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

THE UNIVERSITY OF TEXAS AT DALLAS

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

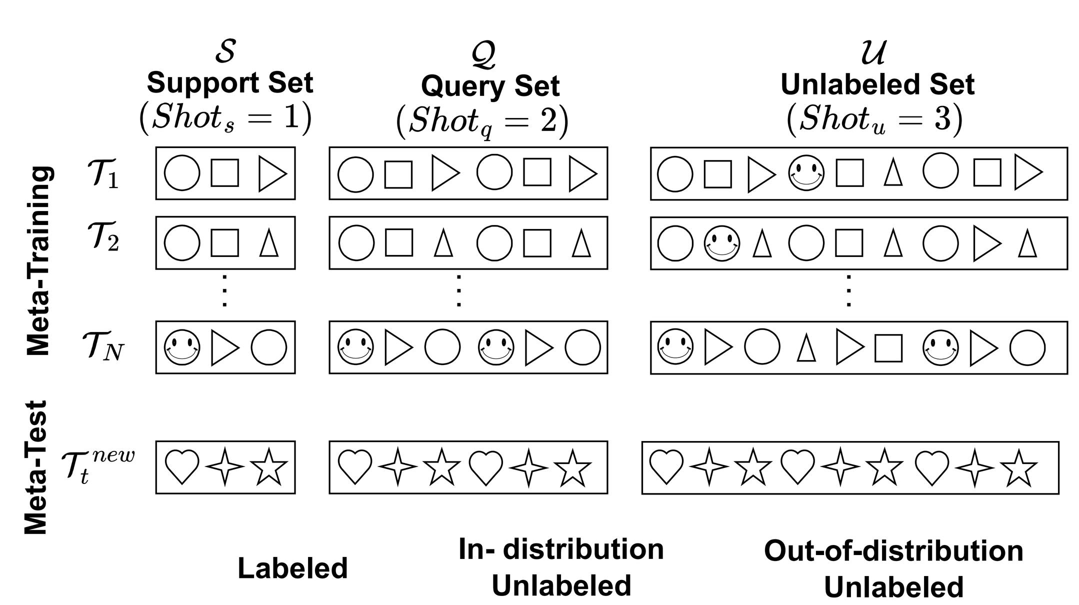
ICML International Conference On Machine Learning

INTRODUCTION

- Unlabeled examples could also be used to boost the performance in semi-supervised learning. Similarly can we use unlabeled information in semi-supervised few shot learning (SSFSL)?
- Current SSFSL: 1) transfer-learning based (pretrained features); 2) meta-learning based.
- Submodular information measures have been used as acquisition functions for active learning in scenarios with class imbalance, redundancy and OOD data.
- In this work, we propose PLATINUM (semi-su**P**ervised mode**L** **A**gnostic me**T**a learn**I**ng usi**N**g s**U**bmodular **M**utual information), a novel semi-supervised model agnostic meta learning framework that uses the submodular mutual information (SMI) functions to boost the performance of FSC.
- We study the performance of PLATINUM in two scenarios: 1) where the unlabeled data points belong to the same set of classes as the labeled set of a certain episode. 2) where there exist out-of-distribution classes that do not belong to the labeled set.

