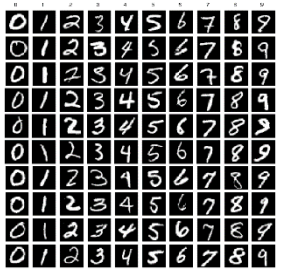
RSAC: Regularized Subspace Approximation Classifier for Lightweight Continuous Learning

Chih-Hsing Ho

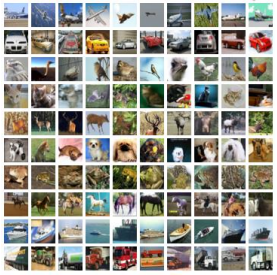
Advisor : Shang-Ho (Lawrence) Tsai

Supervised Learning

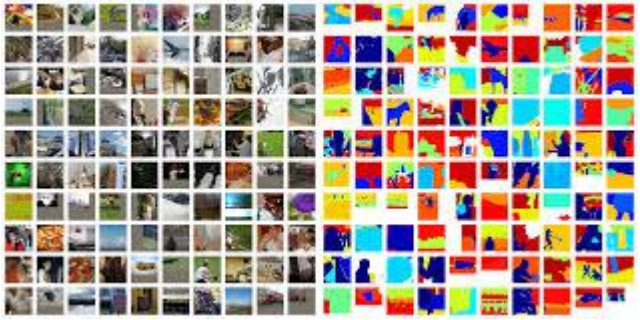
- Supervised learning achieves great success
- All tasks are assigned at once
- Entire labeled dataset is provided to deep network



