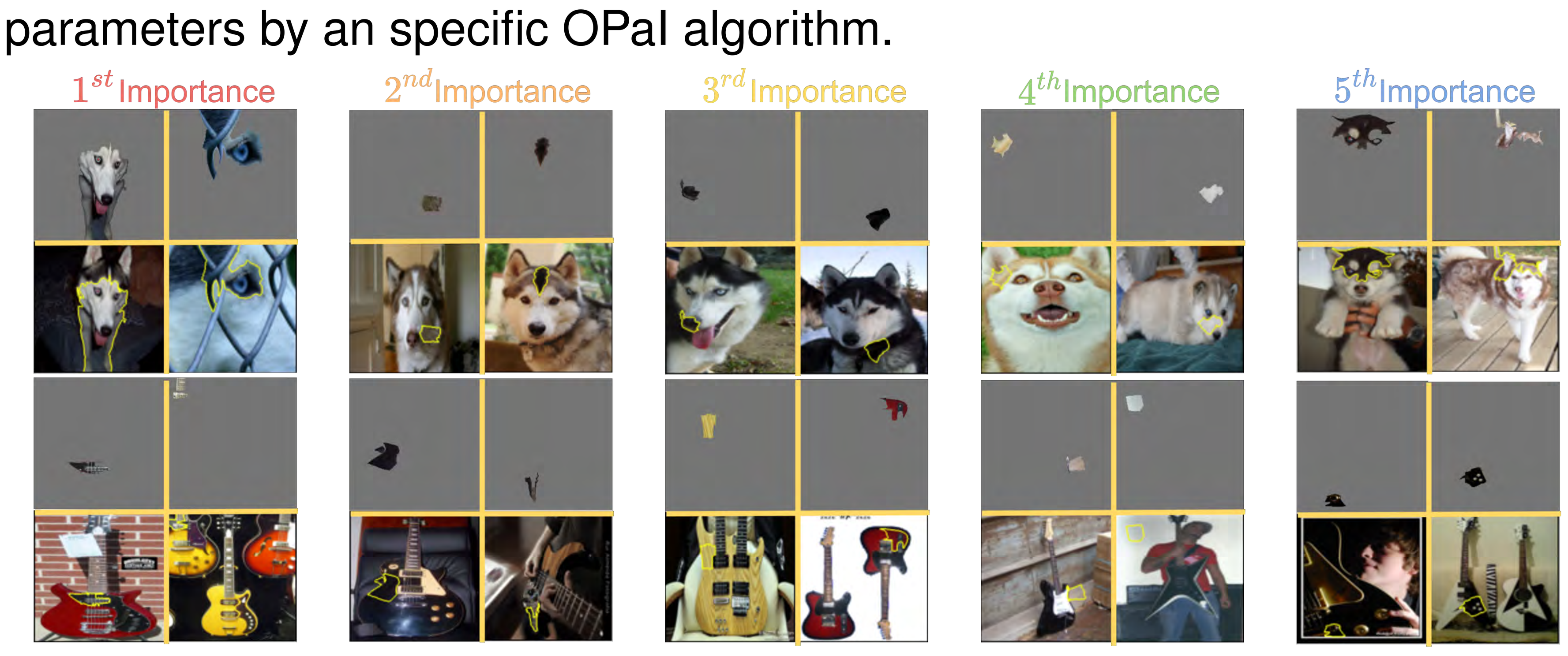

Figure: Concept maps extracted by pre-trained ResNet-50 from the ImageNet. Here we show are Siberian husky and electric guitar with two examples in each class's Top-5 TCAV important concept clusters.

Figure: Super Stitching. From left to right: zebra, goldfish, Siberian husky, ambulance, cash machine, and window screen.

Experiments and Results

- Pruning ResNet-50 with Varying Levels of Sparsity with Discriminative Image Patches.

Sparsity percentage	60%	80%	90%	95%
(Baseline)	76.47%			
SNIP	74%	70.94%	61.06%	36.43%
SNIP with DOP (Ours)	**74.29%**	**71.15%**	**64.12%**	**48.14%**
GraSP	73.87%	71.14%	67.07%	**61.76%**
GraSP with DOP (Ours)	**74.19%**	**71.76%**	**67.65%**	60.02%

Table: DOP: Top-1 Test Accuracy of ResNet-50 on ImageNet.

- We design three ablation experiments to investigate the function of data in OPaI by gradually changing the content of input data.

