ID:1699

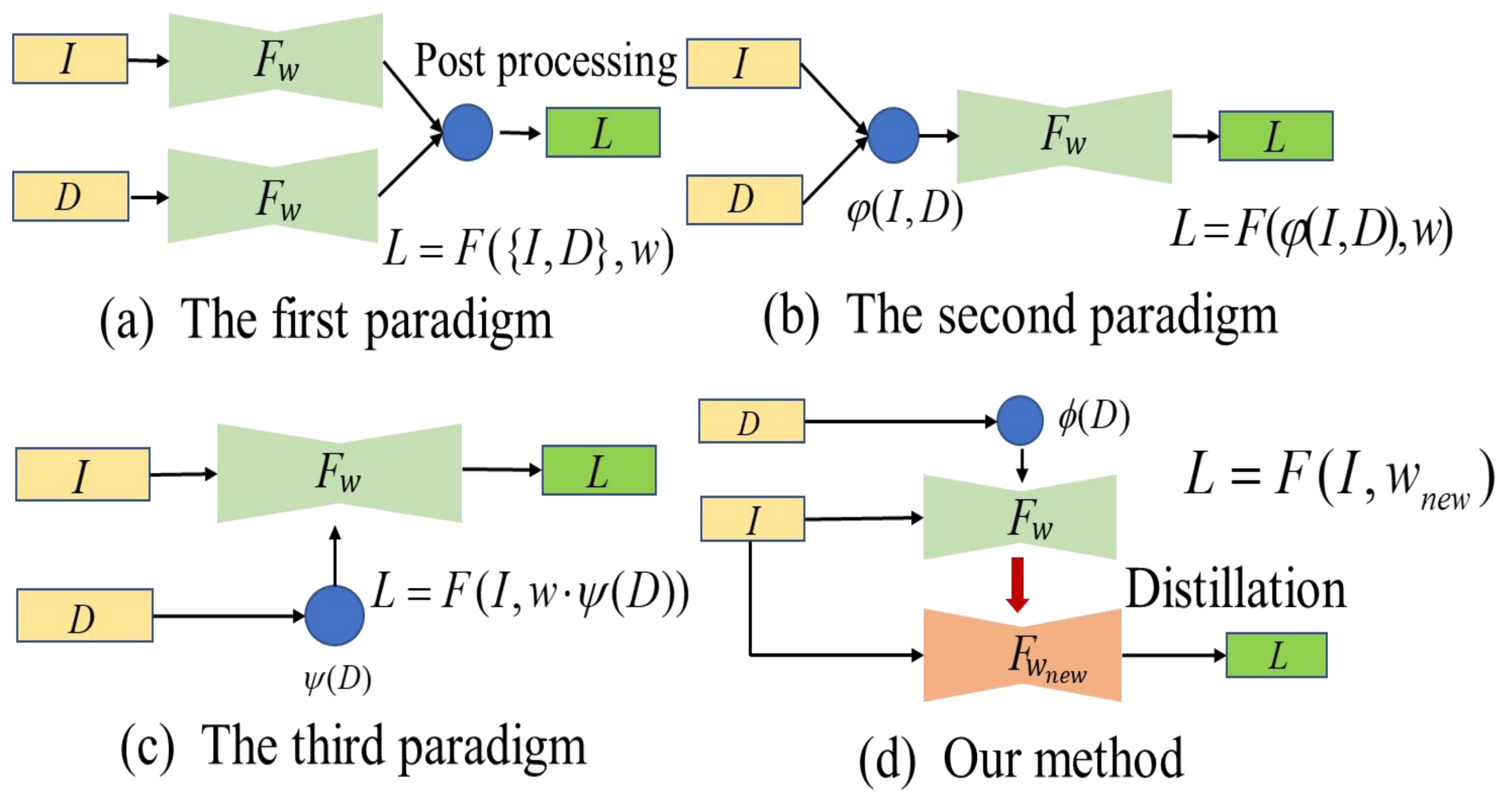
Depth Removal Distillation for RGB-D Semantic Segmentation

Tiyu Fang[1], Zhen Liang[1], Xiuli Shao[2], Zihao Dong[1]*, Jinping Li[1]*
[1]School of Information Science and Engineering, University of Jinan, Jinan 250022, China
[2]College of computer science, Nankai University, Tianjin 300350, China

Abstract

Most of RGB-D semantic segmentation methods need to acquire the real depth information for segmenting RGB images effectively. Therefore, it is extremely challenging to take full advantage of RGB-D semantic segmentation methods for segmenting RGB images without the depth input. To address this challenge, a general depth removal distillation method is proposed to remove depth dependence from RGB-D semantic segmentation model by knowledge distillation, which can be employed to any CNN-based segmentation network structure.

