Cross-Dataset Collaborative Learning for Semantic Segmentation

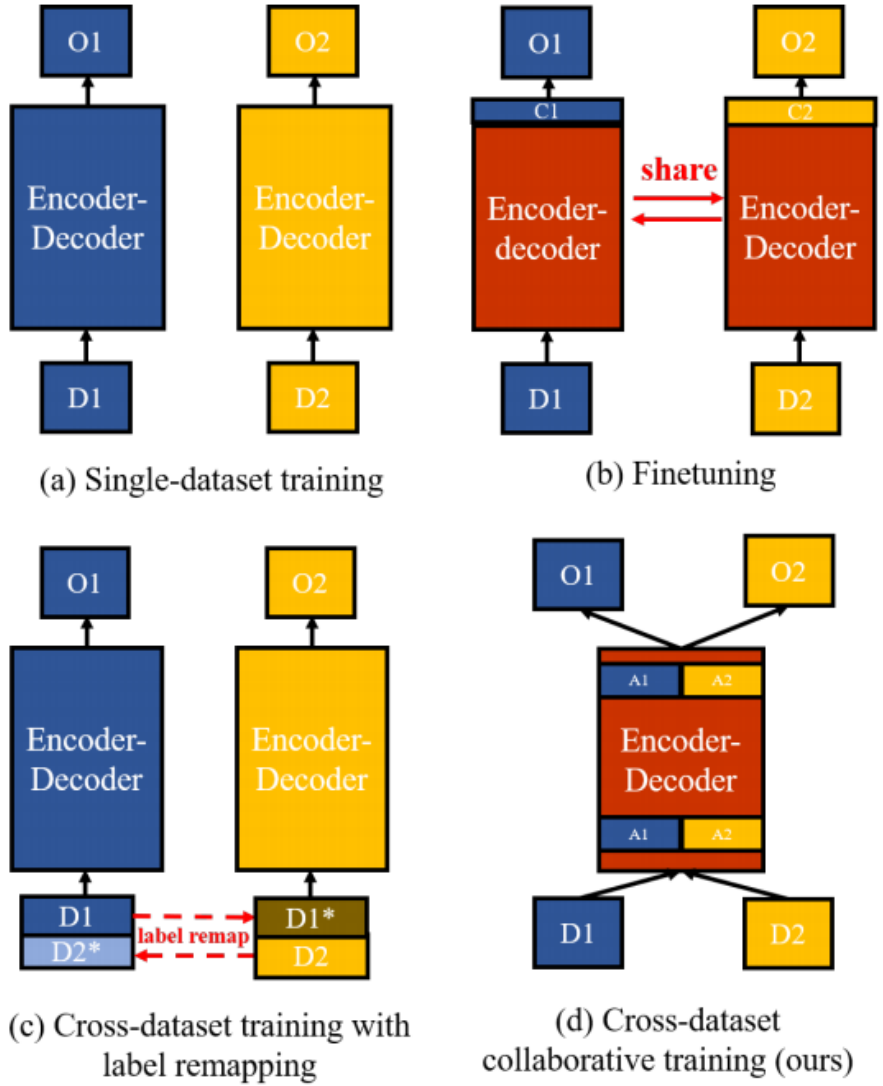
Li Wang [1], Dong Li [1], Yousong Zhu[2], Lu Tian [1], Yi Shan [1]
[1] Xilinx Inc., Beijing, China.
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, China.
{liwa, dongl, lutian, yishan}@xilinx.com, yousong.zhu@nlpr.ia.ac.cn

