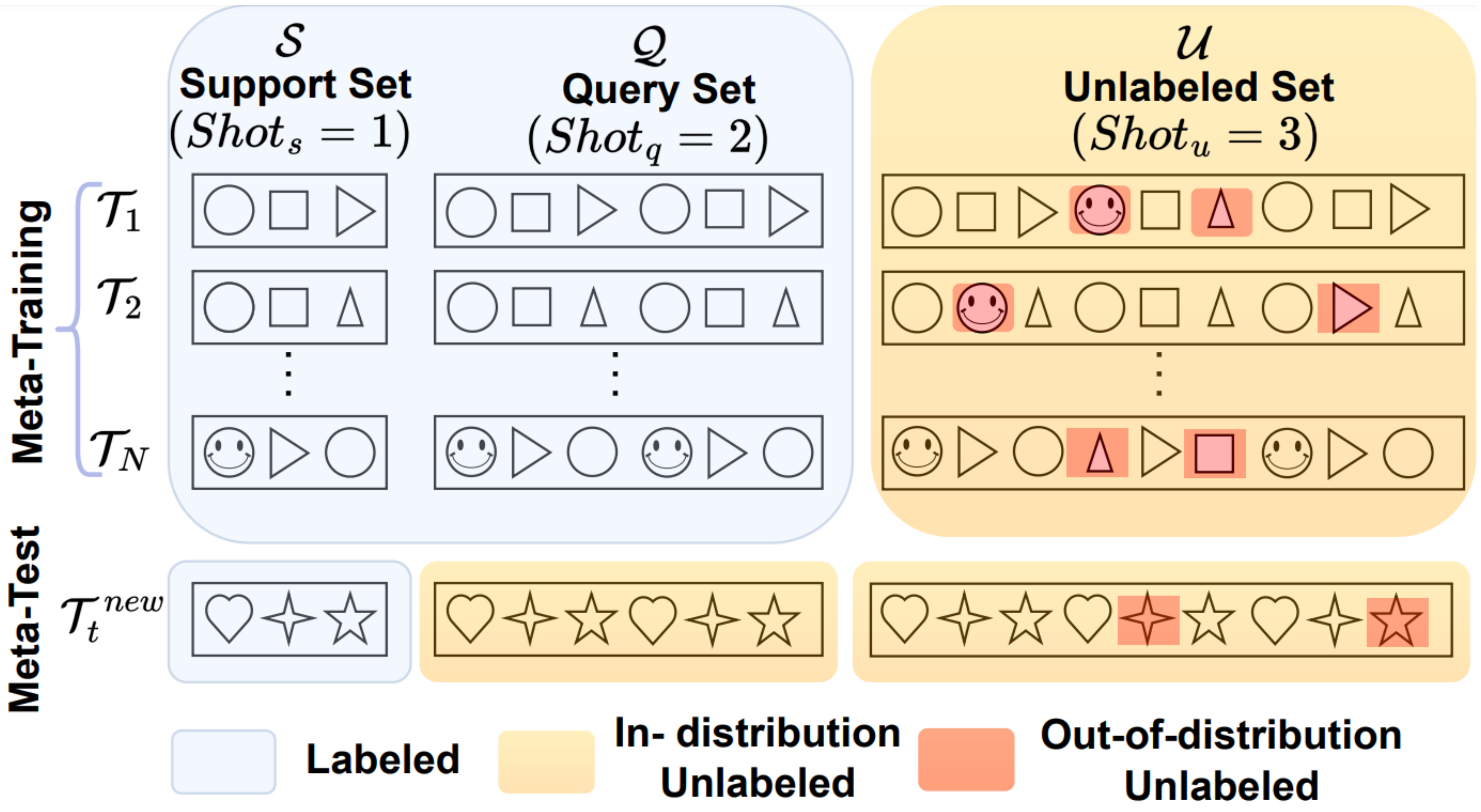
PLATINUM: Semi-Supervised Model Agnostic Meta-Learning using Submodular Mutual Information