Introduction

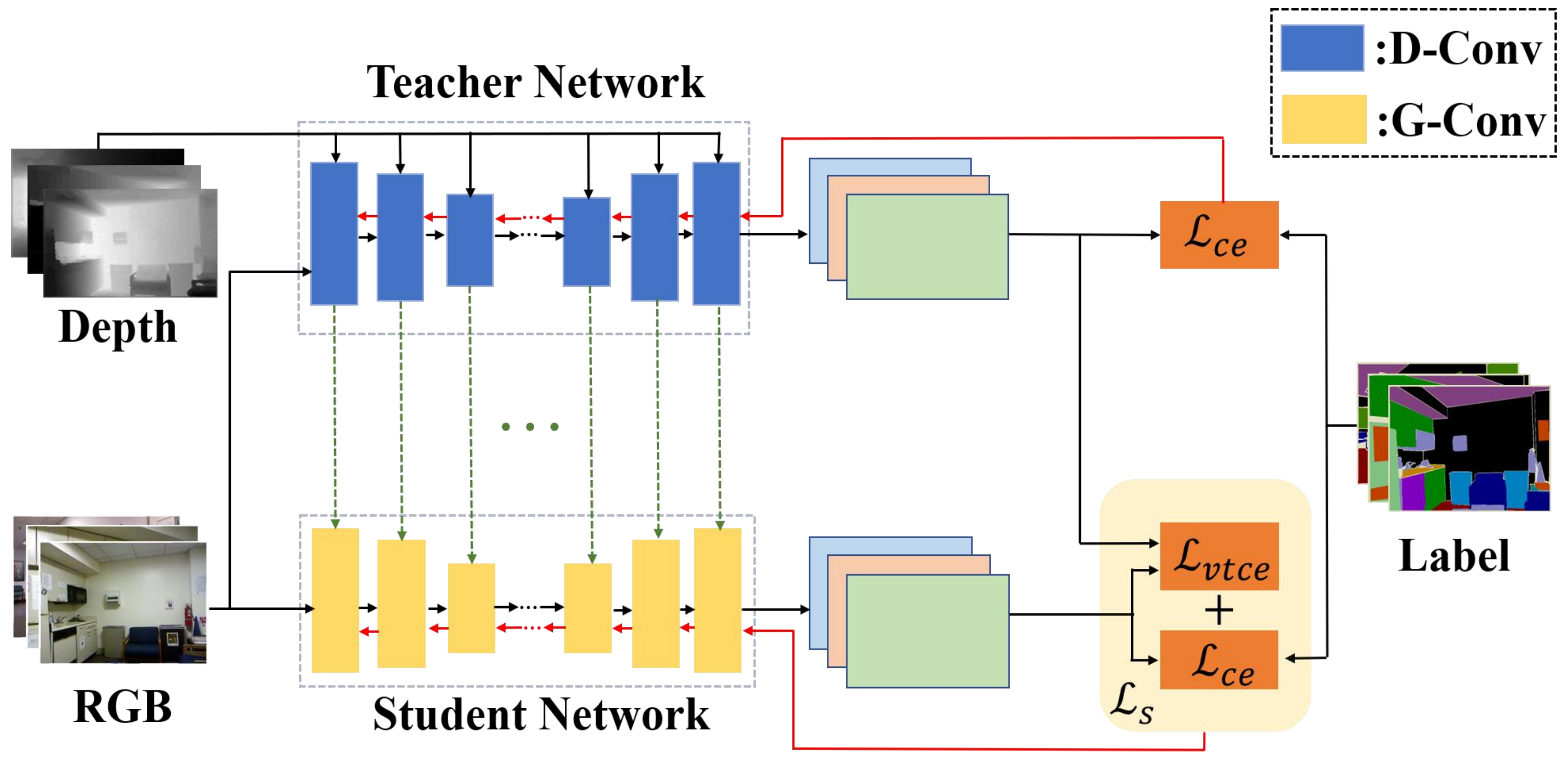


(a) The first paradigm
(b) The second paradigm
(c) The third paradigm
(d) Our method

The paradigms of RGB-D semantic segmentation. *I*,*D*,*L* denote RGB Image, Depth, Label.

(a) I and D are used as the input of network respectively, the segmentation results are obtained by combining the output of I and D.
(b) I and D are fused as the input of network by preprocessing operation, such as HHA image.
(c) D becomes an auxiliary factor to optimize the weight w instead of the input of the network.

