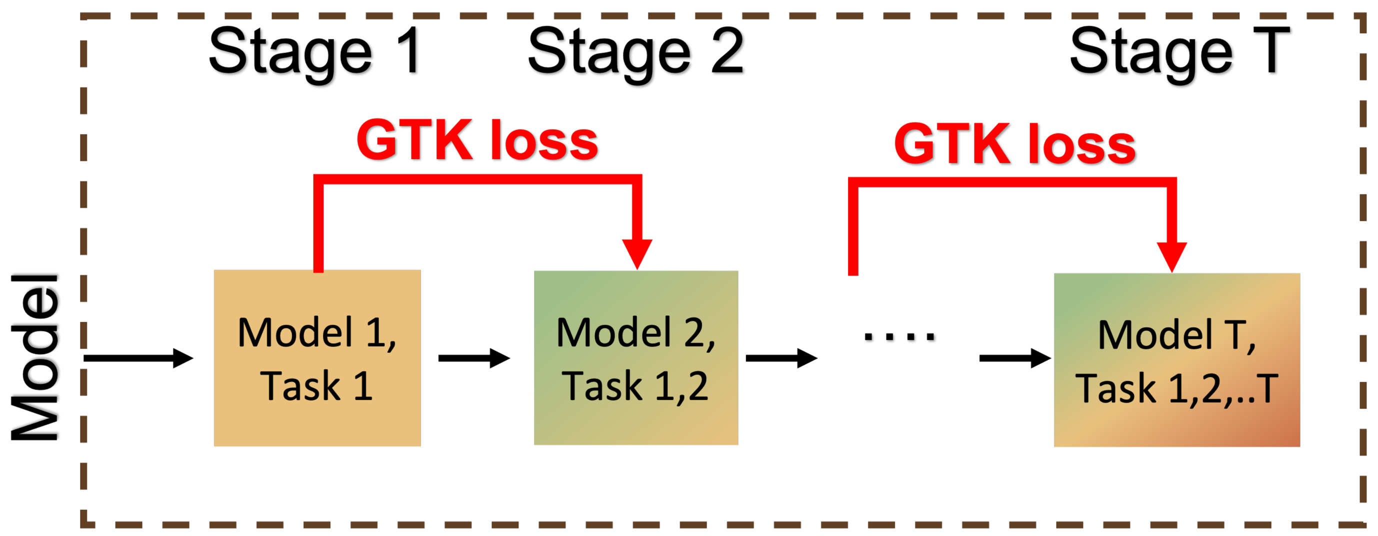
TKIL: Tangent Kernel Optimization for Class Balanced Incremental Learning

Jinlin Xiang, Eli Shlizerman



Overview:

TKIL is a Tangent Kernel optimization for optimal class-balanced Incremental Learning with a novel **Gradient Tangent Kernel** loss.

Motivation:

> Learning balanced Data representations is crucial for Class Incremental Learning (CIL).
> We proposed learning a set of task-specific parameters instead of a single fixed model.
> Tangent kernel loss alleviates catastrophic forgetting of previous representations.

Highlight

> Address data imbalances in memory-based Class Incremental Learning
> A novel gradient tangent kernel loss for finite-width neural networks
> Low training computational load and constant memory usage

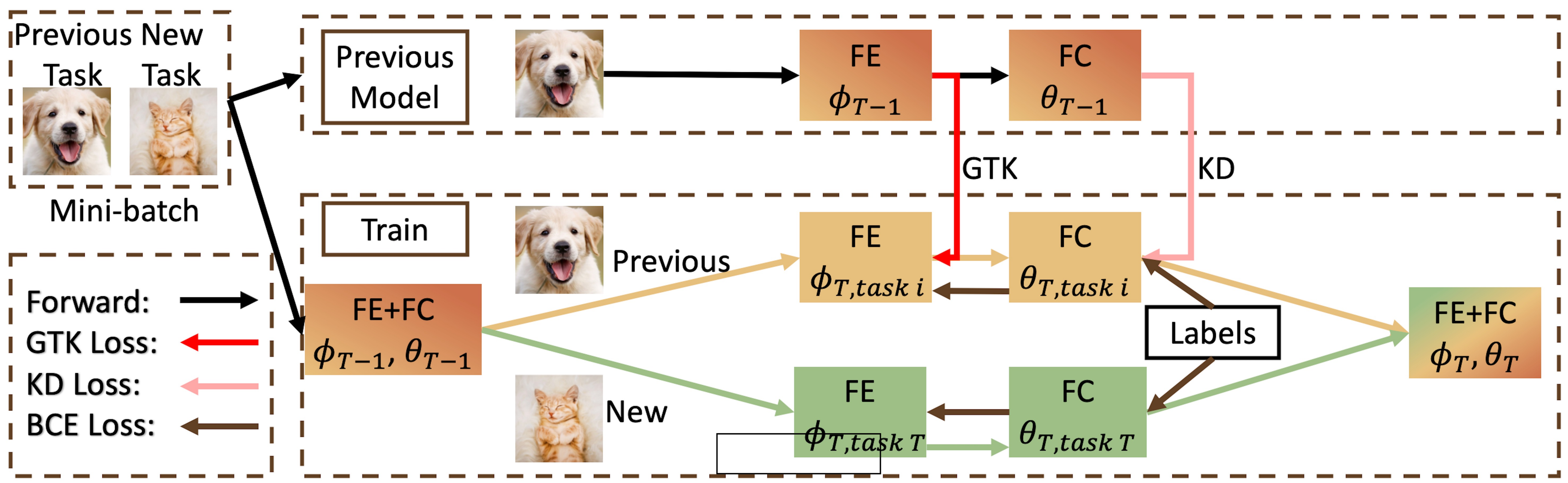
Methods:

