Survey on Segmentation techniques

sgc

April 3, 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 add more Class tokens to produce class-specific object localization maps as a kind of pseudo label to help do semantic segmentation.

4 Paper

[1],Jonathan Long, Evan Shelhamer, Trevor Darrel (2016)"Fully Convolutional Networks for Semantic Segmentation":A pioneering work in semantic segmentation. Use fully convolutional network and upsampling to use any size of input and get classification image.
[2],Fabian Isensee, Paul F.Jaeger, Simon A. A. Kohl, Jens Petersen and Klaus H. Maier-Hein (2021) "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation":nnU-Net("no new net") underscores the relative importance of method configuration over architectural variations. The strong performance of nnU-Net is not achieved by a new network architecture, loss function or training scheme, but by systematizing the complex process of method configuration.
[3], Alexey Dosovitskiy;y, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai(2021) "AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE"
[4], Mathilde Caron, Hugo Touvron, Ishan Misra, Herve Jegou ,Julien Mairal, Piotr Bojanowsk,i Armand Joulin (2021)"Emerging Properties in Self-Supervised Vision Transformers" [5], Jiarui Xu1* Shalini De Mello, Sifei Liu, Wonmin Byeon, Thomas Breuel, Jan Kautz, Xiaolong Wang (2022)"GroupViT: Semantic Segmentation Emerges from Text Supervision" [6], Lian Xu, Wanli Ouyang, Mohammed Bennamoun, Farid Boussaid, and Dan Xu (2022) "Multi-class Token Transformer for Weakly Supervised Semantic Segmentation"