A Real-time Bilateral Segmentation Network from Scene to Instance with the Connected Components Postprocessing

NO.50   21821209 Shuai Xie 11821042 Na Yu
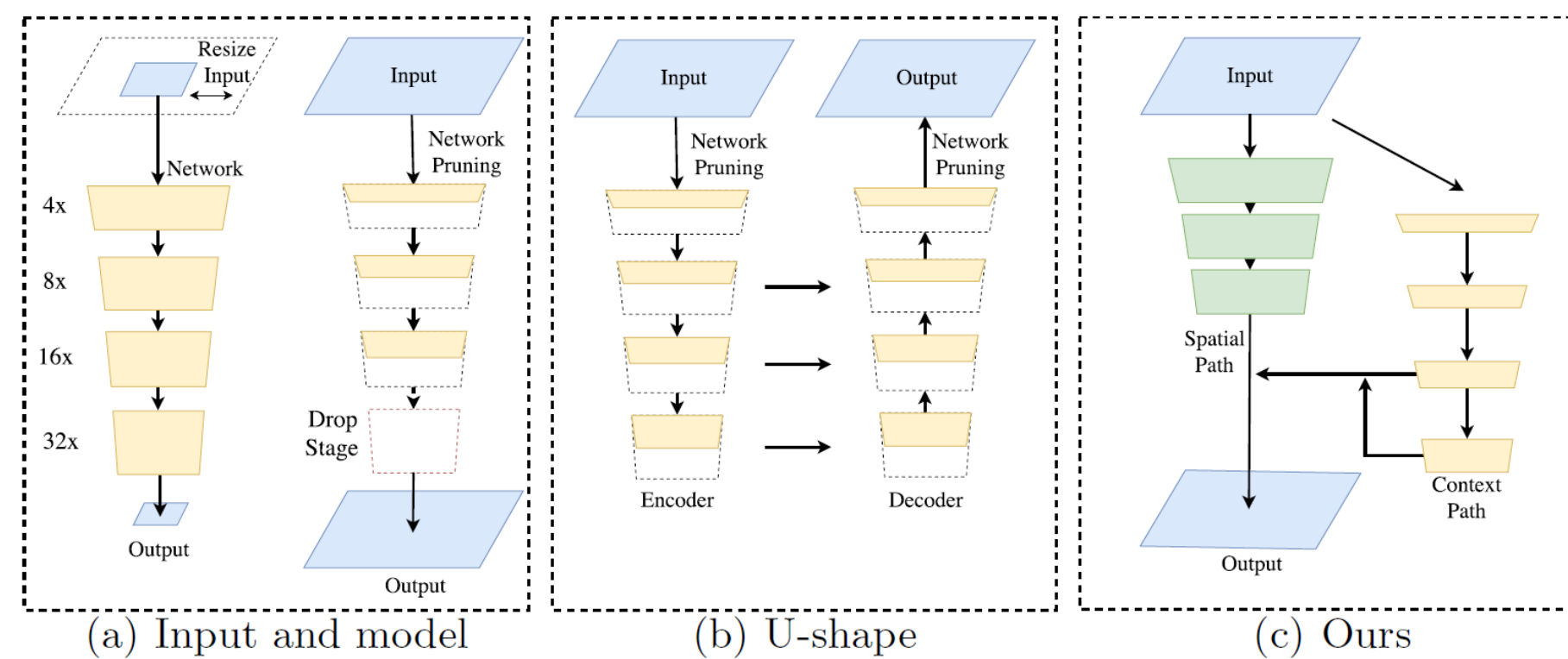
Computer Science Department, Zhejiang University

VIPA
Visual Intelligence and Pattern Analysis

MOTIVATION

➢ **Accelerate the model**

- Cropping → Loss spatial details → Accuracy decrease
- Prune → Weakens the spatial capacity
- Drop the last stage of the model → Narrow the receptive field → A poor discriminative ability

➢ **U-shape structure**

- High-resolution feature maps → Reduce the speed
- Most spatial information lost cannot be easily recovered by involving the shallow layers.

