

Lifelong Learning via Progressive Distillation and Retrospection

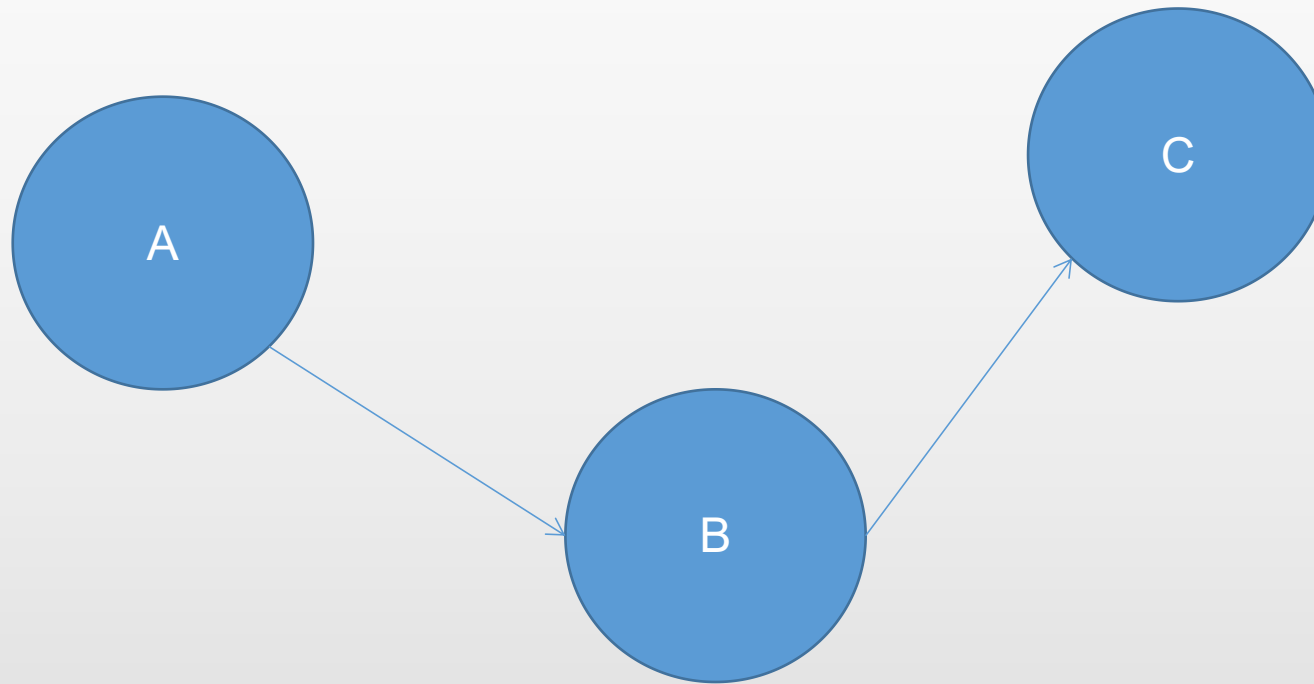
ECCV2018

胡超杰

2019.3.15

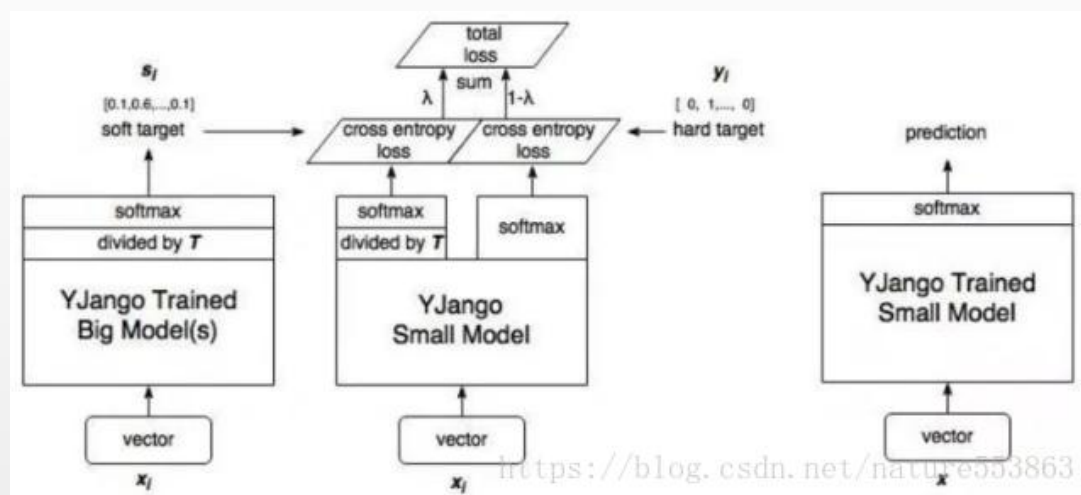
Lifelong Learning

- fine-tuning
- joint training
- Feature Extraction

