Class-Incremental Learning with Cross-Space Clustering and Controlled Transfer

Arjun Ashok, K J Joseph, Vineeth N Balasubramanian

Indian Institute of Technology Hyderabad, India

IIT Hyderabad

ECCV TEL AVIV 2022

Objectives

- **Class-Incremental Learning (CIL)**:
  - learning new classes continually
  - while maintaining knowledge on previous classes
- Two knowledge-distillation based objectives that leverage the **feature space structure** for effective CIL

Motivation and Contributions

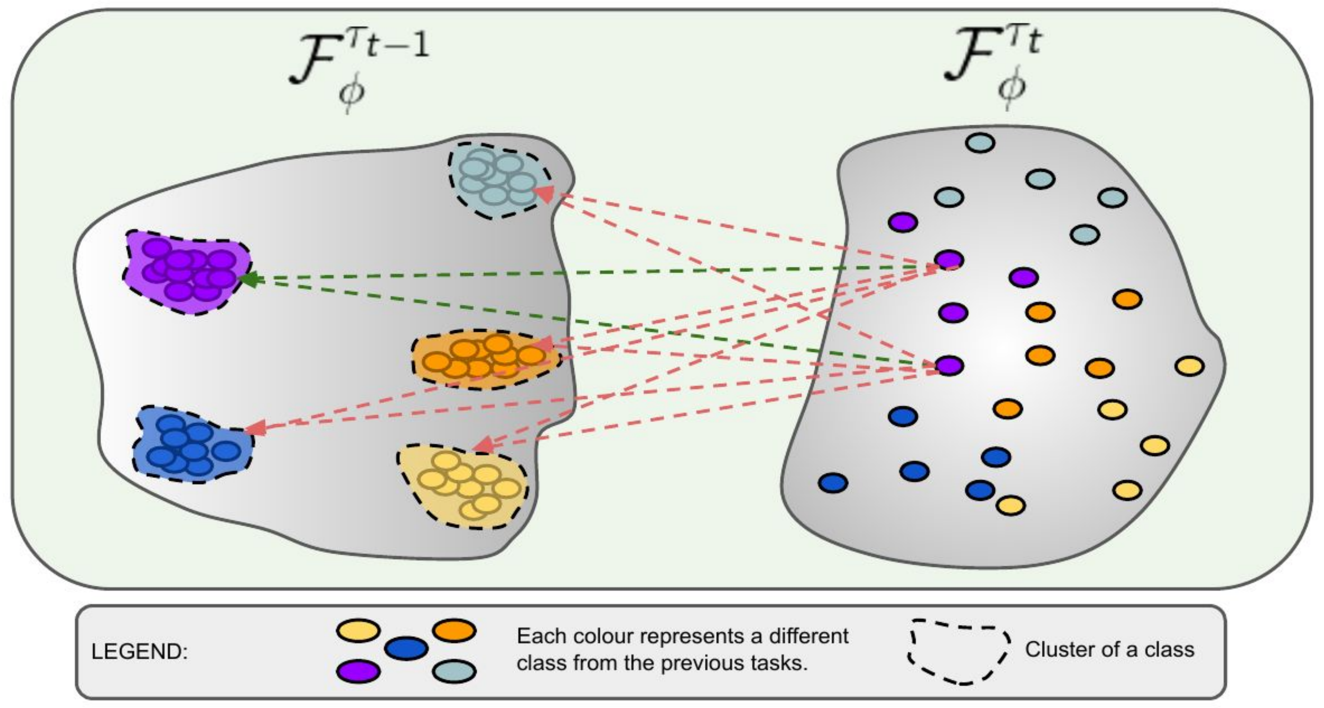
- **Cross-Space Clustering (CSC)**:
  - **Distillation is flawed:**
    - Preserves features of individual instances **independently**
    - Does not characterize the **holistic properties** of each class
  - **Fix:**
    - Distill shared class-level semantics
- **Controlled Transfer (CT):**
  - **Controlling inter-class transfer** is critical for class-incremental learning:
    - Negative backward transfer decreases stability
    - Positive forward transfer increases plasticity
  - **Implementation:**
    - Explicitly regularize inter-class transfer by conditioning learning on inter-class similarities

Cross-Space Clustering (CSC)

- Identifies
  - **class-specific regions**
    - all instances of a class are optimized to stay within these regions
  - **harmful regions**
    - all instances of a class are prevented from drifting towards these regions

$$L_{Cross-Cluster} = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\Big(1-\cos(\mathcal{F}_\phi^{\mathcal{T}_t}(x_i),\mathcal{F}_\phi^{\mathcal{T}_{t-1}}(x_j))\Big) * ind(y_i == y_j)$$