⇨ Mnist



⇨ Cifar 10

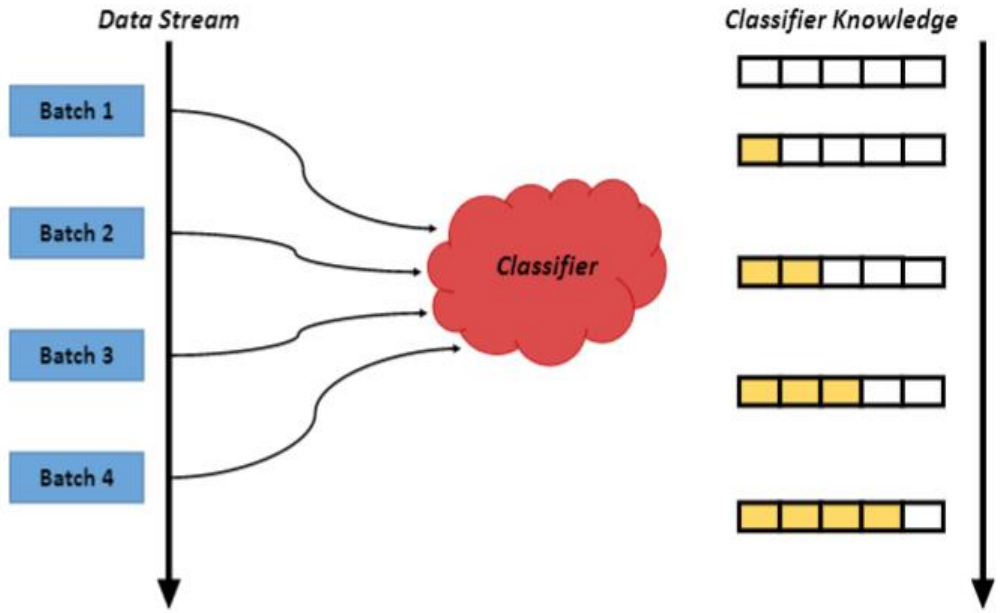


⇨ COCO Dataset

Continuous Learning

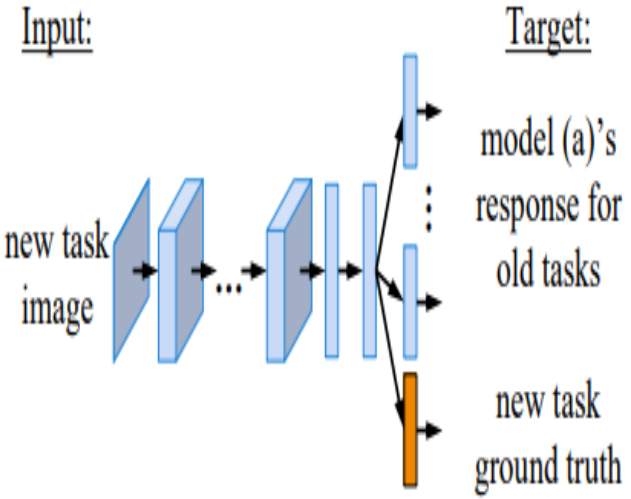
- Human grow their knowledge continually
- Deep network imitates human to continually learning

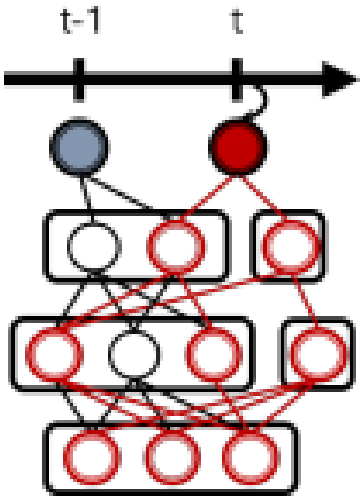
➡ Catastrophic forgetting

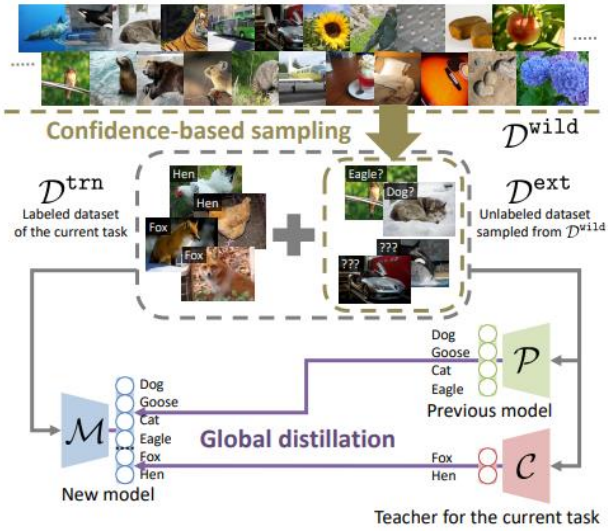


Continuous Learning

- Several approaches are proposed to resolve catastrophic forgetting
  1. Weight consolidation
  2. Architecture expansion
  3. Memory rehearsal