MOTIVATION and SETUP

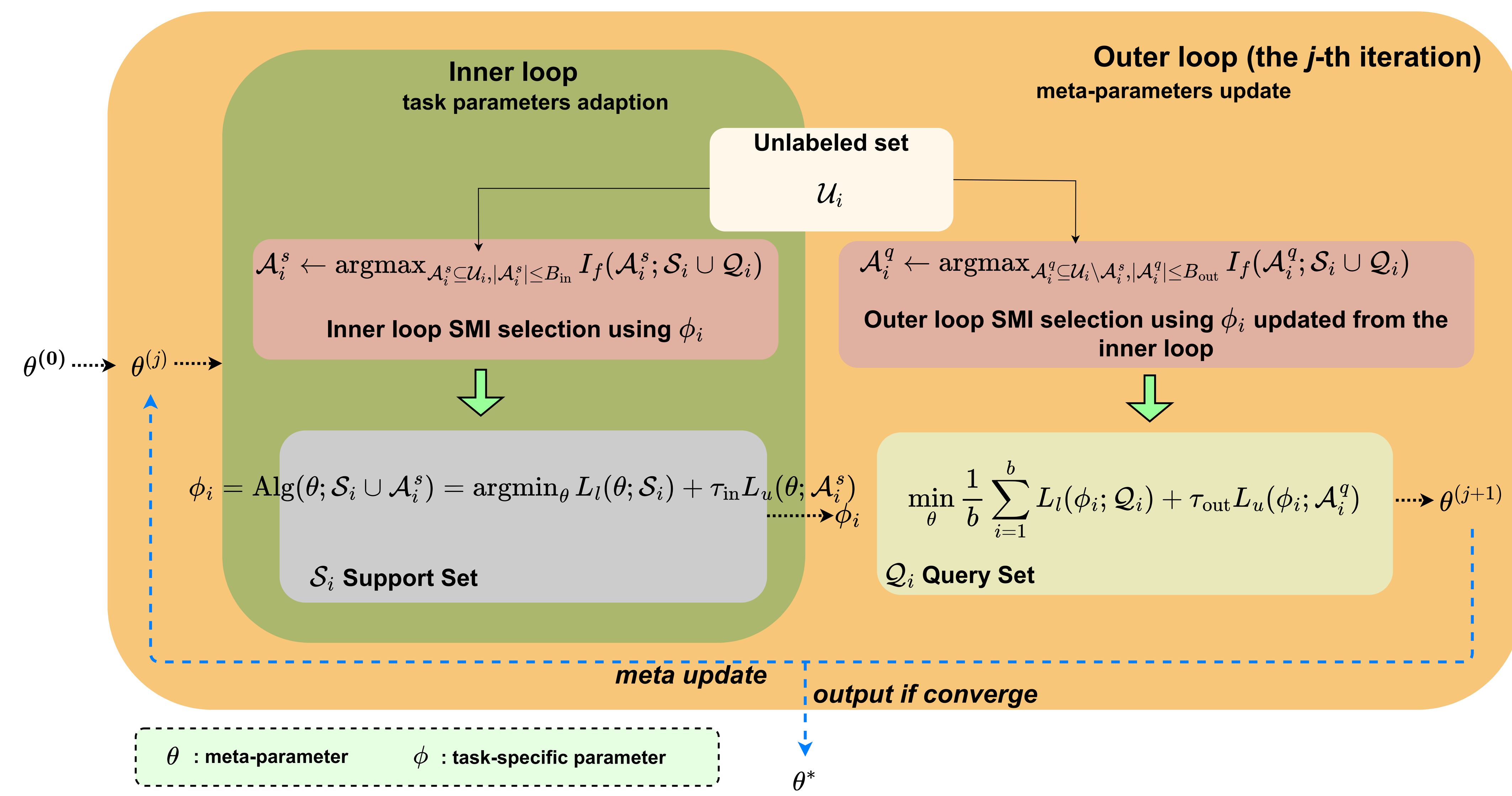
Goal: Leverage unlabeled dataset by embedding semi-supervision in the inner and outer loop of MAML.



Semi-supervised few-shot learning setup

- During meta-training, the goal is to iterate overs tasks $T_1 \cdots T_N$ and meta-learn a parametrization using the support set S, query set Q, and the unlabeled set U.
- During meta-testing, the learned parametrization is used as an initialization and a model is trained using the S and U to perform well on Q.
- In any task, U may contain data points that are out-of-distribution.

The PLATINUM Framework



- For a specific task T_i, in each inner loop and outer loop gradient update step, we select a subset from the unlabeled set by maximizing the per-class SMI function.
- In each inner loop step, the selected subset A_i^s and support set S_i are used to update model parameters ϕ_i.
- In the outer-loop of the meta-training stage, another subset A_i^q will be selected after inner loop selection according to the updated model parameters ϕ_i. Meta-parameters θ would be updated based on A_i^q and the query set Q_i.

RESULTS

Semi-Supervised Few-Shot Classification: *mini*ImageNet, $\rho = 0.01$.

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

*tiered*ImageNet, $\rho = 0.01$.

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

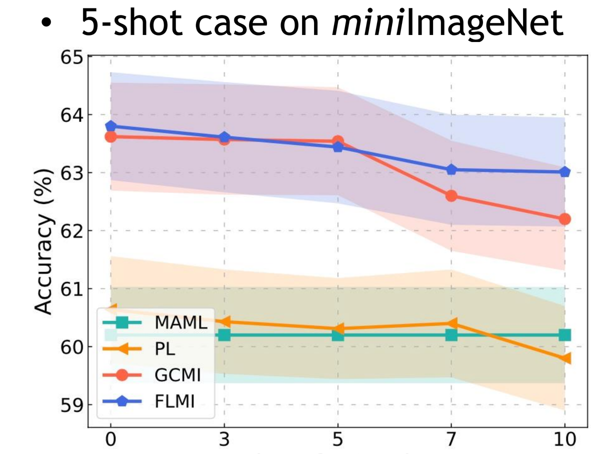
*mini*ImageNet, $\rho = 0.4$.

