Github page

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

Jong-Chyi Su   Zezhou Cheng   Subhransu Maji

UMassAmherst

Motivation

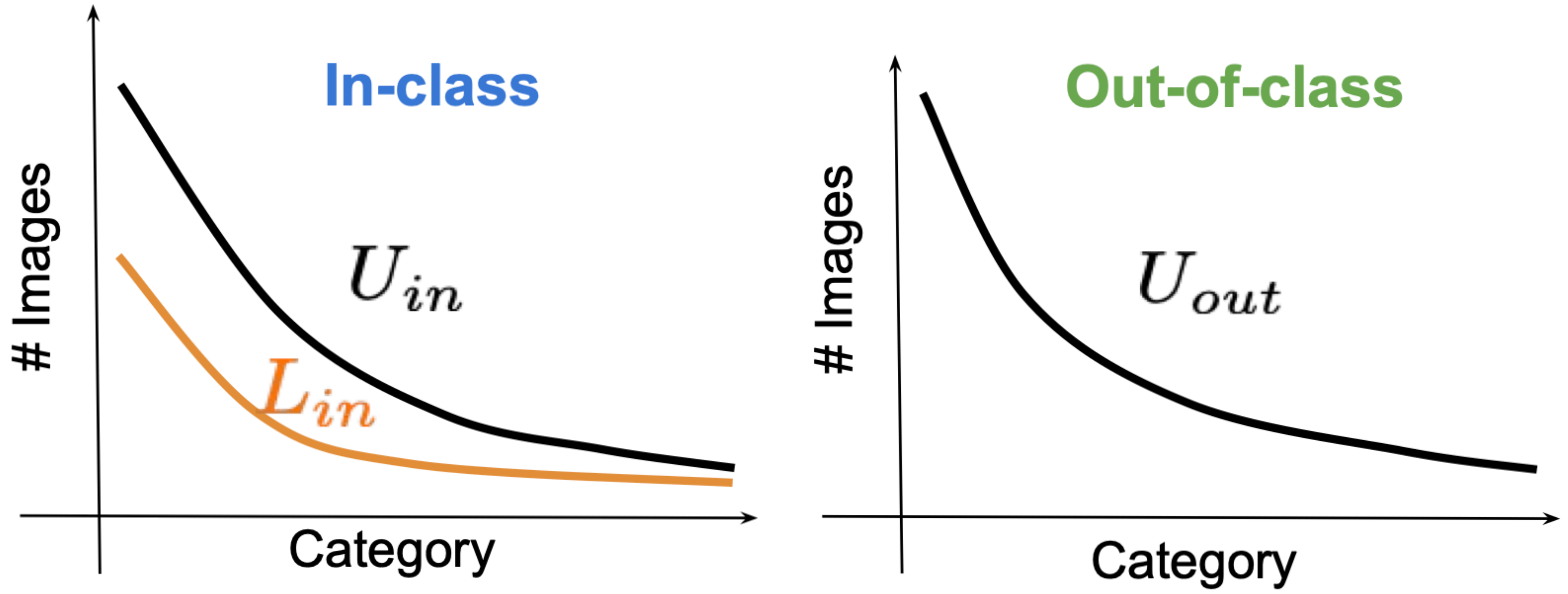
- Existing semi-supervised learning (Semi-SL) benchmarks are lacking [1]:
  - Curated datasets: CIFAR, SVHN, STL-10, ImageNet
  - Uniform class distribution
  - Low-resolution images
  - Unlabeled data does not contain novel class

- Do Semi-SL methods work in realistic datasets?

[1] Oliver et al., **Realistic evaluation of deep semi-supervised learning algorithms**, *NeurIPS* ’18

A Realistic Benchmark for Semi-SL

- Semi-Aves @ FGVC7
- 1000 bird species from iNat-18 [1]