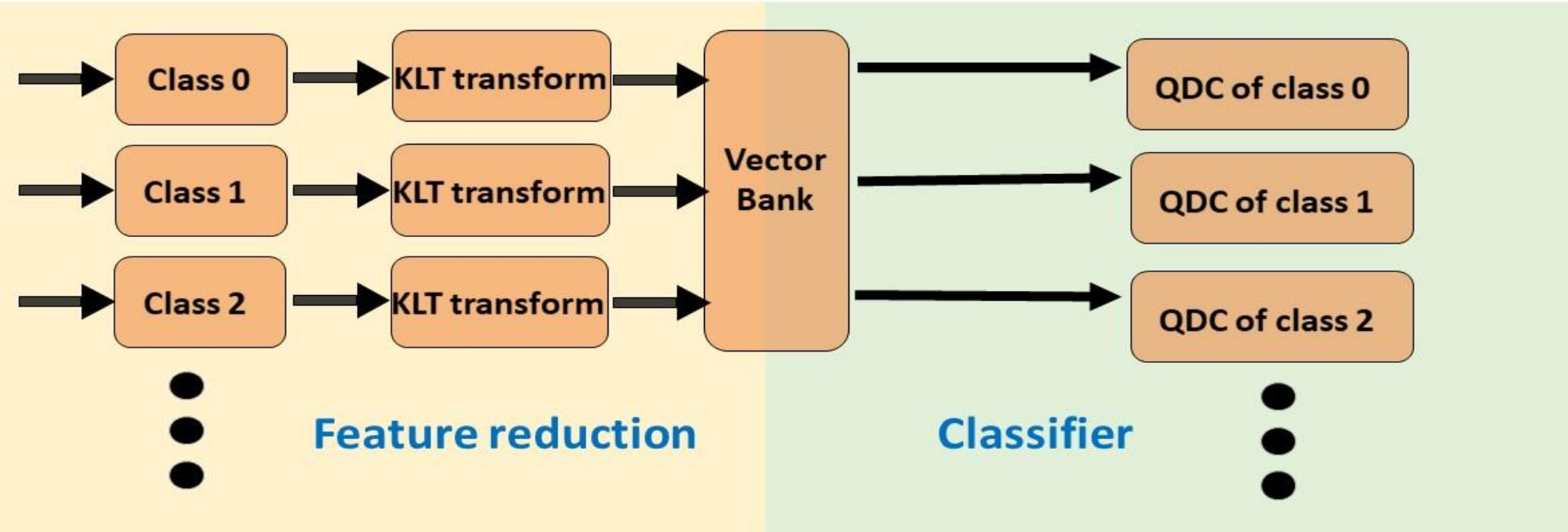
Zhizhong .et al ECCV 2016

Jaehong .et al ICLR 2018

Kibok .et al ICCV 2019

RSAC

- Modules
  1. Feature extraction module
  2. Classifier module



RSAC

- Notation

x is the raw image, which the mean μ_c and the covariance matrix Σ_c of class c established on

$$\mu_c = \frac{1}{N_c}\sum_{j=1}^{N_c} x_j$$

N_c is the number of data belongs to class c, Λ_c is a diagonal matrix with the eigenvalue σ_c^j as the j^{th} entry and Q_c is a $d \times d$ orthonormal matrix

$$\Sigma_c = \frac{1}{N_c}\sum_{j=1}^{N_c}(x_j - \mu_c)(x_j - \mu_c)^T = Q_c \Lambda_c Q_c^T$$

