Survey on Segmentation techniques

sgc

April 10, 2022

1 Abstract

As computer vision becoming increasingly widely used in all kinds of fields, segmentation which is a important subject in this field also is a main hotspot draws many scholars' attention. And lots of solution to this problem is popping out.

2 Introduction

Segmentation is a method to extract or separate different part of an image or divide an image into constituent parts. There are 3 main types of segmentation: general segmentation refers to just separate pixels of different objects. Semantic segmentation gives semantic class to each region. Instance segmentation labels each object. And recently a novel and important research topic named panoptic segmentation appeared. It's main task is to give each pixel a semantic label and instance ID. Panoptic segmentation integrates semantic segmentation and instance segmentation in a sense, or it adds process on backgrounds compare with instance segmentation.
Top-level conferences and journals in Semantic segmentation or CV CVPR, ICCV, ECCV, Neurips, ICML, ICLR, AAAI, IJCAI.
Mostly used data set: Pascal VOC 2012consists of 20 kinds of objects including humans, mobiles, and others, can be used in segmentation.
Cityscapescity scene pictures of 50 cities.
Pascal Context400 indoors or outdoors pictures.
Stanford Background DatasetA set of outdoor scenes with at least one foreground object.
The standard index used to evaluate the performance of semantic segmentation model is the average IOU (intersection over union). IOU is defined as follows:

$$IOUs = \frac{Areaofoverlap}{AreaofUnion} = \frac{A_{pred} \cap A_{true}}{A_{pred} \cup A_{true}}$$

It can judge the accuracy of the model.

3 Methods of segmentaion

A general semantic segmentation architecture can be widely considered as an encoder network, followed by a decoder network: the task of the decoder is to project the recognition feature semantics learned by the encoder onto the pixel space to obtain dense classification.

Three main methods:

1-region based semantic segmentation:

The region based method first extracts and describes the free-form region from the image, and then classifies it based on the region. During testing, region based prediction is converted to pixel prediction, usually by marking pixels according to the highest scoring region containing the prediction.

2-full convolution network:

The original complete convolution network (FCN) learns the mapping from pixel to pixel without extracting region recommendations. FCN network pipeline is an extension of classic CNN. The main idea is to make the classic CNN take images of any size as input. CNN only accepts and produces labels with specific size input. The restriction comes from the fully connected layer. In contrast, FCN has only convolution layer and pooling layer, which can take the input of any size.

3-weakly supervised semantic segmentation:

Most related methods in semantic segmentation rely on a large number of images with pixel level segmentation masks. However, manually annotating these masks is quite time-consuming, frustrating and commercial cost. Therefore, some weakly supervised methods have been proposed recently, which are committed to semantic segmentation by using annotated bounding boxes.

Recently, more and more method based on Transformer merged and have good performance when applied to semantic segmentation.

The most important method ViT [3] proposed to implement transformer to CV and got great success. And many method find self-supervision(DINO[4]) and weak-supervision(Examples:[5]) help to improve performance and lessen dependence on abundant human annotation or pixel level labels. MCTformer[6] improved ViT by adding more Class tokens to produce class-specific object localization maps as a kind of pseudo label to help do semantic segmentation.

3.1 Recent Works

RCA[9] is designed to take use of Inter-image information with a memory bank to store object pattern appearing in training data.
Sparsely annotation semantic segmentation(SASS) aim to train a segmentation networks from partly labeled pixels. TEL[10] provide Semantic Affinity of low-level and high-level for labeling and it is effective and easy to be incorporated into existing frameworks by combining it with a traditional segmentation loss.
A recent work[11] proposed pixel-to-prototype contrastive learning method to WSSS. This method uses 2 networks and extract Intra-View and Cross-View loss from their prototypes and values.
There are usually unreliable pixels remain unused, so U^2PL[13] is designed to keep a memory bank to store unreliable pixels as negative samples to generate contrastive loss.
The first method to implement Transformer to WSSS is AFA[13]. AFA using affinity information from MHSA to refine pseudo labels generation. AFA is complemented with PAR which can refine labels considering information in its neighbor field.

4 Related Works in This Field

There are lots of detail problems can be handled and recently some excellent works revealed some intrinsic shortage of some existing methods; For conventional knowledge distillation, DKD[7] divided KL Loss into 2 parts (TCKD and NCKD) and found that NCKD is depressed and solve this by decouple $(1 - p_t)$ and NCKD.
As Siamese Network being widely used in SSL, people find that using random crop has 2 major problems: 1, It may generate bad pairs containing useless images(background); 2, It may generate similar pairs. So Contrastive crop [8] is designed to deal with that by implementing Semantic-aware localization and Center-suppressing sampling.

References

[1] Jonathan Long, Evan Shelhamer, Trevor Darrel (2016)"Fully Convolutional Networks for Semantic Segmentation". In CVPR 2015

[2] Fabian Isensee, Paul F.Jaeger, Simon A. A. Kohl, Jens Petersen and Klaus H. Maier-Hein "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In Nature Method 2021

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai "An image is worth 16X16 words: transformers for image recognition at scale". In ICLR 2021.

[4] Mathilde Caron, Hugo Touvron, Ishan Misra, Herve Jegou, Julien Mairal, Piotr Bojanowsk, Armand Joulin "Emerging Properties in Self-Supervised Vision Transformers". In ICCV 2021.

[5] Jiarui Xu Shalini De Mello, Sifei Liu, Wonmin Byeon, Thomas Breuel, Jan Kautz, Xiaolong Wang "GroupViT: Semantic Segmentation Emerges from Text Supervision". In arXiv 2022.

[6] Lian Xu, Wanli Ouyang, Mohammed Bennamoun, Farid Boussaid, and Dan Xu "Multi-class Token Transformer for Weakly Supervised Semantic Segmentation". In arXiv 2022.

[7] Borui Zhao, Quan Cui, Renjie Song, Yiyu Qiu, Jiajun Liang " Decoupled Knowledge Distillation " In CVPR 2022.

[8] Xiangyu Peng, Kai Wang, Zheng Zhu, Mang Wang, Yang You "Crafting Better Contrastive Views for Siamese Representation Learning" In CVPR 2022.

[9] Tianfei Zhou1, Meijie Zhang2, Fang Zhao, Jianwu Li2 "Regional Semantic Contrast and Aggregation for Weakly Supervised Semantic Segmentation" In CVPR 2022.

[10] Zhiyuan Liang, Tiancai Wang, Xiangyu Zhang, Jian Sun, Jianbing Shen "Tree Energy Loss: Towards Sparsely Annotated Semantic Segmentation" In CVPR 2022.

[11] Ye Du, Zehua Fu, Qingjie Liu, Yunhong Wang"Weakly Supervised Semantic Segmentation by Pixel-to-Prototype Contrast" In CVPR 2022.

[12] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, Xinyi Le1 "Semi-Supervised Semantic Segmentation Using Unreliable Pseudo-Labels" In CVPR 2022.

[13] Lixiang Ru, Yibing Zhan, Baosheng Yu, Bo Du "Learning Affinity from Attention: End-to-End Weakly-Supervised Semantic Segmentation with Transformers" In CVPR 2022.