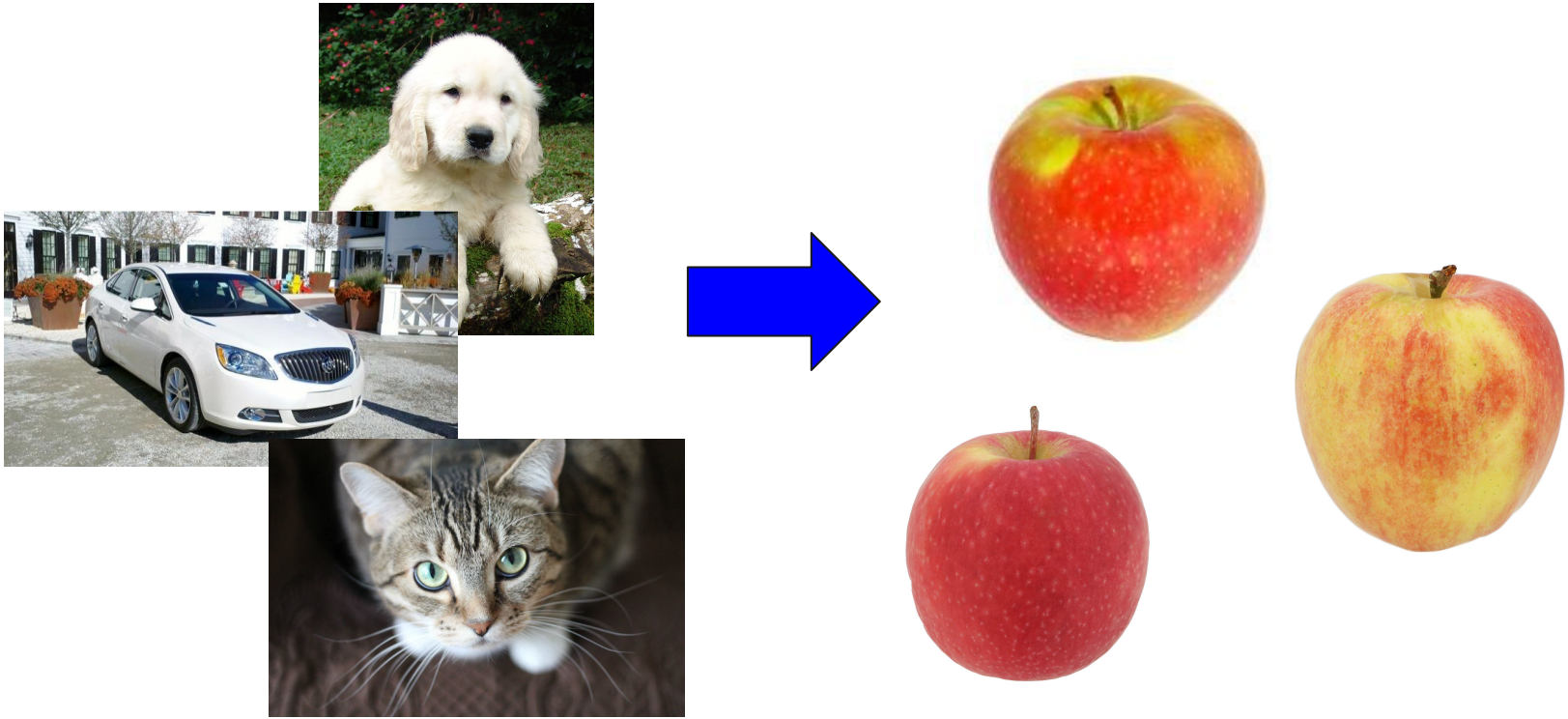
A Baseline for Few-Shot Image Classification