RSAC

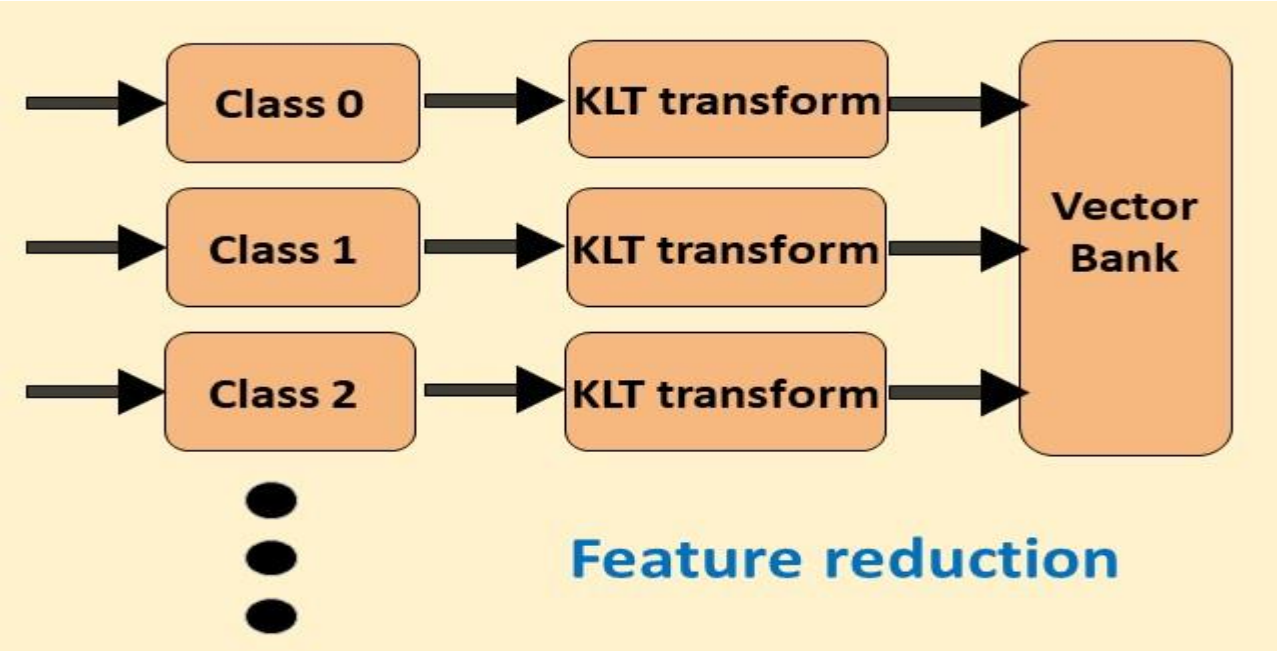
- Feature reduction

x is the raw image, which the mean μ_c and the covariance matrix Σ_c of class c established on

N_c is the number of data belongs to class c, Λ_c is a diagonal matrix with the eigenvalue σ_c^j as the j^{th} entry and Q_c is a $d \times d$ orthonormal matrix

