A Realistic Evaluation of Semi-Supervised Learning for Fine-Grained Classification

University of Massachusetts Amherst

Jong-Chyi Su Zezhou Cheng Subhransu Maji

University of Massachusetts, Amherst

CVPR VIRTUAL JUNE 19-25

Motivation and Contributions

Existing semi-supervised benchmarks are lacking[1]:

- Curated datasets: CIFAR, SVHN, STL-10, ImageNet
- Uniform class distribution
- Low-resolution images
- Unlabeled data does not contain novel class

Does semi-supervised learning (SSL) work in realistic datasets?

A Realistic Benchmark

Semi-Aves Dataset @ FGVC7

Images from:

- L_{in}: 200 species of birds, where 10% are labeled images
- U_{in}: same set of classes as L_{in}
- U_{out}: different set of classes in the Aves taxa

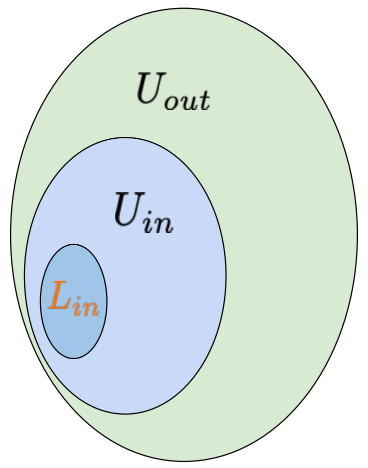
Differences from existing benchmarks:

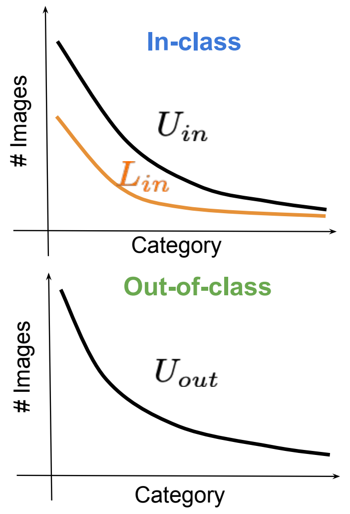
- Long-tailed distribution of classes
- Unlabeled data contains novel classes
- Fine-grained similarity between classes

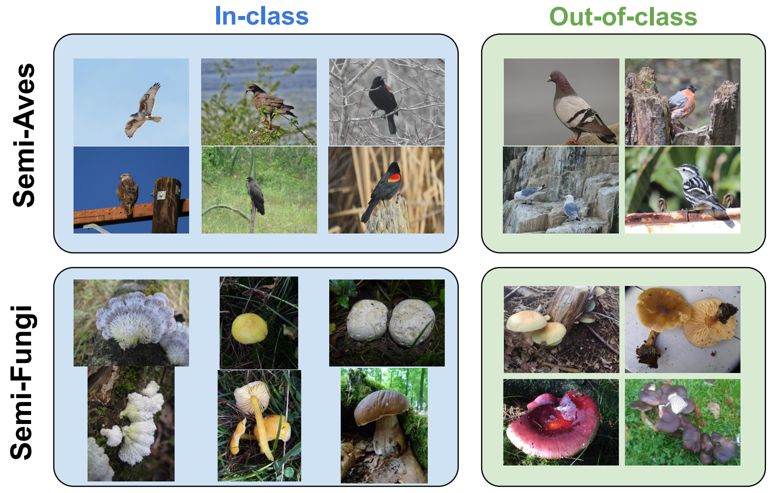
Variations:

- **Semi-Fungi @** FGVC5 & **Semi-iNat @** FGVC8

Dataset	#Classes L_{in} / U_{in} / U_{out}	#Images	Imbalance Ratio
Semi-Aves	200/200/800	6k/27k/122k	7.9
Semi-Fungi	200/200/1194	4k/13k/65k	10.1
Semi-iNat	810/ (2438)	10k/ (313k)	12.9







Experiments

Methods for semi-supervised learning (SSL)

- Pseudo-Labeling [2] and Curriculum Pseudo-Labeling [3]
- FixMatch [4]
- Self-Training via Distillation [5]
- Self-Supervised Learning (MoCo) [6] + Baseline
- Self-Supervised Learning (MoCo) [6] + Self-Training [7]
- **Baseline:** Train w/ labeled data
- **Oracle:** Train w/ fully labeled data

