Note: Generalizable model-agnostic semantic segmentation via target-specific normalization

sgc

May 22, 2022

1 Introduction

This paper proposed a new method including new paradigm and test strategy to alleviate the problem that model works worse in unseen domains. And furthermore a memory bank is developed to obtain more accurate statistics for normalization. First this method compared UDA(unsupervised domain adaptation) and DG(domain generalization), the key difference of this two method is that whether the target domain data can be used or not while training.

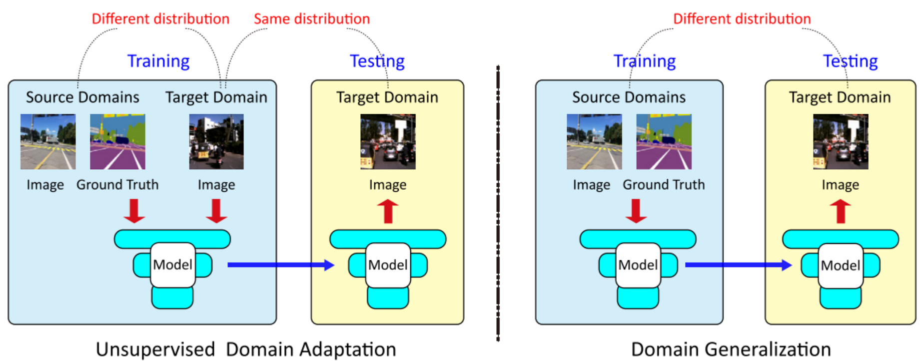

Fig. 1. Comparison between UDA and DG.

Contributions:
1. A novel domain generalization framework that jointly exploits the model-agnostic training scheme and the target-specific normalization test strategy is developed to address the generalizable semantic segmentation task. 2. An image bank is utilized to deal with the inaccurate statistics estimation for a single sample in the test stage. 3. This proposed method achieved a new benchmark.

2 Method

2.1 Model-agnostic meta-learning in DG

Sample data from each seen domain and separate into 2 sets (training and testing). Then there are 2 types of loss: $L_{ds} = L(B_{tr}; \theta, \phi)$ where θ and ϕ are the parameter of feature extractor and classifier of the network to be trained. And $L_{dg} = L(B_{te}; \theta', \phi')$ where θ', ϕ' are updated parameters through back-propagation of L_{ds}.
The final meta-learning loss is: $L_{meta} = L_{dg} + \alpha L_{ds}$

Algorithm 1. Model-agnostic learning for generalizable semantic segmentation.

Input source data $\mathcal{D}_{src}$, network parametrized by θ, ϕ, hyperparameters η, α, γ

Output the trained network

```
repeat
    sample a mini-batch $\mathcal{B}$ (i.e., a task) from $\mathcal{D}_{\int\nabla\rfloor}$
    random split $\mathcal{B}$ into meta-train $\mathcal{B}_{\sqcup\nabla}$ and meta-test data $\mathcal{B}_{\sqcup\rceil}$
    $\mathcal{B}_{\sqcup\nabla} \cap \mathcal{B}_{\sqcup\rceil} = \emptyset, \mathcal{B}_{\sqcup\nabla} \cup \mathcal{B}_{\sqcup\rceil} = \mathcal{B}_{\int\nabla\rfloor}$
    compute domain-specific loss: $\mathcal{L}_{\lceil\int} = \mathcal{L}(\mathcal{B}_{\sqcup\nabla}; \theta, \phi)$
    $\theta = \theta - \eta * \nabla_\theta \mathcal{L}_{\lceil\int}$
    $\phi = \phi - \eta * \nabla_\phi \mathcal{L}_{\lceil\int}$
    compute domain-generalization loss: $\mathcal{L}_{\lceil\}} = \mathcal{L}(\mathcal{B}_{\sqcup\rceil}; \theta, \phi)$
    compute the overall loss: $\mathcal{L}_{\Updownarrow\rceil\sqcup\dashv} = \mathcal{L}_{\lceil\int} + \alpha \mathcal{L}_{\lceil\}}$
    $\theta = \theta - \gamma * \nabla_\theta \mathcal{L}_{\Updownarrow\rceil\sqcup\dashv}$
    $\phi = \phi - \gamma * \nabla_\phi \mathcal{L}_{\Updownarrow\rceil\sqcup\dashv}$
until converge
```