Changbin Li*, Suraj Kothawade*, Feng Chen, Rishabh Iyer

ICML 2022
July 2022, Baltimore MD

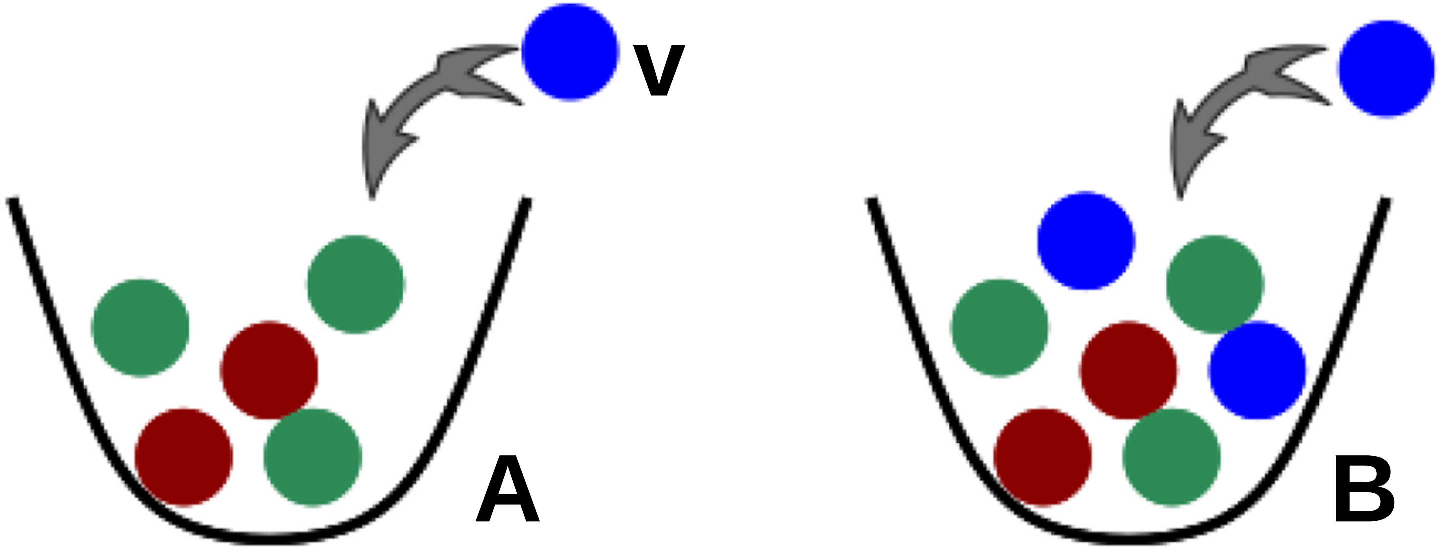
*Equal Contribution

Semi-supervised Few-shot Learning



Submodular Functions

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B$$

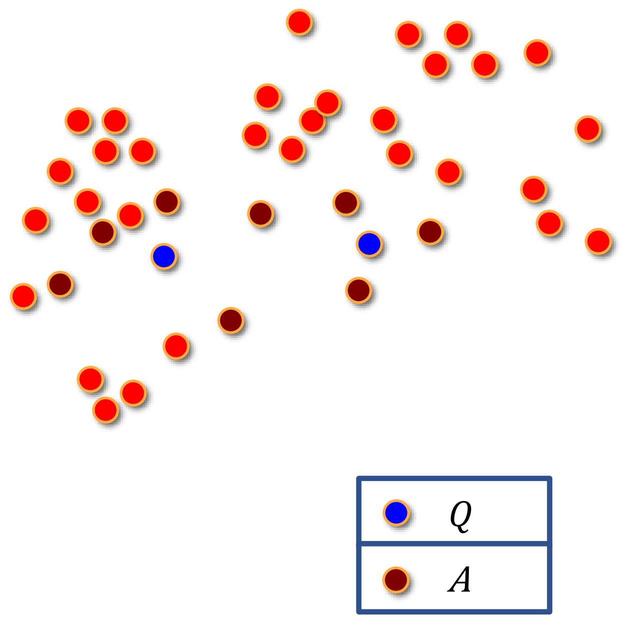


$f = \#$ of distinct colors of balls in the urn.