Tangent Kernel and Gradient Tangent Kernel Loss:

$$\min_{\phi} \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})}[\mathcal{L}_{GTK}(1-\frac{\boxed{\langle G_T, G_{T-1}\rangle}}{\|G_T\|\|G_{T-1}\|})]$$

Minimize cosine distances between tangent kernels

Tangent Kernel Incremental Learning (TKIL) Approach:

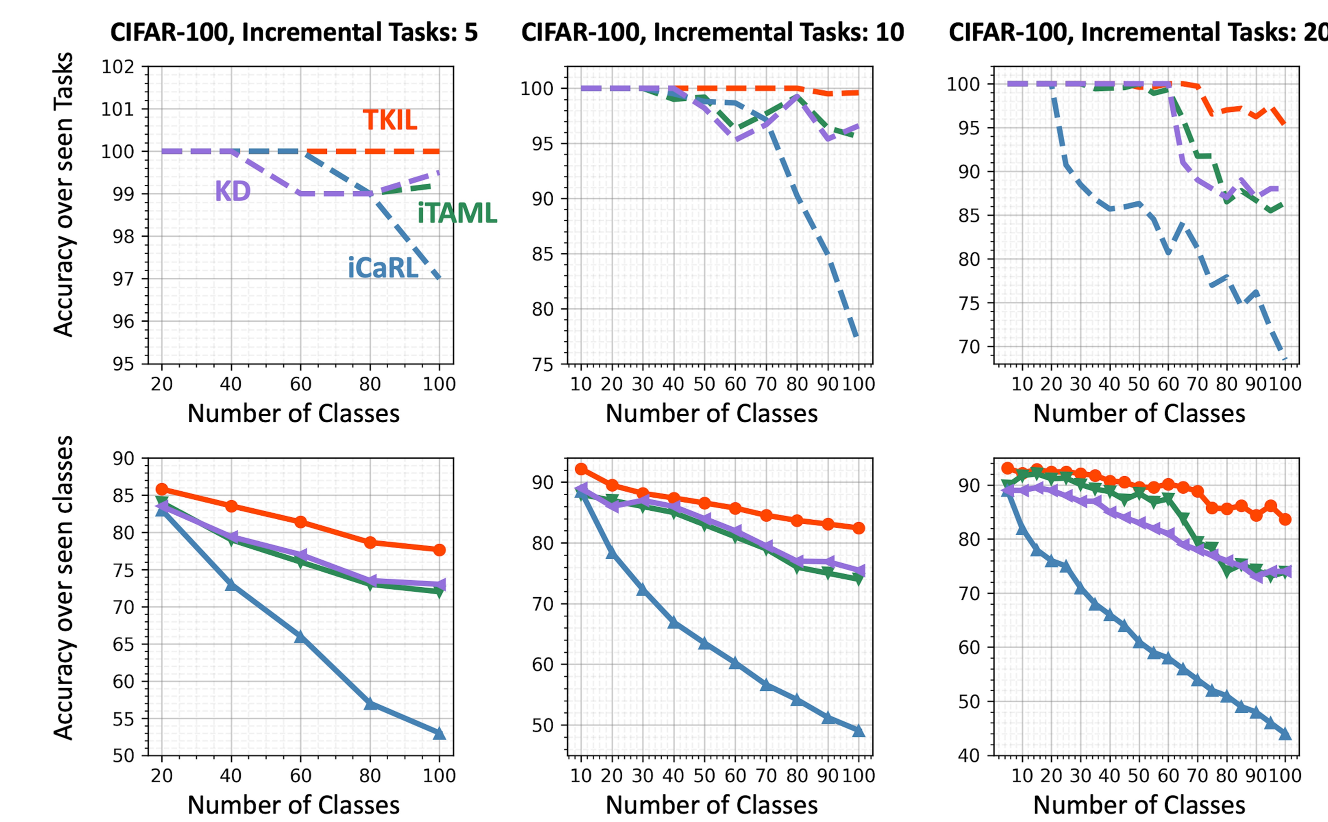


Main Results:

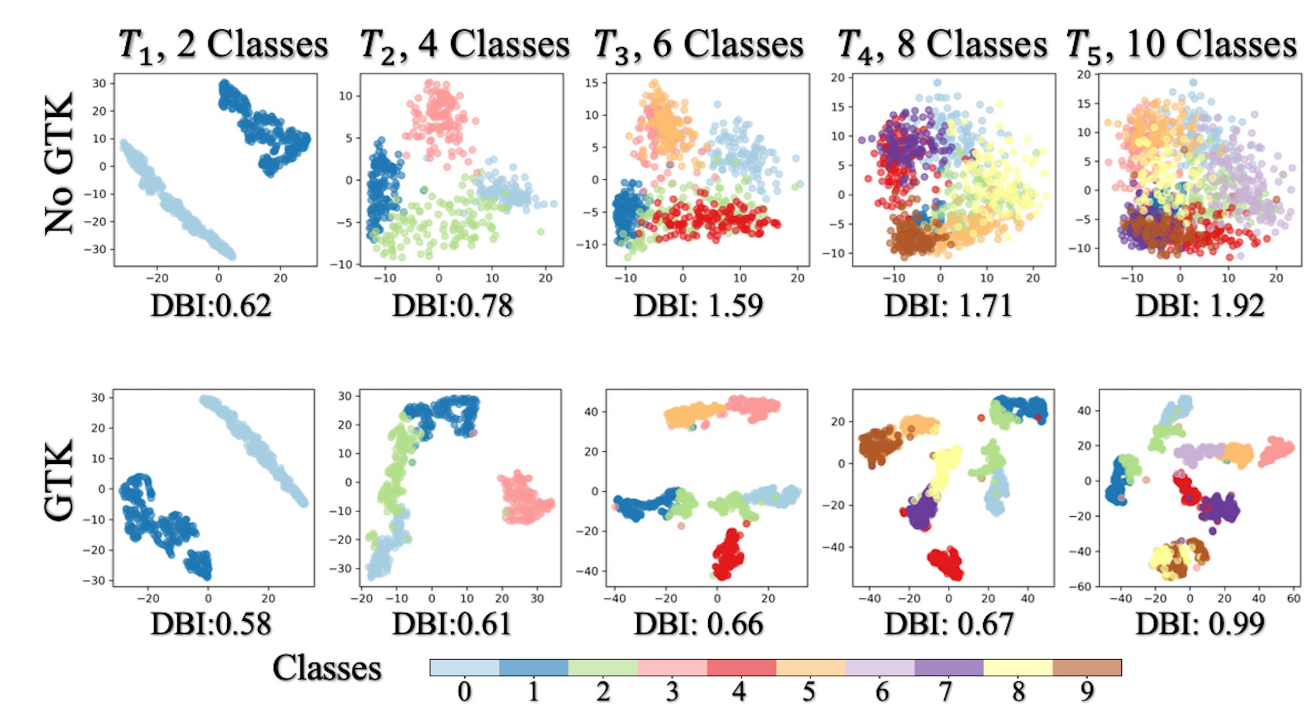
Table 1. Performance comparison between TKIL and other SOTA methods on CIFAR-100 (left-half) and ImageNet-100 (right-half)

Methods	**CIFAR-100**, Memory size $\mathcal{M}=2k$					**ImageNet-100**, Memory size $\mathcal{M}=2k$				
Stages	25	10	5	5	10	25	10	5	5	10
New classes per stage	2	5	10	20	10	2	5	10	20	10
Joint Training (Upper Bound)	86.3%			84.6%		81.3%			76.8%	
iCaRL	50.6%	53.8%	58.1%	57.2%	52.6%	54.6%	60.8%	65.6%	60.1%	59.6%
iTAML	55.9%	74.9%	75.4%	74.5%	74.6%	64.7%	69.5%	71.9%	69.3%	70.4%
RMM	59.5%	60.9%	69.5%	62.7%	60.6%	68.8%	71.4%	73.8%	70.5%	69.4%
SS-IL	58.0%	71.5%	75.1%	74.8%	71.1%	69.5%	71.7%	73.5%	68.8%	67.6%
Mnemonics	61.0%	62.3%	64.1%	63.3%	62.2%	69.7%	71.4%	72.6%	70.6%	70.4%
PODNet	62.7%	64.1%	64.5%	58.9%	59.7%	68.3%	74.3%	75.6%	72.5%	71.5%
AFC	64.1%	64.3%	65.9%	64.9%	64.4%	73.4%	75.8%	75.9%	72.9%	71.7%
TKIL (Ours)	**73.5%**	**80.5%**	**83.6%**	**80.6%**	**82.5%**	**77.3%**	**78.5%**	**79.7%**	**75.7%**	**75.3%**

UNIVERSITY of WASHINGTON NeuroAI



Visualizations



Conclusion

> We formulate gradient tangent kernel loss over feature layers to learn balanced representations and alleviated forgetting for CIL.
> Experiments on multiple benchmarks show that TKIL outperforms existing state-of-the-art methods in various settings, especially in large-volume task scenarios.