PIXEL-LEVEL DATA AUGMENTATION FOR SEMANTIC IMAGE SEGMENTATION USING GENERATIVE ADVERSARIAL NETWORKS

Shuangting Liu[1], Jiaqi Zhang[1], Yuxin Chen[1], Yifan Li[3], Zengchang Qin[1], Tao Wan[2]

[1] Intelligent Computing and Machine Learning Lab; [2] School of Biological Science and Medical Engineering, Beihang University, Beijing, 100191, China;
[3] The University of Adelaide, Adelaide, South Australia 5005 Australia
Contact info: zcqin@buaa.edu.cn

Introduction

Objective

To improve the accuracy of semantic segmentation by data augmentation approach to balance the label distribution using Generative Adversarial Networks (GANs).

Context

- The label distribution in most datasets are unbalanced.
- Manual pixel-level annotations of dataset have a huge cost.

Previous Work

- Photographic image generation by conditional GANs, e.g. Ting-Chun Wang *et al.* [1].
- Image-level label balance to improve the emotion classification accuracy[2].

Contributions

- Data augmentation pipeline by using GANs to generate supplementary data for semantic segmentation task.
- New method for image augmentation at pixel-level.

Feasibility Analysis

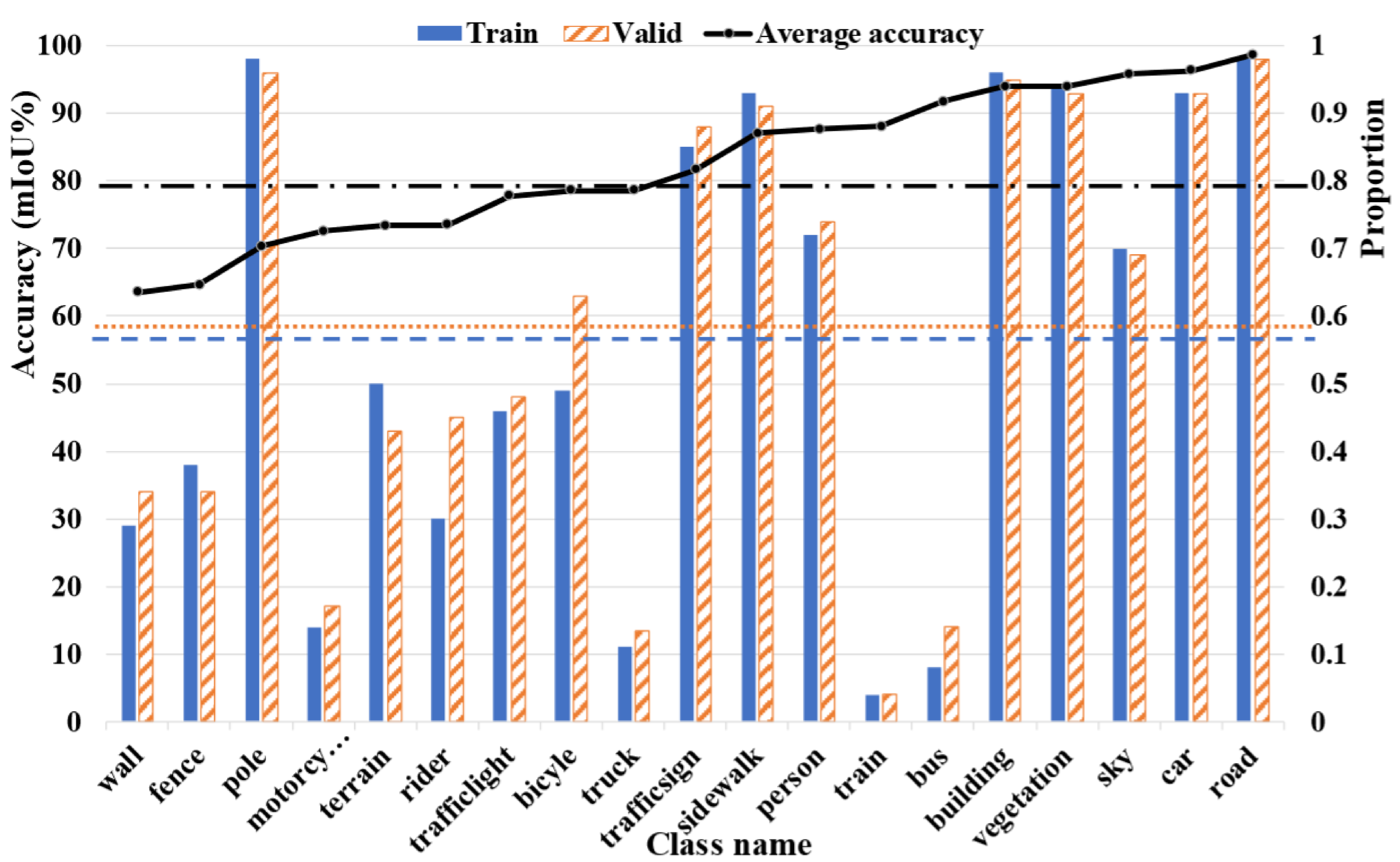

Figure 1. Average segmentation accuracy of top 5 ranked models on Cityscapes website and the proportion of each class in Cityscapes dataset.

Correlation:

Comparing those classes with low appearance frequency and those have low segmentation accuracy, we find out that two groups are highly overlapped. In other word, it is possible to balance the dataset and further improve segmentation accuracy by increasing the appearance frequency of some specific classes.