How to select subsets from unlabeled data to augment each task?

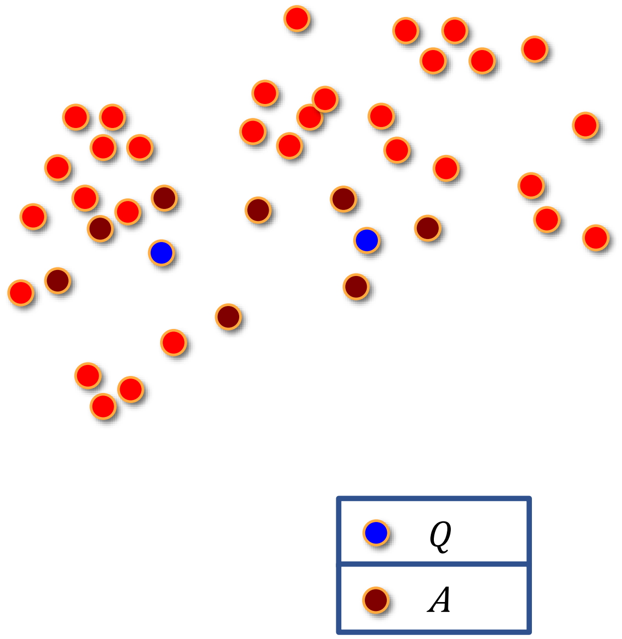
Submodular Mutual Information (SMI)

- Given a set of data points $V = \{1, \cdots, n\}$, and sets $A, Q \subseteq V$, the Submodular Mutual Information $I_F(A; Q) = F(A) + F(Q) - F(A \cup Q)$, where the information of a **set** of points is $F(A)$ and F is a submodular function.



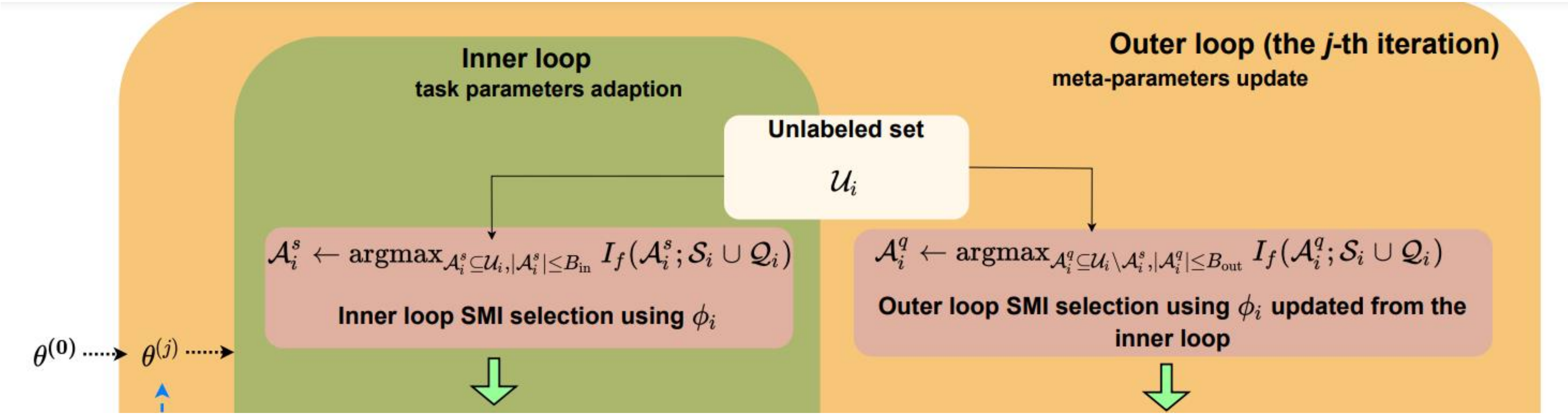
Submodular Mutual Information (SMI)

- Given a set of data points $V = \{1, \cdots, n\}$, and sets $A, Q \subseteq V$, the Submodular Mutual Information $I_F(A; Q) = F(A) + F(Q) - F(A \cup Q)$, where the information of a **set** of points is $F(A)$ and F is a submodular function.



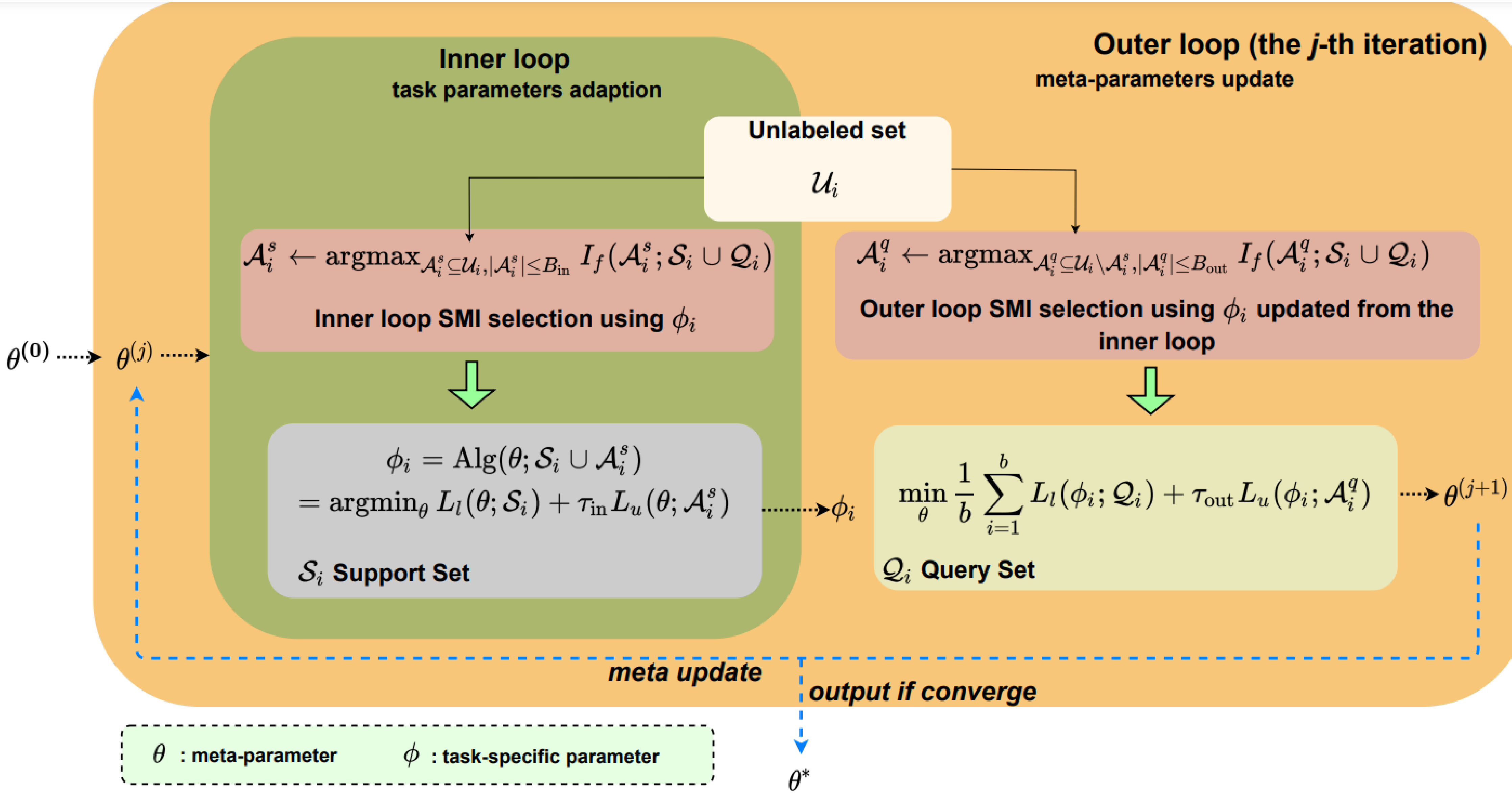
Name	$I_F(A; Q)$
Graph Cut MI **(GCMI)**	$2 \sum_{i \in A} \sum_{j \in Q} s_{ij}$
Facility Location MI (FLMI)	$\sum_{i \in Q} \max_{j \in A} s_{ij} + \eta \sum_{i \in A} \max_{j \in Q} s_{ij}$

Overview of PLATINUM



- ❖ Use Submodular Mutual Information (SMI) for semi-supervision.
- ❖ Augment both Support and Query sets in Inner and Outer loop of MAML.
- ❖ Support and Query sets are augmented using per-class instantiations of SMI.

Overview of PLATINUM



Experimental Setting

Datasets

- *mini*ImageNet, *tiered*ImageNet, CIFAR-FS

Semi-supervised few-shot classification (Ren et al., 2018)

- 5-way 1-shot (5-shot)
- backbones: 4-layer CONV (for all approaches)
- Two scenarios
  - There exist OOD examples in unlabeled set
  - There’s no OOD examples in unlabeled set
- Smaller ρ $(1\%, 10\%, 20\%, \dots)$, $\rho = \dfrac{Count(labeled\ examples\ per\ class)}{Count(Total\ examples\ per\ class)}$

  ($\rho = 40\%$ for *mini*ImageNet, 10% for *tiered*ImageNet in Ren et al., 2018)

Experiments

PLATINUM (ours)

- SMI functions: GCMI, FLMI
- On the top of first-order MAML

Meta-learning based baselines:

- Extended prototypical network (Ren et al., 2018)
- TPN-semi (Liu et al., 2019)
- LST (Li et al., 2019)
- MAML: only supervised setting is considered.

Note: we did not consider transfer-learning based approaches for fair comparison.

5-way classification accuracy

- *mini*ImageNet
- $\rho = 1\%$

Methods	1-shot w/o OOD	1-shot w/ OOD	5-shot w/o OOD	5-shot w/ OOD
Soft k-Means (Ren et al., 2018)	24.61±0.64	23.57±0.63	38.20±1.64	38.07±1.53
Soft k-Means+Cluster (Ren et al., 2018)	15.76±0.59	9.77±0.51	33.65±1.53	30.47±1.42
Masked Soft k-Means (Ren et al., 2018)	25.48±0.67	25.03±0.68	39.33±1.55	38.48±1.74
TPN-semi (Liu et al., 2019)	40.25±0.92	26.70±0.98	46.27±1.67	36.81±0.87
LST(*small*) (Li et al., 2019)	37.65±0.78	37.82±0.91	61.50±0.92	57.67±0.85
LST(*large*) (Li et al., 2019)	41.36±0.98	39.32±0.95	61.51±0.98	59.24±0.95
MAML† (Finn et al., 2017)	35.26±0.85	35.26±0.85	60.22±0.83	60.20±0.83
VAT (Miyato et al., 2018)	36.55±0.86	34.03±0.84	61.60±0.83	61.24±0.88
PL (Lee et al., 2013)	37.71±0.94	35.16±0.85	60.64±0.92	60.31±0.87
GCMI (ours)	41.94±0.96	**42.57**±0.93	63.62±0.95	**63.54**±0.94
FLMI (ours)	**42.27**±0.95	41.53±0.97	**63.80**±0.92	63.44±0.99

- *tired*ImageNet
- $\rho = 1\%$

Methods	1-shot w/o OOD	1-shot w/ OOD	5-shot w/o OOD	5-shot w/ OOD
Soft k-Means (Ren et al., 2018)	27.53±0.74	27.04±0.76	44.63±1.19	44.78±1.05
Soft k-Means+Cluster (Ren et al., 2018)	30.48±0.84	31.30±0.86	46.93±1.18	49.33±1.17
Masked Soft k-Means (Ren et al., 2018)	33.85±0.84	32.99±0.87	47.63±1.12	47.35±1.08
TPN-semi (Liu et al., 2019)	44.13±1.04	31.83±1.09	58.53±1.57	56.92±1.67
LST(*small*) (Li et al., 2019)	42.86±0.86	42.33±0.95	59.55±0.92	58.82±0.93
LST(*large*) (Li et al., 2019)	44.34±0.97	44.59±0.99	61.45±0.90	60.75±0.93
MAML† (Finn et al., 2017)	41.96±0.84	41.96±0.84	61.30±0.85	61.30±0.85
VAT (Miyato et al., 2018)	41.52±0.82	41.51±0.79	59.98±0.83	60.01±0.87
PL (Lee et al., 2013)	41.22±0.89	40.87±0.83	61.70±0.77	60.57±0.87
GCMI (ours)	45.49±0.91	45.55±0.90	63.67±0.83	**62.59**±0.85
FLMI (ours)	**45.63**±0.86	**46.19**±0.94	**63.75**±0.87	62.19±0.91

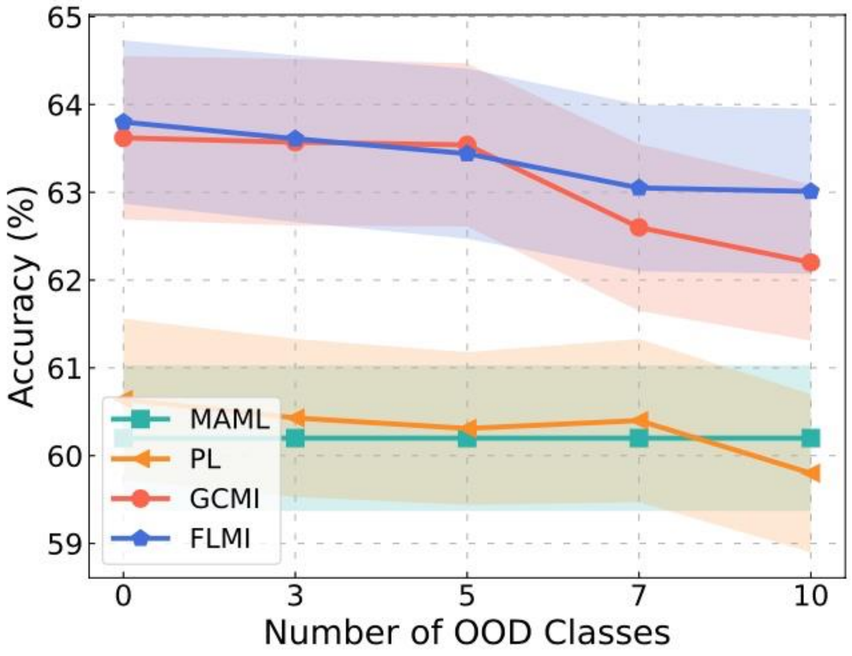
5-way classification accuracy

- *mini*ImageNet
- $\rho = 40\%$, exactly the same setting as previous works.

