KAUST

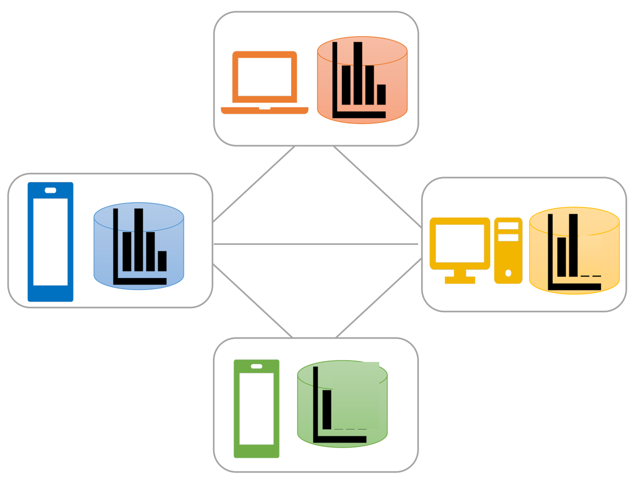
SANDS Lab

A First Look at the Impact of Distillation Hyper-Parameters in Federated Knowledge Distillation

Norah Alballa, Marco Canini

JOINT DISTILLATION FOR DECENTRALIZED TRAINING

Joint knowledge distillation



- transfer knowledge between already pre-trained models
- suitable for federated / decentralized training setting (P2P)

Pros:

+ communication reduction
+ mitigate system and data heterogeneity
+ model architecture flexibility

Cons:

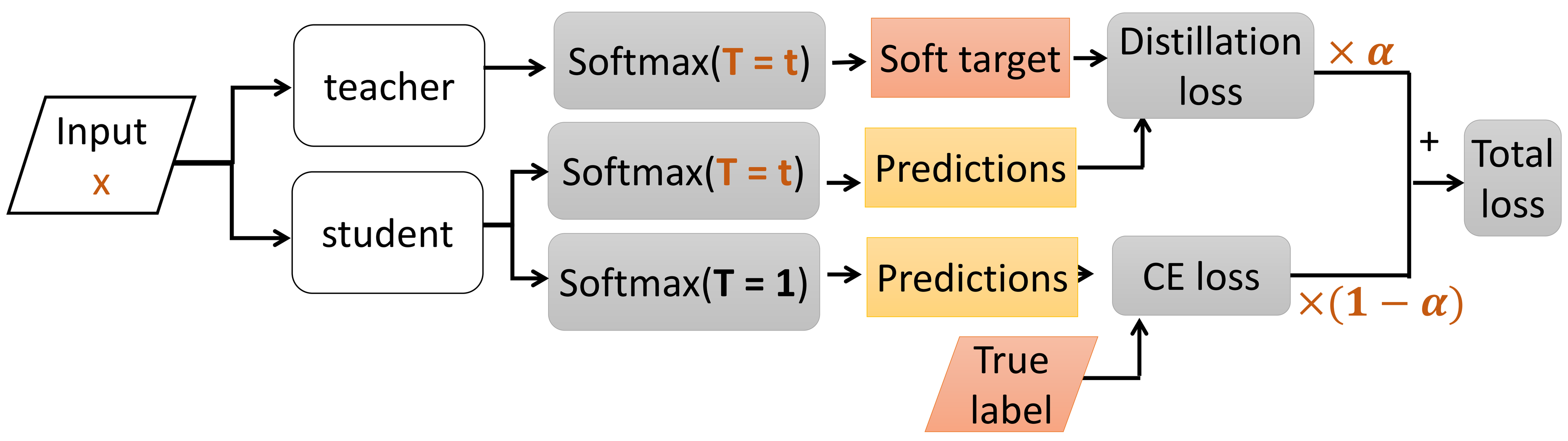
- setting not well studied, default hyper-params not effective
- KD is comp. intensive and not always an accuracy gain

How sensitive is performance to hyper-params?

What's the room for tuning them?