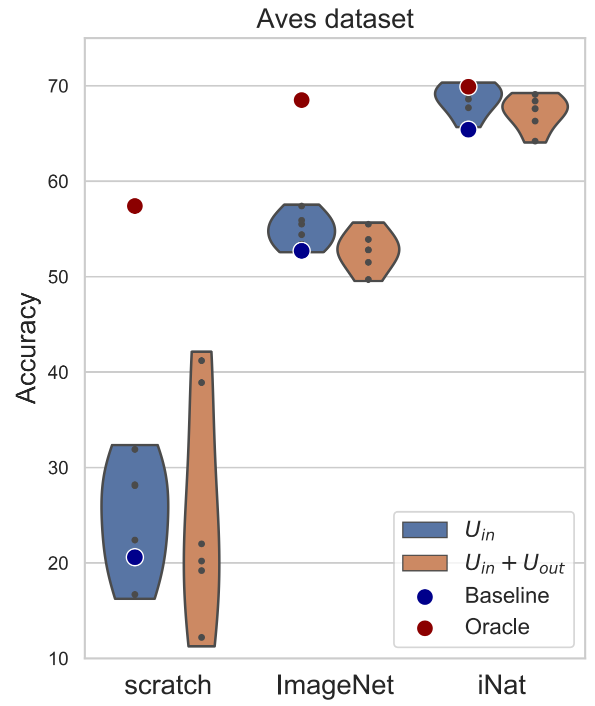
Investigate the role of:

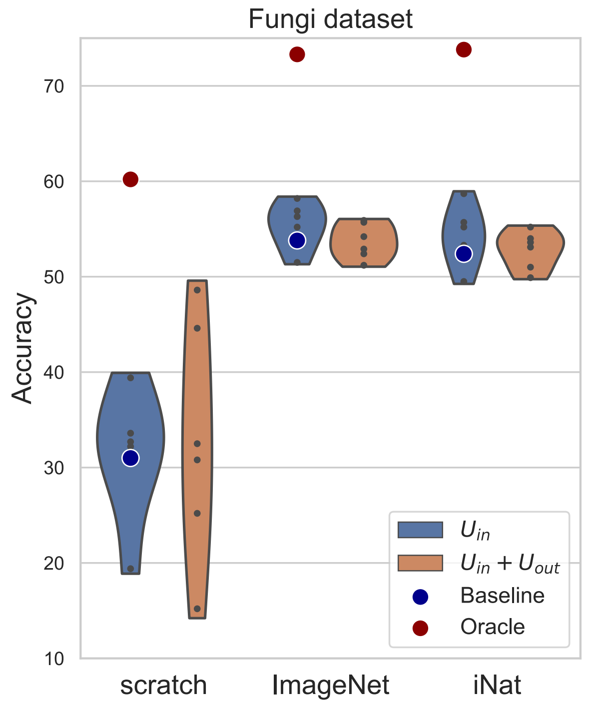
- **Initialization:** scratch / ImageNet / iNat18 pre-trained models
- **Out-of-domain data: U_{in}** only or **U_{in} + U_{out}**

on the performance of ResNet50 w/ 224x224 images

Key Takeaways

- Training from scratch with SSL is worse than supervised transfer learning (Baseline) from ImageNet or iNat (see below).
- Several state-of-the-art SSL techniques are not robust to the presence of out-of-domain data (see right)
- When evaluated w/ transfer learning, contrastive self-supervised learning is not as effective.
- Performance of current methods are still far below the oracle — big room for improvement!