(a) Input and model   (b) U-shape   (c) Ours

CONTRIBUTIONS

① Train Bilateral Segmentation Network (BiSeNet) on different scene segmentation datasets.

② Achieve impressive results on **ADE20K, CamVid, and SUN-RGBD datasets.**

③ Expand BisSeNet from Scene Segmentation to Instance Segmentation via Connected Components Postprocessing.

METHODOLOGY

➢ **BiSeNet**

Bilateral Segmentation Network (BiSeNet) is designed to segment the scenes in Real-time, proposed by Face++.

Architecture

(a) Network Architecture   (b) Attention Refinment Module   (c) Feature Fusion Module
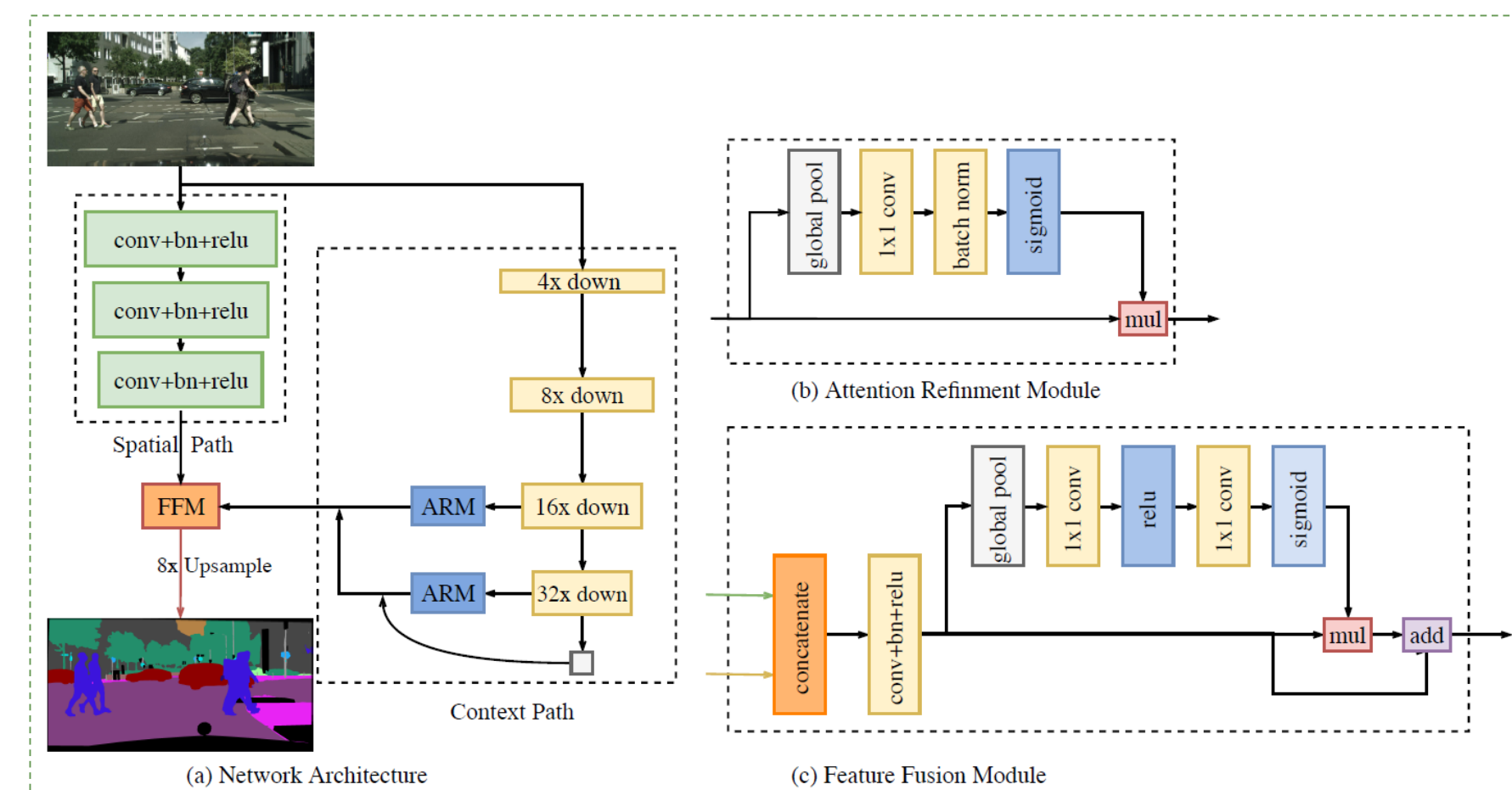
Loss Definition

