



## 电子回条

此回条认定 Turnitin 已收到您的文稿。您能在下方看到有关您的提交物件的回条资讯。

您的提交物件的第一页在下方显示。

提交物件作者: C42d2258m8 SID  
作业标题: paper  
提交物件标题: Research on Panoptic Segmentation Algorithm Based on Tra...  
文档名称: arch\_on\_Panoptic\_Segmentation\_Algorithm\_Based\_on\_Trans...  
文档大小: 931.79K  
页面总数: 13  
文字总数: 5,610  
字符总数: 28,978  
提交日期: 2023年12月27日 01:41上午 (UTC-0800)  
提交作业代码: 2265065293

### Research on Panoptic Segmentation Algorithm Based on Transformer

#### Abstract

Panoptic segmentation is a kind of image segmentation task that integrates image semantic segmentation and instance segmentation, which not only focuses on the stuff of the background such as sky, road, etc., but also focuses on the things of different individuals such as people, car, etc. In order to learn more comprehensive image features, the FPN-style mask head in Mask R-CNN is added on top of the End-to-end Detection Transformer, more comprehensive image features, a new FPN-style mask head in Mask R-CNN is added on top of End-to-end Detection Transformer, which converts the output features of Transformer Decoder into instance segmentation map. Finally, the PQ in COCO 2017 Validation is 42.8 for PQ of stuff and 35.8 for PQ of things, which is in the high level compared with the same period.

#### I. Introduction

With the development of computer hardware and Internet technology, not only are computers able to handle more complex tasks, but also major manufacturers and companies are able to collect more types and amounts of data, sometimes text, sometimes speech, and sometimes images. When more of this data is available, the need for more complex computer tasks gradually arises, and this is the case with image recognition. Early image recognition was to read the image as a matrix, where each pixel point was treated as a feature, so the number of pixel points was the number of features. Limited by the hardware of the device, early image recognition can only accomplish the task of classifying image features on pictures with particularly small pixels, and most algorithms are based on traditional statistics and probability theory reasoning, without substantial breakthroughs.

With the emergence of CNNs such as LeNet-5 [1], AlexNet [2], VGG [3], and ResNet [4] at the beginning of the 21st century, the field of image recognition is no longer limited to categorizing images, and more complex tasks such as object detection and image segmentation have been proposed. Earlier in image segmentation task is to classify images at pixel level, J. Long, et al. [5] proposed Full Convolutional Neural Network FCN in 2015, which is the first deep learning based semantic segmentation model. The model can be divided into the structure of encoder-decoder, the encoder, i.e., the structure of CNN is used to extract the feature map; the decoder consists of multiple up-sampling layers, which are up-sampled by the feature map to the original image size of the segmentation map, so the subsequently proposed U-Net [6], V-Net [7] and other deep learning models for semantic segmentation are designed based on the encoder-decoder structure. With the development of image semantic segmentation, scholars believe that the sky, roads and pedestrians, vehicles can't be confused, B. Hariharan, et al. [8] that there should exist a method to be able to segment the target object Things for pedestrians, vehicles, that is, instance segmentation in the middle of the development of image segmentation. Instance segmentation is different from semantic segmentation, which is to form a segmentation map of target objects on the detected target, which can be understood as a two-stage task of semantic segmentation of objects of object frames after generating object prediction frames from object detection, so scholars add FCN decoder similar structure on the object detection model in order to generate target segmentation map. Until 2017 K. He, et al. [9] proposed Mask R-CNN more flexible and simple two-stage instance segmentation model based on R-CNN framework. The model achieves better experimental results on MS COCO [10] dataset, adopts ResNet-FPN structure, i.e., multi-layer feature