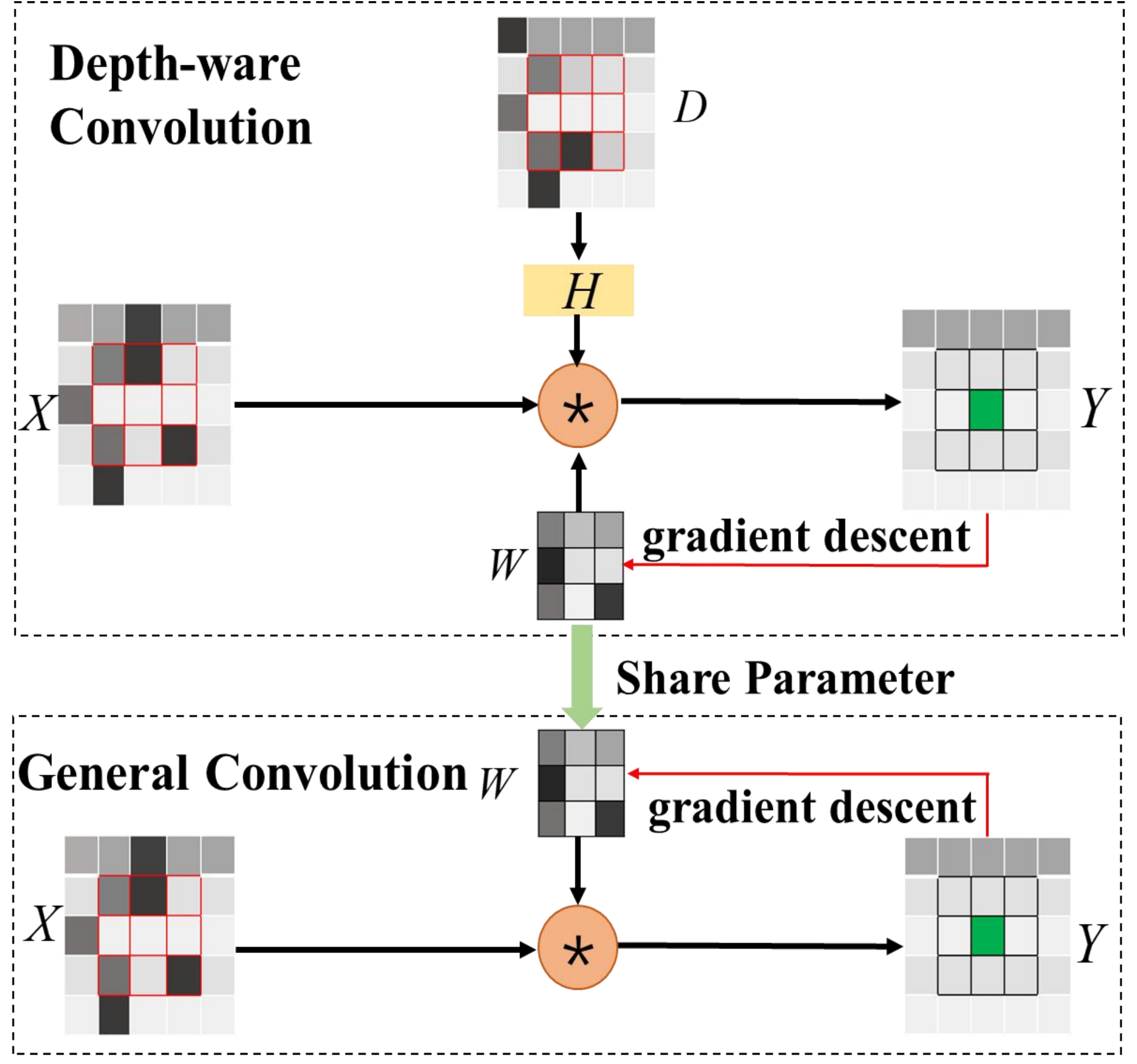
Overall Architecture



The overall architecture of our proposed method.

The proposed method is divided into two parts: teacher network and student network. Depth-aware convolution (D-Conv) is adopted to construct teacher network and general convolution (G-Conv) is used to construct student network with the same structure as teacher network.