$$loss = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -\log\left(\frac{e^{p_i}}{\sum_j e^{p_j}}\right)$$

$$L(X;W) = l_p(X;W) + \alpha \sum_{i=2}^{K} l_i(X_i;W)$$

➢ **Connected Components Postprocessing**

As the segmentation result of each class is a Binary Image, Connected Components Decision can be used to segment the class superpixel to instance superpixels.
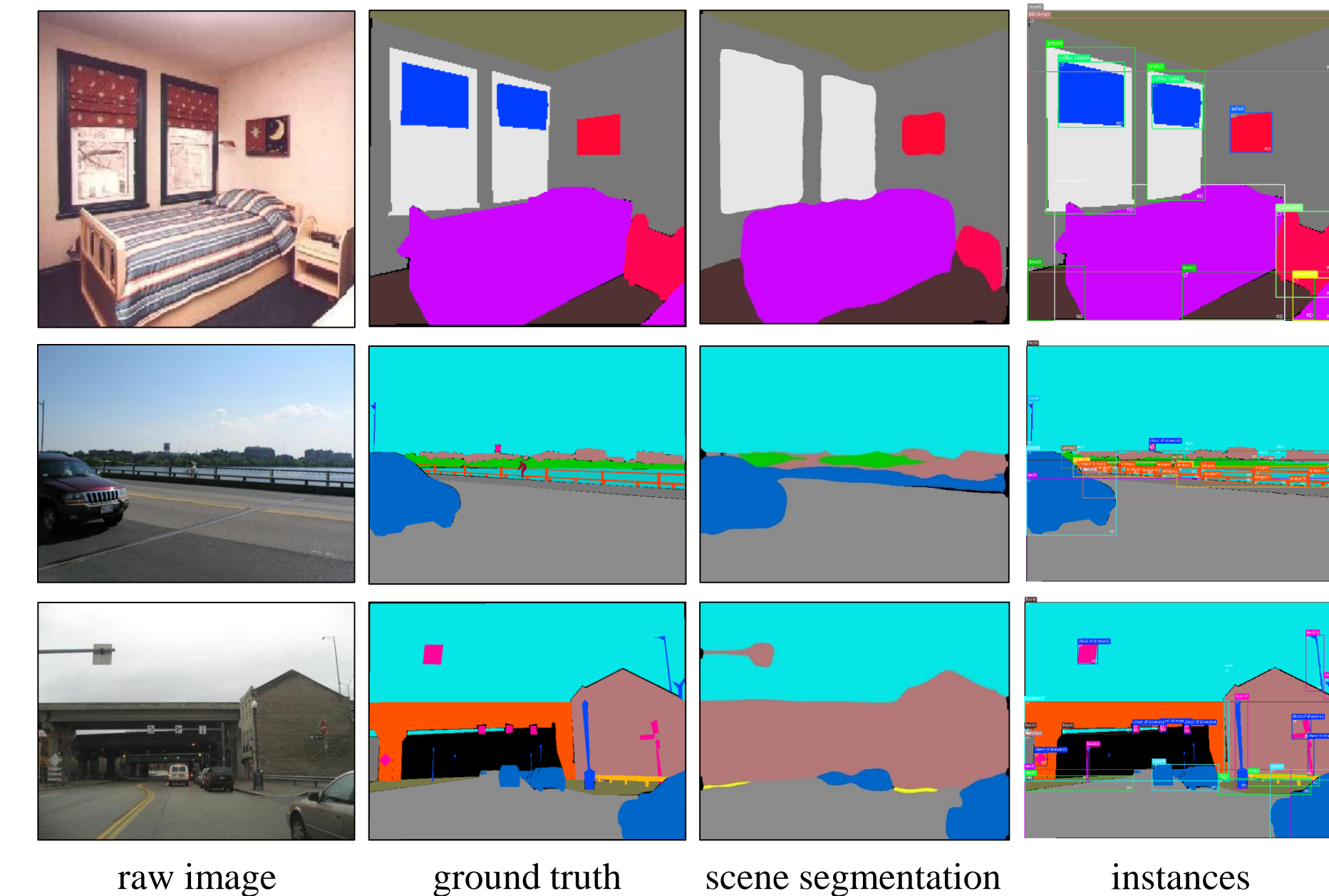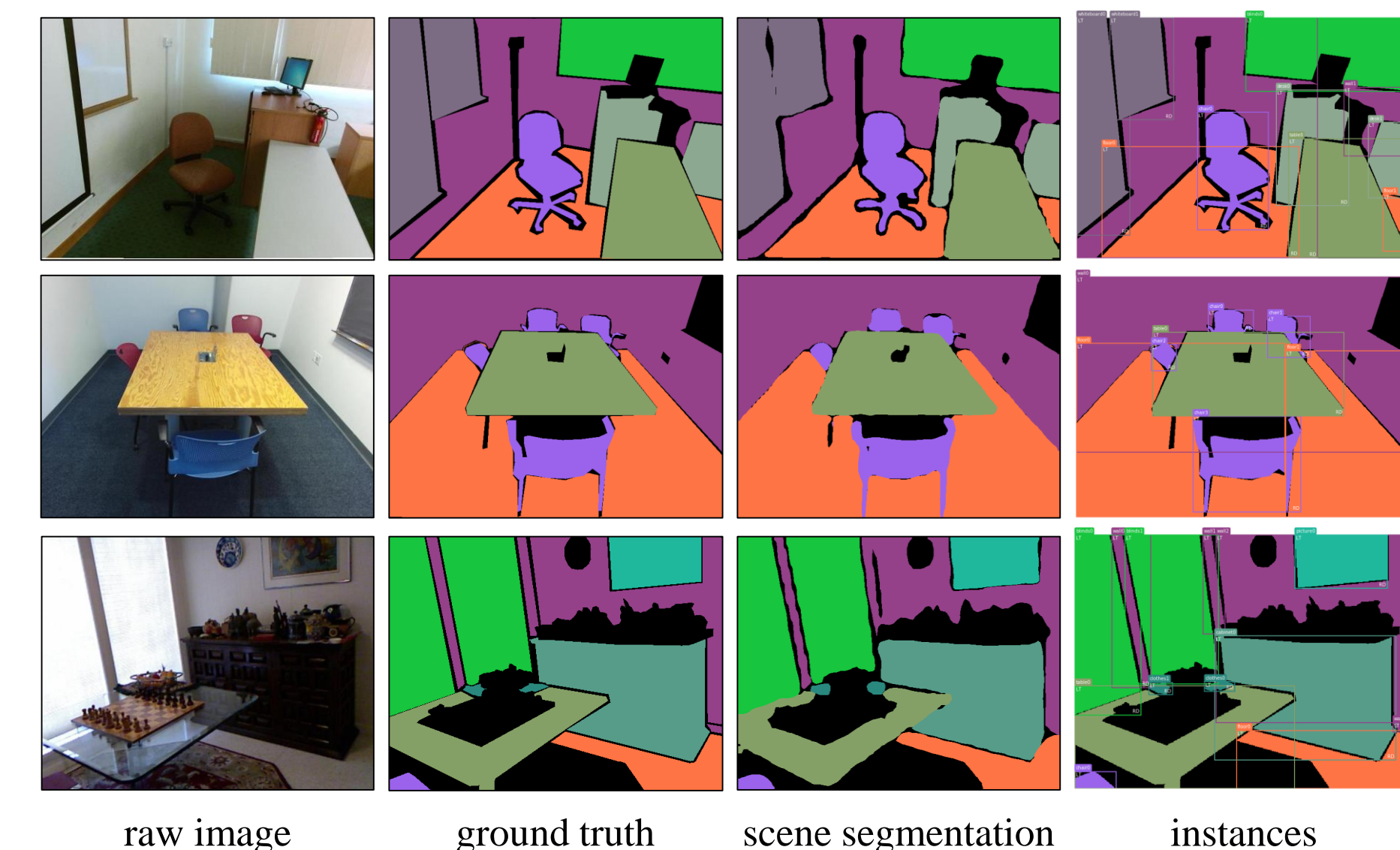