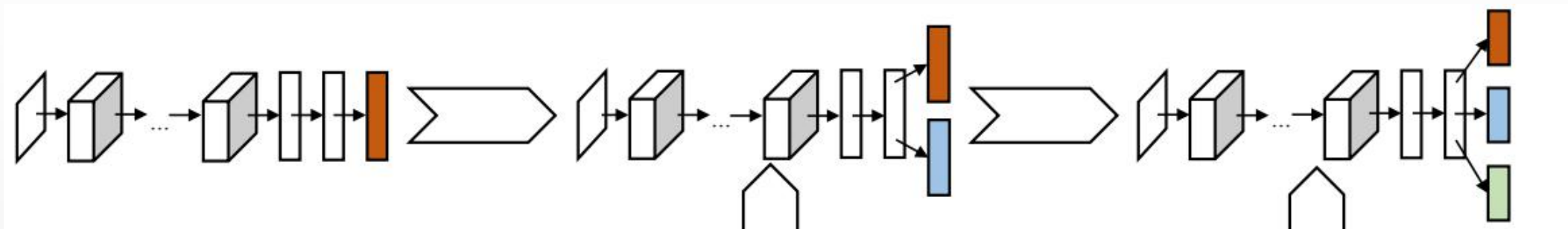


Distillation

- hard supervision -> soft supervision

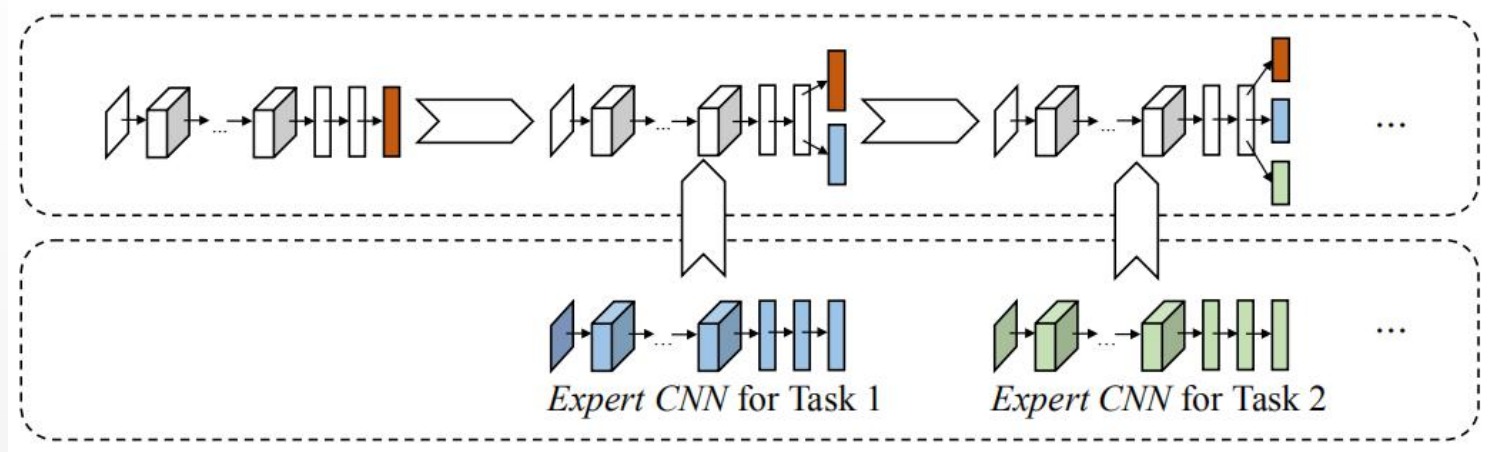


learning without forgetting



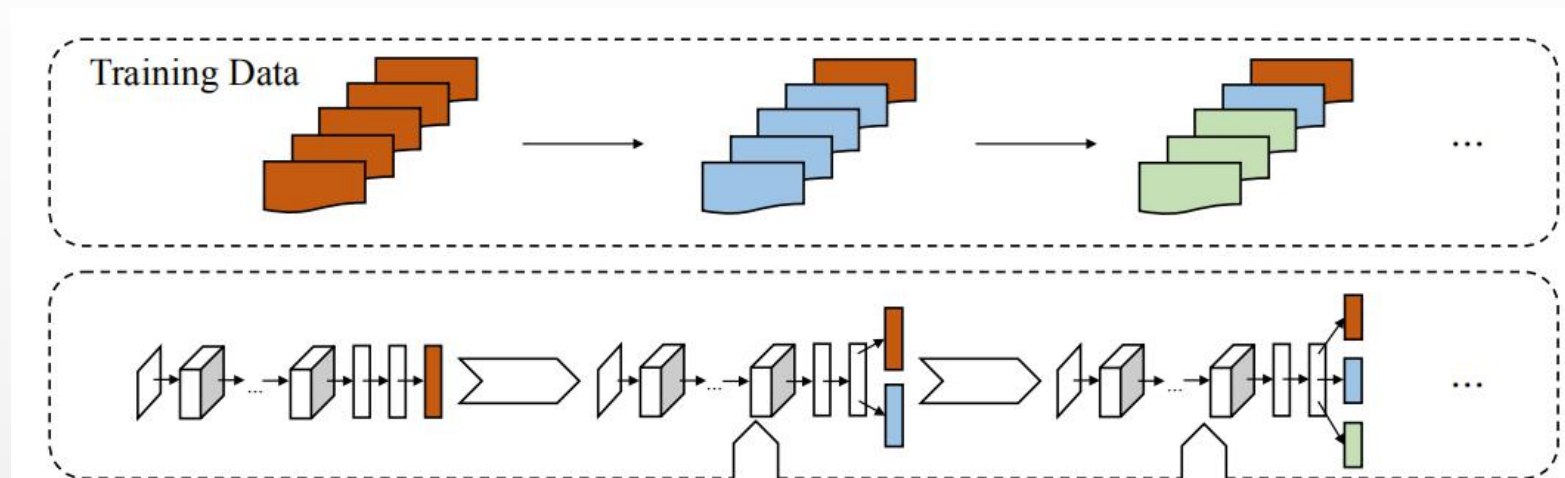
- (1)旧模型在新的数据集上跑，得出这个时候的旧分类的相应的输出，保存下来当作接下来蒸馏训练中的老师
- (2)对于新的数据集的新增加的分类采用one-hot“硬监督”的形式进行训练，作为第一个loss
- (3)训练过程中旧分类的输出按照第一步中的结果以“软监督”的形式进行训练，作为第二个loss

Progressive Distillation and Retrospection



- (1) 在每个新的数据集上都训练一个模型作为教师模型
- (2) 第一个loss同learning without forgetting，最大限度的保留旧知识
- (3) 第二个loss采用的是本地的教师模型的输出作为“软监督”来进行训练，不使用one-hot“硬监督”

