Note: Context Decoupling Augmentation for Weakly Supervised Semantic Segmentation Regularization.

sgc

April 26, 2022

1 Introduction

Merely increasing the same contextual semantic data does not bring much gain to the networks to distinguish the objects, e.g., the correct image-level classification of aeroplane may be not only due to the recognition of the object itself, but also its cooccurrence context like sky, which will cause the model to focus less on the object features.
This paper proposed Context Decoupling Augmentation dubbed CDA to change context in which the objects appear and drive the network to remove the dependence between object instances and contextual information.
Contributions: This paper proposed CDA that does not require additional data and it can remove the correlation between foreground object instances and background context information. CDA can boost the performance of different WSSS methods to the new state-of-the-art by a large margin.

2 Method

2.1 Object Instances Collection

In the first stage, a network is trained to qualify object instances. There are 2 criteria: 1) the current image should only have a single class, 2) $\epsilon_1 < \frac{m}{n} < \epsilon_2$. The reasons are described in the paper.

3 Online Augmentation Training

Blending The instance blended into the image can be rescaled, rotated, and can add Gaussian smoothing. Notably, only instances do not appearing in the image are pasted.

There 3 grades for the quality of augmentation: "perfect", "good" and "noise". However this pasting method is robust. The "Good" pasting can actually help learning more regions of the object and "noise" does not happen much.

Online training In each batch, sample N/2 images from the training dataset and the same number object instances images from the subset which is provided from stage-I. Then randomly pasting the segmented objects into the input images, which creates a N/2 batch new images. Thus, a batch of size N is generated online for each augmentation iteration.

4 Ablation study

1, The best performance to 50.8% mIoU on PASCAL VOC training set is obtained with combining rescale and rotation.
2, Applying pairwise training strategy outperforms the one in single augmented images, which illustrates that this helps the network classifier to learn more discriminative features.
3, over-pasted objects may cover the objects in the original image, making the noise sample dominant. 1 is enough.

Algorithm 1 Stage-II: Online Augmentation.

Input:

The training dataset images $\mathcal{I}$ and the corresponding labels $\mathcal{L}$;

The object instances $\mathcal{O}$ and the corresponding labels $\mathcal{T}$.

while not done **do**
  $(\mathcal{I}_i, \mathcal{L}_i) \leftarrow$ Draw one sample from training dataset;
  $(\mathcal{O}_j, \mathcal{T}_j) \leftarrow$ Draw one sample from object instances subset;
  while $\mathcal{T}_j$ in $\mathcal{L}_i$ **do**
    $(\mathcal{O}_j, \mathcal{T}_j) \leftarrow$ Resample;
  end while
  $\mathcal{I}_i^{'} \leftarrow$ Blend $\mathcal{O}_j$ into $\mathcal{I}_i$;
  $\mathcal{L}_i^{'} \leftarrow$ Append $\mathcal{T}_j$ in $\mathcal{L}_i$;
  Train CAM $\leftarrow$ Loss$(\mathbb{C}(\mathcal{I}_i), \mathcal{L}_i)$ + Loss$(\mathbb{C}(\mathcal{I}_i^{'}), \mathcal{L}_i^{'})$;
end while
Expansion.

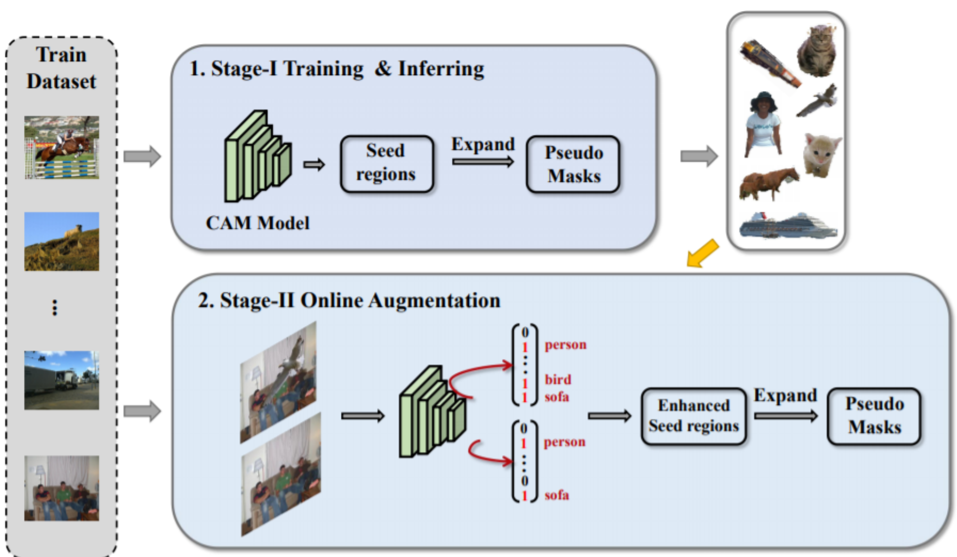

Figure 2. Overview of the proposed augmentation scheme. Stage-I: use the off-the-shelf weakly supervised semantic segmentation methods to obtain some simple object instances with good segmentation. Stage-II: paste the object instances randomly into the raw images to form the new input images, and perform online data augmentation training in a pairwise way with the original input images.

5 Summary

This paper proposed CDA framework, a new augmentation method that does not need extra labels and get rid of contextual information disturbing learning. But I think sometimes contextual information helps so even though this method yield good results, we can not totally deny the role of co-occurring background information.

References