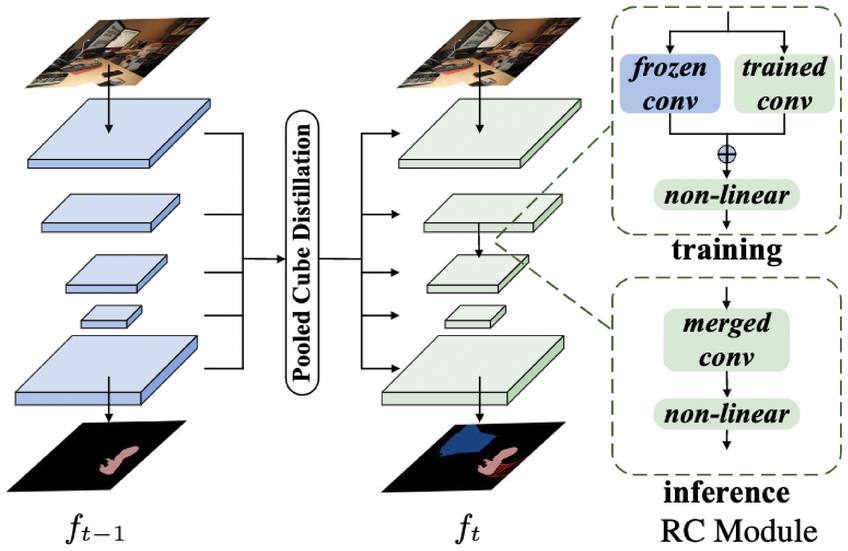
Representation Compensation Networks for Continual Semantic Segmentation

Chang-Bin Zhang, Jia-Wen Xiao, Xialei Liu, Ying-Cong Chen, Ming-Ming Cheng

The code is available at: https://github.com/zhangchbin/RCIL

Motivation

Problem: 1) Continual Class Segmentation: Expanding the number of classes that can be recognized.
2) Continual Domain Segmentation: Improving discrimination for different domains while ignoring classes difference between domains.
Challenge: Alleviating catastrophic forgetting while improving discrimination for new classes.
Motivation: Decouple the old and new knowledge in the parametric space.

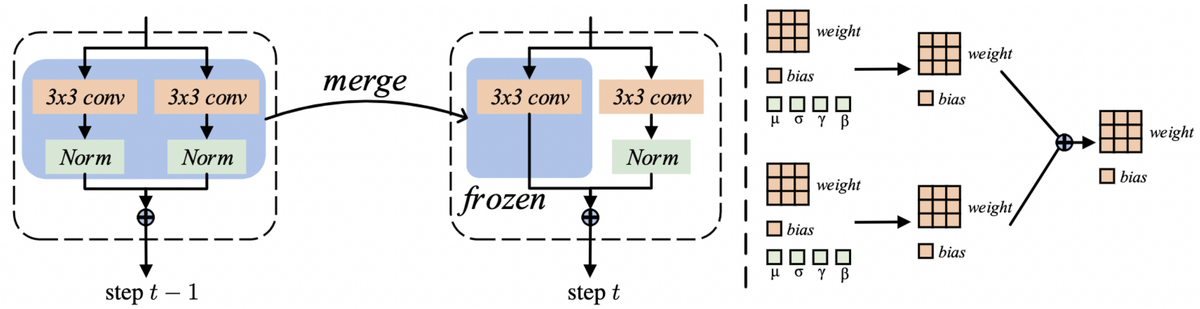


Contribution: 1) Representation Compensation (RC) Module; 2) Pooled Cube Distillation Strategy

Evaluation protocols

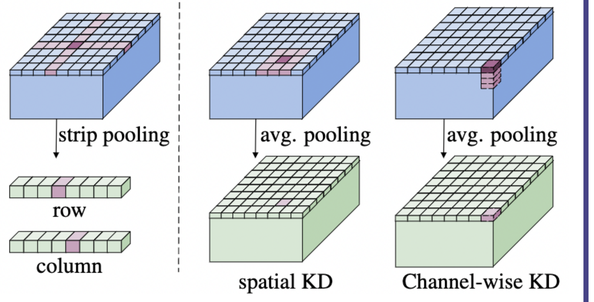
- **Disjoint**: Training images in the current step do not contain classes in the future steps.
- **overlapped**: Training images in the current step may contain potential classes in the future.
- **X-Y**: The model can recognize X classes/domains in the first step and learn Y classes/domains in the subsequent steps.