Methodology

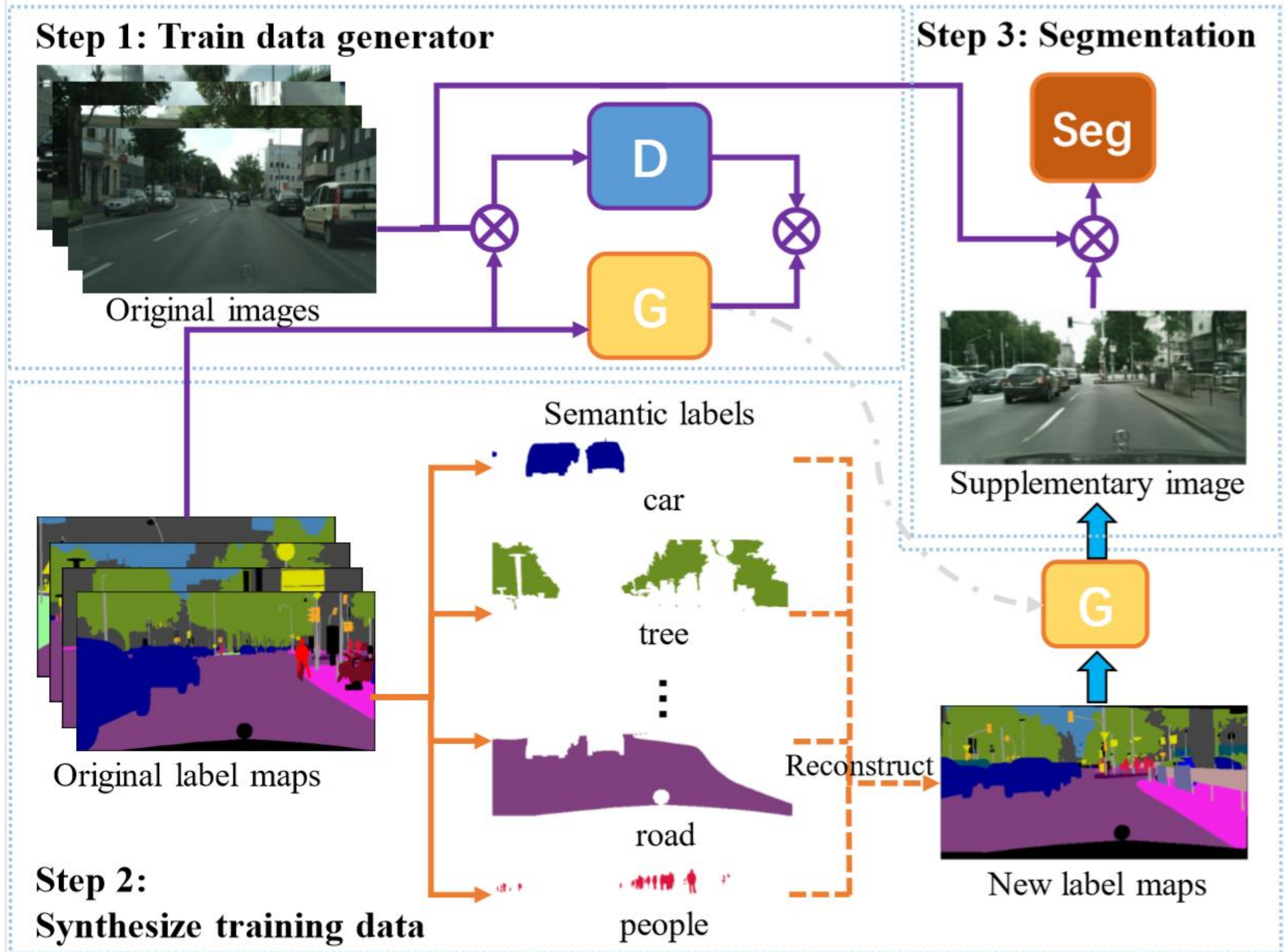

Figure 2. Pipeline of proposed method for generating supplementary data for semantic segmentation to balance the distribution of semantic labels and improve the segmentation results.

Step 1: Train data generator

- Pix2pix HD[1] model is used to generate realistic images on condition of specific semantic label map.
- Minimax algorithm is used to train the generator G and discriminator D.

Step 2: Synthesize training-pair

- To begin with, separating each semantic label map in training set according to class of labels to form single-class label maps.