KNOWLEDGE DISTILLATION PIPELINE

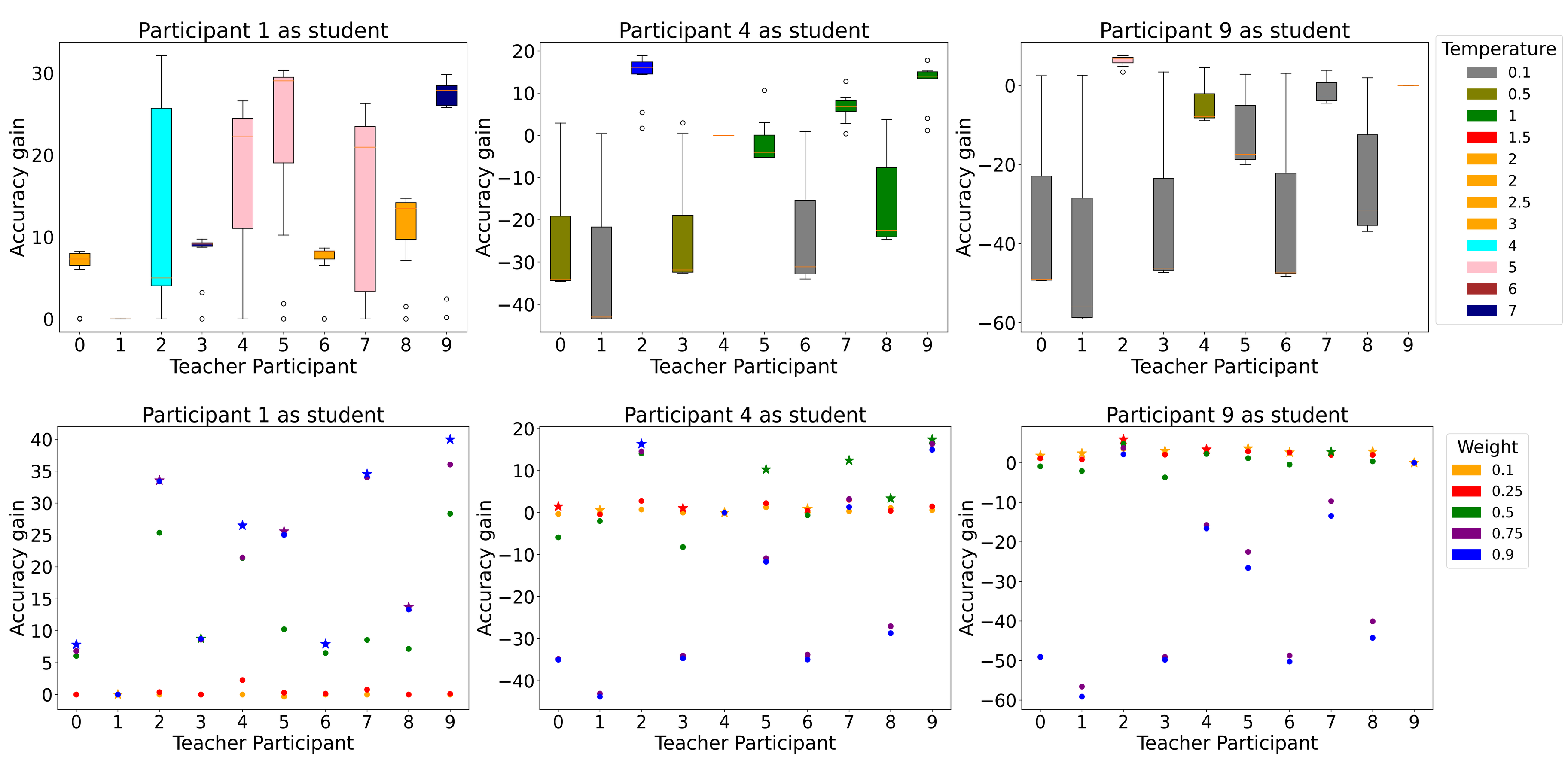
Hyper-params: temperature T, weight α, transfer set (input x), and position (teacher vs. student)