CVPR 2021

韩坤洋

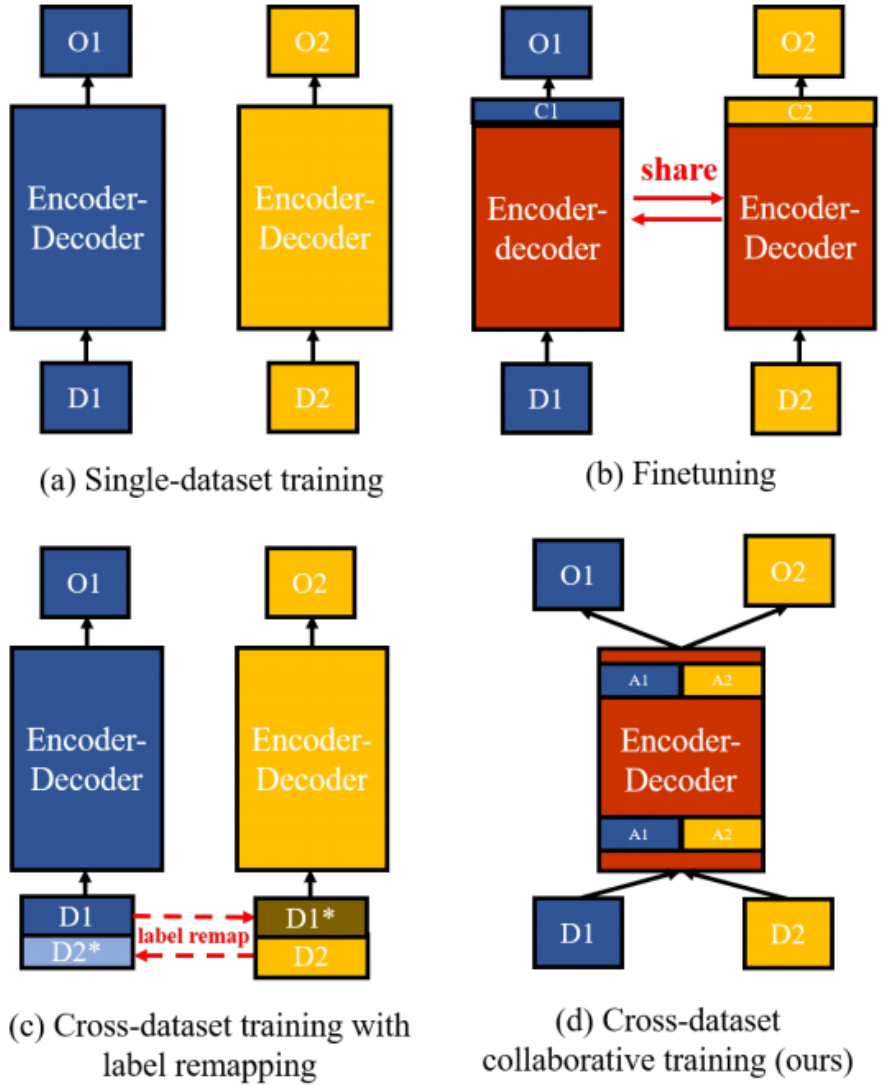
Cross-Dataset Collaborative Learning for Semantic Segmentation

- Proposed a simple, flexible, and general cross-dataset collaborative learning algorithm.



Analysis

- (a) Train and evaluate a network for each dataset separately
- (b) Finetuning
  - Time-consuming
  - Not applicable for joint optimization
  - Manually adjust hyper-parameters
- (c) Generate a hybrid dataset
  - Simple label concatenation
  - Label mapping



Label mapping

- Prior knowledge needed
  - Class duplication and conflict
- Poor performance
  - Discrepancy of camera viewpoints, scenes, etc.