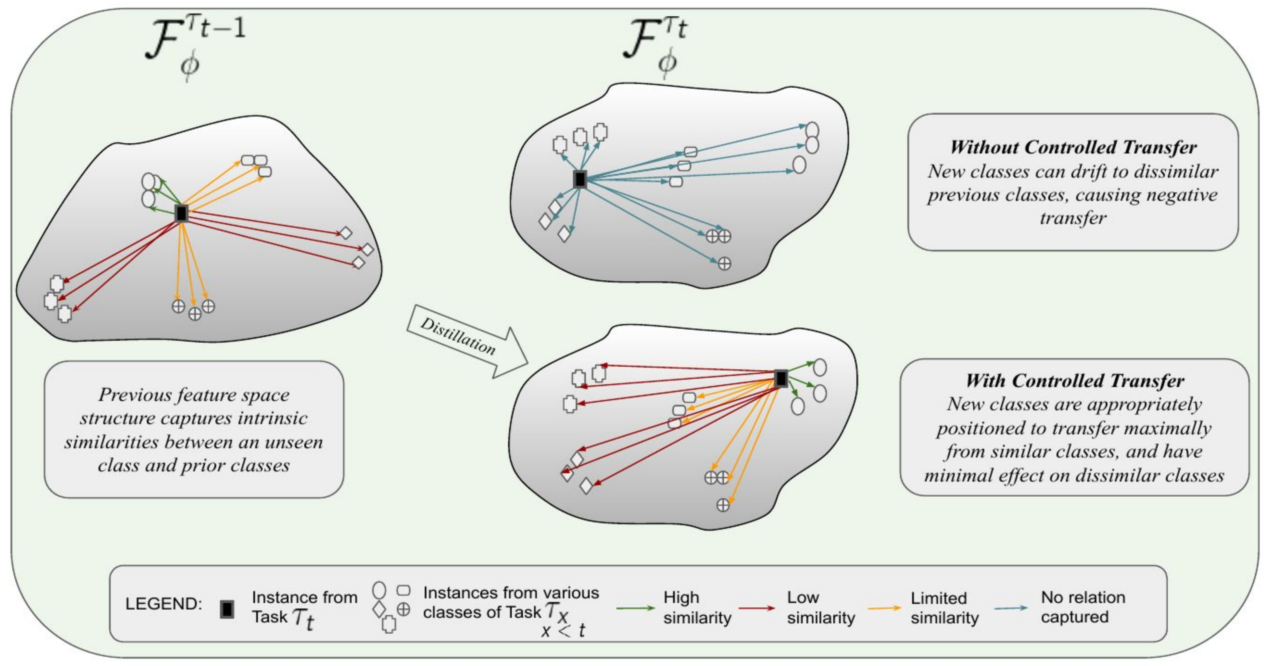
Applied on batches of size k



- Sample is optimized to be
  - **close** to previous positions of **all samples of the class**
  - **far** from previous positions of **other samples**
- Applies the **exact same constraint** on **all samples of a class**
  - Instances of a class **unite and jointly preserve the class** from forgetting

Controlled Transfer (CT)

- Given a new class, when a previous class is
  - **dissimilar**: treat that class distinctively; prevent its forgetting
  - **similar**: maximally transfer features from it; enable plasticity



- Previous feature space contains **dark knowledge** on inter-class similarities
- Cosine similarity between two instances in the same feature space $\mathcal{F}_\phi^{\mathcal{T}_k}$

$$z^{\mathcal{T}_k}_{x_i,x_j} = \cos(\mathcal{F}_\phi^{\mathcal{T}_k}(x_i), \mathcal{F}_\phi^{\mathcal{T}_k}(x_j))$$

- *Softmax-normalized similarity distribution* of x_i in $\mathcal{F}_\phi^{\mathcal{T}_k}$ over D_t

$$H^{\mathcal{T}_k}_{x_i,D_k,T} = \left\{ \frac{(z^{\mathcal{T}_k}_{x_i,x_j}/T)}{\sum_{g=1}^{|D_k|}(z^{\mathcal{T}_k}_{x_i,x_g}/T)} \right\}_{j=1}^{|D_k|}$$

$$L_{Transfer} = \frac{1}{p}\sum_{i=1}^{p} KL(H^{\mathcal{T}_t}_{x_i,Q,T} || H^{\mathcal{T}_{t-1}}_{x_i,Q,T})$$