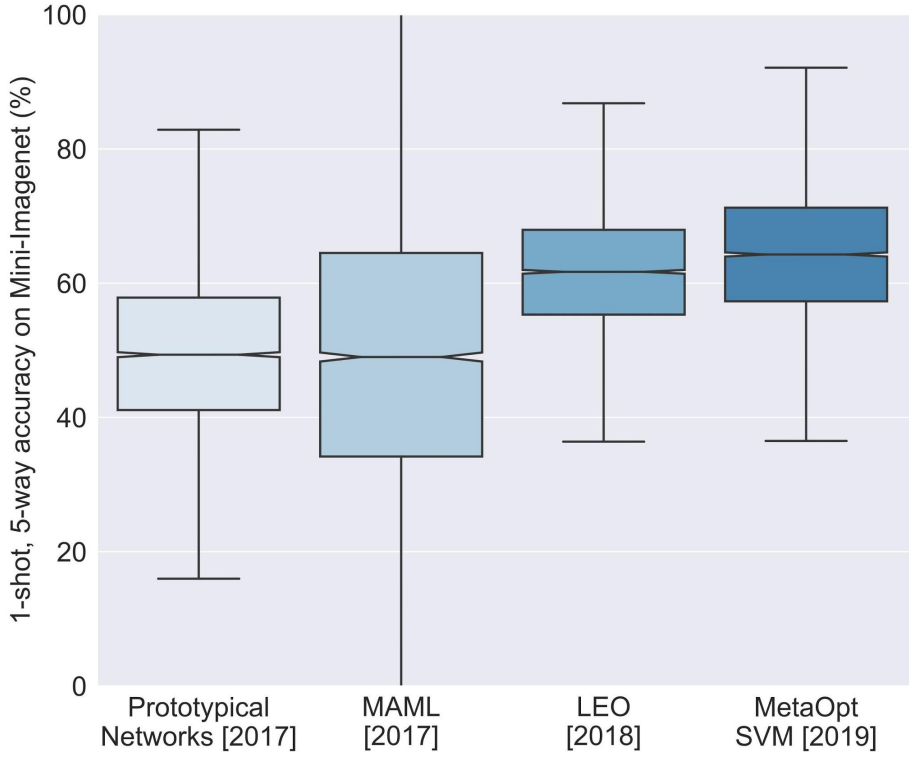
Guneet S. Dhillon[1], Pratik Chaudhari[2], Avinash Ravichandran[1], Stefano Soatto[1, 3]

[1]*Amazon Web Services*, [2]*University of Pennsylvania*, [3]*University of California, Los Angeles*

What is few-shot learning?



Are we making progress?



Goals

- Establish a simple baseline for few-shot image classification
- Provide a systematic evaluation methodology to compare different few-shot algorithms

Proposed baseline

- Standard cross-entropy meta-training / pre-training
- Initialization of classifier for few-shot classification[1]
- Fine-tuning the classifier on the few-shot dataset
  - Vanilla : minimize cross-entropy loss on train data
  - Transductive : minimize entropy loss on test data

[1] Nicholas Frosst, Nicolas Papernot, Geoffrey Hinton. Analyzing and Improving Representations with the Soft Nearest Neighbor Loss. In Proc. of the International Conference on Machine Learning (ICML), 2019.

Results (standard benchmarks)

		Mini-ImageNet		Tiered-ImageNet	
Algorithm	**Architecture**	**1-shot (%)**	**5-shot (%)**	**1-shot (%)**	**5-shot (%)**
Prototypical Networks	$\text{conv}(64)_{\times 4}$	49.42 ± 0.78	68.20 ± 0.66		
MAML	$\text{conv}(32)_{\times 4}$	48.70 ± 1.84	63.11 ± 0.92		
TADAM	ResNet-12	58.5 ± 0.3	76.7 ± 0.3		
Transductive Propagation	$\text{conv}(64)_{\times 4}$	55.51 ± 0.86	69.86 ± 0.65	59.91 ± 0.94	73.30 ± 0.75
LEO	WRN-28-10	61.76 ± 0.08	77.59 ± 0.12	66.33 ± 0.05	81.44 ± 0.09
MetaOpt SVM	ResNet-12*	64.09 ± 0.62	$\mathbf{80.00 \pm 0.45}$	65.81 ± 0.74	81.75 ± 0.53
Support-based initialization	WRN-28-10	58.47 ± 0.66	75.56 ± 0.52	$67.34 \pm 0.69^{\dagger}$	$83.32 \pm 0.51^{\dagger}$
Fine-tuning	WRN-28-10	59.62 ± 0.66	$\mathbf{79.93 \pm 0.47}$	66.23 ± 0.68	$\mathbf{86.08 \pm 0.47}$
Transductive fine-tuning	WRN-28-10	$\mathbf{68.11 \pm 0.69}$	$\mathbf{80.36 \pm 0.50}$	$\mathbf{72.87 \pm 0.71}$	$\mathbf{86.15 \pm 0.50}$