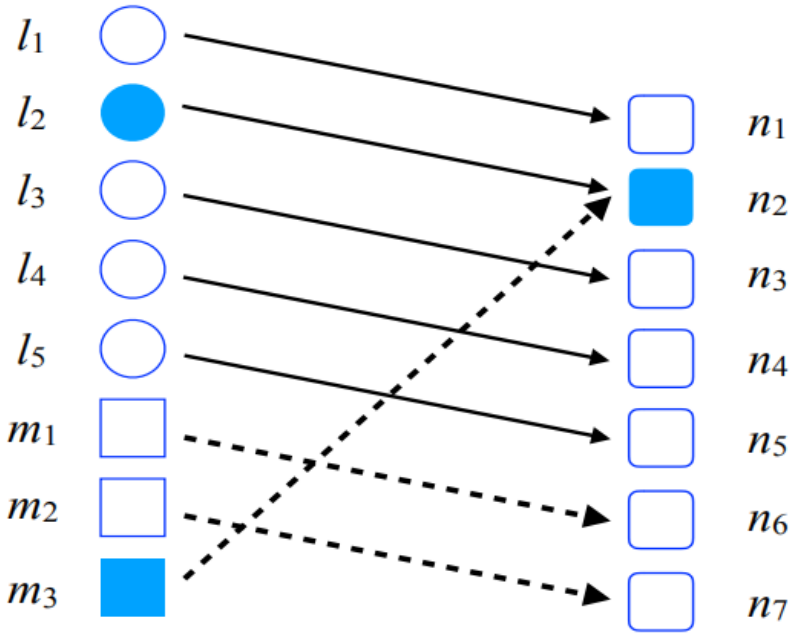


Figure 3: A simple example of label mapping. Two datasets with labels l_1, l_2, l_3, l_4, l_5 (labels are represented by circles) and m_1, m_2, m_3 (represented by rectangles) are merged into a new hybrid dataset with labels $n_1, n_2 .... n_7$ (represented by round rectangles). The solid shapes in blue indicate that the labels belong to the same class that can be merged into one class in label mapping operation. Other labels stay unchanged during mapping.

Cross-Dataset Collaborative Learning

- Different segmentation datasets vary greatly
  - (e.g., scenes and illumination)

(a) Visualization for samples of different datasets

Rethinking BatchNorm and Convolution

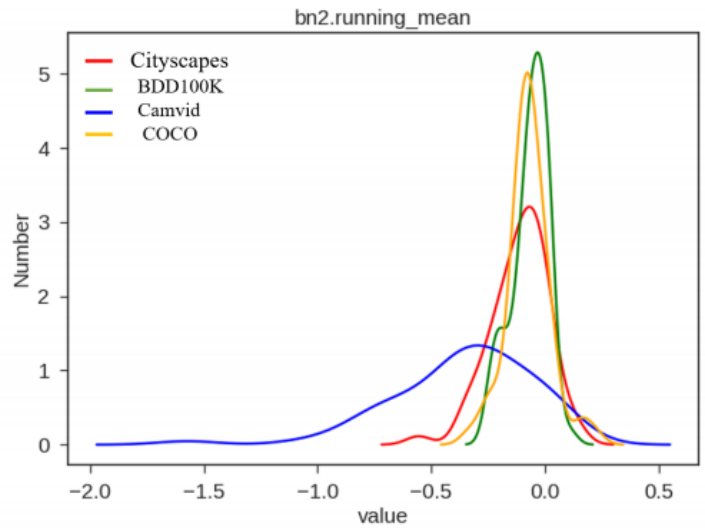
- Analyze the **parameter** distributions of conv and bn layers.
- Conv,
  - **weight**, bias from conv kernel
- BN,
  - **running_var, running_mean**, weight, bias

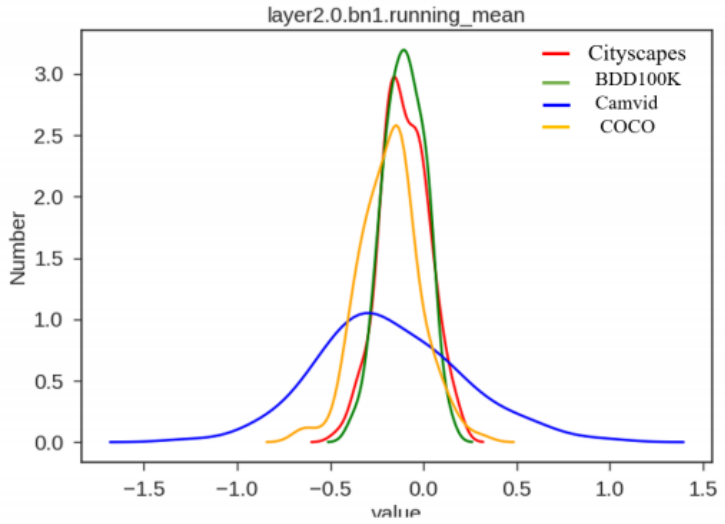
$$y = \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x] + \epsilon}} * \gamma + \beta$$

