Joint COCO and Mapillary Workshop at ICCV 2019: Detection Challenge Track

Technical Report: Deformable convolution hybrid task cascade semantic segmentation

Wang jian
Jilin University
No. 5372, Nanhu Road,
Changchun City

Wu jinyang
Jilin University
No. 5372, Nanhu Road,
Changchun City

Abstract

The invention proposes an embedding balanced transformable convolution hybrid task cascaded semantic segmentation algorithm. The Mask RCNN[1] model proposed in 2017 is simple and flexible. It is a general object instance segmentation framework. The Cascade RCNN[2] proposed in 2018 is also a classic architecture, and it is superior to Mask RCNN and can be improved. Performance of various tasks. Based on the former two, we propose a new framework, namely the embedded balanced deformable convolution hybrid task cascading framework, which is a new cascade architecture for instance segmentation. The innovations of the architecture are as follows: 1 The overall structure of the model is to cascade the Mask RCNN and the Cascade RCNN, improve the information flow by combining cascading and multi-task parallelism in each stage, and use the spatial background to further improve the accuracy. In the cascading mode, an improved progressive cascade pipeline is designed. At each stage, the bounding box regression and mask prediction are combined in a multitasking manner. 3 In the backbone network that extracts features, using Deformable Convolutional Networks for sampling, we introduce two new modules to enhance CNN's transformation modeling capabilities, ie, anamorphic convolution and deformable interest. The Region of Interest (ROI) pooling, the new module can easily replace the common modules in the existing CNN, and can easily end-to-end training through standard backpropagation, resulting in a deformable convolution network. . In addition, the framework integrates three novel components: IoU balanced sampling, balanced feature pyramids, and balanced L1 norm, which are used to reduce imbalances in samples, extracted features, and target detection, respectively, in an overall balanced design. Benefit. Taking the COCO data set as the training set and test set, the experimental results show that this framework significantly improves the accuracy of segmentation.

1. Content Recommendations

The coding end of the model builds a series of convolutional layers and maximum pooling layers based on the Cascade RCNN model for extracting features, and doubles the number of channels of the convolutional feature map, while halving the length and width of the maximum pooled feature map. We introduced two new modules to enhance CNN's transformation modeling capabilities, namely morphing convolution and deformable ROI. Both are based on the following ideas: increasing the spatial sampling position in the module and additional offsets, and learning the offset of the target task without additional supervision. The new module is easy to replace the common module

in the existing CNN, and can be end-to-end training through standard backpropagation, resulting in a deformable convolution network. Therefore, applying a regular convolution on a deformable feature image can more effectively reflect complex structures.The invention proposes an embedding balanced transformable convolution hybrid task cascaded semantic segmentation algorithm. The Mask RCNN[1] model proposed in 2017 is simple and flexible. It is a general object instance segmentation framework. The Cascade RCNN[2] proposed in 2018 is also a classic architecture, and it is superior to Mask RCNN and can be improved. Performance of various tasks. Based on the former two, we propose a new framework, namely the embedded balanced deformable convolution hybrid task cascading framework, which is a new cascade architecture for instance segmentation. The innovations of the architecture are as follows: 1 The overall structure of the model is to cascade the Mask RCNN and the Cascade RCNN, improve the information flow by combining cascading and multi-task parallelism in each stage, and use the spatial background to further improve the accuracy. In the cascading mode, an improved progressive cascade pipeline is designed. At each stage, the bounding box regression and mask prediction are combined in a multitasking manner. 3 In the backbone network that extracts features, using Deformable Convolutional Networks for sampling, we introduce two new modules to enhance CNN's transformation modeling capabilities, ie, anamorphic convolution and deformable interest. The Region of Interest (ROI) pooling, the new module can easily replace the common modules in the existing CNN, and can easily end-to-end training through standard backpropagation, resulting in a deformable convolution network. . In addition, the framework integrates three novel components: IoU balanced sampling, balanced feature pyramids, and balanced L1 norm, which are used to reduce imbalances in samples, extracted features, and target detection, respectively, in an overall balanced design. Benefit. Taking the COCO data set as the training set and test set, the experimental results show that this framework significantly improves the accuracy of segmentation.

References

[1]Kaiming H , Georgia G , Piotr D , et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018:1-1.
[2]Cai Z , Vasconcelos N . Cascade R-CNN: Delving into High Quality Object Detection[J]. 2017.
[3]Long J , Shelhamer E , Darrell T . Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.
［4］Ghiasi G , Fowlkes C C . Occlusion Coherence: Localizing Occluded Faces with a Hierarchical Deformable Part Model[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014.
［5］Qing L , Shao-Zi L I , Duan-Sheng C . Using Poselet and Scene Information for Action Recognition in Still Images[J]. Journal of Chinese Computer Systems, 2015.