Progressive Distillation and Retrospection



转向新的数据集训练的时候捎带一些旧的数据以及对应的标签等信息参与训练

Experiments

	ImageNet→Birds		ImageNet→Flowers		ImageNet→Scenes	
<i>Feature Extraction</i>	57.44 (ref)	50.12 (-7.07)	57.44 (ref)	83.10 (-3.99)	57.44 (ref)	60.22 (-2.61)
<i>Finetuning</i>	43.20 (-14.25)	57.19 (ref)	48.45 (-8.99)	87.09 (ref)	46.61 (-10.84)	62.84 (ref)
<i>LwF</i> [15]	54.49 (-2.95)	57.45 (+0.26)	55.77 (-1.67)	85.87 (-1.22)	55.01 (-2.43)	64.03 (+1.19)
<i>D</i> (ours)	55.34 (-2.11)	58.21 (+1.02)	55.95 (-1.49)	86.19 (-0.89)	55.65 (-1.79)	64.70 (+1.87)
<i>LwF + R</i>	55.61 (-1.83)	57.79 (+0.60)	56.48 (-0.96)	86.53 (-0.55)	55.71 (-1.73)	64.70 (+1.87)
<i>D + R</i> (ours)	55.85 (-1.59)	59.55 (+2.36)	56.53 (-0.92)	87.02 (-0.07)	56.02 (-1.43)	65.00 (+2.16)

	Flowers→Birds		Flowers→Scenes		Flowers→Aircrafts	
<i>Feature Extraction</i>	87.09 (ref)	48.29 (-8.72)	87.09 (ref)	57.09 (-5.07)	87.09 (ref)	40.98 (-26.13)
<i>Finetuning</i>	72.97 (-14.12)	57.02 (ref)	72.97 (-14.12)	62.16 (ref)	70.88 (-16.20)	67.12 (ref)
<i>LwF</i> [15]	85.08 (-2.00)	54.55 (-2.46)	84.86 (-2.23)	61.87 (-0.30)	81.69 (-5.40)	66.10 (-1.02)
<i>D</i> (ours)	85.30 (-1.79)	56.64 (-0.38)	85.36 (-1.72)	62.31 (+0.15)	82.14 (-4.94)	67.57 (+0.45)
<i>LwF + R</i>	85.15 (-1.93)	56.79 (-0.22)	85.31 (-1.77)	62.54 (+0.37)	85.07 (-2.02)	66.88 (-0.24)
<i>D + R</i> (ours)	85.38 (-1.71)	58.16 (+1.14)	85.73 (-1.35)	64.03 (+1.87)	85.57 (-1.51)	68.38 (+1.26)

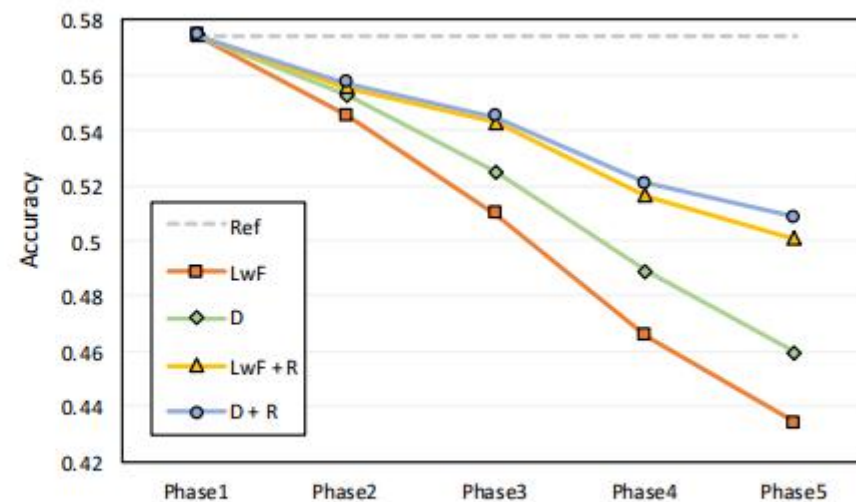
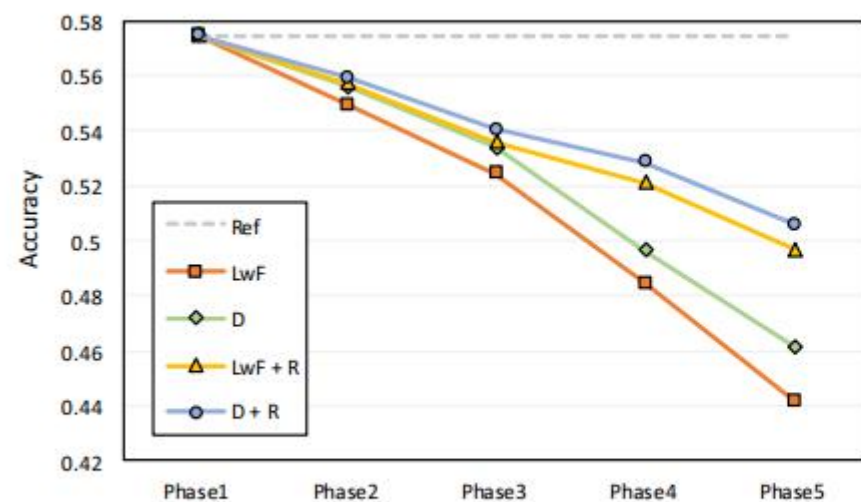
Experiments