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

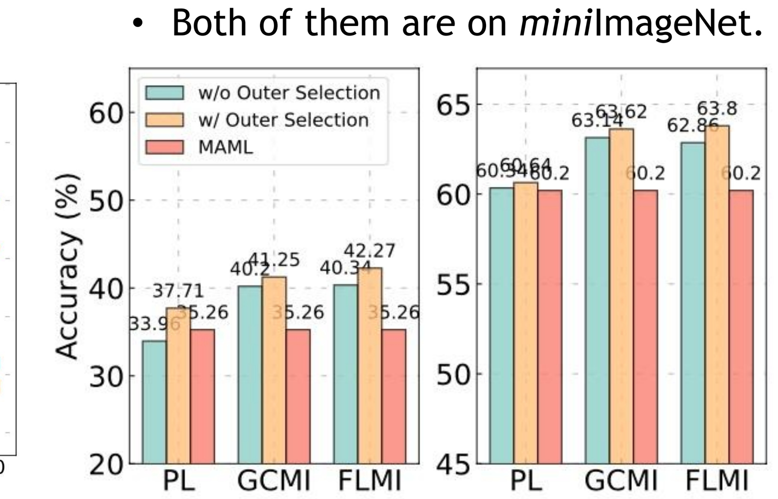
- Embedding semi-supervision on the top of first-order MAML could boost the performance.
- Especially for small ratio of labeled to unlabeled samples, also works for high labeled ratio.

Ablation

- Comparison under different number of OOD classes in the Unlabeled Set
- 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.



Semi-supervision in Inner and Outer Loop

for each task $\mathcal{T}_i = \{\mathcal{S}_i, \mathcal{Q}_i, \mathcal{U}_i\}, i \in [b]$ **do**
  Initialize model parameters $\phi_i \leftarrow \theta$
  for each inner step t **do**
    $\mathcal{P}_{\mathcal{U}_i} \leftarrow \phi_i(\mathcal{U}_i)$
    $\mathcal{X} \leftarrow \textsc{Cosine\_Similarity}(\mathcal{P}_{\mathcal{U}_i}, \{\mathcal{P}_{\mathcal{S}_i} \cup \mathcal{P}_{\mathcal{Q}_i}\})$
    Instantiate a submodular function f based on $\mathcal{X}$.
    / inner loop selection */*
    $\mathcal{A}_{it}^s \leftarrow \mathrm{argmax}_{\mathcal{A}_{it}^s \subseteq \mathcal{U}_i, |\mathcal{A}_{it}^s| \le B_{\text{in}}} I_f(\mathcal{A}_{it}^s; \mathcal{S}_i \cup \mathcal{Q}_i)$
    $\phi_i \leftarrow \phi_i - \nabla_\phi L(\theta; \mathcal{S}_i \cup \mathcal{A}_i^s)$
    $\mathcal{A}_i^s \leftarrow \mathcal{A}_i^s \cup \mathcal{A}_{it}^s$
  end for
  $\mathcal{P}_{\mathcal{U}_i \setminus \mathcal{A}_i^s} \leftarrow \phi_i(\mathcal{U}_i \setminus \mathcal{A}_i^s)$
  $\mathcal{X} \leftarrow \textsc{Cosine\_Similarity}(\mathcal{P}_{\mathcal{U}_i \setminus \mathcal{A}_i^s}, \{\mathcal{P}_{\mathcal{S}_i} \cup \mathcal{P}_{\mathcal{Q}_i}\})$
  */*outer loop selection*/*
  $\mathcal{A}_i^q \leftarrow \mathrm{argmax}_{\mathcal{A}_i^q \subseteq \mathcal{U}_i \setminus \mathcal{A}_i^s, |\mathcal{A}_i^q| \le B_{\text{out}}} I_f(\mathcal{A}_i^q; \mathcal{S}_i \cup \mathcal{Q}_i)$
end for
*/*meta update (outer loop) */*
Obtain $\theta^{(t+1)}$ by using $\{\mathcal{Q}_i \cup \mathcal{A}_i^q\}_{i=1}^b$

CONCLUSIONS

- 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 SSFSL experiments on the top of first-order MAML validates the effectiveness, especially for small ratio of labeled to unlabeled samples.
- Future work: involve some diversity measurements for the selected subset to do quantitative analysis

PAPER

SCAN ME

Get the paper for more technical details and results.