$$\frac{\sum_{j=1}^{k} \sigma_c^j}{\sum_{j=1}^{d} \sigma_c^j} \geq t$$

$$f(x) = \hat{Q}_c^T x$$

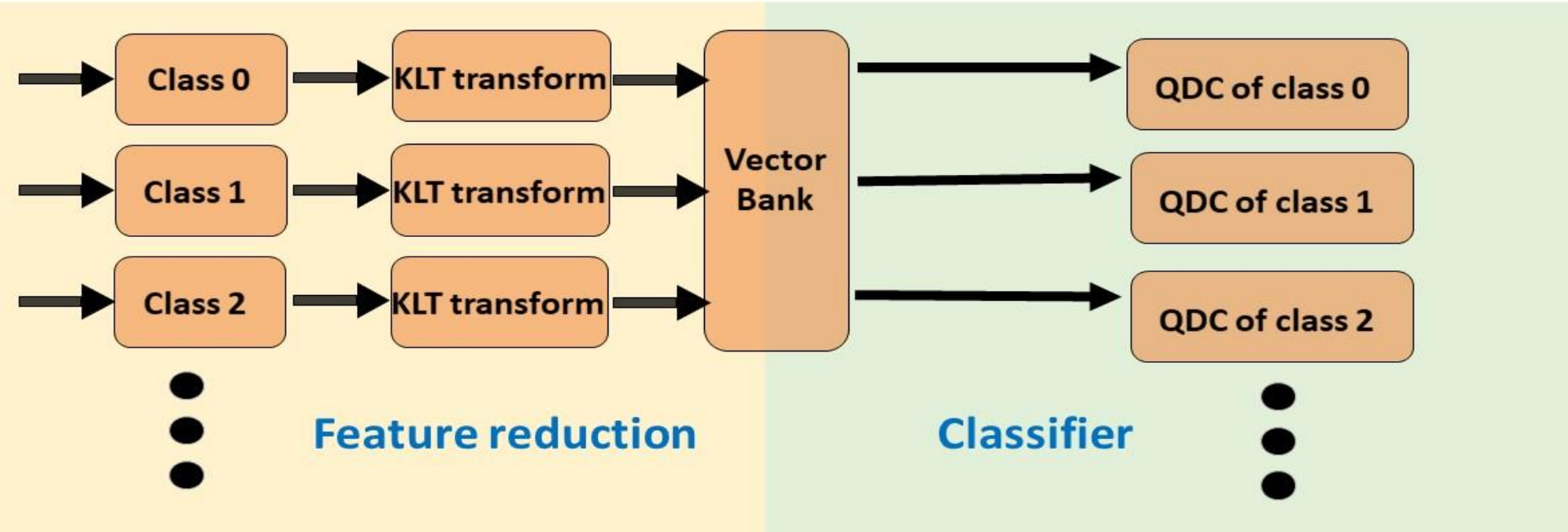


RSAC

- Classifier module

$$\arg\max_c \ln\left(P_{Y|X}(c|f(x))\right)$$

$$= \arg\max_c \ln\left(\frac{1}{|\hat{\Sigma}_c|}\right) - \frac{1}{2}(f(x) - \hat{\mu}_c)^T \hat{\Sigma}_c^{-1}(f(x) - \hat{\mu}_c) + \ln\left(\frac{N_c}{\sum_{j=1}^{C} N_j}\right)$$

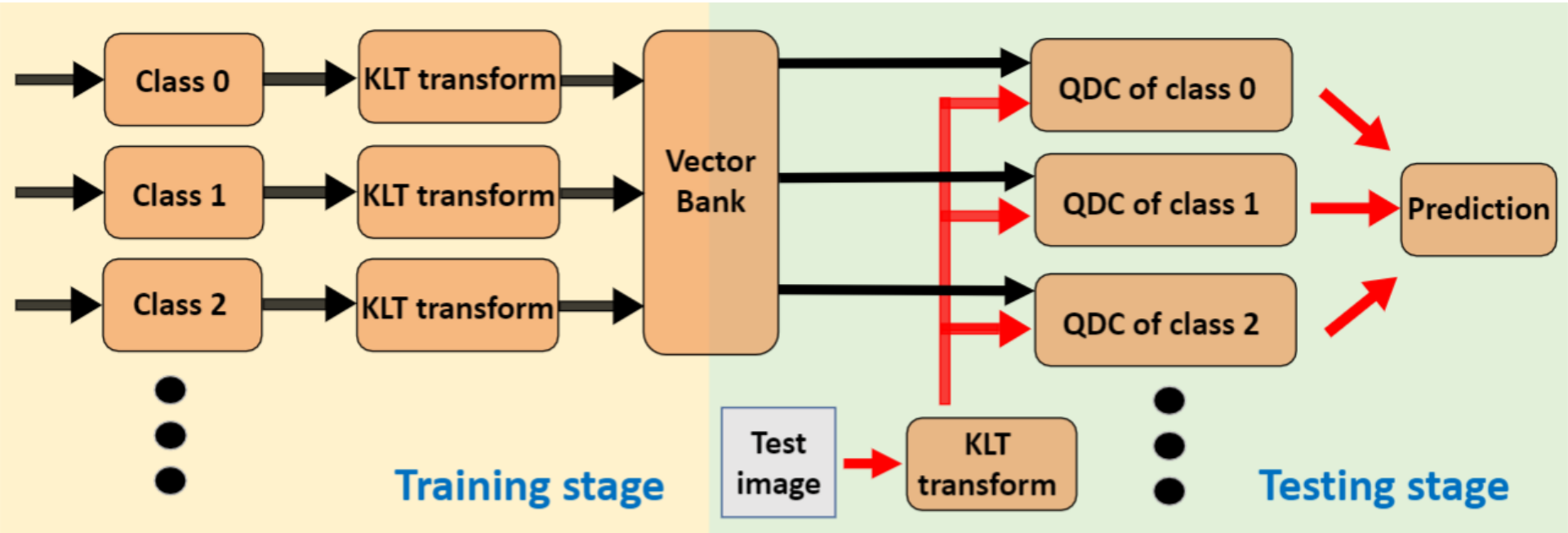


RSAC

- Regularization

$$\hat{\Sigma}_c = Cov(f(x)) = \hat{Q}_c^T Cov(x)\hat{Q}_c = \hat{Q}_c^T \Sigma_c \hat{Q}_c = \hat{Q}_c^T Q_c \Lambda_c Q_c^T \hat{Q}_c = \hat{\Lambda}_c$$

