1. Training Settings and Hyper-parameters

Datasets The Cityscapes dataset is the target domain dataset with 2,957 2048 × 1024 training images and 500 validation images of the same resolution. Cityscapes has 19 categories in total. GTA5 and Synthia are two source domain datasets of computer generated synthetic images, which contain 24,966 1914 × 1052 training images and 9400 1280 × 760 training images respectively.

Category Centers The category center \( f_c \) for each category is updated using source features at the beginning of each learning step defined in Section 3.1 (Coarse-to-Fine Pipeline). They are introduced to make the calculation of category-oriented triplet loss practical. If we use the original triplet loss without category centers, we need to store pairwise distances among all pixels with a tremendous GPU memory overload (HW x HW x c).

Hyper-parameters The values of \( P_h \) and \( p \) are chosen according to the study in [2]. Experiments suggest mIoU is not sensitive to the values of hyper-parameters. For example, when \( P_h \) varies between 0.7 and 0.9 and \( p \) varies between 5 and 20, the final performance ranges from 55.3% to 56.1%. The degree of performance improvement varies from 4.3% in the first learning step to 0.1% in the last step.

2. Cross-Category Margin for the Target Domain

Two category-level feature distribution regularization methods are proposed in our work. In the source domain, a category-oriented triplet loss is introduced to enlarge category-level margins. In the target domain, a domain consistency regularization is used to minimize intra-domain variances. The category-oriented triplet loss is applied to the source domain samples only because it relies heavily on labeled hard samples which are only available in the source domain. Although the category-oriented triplet loss is only applied to the source domain, it cooperates with the consistency regularization to enlarge category-level margins in the target domain as well. We propose a metric called cross-category margin (CCM) for the target domain to verify our claim. Let \( x_{i,j} \) denote the pixel feature at location \((i, j)\) in a target domain image, the corresponding ground-truth label is denoted as \( y_{i,j} \). The CCM is defined as

\[
CCM(c) = \frac{1}{S_c} \sum_{S_c} \min_{k \neq c} (\delta_{i,j,k})
\]

\[
\delta_{i,j,k} = \frac{||x_{i,j} - f_{c=k}\| - ||x_{i,j} - f_{c=y_{i,j}}\||}{||f_{c=y_{i,j}} - f_{c=k}\||}
\]

where \(CCM(c)\) is the cross-category margin for category \(c\), \(S_c\) is the set of pixels in category \(c\) across all images in the target domain, \(f_c\) is the category center in the source domain for category \(c\), and \(\delta_{i,j,k}\) is the category-level margin at location \((i,j)\) between groundtruth category \(y_{i,j}\) and another category \(k\). This margin is a normalized one with respect to the category center distance between categories \(y_{i,j}\) and \(k\). We choose the minimum margin, which corresponds to the most confusing class, to be the category-level margin at a certain pixel. Then such margins for all pixels in category \(c\) are averaged to define the margin for category \(c\). Note that the target domain groundtruth labels, \(y_{i,j}\)’s, are not used during training, and are only used to evaluate cross-category margins in the target domain. This metric can be considered as an indicator of category-level feature distributions in the feature space.

We compare cross-category margins computed for the models trained in the ablation study. Compared to the model with coarse alignment only, the model trained with additional target domain consistency regularization achieves improved cross-category margins, and the further integration of the category-oriented triplet loss improves cross-category margins further as shown in Figure 1. By

\*These authors have equal contribution.
\*Corresponding author.
comparing to the experimental results in Table 1 of the main paper, we can draw the conclusion that both our category-level feature distribution regularization methods improve cross-category margins in the target domain.

3. Experiments with DeepLabV2

In addition to the experiments discussed in the main paper, here we discuss experiments using another commonly adopted segmentation backbone DeepLabV2 [6, 7, 2]. As shown in Table 1, although the performance improvement achieved by our proposed framework using this backbone is not as significant as DeepLabV3+, our proposed method still outperforms all previous methods, achieving a new state-of-the-art mIoU (51.9%), which is 1.5% higher than the previous best result using the same backbone on the GTA5→Cityscapes task. On the Synthia→Cityscapes task, our proposed method also achieves a new state-of-the-art mIoU (54.6%), which is 2.1% higher than the previous best result using the same backbone (Table 2). Although DeepLabV3+ is more powerful than DeepLabV2, the performance of our method on the Synthia→Cityscapes task is similar using both backbones. This is because Synthia mostly consists of large objects where high-resolution feature maps of DeepLabV3+ provide limited improvement.

We also compare the mean cross-category margins (CCM) of different methods on both DeepLab V2 and DeepLab V3+ models in the following table. We compare CCM with other SOTAs[7, 9, 6]. Our proposed method improves CCM on different segmentation models. Note that the magnitude of CCM highly depends on the choice of models, and we still use mIoU as our main evaluation metric.

4. More Qualitative Samples

More segmentation results are given in this section in addition to the qualitative results in the paper. In the following part, we demonstrate (1) our photometric alignment results for the GTA5→Cityscapes task with frequency alignment results from FDA [8] as a comparison (Figure 2-Figure 17); (2) our photometric alignment results for the Synthia→Cityscapes task (Figure 18-Figure 33); (3) our segmentation results for the GTA5 [4]→Cityscapes [1] task with segmentation results from CAG [9] as a comparison (Figure 34-Figure 49); (4) our segmentation results for the Synthia [5]→Cityscapes [1] task (Figure 50-Figure 65.).
Cross-Category Margin

Normalized Margin

Figure 1: Our proposed category-oriented triplet loss exploits hard samples and further enlarge category margins.

References


Section 4.1
Global Photometric Alignment on GTA2Cityscapes
Figure 2: Aligned sample from GTA5 [4]→Cityscapes [1] task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 3: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 4: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 5: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 6: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 7: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 8: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 9: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 10: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment.
Figure 11: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 12: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 13: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 14: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 15: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 16: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 17: Aligned sample from GTA5→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Section 4.2
Global Photometric Alignment on Synthia2Cityscapes
Figure 18: Aligned sample from Synthia [5] → Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 19: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 20: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 21: Aligned sample from Synthia → Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 22: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment.
Figure 23: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 24: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment.
Figure 25: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 26: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment.
Figure 27: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 28: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 29: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 30: Aligned sample from Synthia → Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 31: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 32: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Figure 33: Aligned sample from Synthia→Cityscapes task: (a) source domain image, (b) target domain reference image, (c) FDA [8], (d) our global photometric alignment
Section 4.3
Sample Segmentation Results on GTA2Cityscapes
Figure 34: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 35: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 36: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 37: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 38: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 39: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 40: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 41: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 42: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 43: Sample segmentation result from GTA5 → Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 44: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 45: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 46: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 47: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 48: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 49: Sample segmentation result from GTA5→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Section 4.4
Sample Segmentation Results on Synthia2Cityscapes
Figure 50: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 51: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 52: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 53: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 54: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 55: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 56: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 57: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 58: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 59: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 60: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 61: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 62: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 63: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours
Figure 64: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d)ours
Figure 65: Sample segmentation result from Synthia→Cityscapes task: (a) input image, (b) ground-truth, (c) CAG[9], (d) ours