Methods	1-shot w/o OOD	1-shot w/ OOD	5-shot w/o OOD	5-shot w/ OOD
Soft k-Means (Ren et al., 2018)	50.09±0.45	48.70±0.32	64.59±0.28	63.55±0.28
Soft k-Means Cluster (Ren et al., 2018)	49.03±0.24	48.86±0.32	63.08±0.18	61.27±0.24
Masked Soft k-Means (Ren et al., 2018)	50.41±0.31	49.04±0.31	64.39±0.24	62.96±0.14
TPN-semi (Liu et al., 2019)	**52.78**±0.27	**50.43**±0.84	66.42±0.21	64.95±0.73
GCMI (*large*, ours)	**51.35**±0.93	**50.85**±0.89	**66.65**±0.75	**66.66**±0.74
FLMI (*large*, ours)	51.06±0.96	49.83±0.91	**67.34**±0.72	**66.20**±0.73

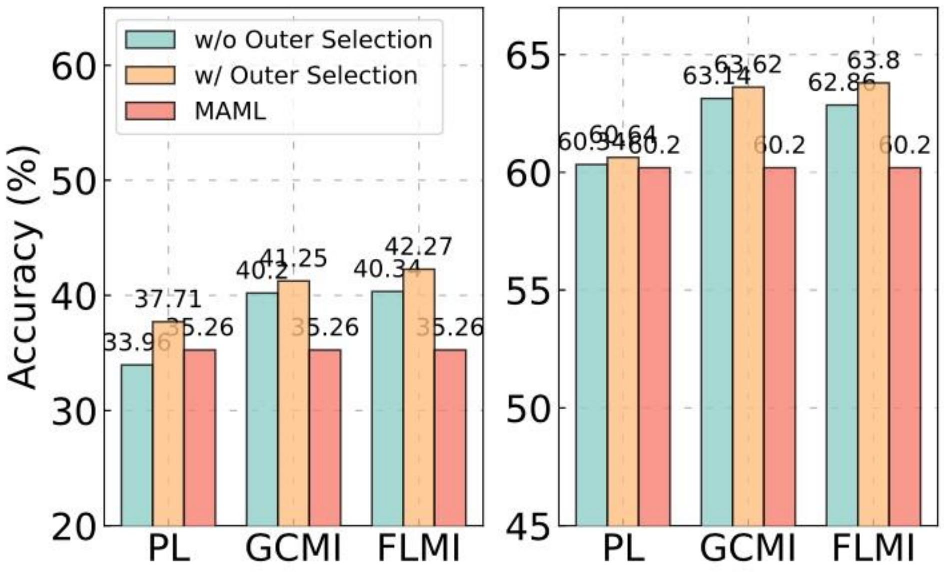
Ablation

- Different number of OOD classes



Comparison under different number of OOD classes in the Unlabeled Set for 5-shot case on *mini*ImageNet

- *w/* vs. *w/o* outer selection



Left: 1-shot, Right: 5-shot.
Both of them are on *mini*ImageNet.

Ablation

Other Backbones?

The accuracy (%) of 5-way 5-shot experiment

- on *mini*ImageNet
- Pretrained ResNet-12
- $\rho = 40\%$ (the same ratio from Ren et al., 2018 and Li et al., 2019)

MAML	LST (Li et al., 2019)	GCMI (*large*, ours)
$75.21_{\pm 0.65}$	$78.70_{\pm 0.80}$	$79.44_{\pm 0.76}$

Conclusion

- PLATINUM: A novel semi-supervised model-agnostic meta-learning framework.
- It leverages submodular mutual information functions as per-class acquisition functions to select more data from unlabeled data in the inner and outer loop of meta-learning.
- Meta-learning based semi-supervised few-shot learning experiments validates the effectiveness of embedding semi-supervision on the top of first-order MAML, especially for small ratio of labeled to unlabeled samples.

Thank You

For more details, do visit our poster.