4-connect   8-connect   example
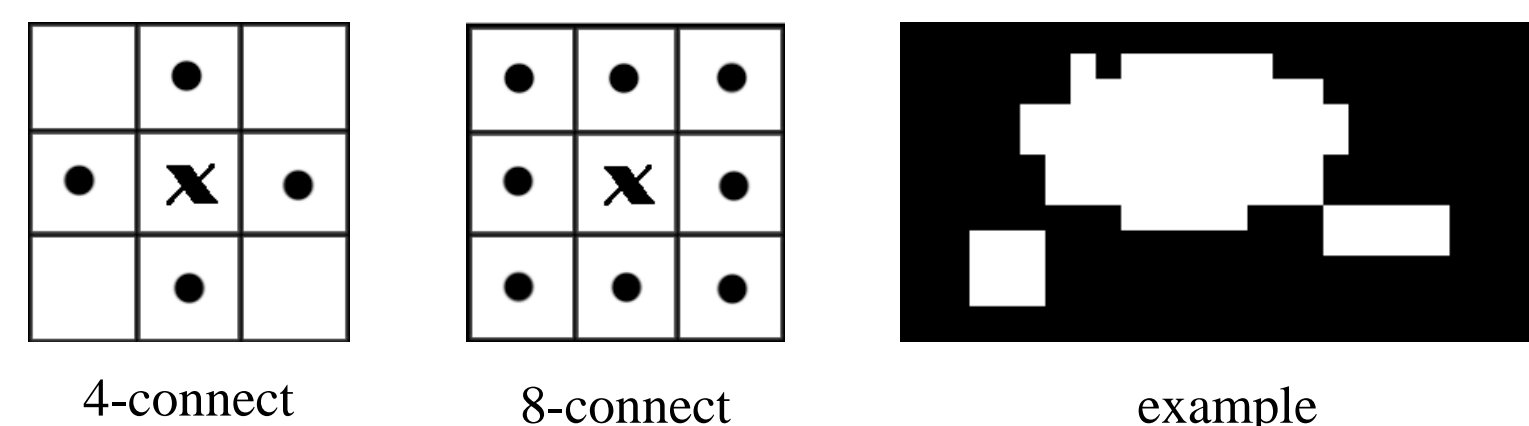
EXPERIMENTS

➢ **Accuracy & Speed**

DataSet	Accuracy	Speed/fps	Memory Usage
ADE20K	0.648	12.7	2100 MB
SUN-RGBD	0.769	12.6	2100 MB

(Note: image size = 480x640, train ADE20K 44 epochs, SUN-RGBD 290 epochs)

➢ **ADE20K**

raw image   ground truth   scene segmentation   instances

➢ **SUN-RGBD**

raw image   ground truth   scene segmentation   instances