Method Innovation



Sharing Parameters

Variable Temperature Cross Entropy

$$L_{vtce} = -\sum_{i=1}^{n} T q_i^t \log(q_i^s)$$

$$q_i = \frac{e^{z_i/T}}{\sum_{j=1}^{n} e^{z_j/T}}$$

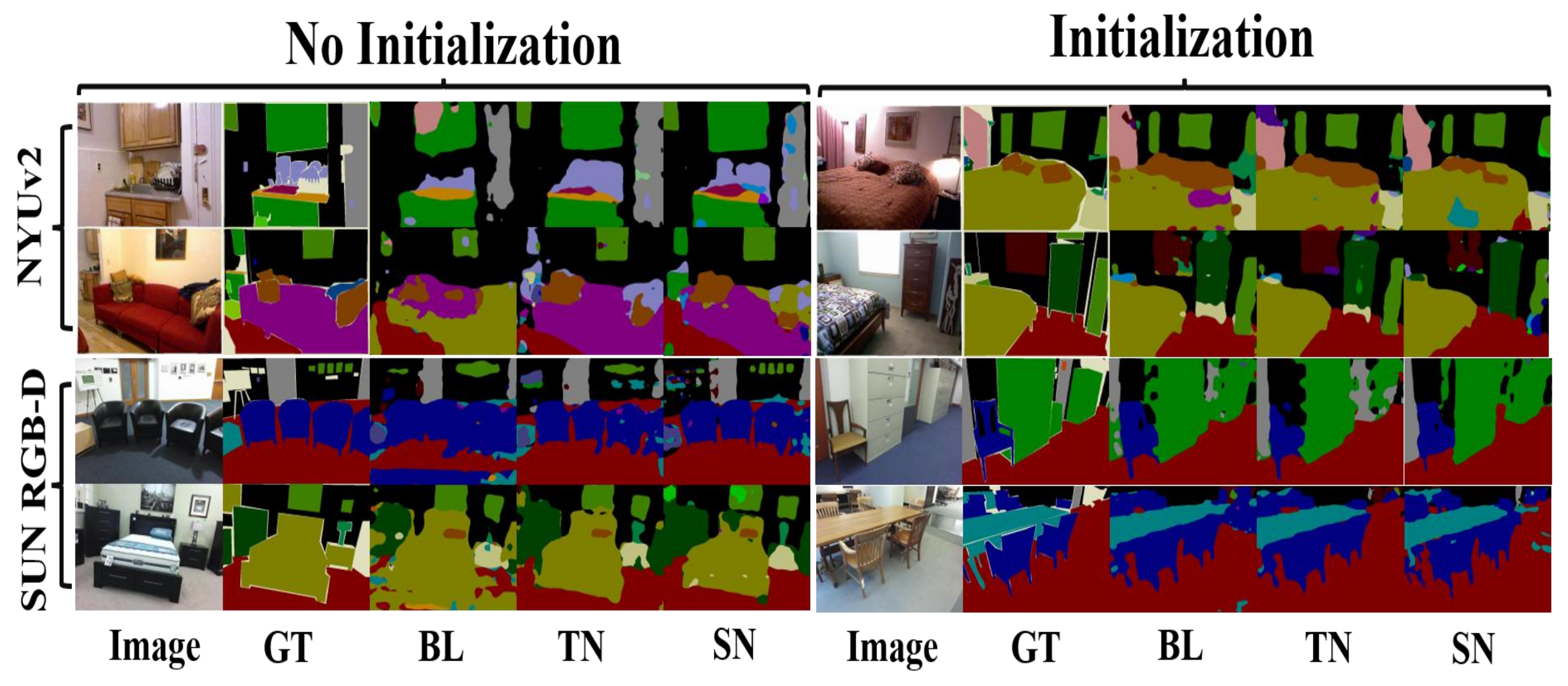
T is a variable to control the impact of teacher network on the student network

Loss Distillation

A depth-aware convolution is adopted to construct the teacher network for getting knowledge from RGB-D images. the teacher network is used to transfer the learned knowledge to the student network with general convolutions by sharing parameters and loss distillation.

Experiments

	NYUv2					
	No Initialization			Initialization		
	BL	TN	SN	BL	TN	SN
Input Data	RGB	RGB-D	RGB	RGB	RGB-D	RGB
mPA(%)	22.8	48.2	**51.0**	35.0	51.6	**51.0**
mIoU(%)	15.2	32.3	**38.1**	24.6	39.1	**38.2**
	SUN RGB-D					
	No Initialization			Initialization		
	BL	TN	SN	BL	TN	SN
Input Data	RGB	RGB-D	RGB	RGB	RGB-D	RGB
mPA(%)	31.6	40.4	**39.3**	39.8	50.8	**48.9**
mIoU(%)	22.9	30.8	**28.5**	31.7	41.0	**39.5**



GT: Ground Truth, BL: BaseLine, TN: Teacher Network, SN: Student Network