INCREMENTAL LEARNING TECHNIQUES FOR SEMANTIC SEGMENTATION

Umberto Michieli and Pietro Zanuttigh {michieli, zanuttigh}@dei.unipd.it lttm.dei.unipd.it/paper_data/IL/

DEPARTMENT OF INFORMATION ENGINEERING UNIVERSITY OF PADOVA

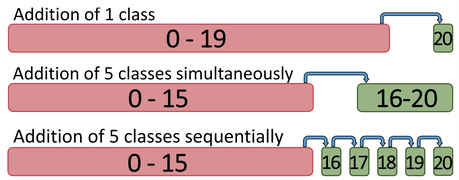
ICCV 2019 Seoul, Korea

Abstract

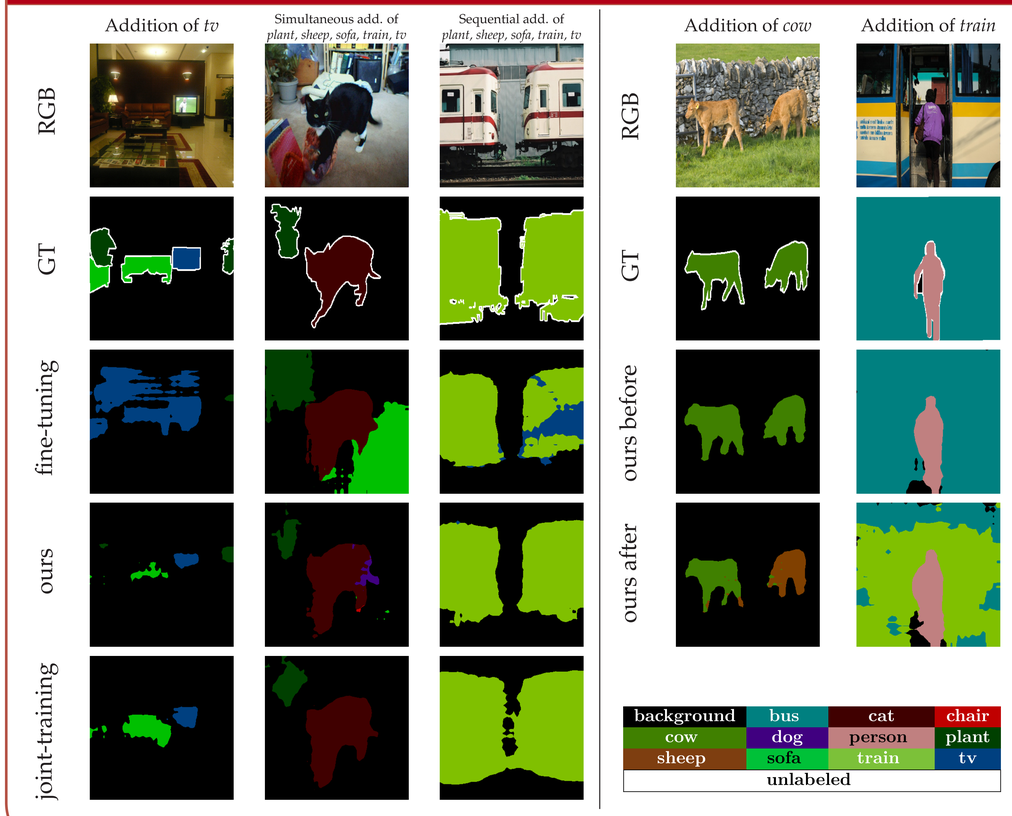
Deep learning architectures exhibit a critical drop of performance due to catastrophic forgetting when they are required to incrementally learn new tasks. Contemporary incremental frameworks focus on object classification or detection while in this work we formally introduce the incremental learning problem for semantic segmentation in which a pixel-wise labeling is considered. To tackle this task we propose to freeze part of the network and to distill the knowledge from output logits or from intermediate features of the previous model to retain the information about previously learned classes, whilst updating the current model to learn new ones. In opposition to recent methods, we do not store any image from previous training steps and only the last model is needed for adaptation. The experimental evaluation on VOC2012 shows the validity of the proposed methods.