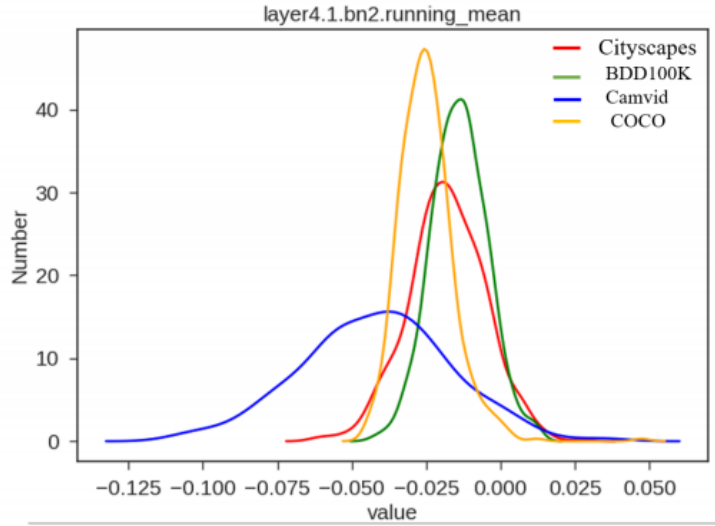
$$\hat{x}_{\text{new}} = (1 - \text{momentum}) \times \hat{x} + \text{momentum} \times x_t$$

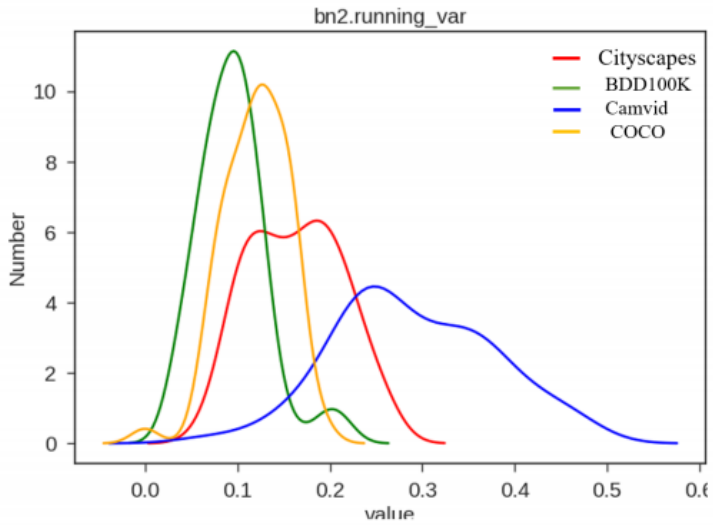
where *x*^ is the estimated statistic and x_t is the new observed value