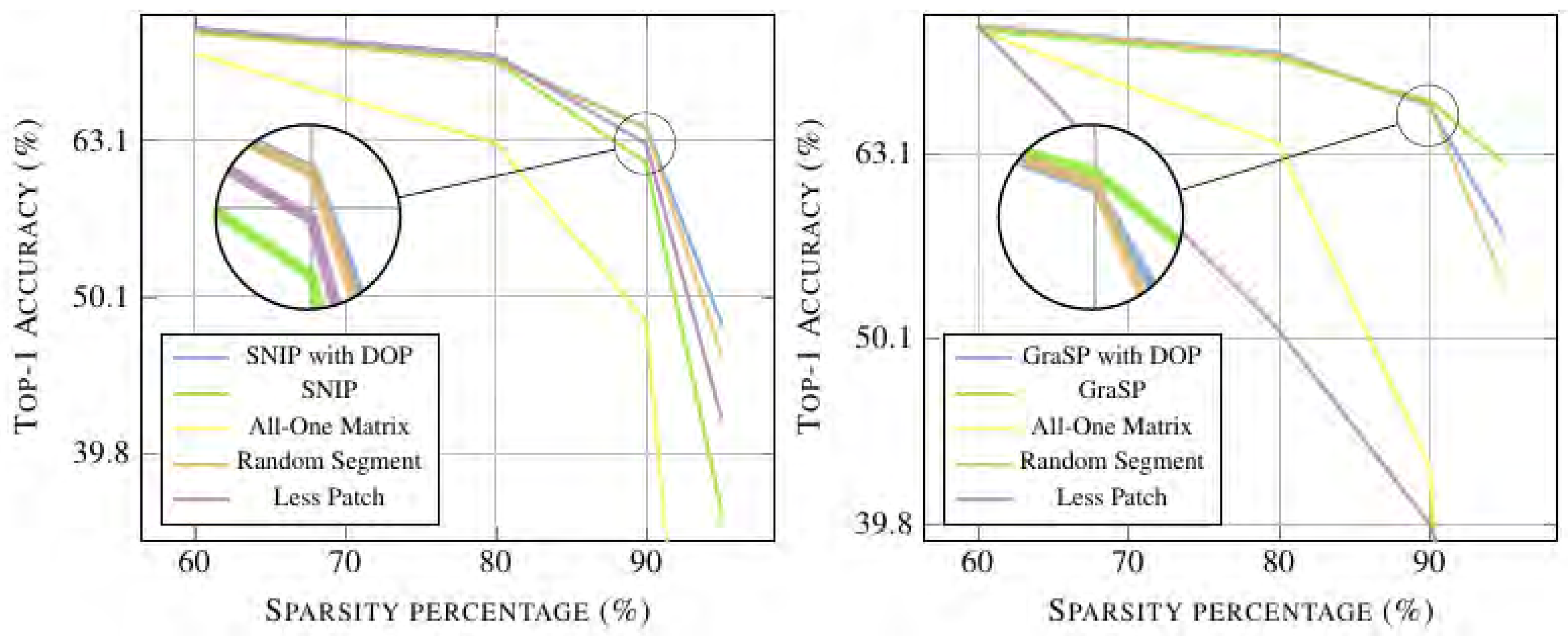

Figure: Ablations at DOP Left: SNIP with DOP. **Right:** GraSP with DOP. (1) All-One Matrix: replace images with all-one matrix. (2) Random Segment: use the same super-pixel algorithm segment images, and prune the network with those segments. (3) Less Concept: use less concept to prune. If the OPaI is data-independent, then the same results should be obtained in the ablation experiments. However, we still observe disparity in ablation experiments.

- We create Super Stitching to further improve the discriminative image patches in OPaI. We aim to use fewer but more informative samples to enhance the gradient flow.

Method	Material Number	Material Type	Sparsity percentage 60%	80%	90%	95%
GraSP	30,000	image	73.87%	71.14%	67.07%	61.76%
GraSP with DOP (Ours)	30,000	discriminative patch	**74.19%**	**71.76%**	67.65%	60.02%
GraSP with Super Stitching (Ours)	30,000	stitching patch $\sigma = 0.5$	74.14%	71.57%	67.59%	**62.70%**
GraSP with Super Stitching (Ours)	10,000	stitching patch $\sigma = 0.75$	73.94%	71.65%	**68.02%**	62.06%
GraSP (Wang et al.)	37,500	image	74.02%	72.06%	68.14%	-
GraSP (Jorge et al.)	614,440	image	-	-	65.4%	46.2%
GraSP (Frankle et al.)	10,000	image	73.4%	71.0%	67%	-
GraSP (Hayouet al.)	-	image	-	-	66.41%	62.1%
FORCE (Jorge et al.)	614,440	image	-	-	64.9%	59.0%
Iter SNIP (Jorgeet al.)	614,440	image	-	-	63.7%	54.7%
SynFlow (Hayou et al.)	-	all-one matrix	-	-	66.2%	62.05%
SBP-SR (Hayou et al.)	-	image	-	-	67.02%	62.66%
ProsPr (Alizadeh et al.)	-	image	-	-	66.86%	59.62%

Table: DOP and Super Stitching test accuracy of ResNet-50 on ImageNet based on GraSP pruning criterion. Above: results under fair conditions. Blow: results in corresponding paper. At high sparsity, we can observe that Super Stitching has a significant advantage. Also, DOP has advantages at low sparsity. We notice that SBP-SR claims a higher performance for ResNet-50 on ImageNet recently, which is the SOTA OPaI method. It is noticed that DOP and Super Stitching achieved their best results at 60% and 95% sparsity, respectively, outperforming the original SOTA.

Conclusion

- Our research reveals that informative data is helpful in OPaI, and provides a new route for OPaI advancement.
- Our novel proposed methods DOP and Super Stitching can significantly improve pruning performance.
- Our work refreshes our typical views of the OPaI methods.

Acknowledgments

This work is supported by the Initiative for Realizing Diversity in the Research Environment (Advanced Type).