- Then, recombining these single-class label maps to form reconstructed label maps (Using three different reconstructed ways).
- Lastly, data generator transfer reconstructed label maps to corresponding photographic images.

Step 3: Segmentation validation

- Train semantic segmentation network with/without the generated supplementary dataset. PSPNet [3] is used as the segmentation model.

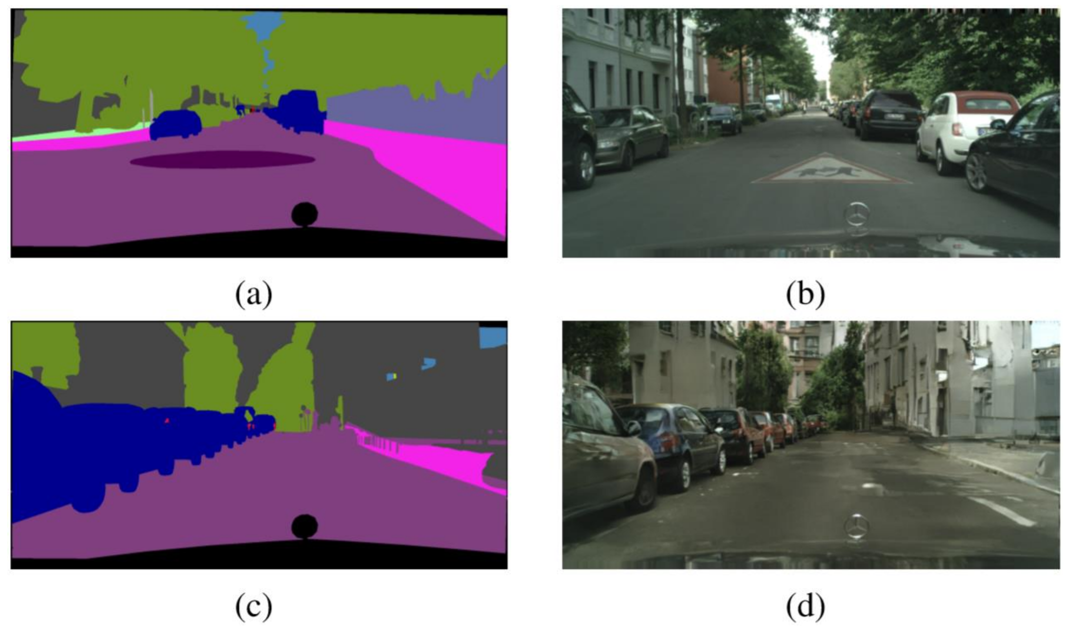

Figure 3. (a)(b) Original image and its corresponding semantic label map;
(c)(d) Reconstruct new semantic label map and its corresponding realistic image.