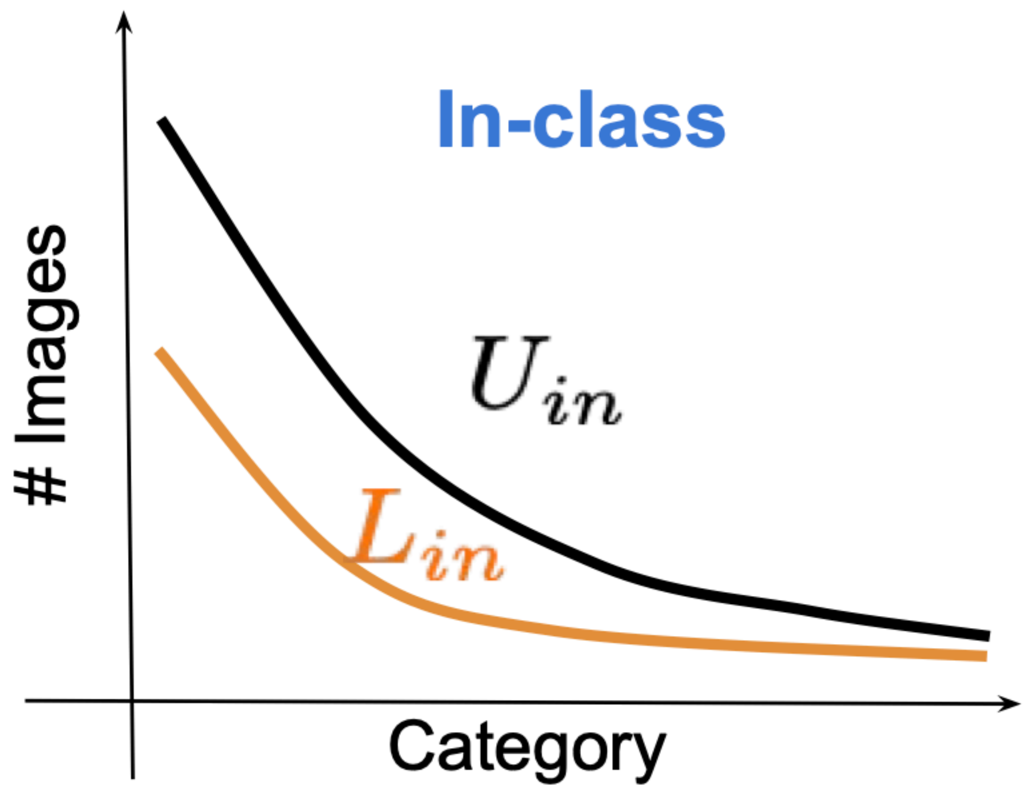


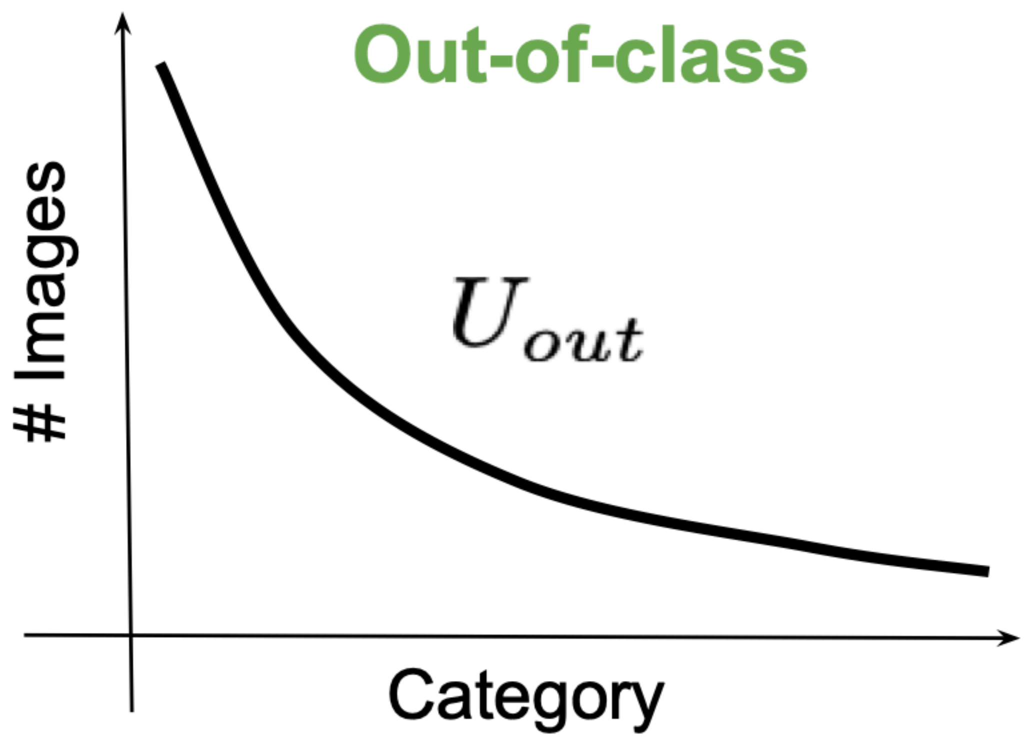
- Data splits:
  - ***L_{in}***: 200 species of birds, where 10% are labeled images
  - ***U_{in}***: same set of classes as *L_{in}*
  - ***U_{out}***: different set of classes in the Aves taxa

[1] van Horn et al., **The iNaturalist species classification and detection dataset**, *CVPR*, '18.