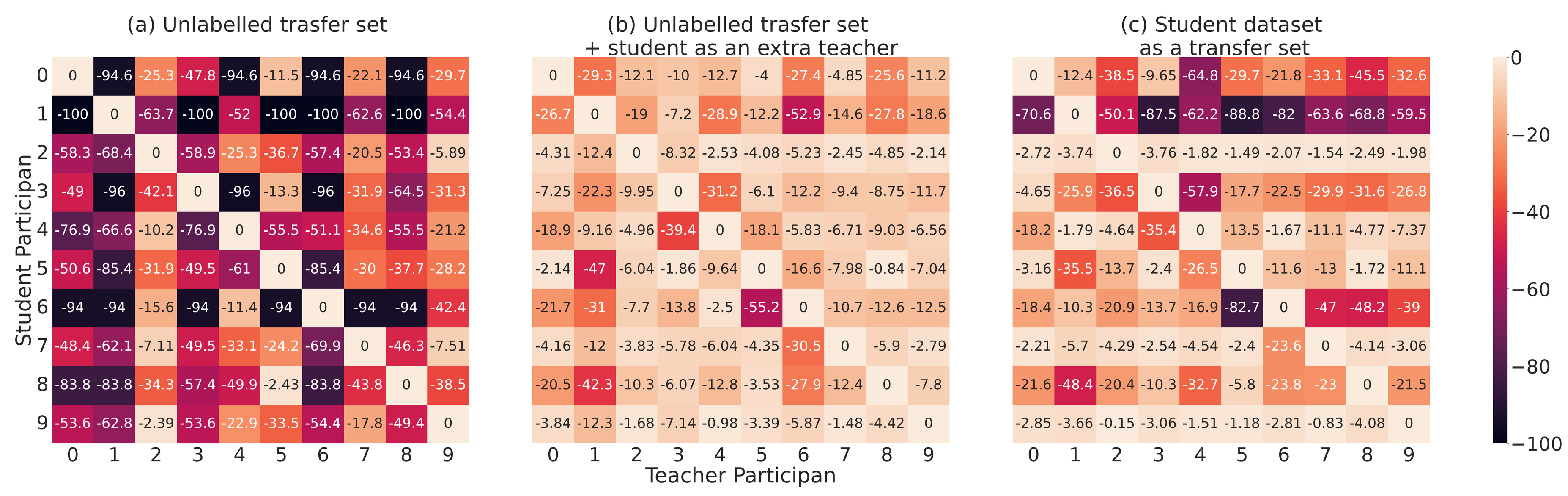
SETUP

- Simple scenario: 10 participants, ResNet-18 on CIFAR10, heterogeneous data partitions
- One round joint KD for all possible pairwise interactions
- Full sweep on space of hyper-params

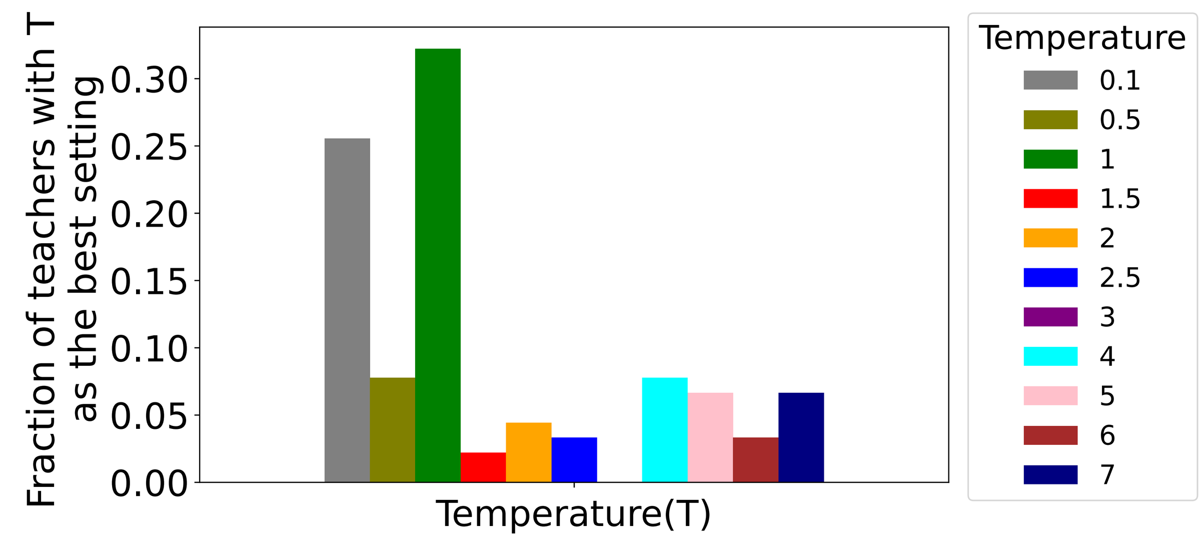
NO SINGLE SETTING IS THE BEST FOR ALL

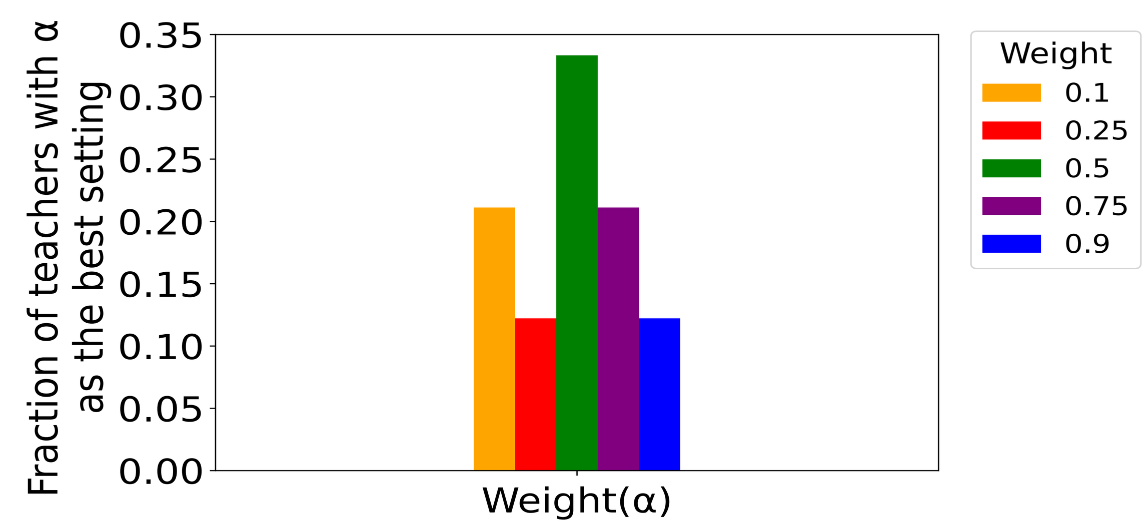


MAIN OBSERVATIONS



Transfer set affects student's forgetting





2/3 of the pairs require temperature and weight other than the common default T=1 and α =0.5

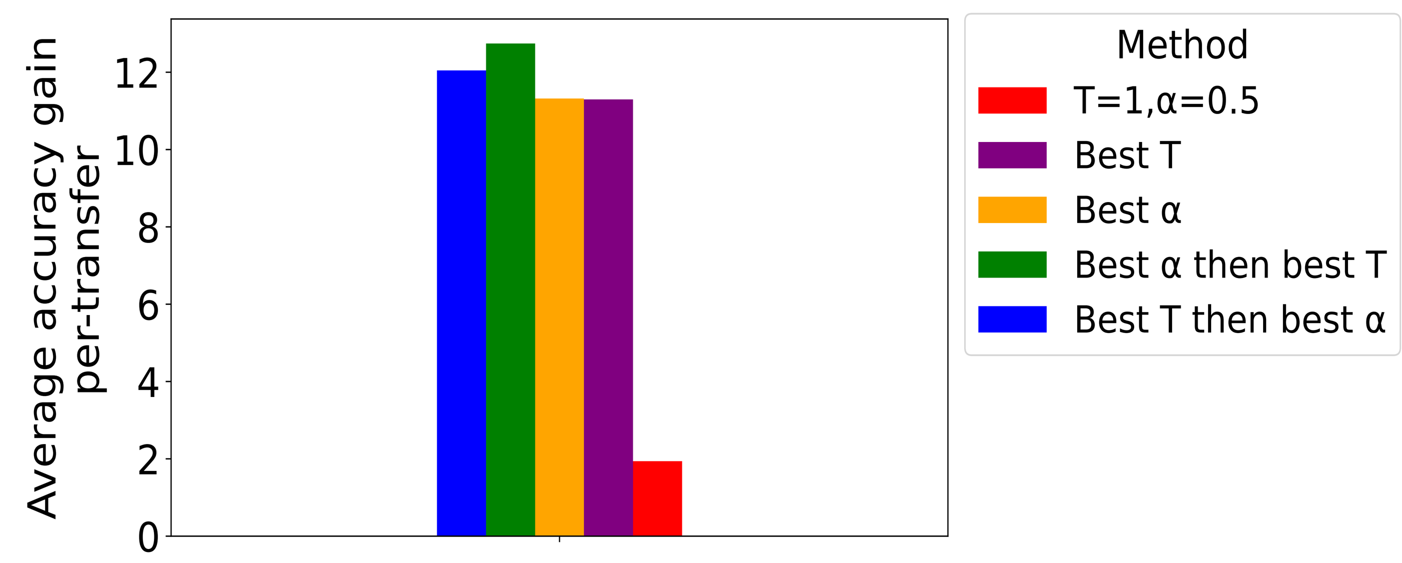
With participant:

	Participant 1		Participant 4		Participant 9	
	As student	**As teacher**	**As student**	**As teacher**	**As student**	**As teacher**
0	8.22	11.70	3.31	-9.74	3.85	-13.47
1	-	-	1.26	-11.52	3.59	-17.87
2	42.65	67.16	19.86	24.37	7.09	8.92
3	9.59	10.51	2.79	-13.39	4.00	-11.27
4	32.27	45.05	-	-	4.46	0.30
5	28.61	33.97	8.91	0.07	4.21	-3.19
6	8.33	9.88	1.07	-12.57	3.86	-14.58
7	37. 10	48.02	10.57	10.50	3.30	-0.05
8	14.06	16.09	0.77	-4.92	3.91	-10.64
9	41.25	62.71	15.63	19.80	-	-

Setting the right position is important

APPROPRIATE TUNING

Up to 5+ times improvement on average



TAKE AWAY

- Unlike offline and online distillation, requires careful tuning. Appropriate tuning in joint distillation can significantly improve the accuracy gain.
- Automating the tuning is an important direction.