Representation Compensation Strategy



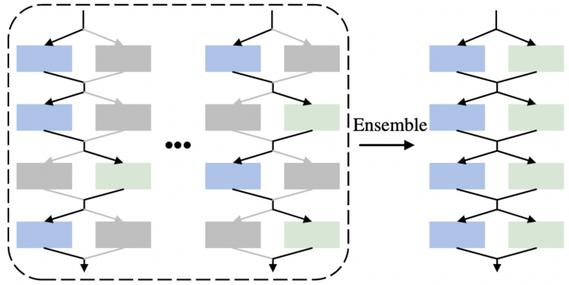
- **Re-parameter scheme**: Two linearly added branches can be equivalently merged as a convolution path.
- **Representation Compensation**: The merged branch is frozen to memorize the old knowledge.

Pooled Cube Distillation



- **Previous approach**: Imposing long-range dependency with strip pooling.
- **Our approach**: Reducing unrelated noise with local average pooling. Meanwhile, it performs on both the spatial and channel-wise dimensions.

Effectiveness Analysis



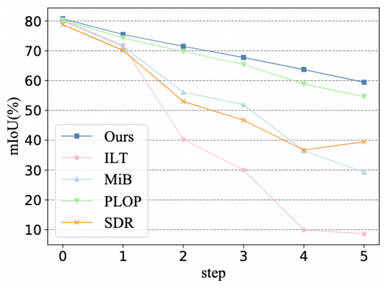
Blue : frozen Green : trainable Gray : ignored

- **Model Ensemble**: The parallel structure can be understanded as integration of many sub-networks.
- **Implicit Distillation**: The frozen path distill knowledge to the trainable path implicitly.

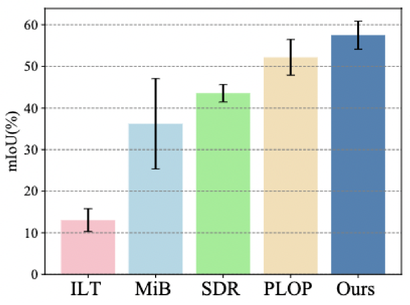
Experimental results

	15-5 (2 steps)						**15-1** (6 steps)						**10-1** (11 steps)					
	Disjoint			**Overlapped**			**Disjoint**			**Overlapped**			**Disjoint**			**Overlapped**		
Method	*0-15*	*16-20*	*all*	*0-15*	*16-20*	*all*	*0-15*	*16-20*	*all*	*0-15*	*16-20*	*all*	*0-10*	*11-20*	*all*	*0-10*	*11-20*	*all*
Fine-tuning	5.7	33.6	12.3	6.6	33.1	12.9	4.6	1.8	3.8	4.6	1.8	3.9	6.3	1.1	3.8	6.4	1.2	3.9
Joint	79.8	72.6	78.2	79.8	72.6	78.2	79.8	72.6	78.2	79.8	72.6	78.2	78.2	78.0	78.2	78.2	78.0	78.2
LwF	60.4	37.4	54.9	60.8	36.6	55.0	5.8	3.6	5.3	6.0	3.9	5.5	7.2	1.2	4.3	8.0	2.0	4.8
ILT	64.9	39.5	58.9	67.8	40.6	61.3	8.6	5.7	7.9	9.6	7.8	9.2	7.3	3.2	5.4	7.2	3.7	5.5
MiB	73.0	43.3	65.9	76.4	49.4	70.0	48.4	12.9	39.9	38.0	13.5	32.2	9.5	4.1	6.9	20.0	20.1	20.1
SDR	74.6	44.1	67.3	76.3	50.2	70.1	59.4	14.3	48.7	47.3	14.7	39.5	17.3	11.0	14.3	32.4	17.1	25.1
PLOP	71.0	42.8	64.3	75.7	51.7	70.1	57.9	13.7	46.5	65.1	21.1	54.6	9.7	7.0	8.4	44.0	15.5	30.5
Ours	75.0	42.8	67.3	78.8	52.0	72.4	66.1	18.2	54.7	70.6	23.7	59.4	30.6	4.7	18.2	55.4	15.1	34.3

Performance evaluation of continual class segmentation on the PASCAL VOC 2012 dataset.



Performance of 15-1 overlapped procedure.



Robustness analysis under five random class order.