Differences from existing benchmarks:

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





In-class

Out-of-class

Variants

- Semi-Fungi @ FGVC5
- Semi-iNat @ FGVC8 @ CVPR '21
    - Animalia, Plantae, Fungi
    - Combine $\boldsymbol{U_{in}}$ and $\boldsymbol{U_{out}}$
    - Coarse label

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

Semi-SL methods

- Pseudo-Label [1] and Curriculum Pseudo-Label [2]
- FixMatch [3]
- Self-Training via Distillation [4]
- Self-Supervised Learning (MoCo) [5] + Baseline
- Self-Supervised Learning (MoCo) [5] + Self-Training [6]

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

We investigate the effects of

- **Initialization:** scratch / ImageNet / iNat-18 pre-trained models
- **Out-of-domain data:** U_{in} only or $U_{in} + U_{out}$

- **Baseline**: Train w/ labeled data
- **Oracle**: Train with fully labeled data

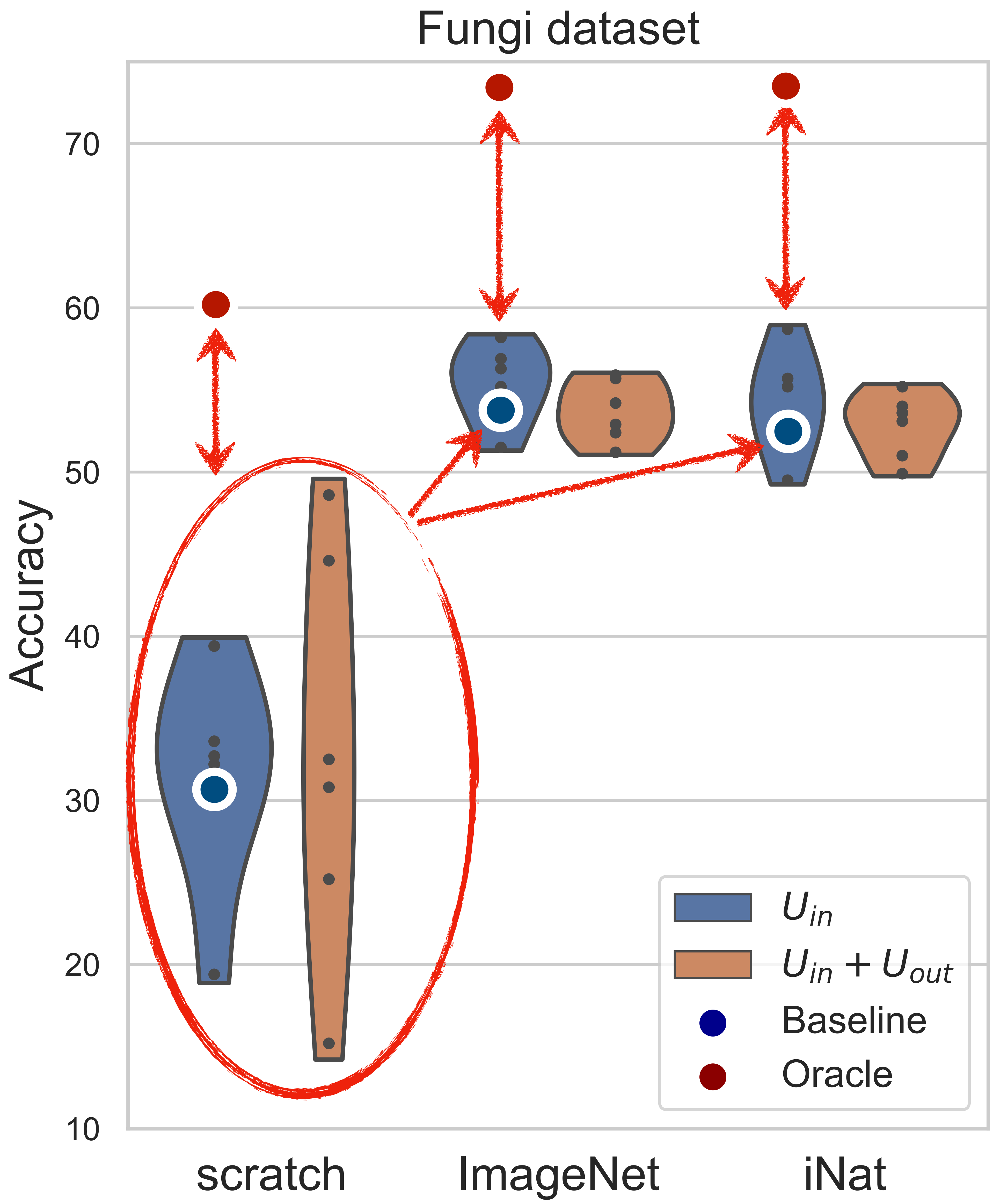
- on the performance of ResNet50 w/ 224x224 image

Overall performances of Semi-SL methods

- Training from scratch with SSL is worse than supervised transfer learning (Baseline) from ImageNet or iNat.