Experimental Results

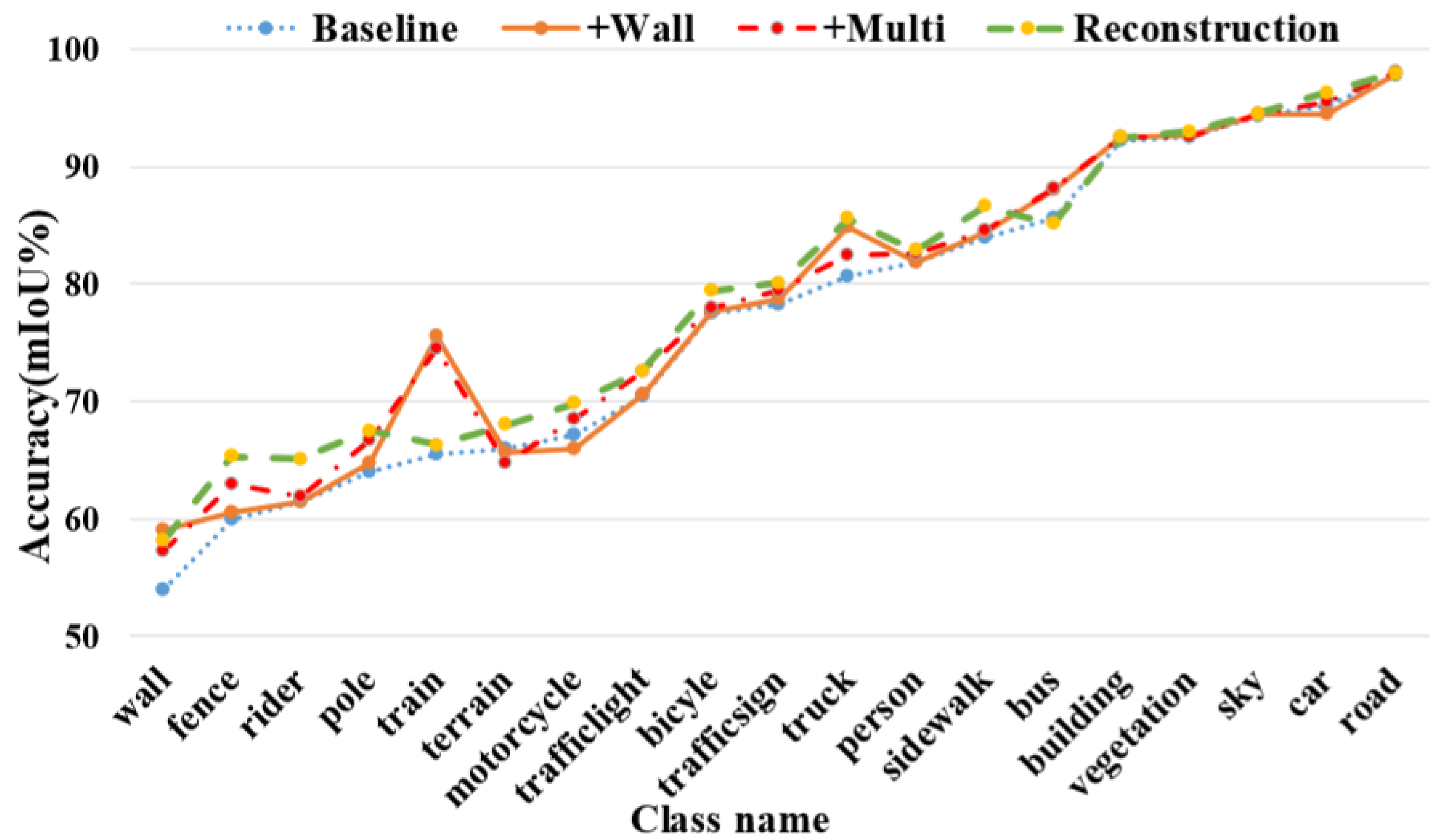

Figure 4. Detailed results comparison. Baseline: only use original Cityscpaes dataset. +Wall: add wall to each image. +Multi: +Wall +Fence +Pole +Traffic light +Train. Reconstruction: a new label map.

Method	Baseline	Add single	Add multi	Recon
mIoU	77.31	78.65	78.82	79.41

Table 1. Results of different generation method.

Method	Tradition	Style Trans	Recon
mIoU	77.31	78.82	79.41

Table 2. Comparison with traditional data augmentation method and style transfer.

Concluding Remarks

- By adding generated label classes to original images as supplementary data, we can improve the diversity of data and balance the data distribution.
- The results shown that by our method, the mean accuracy of a specific class can increase up to 5.5% and the average segmentation accuracy can increase 2.1%.

Reference

[1] Wang T C, Liu M Y, Zhu J Y, et al. "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," [J]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8798-8807

[2] Zhu X, Liu Y, Li J, et al. "Emotion Classification with Data Augmentation Using Generative Adversarial Networks," [C] Pacific-Asia Conference on Knowledge Discovery & Data Mining (PAKDD), 2018, pp. 349-360

[3] Zhao H , Shi J , Qi X , et al. "Pyramid scene parsing network," [C] The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2881-2890