Uniform Sampling Over Episodic Difficulty

S. M. R. Arnold*, G. S. Dhillon*, A. Ravichandran, S. Soatto

October 18, 2021

USC University of Southern California

aws

Episodic Training in Few-Shot Learning

Few-Shot Learning

- Learn a model to solve new tasks with limited labelled data.

Episodic Training in 3 Steps

1. Sample an episode from distribution.
2. Solve the episode with limited data.
3. Update the model to improve generalization of solution.

Plenty of Recent Methods

- Gradient-based: MAML, ANIL, MetaCurvature, KFO, MT-Nets, ...
- Metric-based: ProtoNets, MetaOptNet, FEAT, DeepEMD, ...

This Paper: A Closer Look at Episodic Sampling

Driving Question

- How should we sample episodes for best transfer accuracy?

Contributions

- Analysis of episode difficulty and its distribution.
- Simple method to approximate any episodic sampling distribution.
- Main result: uniform sampling over episode difficulty improves episodic training.

Which Is Best?

Few-Shot Classification Episodes

Baseline Episode Sampling

1. Sample *n* classes from base dataset.
2. Sample *k* samples / class for support set τ_S.
3. Sample *k'* samples / class for query set τ_Q.

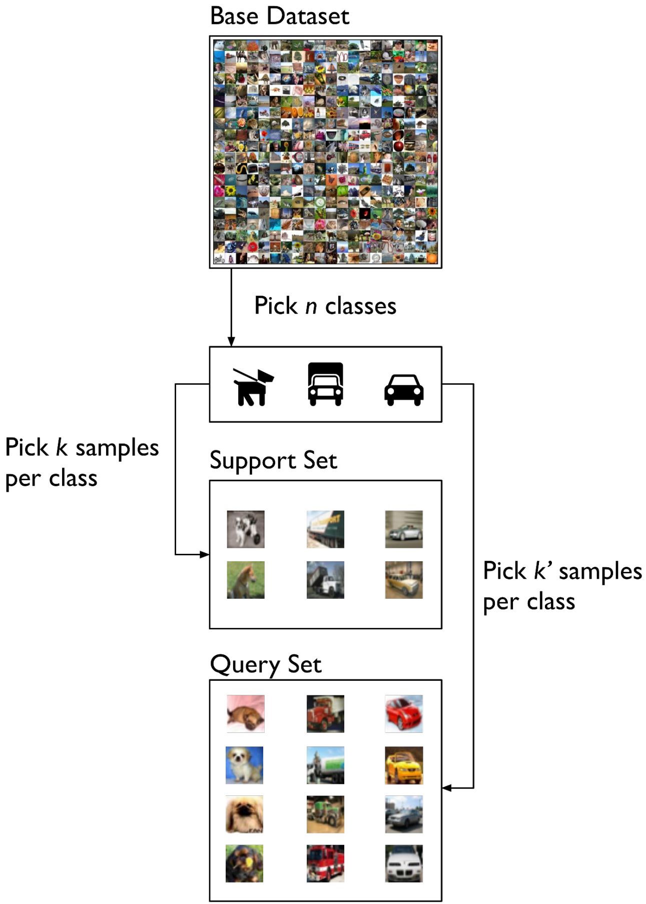
Solving an Episode with ProtoNets

- Compute class centroids on support set:

$$\phi_\theta^c = \frac{1}{k} \sum_{\substack{(x,y) \in \tau_S \\ y=c}} \phi_\theta(x)$$

- Classify query set with nearest centroid:

$$l_\theta(y \mid x, \tau_S) = \frac{\exp\left(-d(\phi_\theta(x), \phi_\theta^y)\right)}{\sum_{y' \in \mathcal{C}_\tau} \exp\left(-d(\phi_\theta(x), \phi_\theta^{y'})\right)}$$



"Prototypical Networks for Few-Shot Learning", Snell et al., NeurIPS 2017.

Distribution of Episode Difficulty

Definition

- The difficulty of an episode τ is given by:

$$\Omega_{l_\theta}(\tau) = -\log l_\theta(\tau)$$

for model likelihood l, support set S, and query set Q.

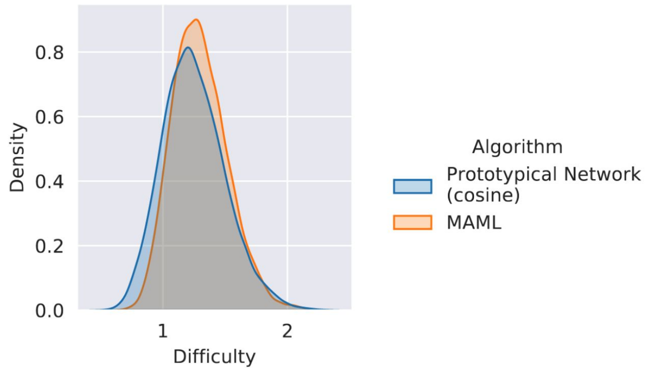
Why This Definition?

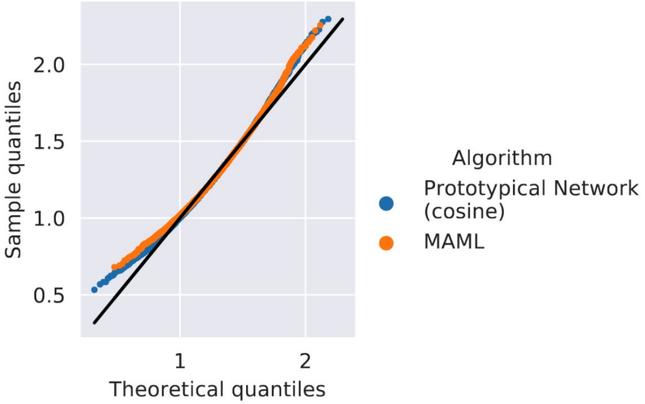
- Easy to compute, readily available during training.
- Model-agnostic — applies to all methods.
- No discretization artifacts (unlike, say, accuracy).

Empirical Analysis

For many algorithms and models:

Episode difficulty is approximately normally distributed.





Implicit Dependence on Modelling Choices

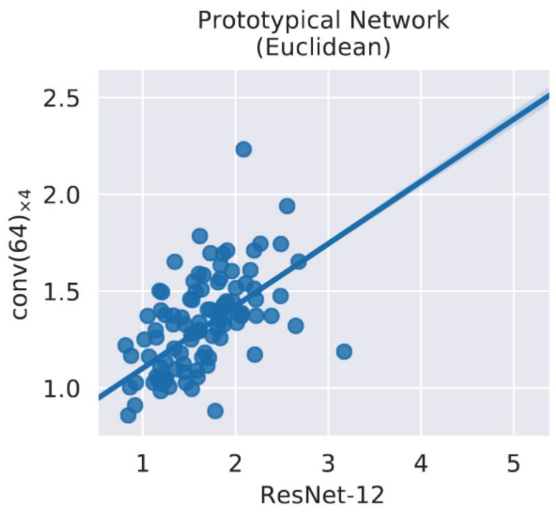
Is episode difficulty the same across different...

... model architectures? ✔

- Compare CNN4 *v.s.* ResNet12.
- Average Spearman correlation: 0.59.

... model parameters?

... training algorithms?



Implicit Dependence on Modelling Choices

Is episode difficulty the same across different…

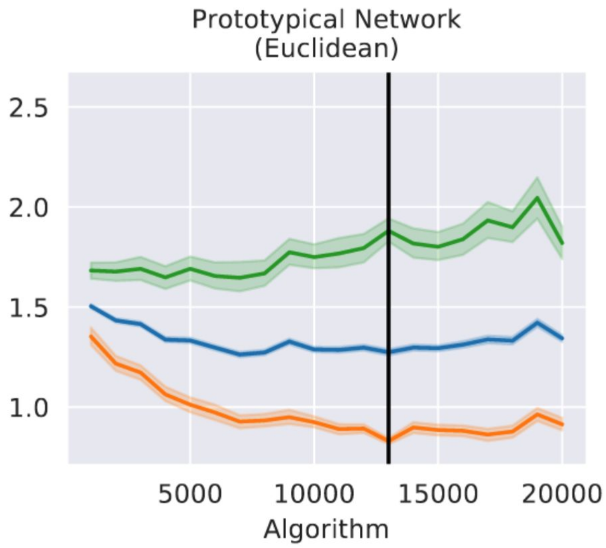
… model architectures? ✔

- Compare CNN4 *v.s.* ResNet12.
- Average Spearman correlation: 0.59.

… model parameters? ✔

- Compare across iterations in training run.
- Hard episodes remain hard; easy remain easy.

… training algorithms?



Implicit Dependence on Modelling Choices

Is episode difficulty the same across different…

… model architectures? ✔

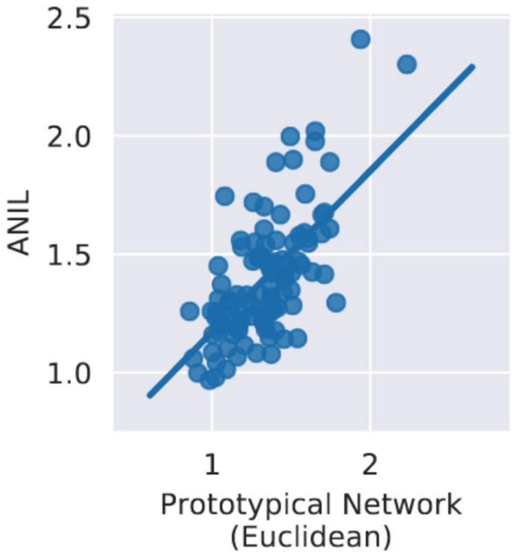
- Compare CNN4 *v.s.* ResNet12.
- Average Spearman correlation: 0.59.

… model parameters? ✔

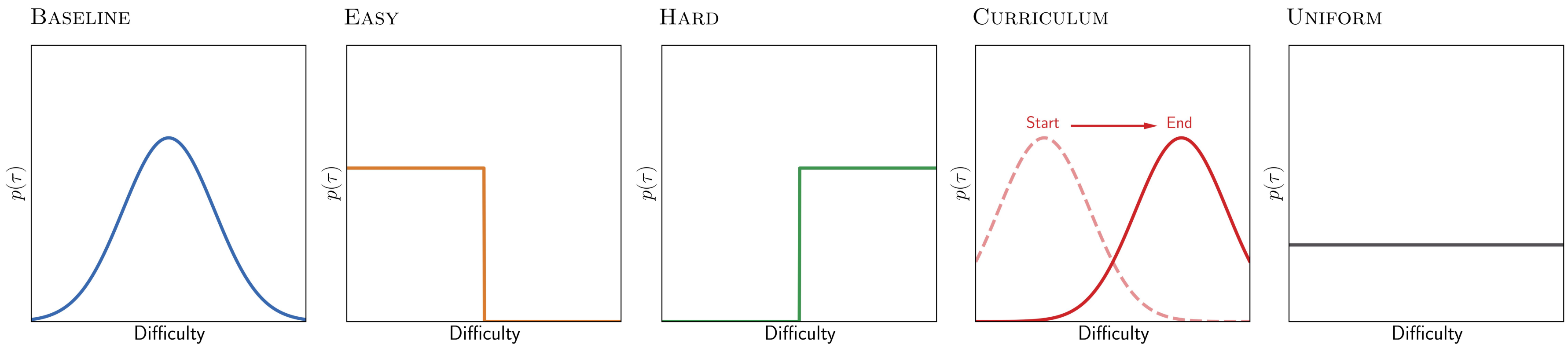
- Compare across iterations in training run.
- Hard episodes remain hard; easy remain easy.

… training algorithms? ✔

- Compare across MAML, ANIL, ProtoNet (Euclidean & Cosine).
- Average Spearman correlation: 0.65.



How should we sample episodes?



5 Candidate Distributions

- Baseline — Normal distribution over difficulty.
- Easy — Sample uniformly over the easier 50% of episodes.
- Hard — Sample uniformly over the harder 50% of episodes.
- Curriculum — Sample easier episode when training starts, harder episodes towards end.
- Uniform — Sample uniformly over the difficulty range.

A Simple Method to Study Episode Sampling

aws USC University of Southern California

Importance Sampling for Episodic Training

- Reweight episodes to approximate target distribution *p(τ)*:

$$\mathop{\mathrm{E}}_{\tau\sim q(\cdot)}\left[w(\tau)\log l_\theta(\tau)\right] \quad \text{where} \quad w(\tau)=\frac{p(\tau)}{q(\tau)}$$

and *q(τ)* is the distribution induce by Baseline sampling.

Adjusted Mini-Batching with Expected Sample Size

- Get the right number of sample from the target distribution:

$$\mathop{\mathrm{E}}_{\tau\sim p(\cdot)}\left[\log l_\theta(\tau)\right] \approx \frac{1}{\mathrm{ESS}(\mathcal{B})}\sum_{\tau\in\mathcal{B}} w(\tau)\log l_\theta(\tau)$$

$$\text{where} \quad \mathrm{ESS}(\mathcal{B})=\frac{\left(\sum_{\tau\in\mathcal{B}} w(\tau)\right)^2}{\sum_{\tau\in\mathcal{B}} w(\tau)^2}$$

```
sampler.reset()
for episode in batch:
  loss = compute_loss(episode)
  is_weight = sampler.weight(loss)
  sampler.update(loss)
  (is_weight * loss).backward()
model.parameters /= sampler.ess()
```

Sampling Matters for Episodic Training

Experimental Setup

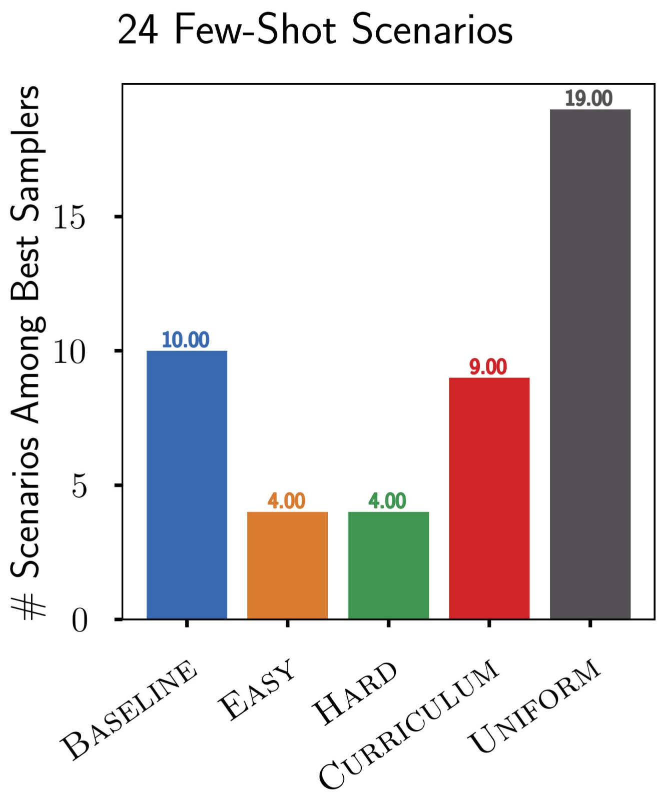
Compare 5 candidate distributions on different:

- Architectures: CNN4, ResNet12
- Algorithms: MAML, ANIL, ProtoNet - Euclidean & Cosine
- Datasets: mini-ImageNet, tiered-ImageNet
- Setting: 5-ways 1-shot, 5-ways 5-shots

Total: 24 different few-shot scenarios.

Results

Uniform sampling dominates, Baseline second best.



Sampling Improves Cross-Domain Transfer

Experimental Setup

Compare Baseline *v.s.* Uniform when:

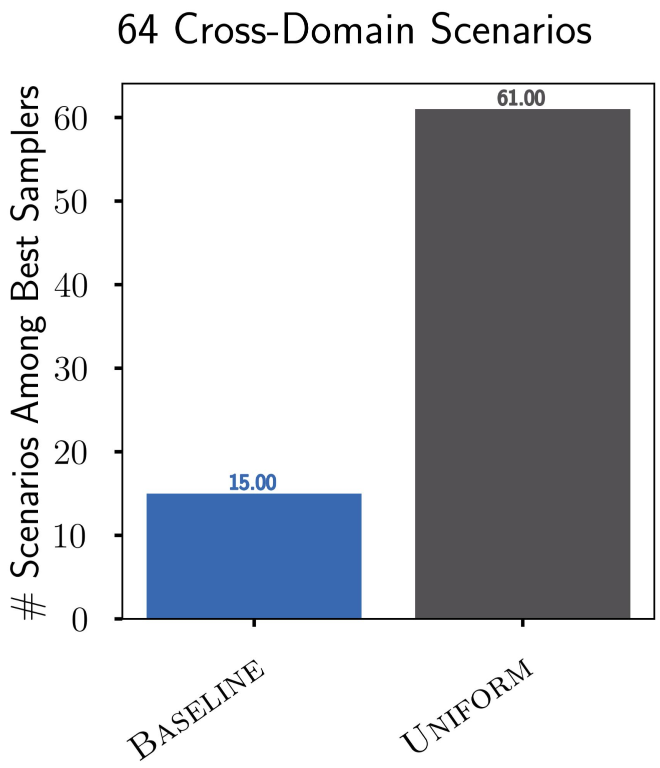
- training on mini-ImageNet or tiered-ImageNet
- testing on:
    - CUB-200,
    - Describable Textures,
    - FGVC-Aircraft, and
    - VGG Flowers.

Total: 64 Cross-Domain Scenarios.

Results

Uniform sampling dominates.

Thanks Reviewer KvDG!



... And Improves Upon SOTA

	Mini-ImageNet		Tiered-ImageNet	
	1-shot (%)	5-shot (%)	1-shot (%)	5-shot (%)
FEAT	66.02±0.20	81.17±0.14	**70.50±0.23**	84.26±0.16
+ UNIFORM (Online)	**66.27±0.20**	**81.54±0.14**	**70.61±0.23**	**84.42±0.16**

Experimental Setup

- Compare Baseline *v.s.* Uniform when training with FEAT.
- ResNet12 on mini-ImageNet & tiered-ImageNet.

Results

Uniform sampling also improves SOTA algorithms.

"Few-Shot Learning via Embedding Adaptation with Set-to-Set Functions", Ye, Hu, et al., CVPR 2020.

Thank You

Takeaways

1. Sampling matters in episodic training.
2. Episodic difficulty is (mostly) agnostic to architecture, algorithm, and parameter choice.
3. Uniform sampling outperforms other sampling schemes.

Collaborators

Guneet S. Dhillon*

Avinash Ravichandran

Stefano Soatto

Learn More

- PDF, poster, slides: sebarnold.net/projects/eis
- Code: bit.ly/3p7cYz5 or learn2learn.net
- Contact: smr.arnold@gmail.com