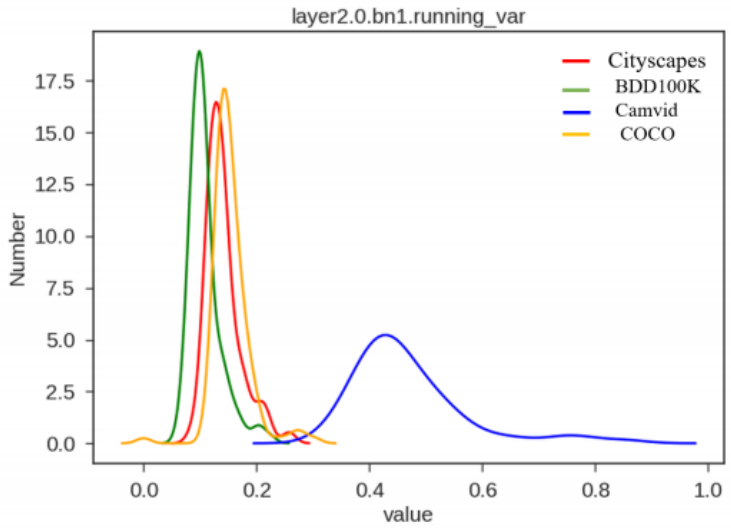
Rethinking BatchNorm and Convolution

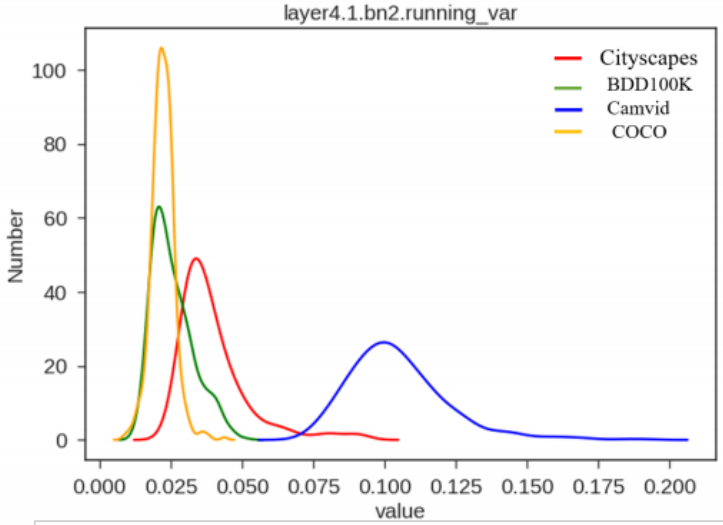




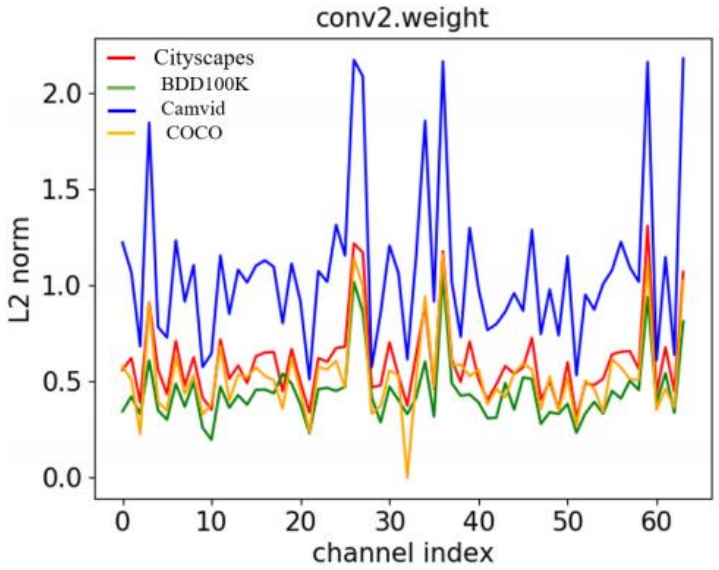


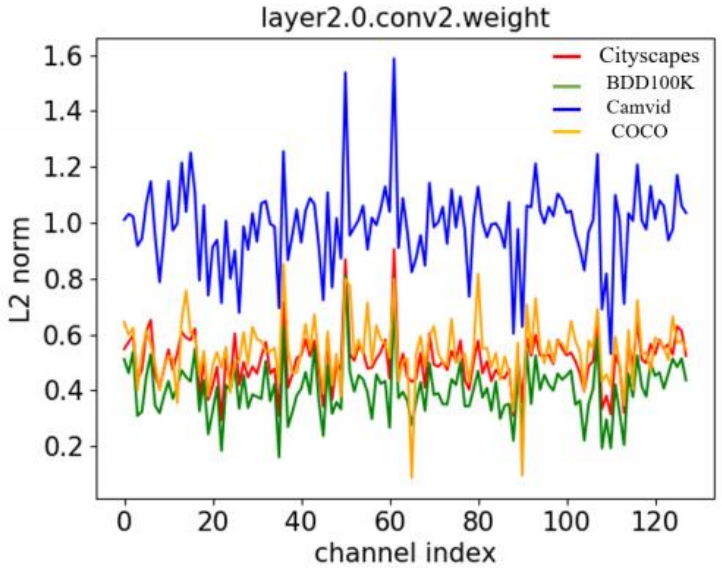


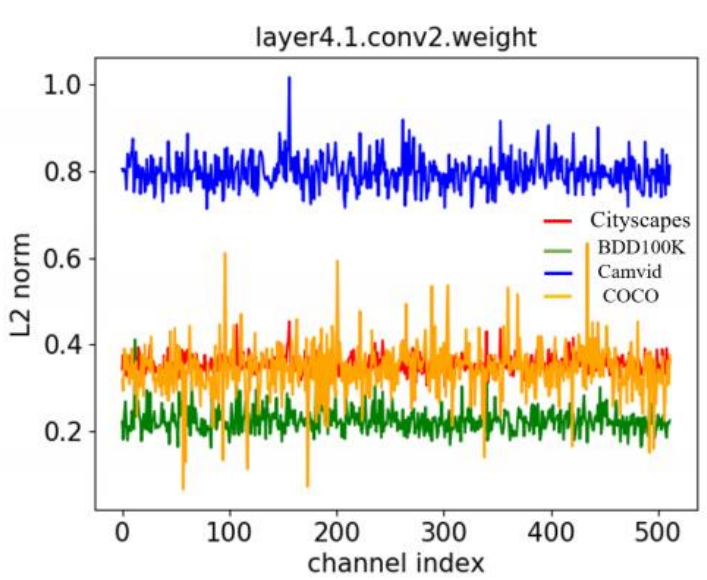