	Imagenet \rightarrow Scenes \rightarrow Birds \rightarrow Flowers \rightarrow Aircrafts					
	Imagenet	Scenes	Birds	Flowers	Aircrafts	Average
<i>LwF</i> [15]	44.20 (ref)	55.90 (ref)	52.22 (ref)	81.64 (ref)	65.80 (ref)	59.95 (ref)
<i>D</i> (ours)	46.15 (+1.95)	55.67 (-0.22)	53.17 (+0.95)	82.37 (+0.73)	66.79 (+0.99)	60.83 (+0.88)
<i>LwF</i> + <i>R</i>	49.70 (+5.49)	59.25 (+3.36)	56.45 (+4.22)	85.49 (+3.85)	66.82 (+1.02)	63.54 (+3.59)
<i>D</i> + <i>R</i> (ours)	50.58 (+6.38)	60.52 (+4.63)	56.84 (+4.62)	86.00 (+4.36)	68.41 (+2.61)	64.47 (+4.52)

	Imagenet \rightarrow Birds \rightarrow Flowers \rightarrow Aircrafts \rightarrow Scenes					
	Imagenet	Birds	Flowers	Aircrafts	Scenes	Average
<i>LwF</i> [15]	43.37 (ref)	52.26 (ref)	79.91 (ref)	63.25 (ref)	60.82 (ref)	59.92 (ref)
<i>D</i> (ours)	45.94 (+2.57)	51.90 (-0.36)	81.21 (+1.30)	64.30 (+1.05)	60.90 (+0.07)	60.85 (+0.93)
<i>LwF</i> + <i>R</i>	50.05 (+6.67)	55.60 (+3.34)	85.12 (+5.20)	66.43 (+3.18)	62.39 (+1.57)	63.92 (+3.99)
<i>D</i> + <i>R</i> (ours)	50.84 (+7.47)	57.05 (+4.79)	85.72 (+5.81)	67.42 (+4.17)	62.91 (+2.09)	64.79 (+4.87)

	Imagenet \rightarrow Flowers \rightarrow Aircrafts \rightarrow Scenes \rightarrow Birds					
	Imagenet	Flowers	Aircrafts	Scenes	Birds	Average
<i>LwF</i> [15]	44.49 (ref)	77.50 (ref)	61.57 (ref)	60.30 (ref)	56.02 (ref)	59.98 (ref)
<i>D</i> (ours)	46.37 (+1.88)	79.25 (+1.74)	62.47 (+0.90)	60.00 (-0.30)	57.22 (+1.21)	61.06 (+1.08)
<i>LwF</i> + <i>R</i>	50.26 (+5.77)	84.48 (+6.98)	65.38 (+3.81)	62.31 (+2.01)	57.54 (+1.52)	63.99 (+4.02)
<i>D</i> + <i>R</i> (ours)	50.76 (+6.26)	85.07 (+7.56)	65.83 (+4.26)	62.54 (+2.24)	59.52 (+3.50)	64.74 (+4.76)

Experiments



(a) Imagenet→Scenes→Birds→Flowers→Aircrafts. (b) Imagenet→Birds→Flowers→Aircrafts→Scenes.

thanks!