- Several state-of-the-art SSL techniques are not robust to the presence of $\boldsymbol{U_{out}}$

- Performance of current methods are still far below the oracle — big room for improvement!

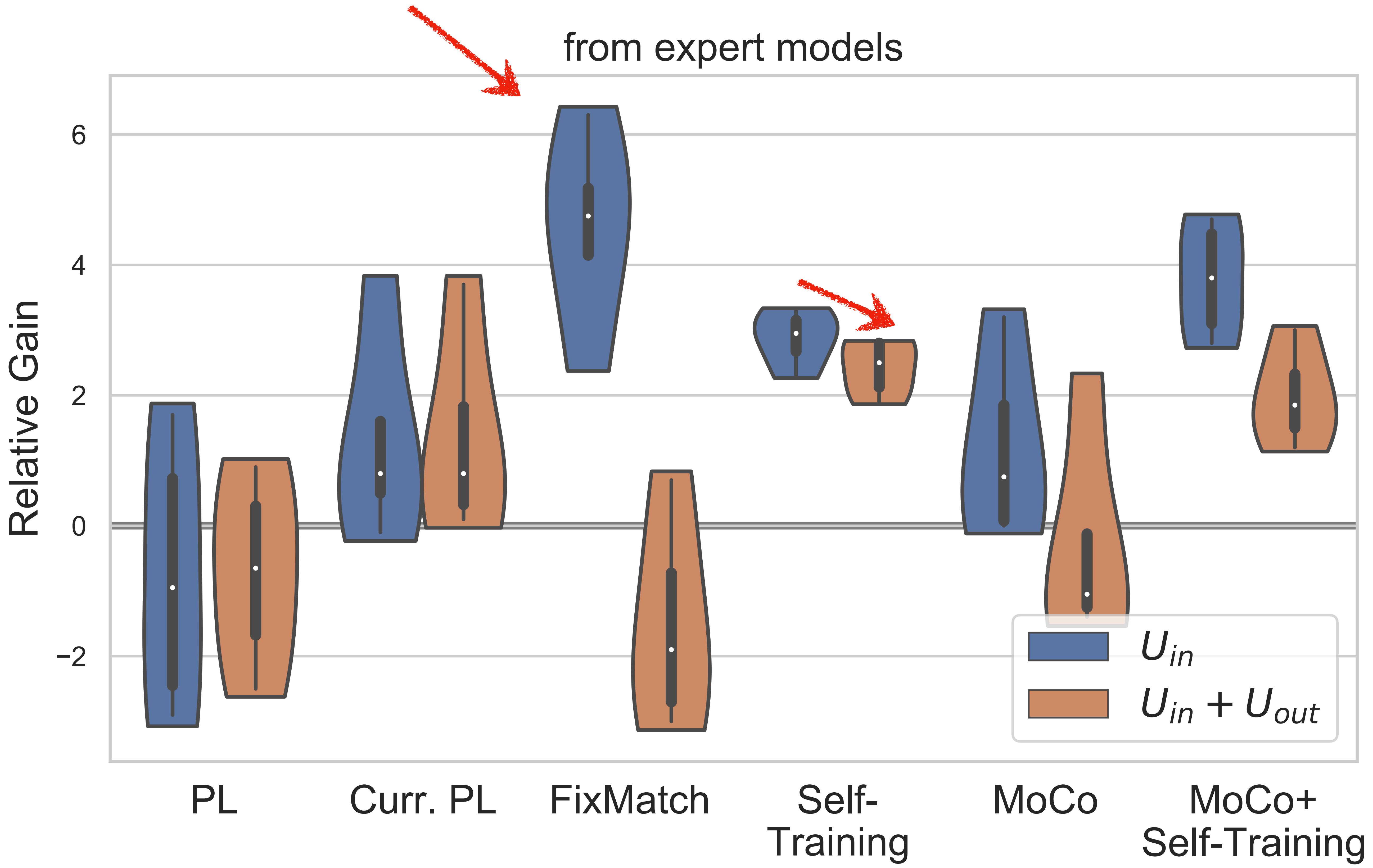


Training from scratch

- FixMatch and Self-Training provide improvements
- Self-Supervised methods can further benefit from U_{out}
- MoCo + Self-Training performs the best

Training from expert models (ImageNet or iNat)

- U_{in} only: FixMatch is the best
- $U_{in} + U_{out}$: Self-Training is more robust
- Self-Supervised learning is not as effective here
- No method was able to reliably use U_{out}
  - even though the domain shift is relatively small



Conclusion

- We created a realistic Semi-SL benchmark for fine-grained classification

- We found that:
    - Existing Semi-SL methods do not work well in realistic settings
    - Transfer learning performs better
    - U_{out} often hurts the performance
    - There is still big room for improvement!