Rethinking BatchNorm and Convolution

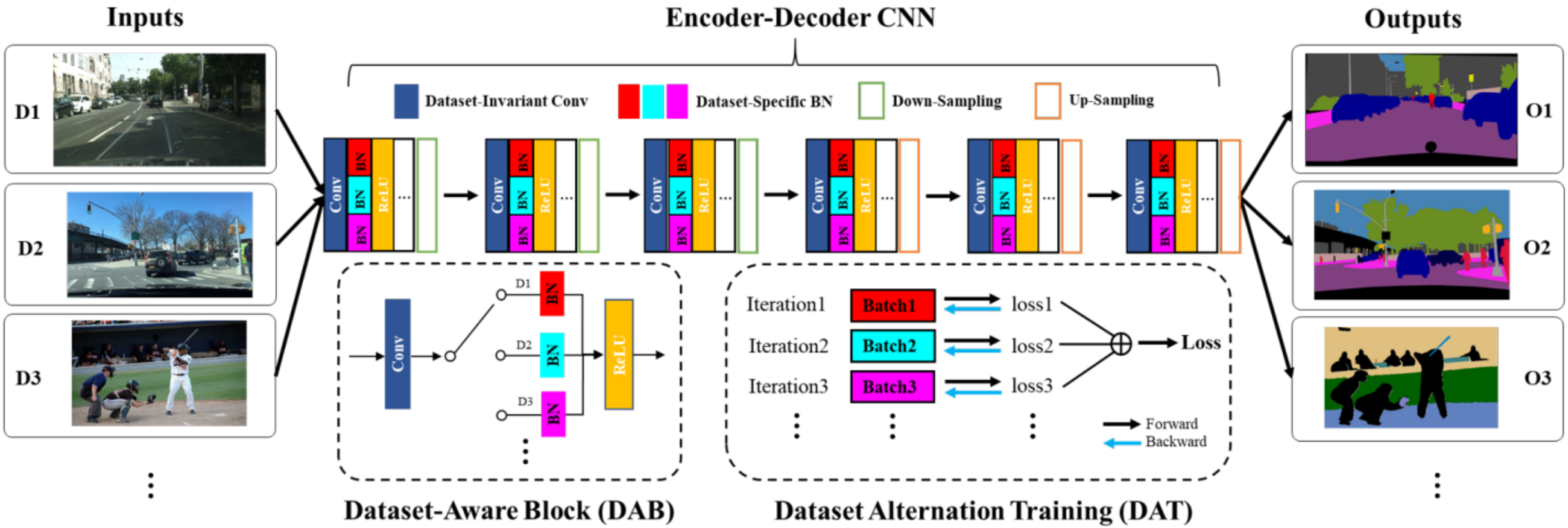






- For Conv, weights hold the same distributions
- For BN, distributions of both running mean and running var have different shapes for different datasets