- Same hyper-parameters for all experiments

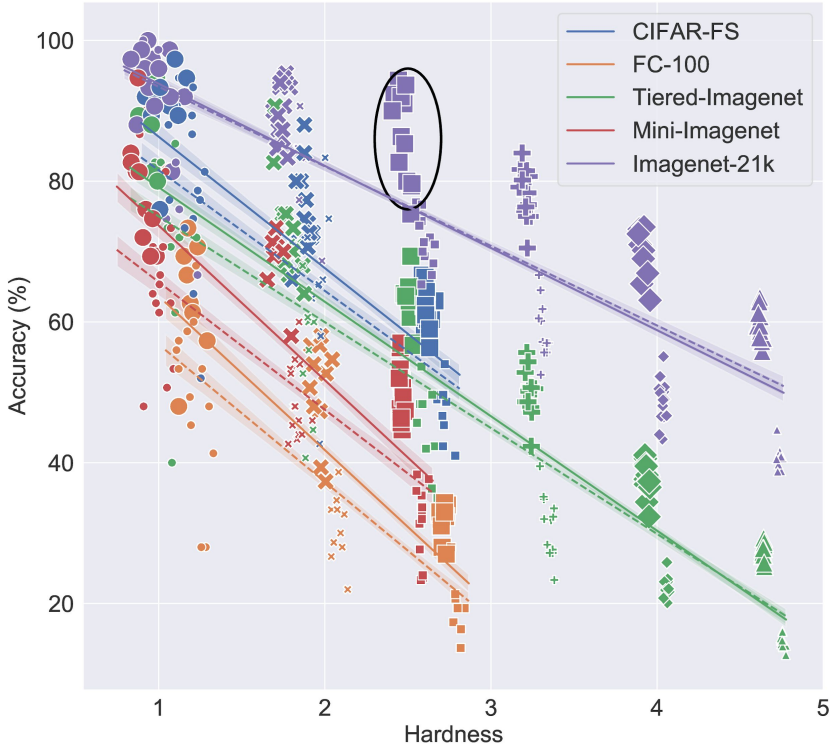
Results (ImageNet-21k)

		ImageNet-21k					
		Way					
Algorithm	**Shot**	**5**	**10**	**20**	**40**	**80**	**160**
Support-based initialization	1	87.20 ± 1.72	78.71 ± 1.63	69.48 ± 1.30	60.55 ± 1.03	49.15 ± 0.68	40.57 ± 0.42
Transductive fine-tuning	1	89.00 ± 1.86	79.88 ± 1.70	69.66 ± 1.30	60.72 ± 1.04	48.88 ± 0.66	40.46 ± 0.44
Support-based initialization	5	95.73 ± 0.84	91.00 ± 1.09	84.77 ± 1.04	78.10 ± 0.79	70.09 ± 0.71	61.93 ± 0.45
Transductive fine-tuning	5	95.20 ± 0.94	90.61 ± 1.03	84.21 ± 1.09	77.13 ± 0.82	68.94 ± 0.75	60.11 ± 0.48

- 7,491 meta-training classes, 13,007 classes for few-shot training / testing
- 1-shot 5-way accuracy : 89%
- 1-shot 20-way accuracy : 70%

A proposal for systematic evaluation

- Hardness measures how hard is it to correctly classify a test set, given the labeled train set?



Come to the poster

Link to the full paper: