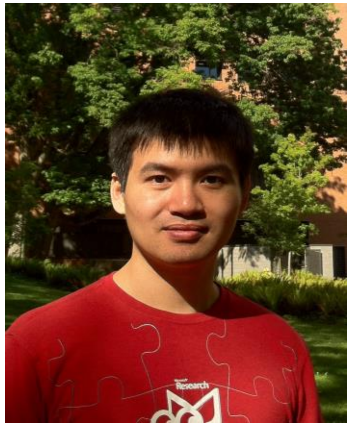
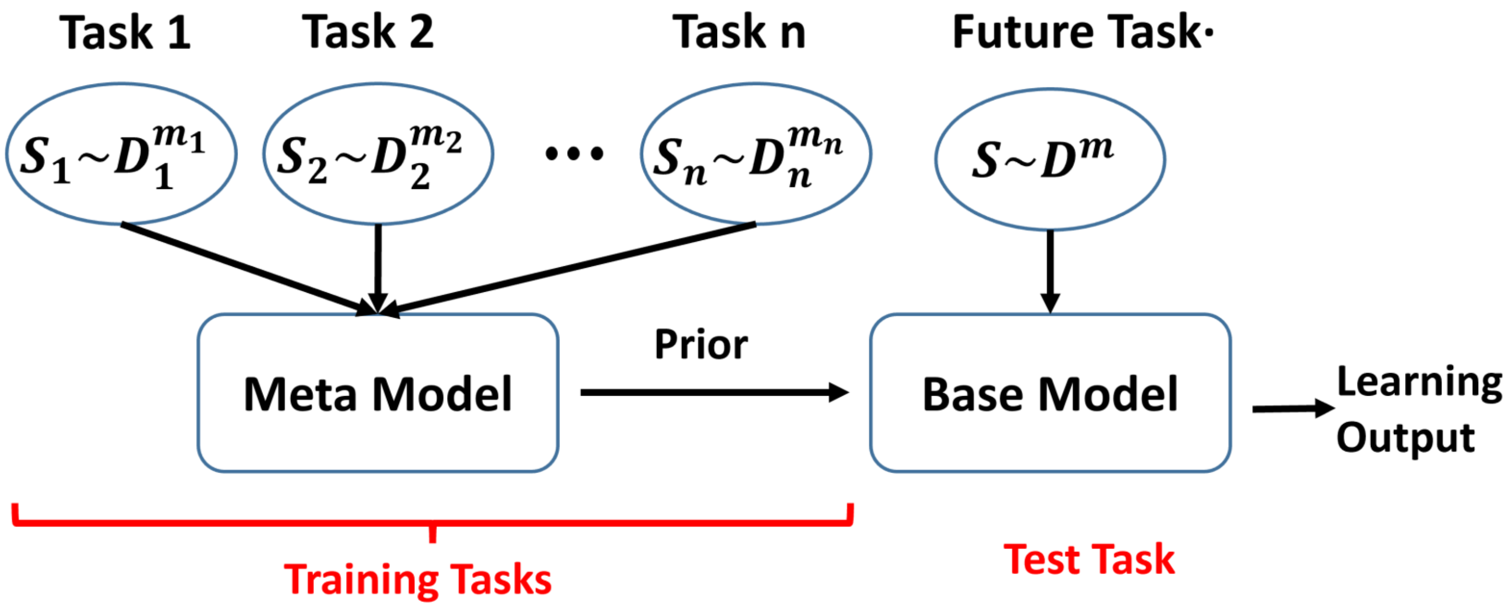
Adaptive Task Sampling for Meta-Learning

Chenghao Liu, Zhihao Wang, Doyen Sahoo, Yuan Fang, Kun Zhang, Steven C.H. Hoi

Meta-Learning for Few-shot Learning

- Recent approaches to few-shot learning follow a **meta-learning** principle
- **SIMULATE** few-shot tasks on the meta-train data to acquire ability to do few-shot learning
- Each task in meta-train set has its own
  - Training Data: Support Set
  - Test Data: Query Set
- Minimize the loss of the prediction for each sample in the query set, given the support set



Meta-Learning for Few-shot Learning

- Construct a collection of *K*-way-*M*-shot classification tasks sampled from the amply labelled set
- Procedure:
  - **Randomly sample** *K* classes
  - **Randomly sample** *M* and *N* labelled images per class to construct support set and query set
- The **goal** is to learn from support set and improve performance over query set

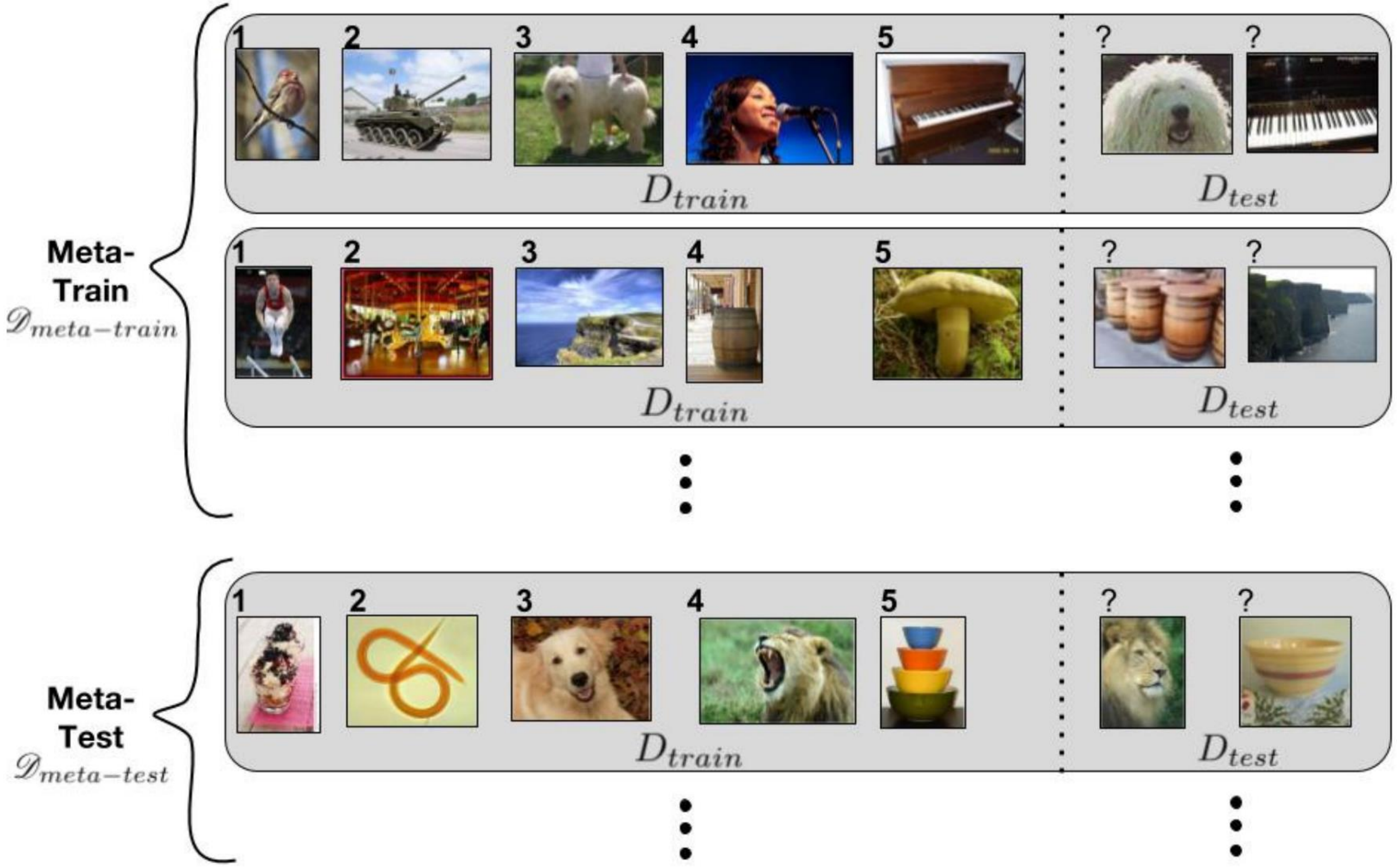
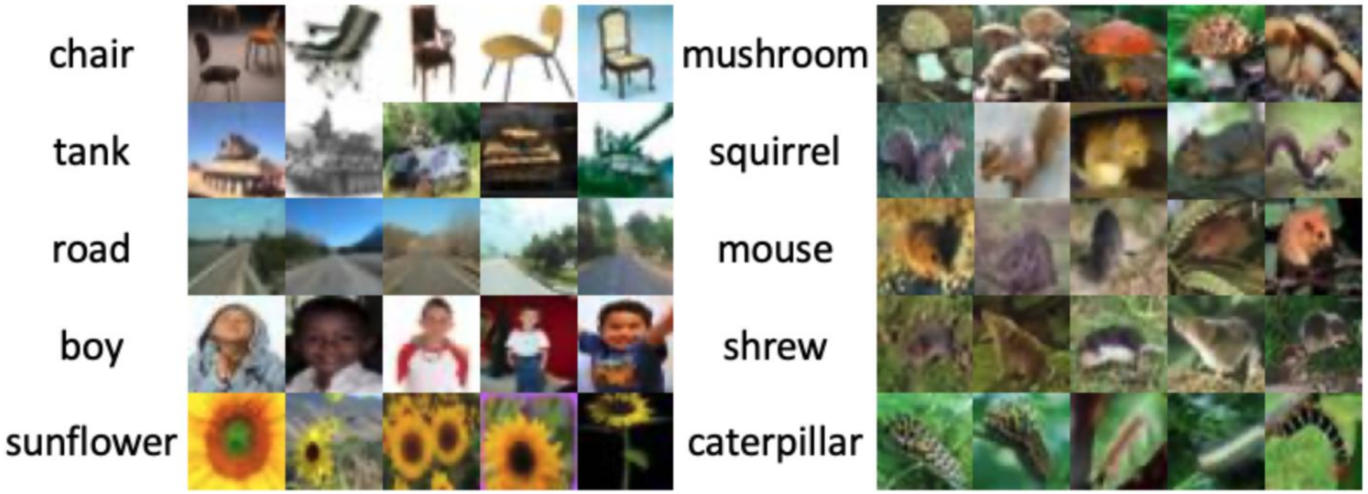


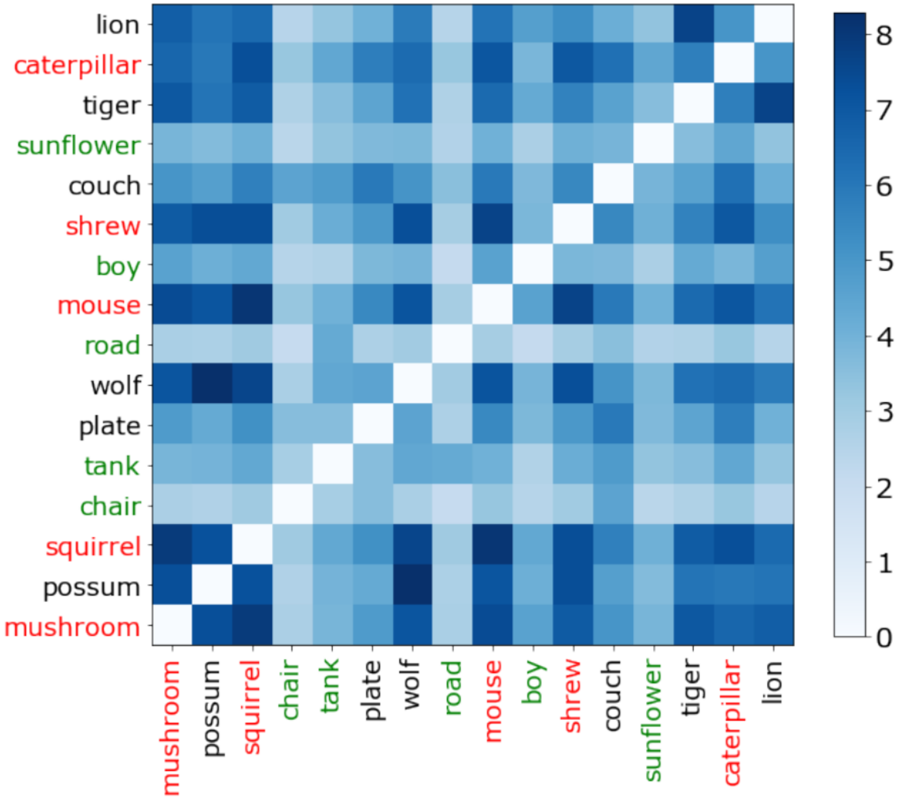
Figure: Sachin Ravi

Random Sampling



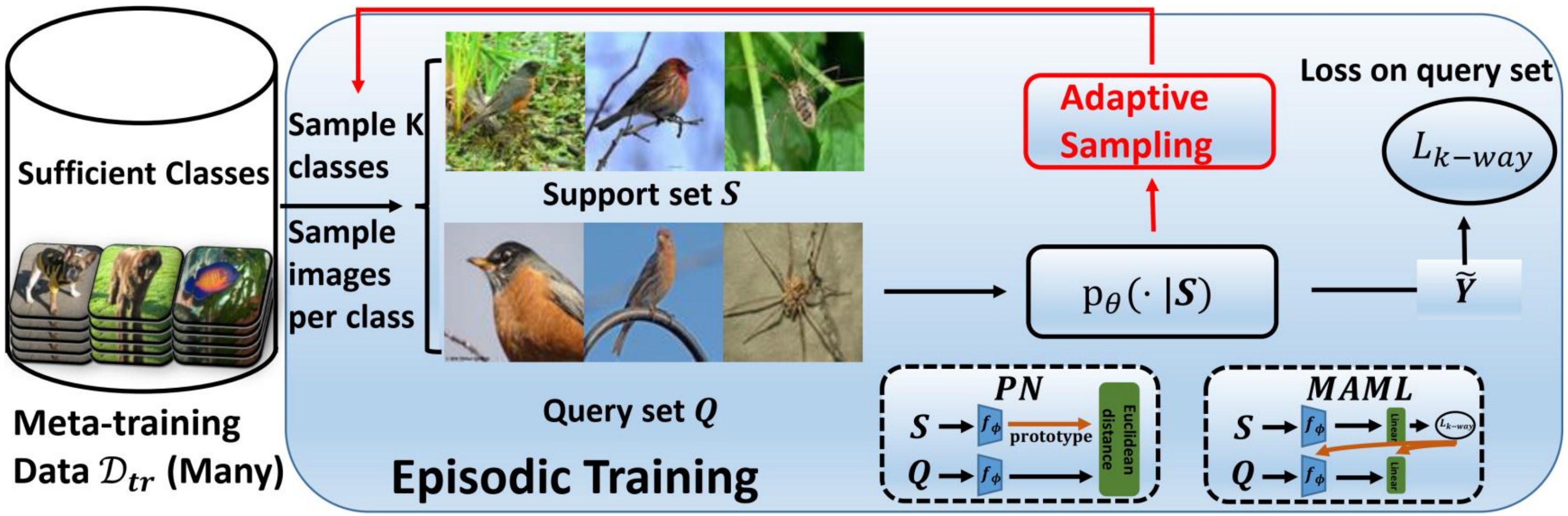
Random Sampling

Adaptive Sampling



ignore the intrinsic relationships between classes!

Motivation



- A randomly sampled task of classifying dogs from laptops may have little effect on the model update due to its simpleness
- Prioritizing challenging training examples could improve the generalization performance

Idea: Can we perform adaptive task sampling and create more difficult tasks for meta-learning?

Challenge: Define the difficulty of a task

Contribution

- A **class-pair based** adaptive task sampling method for meta-learning
- A **greedy class-pair** based approach that reduces the complexity of task distribution computation and guarantees the generation of an identical task distribution
- Comprehensive experiments about the impact of the adaptive task sampling method by integrating it with various meta-learning approaches
- Extensive investigation of different sampling strategies, including class-based method, easy class-pair based method and uncertain class-pair based method

Class-based Sampling

Traditional Supervised Learning

Initial: Probability of selecting each sample is equal. $p_0(i|\mathbb{D}) = \frac{1}{|\mathbb{D}|}$

Episode t: Update the selection probability according to the current prediction probability and the selection probability at previous iteration

$$p^{t+1}(i) \propto (p^t(i))^\tau e^{\alpha(1-p(y_i|x_i))}$$

Meta-Learning for Few-shot Learning

Idea: Direct task based sampling is infeasible, adaptively sample classes for each *K*-way classification task.

Given S^t and Q^t at episode *t,* update the class selection probability

$$p_C^{t+1}(c) \propto (p^t(c))^\tau e^{\alpha \frac{\sum_{(q_n,y_n)\in \mathbb{Q}^t} \mathbb{I}[c\neq y_n] p(c|q_n,\mathbb{S}^t) + \mathbb{I}[c=y_n](1-p(c|q_n,\mathbb{S}^t))}{NK}}$$

It implicitly assumes that the difficulty of each class is independent!

Class-Pair Based Sampling

- Exploits the pairwise relationships between classes
- Formulate the task selection probability by leveraging the Markov random field over class pairs

$$p_{CP}^{t+1}(\mathbb{L}_K^{t+1}) \propto \prod_{(i,j)\in\mathbb{L}_K^{t+1}} C^t(i,j), \qquad \text{s.t. } i,j \in \mathbb{C}_{tr}$$

  - $C_t(i,j)$ is a potential function over class pair (*i, j*) at episode *t*

- Adaptively update the potential function

$\binom{K}{2} \cdot \binom{|\mathbb{C}_{tr}|}{K}$ multiplication operations need to be performed for different combinations of *K*-class in the category set!

$$C^{t+1}(i,j) \leftarrow (C^t(i,j))^\tau e^{\alpha \bar{p}((i,j)|\mathbb{S}^t,\mathbb{Q}^t)}, \quad i \neq j$$

$$\bar{p}((i,j)|\mathbb{S}^t,\mathbb{Q}^t) = \frac{\sum_{(q_n,y_n=j)\in\mathbb{Q}^t} p(c=i|q_n,\mathbb{S}^t)}{N} + \frac{\sum_{(q_n,y_n=i)\in\mathbb{Q}^t} p(c=j|q_n,\mathbb{S}^t)}{N}$$

Greedy Class-Pair Based Sampling

- Iteratively sample a new class based on the already sampled classes (cost O(K))

$$p_{GCP}^{t+1}(\mathbb{L}_{k+1}^{t+1}) \propto \begin{cases} C^t(i,j), & k=1 \\ p(c|\mathbb{L}_k^{t+1}, C^t), & k>1 \end{cases} \qquad p(c=i|\mathbb{L}_k^{t+1}, C^t) \propto \prod_{j\in\mathbb{L}_k^{t+1}} C^t(i,j)$$

- **Proposition** *The greedy class-pair based sampling strategy is identical to the class-pair based sampling*

Class	1	2	3	4	5
1	0	2	5	6	3
2	2	0	9	8	2
3	5	9	0	1	1
4	6	8	1	0	1
5	3	2	1	1	0

Class-pair potential C^t

$\mathbb{L}_0^{t+1} = \{\}.$

$\mathbb{L}_2^{t+1} = \{2, 3\}$

$p(c|\mathbb{L}_2^{t+1}, C^t) = C^t_{2\cdot} \odot C^t_{3\cdot} = (10,0,0,8,2).$

$\mathbb{L}_3^{t+1} = \{2, 3, 1\}$

$p(c|\mathbb{L}_3^{t+1} C^t) = p(c|\mathbb{L}_2^{t+1}, C^t) \odot C^t_{1\cdot} = (0,0,0,48,6).$

$\mathbb{L}_4^{t+1} = \{2, 3, 1, 4\}$

Empirical Results

- Compatibility with different meta-learning algorithms

	miniImageNet		CIFAR-FS	
Model	1-shot	5-shot	1-shot	5-shot
Matching Network †	48.26 ± 0.76	62.27 ± 0.71	53.14 ± 0.85	68.16 ± 0.76
Matching Network with gcp-sampling	$\mathbf{49.61} \pm 0.77$	$\mathbf{63.23} \pm 0.75$	$\mathbf{54.72} \pm 0.87$	$\mathbf{69.28} \pm 0.74$
PN †	44.15 ± 0.76	63.89 ± 0.71	54.87 ± 0.72	71.64 ± 0.58
PN with gcp-sampling	$\mathbf{47.13} \pm 0.81$	$\mathbf{64.75} \pm 0.72$	$\mathbf{56.12} \pm 0.81$	$\mathbf{72.77} \pm 0.64$
Reptile †	46.12 ± 0.80	63.56 ± 0.70	55.86 ± 1.00	71.08 ± 0.74
Reptile with gcp-sampling	$\mathbf{47.60} \pm 0.80$	$\mathbf{64.56} \pm 0.69$	$\mathbf{57.25} \pm 0.99$	$\mathbf{71.69} \pm 0.71$
MAML †	48.25 ± 0.62	64.09 ± 0.70	56.93 ± 0.99	72.10 ± 0.74
MAML with gcp-sampling	$\mathbf{49.65} \pm 0.85$	$\mathbf{65.37} \pm 0.70$	$\mathbf{57.62} \pm 0.97$	$\mathbf{72.51} \pm 0.72$
MAML++ †	50.60 ± 0.82	68.24 ± 0.68	58.87 ± 0.97	73.86 ± 0.76
MAML++ with gcp-sampling	$\mathbf{52.34} \pm 0.81$	$\mathbf{69.21} \pm 0.68$	$\mathbf{60.14} \pm 0.97$	$\mathbf{73.98} \pm 0.74$

Empirical Results

- Efficacy of different adaptive task sampling strategies

	miniImageNet		CIFAR-FS	
Sampling Strategy	5-way-1-shot	5-way-5-shot	5-way-1-shot	5-way-5-shot
random sampling	50.60 ± 0.82	68.24 ± 0.68	58.87 ± 0.97	73.36 ± 0.76
c-sampling with hard class	51.43 ± 0.75	68.74 ± 0.67	58.61 ± 0.92	73.98 ± 0.72
gcp-sampling with easy class	50.88 ± 0.88	68.22 ± 0.72	58.73 ± 1.14	73.41 ± 0.76
gcp-sampling with uncertain class	51.73 ± 0.87	69.01 ± 0.72	59.43 ± 1.02	73.84 ± 0.82
gcp-sampling with hard class	$\mathbf{52.34} \pm 0.81$	$\mathbf{69.21} \pm 0.68$	$\mathbf{60.14} \pm 0.97$	$\mathbf{74.58} \pm 0.74$

Conclusion

- An adaptive task sampling method for meta-learning
- It is essential for the sampling process to be dependent on tasks
- The proposed method could be applied to any meta-learning algorithms that follow episodic training

Thanks for watching !