Network


Inputs
Encoder-Decoder CNN
Outputs
D1
D2
D3
Dataset-Invariant Conv
Dataset-Specific BN
Down-Sampling
Up-Sampling
Conv
BN
ReLU
O1
O2
O3
Iteration1
Iteration2
Iteration3
Batch1
Batch2
Batch3
loss1
loss2
loss3
Loss
Forward
Backward
Dataset-Aware Block (DAB)
Dataset Alternation Training (DAT)

Dataset-Aware Block

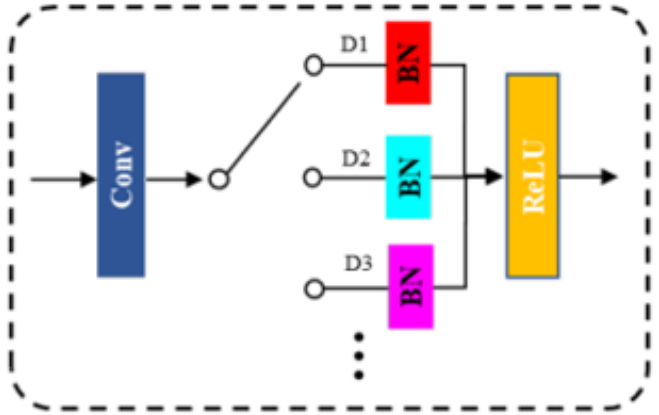
- Dataset-invariant conv layer
- Dataset-specific bn layers
- A switch is automatically to determine which BN should be activated based on the data source (if statement?)
- Dataset-aware classifiers (conv with different out channel?)

$$\mathrm{DSBN}\{D_i\}(X_i;\gamma_i,\beta_i) = \gamma_i \hat{X}_i + \beta_i \tag{1}$$

where,

$$\hat{X}_i = \frac{X_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \tag{2}$$

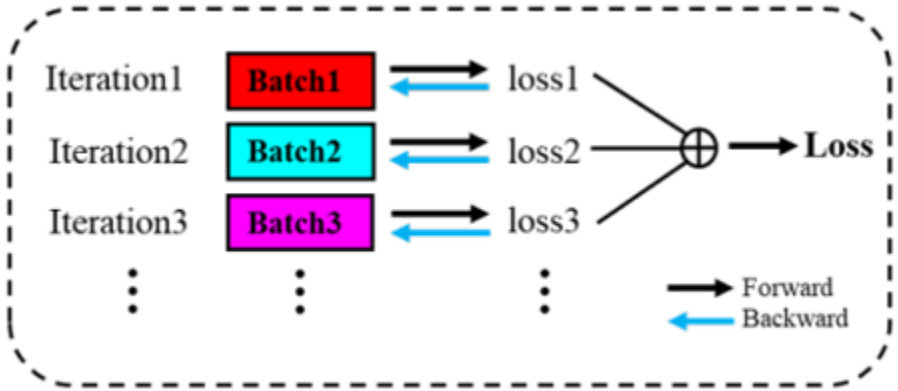
$$\mu_i = \frac{1}{B}\sum_{j=1}^{B} X_i^j,\ \sigma_i^2 = \frac{1}{B}\sum_{j=1}^{B}(X_i^j - \mu_i)^2$$



Dataset-Aware Block (DAB)