$$\hat{\Sigma}_c' = \hat{\Sigma}_c + \alpha * I = \hat{\Lambda}_c + \alpha * I$$



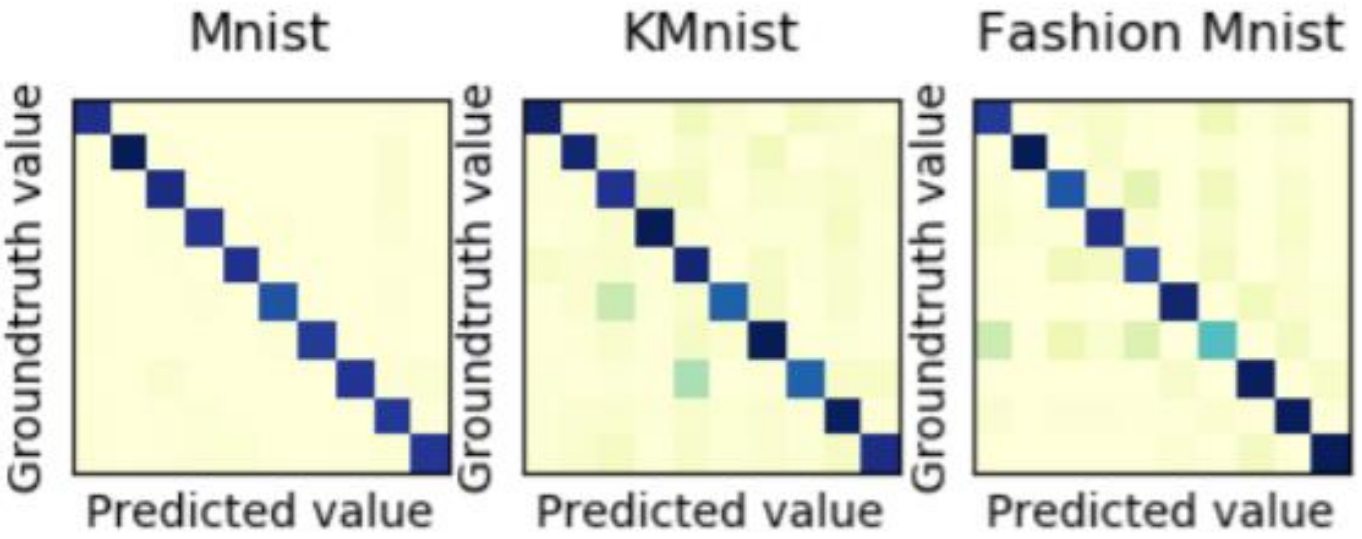
Experiment

- Class incremental

Methods	Datasets (Accuracy)			Datasets (Training Time (sec))		
	Mnist	KMnist	Fashion Mnist	Mnist	KMnist	Fashion Mnist
DGR [28]	90.44±1.56	69.25±2.94	74.83±5.50	315.99±2.25	748.75±51.17	760.21±21.72
DGR+distill [20], [28]	92.31±0.74	64.42±1.12	76.03±4.12	314.12±12.79	819.52±14.52	800.81±3.69
EWC [21]	20.45±1.15	19.54±0.12	19.97±0.02	398.86±11.04	719.89±21.95	697.24±53.39
Online EWC [61]	20.69±1.53	19.54±0.12	19.97±0.03	371.87±12.35	665.04±3.40	692.49±29.20
iCaRL [13]	93.24±0.70	70.83±2.78	79.61±0.79	200.16±9.83	468.38±4.98	466.60±11.09
LwF [20]	20.98±0.85	20.16±0.24	19.42±2.54	198.40±9.09	495.62±31.48	499.49±8.77
RtF [46]	93.75±1.28	66.16±3.06	74.11±4.82	253.37±9.22	639.66±25.56	678.42±34.04
SI [22]	19.85±0.10	19.53±0.09	19.97±0.02	194.16±87.6	503.72±5.15	498.37±3.28
CNDPM [62]	93.54±0.13	74.35±1.4	44.62±2.1	> 3600	> 3600	> 3600
Saak [32]	95.21	76.25	73.51	> 3000	> 3000	> 3000
Ours	**95.59**	**77.35**	**80.32**	**5.90**	**5.72**	**5.48**