- **Similarity Regularization:** Constrain similarity in current space using estimated similarities from previous space

Objective Function

where L_{method} refers to the objective of a base method

$$L_{CSCCT} = L_{method} + \alpha * L_{Cross-Cluster} + \beta * L_{Transfer}$$

Experimental Results

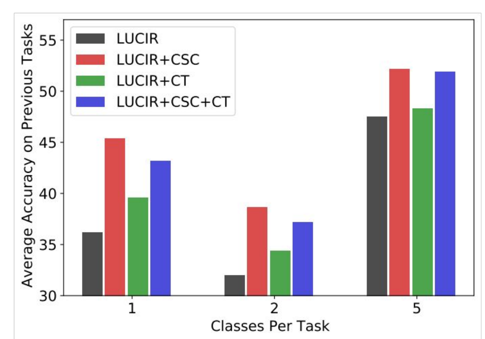
- Experiments on two protocols
  - training with 50% of classes in the first task
  - training with equal classes in all tasks
- Evaluated as an **add-on** on top of existing methods

Dataset	CIFAR100					
Settings	$\mathcal{B}=50$			$\mathcal{B}=\mathcal{C}$		
Methods	$\mathcal{C}=1$	$\mathcal{C}=2$	$\mathcal{C}=5$	$\mathcal{C}=1$	$\mathcal{C}=2$	$\mathcal{C}=5$
iCaRL [36]	43.39	48.31	54.42	30.92	36.80	44.19
iCaRL + CSCCT	$46.15_{+2.76}$	$51.62_{+3.31}$	$56.75_{+2.33}$	$34.02_{+3.1}$	$\mathbf{39.60}_{+2.8}$	$46.45_{+2.26}$
LUCIR [18]	50.26	55.38	59.40	25.40	31.93	42.28
LUCIR + CSCCT	$52.95_{+2.69}$	$56.49_{+1.13}$	$62.01_{+2.61}$	$28.12_{+2.72}$	$34.96_{+3.03}$	$44.03_{+1.55}$
PODNet [13]	56.88	59.98	62.66	33.58	36.68	45.27
PODNet + CSCCT	$\mathbf{58.80}_{+1.92}$	$\mathbf{61.10}_{+1.12}$	$\mathbf{63.72}_{+1.06}$	$\mathbf{36.23}_{+2.65}$	$39.3_{+2.62}$	$\mathbf{47.8}_{+2.53}$

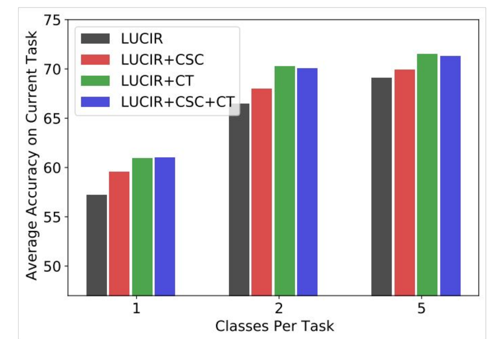
Dataset	ImageNet-Subset					
Settings	$\mathcal{B}=50$			$\mathcal{B}=\mathcal{C}$		
Methods	$\mathcal{C}=2$	$\mathcal{C}=5$	$\mathcal{C}=10$	$\mathcal{C}=2$	$\mathcal{C}=5$	$\mathcal{C}=10$
iCaRL [36]	55.81	57.34	65.97	40.75	55.92	60.93
iCaRL + CSCCT	$57.01_{+1.2}$	$58.37_{+1.03}$	$66.82_{+0.8}$	$42.46_{+1.71}$	$57.45_{+1.53}$	$62.60_{+1.67}$
LUCIR [18]	60.44	66.55	70.18	36.84	46.40	56.78
LUCIR + CSCCT	$61.52_{+1.08}$	$67.91_{+1.36}$	$71.33_{+1.15}$	$37.86_{+1.02}$	$47.55_{+1.15}$	$58.07_{+1.29}$
PODNet [13]	67.27	73.01	75.32	44.94	58.23	66.24
PODNet + CSCCT	$\mathbf{68.91}_{+1.64}$	$\mathbf{74.35}_{+1.34}$	$\mathbf{76.41}_{+1.09}$	$\mathbf{46.06}_{+1.12}$	$\mathbf{59.43}_{+1.2}$	$\mathbf{67.49}_{+1.25}$

Ablation Studies

Average Accuracy on Previous Tasks



Average Accuracy on Current Task



CSC mainly affects **stability**

CT mainly affects **plasticity**