Dataset Alternation Training

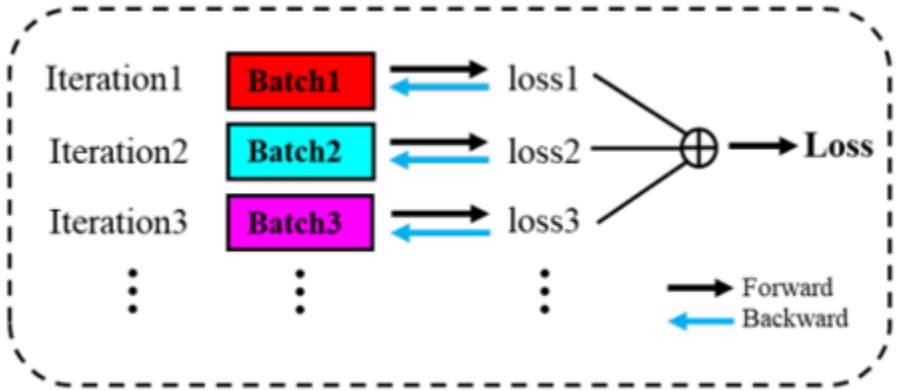
- One iteration
  - Construct the batch with samples from a **single dataset,** compute loss
- Next iteration
  - Samples from **another** dataset, compute loss
- Dataset-number iteration
  - …
  - Accumulate all losses and backpropagate



Dataset Alternation Training (DAT)

Dataset Alternation Training

- Backpropagate the loss of each dataset in each iteration will incur training instability.
- Efficient way to train the samples from multiple datasets



Dataset Alternation Training (DAT)

Result

Method	Cityscapes (%)		BDD100K(%)
	Val.	Test	Val.
Single-dataset	67.52	67.75	53.88
Finetuning	67.79	66.52	58.30
Label remapping	66.23	66.39	58.74
CDCL (Ours)	**72.63**	**71.55**	**60.47**

(a) Cityscapes + BDD100K

Method	Cityscapes (%)		CamVid (%)	
	Val.	Test	Val.	Test
Single-dataset	67.52	67.75	73.05	70.41
Finetuning	67.35	67.87	74.83	71.16
Label remapping	67.13	68.22	78.03	76.86
CDCL (Ours)	**69.77**	**68.56**	**78.45**	**77.34**

(b) Cityscapes + CamVid

Method	Cityscapes (%)		COCO (%)
	Val.	Test	Val.
Single-dataset	67.52	67.75	32.86
Finetuning	70.44	70.23	32.10
Label remapping	50.35	52.39	32.67
CDCL (Ours)	**72.63**	**72.52**	**32.87**

(c) Cityscapes + COCO

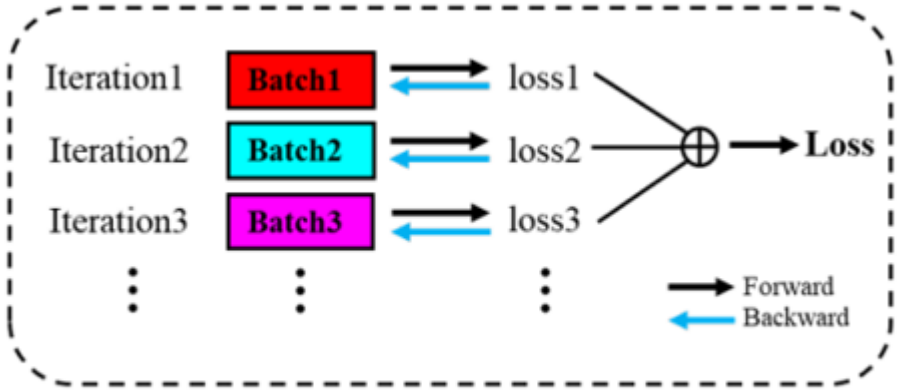
Result

Method	Cityscapes (%)		CamVid (%)		BDD100K (%)
	Validation	Test	Validation	Test	Validation
Single-dataset	67.52	67.75	73.05	70.41	53.88
CDCL (Ours)	73.17 **(+5.65)**	70.98 **(+3.23)**	78.84 **(+5.79)**	75.52 **(+5.11)**	60.45 **(+6.57)**

Result

Method	Norm	DAT	Cityscapes	BDD100K
Single-dataset	BN		67.75%	53.88%
Cross-dataset	BN		62.10%	53.34%
Cross-dataset	BN	✓	64.69%	56.33%
Cross-dataset	DSBN		68.55%	58.93%
CDCL (Ours)	DSBN	✓	**72.63%**	**60.47%**

Table 3: Ablation studies on DSBN and DAT with the ResNet-18 backbone on the Cityscapes and BDD100K validation sets.



Dataset Alternation Training (DAT)