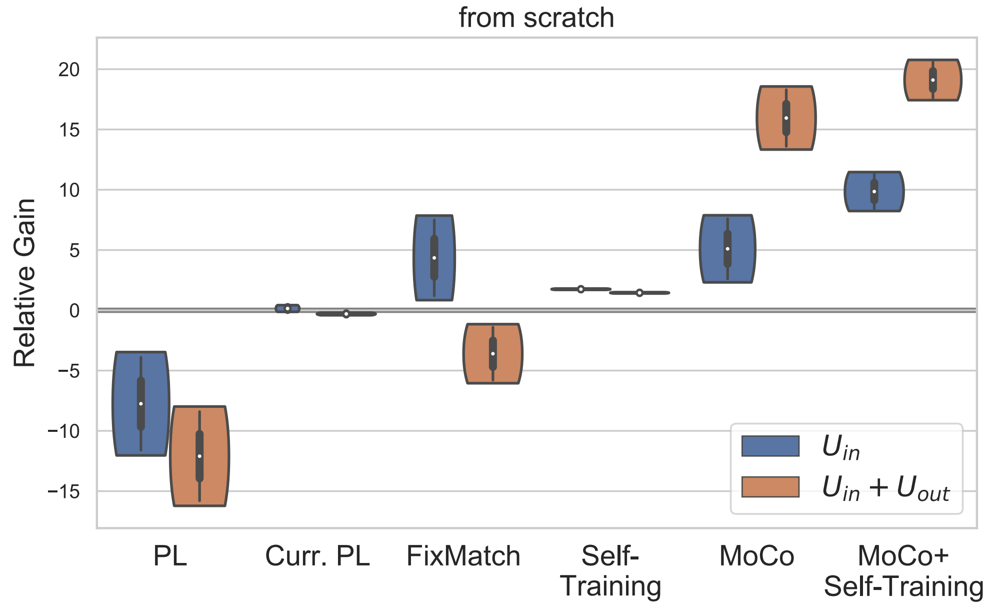


Click here for code

How Effective is Transfer Learning?

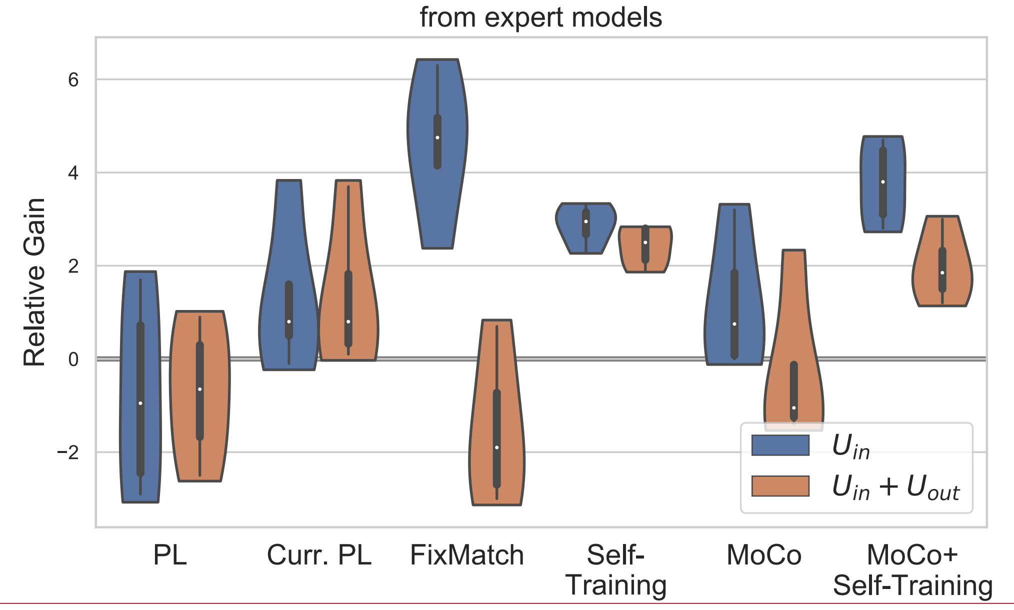
Training from scratch

- FixMatch and Self-Training provide improvements, but self-supervised methods can further benefit from U_{out}
- Overall, MoCo + Self-Training performs the best



Training from experts (ImageNet or iNat)

- FixMatch performs the best when using only U_{in}, while Self-Training is more robust to the presence of U_{out}
- No method was able to reliably use out-of-class data even though the domain shift is relatively small



References

[1] Oliver et al., **Realistic evaluation of deep semi-supervised learning algorithms**, *NeurIPS* '18
[2] Lee, **Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks**, *ICML Workshop* '13
[3] Cascante-Bonilla et al., **Curriculum labeling: Self-paced pseudo labeling for semi-supervised learning**, *arXiv* '20
[4] Sohn et al., **FixMatch: Simplifying semi-supervised learning with consistency and confidence**, *NeurIPS* '20
[5] Xie et al., **Self-training with noisy student improves ImageNet classification**, *CVPR* '20
[6] He et al., **Momentum contrast for unsupervised visual representation learning**, *CVPR* '20
[7] Chen et al., **Big self-supervised models are strong semi-supervised learners**, *NeurIPS* '20