Experiment

- Class incremental



Experiment

- Ablation Study

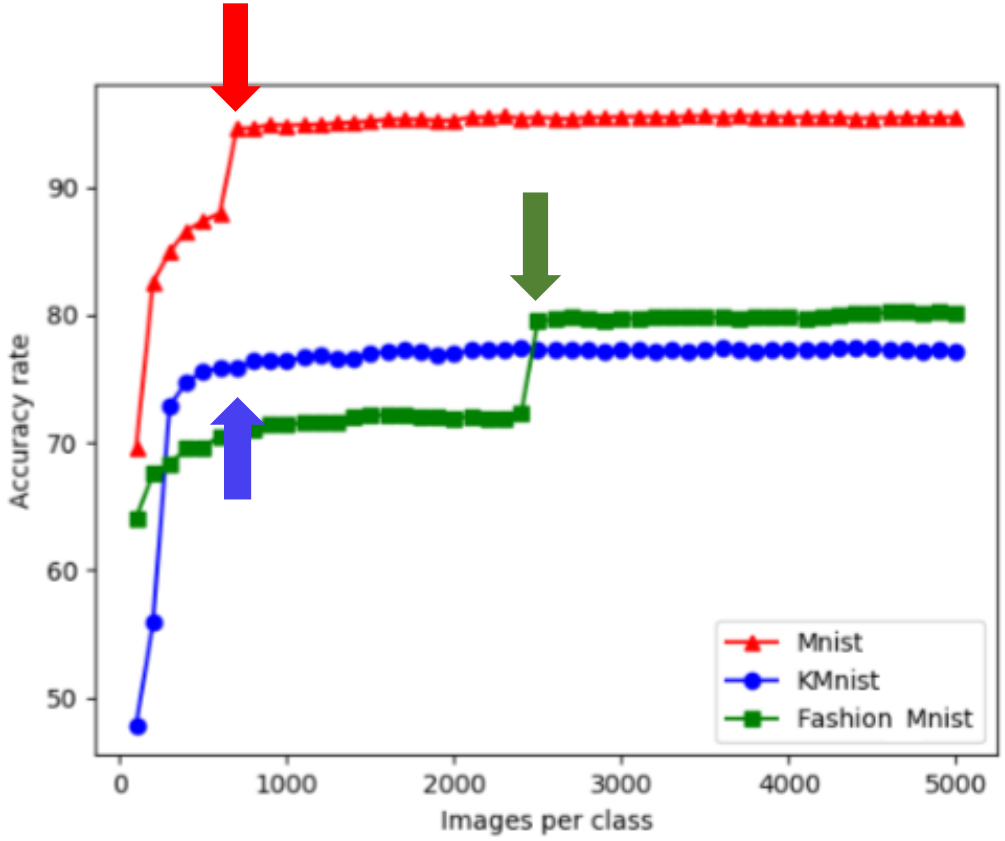
  t : power threshold

  k : the amount of features after reduction

Power threshold t	Mnist		KMnist		Fashion Mnist	
	k	acc	k	acc	k	acc
0.8	31	67.75	64	61.30	26	64.90
0.9	68	93.22	126	76.16	77	73.98
0.95	121	95.41	211	77.13	156	79.74
0.96	141	95.43	243	76.87	185	80.25
0.97	168	95.43	285	74.84	224	73.56
0.98	206	91.66	346	75.09	278	73.95
Best	150	95.59	192	77.35	183	80.32

Experiment

- Ablation Study



Thank you for your listening

https://arxiv.org/abs/2007.01480

RSAC: Regularized Subspace Approximation Classifier for Lightweight Continuous Learning