Tsinghua University

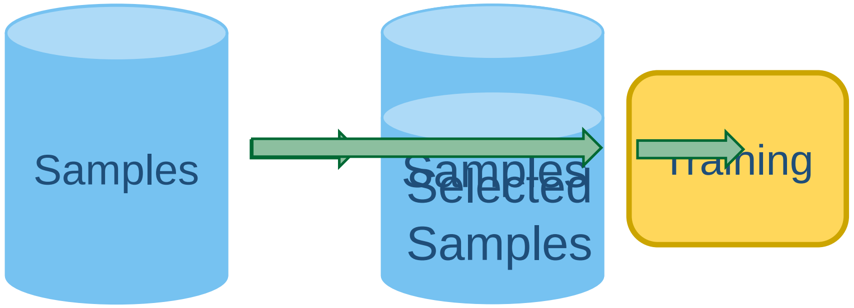
Structural-Entropy-Based Sample Selection for Efficient and Effective Learning

Tianchi Xie[1,*], Jiangning Zhu[1,*], Guozu Ma[2], Minzhi Lin[1],
Wei Chen[3], Weikai Yang[4], Shixia Liu[1]

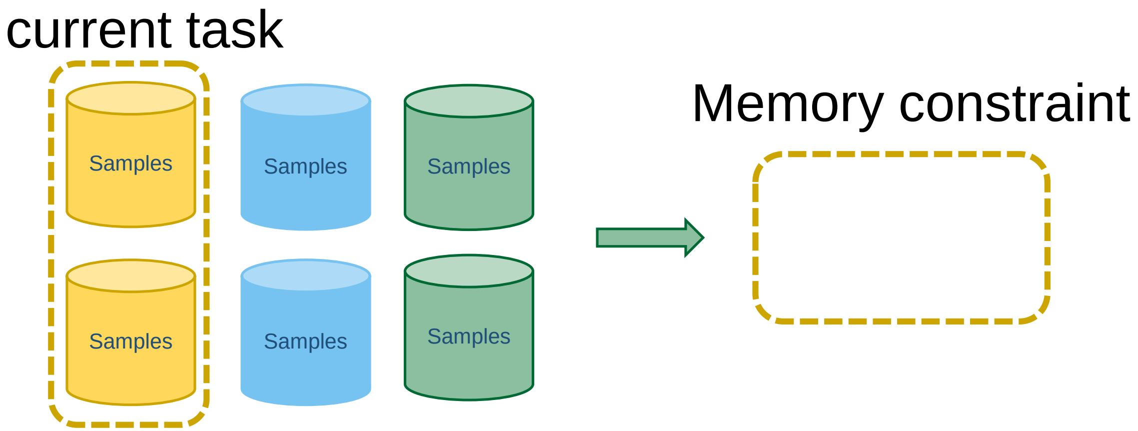
1 Tsinghua University
2 China Telecom Wanwei Information Technology Co., Ltd
3 Microsoft Research
4 Hong Kong University of Science and Technology (Guangzhou)

Motivation

- Data budgets that **limit sample sizes** are pervasive.



Select samples for
training efficiency



Select samples for
training effectiveness

Related Work

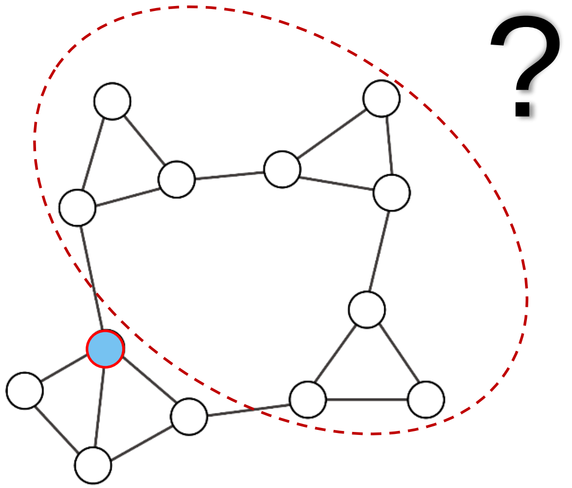
- Existing sample selection methods primarily utilize **local information**.