2.2 Target-specific normalization

In conventional normalization, accumulated mean $\hat{\mu}$ and variance $\hat{\delta}^2$ in the training stage are use in test stage. But in this paper, due to the difference between seen domain and unseen domain authors use the target-specific normalization by obtaining new mean and variance for mini-batch with M samples in the test stage:

$$\begin{aligned} \bar{\mu}_c &= \frac{1}{MHW} \sum_{n=1}^{M} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{n,c,w,h}; \\ \bar{\sigma}_c^2 &= \frac{1}{MHW} \sum_{n=1}^{M} \sum_{h=1}^{H} \sum_{w=1}^{W} (x_{n,c,w,h} - \bar{\mu}_c)^2. \end{aligned}$$

$$\hat{x}_{n,c,h,w} = \frac{x_{n,c,h,w} - \bar{\mu}_c}{\sqrt{\bar{\sigma}_c^2 + \epsilon}} w_c + b_c.$$

Authors issued a remark that TN dose not help to improve the performance of a model trained by the standard method. And the size of mini-batch has impact on final performance.

2.3 Image bank

This paper utilized a image bank to aid the estimation of a new image. The image bank has a fixed size of Q. When a new image comes, first combine it with several precious images to produce its statistics.

Style-based selection policy To deal with the problem that images in the image bank from different domain would lower the accuracy for current image, Style-based selection policy is introduced to select images that are similar to the current image.
Adopt the statistics $s = (\mu, \delta)$ to represent the style information. To maintain the detailed information, Authors extract the features generated by the first layer of ResNet. Generate the symmetric KL Divergence between the styles of current images and images in image bank. Selecting top-M images with highest similarity.

$$\begin{aligned} D(s_i||s_j) &= KL(s_i||s_j) + KL(s_j||s_i); \\ KL(s_i||s_j) &= \log\frac{\sigma_j}{\sigma_i} + \frac{\sigma_i^2 + (\mu_i - \mu_j)^2}{2\sigma_j^2} - \frac{1}{2}. \end{aligned}$$

Concatenate them with input image to generate the statistics.

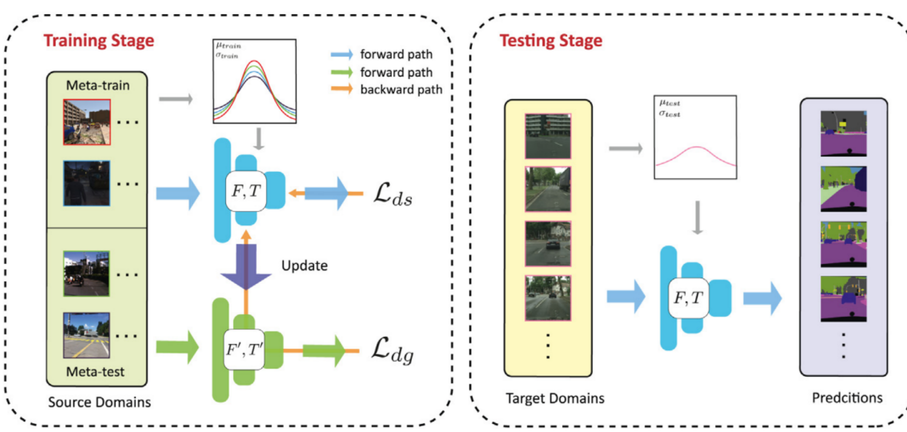

Fig. 2. The framework of our proposed method.

3 Results

This proposed method outperforms largely the existing SOTA. And notably this method is comparable to UDA methods.

4 Ablation Study

1, Authors found the best ratio to divide 4 source domain is 2:2.
2, When queue size in image bank increase, the performance first increase and then decrease, the best choice in this paper is 128.
3, More source domains do improve the performance, and those domain similar to the target domain help even more.
4, A stronger backbone improves the performance.

5 Summary

This paper proposed a novel domain model for generalizable semantic segmentation. Its key ideas are meta-learning, TN, Image bank. This method set the new SOTA in DG.

References