Problem Formulation

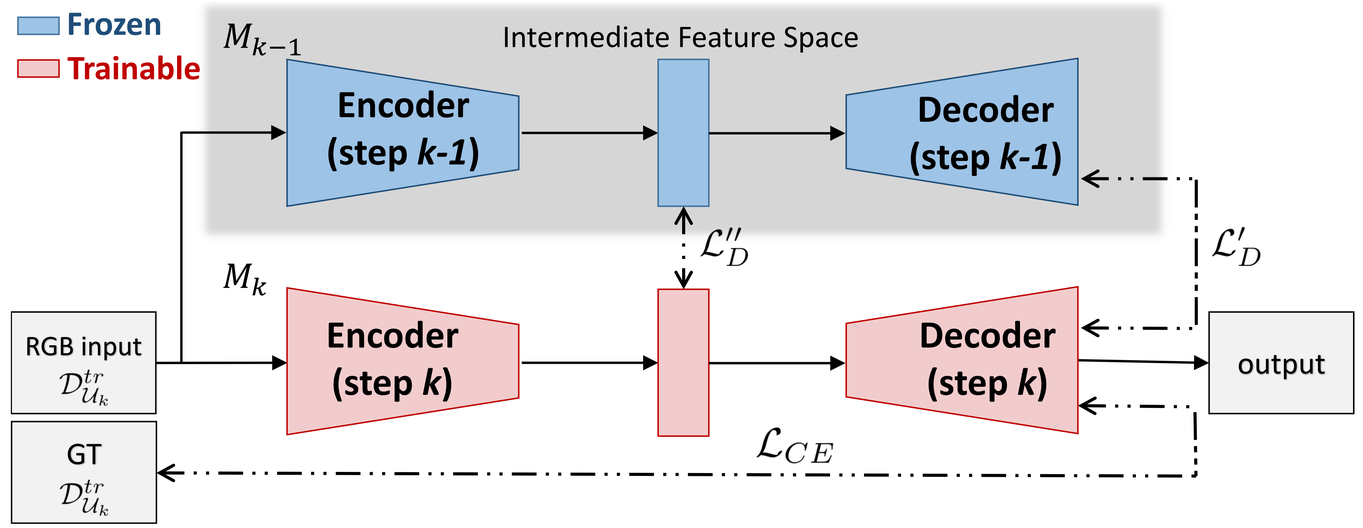
- Initial training on $\mathcal{S}_0$ classes using dataset $\mathcal{D}_0^{tr}$ (pixels only belong to classes in $\mathcal{S}_0$)
- Incremental step k to learn *unseen* classes $\mathcal{U}_k$ using dataset $\mathcal{D}_k^{tr}$ which could contain few and correlated pixels of previous classes



Qualitative Results



Proposed Approach



Cross-entropy (CE) loss

$$\mathcal{L}_{CE} = -\frac{1}{|\mathcal{D}_k^{tr}|} \sum_{\mathbf{X}_n \in \mathcal{D}_k^{tr}} \sum_{c \in \mathcal{S}_{k-1} \cup \mathcal{U}_k} \mathbf{Y}_n[c] \cdot \log\left(M_k\left(\mathbf{X}_n\right)[c]\right)$$

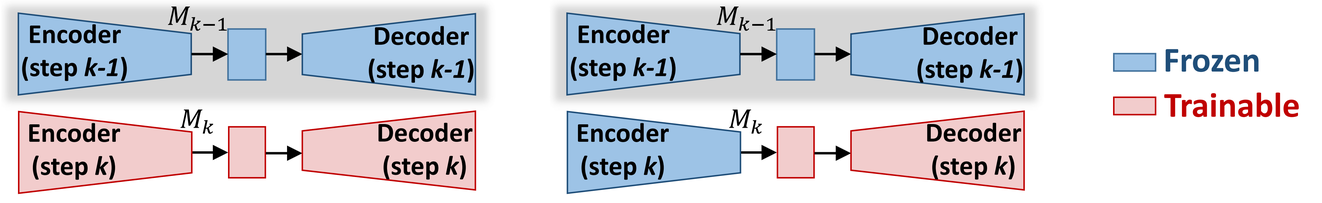
Distillation as CE on previous and current softmax

$$\mathcal{L}'_D = -\frac{1}{|\mathcal{D}_k^{tr}|} \sum_{\mathbf{X}_n \in \mathcal{D}_k^{tr}} \sum_{c \in \mathcal{S}_{k-1}} M_{k-1}(\mathbf{X}_n)[c] \cdot \log\left(M_k(\mathbf{X}_n)[c]\right)$$

Distillation as L2-norm on previous and current features

$$\mathcal{L}''_D = \frac{\|E_{k-1}(\mathbf{X}_n) - E_k(\mathbf{X}_n)\|_2^2}{|\mathcal{D}_k^{tr}|}$$

Freezing Schemes



Quantitative Results

	Addition of *tv*			Simultaneous addition of *plant, sheep, sofa, train, tv*			Sequential addition of *plant, sheep, sofa, train, tv*		
mIoU	old	new	all	old	new	all	old	new	all
Fine-tuning	67.3	20.1	65.1	60.6	38.7	55.4	47.8	33.0	44.2
$\mathcal{L}'_D$	70.0	35.3	68.4	**71.0**	**49.0**	**65.7**	57.4	34.9	52.0
$E_F, \mathcal{L}'_D$	72.5	**51.4**	71.5	70.0	45.8	64.2	**70.5**	44.2	**64.3**
$\mathcal{L}''_D$	**72.6**	50.6	**71.6**	65.6	43.9	60.5	65.9	**45.4**	61.0
Joint-training	73.7	70.5	73.6	75.1	68.5	73.6	75.1	68.5	73.6