➢ Attribute-based methods:

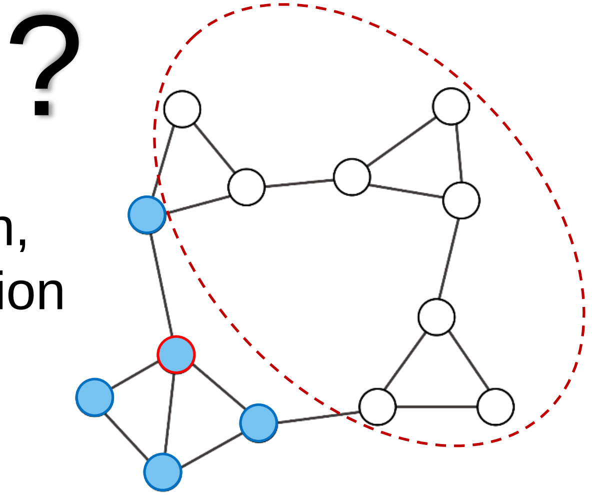
- EL2N (Paul et al., 2021)
- AUM (Pleiss et al., 2020)
- Forgetting (Toneva et al., 2019)

➢ Connection-based methods:

- $\mathbb{D}^2$ Pruning (Maharana et al., 2024)
- Moderate corset (Xia et al., 2023)
- GraphCut (Iyer et al., 2021)



Focus on local information,
but ignore global information



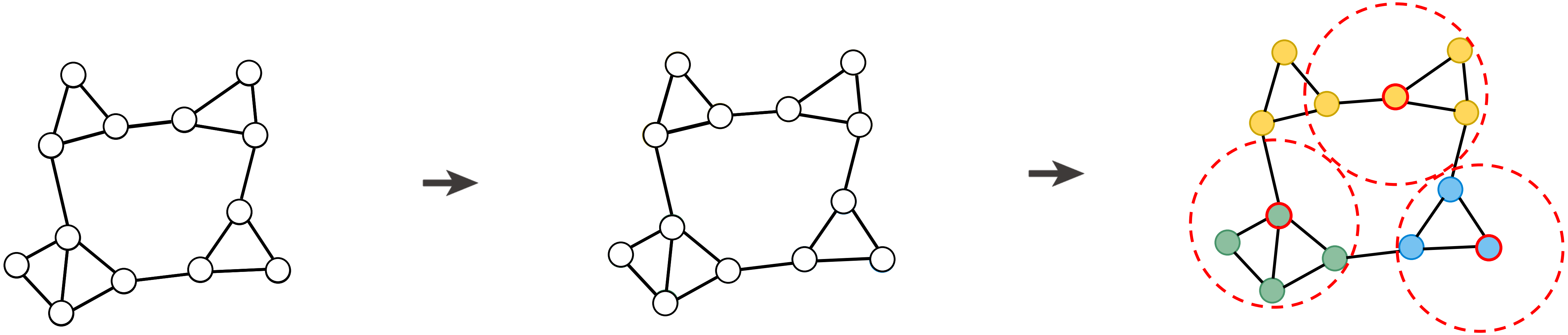
Challenges & Solution

- How to decompose the global information metric to the level of individual nodes?
  - A node-level structural entropy metric that quantifies the importance of nodes in preserving the global structure

- How to employ the node-level metric for sample selection?
  - A structural-entropy-based sample selection method that integrates both global and local metrics

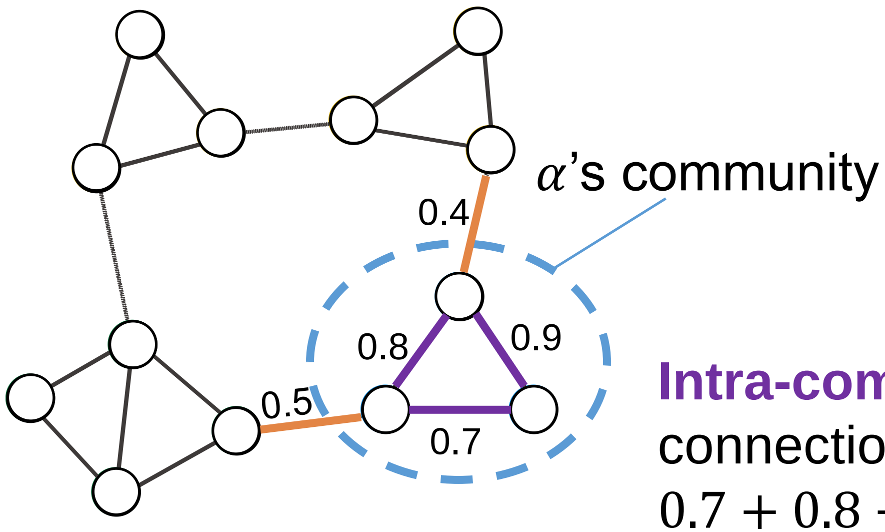
Overview

Graph Construction Node-Level Structural Entropy Importance-Biased Blue Noise Sampling

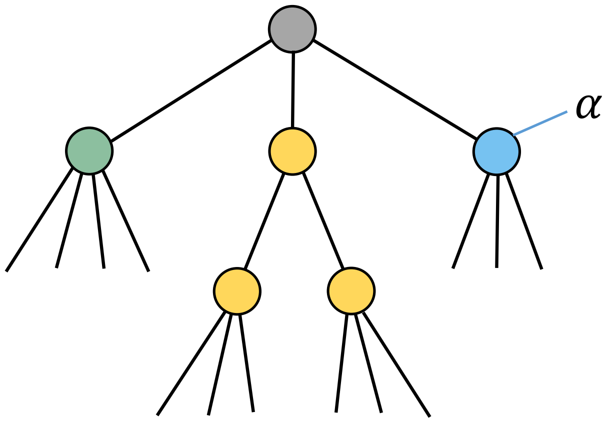


Structural Entropy

Graph G



Encoding Tree T



Intra-community connections $vol(\alpha) = 0.7 + 0.8 + 0.9$

Inter-community connections $g(\alpha) = 0.5 + 0.4$

Structural entropy: $\mathcal{H}(G, \mathcal{T}) = -\sum_{\alpha \in \mathcal{T}} \frac{g(\alpha)}{vol(V)} \log \frac{vol(\alpha)}{vol(\alpha^-)}$

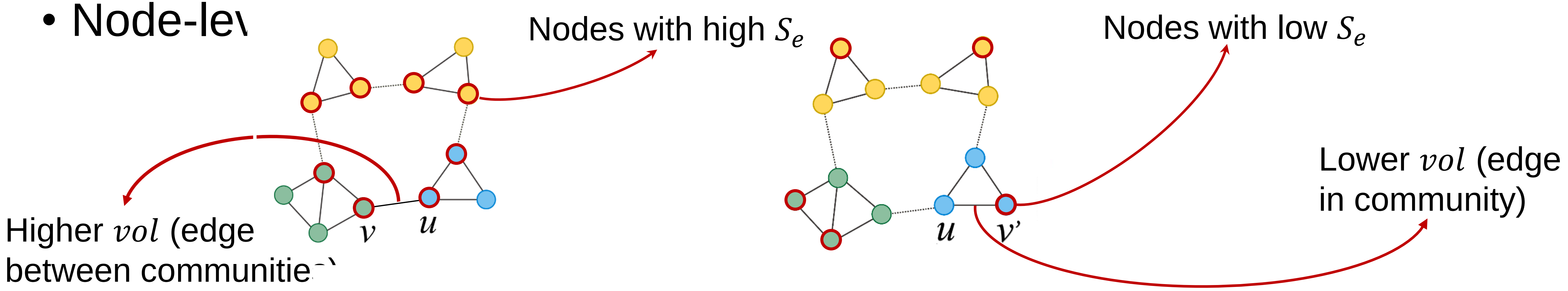
Only a single value for the whole graph!

Node-Level Structural Entropy

- Use the Shapley value to decompose structural entropy:

$$\phi(u) = \frac{1}{|V|} \overset{\text{Intractable}}{\boxed{\sum_{V_s \subseteq V \setminus \{u\}}}} \binom{|V|-1}{|V_S|}^{-1} \Big(\mathcal{H}(G[V_S \cup \{u\}], \mathcal{T}) - \mathcal{H}(G[V_S], \mathcal{T}) \Big)$$

- Node-lev



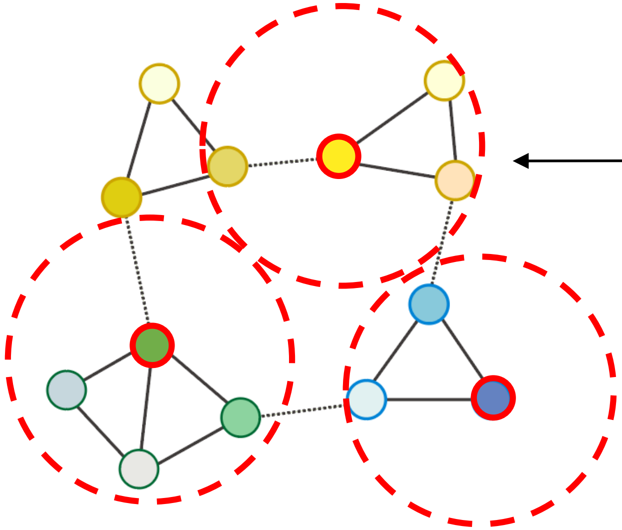
Importance-Biased Blue Noise Sampling

- Importance score:

$$S(u) = \underbrace{S_e(u)}_{\substack{\text{Node-level} \\ \text{structural entropy} \\ \text{(global metric)}}} \cdot \underbrace{S_t(u)}_{\substack{\text{Training} \\ \text{difficulty} \\ \text{(local metric)}}}$$

- Blue noise sampling:

Rejection region



Enhance diversity by rejecting similar samples

Experiment

- Three learning scenarios:
  - Supervised learning: Image classification, text classification, object detection, and VQA
  - Active learning: Image classification
  - Continual learning: Image classification
- Baselines: Random, AUM, EL2N, Forgetting, CCS, $\mathbb{D}^2$ pruning…

(a) ImageNet-1K

Dataset / Sampling rate	ImageNet-1K (100%:73.63) 70%	50%	20%	10%	5%	2%	1%
Random	71.12	69.43	58.77	47.36	33.41	16.41	5.41
Moderate	71.48	68.68	56.35	42.29	26.77	11.37	4.22
CCS	71.46	69.50	58.85	45.06	28.02	9.03	2.33
GraphCut	71.50	69.16	56.08	40.99	24.30	7.90	2.46
$\mathbb{D}^2$ Pruning	71.62	69.02	59.65	45.97	28.08	14.24	4.79
Prototypicality	70.12	66.00	49.20	35.27	24.14	13.88	4.95
SES (Ours)	**72.11**	**70.15**	**60.22**	**48.10**	**34.82**	**17.97**	**6.69**

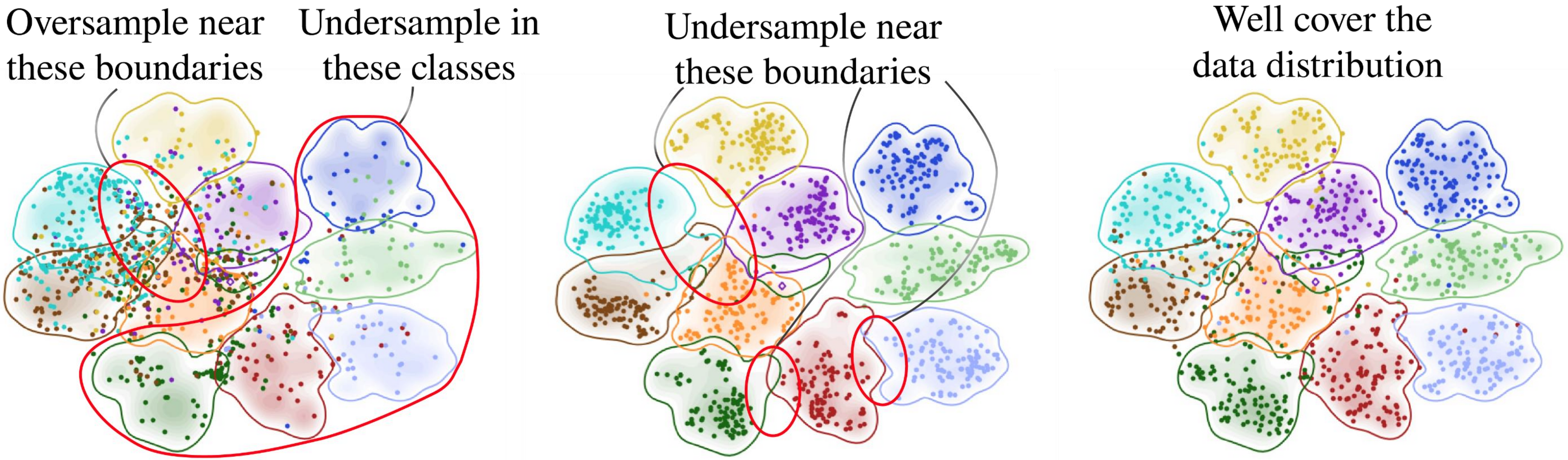
(b) ANLI

Dataset / Sampling	... 2%	1%
Random	7.50	35.96
Moderat	5.83	33.91
CCS	7.41	36.82
$\mathbb{D}^2$ Pruni	7.92	36.29
GraphCu	5.44	34.02
Entropy	6.69	36.40
Forgettin	5.78	35.03
EL2N	4.33	34.27
AUM	4.16	33.62
Variance	3.50	33.17
k-means	6.74	36.11
k-DPP	7.44	36.66
SES (Our	9.88	38.16

Dataset / Memory size	Permuted MNIST 100	Permuted MNIST 200	Split MNIST 100	Split MNIST 200	Split CIFAR10 100	Split CIFAR10 200	Split CIFAR100 100	Split CIFAR100 200	Split Tiny-ImageNet 100	Split Tiny-ImageNet 200
$\mathbb{D}^2$ Pruning	78.25	79.94	96.79	97.69	64.54	66.08	51.86	54.50	19.08	19.50
GraphCut	76.98	78.61	91.34	94.25	61.02	61.66	53.35	54.66	19.76	20.53
k-center	78.17	79.75	94.39	96.61	61.47	62.74	51.16	53.10	18.87	18.90
Gradient Matching	77.30	79.27	95.39	97.54	61.65	62.65	54.13	56.29	19.19	19.00
FRCL	77.33	79.21	94.48	97.10	61.67	62.93	51.40	54.28	18.86	19.01
iCaRL	78.94	80.65	89.50	97.59	62.33	64.08	54.62	56.11	19.58	19.85
Greedy Coreset	78.71	80.13	96.07	97.76	63.18	62.98	56.17	57.72	19.24	19.98
BCSR	77.74	79.51	94.77	96.98	63.23	64.59	50.21	51.49	18.75	18.74
SES (Ours)	**79.92**	**81.18**	**96.94**	**98.28**	**68.26**	**69.32**	**57.60**	**59.69**	**20.80**	**21.20**

1%
5.11
6.54
5.59
6.29
6.54
4.60
5.60
6.13
7.61

Qualitative Analysis



AUM $\mathbb{D}^2$ pruning Ours

Conclusion and Future Work

- A structural-entropy-based sample selection method
  - Decompose graph-level structural entropy to the node level
  - Combine global metric and local metric to select informative and representative samples

- Future Work
  - Automate the hyperparameter selection
  - Improve the support for multimodal data

Tsinghua University

Thank you!

Structural-Entropy-Based Sample Selection for Efficient and Effective Learning

Tianchi Xie[1,*], Jiangning Zhu[1,*], Guozu Ma[2], Minzhi Lin[1], Wei Chen[3], Weikai Yang[4], Shixia Liu[1]

1 Tsinghua University
2 China Telecom Wanwei Information Technology Co., Ltd
3 Microsoft Research
4 Hong Kong University of Science and Technology (Guangzhou)